BabelNet and Friends

Roberto Navigli

Dipartimento
di Informatica

Sapienza
Università di Roma

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

The Luxembourg BabelNet Workshop – Session 1

The Luxembourg BabelNet Workshop at a Glance
2 March, 2016 (today)

9:15-10:45 Session 1 - BabelNet and Friends

(**Roberto Navigli**, Sapienza Univ. of Rome)

10:45-11:15 *Break*

11:15-12:30 Session 2 - Tech Session: Downloading and installing BabelNet. The BabelNet API.

(**Claudio Delli Bovi**, Sapienza Univ. of Rome)

12:30-12:45 *Family picture*

12:45-13:45 *Lunch break*

13:45-15:15 Session 3 - Tech Session: The Babelfy API. Disambiguating text with Babelfy

(**Claudio Delli Bovi**, Sapienza Univ. of Rome)

Lightning talks, part I

The Luxembourg BabelNet Workshop at a Glance
2 March, 2016 (today)

15:15-15:45 *Break*

15:45-17:00 Session 4 - BabelNet on the Linguistic Linked Open Data Cloud; Industrial applications.

(**Roberto Navigli**, Sapienza Univ. of Rome)

Lightning talks, part II

17:15 *Wanderers of the fortifications*

(weather permitting…)

Leave for the restaurant

19:00 Dinner (*) - Porta Nova Restaurant

(*) not offered by the organization

The Luxembourg BabelNet Workshop at a Glance
3 March, 2016 (tomorrow)

9:00-10:30 Session 5 - Invited talk

(**Asunción Gómez-Pérez**, UPM)

Lightning talks, part III

10:30-11:00 *Break*

11:00-12:15 Session 6 - Linking glossaries and thesauri to BabelNet. Case study 0: Hello world (a minimal fake resource)

(**Roberto Navigli**, Sapienza Univ. of Rome)

12:15-13:15 *Lunch break*

The Luxembourg BabelNet Workshop at a Glance
3 March, 2016 (tomorrow)

13:15-14:45 Session 7 - Linking IATE, EUROVOC, Euramis (1/2)

(**Roberto Navigli**, Sapienza Univ. of Rome)

14:45-15:15 *Break*

15:15-16:45 Session 8 - Linking IATE, EUROVOC, Euramis (2/2)

Introduction to the BabelNet ANnotation Group

(**Roberto Navigli**, Sapienza Univ. of Rome)

Lightning talks dedicated to EU themes:

- GlossSearch (**Rodolfo Maslias**, EU Parliament)
- Multilingual Legislative Process (**M.T. Carrasco Benitez**, EU Commission)

Town Hall Meeting: participants ask questions

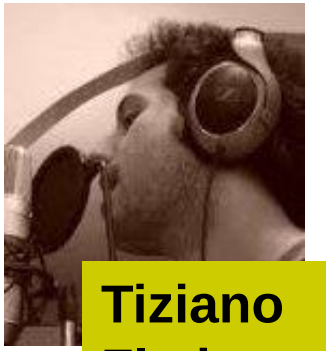
A big THANKS!

A clear example that **joining forces pays off**!!!

The
BabelNet
Workshop
2-3 March 2016

Who is Roberto Navigli?

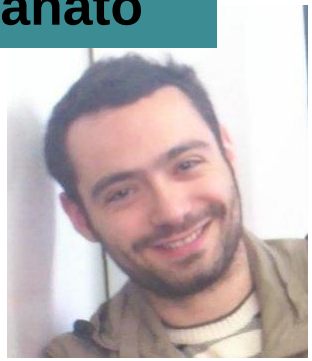
- Associate Professor in the Department of Computer Science (Sapienza University of Rome)
- Home page: http://wwwusers.di.uniroma1.it/~navigli
- Email: navigli@di.uniroma1.it
- Twitter: @RNavigli
- Principal investigator of some (relevant) projects:
  - PI of an ERC Starting Grant (MultiJEDI, 1.3M euros) on **BabelNet** and **multilingual joint disambiguation** and **entity linking**
  - Sapienza PI of an FP7 CSA (LIDER, 1.5M euros) on **Linguistic Linked Open Data**
  - Co-PI of a Google Focused Research Award on **Information Extraction with structured knowledge**

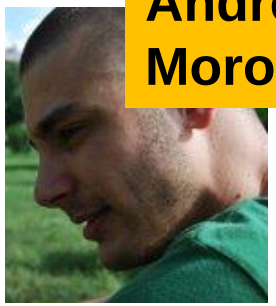


Tiziano Flati

Daniele Vannella

Alessandro Raganato

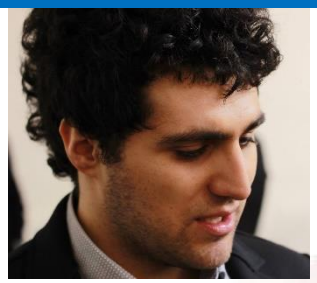




Andrea Moro

Simone Ponzetto

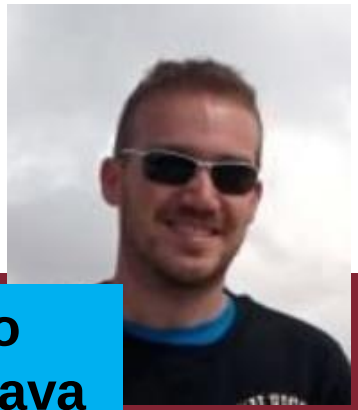
Claudio Delli Bovi



Taher Pilehvar

Francesco Cecconi

Ignacio Iacobacci



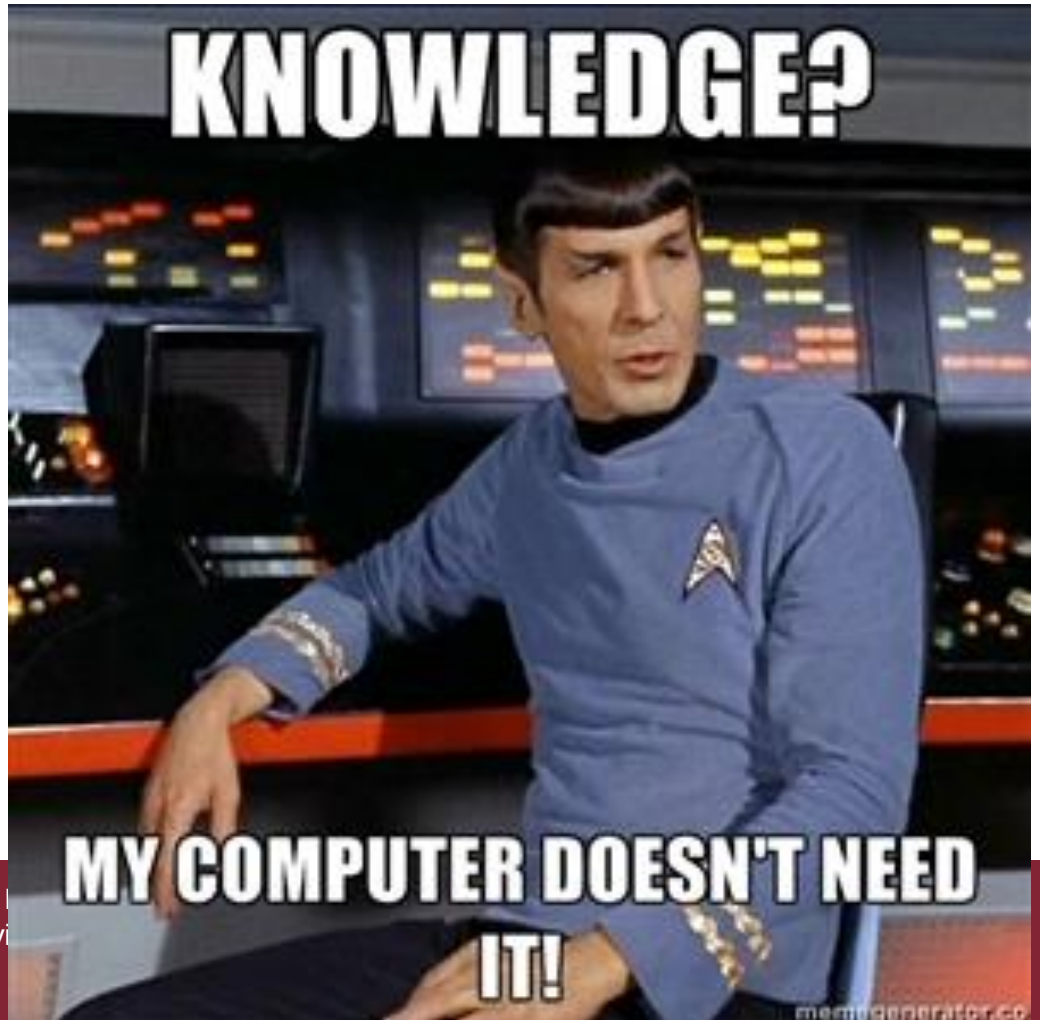
Federico Scozzafava

José Camacho Collados

knowledge

It’s all about knowledge!

- But can we expect computers to know?
- Can’t computers just use, e.g., statistical techniques?



State-of-the-art Machine Translation

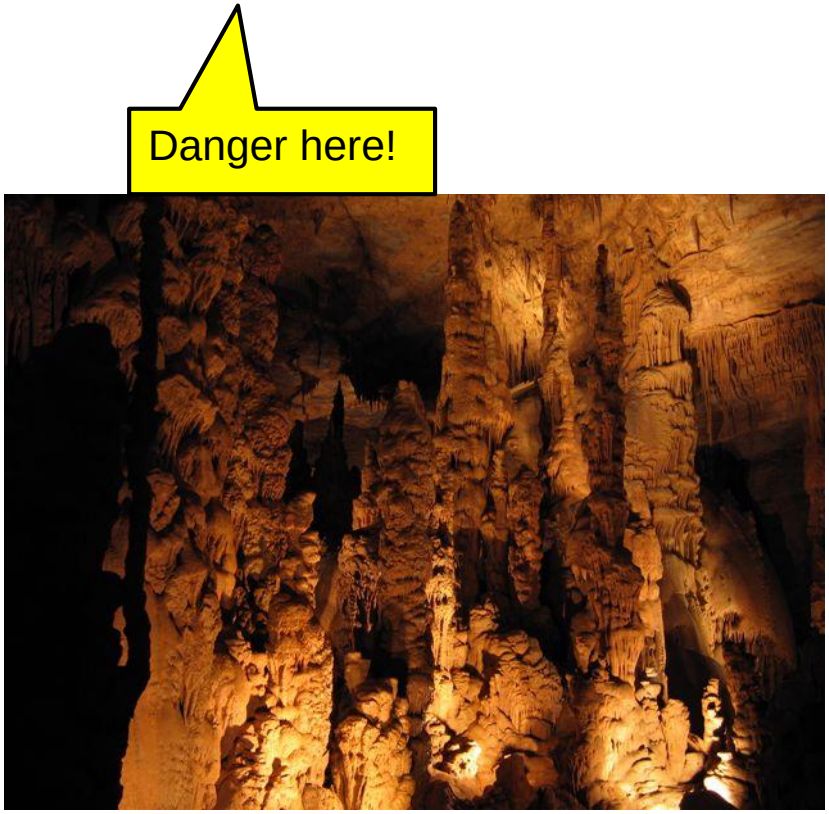
- EN: These are movies in which the music genre, e.g. **rock**, is an important element but not necessarily central to the plot. Examples are Easy Rider (1969), The Graduate (1969), and Saturday Night Fever (1978).

State-of-the-art Machine Translation

- EN: These are movies in which the music genre, e.g. **rock**, is an important element but not necessarily central to the plot. Examples are Easy Rider (1969), The Graduate (1969), and Saturday Night Fever (1978).
- ES: Estas son las películas en las que el género de la música, por ejemplo, **roca**, es un elemento importante, pero no necesariamente el centro de la trama. [...]

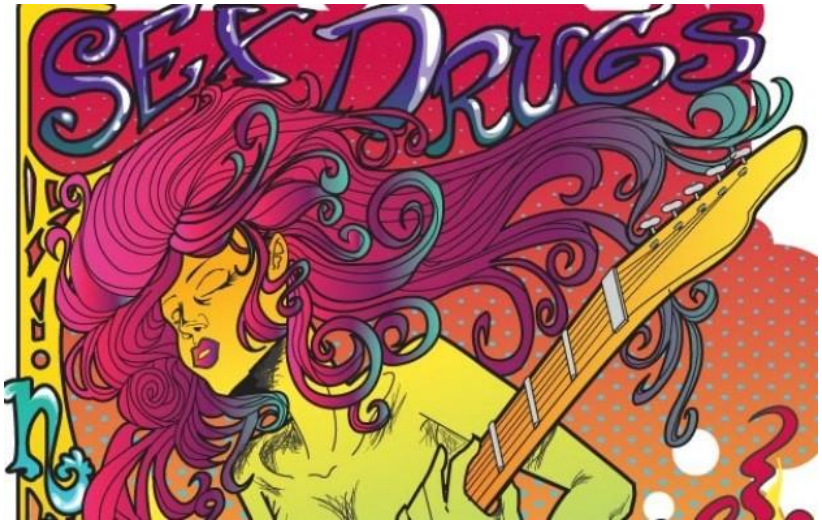
State-of-the-art Machine Translation

- EN: We can look at how this vast slug of molten underground **rock** was injected.



State-of-the-art Machine Translation

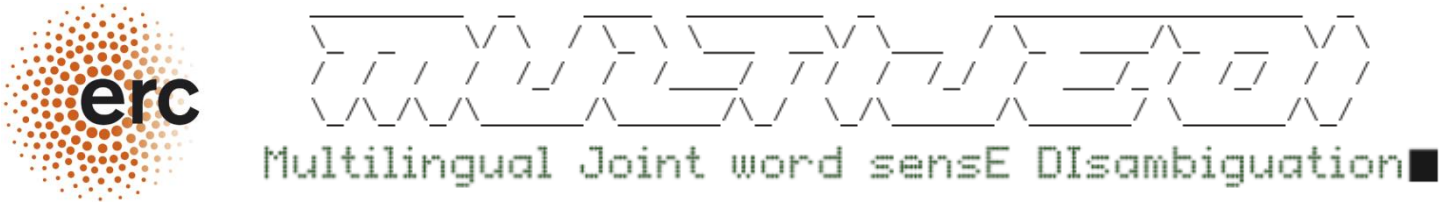
- EN: We can look at how this vast slug of molten underground **rock** was injected.
- FR: Nous pouvons voir comment ce vaste bouchon de **rock** underground fondu a été injecté.
- IT: Possiamo guardare a come è stato iniettato questo vasto slug del **rock** underground fusa.



What we say to dogs and what they hear

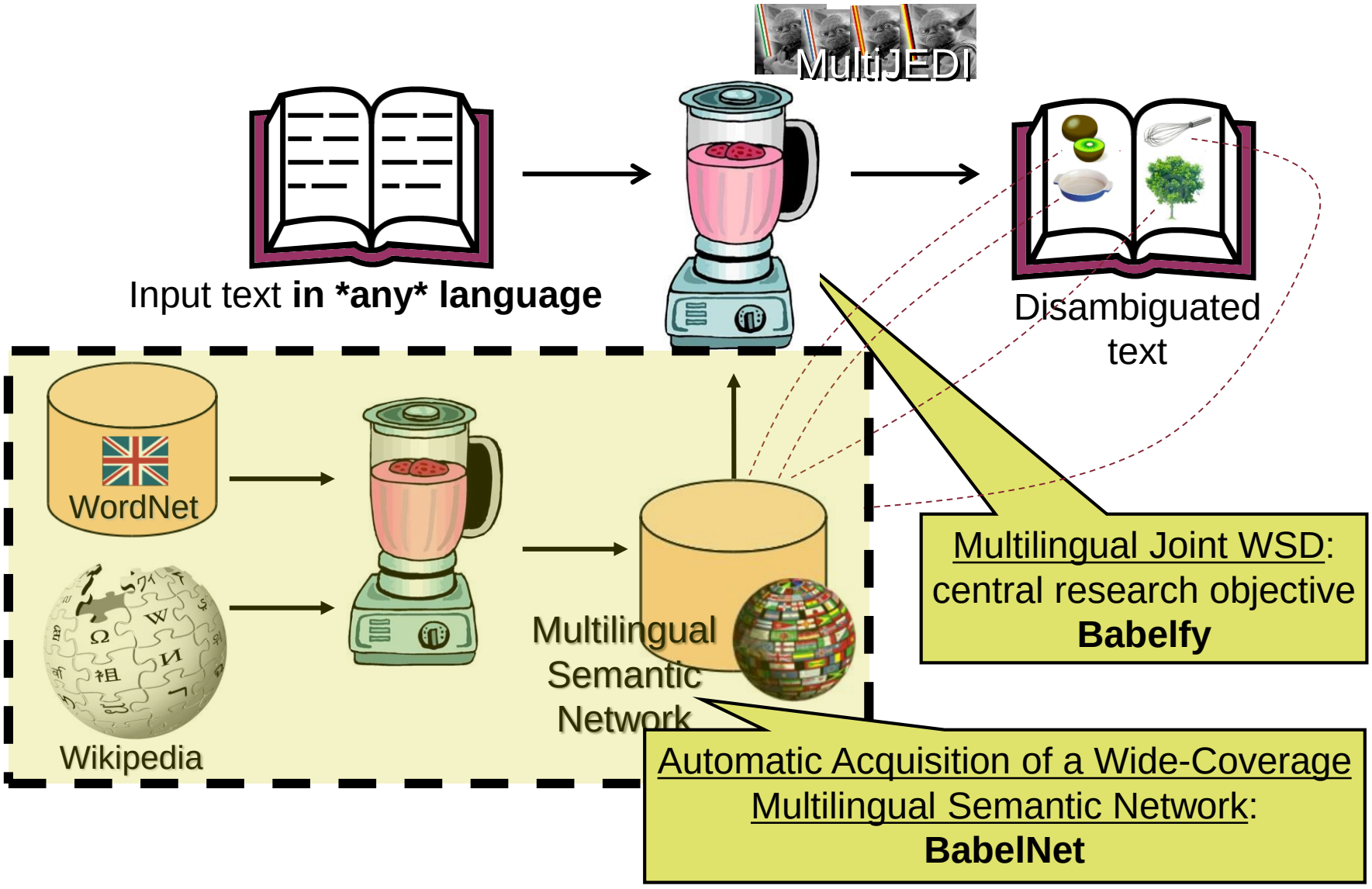
Already 12% accuracy!

What are we talking about?



