Natural Language Processing - 2nd Semester (2023-2024)
1038141

1.10 - Semantics

SAPIENZA
UNIVERSITÀ DI ROMA

Prof. Stefano Faralli
faralli@di.uniroma1.it

Prof. Iacopo Masi
masi@di.uniroma1.it

**credits are reported in the last slide

SAPIENZA
UNIVERSITÀ DI ROMA

1.10 - Semantics

- Semiotic Triangle, From Syntax to Semantics
- Syntax-driven semantic analysis: semantic attachments
- First-Order Logic, lambda calculus for NLP
- Lexical Semantics, lemmatization, word senses, monosemy vs. polysemy, homonymy, metonymy, metaphor, personification, synesthesia, enumerative approach, generative approach.
- WordNet, BabelNet
- Word Sense Disambiguation, Entity Linking, Ontology Learning
- Q&A

Milestones in NLP

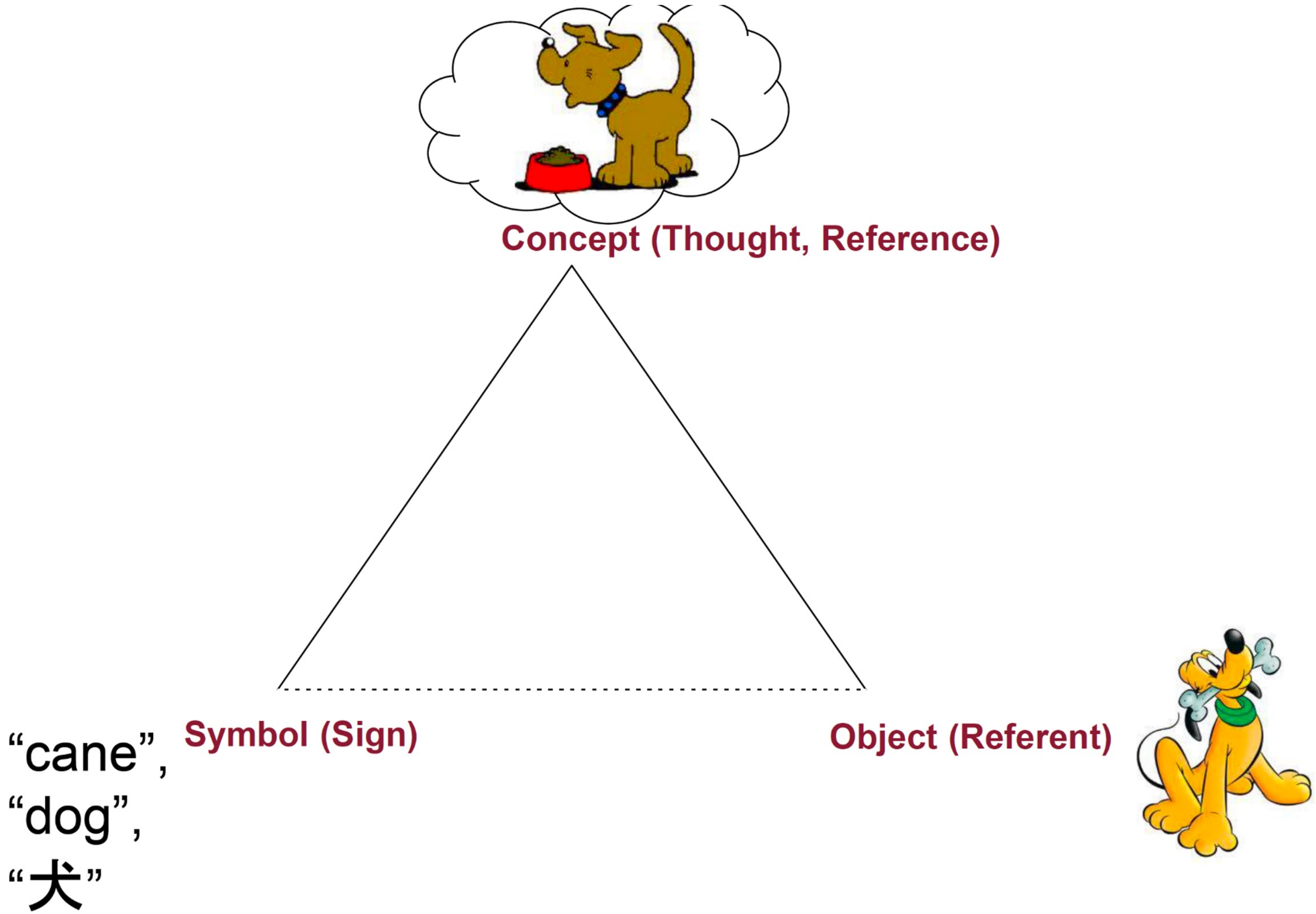
today topic is fundamental for the three areas.

rule-based systems

statistical classical machine learning models

deep learning models

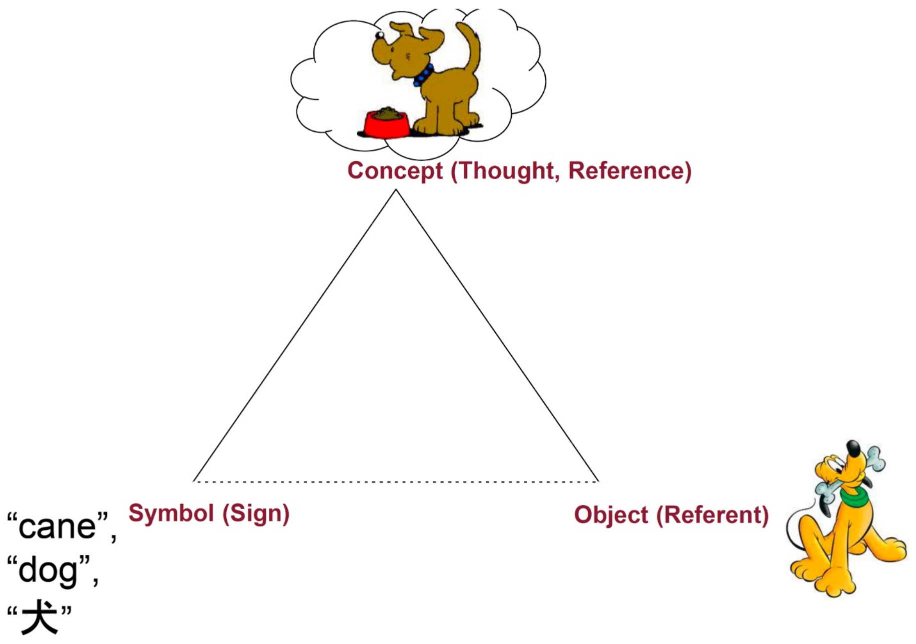
Semantics: Semiotic Triangle (Triangle of meaning)



[Ogden&Richards, 1923]

Semantics: Semiotic Triangle (Triangle of meaning)

The **triangle of meaning** is a model of communication that indicates the **relationship** among a **thought**, **symbol**, and **referent**, <u>and highlights the indirect relationship between the symbol and the referent.</u> The model explains how for any given symbol there can be many different referents, which can lead to misunderstanding.



[Ogden&Richards, 1923]

From Syntax to Semantics

We are now moving from syntax to semantics:

Book the dinner flight

From Syntax to Semantics

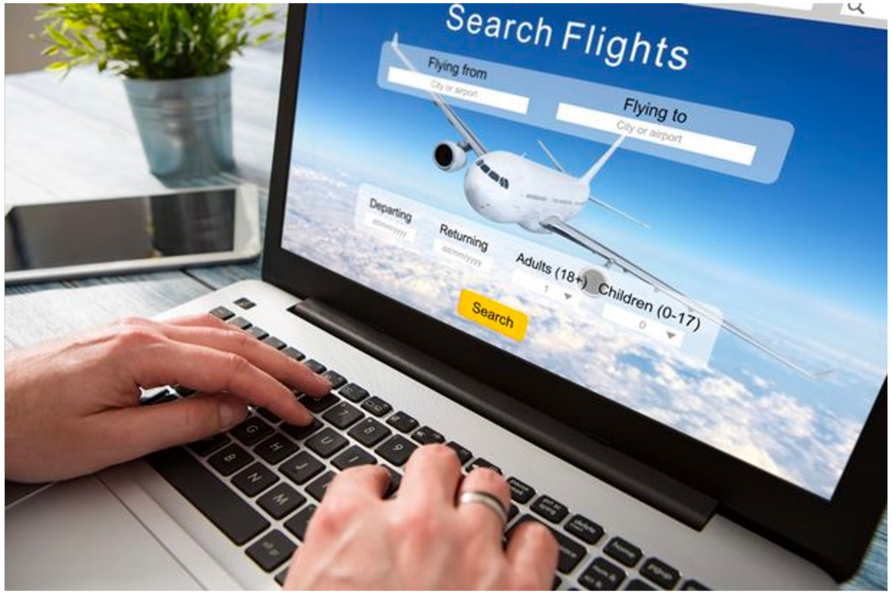
We are now moving from syntax to semantics:

Book the dinner flight

From Syntax to Semantics

We are now moving from syntax to semantics:

Book the dinner flight

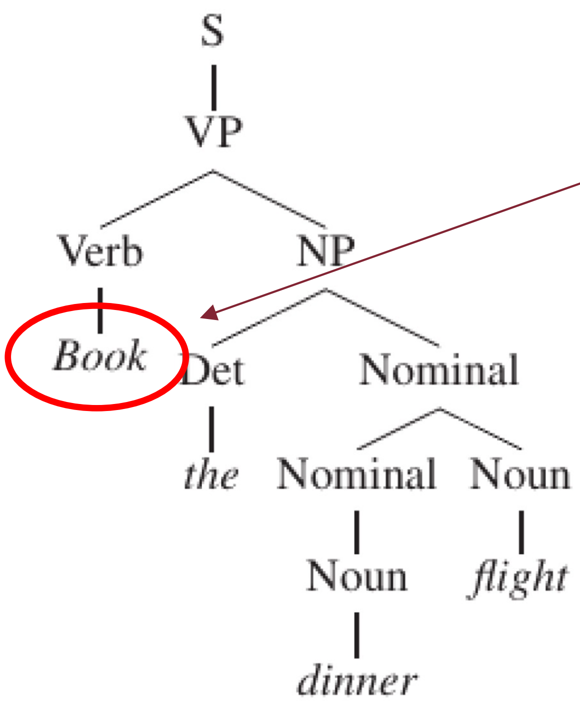


What is the correct semiotic triangle for the word book in this context? -> What is the semantics (the word meaning) for the word book in this context.

From Syntax to Semantics

We are now moving from syntax to semantics:

Book the dinner flight

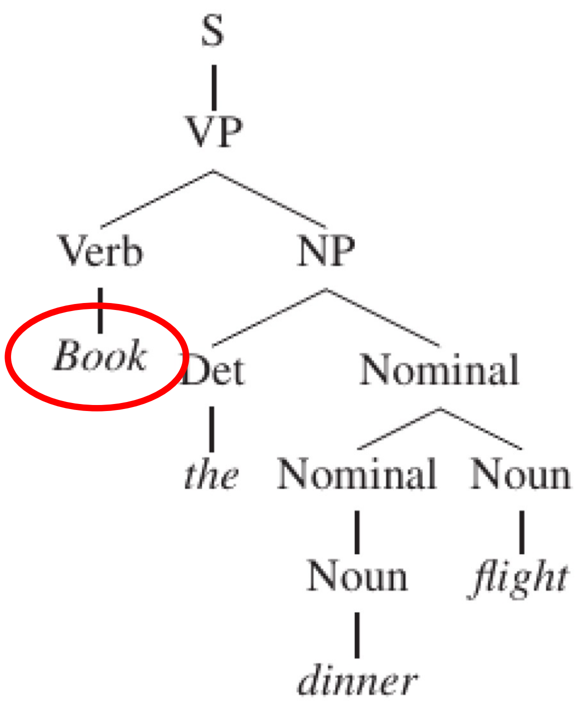


we can understand that, in this context, book is a Verb

From Syntax to Semantics

We are now moving from syntax to semantics:

Book the dinner flight



if we have a sense inventory

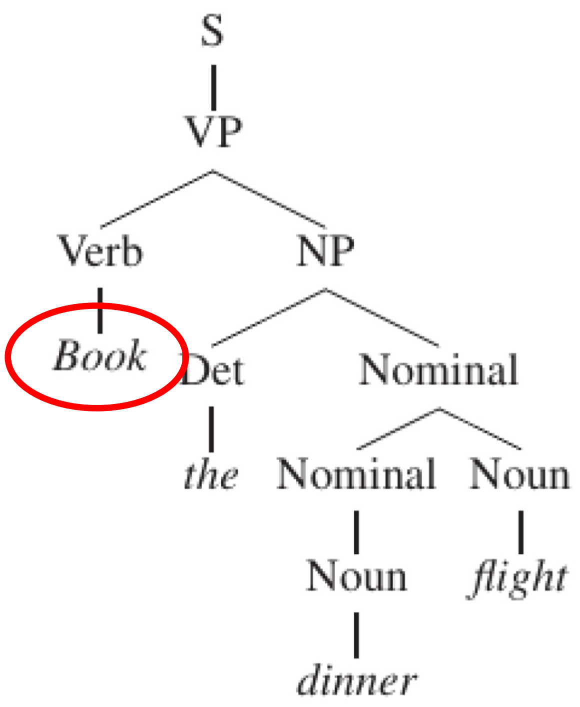
Verb

- S: (v) **book** (engage for a performance) *"Her agent had booked her for several concerts in Tokyo"*
- S: (v) reserve, hold, **book** (arrange for and reserve (something for someone else) in advance) *"reserve me a seat on a flight"; "The agent booked tickets to the show for the whole family"; "please hold a table at Maxim's"*
- S: (v) **book** (record a charge in a police register) *"The policeman booked her when she tried to solicit a man"*
- ~~S: (v) **book** (register in a hotel booker)~~

From Syntax to Semantics

We are now moving from syntax to semantics:

Book the dinner flight



?

if we have a sense inventory

Verb

- S: (v) **book** (engage for a performance) *"Her agent had booked her for several concerts in Tokyo"*
- S: (v) reserve, hold, **book** (arrange for and reserve (something for someone else) in advance) *"reserve me a seat on a flight"; "The agent booked tickets to the show for the whole family"; "please hold a table at Maxim's"*
- S: (v) **book** (record a charge in a police register) *"The policeman booked her when she tried to solicit a man"*
- ~~S: (v) **book** (register in a hotel booker)~~

