Word, Sense and Contextualized Embeddings: Vector Representations of Meaning in NLP

Jose Camacho-Collados

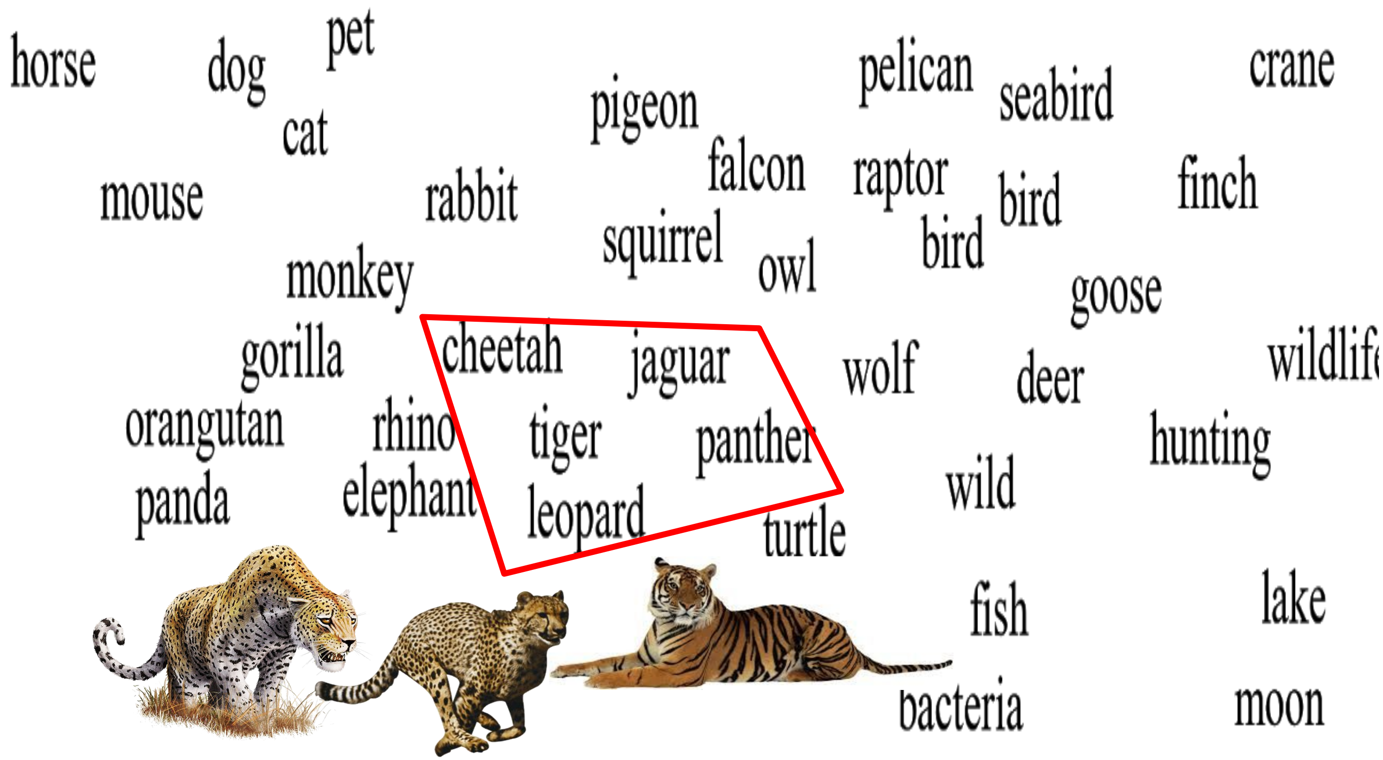
Cardiff University, 18 March 2019

Outline

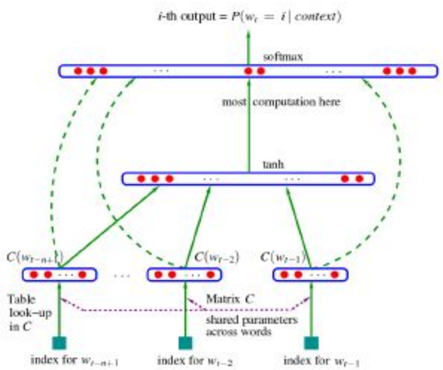
- Background
  - *Vector Space Models (word embeddings)*
  - *Lexical resources*
- Sense representations
  - *Knowledge-based: NASARI, SW2V*
  - *Contextualized: ELMo, BERT*
- Applications

Word vector space models

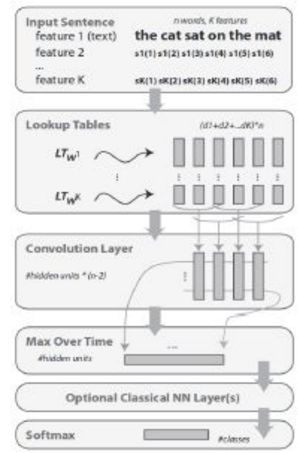
Words are represented as vectors: semantically similar words are close in the vector space



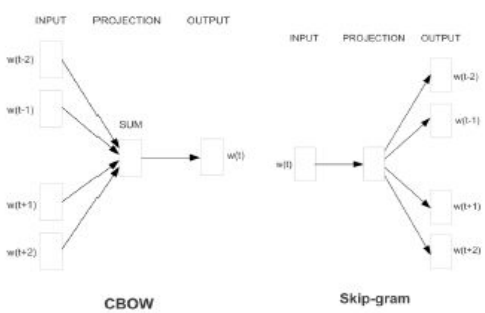
Neural networks for learning word vector representations from text corpora -> **word embeddings**



Bengio et al. (2003)



Collobert & Weston (2008)



Mikolov et al. (2013)

Probability and Ratio	$k = solid$	$k = gas$	$k = water$
$P(k\|ice)$	1.9×10^{-4}	6.6×10^{-5}	3.0×10^{-3}
$P(k\|steam)$	2.2×10^{-5}	7.8×10^{-4}	2.2×10^{-3}
$P(k\|ice)/P(k\|steam)$	8.9	8.5×10^{-2}	1.36

Pennington et al. (2014)

Why word embeddings?

Embedded vector representations:

- are compact and fast to compute
- preserve important relational information between words (actually, meanings):

$$king - man + woman \approx queen$$

- are geared towards general use

Applications for word representations

- Syntactic parsing (Weiss et al. 2015)
- Named Entity Recognition (Guo et al. 2014)
- Question Answering (Bordes et al. 2014)
- Machine Translation (Zou et al. 2013)
- Sentiment Analysis (Socher et al. 2013)

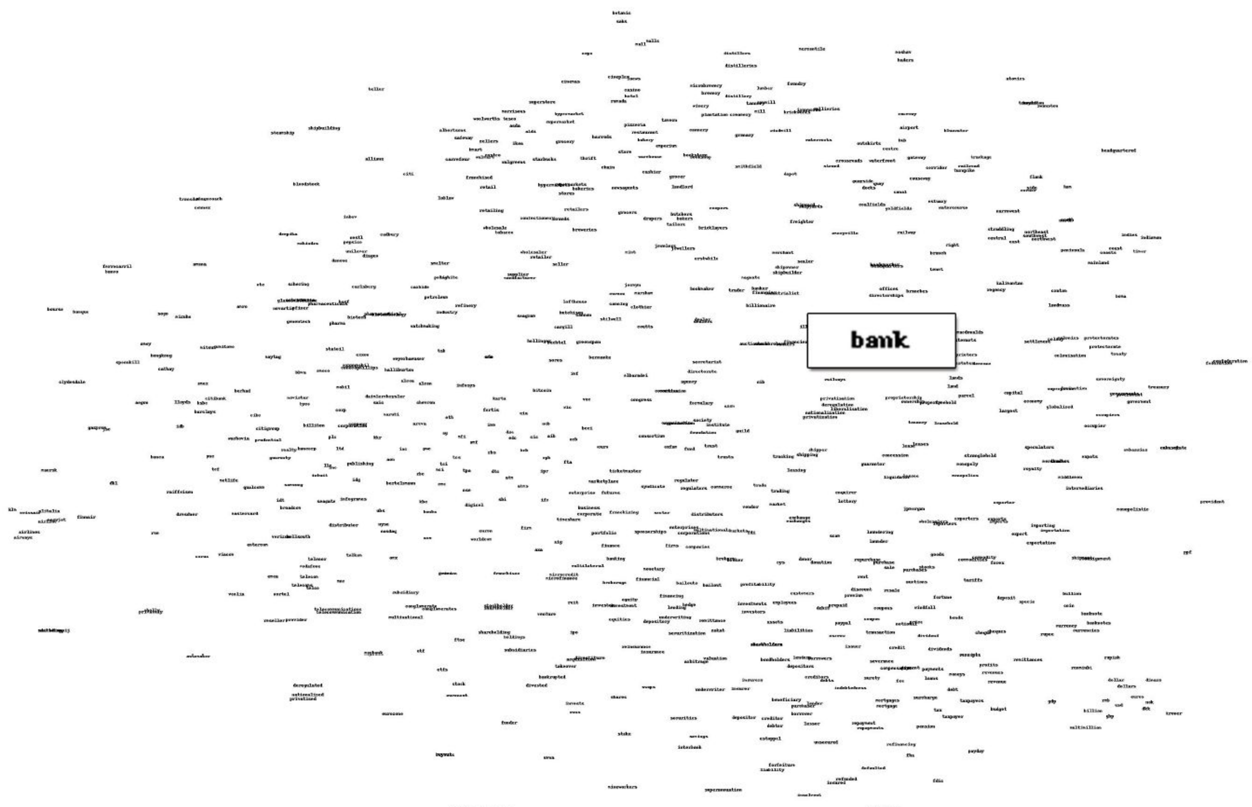
… and many more!