- A 5-year ERC Starting Grant (2011-2016) on Multilingual Word Sense Disambiguation
- Funded by the **European Research Council** with 1.3M euros

MultiJEDI: The Vision



INTEGRATING KNOWLEDGE

[Navigli & Ponzetto, ACL 2010; Pilehvar & Navigli, ACL 2014]

The resource diaspora

InterActive Terminology for Europe

English (en)

Search Screen Help

EuroVoc il thesaurus multilingue dell'Unione europea

Europa > Pagina iniziale EuroVoc > Settori & MT > institutional agreement

Lingua: (en) English

Ricerca

- Ricerca avanzata

Consultazione

- Consultare la versione per argo
- Per s
- Versi perm
- Elen
- Indice alfabetico
- EuroVoc SKOS/RDF
- EuroVoc XML

Proposte

- Contributi
- Nuovi concetti approvati

institutional agreement

RDF/XML TUR

04 POLITICS

MT 0431 politics and public safety

BT1 institutional activity

URI *http://eurovoc.europa.eu/7184*

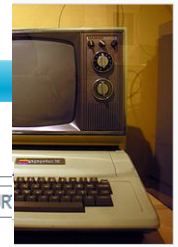
EQUIVALENTI IN ALTRE LINGUE

BG	институционално споразумение
ES	acuerdo institucional
CS	institucionální smlouva
DA	institutionel aftale
DE	institutionelles Abkommen
ET	institutsionaalne kokkulepe
EL	θεσμική συμφωνία
EN	**institutional agreement**
FR	accord institutionnel
HR	institucionalni sporazum
IT	accordo istituzionale
LV	iestāžu nolīgums
LT	institucinis susitarimas
HU	intézményi megállapodás
MT	ftehim istituzzjonali
NL	institutioneel akkoord
PL	porozumienie instytucjonalne
PT	acordo institucional
RO	acord instituțional
SK	inštitucionálna dohoda
SL	institucionalni sporazum
FI	toimielinten sopimus
SV	institutionellt avtal

Euramis DGT Net

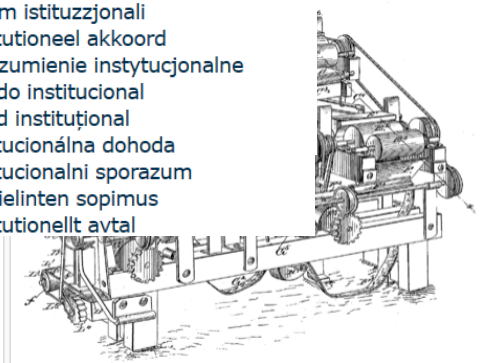
Wikimedia Commons has related media at: *computer*

an article on:



ter (circa early

etrieve large Internet, or



Contents [hide]

1 Etymology
2 History
3 Types

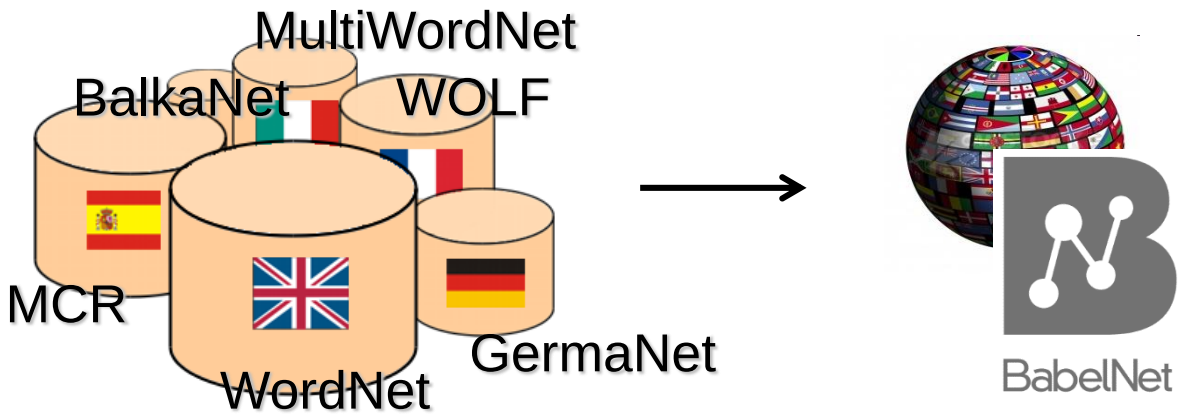
Main page
Contents
Featured c
Current eve
Random ar
Donate to
Wikipedia s

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here

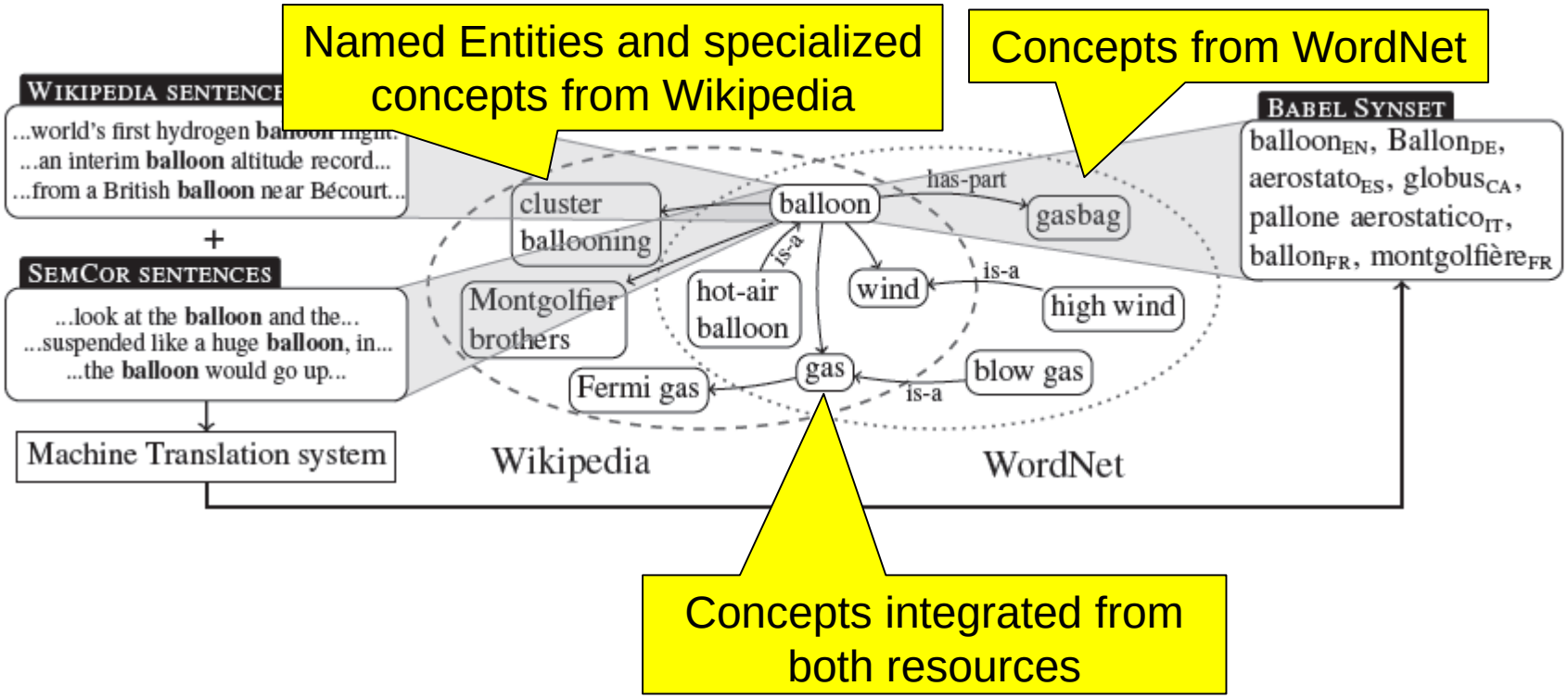
Multilingual Joint Word Sense Disambiguation (MultiJEDI)

Key Objective 1: create **knowledge** for **all languages**



It all started with merging WordNet and Wikipedia [Navigli and Ponzetto, ACL 2010; AIJ 2012]

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



The kind of knowledge we can find in…

WordNet

textile, cloth, fabric

Artifact made by weaving or felting or knitting or crocheting natural or synthetic fibers

3

fabric, framework

The underlying structure

europium, atomic number 63, Eu

A bivalent and trivalent metallic element of the rare earth group

European Union, European Economic Community, Common Market, EU

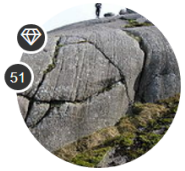
An international organization of European countries formed after World War II to reduce trade barriers and increase cooperation among its members

Wikipedia

Fabric (club)

Fabric is a nightclub in London, United Kingdom.

ID: 01738336n | Named Entity



Fabric (geology)

In geology, a rock's fabric describes the spatial and geometric configuration of all the elements that make it up.

ID: 00912623n | Concept

Eu, Seine-Maritime, Eu

Eu is a commune in the Seine-Maritime department in the Upper Normandy region in northern France.

38

Classifying space for U(n), EU(1)

In mathematics, the classifying space for the unitary group U is a space BU together with a universal bundle EU such that any hermitian bundle on a paracompact space X is the pull-back of EU by a map X → BU unique up to

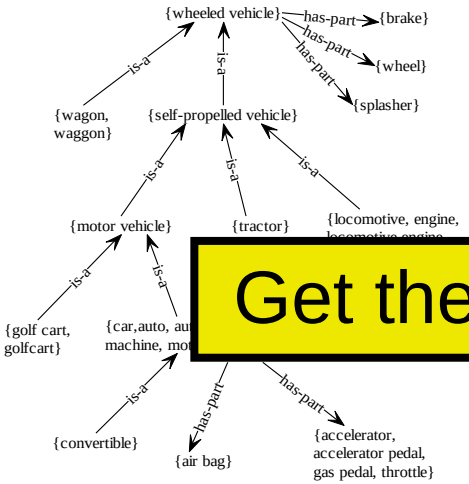
13

EU (group)

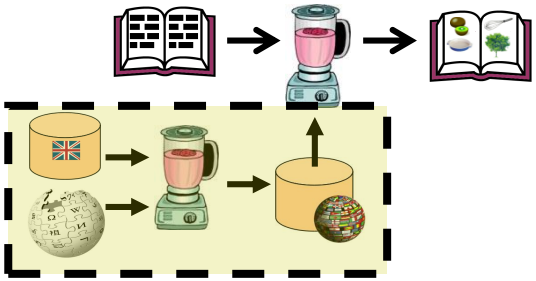
EU or Ёлочные игрушки is a Russian electronic music group, generally considered to be part of the IDM genre, formed in 1997 in Saint Petersburg, and consisting of Sasha Zaitsev and Ilya Baramiya.

Creating a Multilingual Semantic Network

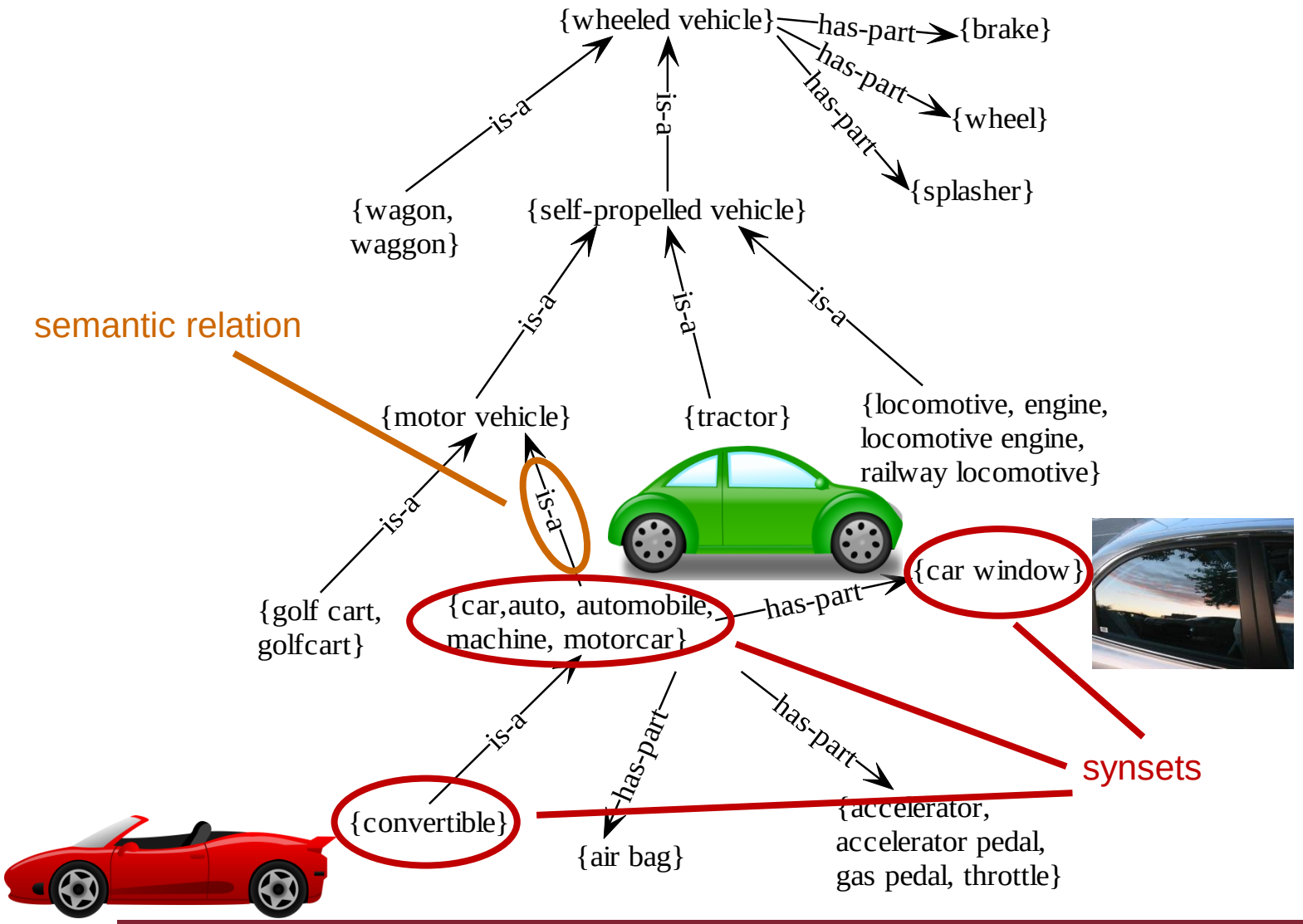
- Start from two large **complementary** resources:
  - WordNet: full-fledged taxonomy
  - Wikipedia: multilingual and continuously updated



Get the best from both worlds



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



Wikipedia [The Web Community, 2001-today]

(unspecified) semantic relation

concepts

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]

Automobile

Benz "Velo" model (1894) – entered into an early automobile race as a motocycle[1][2]

Classification	Vehicle
Industry	Various

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.

Look up *passenger* in Wiktionary, the free dictionary.

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 riding in a company car being driven by another person would be a *passenger*, even if the car was being driven on company

[hide]

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]

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

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

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 [hide]

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

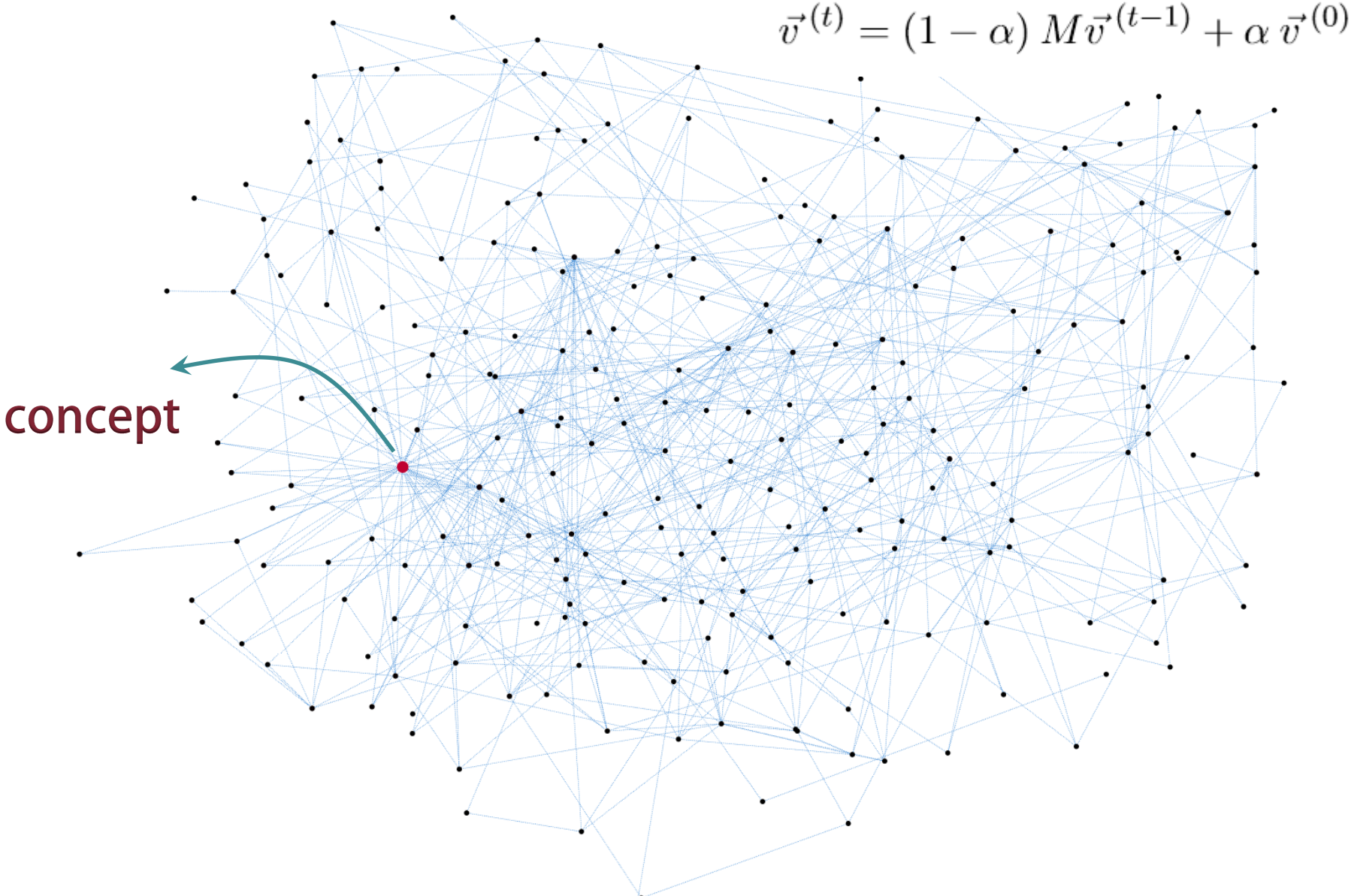
A statue dedicated to the traveler in Oviedo, Spain.