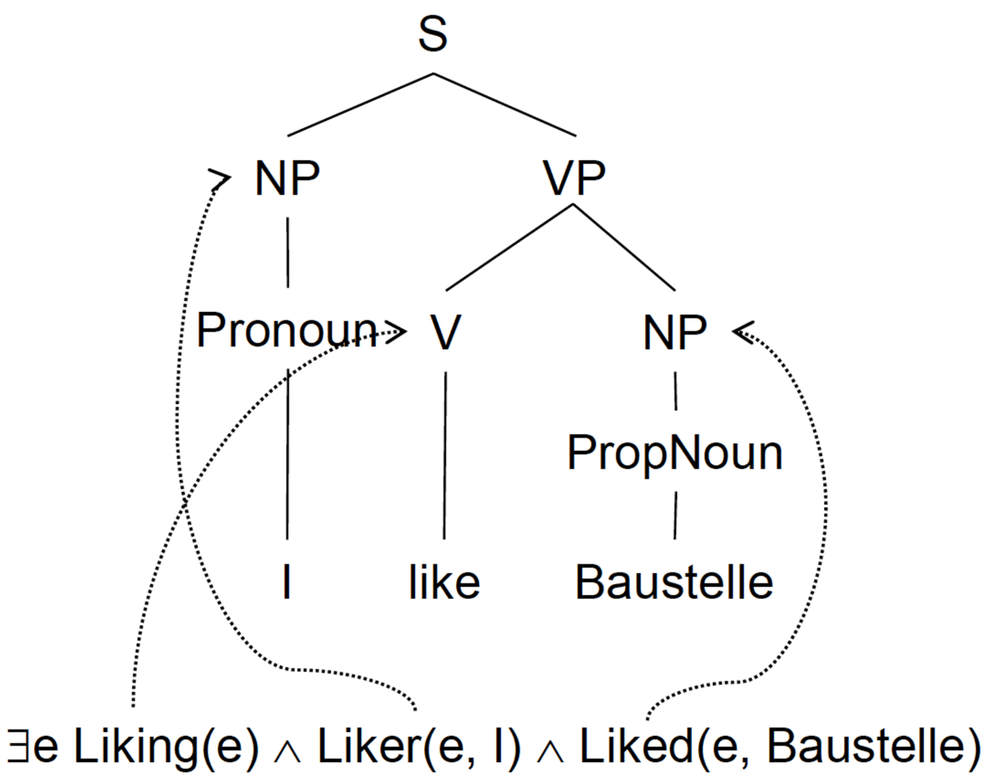
it is a challenging problem to understand the meaning of words in context!

Syntax-driven semantic analysis: Semantic Attachments

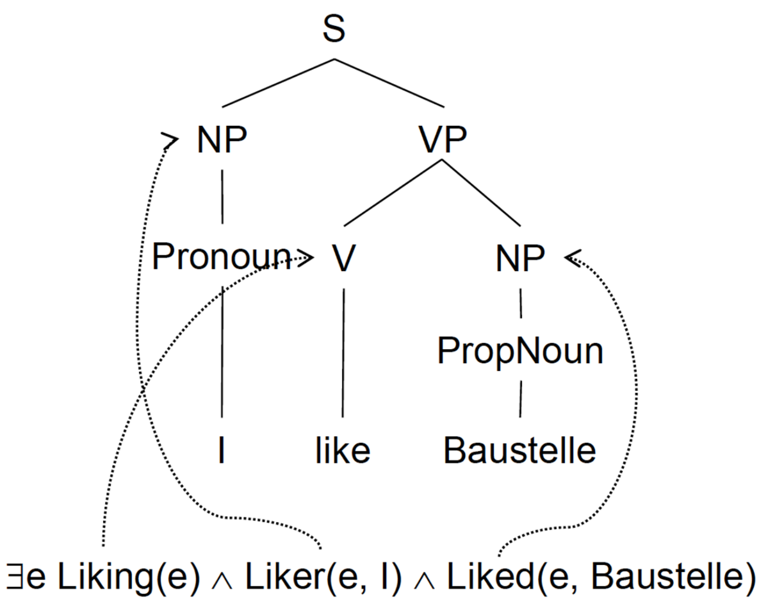
We start from a simplified kind of semantic analysis driven by syntax

–Very complex to use in the real world. Why?

- Example: “I like Baustelle”



Syntax-driven semantic analysis: Semantic Attachments



The assumption here is that we have enough knowledge about this specific example and its parse tree.

But there is an infinite number of such examples and parse trees!

Syntax-driven semantic analysis: Semantic Attachments

We already faced this problem!

E.g., To encode a language, we do not enumerate the strings in the language

- We use a finite formalism that can be used to generate all the strings in the language
- We can do the same for semantics and attach meaning to productions
  - Rule-to-rule hypothesis

Syntax-driven semantic analysis: Semantic Attachments

We can augment CFGs with semantic attachments:

$$A \rightarrow \beta_1 \ldots \beta_n \{ f(\beta_1.sem, \ldots, \beta_n.sem) \}$$

Syntax-driven semantic analysis: Semantic Attachments

We can augment CFGs with semantic attachments:

$$A \rightarrow \beta_1 \ldots \beta_n \{ f(\beta_1.sem, \ldots, \beta_n.sem) \}$$

What can we attach to our context-free rules?

•Our function **f** could be a program fragment

– The semantic interpretation would be obtained as a result of the bottom-up execution of such fragments

– Semantic results would be stored for each non-terminal

–Remember Lex and Yacc?

•But:

– We cannot (immediately) use logical expressions

– How can we design grammars with such programs?

Syntax-driven semantic analysis: Semantic Attachments

One option is to adopt **First-Order Logic (FOL)**,

- you all know what it is…but just to be sure…

$$
\begin{array}{rcl}
\textit{Formula} & \rightarrow & \textit{AtomicFormula} \\
& | & \textit{Formula Connective Formula} \\
& | & \textit{Quantifier Variable}, \ldots \textit{ Formula} \\
& | & \neg \textit{ Formula} \\
& | & (\textit{Formula}) \\
\textit{AtomicFormula} & \rightarrow & \textit{Predicate}(\textit{Term}, \ldots) \\
\textit{Term} & \rightarrow & \textit{Function}(\textit{Term}, \ldots) \\
& | & \textit{Constant} \\
& | & \textit{Variable} \\
\textit{Connective} & \rightarrow & \wedge \mid \vee \mid \Rightarrow \\
\textit{Quantifier} & \rightarrow & \forall \mid \exists \\
\textit{Constant} & \rightarrow & A \mid \textit{VegetarianFood} \mid \textit{Maharani} \cdots \\
\textit{Variable} & \rightarrow & x \mid y \mid \cdots \\
\textit{Predicate} & \rightarrow & \textit{Serves} \mid \textit{Near} \mid \cdots \\
\textit{Function} & \rightarrow & \textit{LocationOf} \mid \textit{CuisineOf} \mid \cdots
\end{array}
$$

Syntax-driven semantic analysis: Semantic Attachments

With such a **First-Order Logic (FOL)** specification we can express facts in the following form:

$$
\begin{array}{rcl}
\textit{Formula} & \rightarrow & \textit{AtomicFormula} \\
& | & \textit{Formula Connective Formula} \\
& | & \textit{Quantifier Variable}, \ldots\ \textit{Formula} \\
& | & \neg\ \textit{Formula} \\
& | & (\textit{Formula}) \\
\textit{AtomicFormula} & \rightarrow & \textit{Predicate}(\textit{Term}, \ldots) \\
\textit{Term} & \rightarrow & \textit{Function}(\textit{Term}, \ldots) \\
& | & \textit{Constant} \\
& | & \textit{Variable} \\
\textit{Connective} & \rightarrow & \wedge \mid \vee \mid \Rightarrow \\
\textit{Quantifier} & \rightarrow & \forall \mid \exists \\
\textit{Constant} & \rightarrow & A \mid \textit{VegetarianFood} \mid \textit{Maharani} \cdots \\
\textit{Variable} & \rightarrow & x \mid y \mid \cdots \\
\textit{Predicate} & \rightarrow & \textit{Serves} \mid \textit{Near} \mid \cdots \\
\textit{Function} & \rightarrow & \textit{LocationOf} \mid \textit{CuisineOf} \mid \cdots
\end{array}
$$

Le Scalette's location is Nemi

LocationOf(LeScalette, Nemi)

Le Scalette is a restaurant and Le Scalette's location is Nemi

Restaurant(LeScalette) ∧ LocationOf(LeScalette, Nemi)

There exists a restaurant in Nemi?

∃x Restaurant(x) ∧ LocationOf(x, Nemi)

All pizzerias serve pizzas

∀x Pizzeria(x) ⇒ Serves(x, Pizza)

Syntax-driven semantic analysis: Semantic Attachments

In order to abstract away from fully specified FOL, we can use the **lambda notation**

It extends the syntax of FOL to include expressions of the form:

$$\lambda x.P(x)$$

You can apply these expressions to other logical terms to yield new FOL expressions

– For instance: $(\lambda x.Restaurant(x))(LeScalette)$ -> $Restaurant(LeScalette)$

This operation is called **beta reduction**

Lambda Calculus

Alonzo Church, 1930s

- Lambda Calculus consists of lambda terms:
  – **Variables:** A variable **x** is a **lambda term**

  – **Abstraction:** If **t** is a lambda term and **x** a variable, **λx.t** is a lambda term:
  - λx represents an anonymous function with a given input x
  - λ binds x in t

  – **Application:**if **t** and **s** are lambda terms, then **ts** is a lambda term

- …and extra operations:
  – **Alpha conversion:** bound variables can be renamed

  – **Beta reduction:**an application of the form (λx.t)s reduces to the term t[x := s]

Lambda Calculus: examples

- Identity: λx.x
- Constant: λx.y
- Square function: λx.x*x
  - What is the square of 4? (λx.x*x) 4 = 4*4
- Multiple-argument functions: λx.λy.x*y

Lambda Calculus: Linguistic examples

- We need constants to build phrases:
  two, red, loves, ate, John, Mary, etc.

- Connectives and quantifiers:
  or, and, not, if … then, for all, there exists

- Examples:
  loves(John, Mary)

  $\exists$ m.mozzarella(m) $\wedge$ ate(John, m)

  $\exists$ m.mozzarella(m) $\wedge$ blue(m) $\wedge$ ate(John, m)

Lambda Calculus: modelings events

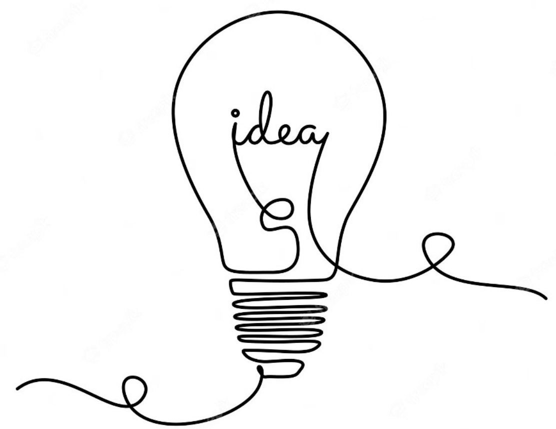
If something happens in a specific moment or period of time, we have an event.

Example:

$\exists e.past(e) \land act(e, eating) \land eater(e, John) \land \exists m (mozzarella(m) \land eatee(e, m))$

How to Get Semantics from sentences

1. Parse sentences into parse trees
2. Get the semantics for each word
3. Use the parse tree to build the semantics bottom-up (from words to more complex constituents)

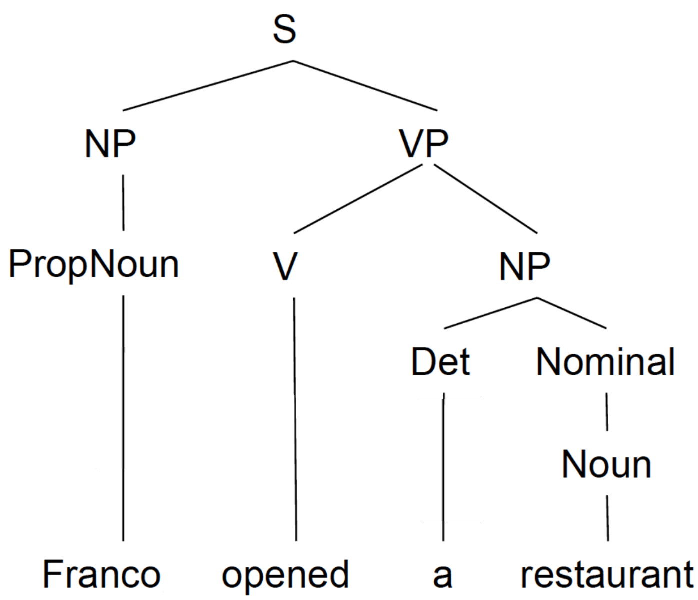


An example excerpt

Grammar Rule	Semantic Attachment
$S \rightarrow NP\ VP$	$\{NP.sem(VP.sem)\}$
$NP \rightarrow Det\ Nominal$	$\{Det.sem(Nominal.sem)\}$
$NP \rightarrow ProperNoun$	$\{ProperNoun.sem\}$
$Nominal \rightarrow Noun$	$\{Noun.sem\}$
$VP \rightarrow Verb$	$\{Verb.sem\}$
$VP \rightarrow Verb\ NP$	$\{Verb.sem(NP.sem)\}$
$Det \rightarrow every$	$\{\lambda P.\lambda Q.\forall x P(x) \Rightarrow Q(x)\}$
$Det \rightarrow a$	$\{\lambda P.\lambda Q.\exists x P(x) \wedge Q(x)\}$
$Noun \rightarrow restaurant$	$\{\lambda r.Restaurant(r)\}$
$ProperNoun \rightarrow Matthew$	$\{\lambda m.m(Matthew)\}$
$ProperNoun \rightarrow Franco$	$\{\lambda f.f(Franco)\}$
$ProperNoun \rightarrow Frasca$	$\{\lambda f.f(Frasca)\}$
$Verb \rightarrow closed$	$\{\lambda x.\exists e Closing(e) \wedge Closed(e,x)\}$
$Verb \rightarrow opened$	$\{\lambda w.\lambda z.w(\lambda x.\exists e Opening(e) \wedge Opener(e,z)$ $\wedge Opened(e,x))$