AI goal: language understanding

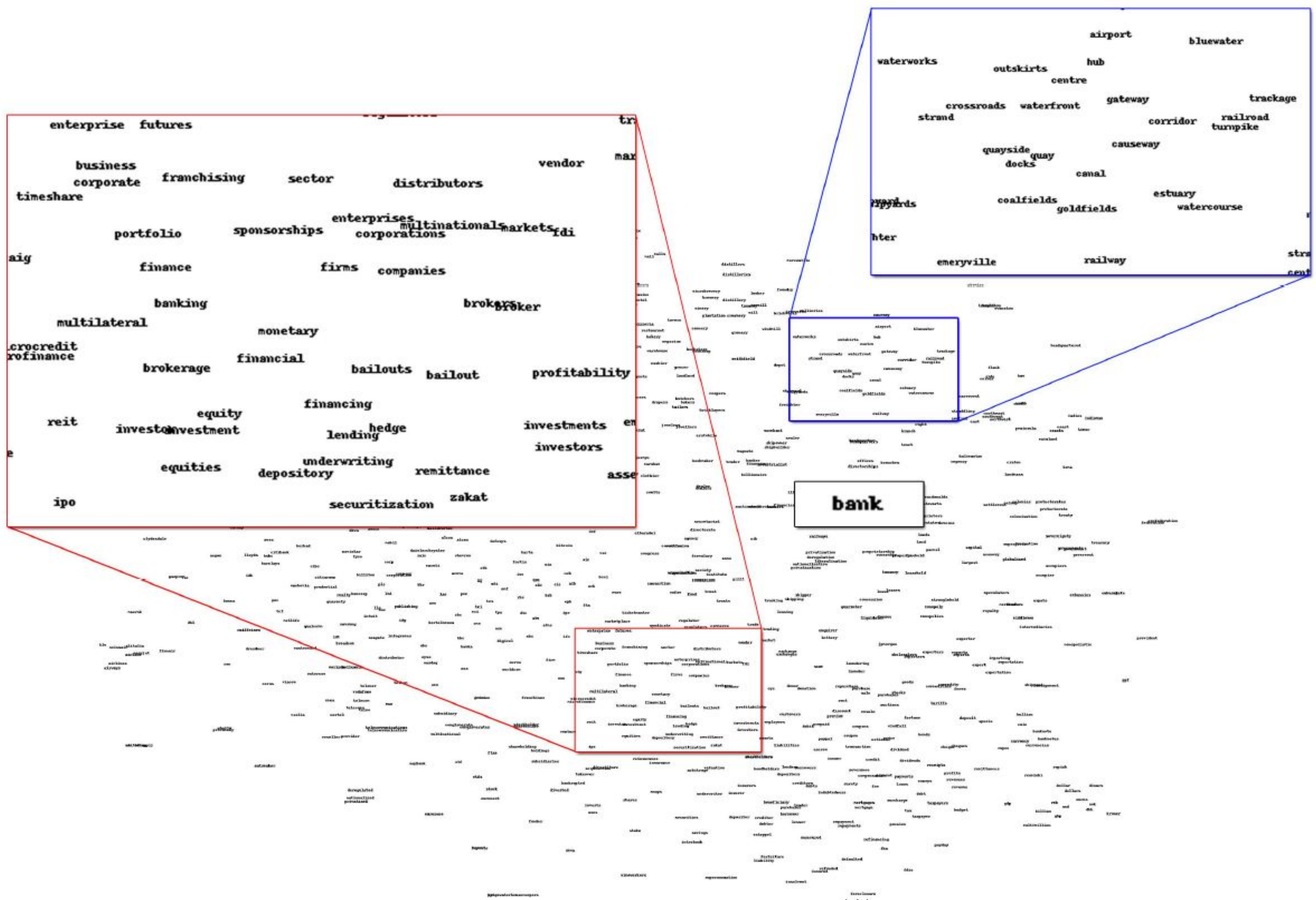
Limitations of word embeddings

- Word representations cannot capture ambiguity. For instance, <u>*bank*</u>

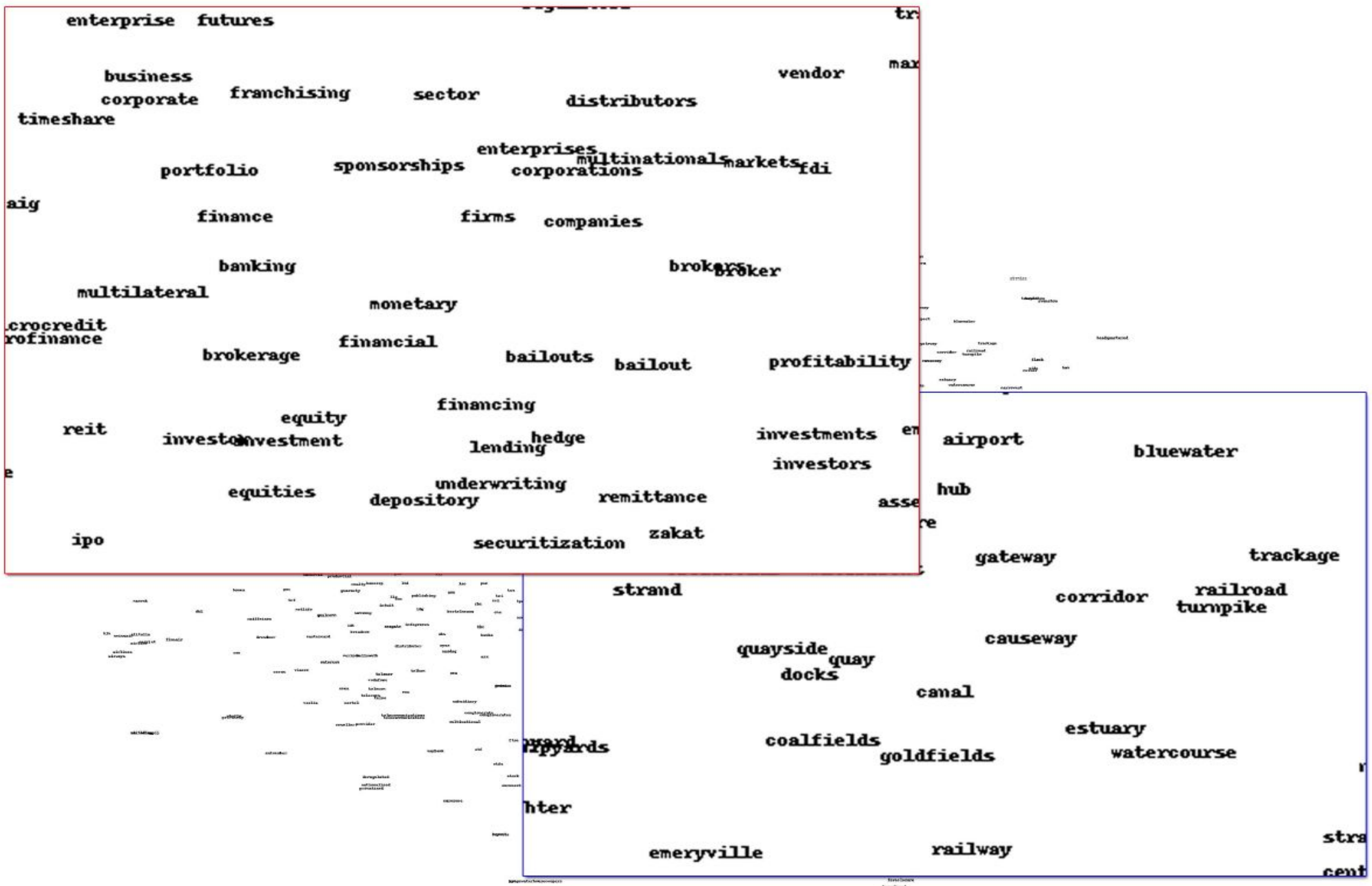
Problem 1: word representations cannot capture ambiguity



Problem 1:
word representations cannot capture ambiguity



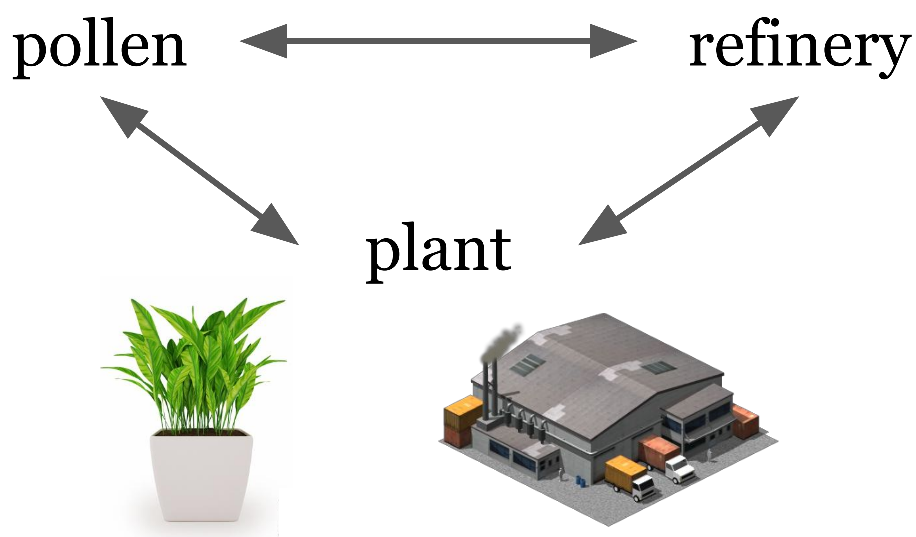
Problem 1: word representations cannot capture ambiguity



Word representations and the triangular inequality

Example from Neelakantan et al (2014)

For distance d, $d(a, c) \leq d(a, b) + d(b, c)$.



Word representations and the triangular inequality

Example from Neelakantan et al (2014)

For distance d, $d(a, c) \leq d(a, b) + d(b, c)$.

pollen ⟷ refinery

plant^1 ⟷ plant^2

Limitations of word representations

- They cannot capture ambiguity. For instance,

bank

-> They neglect rare senses and infrequent words

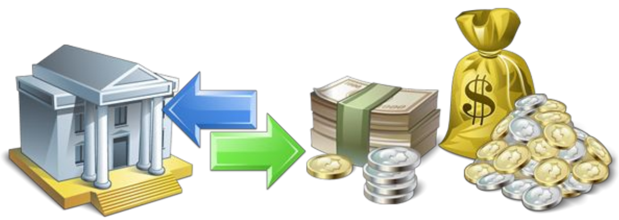
- Word representations do not exploit knowledge from existing lexical resources.

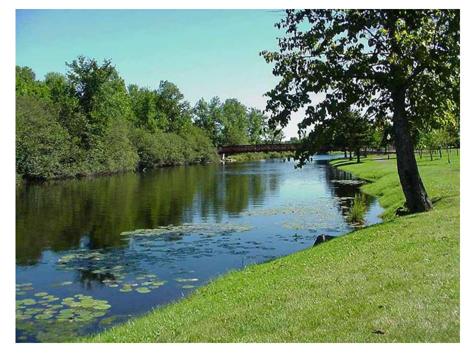
WIKIPEDIA
The Free Encyclopedia

BabelNet

Motivation: Model senses instead of only words

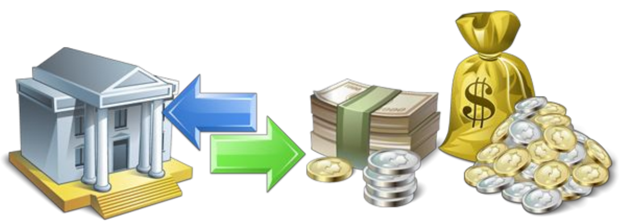
He withdrew money from the ***bank***.