Etymology

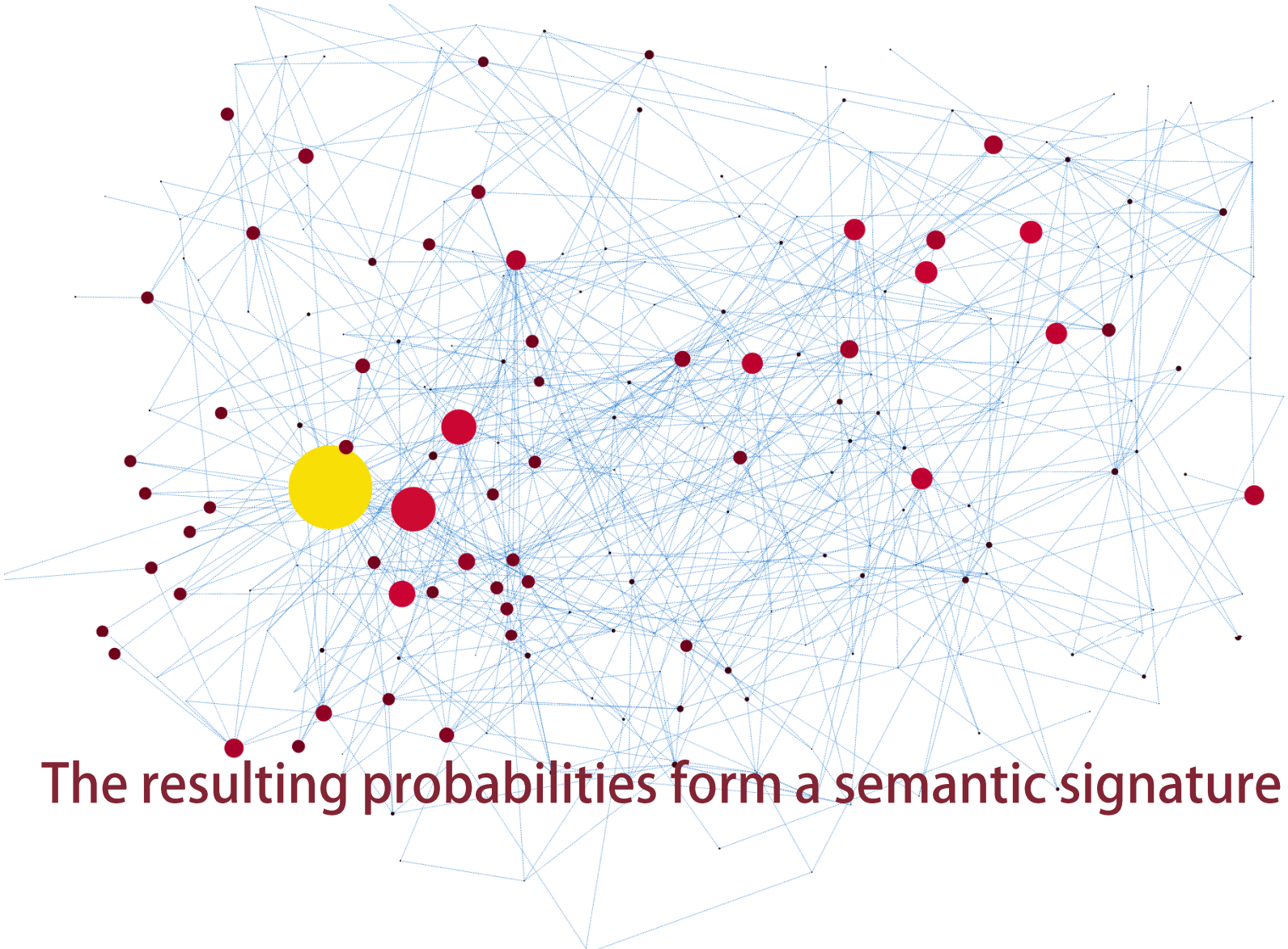
The term "travel" originates from the Old French word *travail*.[3] The term also covers all the activities performed during a travel (movement).[4]

Structural Similarity with Personalized PageRank [Pilehvar and Navigli, ACL 2014]

$$\vec{v}^{(t)} = (1-\alpha)\, M\vec{v}^{(t-1)} + \alpha\, \vec{v}^{(0)}$$



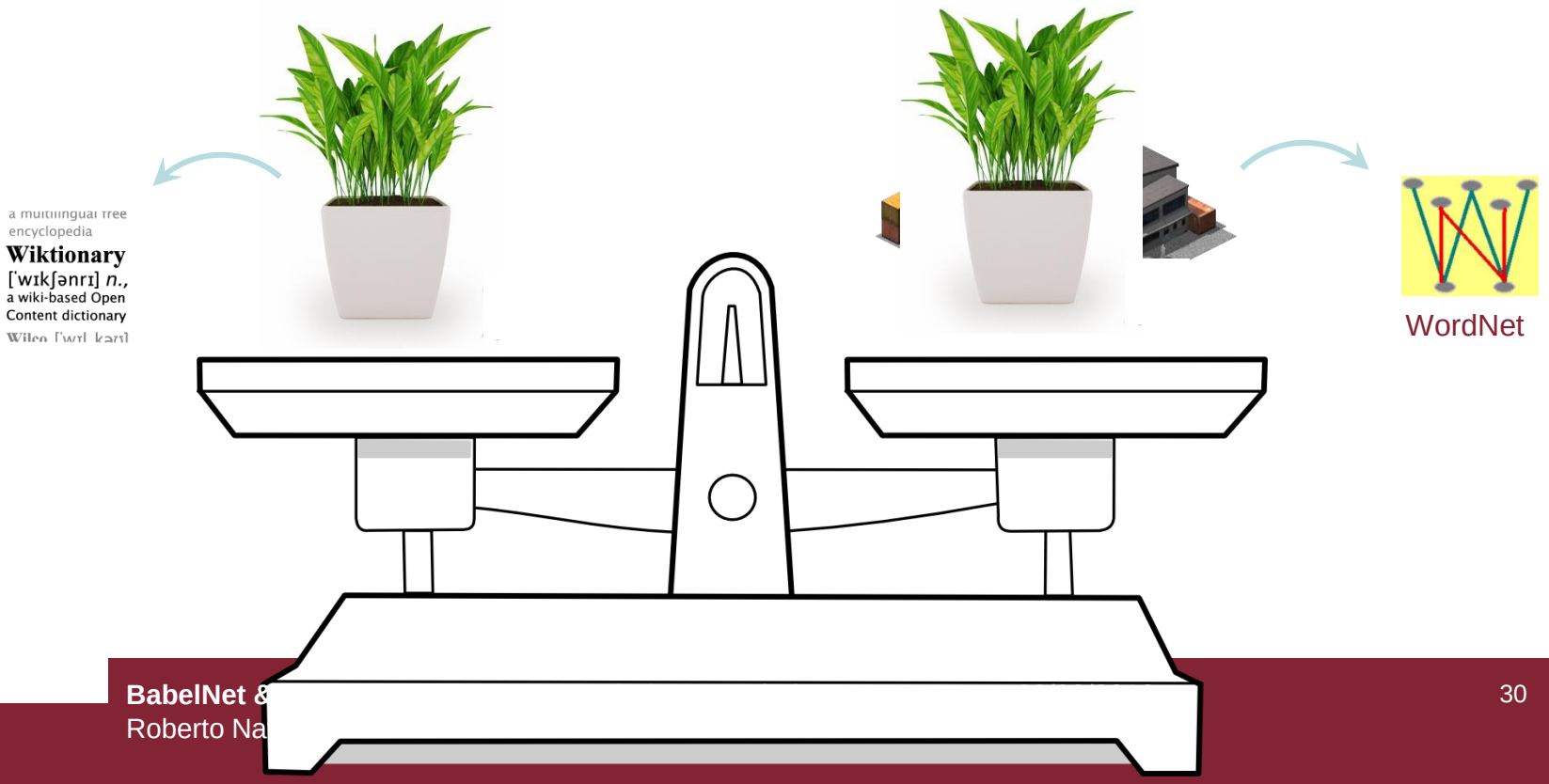
Structural Similarity with Personalized PageRank [Pilehvar and Navigli, ACL 2014]



The resulting probabilities form a semantic signature

To merge or not to merge? [Pilehvar and Navigli, ACL 2014]

- Measure the similarity of senses of the same word (but from different resources)
- If they are similar enough, merge the corresponding two concepts



Comparing Semantic Signatures
Weighted Overlap [Pilehvar et al., ACL 2013]

- We calculate the following formula:

$$Sim(\mathcal{S}_{\mathbf{v}_1}, \mathcal{S}_{\mathbf{v}_2}) = \frac{\sum_{i=1}^{|T|}(r_i^1 + r_i^2)^{-1}}{\sum_{i=1}^{|T|}(2i)^{-1}}$$

- where r_i^k is the ranking of the i-th element in vector k

Merging entries from different resources into BabelNet

- W… fini… n… m…

a multilingual free encyclopedia
Wiktionary
['wɪkʃənrɪ] *n.*,
a wiki-based Open Content dictionary
Wikt ['wɪ kɜrɪ]

Old English *plante* ("young tree or shrub, herb newly planted"), fom Latin *planta* ("sprout, shoot, cutting"). Broader sense of "any vegetable life, vegetation generally" is from French *plante*.

The verb is from Middle English *planten*, from Old English *plantian* ("to plant"), from Latin *plantare*, later influenced by Old French *planter*. Compare also Dutch *planten* ("to plant"), German *pflanzen* ("to plant"), Swedish *planta* ("to plant"), Icelandic *planta* ("to plant").

Pronunciation [edit]

- (*New Zealand, Received Pronunciation*) IPA(key): /plɑːnt/
- (*Australia, US, Canada, Northern England*) IPA(key): /plænt/
- Audio (US) 0:00 MENU
- Rhymes: -ɑːnt

Noun [edit]

plant (*plural* **plants**)

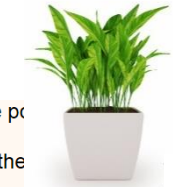
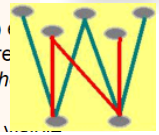
1. An organism that is not an animal, especially an organism capable of photosynthesis. Typically a small or herbaceous organism of this kind, rather than a tree. [quotations ▾]
   *The garden had a couple of trees, and a cluster of colourful **plants** around the border.*
2. (*botany*) An organism of the kingdom *Plantae*; now specifically, a living organism of the *Embryophyta* (land plants) or of the *Chlorophyta* (green algae), a eukaryote that includes double-membraned chloroplasts in its cells containing chlorophyll *a* and *b*, or any organism closely related to such an organism.

organism

tree

flower

S: (n) **plant**, flora, plant life ((botany) a living organism lacking the po… locomotion)
- S: (n) organism, being (a living thing that has (or can develop) the… to act or function independently)
  - S: (n) living thing, animate thing (a living (or once living) …
    - S: (n) whole, unit (an assemblage of parts that is re… a single entity) *"how big is that part compared to th… "the team is a unit"*
      - S: (n) object, physical object (a tangible and … entity; an entity that can cast a shadow) *"it … rackets, balls and other objects"*
        - S: (n) physical entity (an entity that has physical existence)
          - S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))



WordNet

Article Talk Read View source View history

Plant

From Wikipedia, the free encyclopedia

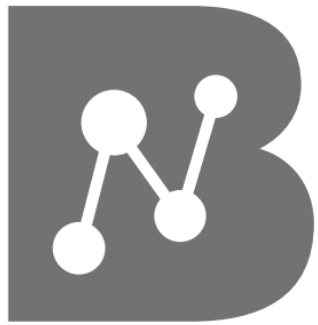
For other uses, see Plant (disambiguation).

Plants, also called **green plants** (**Viridiplantae** in Latin), are multicellular eukaryotes of the kingdom **Plantae**. They form a clade that includes the flowering plants, conifers and other gymnosperms, ferns, clubmosses, hornworts, liverworts, mosses and the green algae. Plants exclude the red and brown algae, the fungi, archaea, bacteria and animals.

Green plants have cell walls with cellulose and characteristically obtain most of their energy from sunlight via photosynthesis by primary chloroplasts, derived from endosymbiosis with cyanobacteria. Their chloroplasts contain chlorophylls a and b which gives them their green color. Some plants are parasitic and have lost the ability to produce normal amounts of chlorophyll or to photosynthesize. Plants are also characterized by sexual reproduction, modular and indeterminate growth, and an alternation of generations, although asexual reproduction is also common.

…s are difficult to determine, but as of 2010, there are thought to be 300–315 thousand species of plants, …at majority, some 260–290 thousand, are seed plants (see the table below).[1] Green plants provide

WIKIPEDIA
The Free Encyclopedia


BabelNet

French	Organisme vivant qui synthétise sa nourriture à partir de substances inorganiques, possède des membranes cellulaires en cel… manière identique à un stimulus, dépourvu d'organes sensoriels, et n'a pas de moyen de locomotion.
	Ogni organismo vivente che sintetizza il suo cibo da sostanze inorganiche, possiede membrane cellulari di cellulosa, risponde… ad uno stimolo, manca di organi di senso specializzati e del sistema nervoso, e non ha poteri di locomozione.
	Organizm o komórkach okrytych ścianą komórkową, samożywny, wytwarzający dzięki ciałkom zieleni w drodze asymilacji złoż… związków nieorganicznych lub cudzożywny, odżywiający się związkami organicznymi.
nian	Vsak živ organizem, ki si hrano sintetizira iz anorganskih snovi, proizvaja celulozne celične stene, se počasi in pogosto trajno … specializirani čutni organi in živčni sistem in se ne more premikati.
lish	En levande organism som typiskt syntetiserar sin föda från oorganiska ämnen, som har cellväggar av cellulosa, som reagerar … stimulus, som saknar specialiserade sinnesorgan och nervsystem, och som saknar rörelseförmåga

…nonyms and translations

…ression		
…guage	Spelling	Annotation
…que	landare	show ▼
…mål	plante	show ▼
…lon	plant	show ▼
…garian	растение	show ▼
…tilian	planta	show ▼
…ch	rostlina	show ▼
…lish	planter	show ▼
…ch	plant	show ▼
…llish	flora	show ▼

OMEGAWIKI

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)

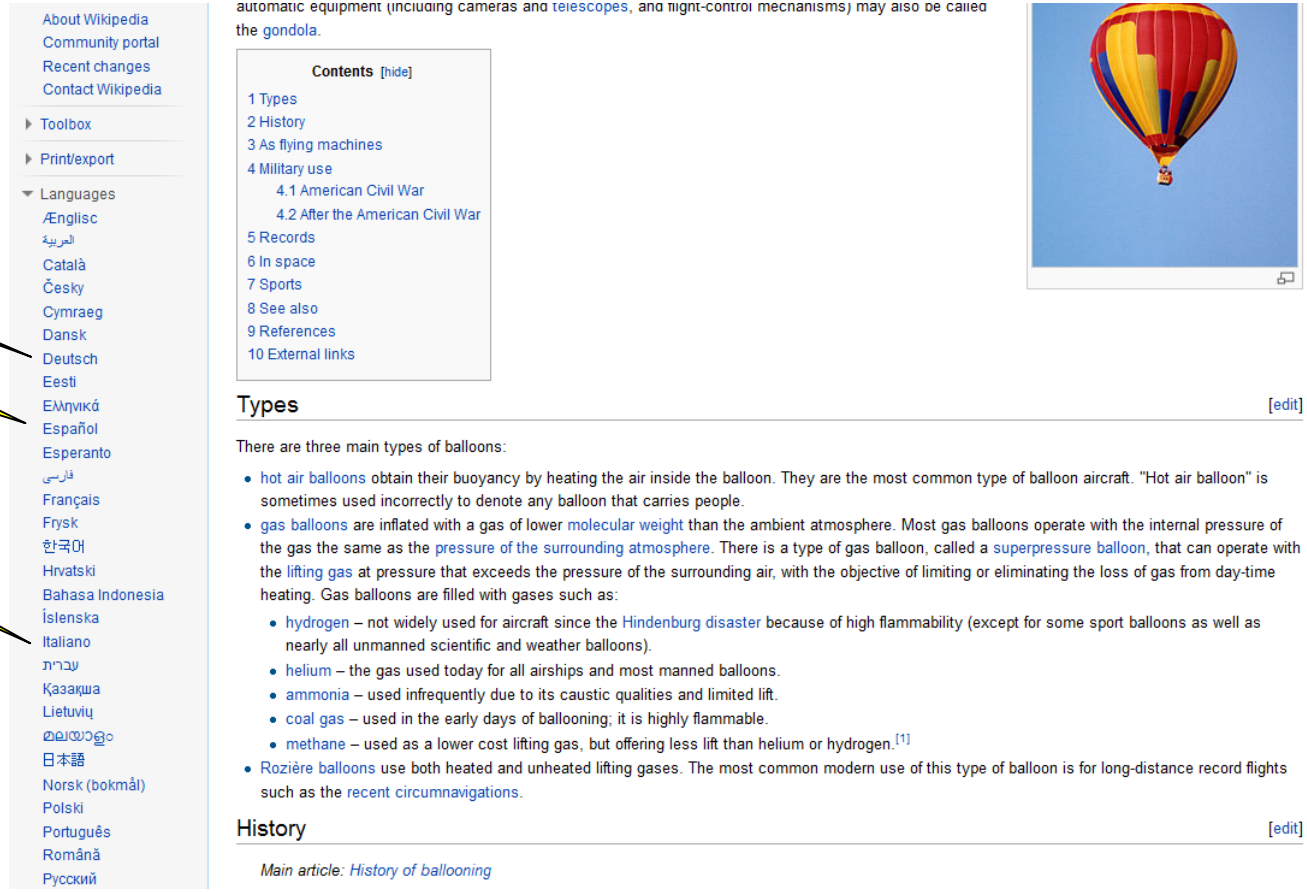
Building BabelNet: Translating Babel synsets

1. Exploiting Wikipedia interlanguage links

Ballon

globo aerostàtico

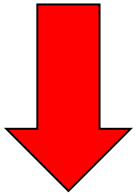
pallone aerostatico



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.



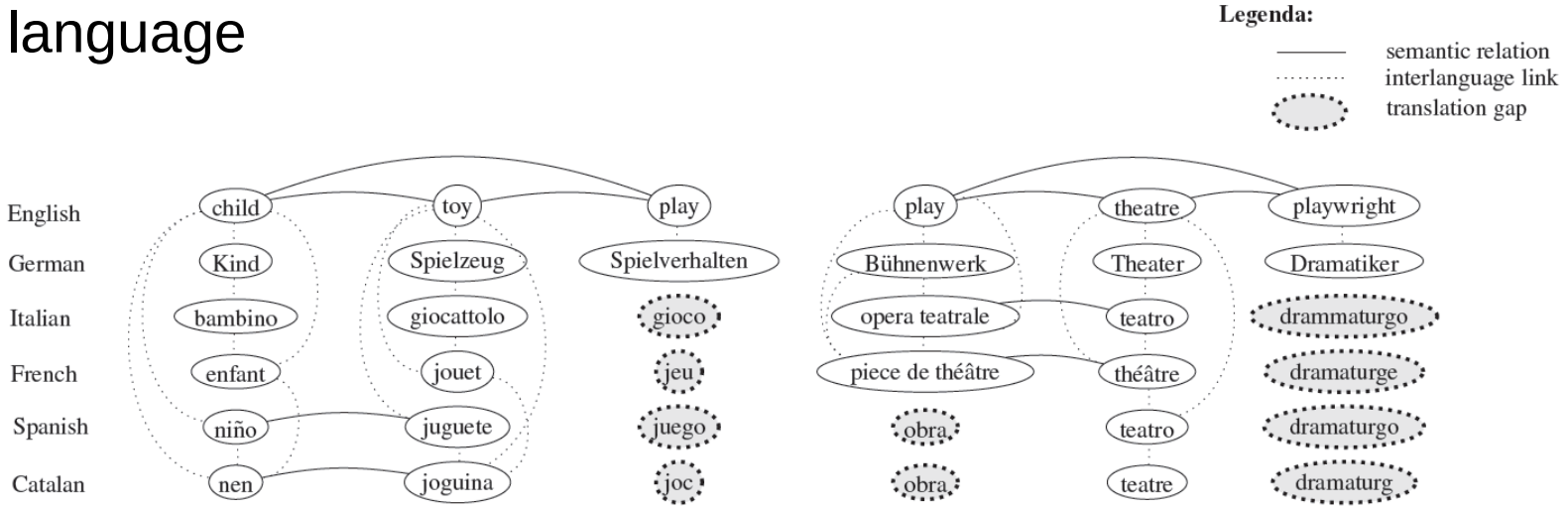
Statistical Machine Translation

- 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
	wikification	may refer to: the…
geoinformatics services' and '	wikification	of GIS by the masses'
the process may be called	wikification	(as in **...**
which is then called "	wikification	and to the related problem
reason needs copyediting,	wikification	, reduction of POV, work on references
huge amount of cleanup,	wikification	, etc. Version of 12 Nov

The most frequent translation of a word in a given meaning

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

The most frequent translation of a word in a given meaning

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

What is BabelNet?