An example excerpt

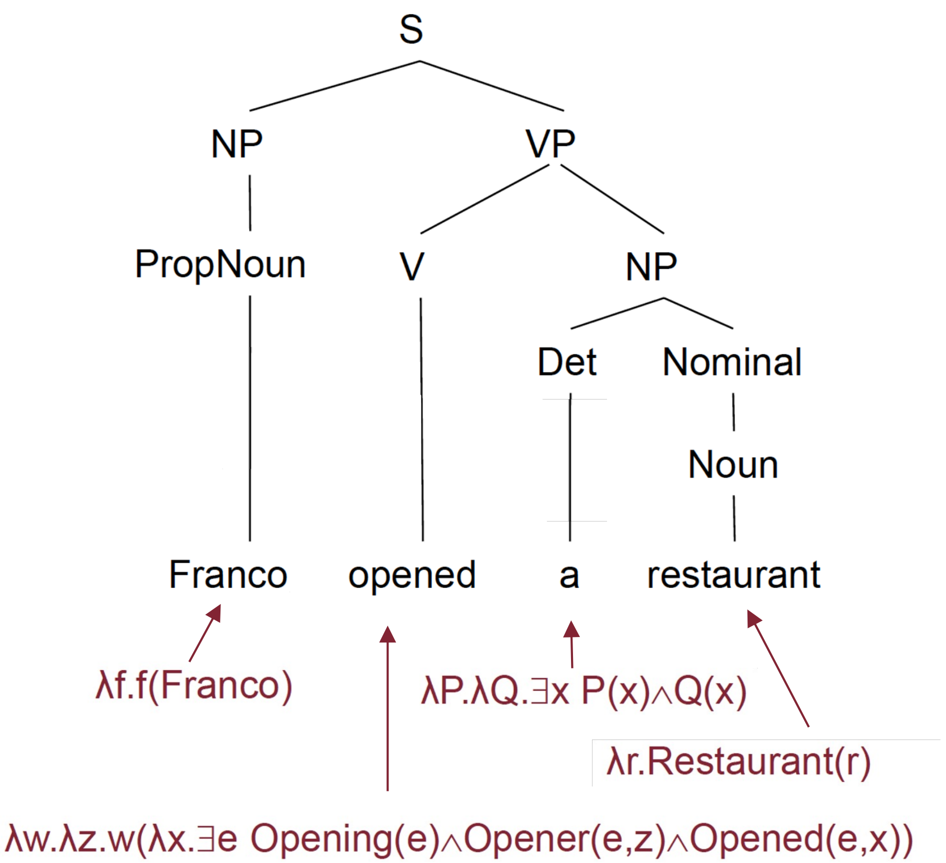
Franco opened a restaurant



Grammar Rule	Semantic Attachment
$S \rightarrow NP\ VP$	$\{NP.sem(VP.sem)\}$
$NP \rightarrow Det\ Nominal$	$\{Det.sem(Nominal.sem)\}$
$NP \rightarrow ProperNoun$	$\{ProperNoun.sem\}$
$Nominal \rightarrow Noun$	$\{Noun.sem\}$
$VP \rightarrow Verb$	$\{Verb.sem\}$
$VP \rightarrow Verb\ NP$	$\{Verb.sem(NP.sem)\}$
$Det \rightarrow every$	$\{\lambda P.\lambda Q.\forall x P(x) \Rightarrow Q(x)\}$
$Det \rightarrow a$	$\{\lambda P.\lambda Q.\exists x P(x) \wedge Q(x)\}$
$Noun \rightarrow restaurant$	$\{\lambda r.Restaurant(r)\}$
$ProperNoun \rightarrow Matthew$	$\{\lambda m.m(Matthew)\}$
$ProperNoun \rightarrow Franco$	$\{\lambda f.f(Franco)\}$
$ProperNoun \rightarrow Frasca$	$\{\lambda f.f(Frasca)\}$
$Verb \rightarrow closed$	$\{\lambda x.\exists e Closing(e) \wedge Closed(e,x)\}$
$Verb \rightarrow opened$	$\{\lambda w.\lambda z.w(\lambda x.\exists e Opening(e) \wedge Opener(e,z) \wedge Opened(e,x))\}$

An example excerpt

Franco opened a restaurant

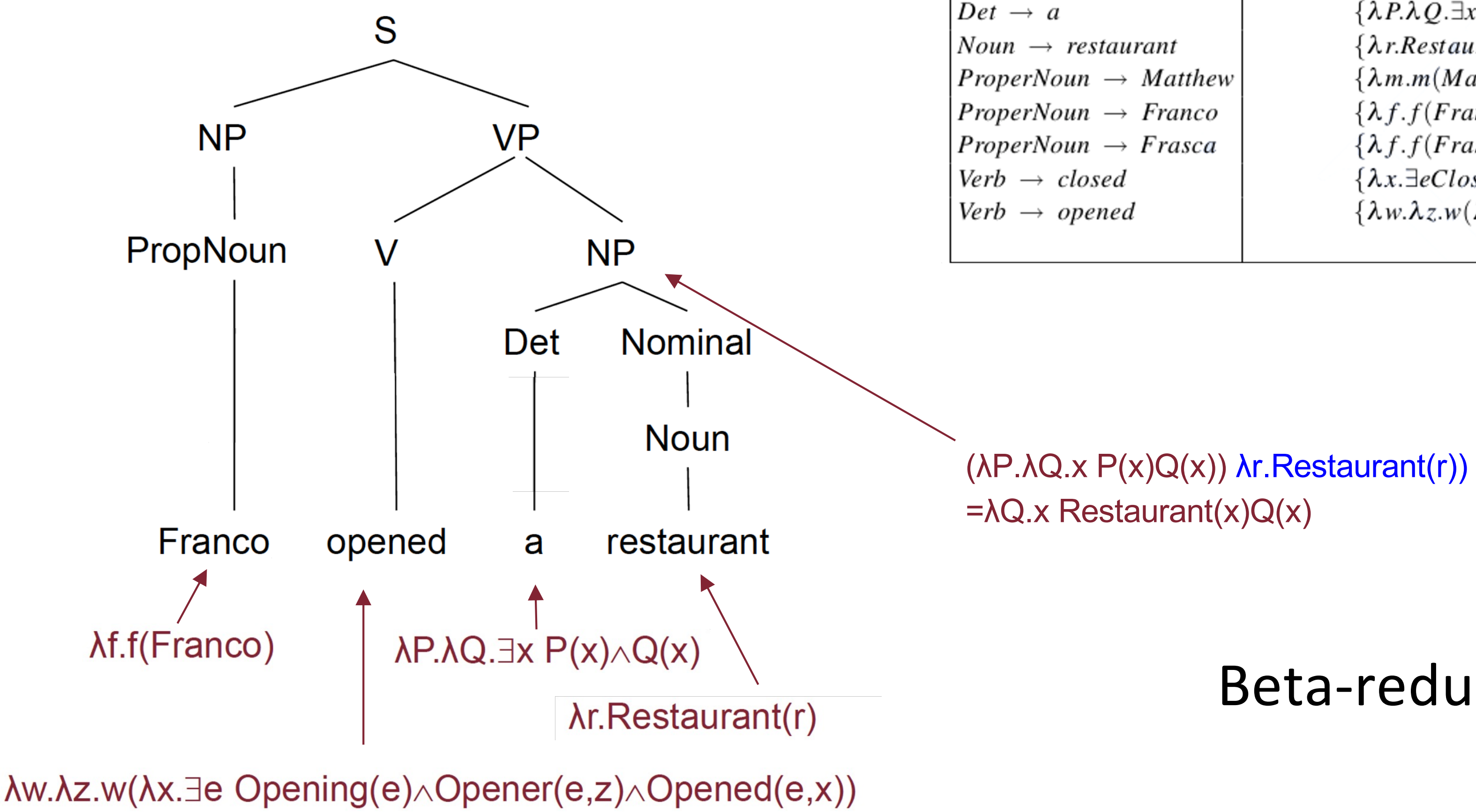


Grammar Rule	Semantic Attachment
$S \rightarrow NP\ VP$	$\{NP.sem(VP.sem)\}$
$NP \rightarrow Det\ Nominal$	$\{Det.sem(Nominal.sem)\}$
$NP \rightarrow ProperNoun$	$\{ProperNoun.sem\}$
$Nominal \rightarrow Noun$	$\{Noun.sem\}$
$VP \rightarrow Verb$	$\{Verb.sem\}$
$VP \rightarrow Verb\ NP$	$\{Verb.sem(NP.sem)\}$
$Det \rightarrow every$	$\{\lambda P.\lambda Q.\forall x P(x) \Rightarrow Q(x)\}$
$Det \rightarrow a$	$\{\lambda P.\lambda Q.\exists x P(x) \wedge Q(x)\}$
$Noun \rightarrow restaurant$	$\{\lambda r.Restaurant(r)\}$
$ProperNoun \rightarrow Matthew$	$\{\lambda m.m(Matthew)\}$
$ProperNoun \rightarrow Franco$	$\{\lambda f.f(Franco)\}$
$ProperNoun \rightarrow Frasca$	$\{\lambda f.f(Frasca)\}$
$Verb \rightarrow closed$	$\{\lambda x.\exists e Closing(e) \wedge Closed(e,x)\}$
$Verb \rightarrow opened$	$\{\lambda w.\lambda z.w(\lambda x.\exists e Opening(e) \wedge Opener(e,z) \wedge Opened(e,x))$

An example excerpt

Grammar Rule	Semantic Attachment
$S \rightarrow NP\ VP$	$\{NP.sem(VP.sem)\}$
$NP \rightarrow Det\ Nominal$	$\{Det.sem(Nominal.sem)\}$
$NP \rightarrow ProperNoun$	$\{ProperNoun.sem\}$
$Nominal \rightarrow Noun$	$\{Noun.sem\}$
$VP \rightarrow Verb$	$\{Verb.sem\}$
$VP \rightarrow Verb\ NP$	$\{Verb.sem(NP.sem)\}$
$Det \rightarrow every$	$\{\lambda P.\lambda Q.\forall x P(x) \Rightarrow Q(x)\}$
$Det \rightarrow a$	$\{\lambda P.\lambda Q.\exists x P(x) \wedge Q(x)\}$
$Noun \rightarrow restaurant$	$\{\lambda r.Restaurant(r)\}$
$ProperNoun \rightarrow Matthew$	$\{\lambda m.m(Matthew)\}$
$ProperNoun \rightarrow Franco$	$\{\lambda f.f(Franco)\}$
$ProperNoun \rightarrow Frasca$	$\{\lambda f.f(Frasca)\}$
$Verb \rightarrow closed$	$\{\lambda x.\exists e Closing(e) \wedge Closed(e,x)\}$
$Verb \rightarrow opened$	$\{\lambda w.\lambda z.w(\lambda x.\exists e Opening(e) \wedge Opener(e,z)$ $\wedge Opened(e,x))$

Franco opened a restaurant

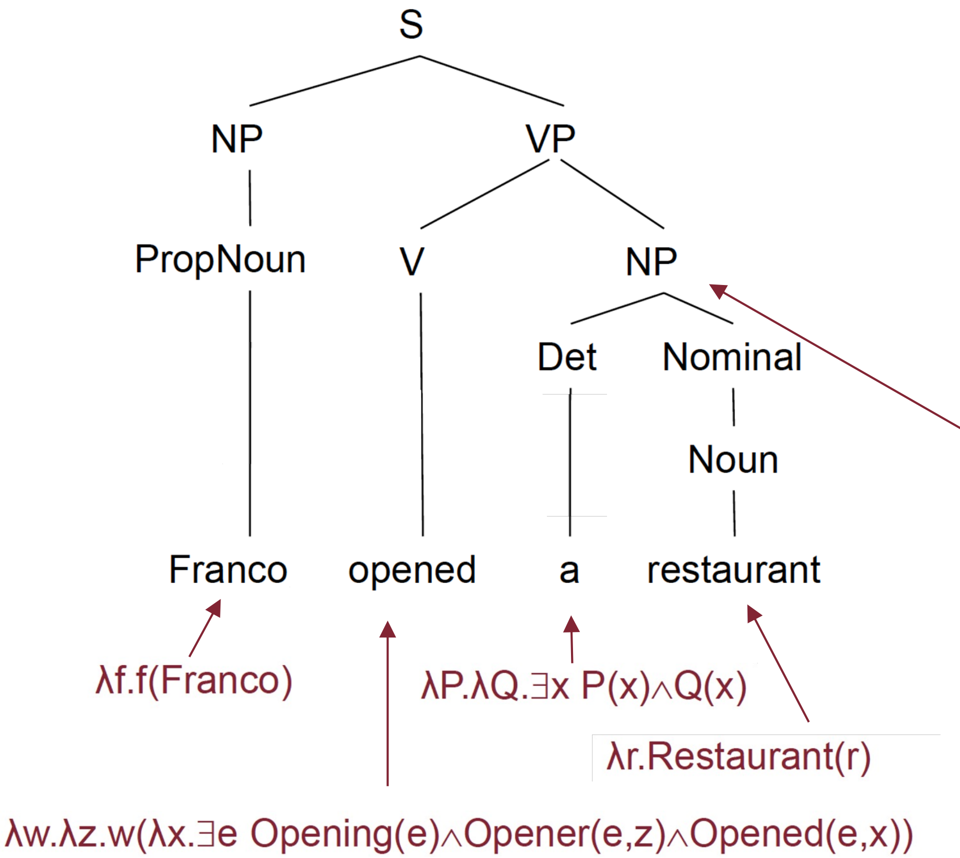


Beta-reduction

An example excerpt

Grammar Rule	Semantic Attachment
$S \rightarrow NP\ VP$	$\{NP.sem(VP.sem)\}$
$NP \rightarrow Det\ Nominal$	$\{Det.sem(Nominal.sem)\}$
$NP \rightarrow ProperNoun$	$\{ProperNoun.sem\}$
$Nominal \rightarrow Noun$	$\{Noun.sem\}$
$VP \rightarrow Verb$	$\{Verb.sem\}$
$VP \rightarrow Verb\ NP$	$\{Verb.sem(NP.sem)\}$
$Det \rightarrow every$	$\{\lambda P.\lambda Q.\forall x P(x) \Rightarrow Q(x)\}$
$Det \rightarrow a$	$\{\lambda P.\lambda Q.\exists x P(x) \wedge Q(x)\}$
$Noun \rightarrow restaurant$	$\{\lambda r.Restaurant(r)\}$
$ProperNoun \rightarrow Matthew$	$\{\lambda m.m(Matthew)\}$
$ProperNoun \rightarrow Franco$	$\{\lambda f.f(Franco)\}$
$ProperNoun \rightarrow Frasca$	$\{\lambda f.f(Frasca)\}$
$Verb \rightarrow closed$	$\{\lambda x.\exists e Closing(e) \wedge Closed(e,x)\}$
$Verb \rightarrow opened$	$\{\lambda w.\lambda z.w(\lambda x.\exists e Opening(e) \wedge Opener(e,z)$ $\wedge Opened(e,x))$