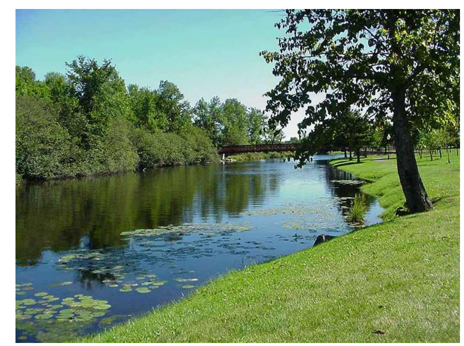


Motivation: Model senses instead of only words

He withdrew money from the ***bank****.*



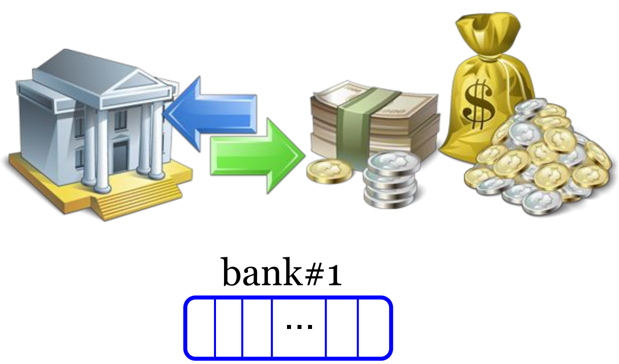
bank#1

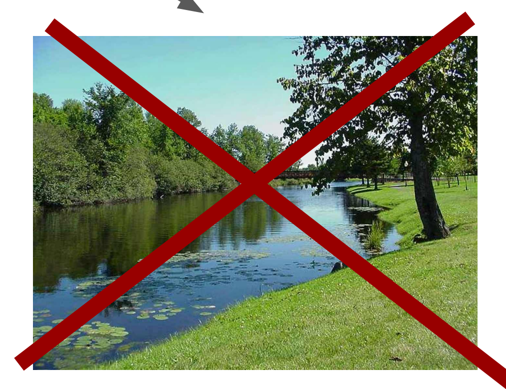


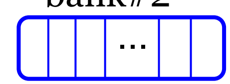
bank#2

Motivation: Model senses instead of only words

*He withdrew money from the **bank**.*





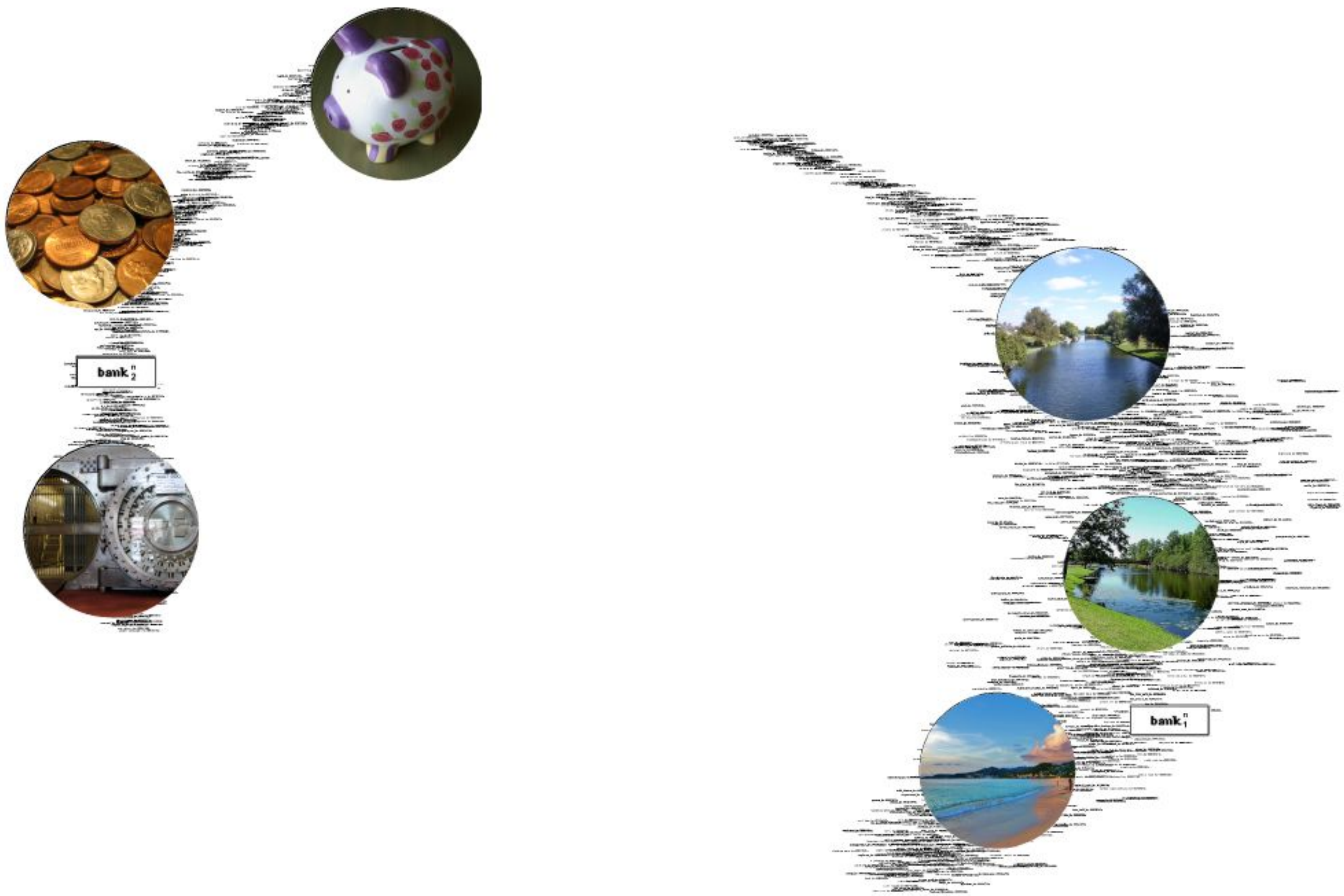


NASARI

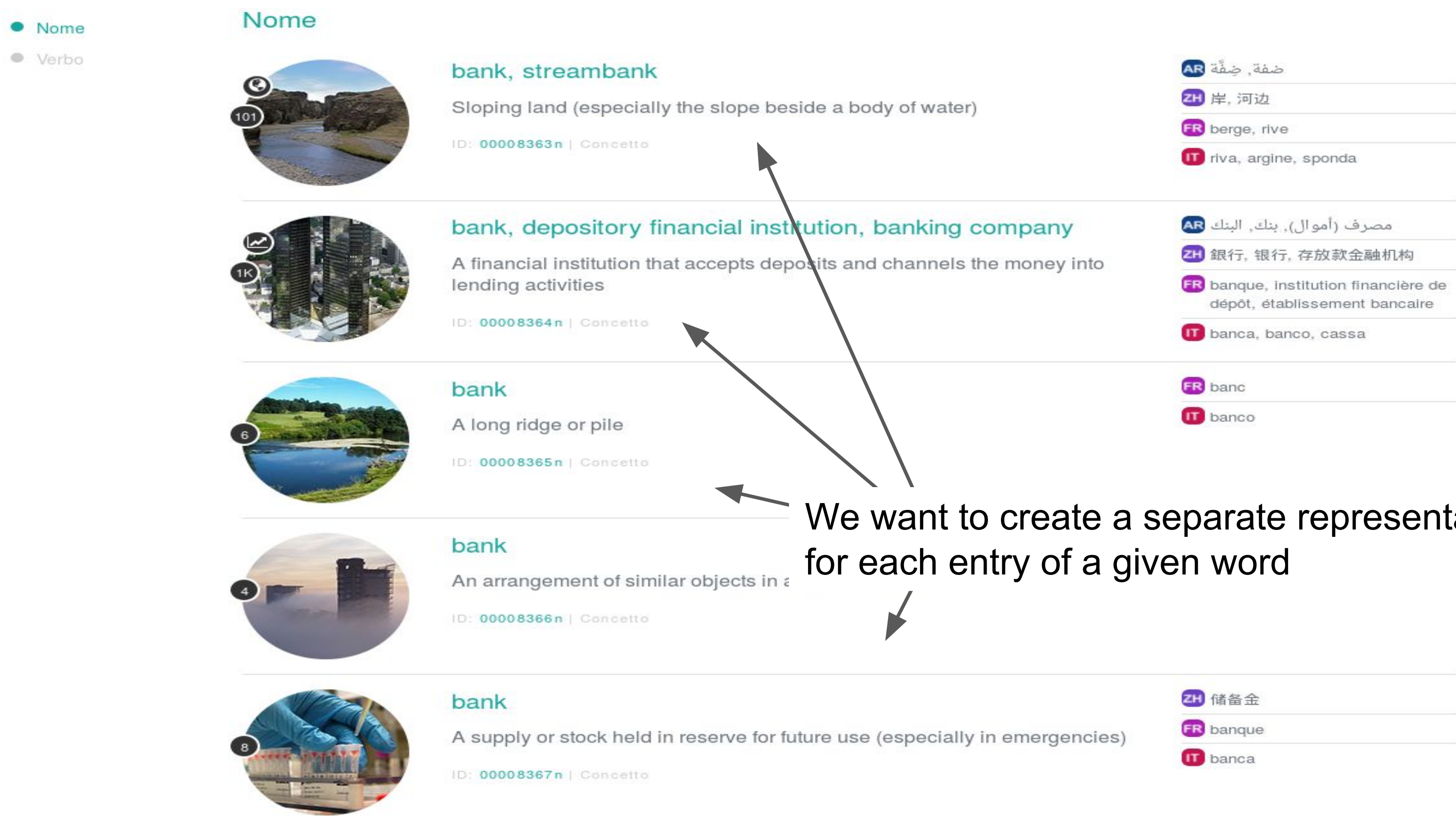
a Novel Approach to a Semantically-Aware Representations of Items