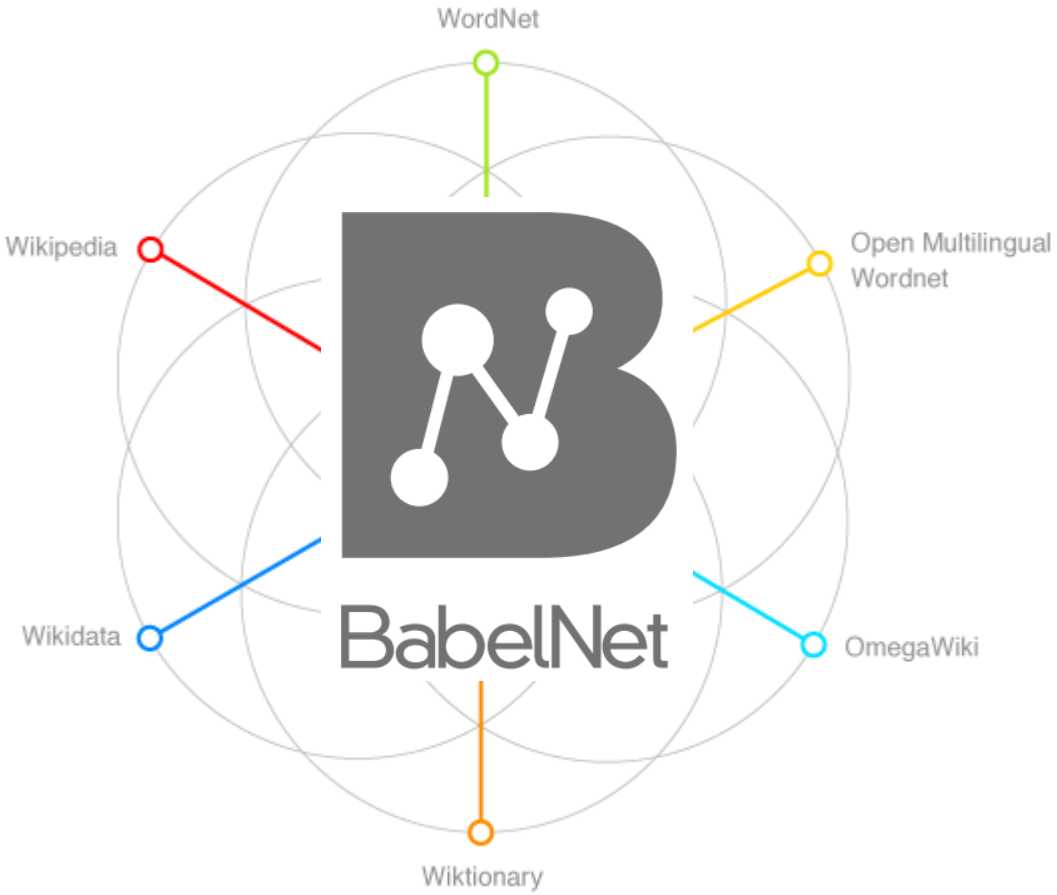
- A merger of resources of different kinds:

What is BabelNet?

- A merger of resources of different kinds:
  - WordNet: the most popular computational lexicon of English
  - Open Multilingual WordNet: a collection of open wordnets
  - WoNeF: a French WordNet
  - ItalWordNet: an Italian WordNet
  - Wikipedia: the largest collaborative encyclopedia
  - Wikidata: the largest collaborative knowledge base
  - Wiktionary: the largest collaborative dictionary
  - OmegaWiki: a medium-size collaborative multilingual dictionary
  - GeoNames: a worldwide geographical database
  - Microsoft Terminology: a computer science thesaurus
  - High-quality automatic sense-based translations

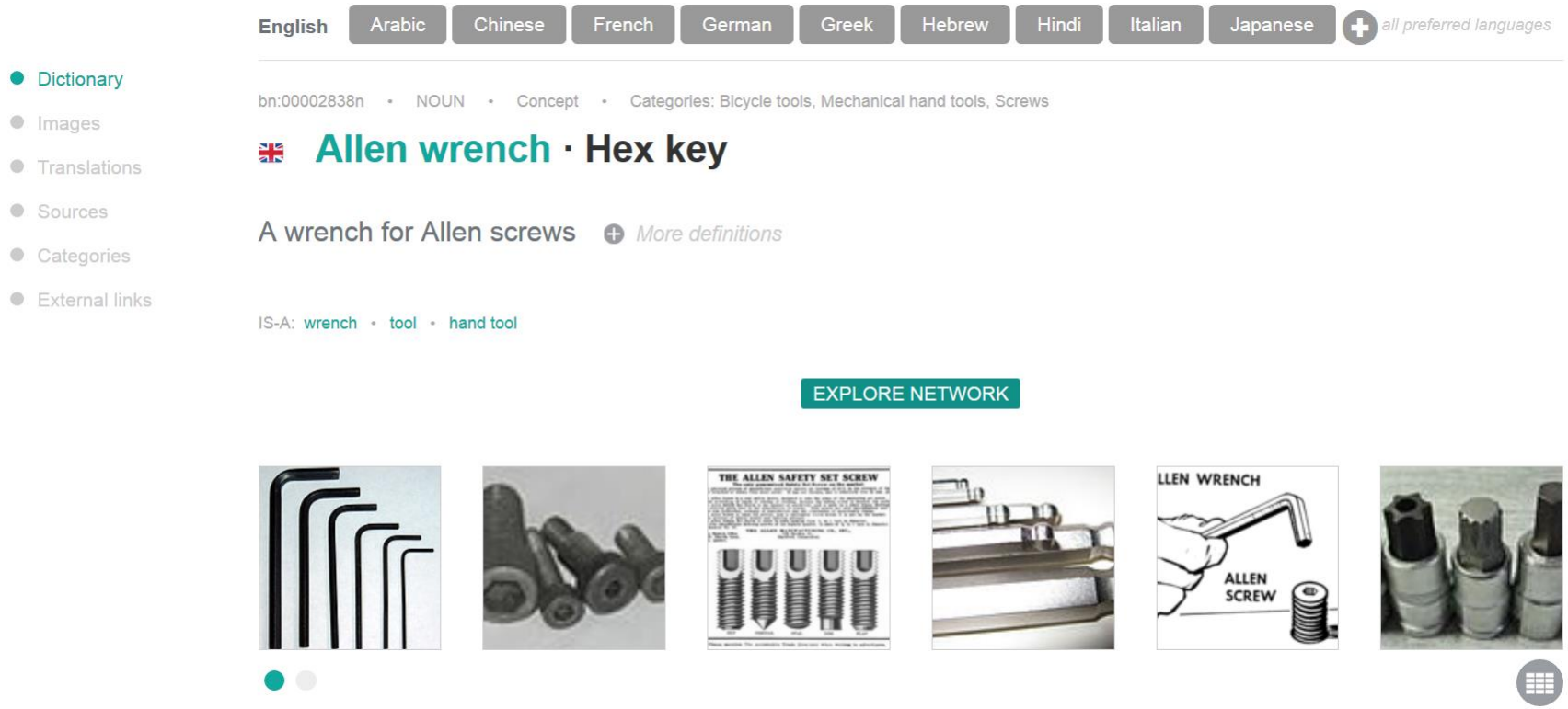
What is BabelNet?

- A merger of resources of different kinds:



Why do we need BabelNet?

- Multilinguality: the same concept is expressed in tens of languages



Why do we need BabelNet?

- Multilinguality: the same concept is expressed in tens of languages

Translations

مفك سداسي, مفتاح سداسي, *ألن وجع, وجع ألين, مفتاح ألين, مفتاح عرافة*

内六角板手, 六角匙, 内六角扳手, *内六角扳手, 六 角 扳 手*

Allen wrench, Hex key, Allen key, Hex head wrench, Allen bolt, Allan keys, Inbus, Alum key, Allan wrench, Zeta key, Allen socket, Hex wrench, Allum key, Unbrako, Alan wrench, Alan key, Allen keys, Imbus, Hex driver, Allan key, Socket head, Umbrako

Clé Allen, clef Allen, Clef six pans, Clé six pans creux, *clé hexagonale*

Inbusschlüssel, Innensechskantschlüssel, Innensechskant, Inbus, Inbusschraube, Innensechskantschraube Bauer und Schaurte, Sechskantschlüssel, Innensechskantschraube, Sechskantschraubendreher

κλειδί allen, εξάγωνο κλειδί

מפתח אלן, *אלן מפתח ברגים, מפתח ברגים, אלן מפתח*

एलन रिंच, हेक्स कुंजी

Chiave a brugola, Brugola, Viti brugola, Imbus, Chiave di Allen, Chiave Allen, *chiave esagonale*

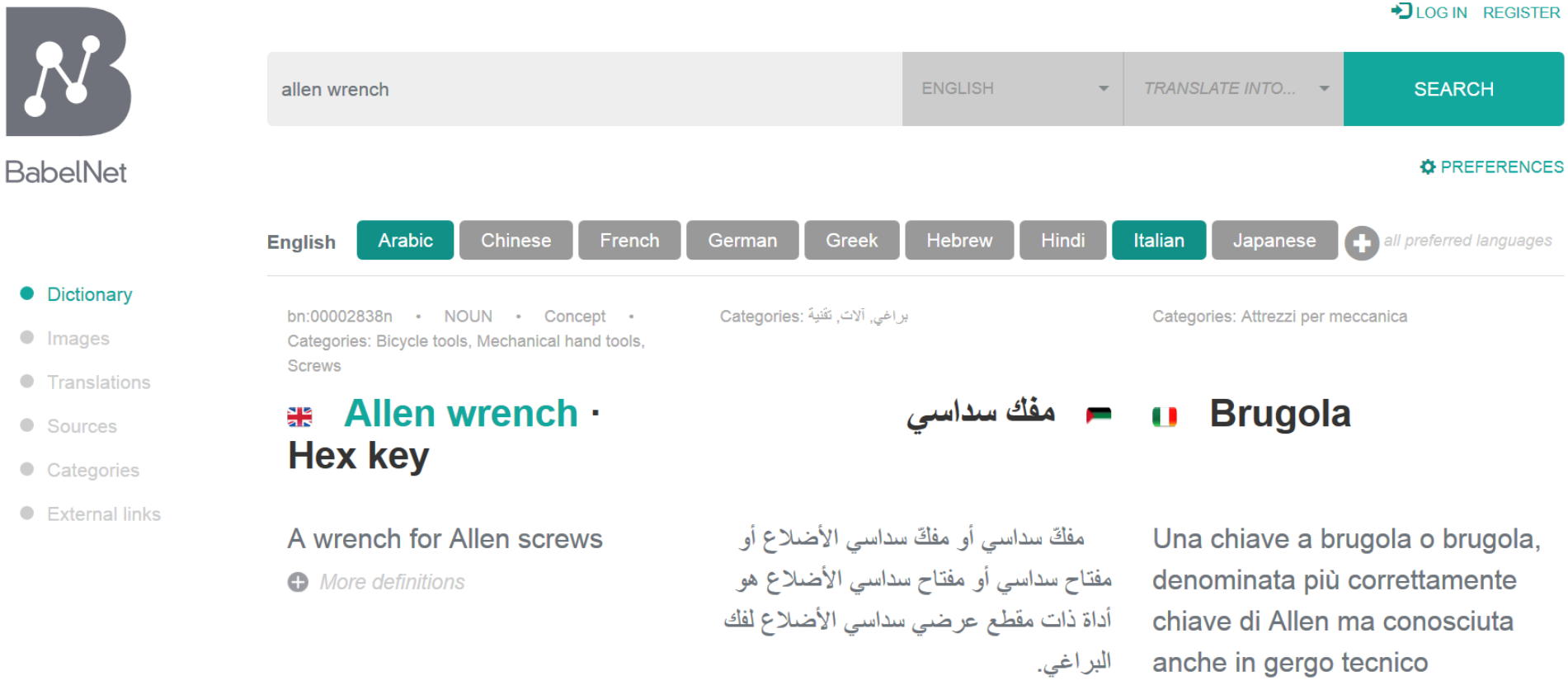
六角棒スパナ, 六角レンチ, 六角棒レンチ, ヘキサゴンレンチ, アーレンキー, *六 角 レ ン チ 。*

Шестигранный ключ, Шестигранный шлиц, Инбусовый ключ, Инбус, Имбусовый ключ, Шестигранник

llave allen, Llaves allen, Tornillo allen, *llave hexagonal*

Why do we need BabelNet?

- Multilinguality: the same concept is expressed in tens of languages

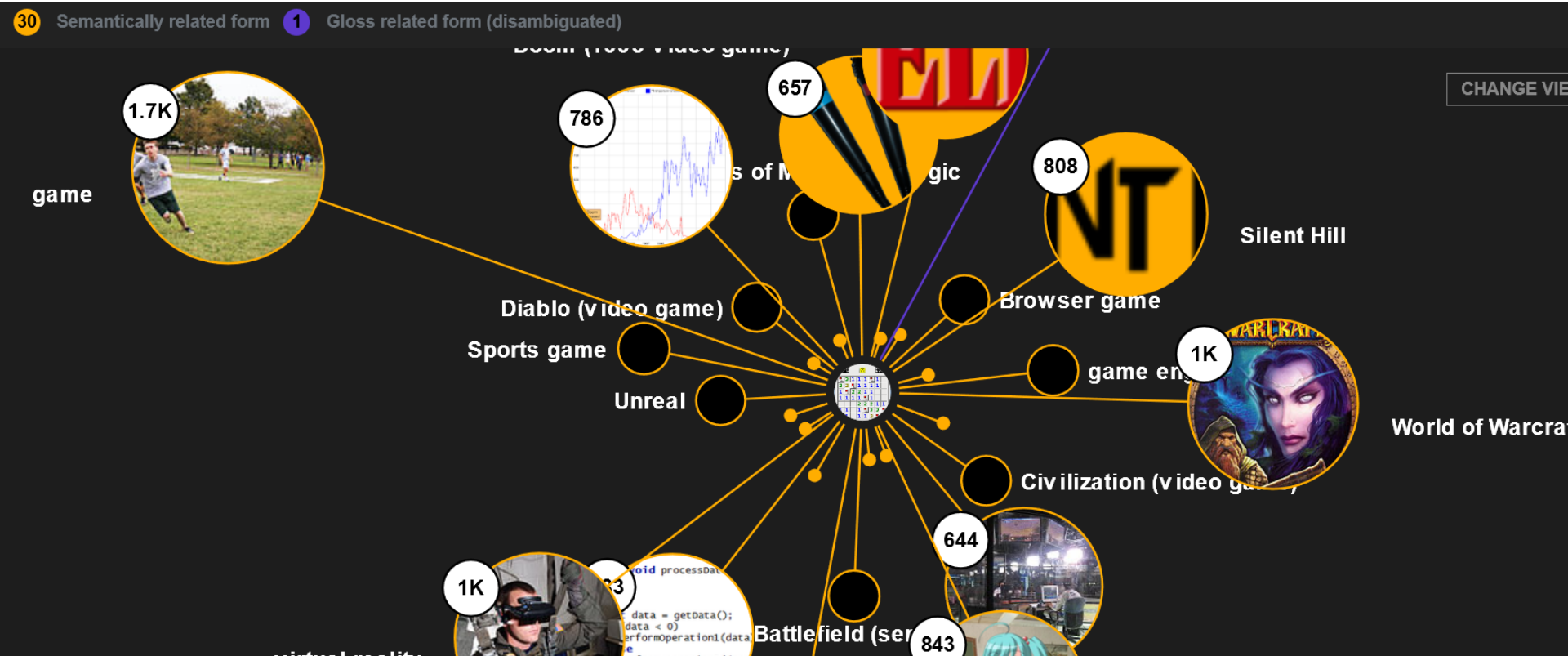


Why do we need BabelNet?

- Multilinguality: the same concept is expressed in tens of languages
- Coverage: 271 languages and 14 million entries!
  - **6M** concepts and **7.7M** named entities
  - **745M** word senses
  - **380M** semantic relations (27 relations per concept on avg.)
  - **11M** images associated with concepts
  - **41M** textual definitions
  - **1.6M** concepts with domains associated

Why do we need BabelNet?

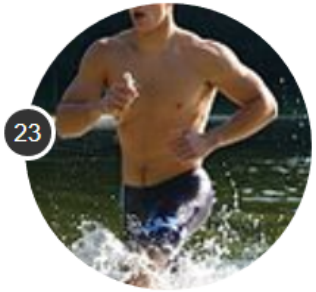
- Multilinguality: the same concept is expressed in tens of languages
- Coverage: 271 languages and 14 million entries!
- Concepts and named entities together: dictionary and encyclopedic knowledge is semantically interconnected



Why do we need BabelNet?

- Multilinguality: the same concept is expressed in tens of languages
- Coverage: 271 languages and 14 million entries!
- Concepts and named entities together: dictionary and encyclopedic knowledge is semantically interconnected
- "Dictionary of the future": semantic network structure with labeled relations, pictures, multilingual synsets

Verb



run

Move fast by using one's feet, with one foot off the ground at any given time

ID: 00093170v | Concept

هَرْوَلَ, جَرَى, رَكَضَ

奔跑, 跑

courir

rennen

τρέχω, κινούμαι

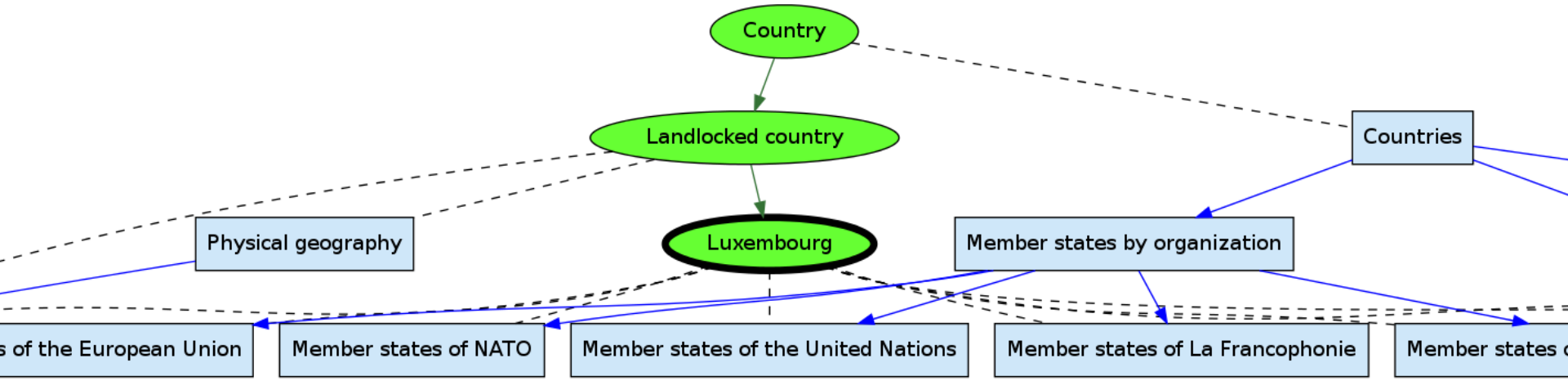
רץ

correre

Why do we need BabelNet?