Franco opened a restaurant

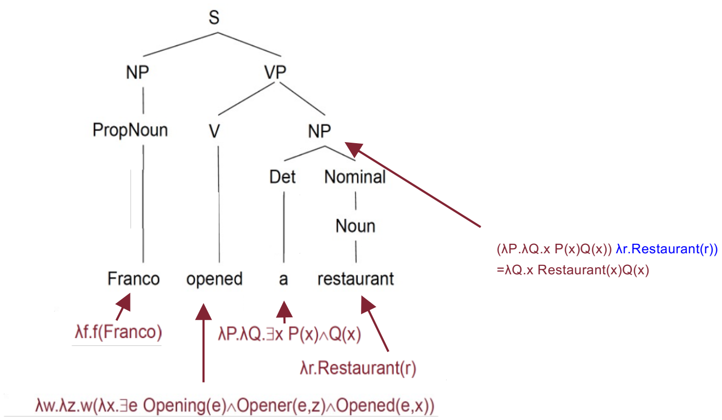


(λP.λQ.x P(x)Q(x)) λr.Restaurant(r))
=λQ.x Restaurant(x)Q(x)

Beta-reduction

and so on …

An example excerpt



Grammar Rule	Semantic Attachment
$S \rightarrow NP\ VP$	$\{NP.sem(VP.sem)\}$
$NP \rightarrow Det\ Nominal$	$\{Det.sem(Nominal.sem)\}$
$NP \rightarrow ProperNoun$	$\{ProperNoun.sem\}$
$Nominal \rightarrow Noun$	$\{Noun.sem\}$
$VP \rightarrow Verb$	$\{Verb.sem\}$
$VP \rightarrow Verb\ NP$	$\{Verb.sem(NP.sem)\}$
$Det \rightarrow every$	$\{\lambda P.\lambda Q.\forall x P(x) \Rightarrow Q(x)\}$
$Det \rightarrow a$	$\{\lambda P.\lambda Q.\exists x P(x) \wedge Q(x)\}$
$Noun \rightarrow restaurant$	$\{\lambda r.Restaurant(r)\}$
$ProperNoun \rightarrow Matthew$	$\{\lambda m.m(Matthew)\}$
$ProperNoun \rightarrow Franco$	$\{\lambda f.f(Franco)\}$
$ProperNoun \rightarrow Frasca$	$\{\lambda f.f(Frasca)\}$
$Verb \rightarrow closed$	$\{\lambda x.\exists e Closing(e) \wedge Closed(e,x)\}$
$Verb \rightarrow opened$	$\{\lambda w.\lambda z.w(\lambda x.\exists e Opening(e) \wedge Opener(e,z) \wedge Opened(e,x))$

Sentence:

Franco opened a restaurant

Semantics:

$$\exists x\ Restaurant(x) \wedge \exists e\ Opening(e) \wedge Opener(e,Franco) \wedge Opened(e,x)$$

Side-effects

1. **It's difficult!**

Side-effects

1. **It's difficult!**

1. **More dangerous:**

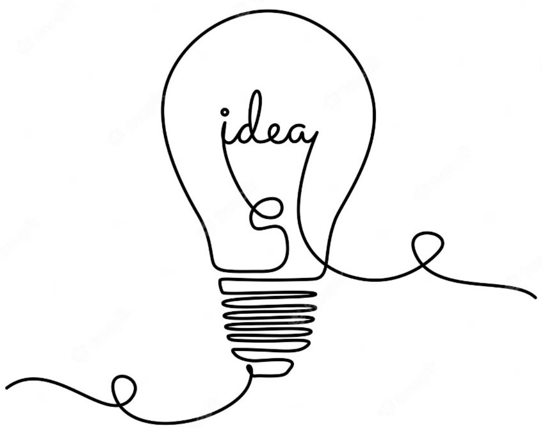
   We made a **simplifying assumption** that each terminal (e.g., Opening, Franco, etc.) represents one **meaning**!

Side-effects

1. **It’s difficult!**

1. **More dangerous:**

   We made a **simplifying assumption** that each terminal (e.g., Opening, Franco, etc.) represents one **meaning**!



Side-effects

1. **It's difficult!**

1. **More dangerous:**

   We made a **simplifying assumption** that each terminal (e.g., Opening, Franco, etc.) represents one **meaning**!

   We need to step back to the **meaning of words** before trying to cope with the **meaning of sentences**.

Lexical-semantics

Lexical-semantics

We need to step back to the **meaning of words** before trying to cope with the **meaning of sentences**.

Some definitions:

– Lexicon: a finite list of given language words

– Lemma: the grammatical base form used in dictionaries

(e.g., peach is the lemma of peaches)

– Word form: any grammatical form of a word

(e.g. eat, eats, ate, etc.)

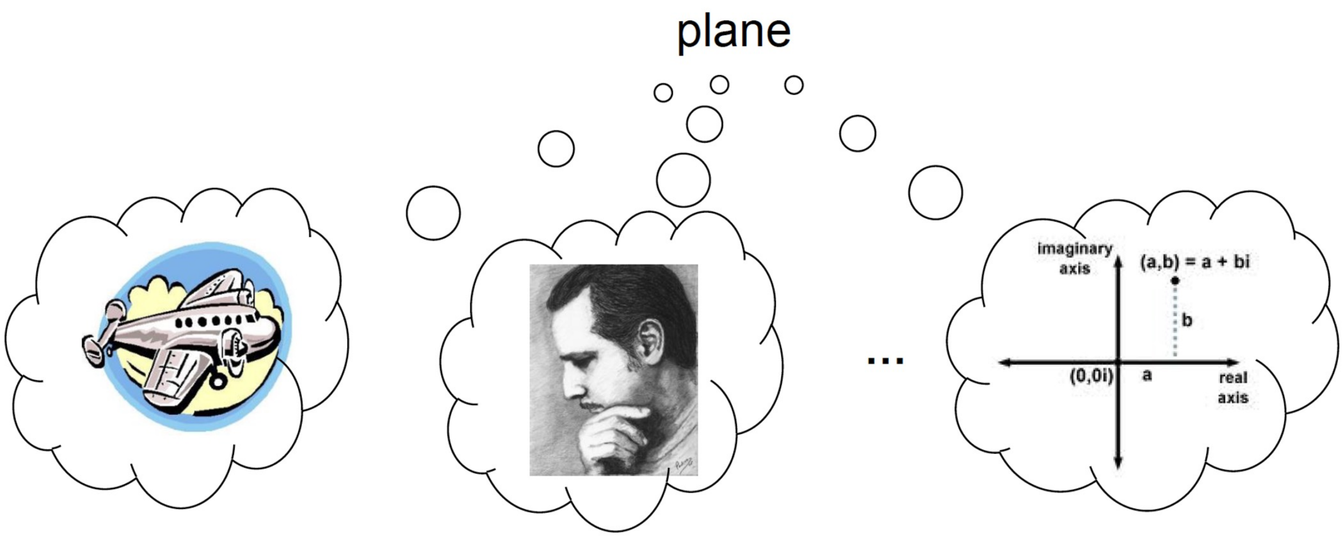
Lexical-semantics: lemmatization

The process of converting word forms to their lemmas is called lemmatization

- Not always deterministic, it may depend on the context
  - It **bound** the parties to observe neutrality
    - Lemma: bind (constrain)
  - The horses were found to **bound** across the meadows
    - Lemma: bound (jumping)
- Plus: lemmas are different across parts of speech:
  - They closed the **purchase** with a handshake
  - They are about to **purchase** a flat

Word Senses

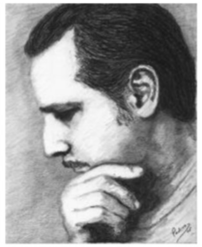
- The meaning of a word depends on the context in which it occurs.
- We call each meaning of a word a word sense.



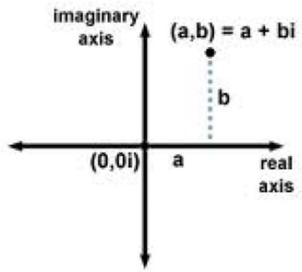
Word Senses in context: the “plane” example

- I am catching the earliest **plane** to Brussels.

- This area probably lies more on the spiritual **plane** than the mental one.



- Let’s represent three-dimensional structures on a two-dimensional **plane**

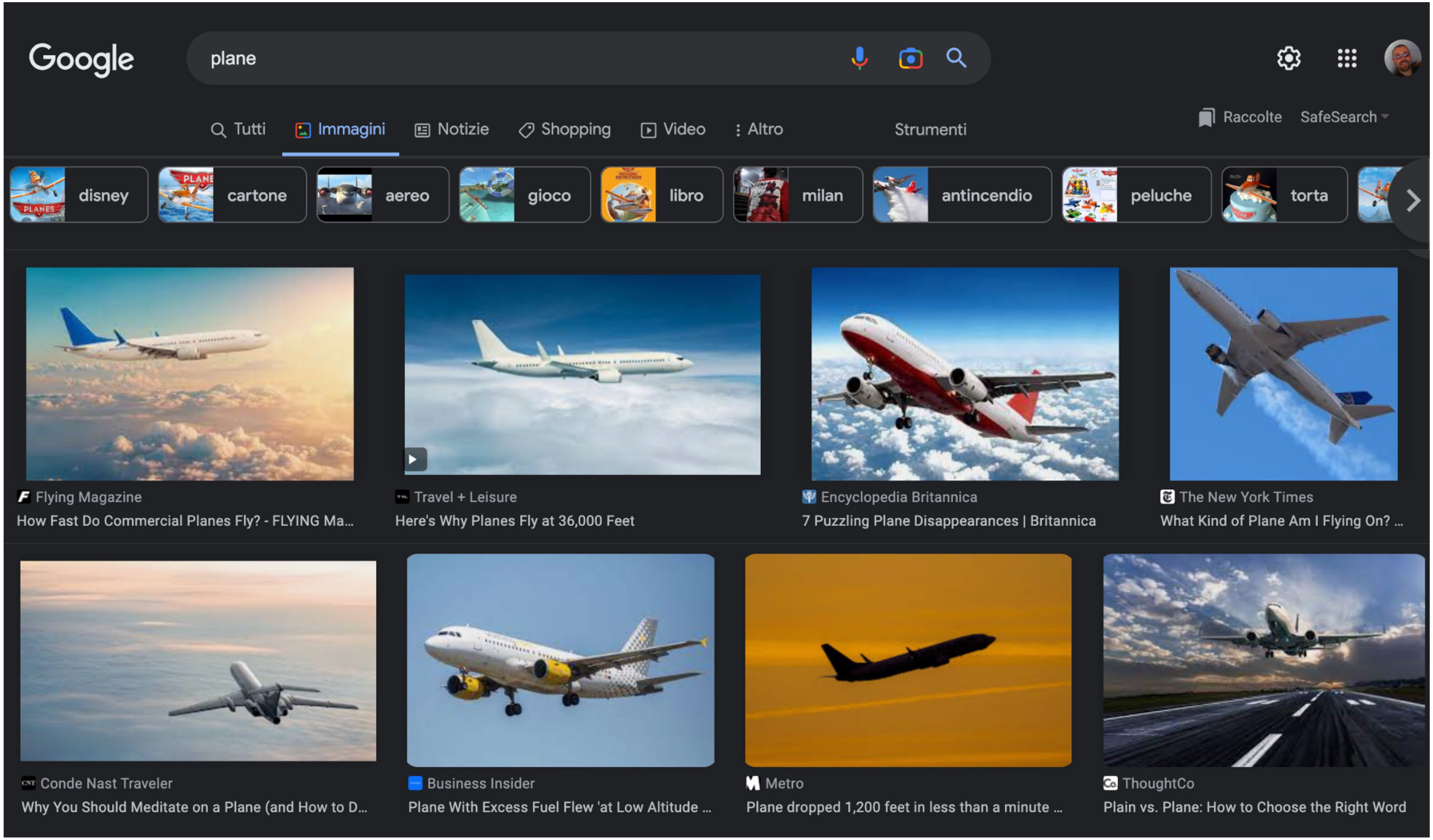


Monosemy vs. Polysemy

- A word is monosemous if it has a single sense
  - Internet
  - plant life
  - …

- Polysemous otherwise
  - bank
  - plane
  - bass
  - …

How do state-of-the-art applications deal with senses?



Homonymy

- That **bank**[1] holds the mortage on my home.

- He sat on the **bank**[2] of the river and watched the currents.

- The **bank**[3] is on the corner of New Oxford Street.

- A huge **bank**[4] of earth was found close to the entrance.

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

Metonymy

- Metonymy: referring to one aspect of a thing instead of the thing itself or vice versa
  - Have another **glass**

Metaphor

- Metaphor: referring to something which is similar to what is meant
  - An **ocean** of trouble

Personification

- Personification (pathetic fallacy): assigning human features to things or animals
  - The **crying** fountain

Synesthesia

- Synesthesia: referring to a sensation in a sense modality different from the normal expected one
  - The **black** yell

Sense Distinctions

The problem is about the word sense representation:

- How to represent word senses?
  - Can we enumerate the senses of a word?

{  ,  ,  , ? }

  - Can we generate the senses of a word?
    - E.g. kiwi is a fruit, the purpose of kiwi is to be eaten, flesh is part of a kiwi, etc.

"**Kiwi** is my mother tongue, but I also speak all other English languages"

Word Senses: the enumerative approach

A fixed sense inventory enumerates the range of all possible meanings of a word.

Word Senses: the enumerative approach

A fixed sense inventory enumerates the range of all possible meanings of a word.

Example: *knife*

1. a cutting tool composed of a blade with a sharp point and a handle
2. an instrument with a handle and blade with a sharp point used as a weapon

Word Senses: the enumerative approach

A fixed sense inventory enumerates the range of all possible meanings of a word.

Example: *knife*

1. a cutting tool composed of a blade with a sharp point and a handle
2. an instrument with a handle and blade with a sharp point used as a weapon

Words in context are assumed to select/activate one of these senses in each context

Word Senses: the enumerative approach

A fixed sense inventory enumerates the range of all possible meanings of a word.

Example: *knife*

1. a cutting tool composed of a blade with a sharp point and a handle
2. an instrument with a handle and blade with a sharp point used as a weapon

Words in context are assumed to select/activate one of these senses in each context

- She chopped the vegetables with a chef's knife
  ➠ KNIFE as a **CUTTING TOOL**
- A man was beaten and cut with a knife.
  ➠ KNIFE as a **WEAPON**

Word Senses: the enumerative approach

A fixed sense inventory enumerates the range of all possible meanings of a word.