http://lcl.uniroma1.it/nasari/

Key goal: obtain sense representations



Key goal: obtain sense representations



We want to create a separate representation for each entry of a given word

Idea

Encyclopedic knowledge

WIKIPEDIA
The Free Encyclopedia

Lexicographic knowledge

WordNet

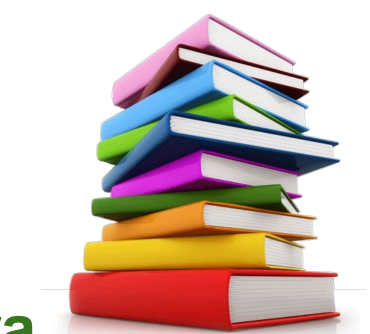
Idea

Encyclopedic knowledge

Wikipedia
The Free Encyclopedia

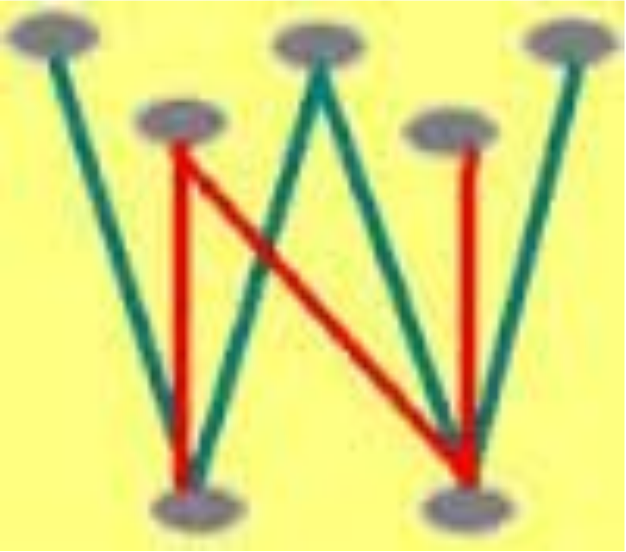
Lexicographic knowledge

WordNet



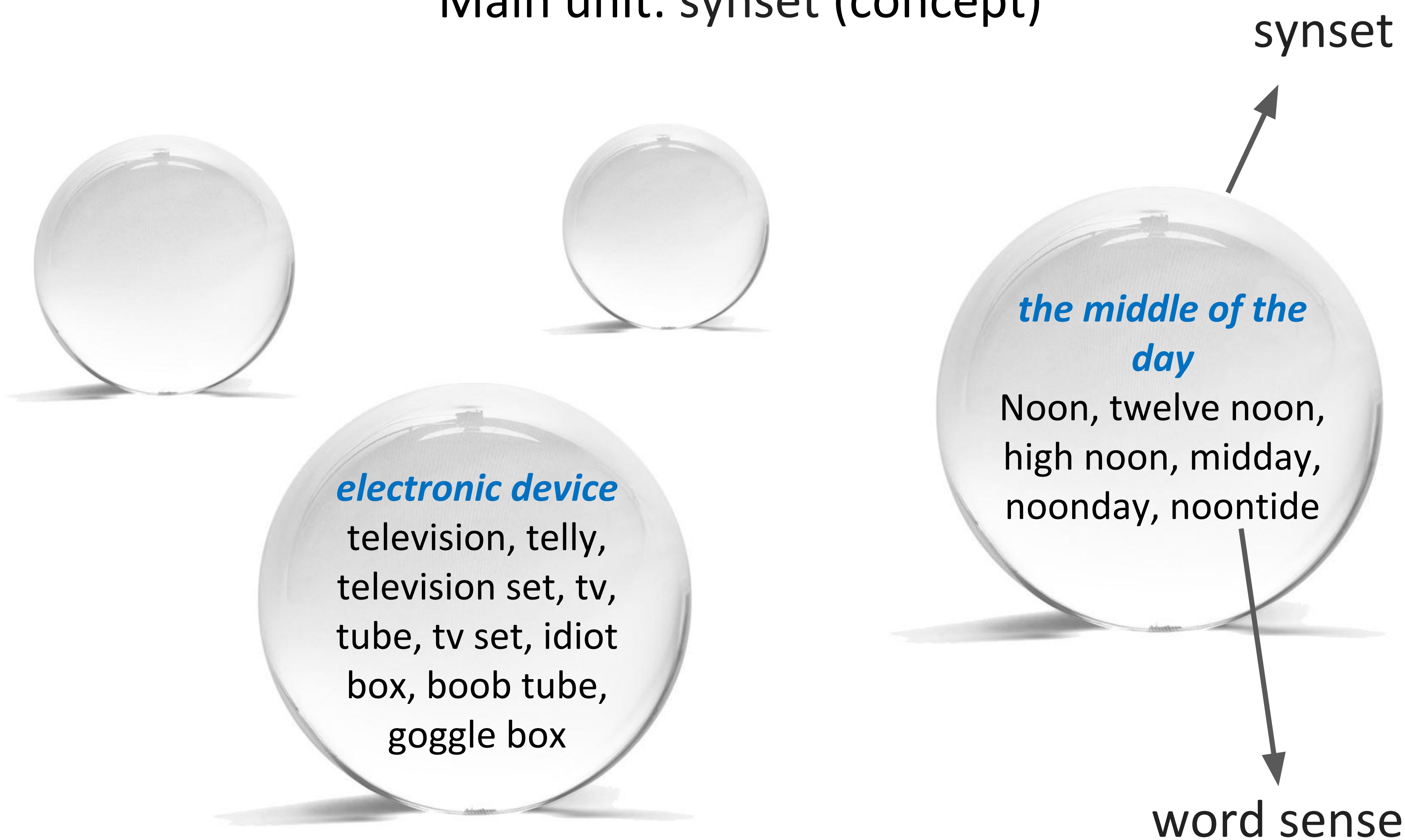
Information from text corpora

WordNet

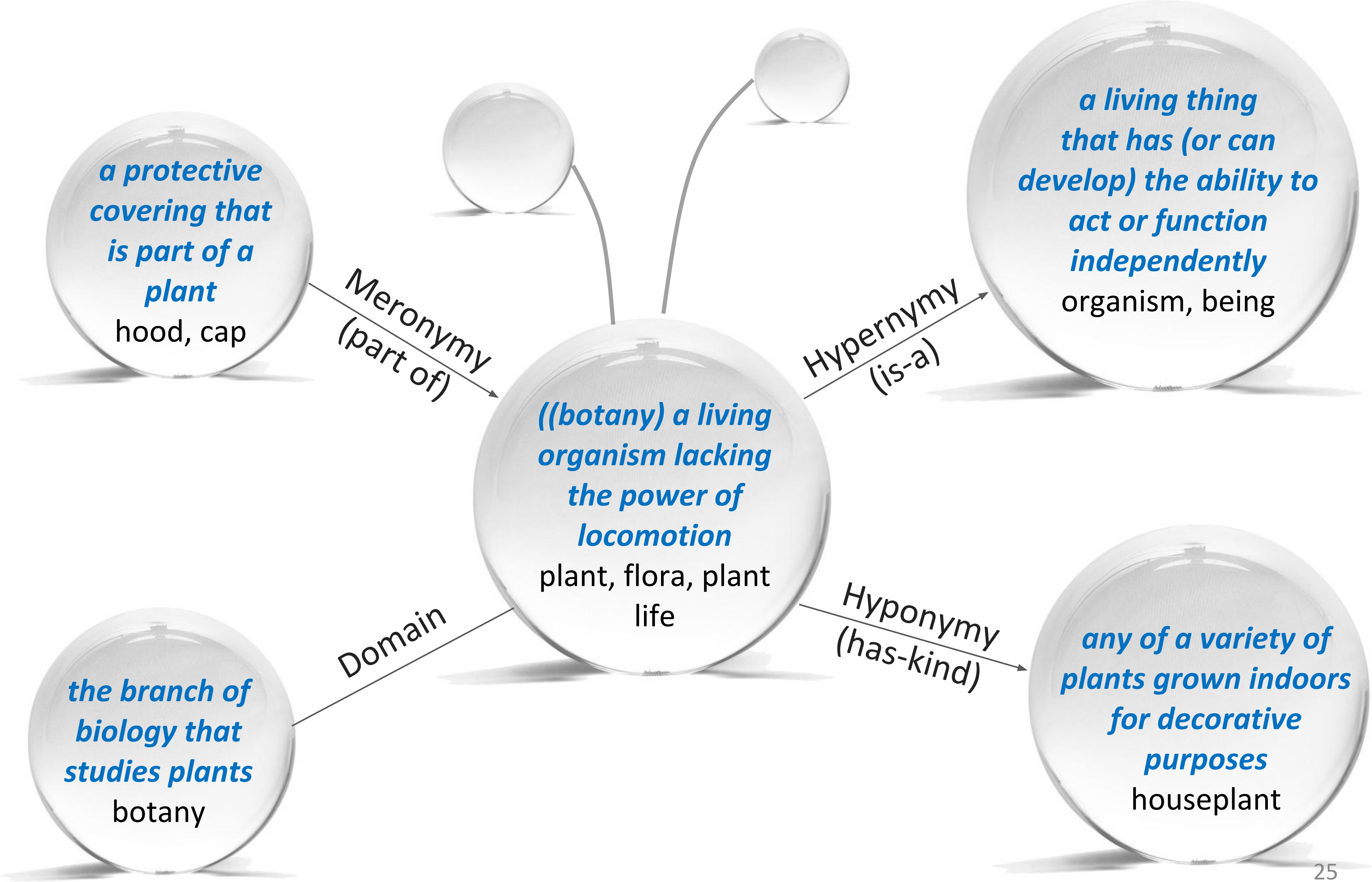


WordNet

Main unit: synset (concept)



WordNet semantic relations



Knowledge-based Representations (WordNet)

X. Chen, Z. Liu, M. Sun: **A Unified Model for Word Sense Representation and Disambiguation** (EMNLP 2014)

S. Rothe and H. Schutze: **AutoExtend: Extending Word Embeddings to Embeddings for Synsets and Lexemes** (ACL 2015)

Faruqui, M., Dodge, J., Jauhar, S. K., Dyer, C., Hovy, E., & Smith, N. A. **Retrofitting Word Vectors to Semantic Lexicons** (NAACL 2015)*

S. K. Jauhar, C. Dyer, E. Hovy: **Ontologically Grounded Multi-sense Representation Learning for Semantic Vector Space Models** (NAACL 2015)

M. T. Pilehvar and N. Collier, **De-Conflated Semantic Representations** (EMNLP 2016)

Wikipedia

WIKIPEDIA
The Free Encyclopedia

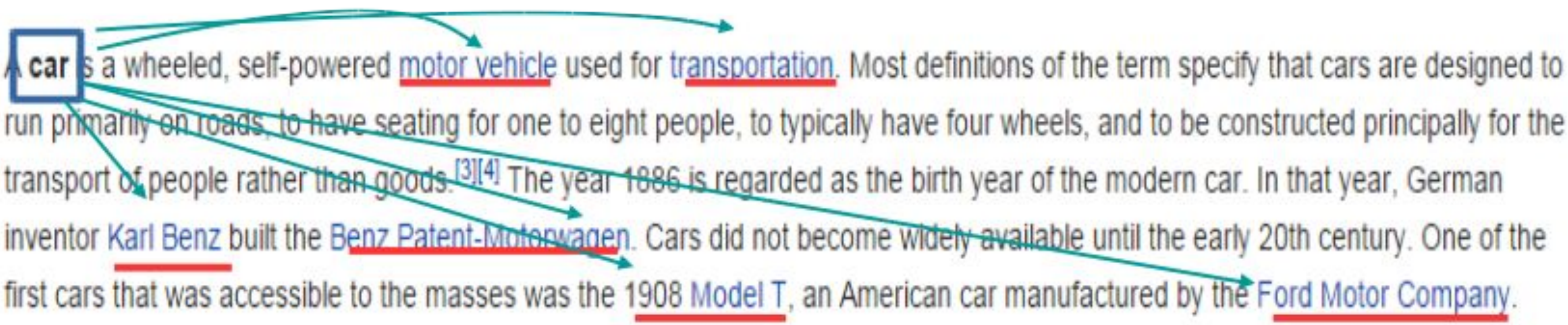
Wikipedia

High coverage of **named entities** and **specialized concepts** from different domains

Wikipedia hyperlinks

A $\boxed{\textbf{car}}$ is a wheeled, self-powered motor vehicle used for transportation. Most definitions of the term specify that cars 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][4] The year 1886 is regarded as the birth year of the modern car. In that year, German inventor Karl Benz built the Benz Patent-Motorwagen. Cars did not become widely available until the early 20th century. One of the first cars that was accessible to the masses was the 1908 Model T, an American car manufactured by the Ford Motor Company.