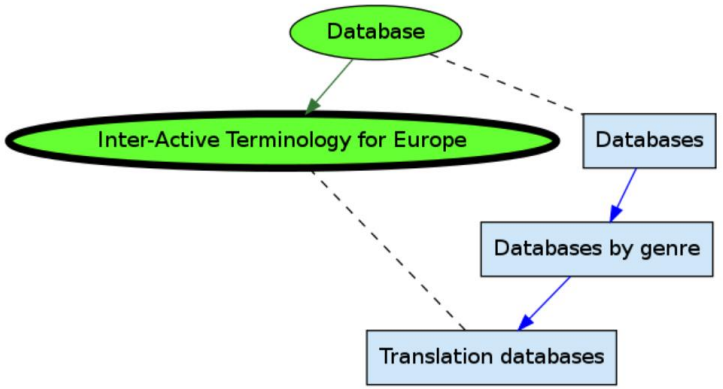
- Multilinguality: the same concept is expressed in tens of languages
- Coverage: 271 languages and 14 million entries!
- Concepts and named entities together: dictionary and encyclopedic knowledge is semantically interconnected
- "Dictionary of the future": semantic network structure with labeled relations, pictures, multilingual synsets
- Full-fledged taxonomy: is-a relations are available for both concepts and named entities (**Wikipedia Bitaxonomy**)

Why do we need BabelNet?



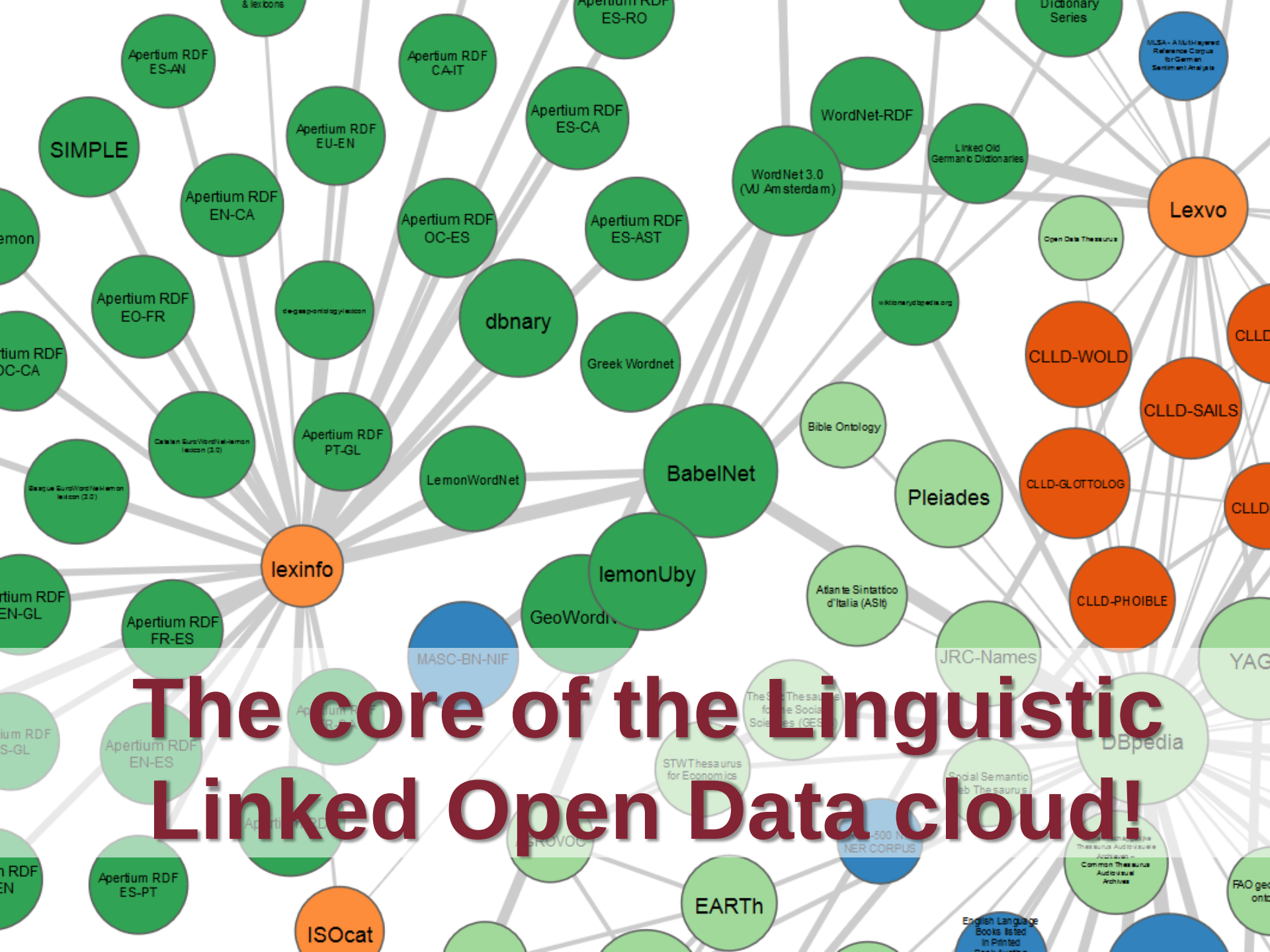
with labeled relations, pictures, mu

- Full-fledged taxonomy: is-a relatio
  both concepts and named entities
  - Luxembourg *is-a* Landlocked country
  - IATE *is-a* database
  - BabelNet *is-a* semantic network & encyclopedic dictionary



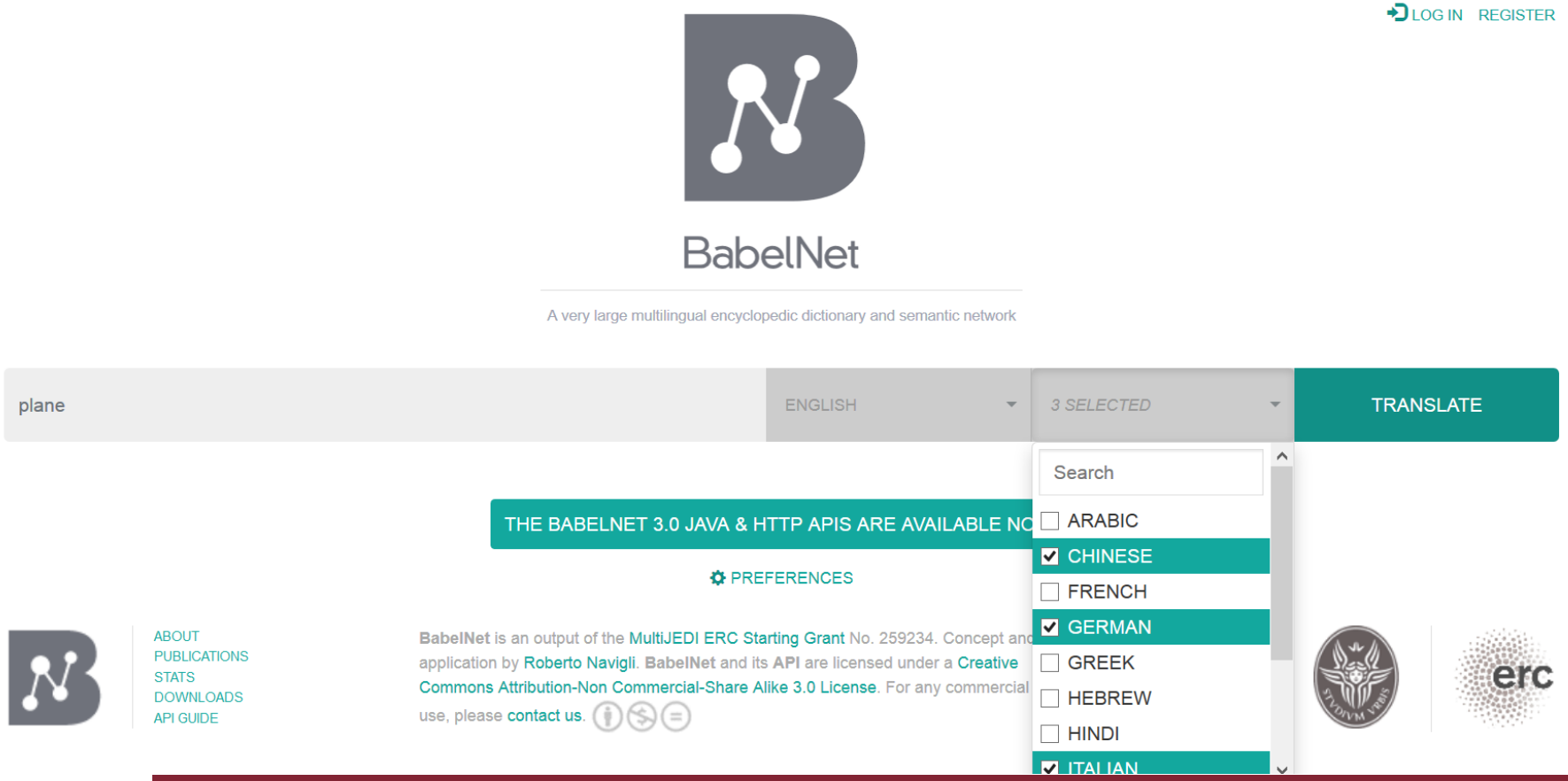
Why do we need BabelNet?

- Multilinguality: the same concept is expressed in tens of languages
- Coverage: 271 languages and 14 million entries!
- Concepts and named entities together: dictionary and encyclopedic knowledge is semantically interconnected
- "Dictionary of the future": semantic network structure with labeled relations, pictures, multilingual synsets
- Full-fledged taxonomy: is-a relations are available for both concepts and named entities (**Wikipedia Bitaxonomy**)
- Easy access: Java and HTTP RESTful APIs; SPARQL endpoint (2 billion triples)


Apertium RDF ES-AN
Apertium RDF CA-IT
Apertium RDF ES-RO
Apertium RDF ES-CA
Apertium RDF EU-EN
SIMPLE
Apertium RDF EN-CA
Apertium RDF OC-ES
Apertium RDF ES-AST
Apertium RDF EO-FR
dbnary
Greek Wordnet
WordNet-RDF
WordNet 3.0 (VU Amsterdam)
Linked Old Germanic Dictionaries
Lexvo
Open Data Thesaurus
CLLD-WOLD
CLLD-SAILS
CLLD-GLOTTOLOG
CLLD-PHOIBLE
Bible Ontology
Pleiades
Apertium RDF PT-GL
LemonWordNet
BabelNet
lexinfo
lemonUby
Atlante Sintattico d'Italia (ASIt)
JRC-Names
Apertium RDF FR-ES
MASC-BN-NIF
Apertium RDF EN-ES
STW Thesaurus for Economics
DBpedia
Apertium RDF ES-PT
ISOcat
EARTh
The core of the Linguistic Linked Open Data cloud!

What can we do with BabelNet?

- Search and translate:



What can we do with BabelNet?

- Noun
- Verb
- Adjective

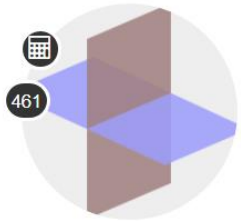
Noun

airplane, plane, aeroplane

An aircraft that has a fixed wing and is powered by propellers or jets

ID: 00001697n | Concept

ZH 飞机, 固定翼飞機, 飛機
DE Flugzeug
EL αεροπλάνο, αεροσκάφος
IT aeroplano, aereo, apparecchio



plane, sheet

(mathematics) an unbounded two-dimensional shape

ID: 00062766n | Concept

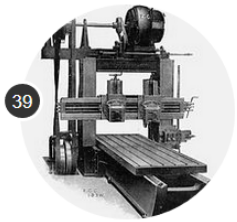
ZH 平面, 二维形状
DE Ebene (Mathematik)
EL επίπεδο
IT piano, piano geometrico

plane

A level of existence or development

ID: 00062767n | Concept

ZH 发展水平
DE Ebene
EL επίπεδο
IT piano, Spostamento della realtà, livello



planer, plane, planing machine

A power tool for smoothing or shaping wood

ID: 00062768n | Concept

DE Hobelmaschine, Hobel
IT piallatrice

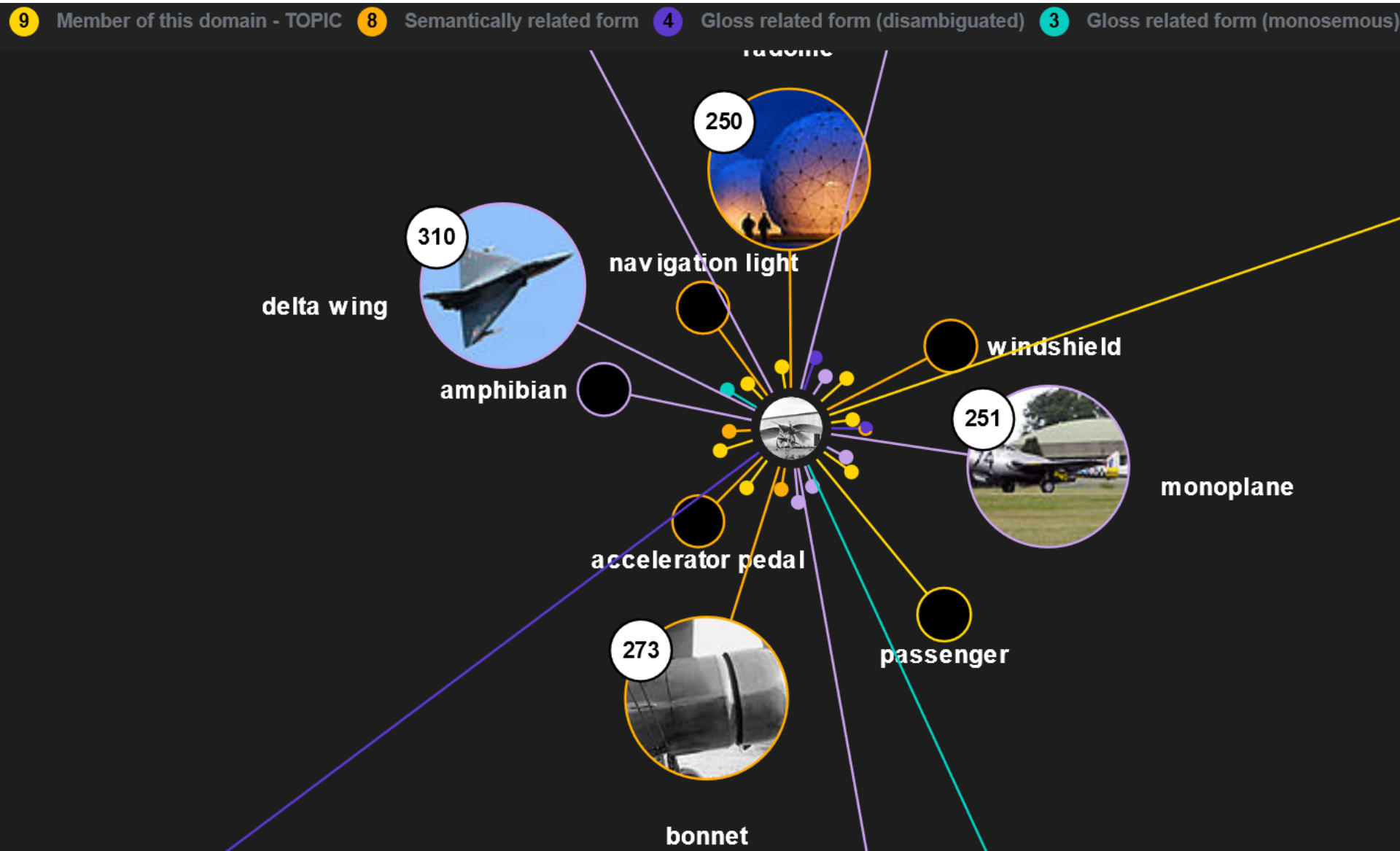
plane, carpenter's plane, woodworking plane

A carpenter's hand tool with an adjustable blade for smoothing or shaping wood

ZH 刨
DE Hobel
EL πλάνη

What can we do with BabelNet?

- Explore the network:



Anatomy of BabelNet

BabelNet

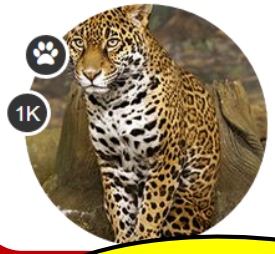
LOG IN REGISTER

jaguar | ENGLISH | ITALIAN | TRANSLATE

PREFERENCES

Noun

Babel synset



jaguar, panther, Felis onca

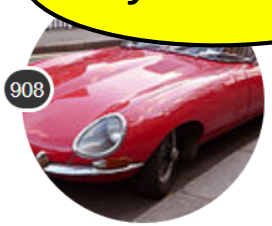
A large spotted feline of tropical America similar to the leopard; in some classifications c genus Felis

Babel sense

ID: 00033987n | C

IT giaguaro, Panthera onca, pantera

synset ID



ar Cars, Jaguar

Jaguar Cars is a brand of Jaguar Land Rover, a British multinational car manufacturer headquartered in Whitley, Coventry, England, owned by Tata Motors since 2008.