Example: *knife*

1. a cutting tool composed of a blade with a sharp point and a handle
2. an instrument with a handle and blade with a sharp point used as a weapon

? should we add a further sense to the inventory for "a cutting blade forming part of a machine"

? are word senses application-independent

Graded word sense assignment [Erk and McCarthy, 2009]

- Similar to category membership in the human mind, the meanings of words in context do not always have clear-cut boundaries
- Given a target word, annotators judge the applicability of each possible word sense on a graded scale

	Senses							
Sentence	1	2	3	4	5	6	7	Annotator
This question provoked **arguments** in America about the Norton Anthology of Literature by Women, some of the contents of which were said to have had little value as literature.	1	4	4	2	1	1	3	Ann. 1
	4	5	4	2	1	1	4	Ann. 2
	1	4	5	1	1	1	1	Ann. 3

Word Senses: the Generative Approach [Pustejovsky, 1991]

Senses are generated from rules which capture regularities in the creation of senses

$$
\left[\begin{array}{ll}
\mathbf{knife} & \\
\text{TYPESTR} = & \left[\text{ARG1} = \boxed{\text{x}}\ \mathbf{artifact\ tool}\right] \\
\text{ARGSTR} = & \left[\begin{array}{l}
\text{D-ARG1} = \boxed{\text{y}}\ \mathbf{physical\ object} \\
\text{D-ARG2} = \boxed{\text{w}}\ \mathbf{human} \\
\text{D-ARG3} = \boxed{\text{z}}\ \mathbf{human} \\
\text{D-E1} = \boxed{e_1}\ \mathbf{transition} \\
\text{D-E2} = \boxed{e_2}\ \mathbf{process}
\end{array}\right] \\
\text{QUALIA} = & \left[\begin{array}{l}
\text{FORMAL} = \boxed{\text{x}} \\
\text{CONSTITUTIVE} = \{\mathbf{blade}, \mathbf{handle}, \ldots\} \\
\text{TELIC} = \mathbf{cut\ act}(\boxed{e_2}, \boxed{\text{w}}, \boxed{\text{x}}, \boxed{\text{y}}) \\
\text{AGENTIVE} = \mathbf{make\ act}(\boxed{e_1}, \boxed{\text{z}}, \boxed{\text{x}})
\end{array}\right]
\end{array}\right]
$$

Qualia structures are the roles describing the meaning of a lexical element:

– FORMAL: distinguishes a word within a larger domain typing information about the object (e.g., hypernym)

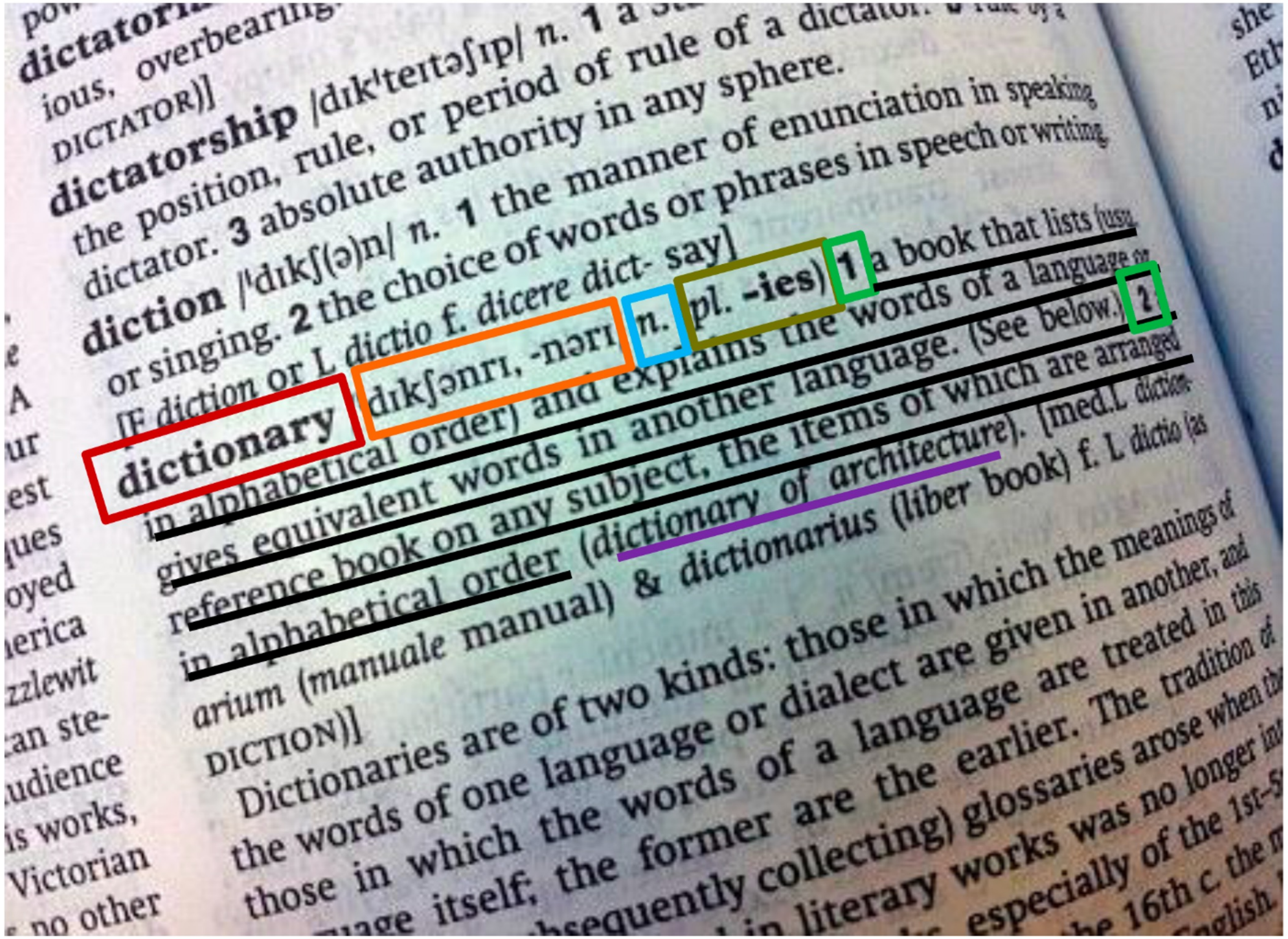
– CONSTITUTIVE: physical properties of a object (e.g., material, parts)

– TELIC: the object’s purpose and function

– AGENTIVE: the object’s origin or “bringing about”, typically a verb denoting an action which brings the object in existence

How do we encode word senses?

- The human way: dictionaries



Lemma

Pronunciation

Part of speech

Morphological info

Senses

Textual definition or gloss

Usage information

How do we encode word senses?

- The computer way:
  - thesauri: lexical resources that list words grouped together according to similarity of meaning (containing synonyms and sometimes antonyms)

IV. WORDS RELATING TO THE INTELLECTUAL FACULTIES; COMMUNICATION OF IDEAS

III. MEANS OF COMMUNICATING IDEAS
2. Conventional means; written language
Book.

[Nouns] booklet; writing, work, volume, tome, opuscule; tract, tractate; livret; brochure, libretto, handbook, codex, manual, pamphlet, enchiridion, circular, publication; chap book.

part, issue, number, livraison; album, portfolio; periodical, serial, magazine, ephemeris, annual, journal.

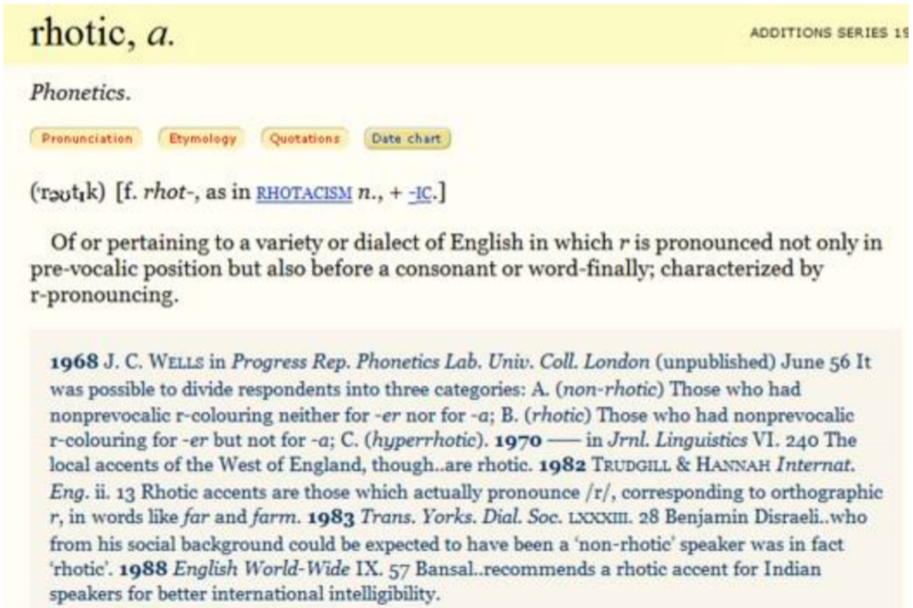
paper, bill, sheet, broadsheet; leaf, leaflet; fly leaf, page; quire, ream

chapter, section head, article, paragraph, passage, clause.

folio, quarto, octavo; duodecimo, sextodecimo, octodecimo.

How do we encode word senses?

- The computer way:
  - thesauri: lexical resources that list words grouped together according to similarity of meaning (containing synonyms and sometimes antonyms)
  - machine-readable dictionaries (electronic counterparts of paper dictionaries)



rhotic, *a.* ADDITIONS SERIES 19

Phonetics.

Pronunciation Etymology Quotations Date chart

(ˈrəʊtɪk) [f. *rhot-*, as in RHOTACISM *n.*, + -IC.]

Of or pertaining to a variety or dialect of English in which *r* is pronounced not only in pre-vocalic position but also before a consonant or word-finally; characterized by r-pronouncing.

1968 J. C. WELLS in *Progress Rep. Phonetics Lab. Univ. Coll. London* (unpublished) June 56 It was possible to divide respondents into three categories: A. (*non-rhotic*) Those who had nonprevocalic r-colouring neither for *-er* nor for *-a*; B. (*rhotic*) Those who had nonprevocalic r-colouring for *-er* but not for *-a*; C. (*hyperrhotic*). **1970** —— in *Jrnl. Linguistics* VI. 240 The local accents of the West of England, though..are rhotic. **1982** TRUDGILL & HANNAH *Internat. Eng.* ii. 13 Rhotic accents are those which actually pronounce /r/, corresponding to orthographic *r*, in words like *far* and *farm*. **1983** *Trans. Yorks. Dial. Soc.* LXXXIII. 28 Benjamin Disraeli..who from his social background could be expected to have been a 'non-rhotic' speaker was in fact 'rhotic'. **1988** *English World-Wide* IX. 57 Bansal..recommends a rhotic accent for Indian speakers for better international intelligibility.

How do we encode word senses?

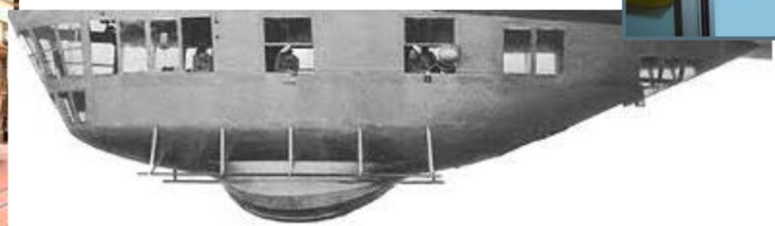
- The computer way:
  - thesauri: lexical resources that list words grouped together according to similarity of meaning (containing synonyms and sometimes antonyms)
  - machine-readable dictionaries (electronic counterparts of paper dictionaries)
  - computational lexicons: highly-structured repositories of syntactic and semantic knowledge about words in a natural language

WordNet [Miller et al. 1990]

- The most popular computational lexicon of English
  - Based on psycholinguistic theories
- Concepts expressed as sets of synonyms (synsets)
  - { car_n^1, $auto_n^1$, $automobile_n^1$, $machine_n^4$, $motorcar_n^1$ }
- A word sense is a word occurring in a synset
  - $\mathbf{machine_n^4}$ is the fourth sense of noun machine

WordNet: the “car” example

$$\begin{aligned} Senses_{WN}(car_n) = \{ & \{car^1_n, auto^1_n, automobile^1_n, machine^4_n, motorcar^1_n\}, \\ & \{car^2_n, rail\ car^1_n, rail\ way\ car^1_n, rail\ road\ car^1_n\}, \\ & \{cable\ car^1_n, car^3_n\}, \\ & \{car^4_n, gondola^3_n\}, \\ & \{car^5_n, elevator\ car^1_n\}\}. \end{aligned}$$