Wikipedia hyperlinks



A **car** is a wheeled, self-powered motor vehicle used for transportation. Most definitions of the term specify that cars 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][4] The year 1886 is regarded as the birth year of the modern car. In that year, German inventor Karl Benz built the Benz Patent-Motorwagen. Cars did not become widely available until the early 20th century. One of the first cars that was accessible to the masses was the 1908 Model T, an American car manufactured by the Ford Motor Company.

BabelNet

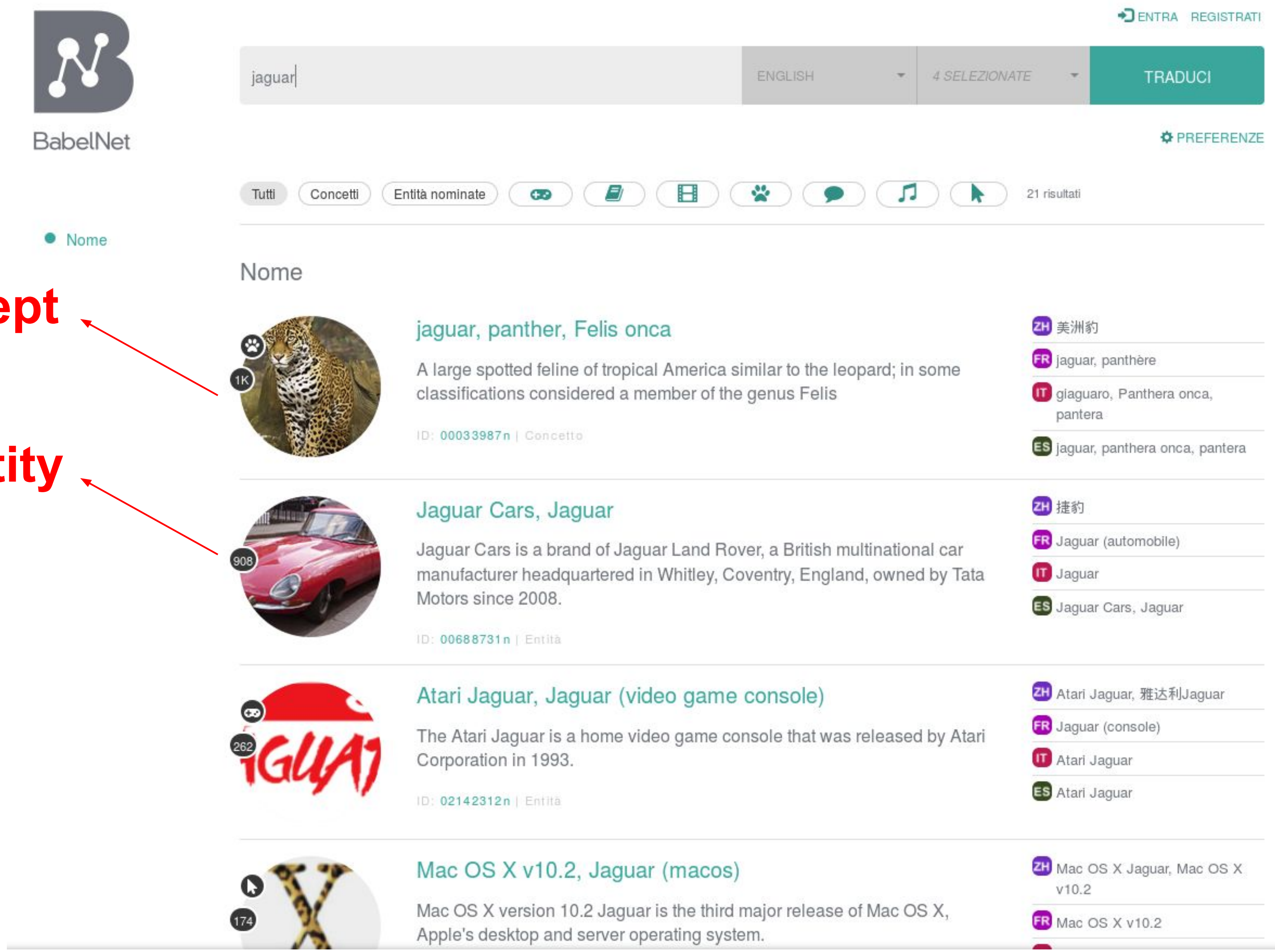
Thanks to an automatic mapping algorithm, **BabelNet integrates Wikipedia and WordNet**, among other resources (Wiktionary, OmegaWiki, WikiData…).

Key feature: **Multilinguality** (271 languages)

BabelNet

Concept

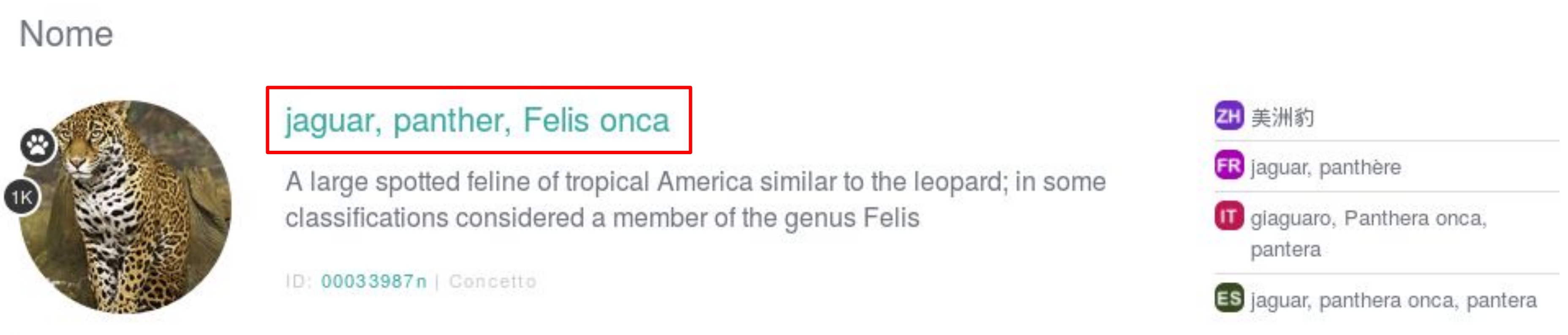
Entity



BabelNet

It follows the same structure of WordNet:

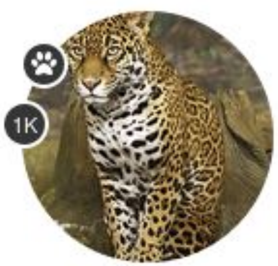
synsets are the main units



BabelNet

In this case, **synsets are multilingual**

Nome



jaguar, panther, Felis onca

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

ID: 00033987n | Concetto

ZH 美洲豹

FR jaguar, panthère

IT giaguaro, Panthera onca, pantera

ES jaguar, panthera onca, pantera

NASARI

(Camacho-Collados et al., AIJ 2016)

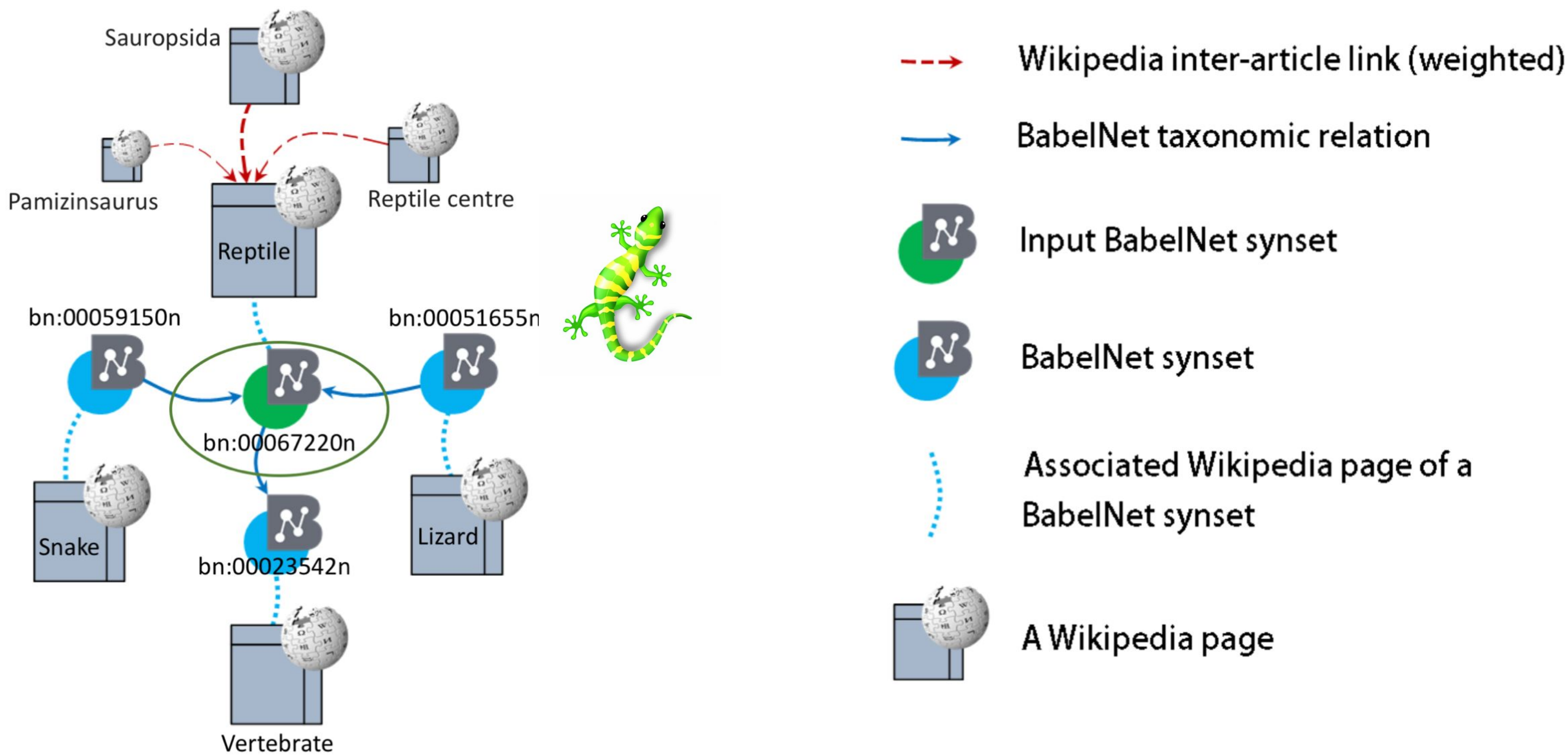
Goal

Build vector representations for multilingual BabelNet synsets.