ID: 00688731n | Named Entity

IT Jaguar

Atari Jaguar, Jaguar (video game console)

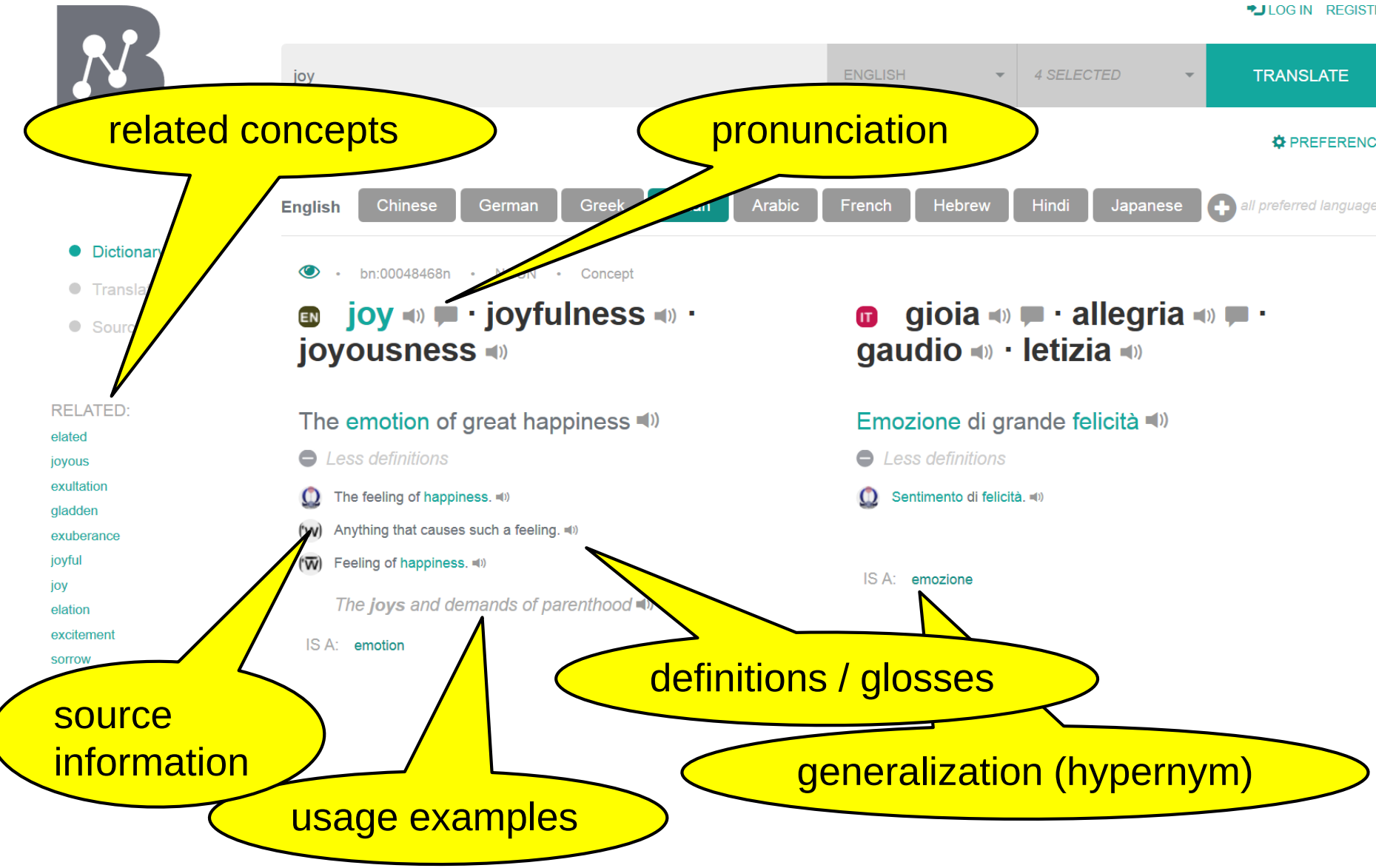
The Atari Jaguar is a home video game console that was released by Atari Corporation in 1993.

ID: 02142312n | Named Entity

IT Atari Jaguar

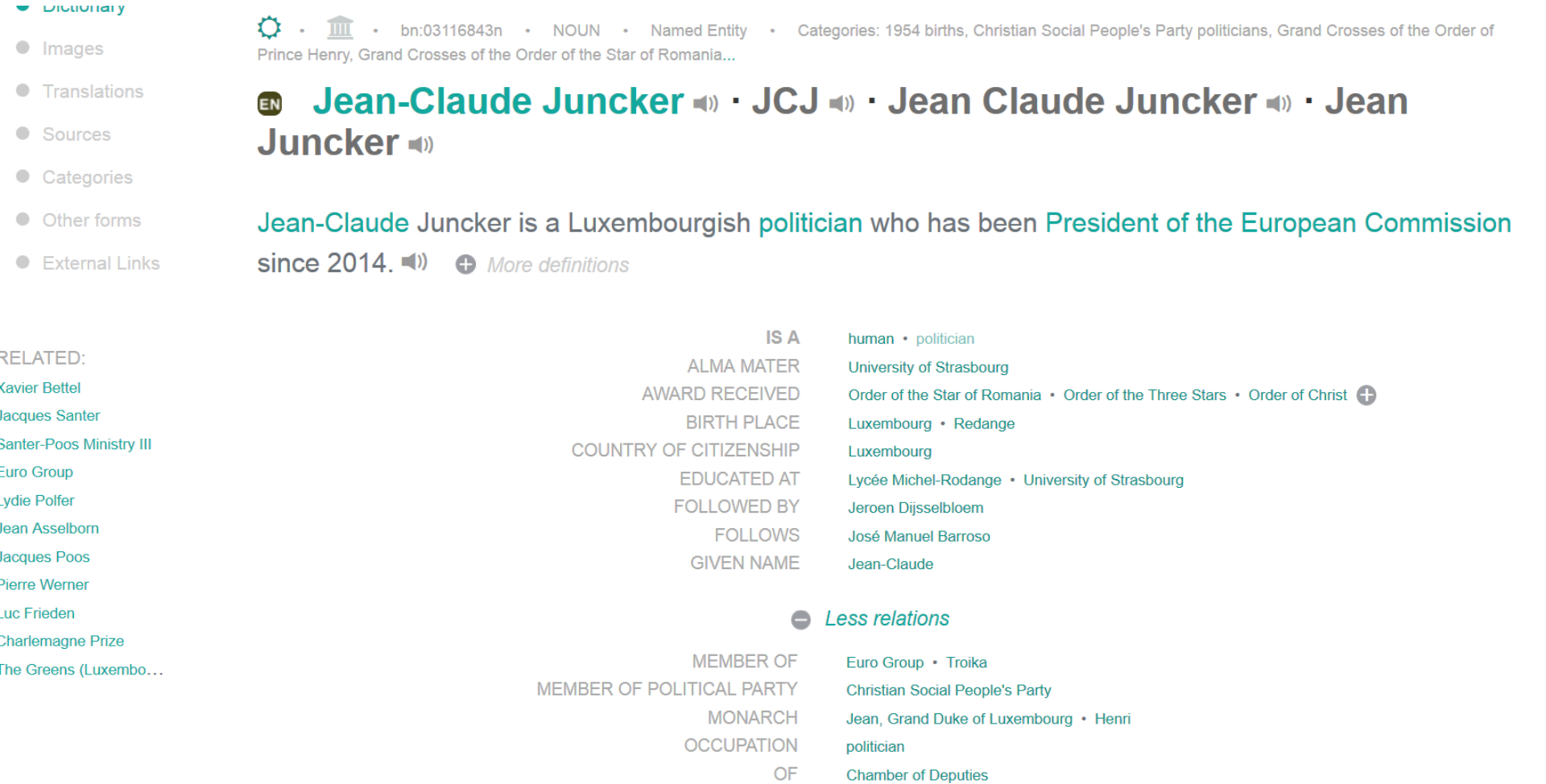
Anatomy of BabelNet

Anatomy of BabelNet



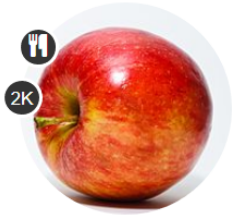
BabelNet 3.6 is now a knowledge base!

- **Semantic relations** from Wikidata + Infoboxes (superset of DBpedia) + relations extracted with Open Information Extraction techniques



BabelNet 3.6 is now a knowledge base!

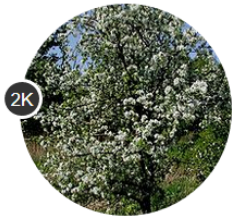
- **Semantic relations** from Wikidata + Infoboxes (superset of DBpedia) + relations extracted with Open Information Extraction techniques
- **Domain labels** for millions of synsets



apple, apple blossom, apple peel

Fruit with red or yellow or green skin and sweet to tart crisp whitish flesh

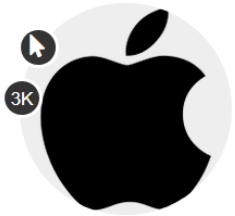
ID: 00005054n | Concept



apple, Malus pumila, orchard apple tree

Native Eurasian tree widely cultivated in many varieties for its firm rounded edible fruits

ID: 00005055n | Concept



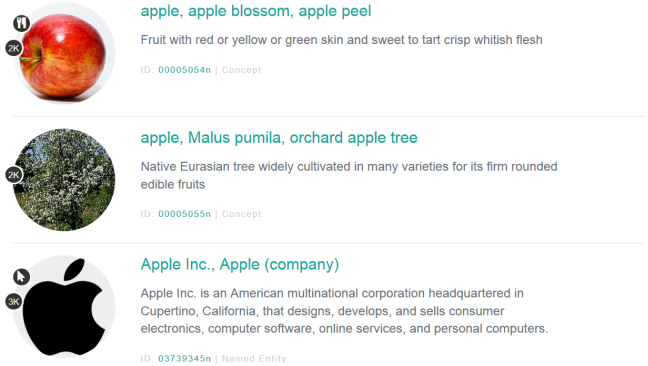
Apple Inc., Apple (company)

Apple Inc. is an American multinational corporation headquartered in Cupertino, California, that designs, develops, and sells consumer electronics, computer software, online services, and personal computers.

ID: 03739345n | Named Entity

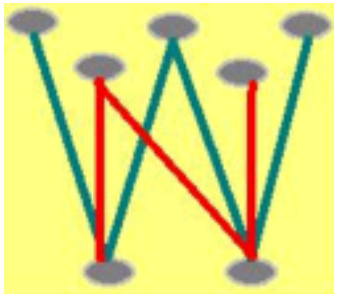
BabelNet 3.6 is now a knowledge base!

- **Semantic relations** from Wikidata + Infoboxes (superset of DBpedia) + relations extracted with Open Information Extraction techniques
- **Domain labels** for millions of synsets



- **Phrases and collocations** for most synsets
  - For bus (**vehicle**) : driver, stop, station, lane, line, depot, schedule, terminal, ticket, timetable, etc.
  - For bus (**computer**): architecture, error, mastering, network, etc.

Integrating WordNet with Wikipedia…



WordNet

Is that all?!?

Open Multilingual WordNet [Bond and Foster, 2013]

- http://compling.hss.ntu.edu.sg/omw/
- 34 languages
- Mappings to the Princeton WordNet synsets
- More than 1,000,000 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)

- Hundreds of languages
- About 54,000 entries («synsets»)

DefinedMeaning:867095 - aerostat

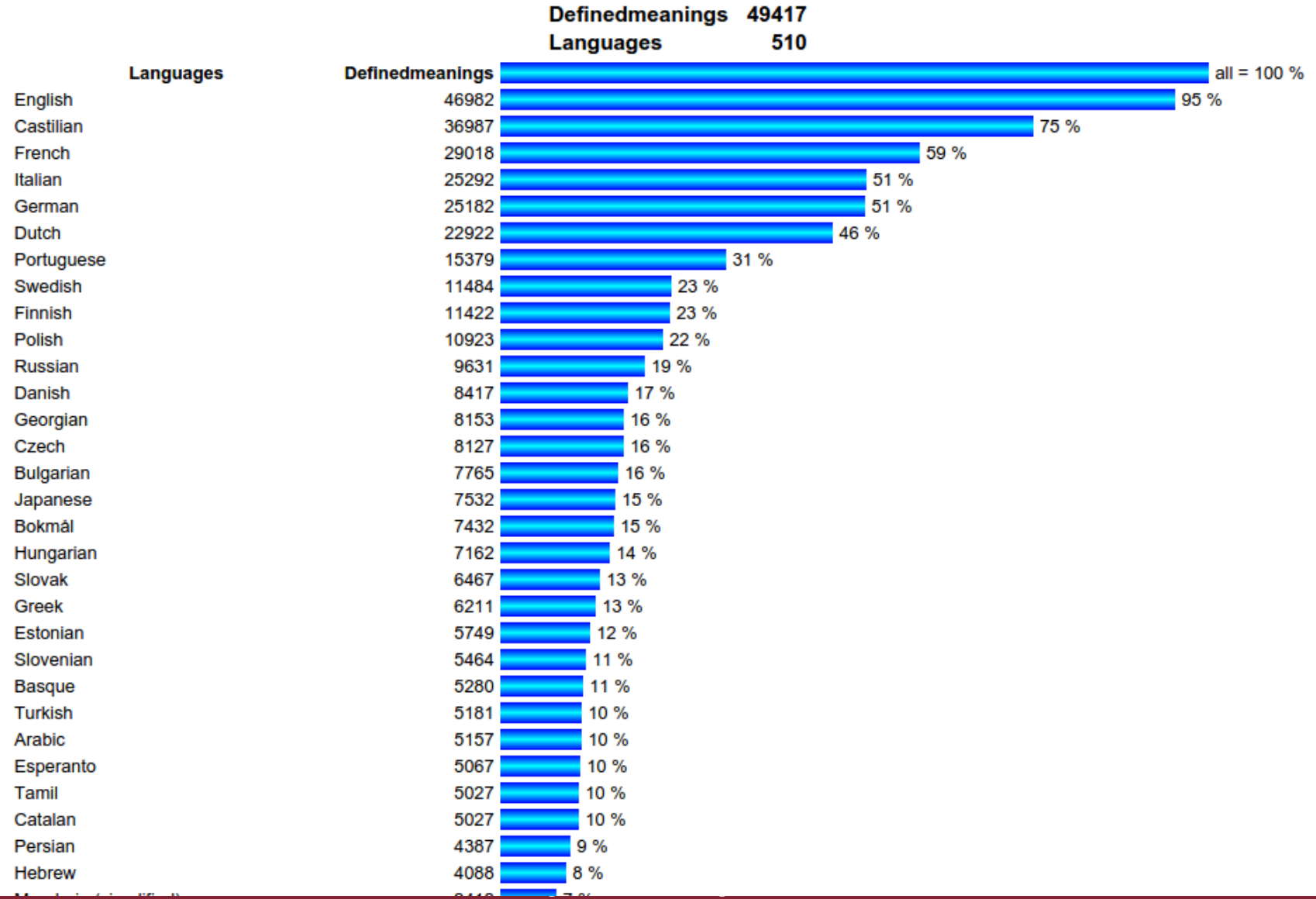
▼ **Definition**

Language	Text
Castilian	Medio de transporte aéreo, una aeronave más ligera que el aire.
Dutch	Een transportmiddel door de lucht, een luchtschip lichter dan de lucht.
English	Means of air transport, an aircraft lighter than air.
French	Un moyen de transport aérien, un aéronef plus léger que l'air.
German	Ein Lufttransportmittel; ein Luftfahrzeug, das leichter als Luft ist.
Polish	Środek transportu lotniczego, pojazd lotniczy lżejszy od powietrza.

▼ **Synonyms and translations**

Expression		
Language	**Spelling**	**Annotation**
Afrikaans	lugballon	show ▼
Castilian	aerostato	show ▼
Castilian	globo	show ▼
Castilian	globo aerostático	show ▼
Danish	ballon	show ▼
Dutch	heteluchtballon	show ▼
Dutch	luchtballon	show ▼
English	aerostat	show ▼
English	balloon	show ▼

Some statistics for OmegaWiki



Wiktionary (http://www.wiktionary.org)

- A collaborative dictionary!
- Hundreds of languages
- About 4.8M entries

a multilingual free encyclopedia

Wiktionary

[ˈwɪkʃənrɪ] *n.*, a wiki-based Open Content dictionary

English [edit]

Etymology [edit]

1570, "a game played with a large, inflated leather ball" (possibly via Middle French *ballon*) from Italian *pallone* "large ball" from *palla* "ball", from Lombardic *palla* "ball" from Proto-Germanic **ballô* ("ball"), from Proto-Indo-European **bholn-* ("bubble"), from Proto-Indo-European **bhel-* ("to blow, swell, inflate"). Akin to Old High German *ballo*, *bal* "ball" (German *Ballen* "bale"; *Ball* "ball"). More at ball.

Colorful toy balloons.

Pronunciation [edit]

- IPA(key): /bəˈluːn/
- Audio (US) 0:00 MENU
- Rhymes: -uːn

Noun [edit]

balloon (*plural* **balloons**)

1. An inflatable buoyant object, often (but not necessarily) round and flexible.
2. Such an object as a child's toy.
3. Such an object designed to transport people through the air.
4. (*medicine*) A sac inserted into part of the body for therapeutic reasons; such as angioplasty.
5. A speech bubble.
6. A type of glass cup, sometimes used for brandy.
7. (*architecture*) A ball or globe on the top of a pillar, church, etc.

   *the **balloon** of St. Paul's Cathedral in London*

Wikipedia has an article on: Balloon

Some statistics for Wiktionary

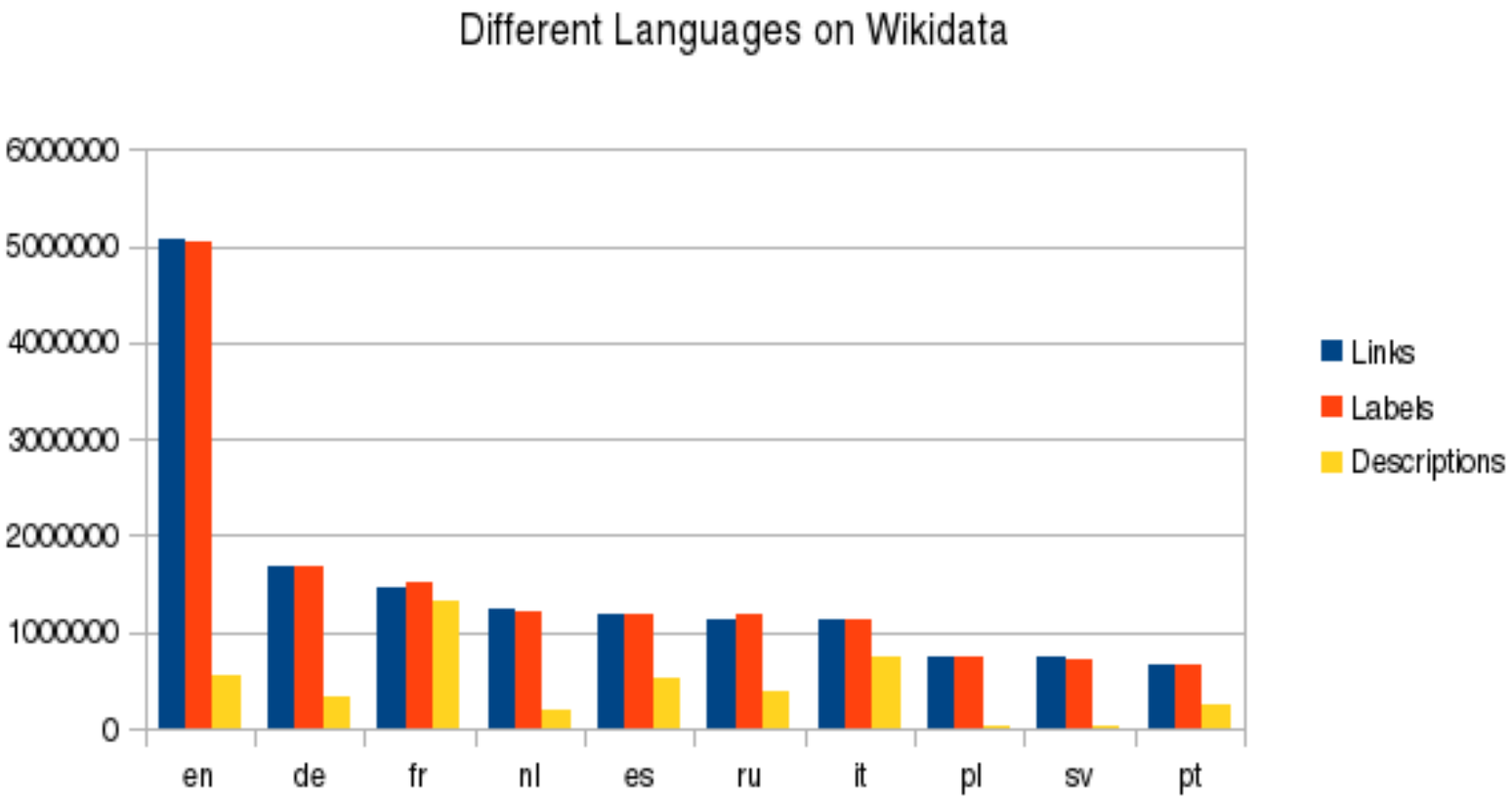
a multilingual free encyclopedia

Wiktionary

language	gloss[1] definitions	change	entries	change	form definitions	change	total definitions	change
Total	**2200562** (10447)	**+12235** (+183)	**4887884** (8638)	**+101487** (+145)	**4589578** (380)	**+115896** (+1)	**6790140** (10827)	**+128131** (+184)
Latin	52908 (98)	+80	679161 (115)	+137	1075458 (50)	+166	1128366 (148)	+246
English	544240	+2497	641292 (1)	+3850	282856	+2006	827096	+4503
Italian	140765	+551	543385	+581	542045	+362	682810	+913
Spanish	61419	+148	364295	+451	459344	+311	520763	+459
French	69793	+863	321681	+5974	346120	+6128	415913	+6991
Portuguese	62885	+586	253433	+495	252233	+172	315118	+758
German	51561	+91	178686	+112	198631	+81	250192	+172
Russian	43387	+460	153531	+83093	150168	+103708	193555	+104168
Finnish	78607 (3)	+138	122230 (3)	+150	73372	+39	151979 (3)	+177
Latvian	11305	-1	121144	-1	237722	-2	249027	-3
Esperanto	14216	+3	109405	-11	96468	-13	110684	-10
Swedish	23152	+28	91738	+21	80499	-1	103651	+27
Dutch	31955	+143	75912	+179	62676	+66	94631	+209
Japanese	93648	+935	73508	+760	14813	+317	108461	+1252
Catalan	13919	+219	66183	+343	70495	+197	84414	+416