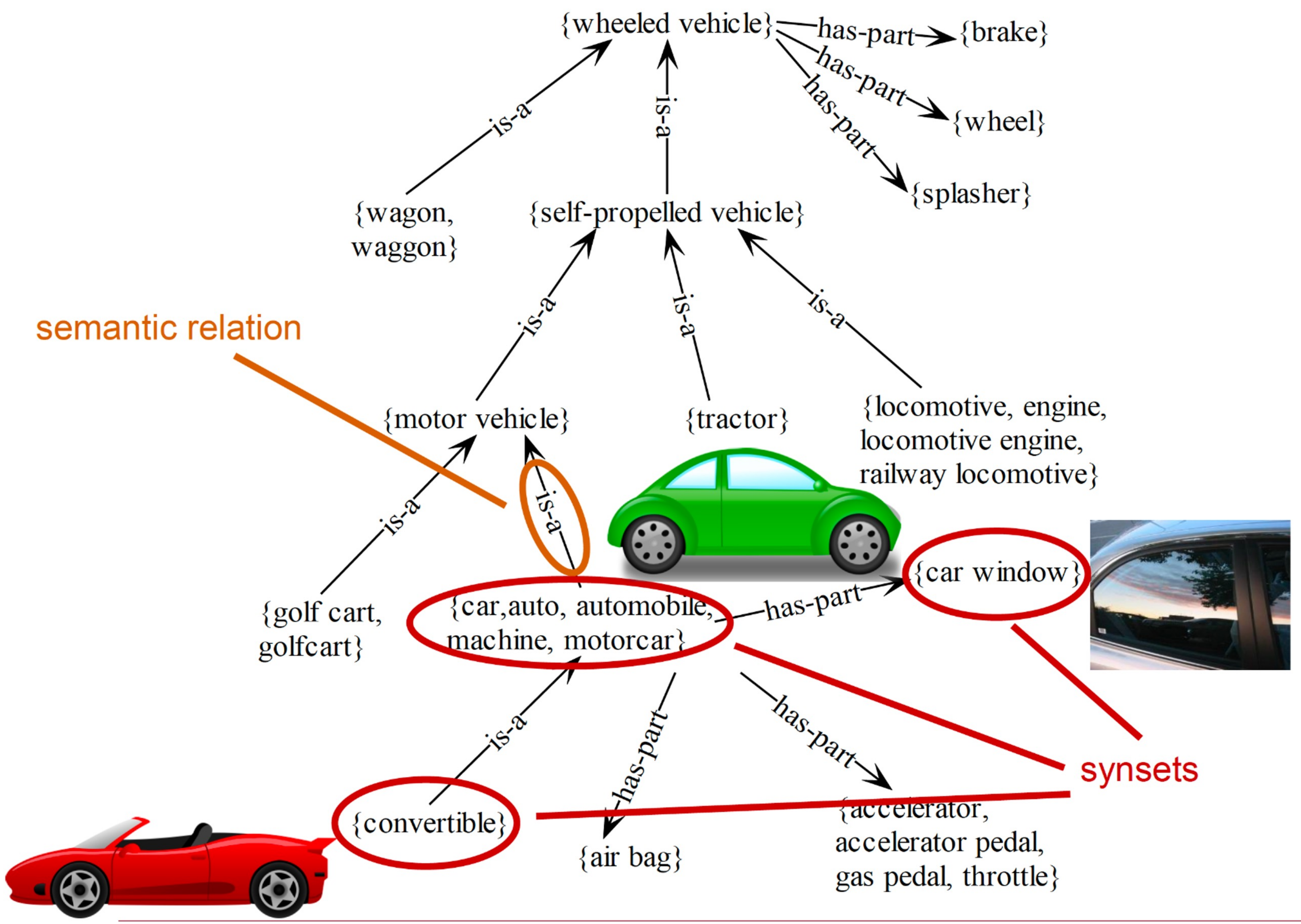
WordNet provides textual definitions

- Called **glosses**
- A textual definition is provided for each synset
- Gloss of car_n^1: 
  - "a 4-wheeled motor vehicle; usually propelled by an internal combustion engine; 'he needs a car to get to work' "
- Gloss of car_n^2: 
  - "a wheeled vehicle adapted to the rails of railroad; 'three cars had jumped the rails' "

WordNet encodes relations!

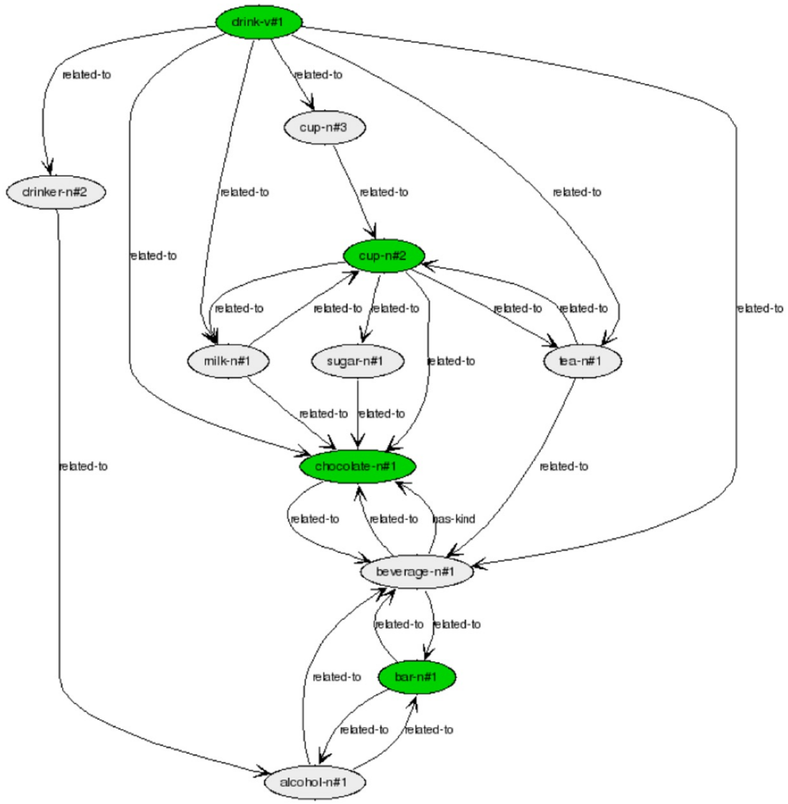
- Semantic relations between synsets
  - **Hypernymy** (car_n^1 is-a motor $vehicle_n^1$)
  - **Meronymy** (car_n^1 has-a car $door_n^1$)
  - **Entailment**, **similarity**, **attribute**, etc.
- Lexical relations between word senses
  - **Synonymy** (i.e., words that belong to the same synset)
  - **Antonymy** ($good_a^1$ antonym of bad_a^1)
  - **Pertainymy** ($dental_a^1$ pertains to $tooth_n^1$)
  - **Nominalization / derivationally related** ($service_n^2$ nominalizes $serve_v^4$)

WordNet as a Graph



But WordNet is more than Simply a Graph!

- It is a semantic network!
- A semantic network is a network which represents semantic relations among concepts
- It is often used as a form of knowledge representation



WordNet [Miller et al. 1990]

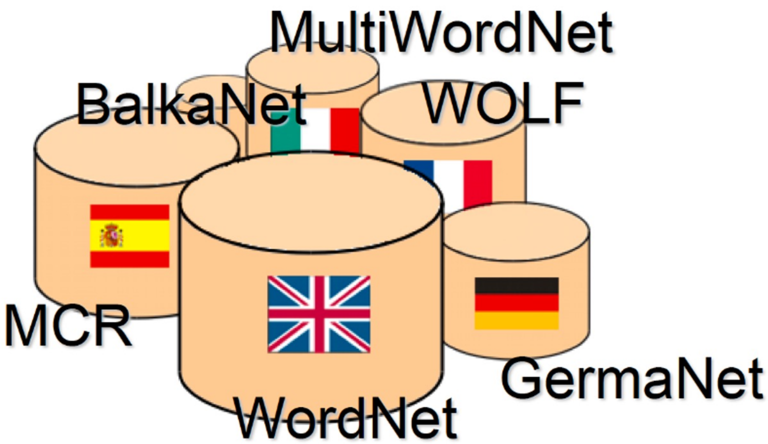
https://wordnet.princeton.edu/

There are several API to programmatically leverage WordNet in your programs.

- Java: (e.g., https://projects.csail.mit.edu/jwi/)
- Python: (e.g., https://www.nltk.org/)
- ...

Wordnets in other languages

- EuroWordNet **[Vossen, 1998]**
- BalkaNet **[Tufis et al., 2004]**
- Multilingual Central Repository **[Atserias et al., 2003]**



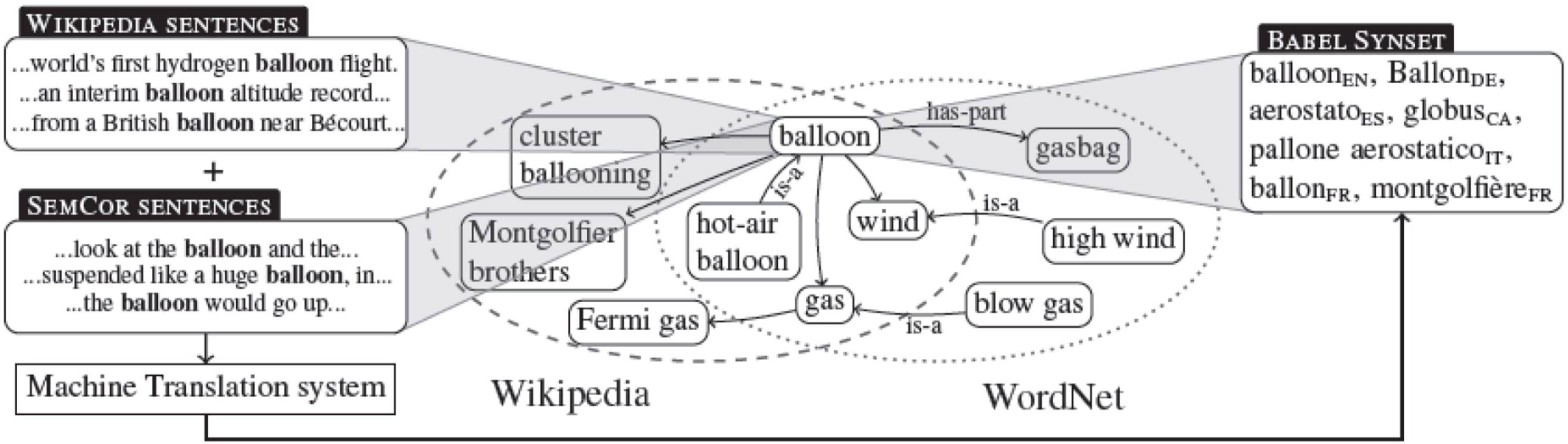
\+ Provide structural knowledge for languages other than English

\- Manually built => limited coverage

\- Available only for a small number of languages

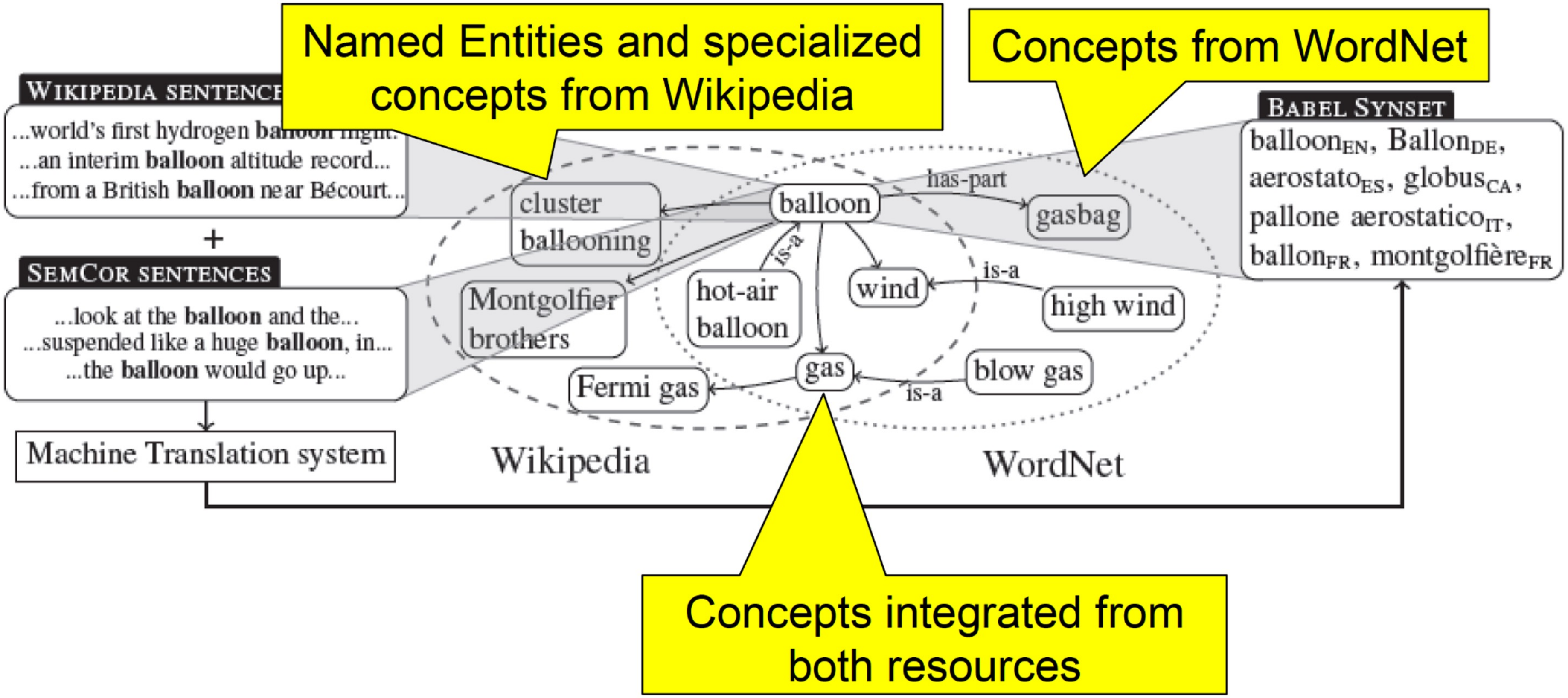
BabelNet [Navigli and Ponzetto, 2010; 2012]

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



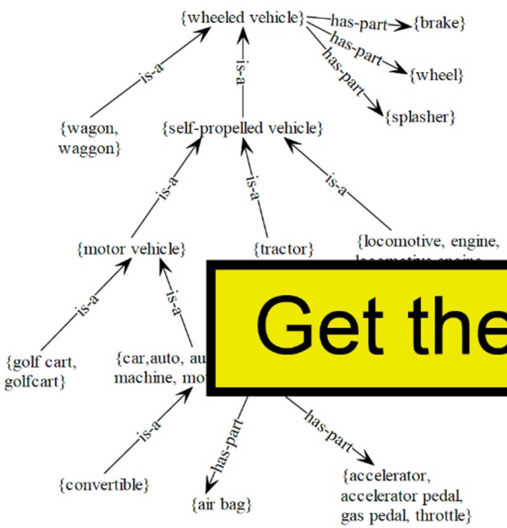
It all started with merging WordNet and Wikipedia [Navigli and Ponzetto, ACL 2010; AIJ 2012]