How?

We exploit **Wikipedia semantic network** and **WordNet taxonomy** to construct a subcorpus (contextual information) for any given BabelNet synset.

Pipeline



Process of obtaining contextual information for a BabelNet synset exploiting BabelNet taxonomy and Wikipedia as a semantic network

Three types of vector representations

Three types of vector representations:

- **Lexical** (dimensions are words)
- **Unified** (dimensions are multilingual BabelNet synsets)
- **Embedded** (latent dimensions)

Three types of vector representations

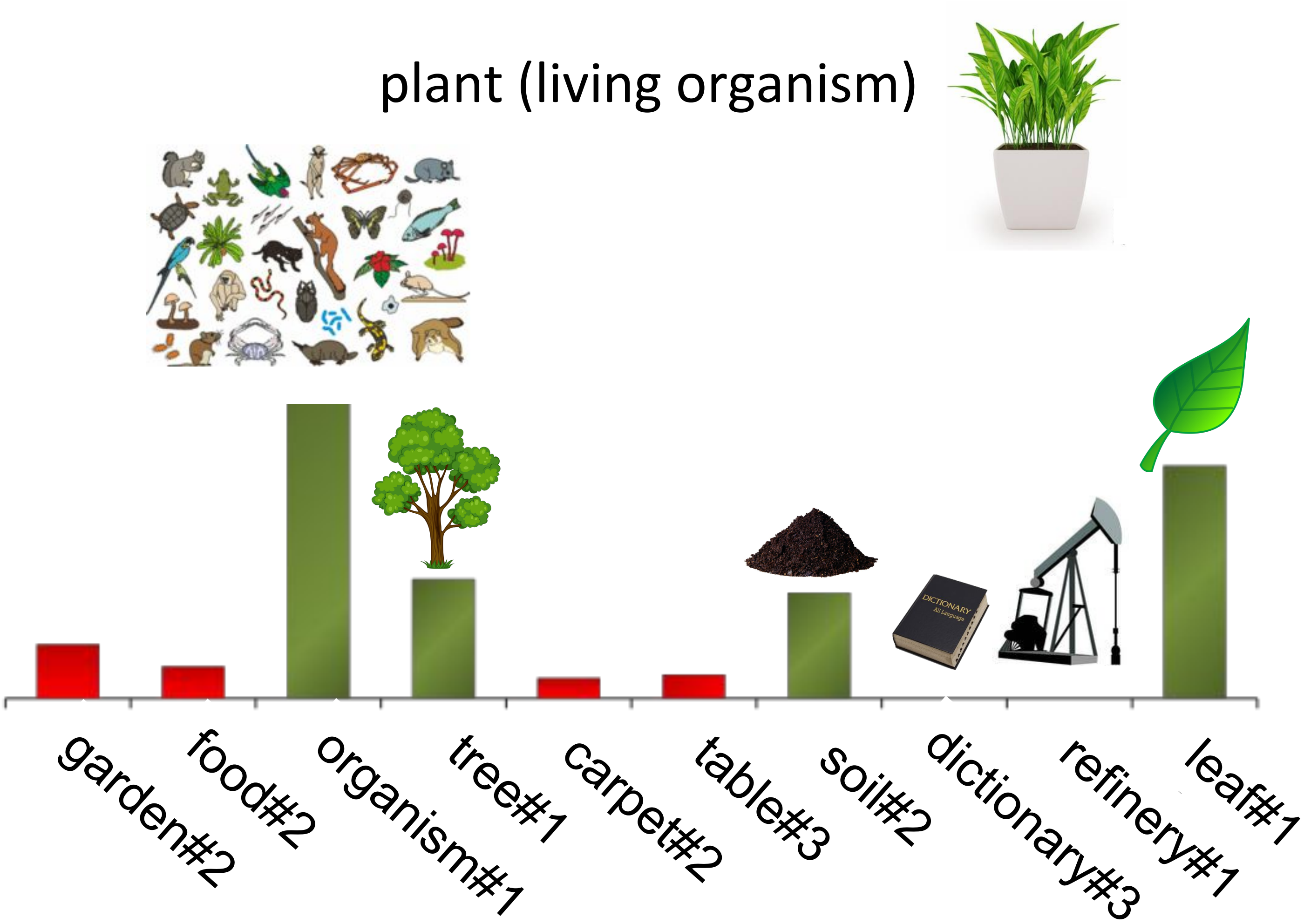
Three types of vector representations:

- **Lexical** (dimensions are words)

- **Unified** (dimensions are multilingual BabelNet synsets)

- **Embedded** (latent dimensions)

Human-interpretable dimensions



Three types of vector representations

Three types of vector representations:

- **Lexical** (dimensions are words)
- **Unified** (dimensions are multilingual BabelNet synsets)

- **Embedded**: Low-dimensional vectors exploiting **word embeddings** obtained from **text corpora**.

Three types of vector representations

Three types of vector representations:

- **Lexical** (dimensions are words)
- **Unified** (dimensions are multilingual BabelNet synsets)

- **Embedded**: Low-dimensional vectors exploiting **word embeddings** obtained from **text corpora**.

Word and synset embeddings share the same vector space!

Embedded vector representation

Closest senses

Bank (financial institution)		Bank (geography)		*bank*	
Closest senses	Cosine	Closest senses	Cosine	Closest senses	Cosine
Deposit account	0.99	Stream bed	0.98	Bank (financial institution)	0.86
Universal bank	0.99	Current (stream)	0.97	Universal bank	0.86
British banking	0.98	River engineering	0.97	British banking	0.86
German banking	0.98	Braided river	0.97	German banking	0.85
Commercial bank	0.98	Fluvial terrace	0.97	Branch (banking)	0.85
Banking in Israel	0.98	Bar (river morphology)	0.97	McFadden Act	0.85
Financial institution	0.98	River	0.97	Four Northern Banks	0.84
Community bank	0.97	Perennial stream	0.96	State bank	0.84

SW2V

(Mancini and Camacho-Collados et al., CoNLL 2017)

A word is the surface form of a sense: we can exploit this intrinsic relationship for **jointly training word and sense embeddings**.

SW2V
(Mancini and Camacho-Collados et al., CoNLL 2017)

A word is the surface form of a sense: we can exploit this intrinsic relationship for **jointly training word and sense embeddings**.

How?

Updating the representation of the word and its associated senses interchangeably.

SW2V: Idea

Given as input a **corpus** and a **semantic network**:

1. Use a semantic network to link to each word its *associated senses in context*.

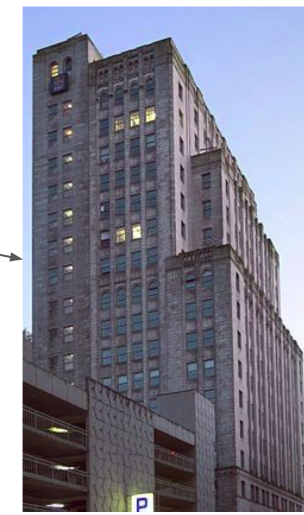
He withdrew money from the ***bank****.*

SW2V: Idea

Given as input a **corpus** and a **semantic network**:

1. Use a semantic network to link to each word its *associated senses in context*.

He withdrew money from the ***bank***.



SW2V: Idea

Given as input a corpus and a semantic network:

1. Use a semantic network to link to each word its *associated senses in context*.

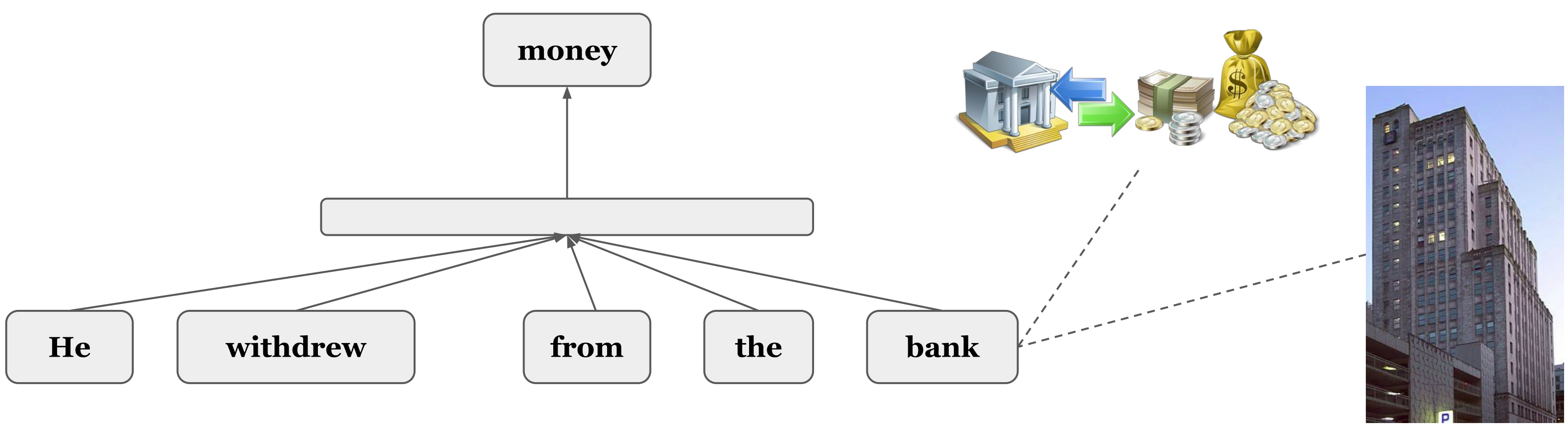
2. Use a **neural network** where the **update of word and sense embeddings is linked**, exploiting *virtual* connections.