Wikidata (http://www.wikidata.org)

- A collaborative **knowledge base**!
- Hundreds of languages
- About 16M entries



Statistics online at http://babelnet.org/stats

BabelNet 3.6: General statistics

ABOUT NEWS PUBLICATIONS STATS DOWNLOADS API GUIDE

Number of languages:	271
Total number of Babel synsets:	13,801,844
Total number of Babel senses:	745,856,326
Total number of concepts:	6,066,396
Total number of Named Entities:	7,735,448
Total number of lexico-semantic relations:	380,239,084
Total number of glosses (textual definitions):	40,705,588
Total number of images:	10,767,833
Total number of Babel synsets with at least one domain:	1,558,806
Total number of compounds:	743,296
Total number of other forms:	6,393,568
Total number of Babel synsets with at least one picture:	2,948,668
Total number of RDF triples:	1,958,820,772

- Statistics between the BabelNet versions
- Languages and Coverage
- Number of polysemous and monosemous words
- Composition of Babel synsets
- Number of lexico-semantic relations
- External Links
- Number of senses by part of speech
- Number of lemmas by part of speech
- Number of concepts, Named Entities and definitions

WordNet-Wikipedia mapping accuracy

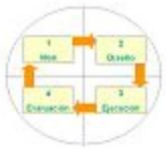
- Accuracy of the resource at the intersection: **98.5%**
  - Calculated on a sample of 2000 potential mappings, only 30 of which are wrongly associated
- If we consider the entire resource (14M entries), the accuracy is **99.99985%**

Key fact!

Annotating with BabelNet: all in one!

- Annotating with **BabelNet** implies annotating with WordNet, Wikipedia, OmegaWiki, Open Multilingual WordNet, Wikidata **and** Wiktionary

Progetto | più | ambizioso | non c'è: | creare | un | dizionario enciclopedico | multil



Progetto
Any piece of work that is undertaken or attempted

più
Used to form the comparative of some adjectives and adverbs

ambizioso
Requiring full use of your abilities or resources

creare
Bring into existence

dizionario enciclopedico
An encyclopedic dictionary typically includes a large number of short

dizionario
A reference book containing an alphabetical list of

multilin
Using or
more tha
language

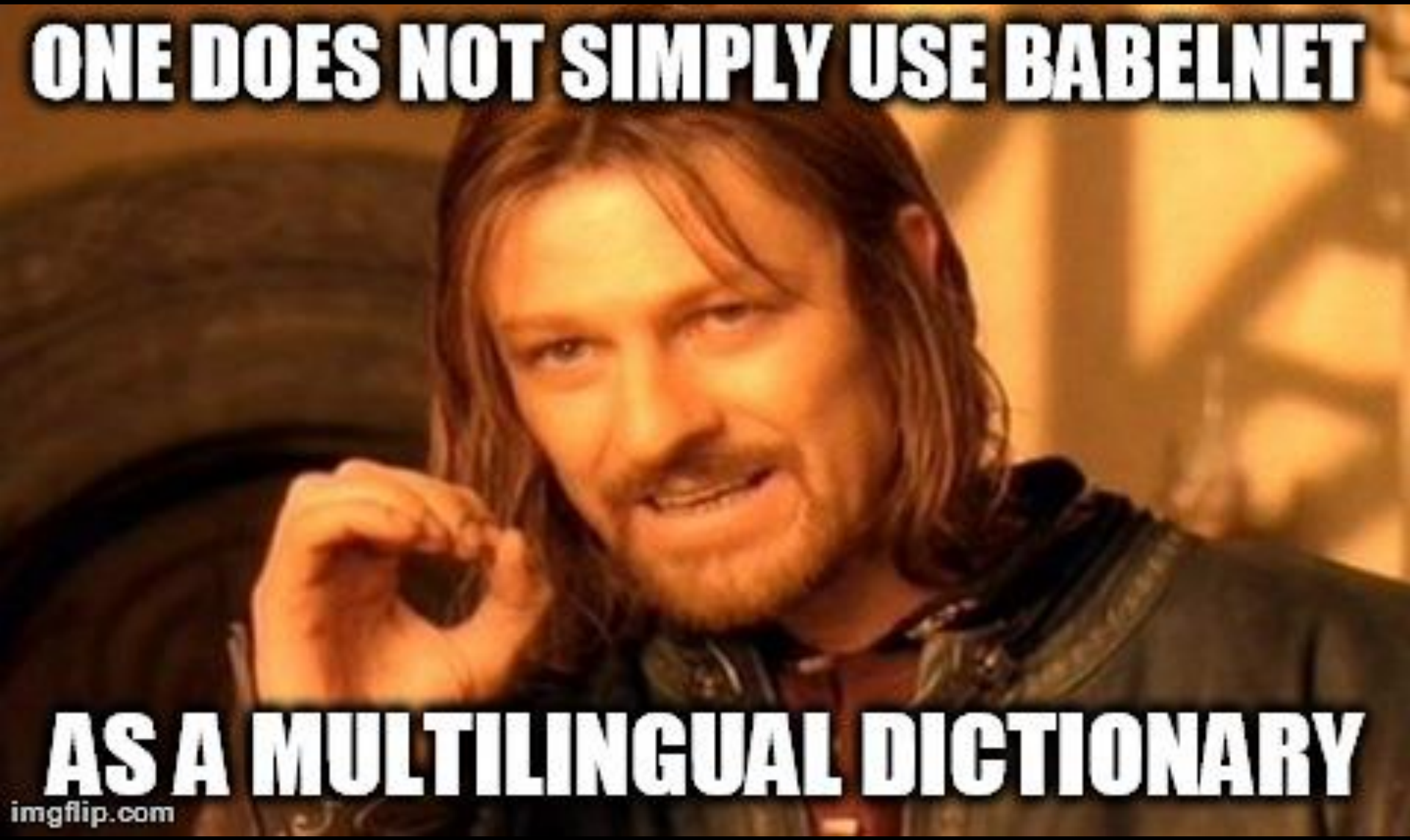
ties • Concepts

We use cookies to offer an improved online experience. By using our services, you agree to our use of cookies. OK

Key fact!

Annotating with BabelNet: all in one!

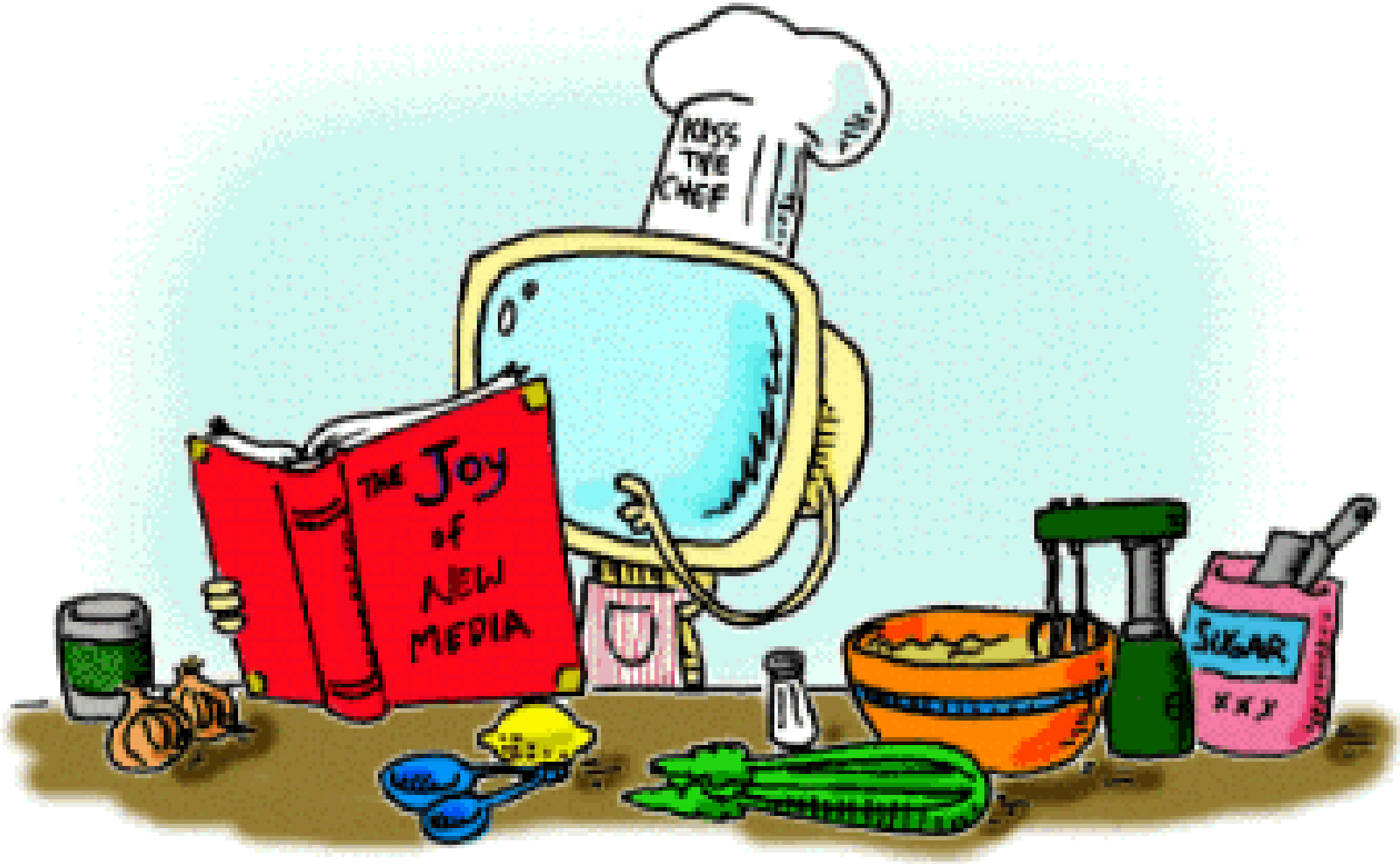
- Annotating with **BabelNet** implies annotating with WordNet, Wikipedia, OmegaWiki, Open Multilingual WordNet, Wikidata **and** Wiktionary


ONE DOES NOT SIMPLY USE BABELNET
AS A MULTILINGUAL DICTIONARY
imgflip.com

ADDRESSING AMBIGUITY

[Moro, Raganato & Navigli, TACL 2014]

Computers can be hungry!



Motivation (1): hungry computers

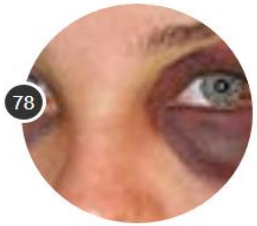
- **EN -** The mouse ate the cheese

mouse, rat, Mus

Any of numerous small rodents typically resembling diminutive rats having pointed snouts and small ears on elongated bodies with slender usually hairless tails

ID: 00056119n | Concept

topo, sorcio, Mus (genere)

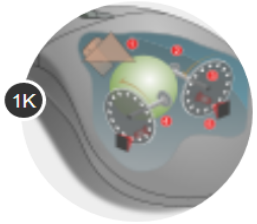


black eye, shiner, mouse

A swollen bruise caused by a blow to the eye

ID: 00010892n | Concept

occhio nero, Ematoma periorbitale, occhio pesto



mouse, computer mouse

A hand-operated electronic device that controls the coordinates of a cursor on your computer screen as you move it around on a pad; on the bottom of the device is a ball that rolls on the surface of the pad

ID: 00021487n | Concept

mouse

Motivation (1): hungry computers

- **EN -** The mouse ate the cheese

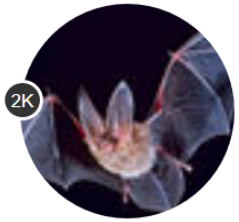
Motivation (1): hungry computers

- **EN -** The mouse ate the cheese
- **IT -** Il mouse ha mangiato il formaggio

Computers can be sadistic/cynical!

Motivation (2): avoiding horror

- **EN -** Squash: a game played in a walled court with soft rubber balls and bats like tennis rackets

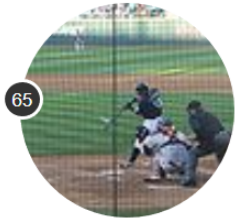


bat, chiropteran, Chiroptera

Nocturnal mouselike mammal with forelimbs modified to form membranous wings and anatomical adaptations for echolocation by which they navigate

ID: 00008977n | Concept

pipistrello, Chiroptera



At bat, at-bat, bat

(baseball) a turn trying to get a hit

ID: 00006690n | Concept

at-bat

7

squash racket, squash racquet, bat

A small racket with a long handle used for playing squash

ID: 00008978n | Concept

racchetta da squash

Motivation (2): avoiding horror

- **EN -** Squash: a game played in a walled court with soft rubber balls and bats like tennis rackets
- **IT -** Squash: una partita giocata in un campo recintato con palle di gomma morbida e pipistrelli come racchette da tennis

More seriously: lexical ambiguity!

- Thomas and Mario played as strikers in Munich.

Thomas and Mario played as strikers in Munich .

Thomas
Thomas Müller is a German footballer who plays for Bayern Munich and the

Mario
Mario Gómez García is a German footballer who plays as a striker for Bayern Munich in

played
participate in games or sport; "We played hockey all afternoon"; "play cards"; "Pele

strikers
a forward on a soccer team

Munich
FC Bayern Munich, is a German sports club based in Munich, Bavaria.

Word Sense Disambiguation and Entity Linking

- Thomas and Mario are strikers playing in Munich

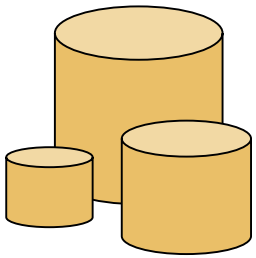
Entity Linking: The task of discovering mentions of entities within a text and linking them in a knowledge base.

WSD: The task aimed at assigning meanings to word occurrences within text.

Word Sense Disambiguation in a Nutshell

strikers
(target word)

"Thomas and Mario are **strikers** playing in Munich"
(context)



knowledge

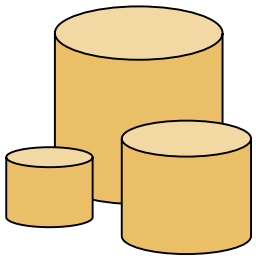
WSD
system

sense of target word

Entity Linking in a Nutshell

Thomas
(target mention)

"**Thomas** and Mario are strikers playing in Munich"
(context)



knowledge

EL
system

Named Entity

Disambiguation and Entity Linking together!

BabelNet is a huge **multilingual inventory**

for both word senses and named entities!

Multilingual Joint Word Sense Disambiguation (MultiJEDI)

Key Objective 2: use **all languages** to disambiguate **one**

So what?

Step 1: Find all possible meanings of words

- "Thomas and Mario are strikers playing in Munich"

Seth Thomas

Mario (Character)

striker (Sport)

Munich (City)

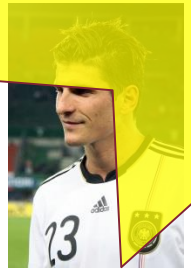
Mario (Album)

Striker (Video Game)

Thomas Müller

FC Bayern Munich

Mario Gómez



Striker (Movie)

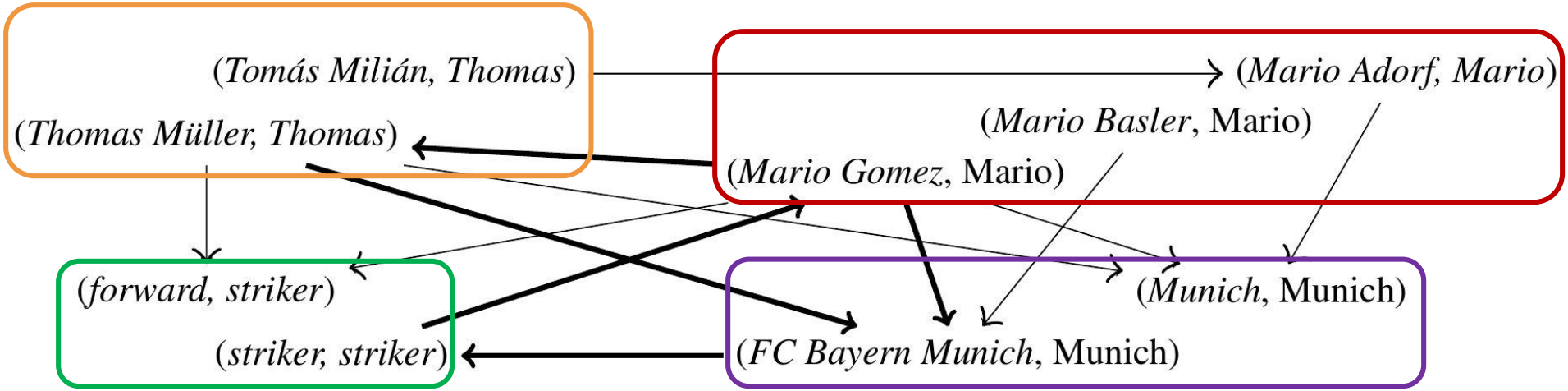
Thomas (novel)

Munich (Song)

Ambiguity!