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

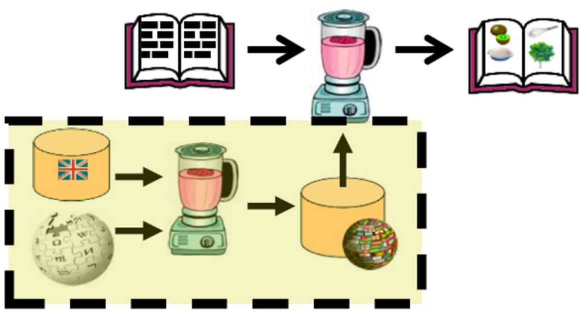


Creating a Multilingual Semantic Network

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



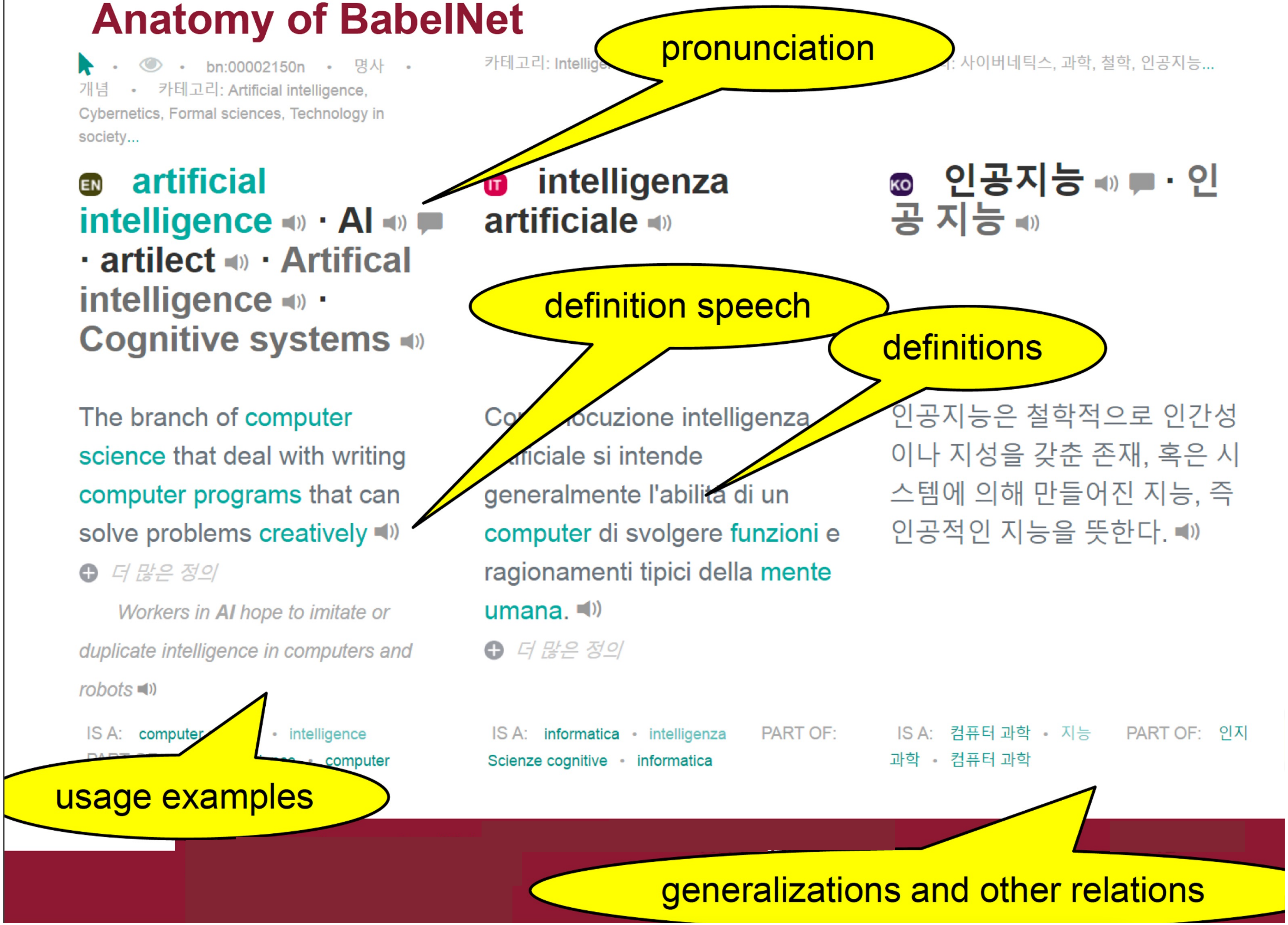
Get the best from both worlds



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
  - 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
  - FrameNet lexical units
  - VerbNet entries
  - Microsoft Terminology: a computer science thesaurus
  - High-quality automatic sense-based translations

Anatomy of BabelNet



Why do we need BabelNet?

- Multilinguality: the same concept is expressed in tens of languages
- Coverage: 284 languages and 16 million entries!
  - **6M** concepts and **7.7M** named entities
  - **119M** word senses
  - **378M** semantic relations (27 relations per concept on avg.)
  - **11M** images associated with concepts
  - **41M** textual definitions
  - **2M** 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
- "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)

Do we need knowledge in some form?

Word Sense Disambiguation (WSD)

Definition:

given a word in context,

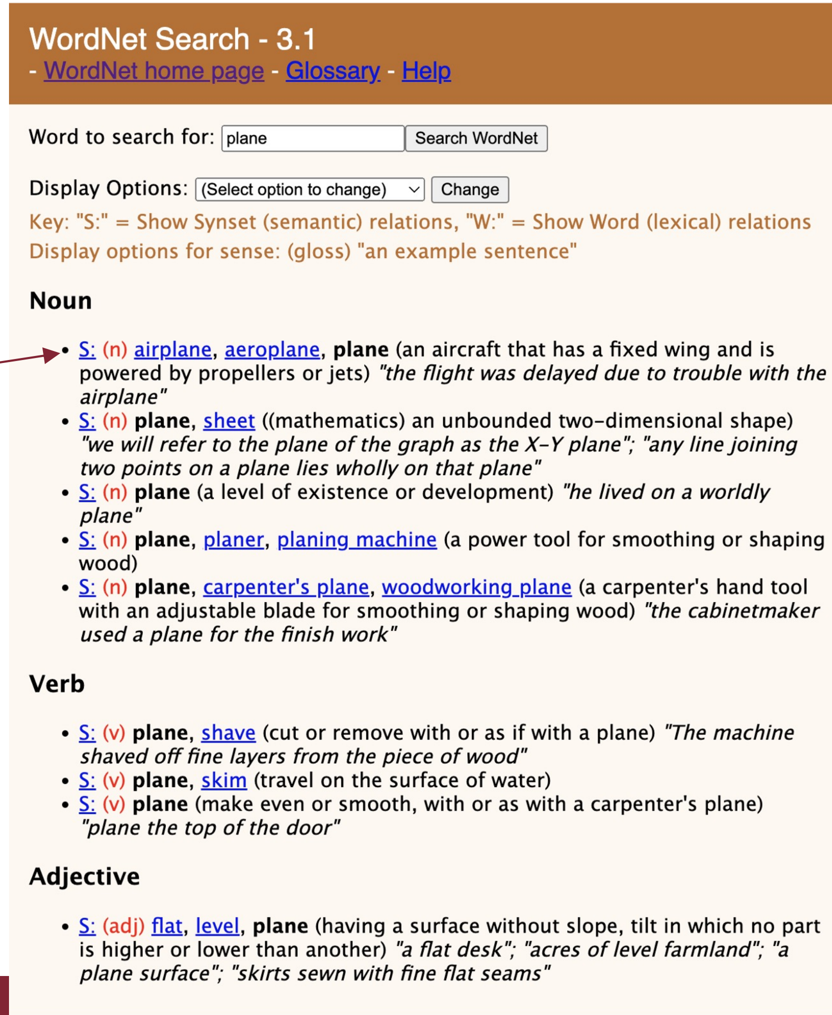
(e.g., “I like **planes** and aeronautics in general”)

and a fixed inventory of potential word senses

(e.g., WordNet)

decide the most suitable synsets, if any.

(e.g., plane_n^1)



WordNet Search - 3.1
- WordNet home page - Glossary - Help

Word to search for: plane [Search WordNet]

Display Options: (Select option to change) [Change]

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) airplane, aeroplane, **plane** (an aircraft that has a fixed wing and is powered by propellers or jets) *"the flight was delayed due to trouble with the airplane"*
- S: (n) **plane**, sheet ((mathematics) an unbounded two-dimensional shape) *"we will refer to the plane of the graph as the X-Y plane"; "any line joining two points on a plane lies wholly on that plane"*
- S: (n) **plane** (a level of existence or development) *"he lived on a worldly plane"*
- S: (n) **plane**, planer, planing machine (a power tool for smoothing or shaping wood)
- S: (n) **plane**, carpenter's plane, woodworking plane (a carpenter's hand tool with an adjustable blade for smoothing or shaping wood) *"the cabinetmaker used a plane for the finish work"*

Verb

- S: (v) **plane**, shave (cut or remove with or as if with a plane) *"The machine shaved off fine layers from the piece of wood"*
- S: (v) **plane**, skim (travel on the surface of water)
- S: (v) **plane** (make even or smooth, with or as with a carpenter's plane) *"plane the top of the door"*

Adjective

- S: (adj) flat, level, **plane** (having a surface without slope, tilt in which no part is higher or lower than another) *"a flat desk"; "acres of level farmland"; "a plane surface"; "skirts sewn with fine flat seams"*

Word Sense Disambiguation (WSD): variants

Definition:

Lexical Sample task

- Small pre-selected set of target words to be disambiguated
- And inventory of senses for each word

All-words task

- Every word in an entire text to be disambiguated
- A lexicon with senses for each word
- Sort of like part-of-speech tagging
- Except each lemma has its own tagset

Word Sense Disambiguation (WSD)

Supervised

Unsupervised

Minimally supervised

Neural WSD

Recent Trends in Word Sense Disambiguation: A Survey

Michele Bevilacqua[1], **Tommaso Pasini**[2], **Alessandro Raganato**[3] and **Roberto Navigli**[1]
[1]Sapienza NLP Group, Department of Computer Science, Sapienza University of Rome
[2]Department of Computer Science, University of Copenhagen
[3]Department of Digital Humanities, University of Helsinki
michele.bevilacqua@uniroma1.it, tommaso.pasini@di.ku.dk
alessandro.raganato@helsinki.fi, roberto.navigli@uniroma1.it

Abstract

Word Sense Disambiguation (WSD) aims at making explicit the semantics of a word in context by identifying the most suitable meaning from a predefined sense inventory. Recent breakthroughs in representation learning have fueled intensive WSD research, resulting in considerable performance improvements, breaching the 80% glass ceiling set by the inter-annotator agreement. In this survey, we provide an extensive overview of current advances in WSD, describing the state of the art in terms of i) resources for the task, i.e., sense inventories and reference datasets for training and testing, as well as ii) automatic disambiguation approaches, detailing their peculiarities, strengths and weaknesses. Finally, we highlight the current limitations of the task itself, but also point out recent trends that could help expand the scope and applicability of WSD, setting up new promising directions for the future.

1 Introduction

Word Sense Disambiguation (WSD) is a historical task in Natural Language Processing (NLP) and Artificial Intelligence (AI) which, in its essence, dates back to Weaver [1949], who recognized the problem of polysemous words in the context of Machine Translation. Even today, word polysemy remains one of the most challenging and pervasive linguistic phenomena in NLP. For example, the ambiguous word *bass* refers to two completely disjoint classes of objects in the following sentences: i) "I can hear *bass* sounds", ii) "They like

and adverbs), as these are the words carrying most of a sentence's meaning. In WSD, the sense inventory for a language can be very large, i.e., in the order of hundreds of thousands of concepts, but also very sparse, in that each *lexeme*[1] is associated with only a small subset of the sense inventory.

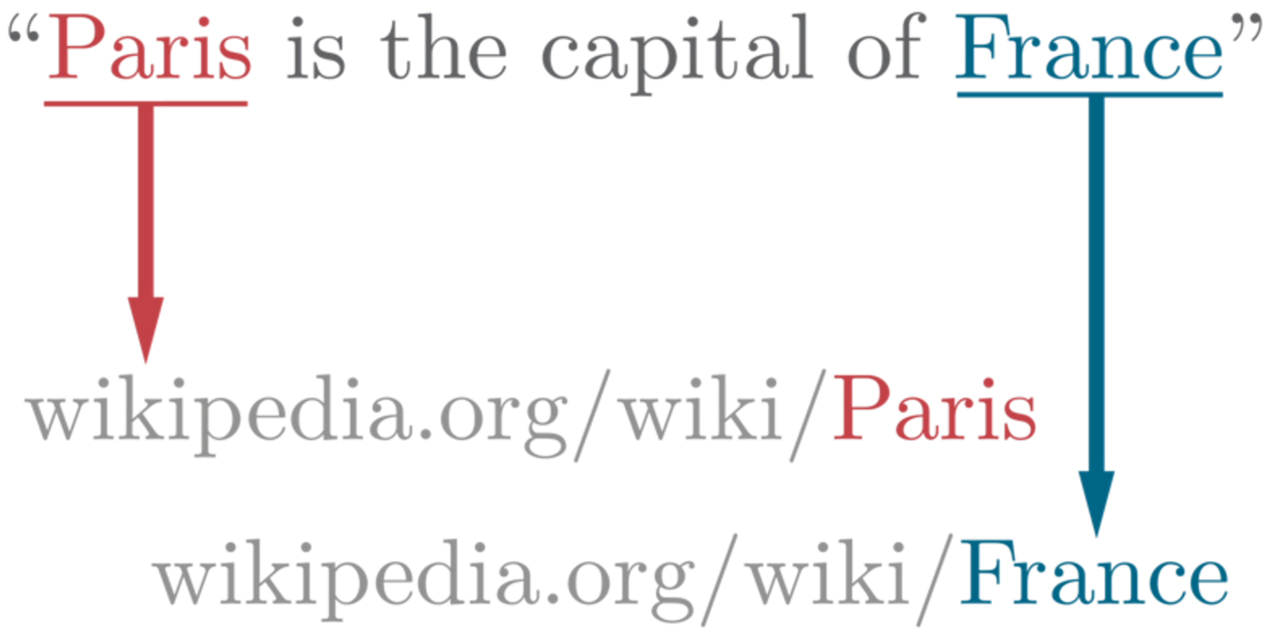
Predefined inventories define the output space for most varieties of past and modern approaches. These exist in many flavors, ranging from purely supervised [Hadiwinoto *et al.*, 2019; Bevilacqua and Navigli, 2019] to knowledge-based [Moro *et al.*, 2014; Agirre *et al.*, 2014; Scozzafava *et al.*, 2020], to hybrid supervised and knowledge-based approaches [Kumar *et al.*, 2019; Bevilacqua and Navigli, 2020; Blevins and Zettlemoyer, 2020; Conia and Navigli, 2021; Barba *et al.*, 2021]. Supervised models, today based on neural architectures, frame the task as a classification problem and take advantage of annotated data to learn the association between words[2] in context and senses. Knowledge-based approaches, instead, often employ graph algorithms on a semantic network, in which senses are connected through semantic relations and are described with definitions and usage examples. Their independence from labeled training data, however, comes at the expense of performing worse than supervised models [Pilehvar and Navigli, 2014; Raganato *et al.*, 2017a; Pasini *et al.*, 2021] which, benefiting from pretrained language models, can now also nimbly scale across different languages. Nonetheless, information in semantic networks, be it unstructured (e.g., definitions) or structured (e.g., relational information), still remains highly relevant. This is demonstrated by hybrid approaches, which, reporting the highest results in literature, are currently attested as the best solution [Barba *et al.*, 2021].