SW2V: Idea

Given as input a corpus and a semantic network:

1. Use a semantic network to link to each word its *associated senses in context*.

2. Use a neural network where the update of word and sense embeddings is linked, exploiting *virtual* connections.

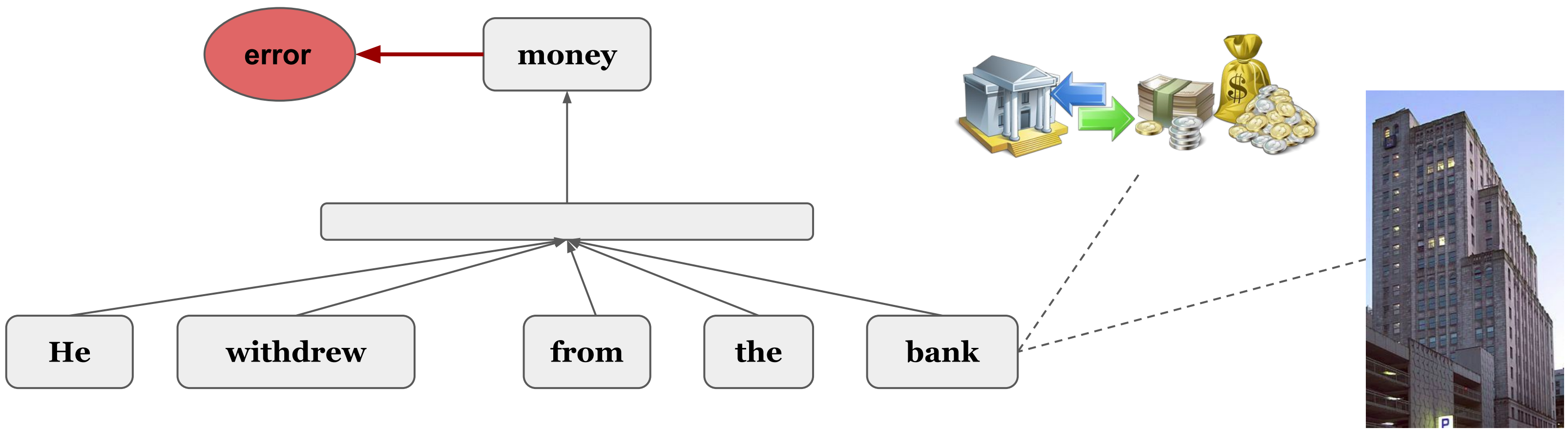


SW2V: Idea

Given as input a corpus and a semantic network:

1. Use a semantic network to link to each word its *associated senses in context*.

2. Use a neural network where the update of word and sense embeddings is linked, exploiting *virtual* connections.

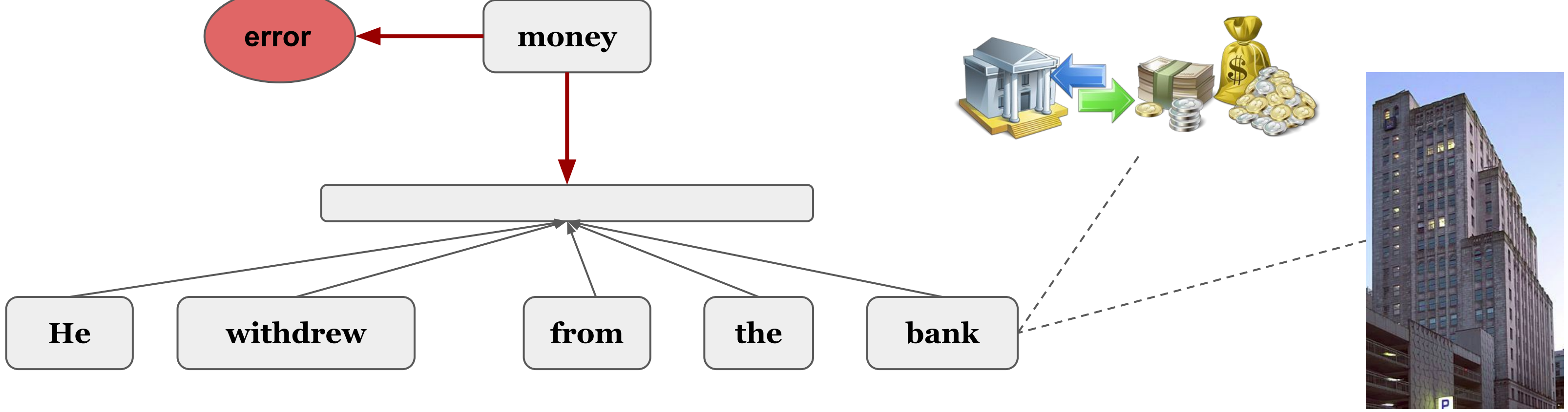


SW2V: Idea

Given as input a corpus and a semantic network:

1. Use a semantic network to link to each word its *associated senses in context*.

2. Use a neural network where the update of word and sense embeddings is linked, exploiting *virtual* connections.

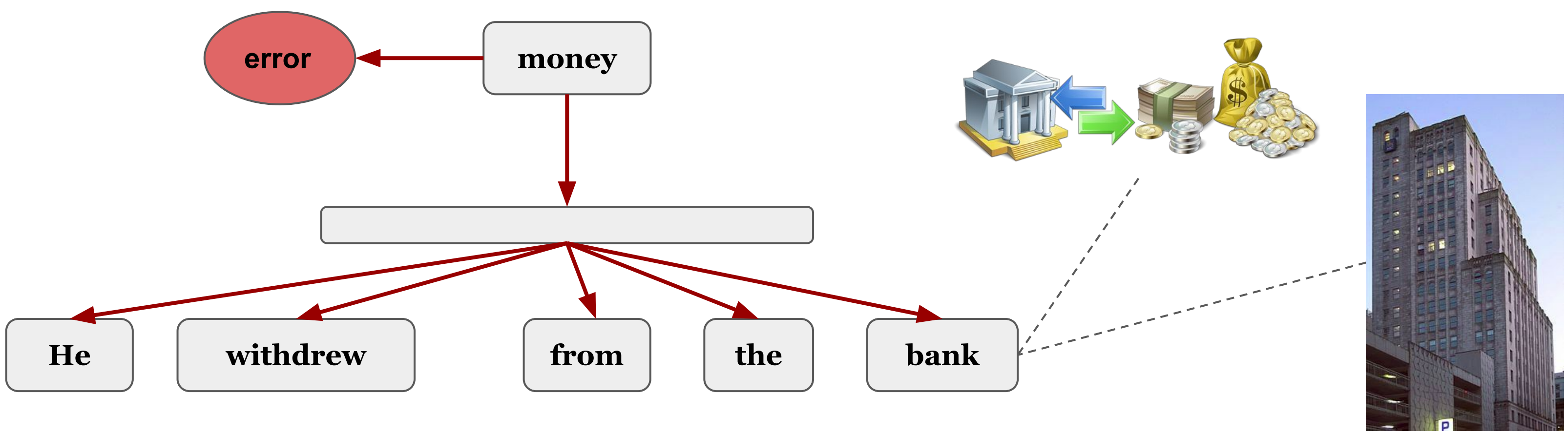


SW2V: Idea

Given as input a corpus and a semantic network:

1. Use a semantic network to link to each word its *associated senses in context*.

2. Use a neural network where the update of word and sense embeddings is linked, exploiting *virtual* connections.

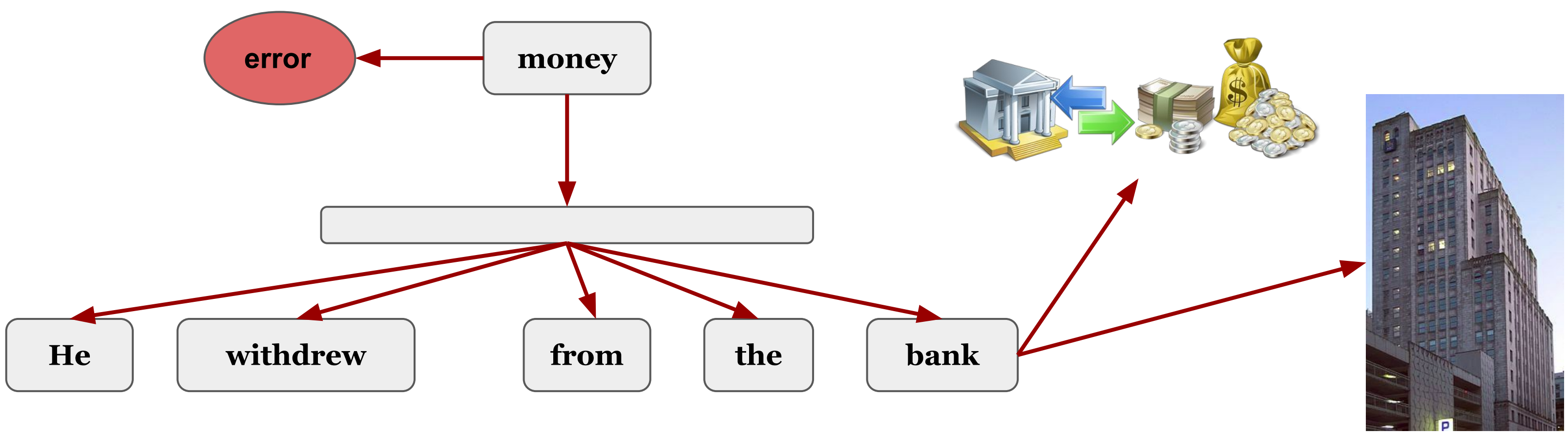


SW2V: Idea

Given as input a corpus and a semantic network:

1. Use a semantic network to link to each word its *associated senses in context*.

2. Use a neural network where the update of word and sense embeddings is linked, exploiting *virtual* connections.



SW2V: Idea

Given as input a corpus and a semantic network:

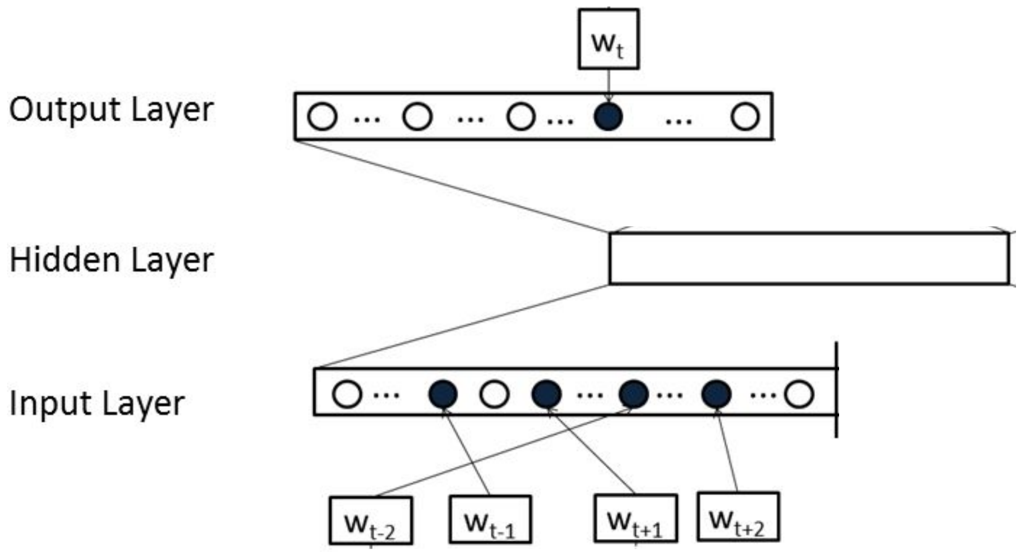
1. Use a semantic network to link to each word its *associated senses in context*.