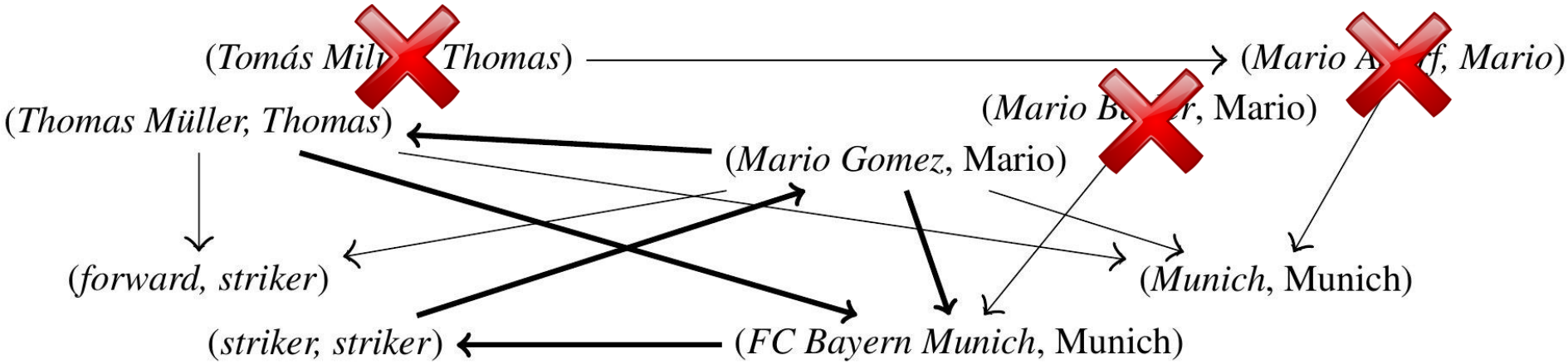
Step 2: Connect all the candidate meanings

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



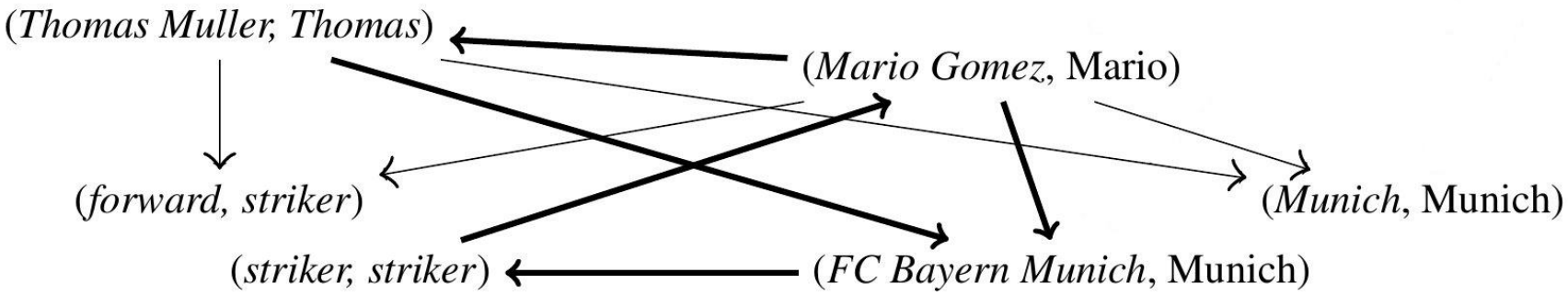
Step 3: Extract a dense subgraph

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



Step 3: Extract a dense subgraph

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



Step 4: Select the most reliable meanings

- “Thomas and Mario are strikers playing in Munich”

Seth Thomas

Mario (Character)

striker (Sport)

Munich (City)

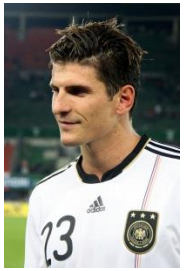
Mario (Album)

Striker (Video Game)

Thomas Müller

FC Bayern Munich

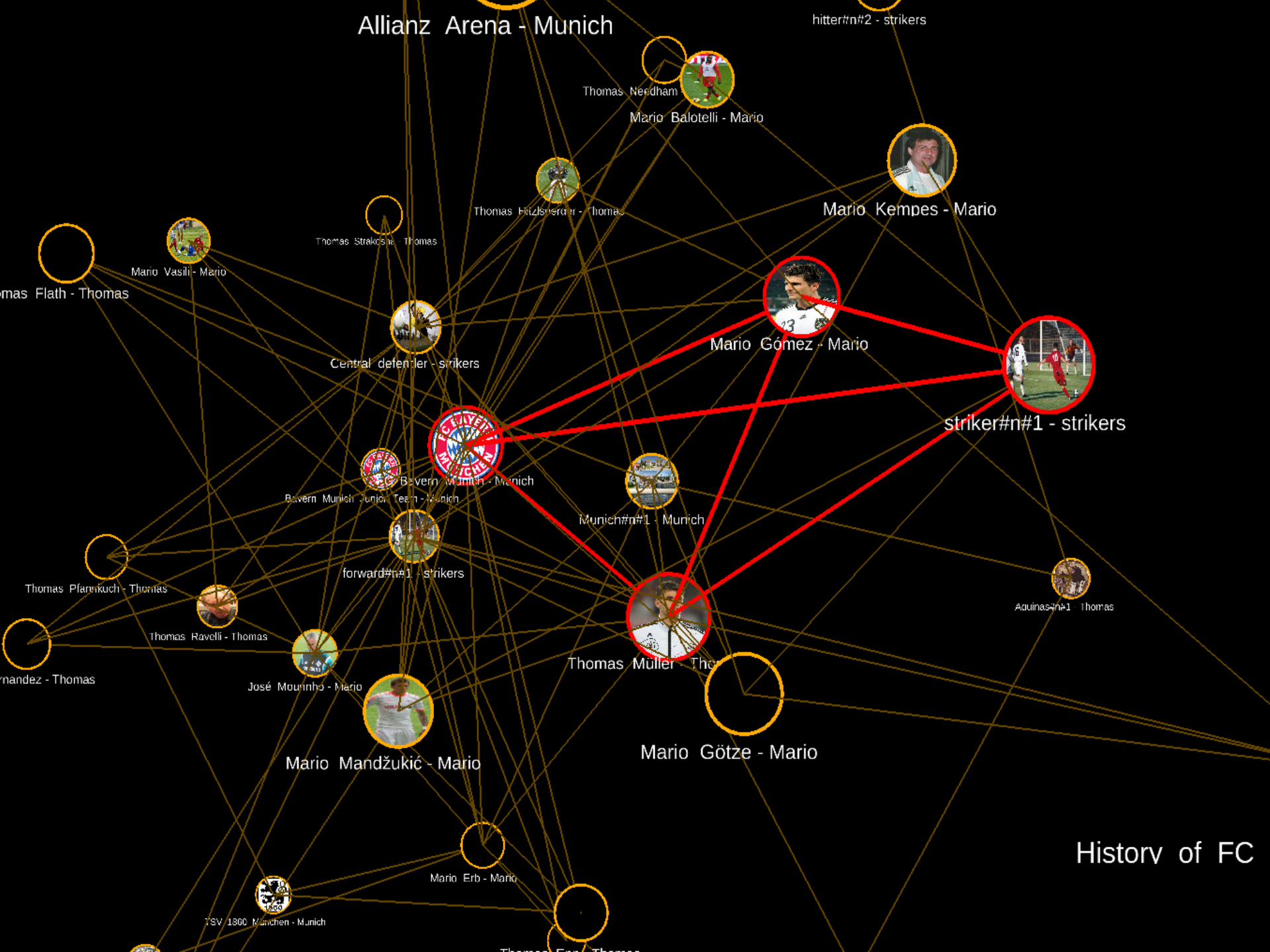
Mario Gómez



Thomas (novel)

Striker (Movie)

Munich (Song)


Allianz Arena - Munich
hitter#n#2 - strikers
Thomas Needham
Mario Balotelli - Mario
Mario Kempes - Mario
Thomas Hitzlsperger - Thomas
Thomas Strakosha - Thomas
Mario Vasili - Mario
mas Flath - Thomas
Central defender - strikers
Mario Gómez - Mario
striker#n#1 - strikers
Bayern Munich - Munich
Bayern Munich Junior Team - Munich
Munich#n#1 - Munich
forward#n#1 - strikers
Thomas Pfannkuch - Thomas
Thomas Ravelli - Thomas
Aquinas#n#1 - Thomas
rnandez - Thomas
José Mourinho - Mario
Thomas Müller - Tho
Mario Mandžukić - Mario
Mario Götze - Mario
Mario Erb - Mario
TSV 1860 München - Munich
Histor v of FC

Experimental Results:
Fine-grained (Multilingual) Disambiguation

Senseval-3

SemEval-2007 task 17

SemEval-2013 task 12

	Sens3	Sem07	SemEval-2013 English			French		German		Italian		Spanish	
System	WN	WN	WN	Wiki	BN	Wiki	BN	Wiki	BN	Wiki	BN	Wiki	BN
Babelfy	68.3	62.7	**65.9**	**87.4**	**69.2**	**71.6**	⋆56.9	81.6	**69.4**	**84.3**	66.6	**83.8**	69.5
IMS	**71.2**	63.3	65.7	–	–	–	–	–	–	–	–	–	–
UKB w2w	⋆65.3	⋆56.0	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	70.3	**65.8**	⋆63.0	⋆80.3	⋆66.5	69.4	45.3	**83.1**	⋆67.4	82.3	57.5	82.4	⋆64.4
Babelfy unif. weights	67.0	65.2	65.0	87.0	68.5	71.9	57.2	81.2	69.8	83.7	66.8	83.8	70.8
Babelfy w/o dens. sub.	68.3	63.3	65.4	87.3	68.7	71.6	57.0	81.7	69.1	84.4	66.5	83.9	69.5
Babelfy only concepts	68.2	62.7	65.5	83.0	68.7	70.2	56.6	79.3	69.3	83.0	66.3	84.0	69.7
Babelfy on sentences	66.0	65.2	63.5	84.0	67.1	70.7	53.6	82.3	68.1	83.8	64.2	83.5	68.7

Experimental Results: KORE50, AIDA-CoNLL

- Two gold-standard Entity Linking datasets:

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	71.8
Illinois Wikifier	41.7	72.4
DBpedia Spotlight	35.4	34.0
Babelfy unif. weights	69.4	81.7
Babelfy w/o dens. sub.	62.5	78.1
Babelfy only NE	68.1	78.8

Babelfy "understands" 'the mouse ate the cheese'!

The mouse ate the cheese .

mouse

Any of numerous small rodents typically resembling diminutive rats having pointed

ate

Take in solid food

cheese

A solid food prepared from the pressed curd of milk

WSD and Entity Linking together win!

I was so **lucky** I could **drive** a **Ferrari Testarossa** !

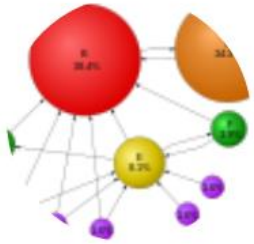
lucky
Occurring by chance

drive
Operate or control a vehicle

Ferrari Testarossa
The Ferrari Testarossa is a 12-cylinder

We **wrote** **PageRank** in **Java** .

wrote
Create code, write a computer program

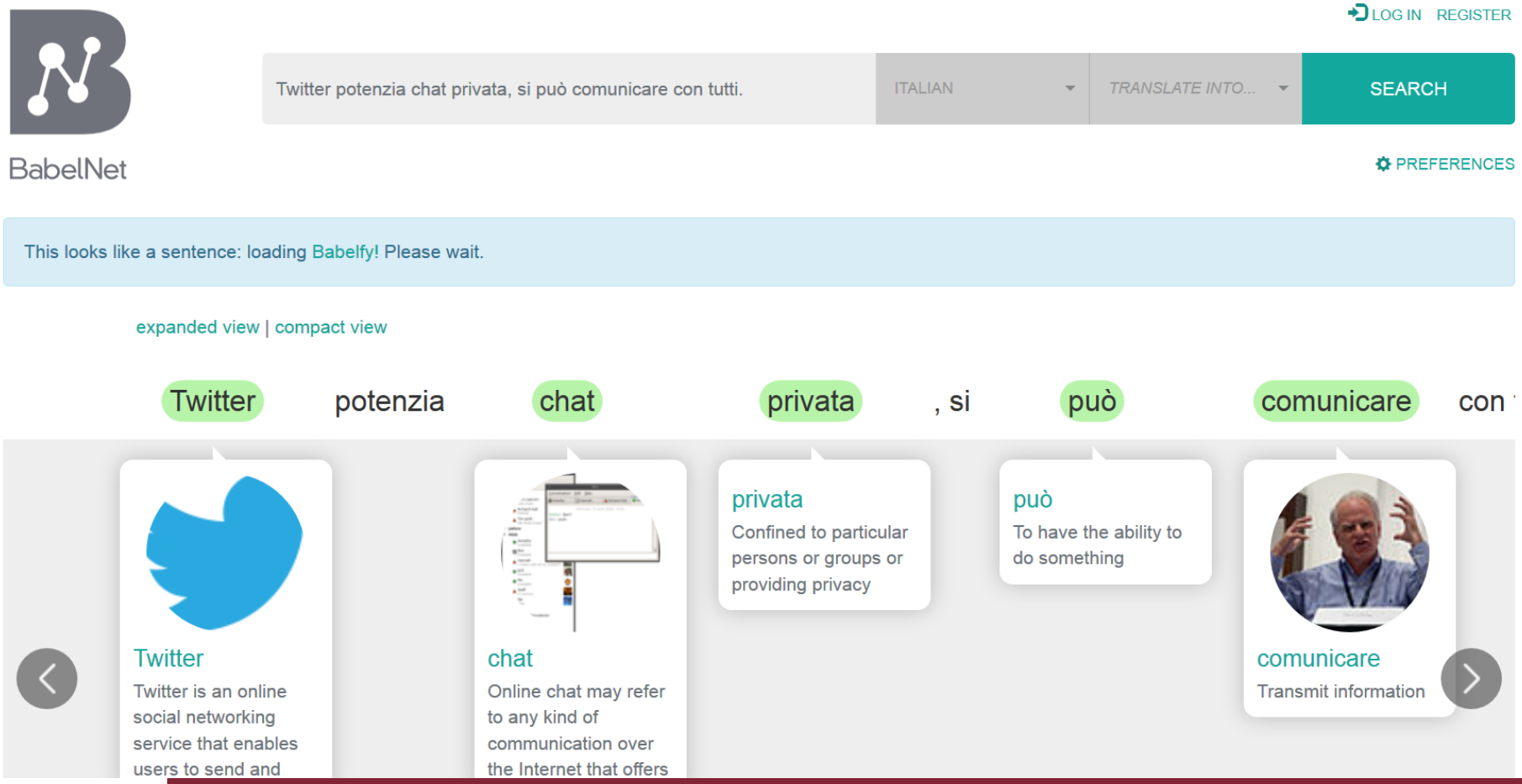


PageRank
PageRank is an algorithm used by Google Search to rank websites in their

Java
A platform-independent object-oriented programming language

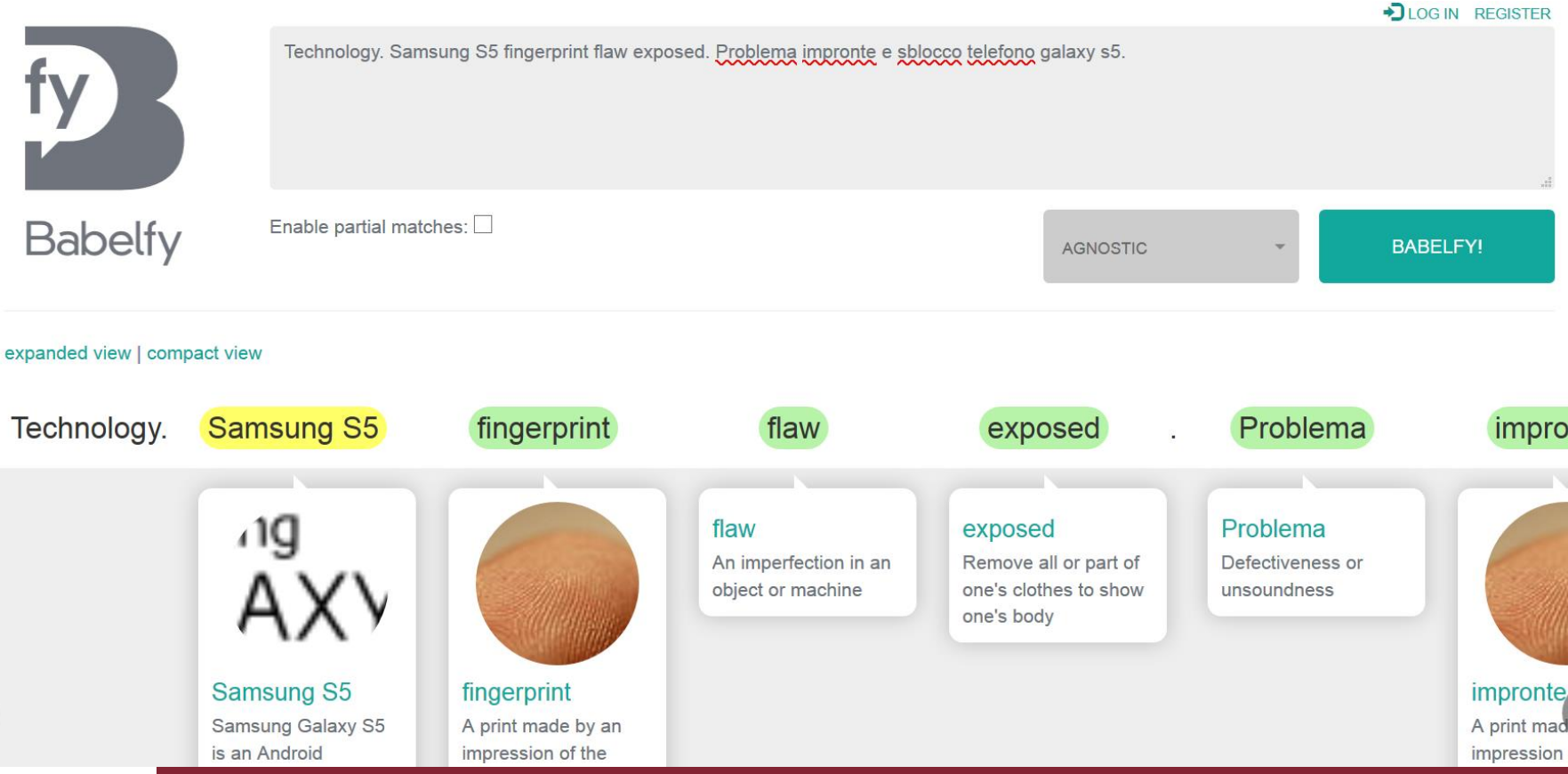
What can we do with Babelfy?

- Disambiguate text written in **any** language!



What can we do with Babelfy?

- Disambiguate in a **language-agnostic** setting!



The Crazy Polyglot!

Live demo (2) – Crazy polyglot!

EN In today's knowledge and information society

FR le paysage lexicographique est plus hétérogène que jamais.

IT Possono le risorse stand-alone competere

ES con múltiples funciones, portale lexicográficas multilingüe y servicios web,

ZH Web服务, 定 制 的 喜 好 和 个 人 用 户 的 个 人 资 料 ？

Coming soon!

- Babelfy 2 with:
  - Improved disambiguation accuracy
  - Use of sense embeddings
  - Use of semantic collocations
  - Improved treatment of Eastern languages
  - Concept and Named Entity extraction and ranking

Summarizing…

NOW...
WHAT WILL YOU DO WITH BABELNET
THIS AFTERNOON?

Thanks or…

(grazie)

SAPIENZA
UNIVERSITÀ DI ROMA

Roberto Navigli

Linguistic Computing Laboratory
http://lcl.uniroma1.it
@RNavigli