Considering the fast pace at which the field is moving, to-

https://www.ijcai.org/proceedings/2021/593

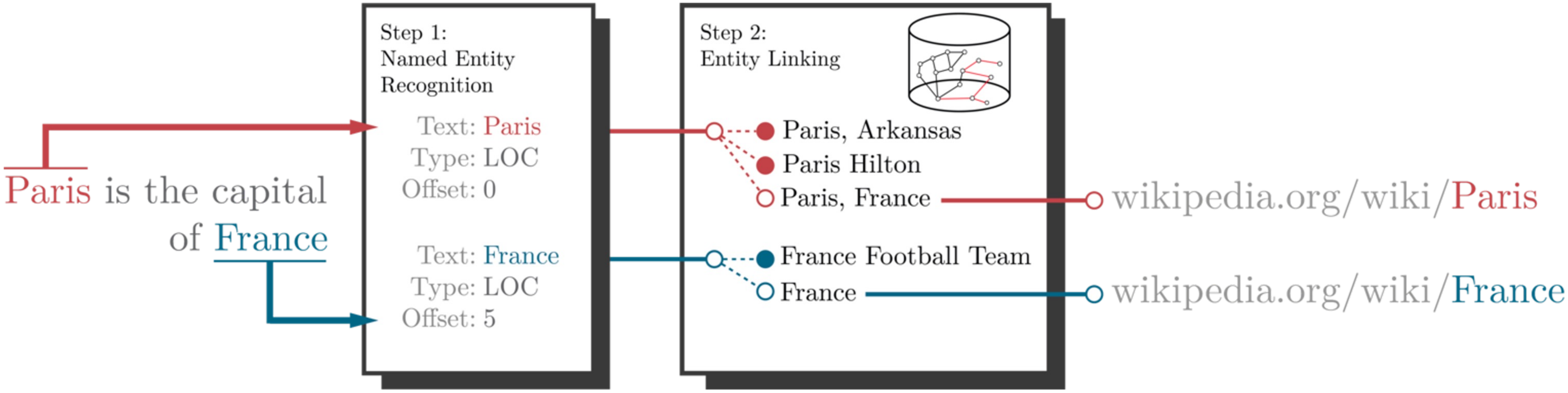
Entity Linking

From Wikipedia:
Entity linking, also referred to as **named-entity linking** (NEL), **named-entity disambiguation** (NED), **named-entity recognition and disambiguation** (NERD) or **named-entity normalization** (NEN) is the task of assigning a unique identity to entities (such as famous individuals, locations, or companies) mentioned in text.



example of text wikification

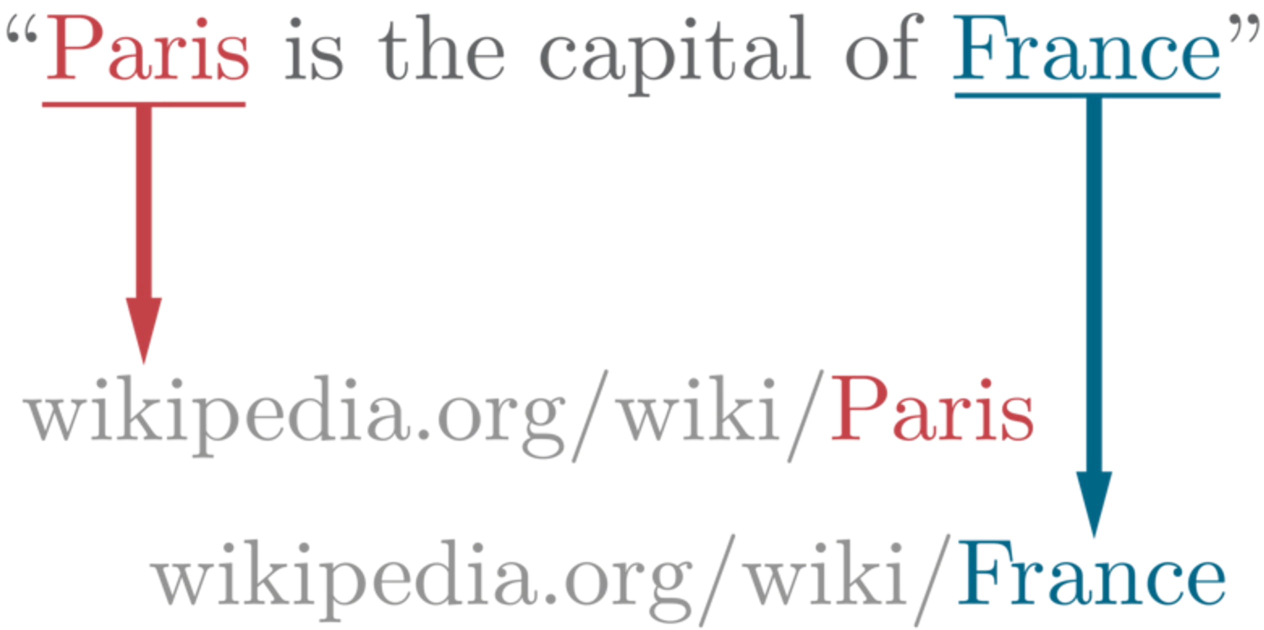
Entity Linking



From Wikipedia:

- **Text-based approaches**, which make use of textual features extracted from large text corpora (e.g. Term frequency–Inverse document frequency (Tf–Idf), word co-occurrence probabilities, etc...).[25][16]
- **Graph-based approaches**, which exploit the structure of knowledge graphs to represent the context and the relation of entities.[3][26]

Relation Extraction



capitalOf(Paris,France)

https://en.wikipedia.org/wiki/Relationship_extraction

Ontology Learning

From Wikipedia:

Ontology learning (**ontology extraction**, **ontology generation**, or **ontology acquisition**) is the automatic or semi-automatic creation of ontologies, including extracting the corresponding domain's terms and the relationships between the concepts that these terms represent from a corpus of natural language text, and encoding them with an ontology language for easy retrieval. As building ontologies manually is extremely labor-intensive and time-consuming, there is great motivation to automate the process.

Typically, the process starts by extracting terms and concepts or noun phrases from plain text using linguistic processors such as part-of-speech tagging and phrase chunking. Then statistical[1] or symbolic[2][3] techniques are used to extract relation signatures, often based on pattern-based[4] or definition-based[5] hypernym extraction techniques.

https://en.wikipedia.org/wiki/Ontology_learning

Ontolearn Reloaded

Computational Linguistics Issues Online Early About Submit

Volume 39, Issue 3

September 2013

Computational Linguistics

< Previous Article Next Article >

Article Contents

September 01 2013

OntoLearn Reloaded: A Graph-Based Algorithm for Taxonomy Induction

In Special Collection: CogNet

Paola Velardi, Stefano Faralli, Roberto Navigli

> Author and Article Information

Computational Linguistics (2013) 39 (3): 665–707.

https://doi.org/10.1162/COLI_a_00146 **Article history**

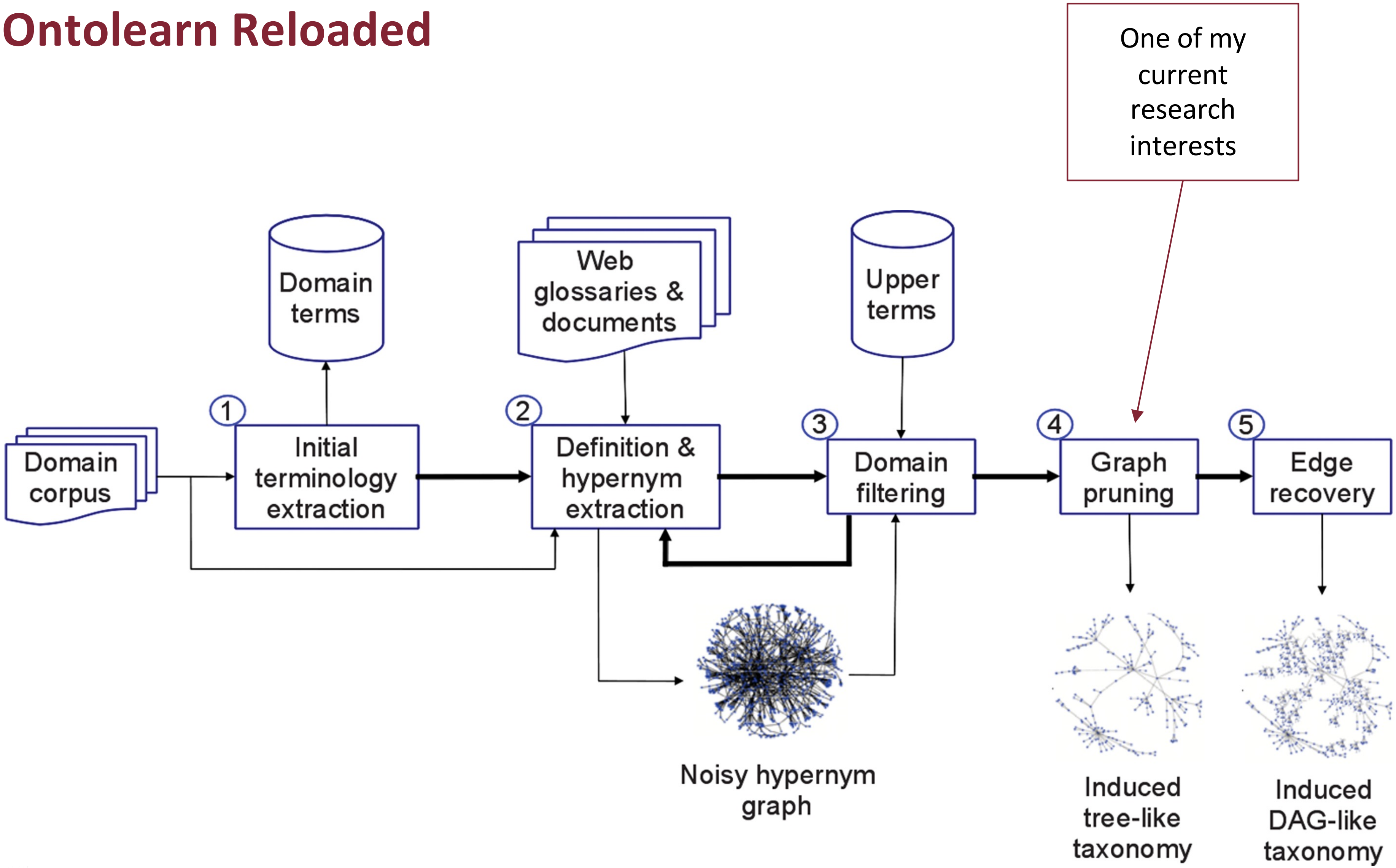
Cite PDF Permissions Share Views

Abstract

In 2004 we published in this journal an article describing OntoLearn, one of the first systems to automatically induce a taxonomy from documents and Web sites. Since then, OntoLearn has continued to be an active area of research in our group and has become a reference work within the community. In this paper we describe our next-generation taxonomy learning methodology, which we name OntoLearn Reloaded. Unlike many taxonomy learning approaches in the literature, our novel algorithm learns both concepts and relations entirely from scratch via the automated extraction of terms, definitions, and hypernyms. This results in a very dense, cyclic and potentially disconnected hypernym graph. The algorithm then induces a taxonomy from this graph via optimal branching and a novel weighting policy. Our experiments show that we obtain high-quality results, both when building brand-new taxonomies and when reconstructing sub-hierarchies of existing taxonomies.

https://direct.mit.edu/coli/article/39/3/665/1442/OntoLearn-Reloaded-A-Graph-Based-Algorithm-for

Ontolearn Reloaded



https://www.youtube.com/watch?v=-k3cOEol_Dk

Q&A

Resources and References

[Jurafsky&Martin, 2022] Jurafsky and Martin. Speech and Language Processing, Prentice Hall, third edition https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf

[Ogden&Richards, 1923] Ogden, Charles Kay & Richards, Ivor Armstrong (1923). The Meaning of Meaning: A Study of the Influence of Language Upon Thought and of the Science of Symbolism. London, England: Kegan, Paul, Trench, Trubner. https://pure.mpg.de/rest/items/item_2366948/component/file_2366947/content

[Erk and McCarthy, 2009] Katrin Erk and Diana McCarthy. 2009. Graded Word Sense Assignment. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, pages 440–449, Singapore. Association for Computational Linguistics. https://aclanthology.org/D09-1046/

[Pustejovsky, 1991] James Pustejovsky. 1991. The Generative Lexicon. Computational Linguistics, 17(4):409–441.

https://aclanthology.org/J91-4003/

[Miller et al., 1990] George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, Katherine J. Miller, Introduction to WordNet: An On-line Lexical Database*, International Journal of Lexicography, Volume 3, Issue 4, Winter 1990, Pages 235–244, https://doi.org/10.1093/ijl/3.4.235

[Vossen,1998] Vossen, P. Introduction to EuroWordNet. Computers and the Humanities 32, 73–89 (1998). https://doi.org/10.1023/A:1001175424222

[Tufis et al.,2004] Dan Tufis, Radu Ion, and Nancy Ide. 2004. Word Sense Disambiguation as a Wordnets' Validation Method in Balkanet. In Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04), Lisbon, Portugal. European Language Resources Association (ELRA). https://aclanthology.org/L04-1119/

[Atserias et al.,2004] Atserias, Jordi & Villarejo, Luis & Rigau, German & Agirre, Eneko & Carroll, J. & Magnini, Bernardo & Vossen, Piek. (2004). The MEANING Multilingual Central Repository. https://adimen.si.ehu.es/~rigau/publications/gwc04-avracmv.pdf

Resources and References

[Navigli&Ponzetto, 2010] Roberto Navigli and Simone Paolo Ponzetto. 2010. BabelNet: Building a Very Large Multilingual Semantic Network. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 216–225, Uppsala, Sweden. Association for Computational Linguistics. https://aclanthology.org/P10-1023/

[Navigli&Ponzetto, 2012] Roberto Navigli, Simone Paolo Ponzetto, BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network, Artificial Intelligence, Volume 193, 2012, Pages 217-250, ISSN 0004-3702, https://doi.org/10.1016/j.artint.2012.07.001

**Credits

The slides of this part of the course are the result of a personal reworking of the slides and of the course material from different sources:

1. The NLP course of Prof. Roberto Navigli, Sapienza University of Rome
2. The NLP course of Prof. Simone Paolo Ponzetto, University of Mannheim, Germany
3. The NLP course of Prof. Chris Biemann, University of Hamburg, Germany
4. The NLP course of Prof. Dan Jurafsky, Stanford University, USA