2. Use a neural network where the update of word and sense embeddings is linked, exploiting *virtual* connections.

In this way it is possible to learn word and sense/synset embeddings jointly on a ***single training****.*

Full architecture of W2V (Mikolov et al., 2013)

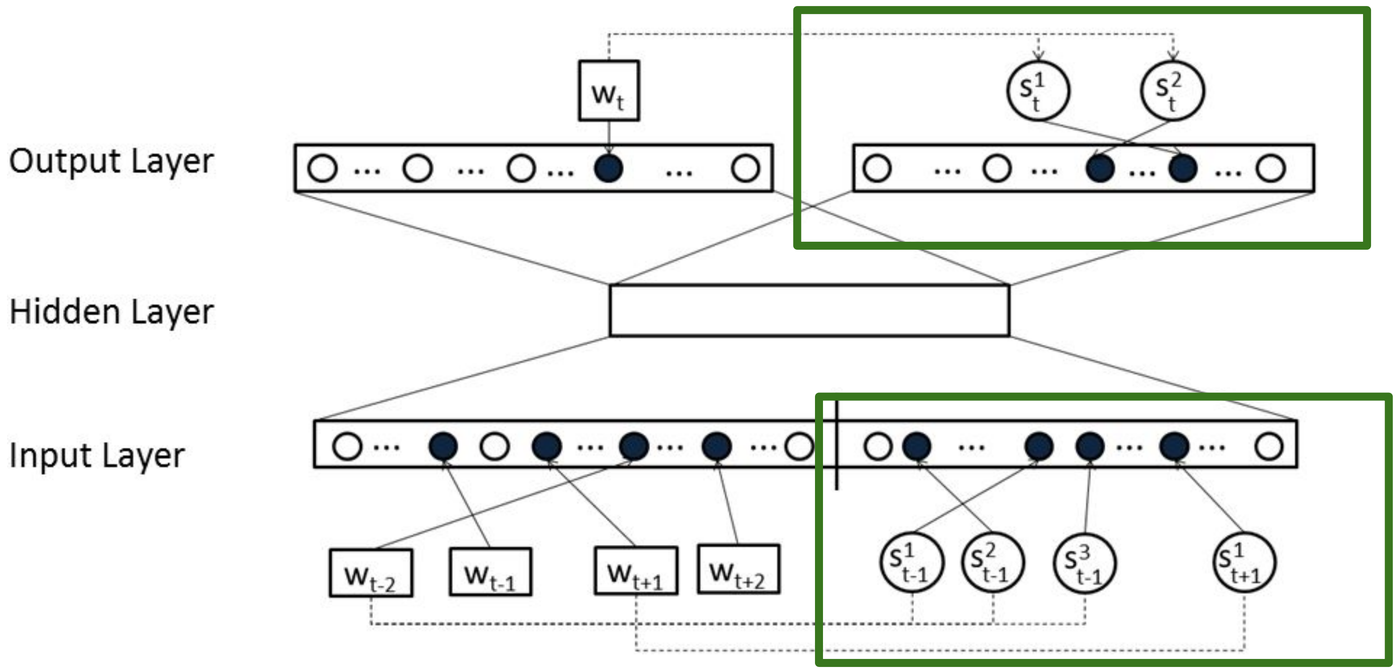
$$E=-\log(p(w_t|W^t))$$



Words and associated senses used both as input and output.

Full architecture of SW2V

$$E=-\log(p(w_t|W^t,S^t)) - \sum_{s\in S^t} \log(p(s|W^t,S^t))$$



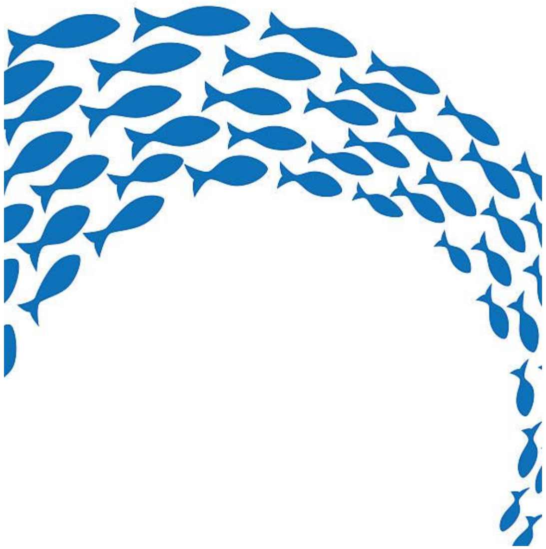
Words and associated senses used both as input and output.

Word and senses connectivity: example 1

company$_n^2$ (military unit)	
AutoExtend	**SW2V**
company$_n^9$	battalion$_n^1$
company	battalion
company$_n^8$	regiment$_n^1$
company$_n^6$	detachment$_n^4$
company$_n^7$	platoon$_n^1$
company$_v^1$	brigade$_n^1$
firm	regiment
business$_n^1$	corps$_n^1$
firm$_n^2$	brigade
company$_n^1$	platoon

Ten closest word and sense embeddings to the sense *company (military unit)*

Word and senses connectivity: example 2



school^7_n (group of fish)

AutoExtend	SW2V
school	schools^7_n
school^4_n	sharks^1_n
school^6_n	sharks
school^1_v	shoals^3_n
school^3_n	fish^1_n
elementary	dolphins^1_n
schools	pods^3_n
elementary^3_a	eels
school^5_n	dolphins
elementary^1_a	whales^2_n

Ten closest word and sense embeddings to the sense *school (group of fish)*

Applications of knowledge-based sense representations

- **Taxonomy Learning** (Espinosa-Anke et al., EMNLP 2016)
- **Open Information Extraction** (Delli Bovi et al. EMNLP 2015).
- **Lexical entailment** (Nickel & Kiela, NIPS 2017)
- **Word Sense Disambiguation** (Rothe & Schütze, ACL 2015)
- **Sentiment analysis** (Flekova & Gurevych, ACL 2016)
- **Lexical substitution** (Cocos et al., SENSE 2017)
- **Computer vision** (Young et al. ICRA 2017)

...

Applications

- ❖ Domain labeling/adaptation
- ❖ Word Sense Disambiguation
- ❖ Downstream NLP applications (e.g. text classification)

Domain labeling

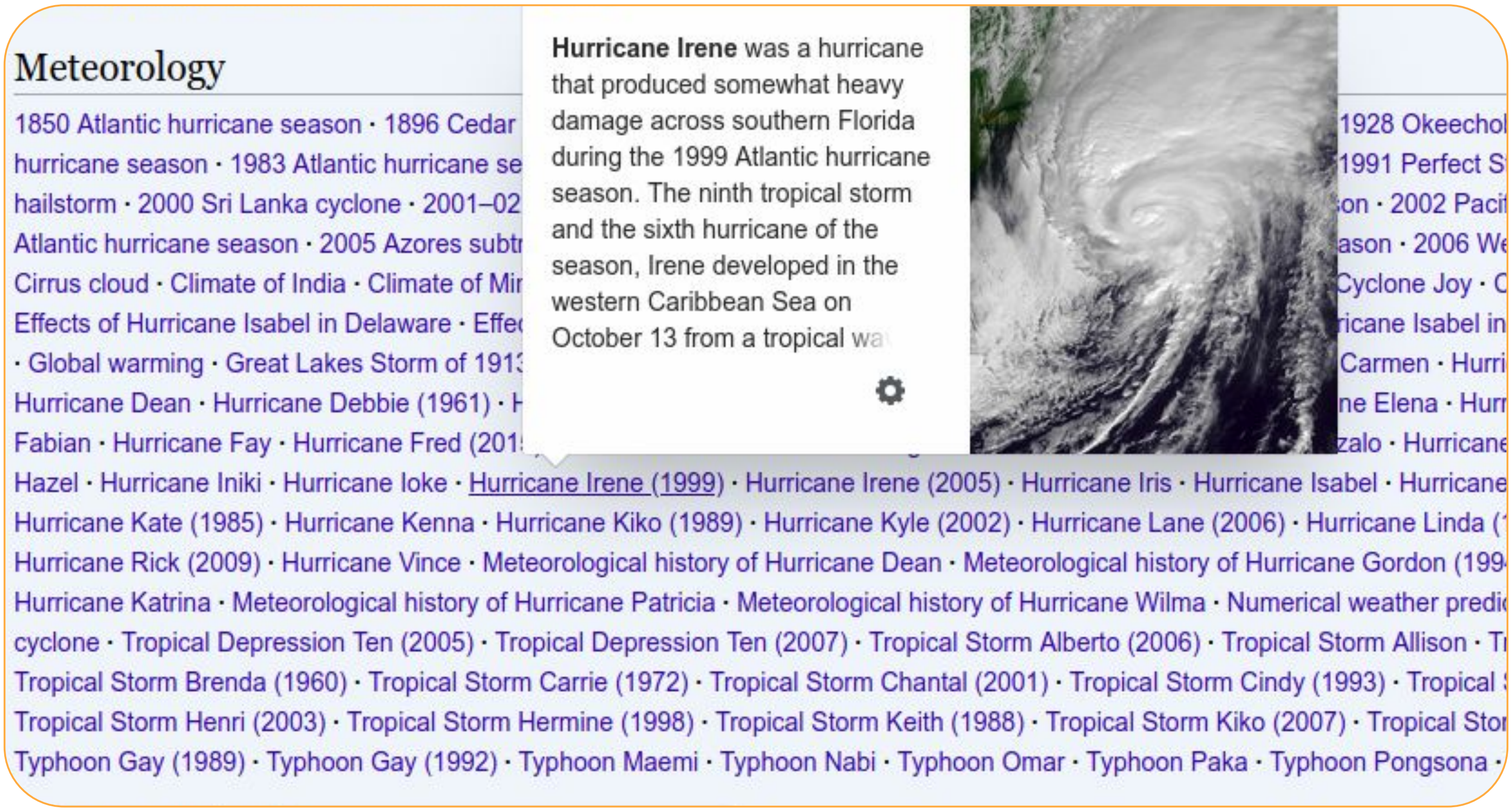
(Camacho-Collados and Navigli, EACL 2017)

Annotate each **concept/entity** with its corresponding **domain of knowledge**.

To this end, we use the Wikipedia featured articles page, which includes 34 domains and a number of Wikipedia pages associated with each domain (*Biology*, *Geography*, *Mathematics*, *Music*, etc.).

Domain labeling

Wikipedia featured articles



Domain labeling

How to associate a concept with a domain?

1. Learn a **NASARI vector** for the concatenation of all Wikipedia pages associated with a given domain.

2. Exploit the **semantic similarity** between knowledge-based vectors and **graph properties** of the lexical resources.

BabelDomains

(Camacho-Collados and Navigli, EACL 2017)

As a result:

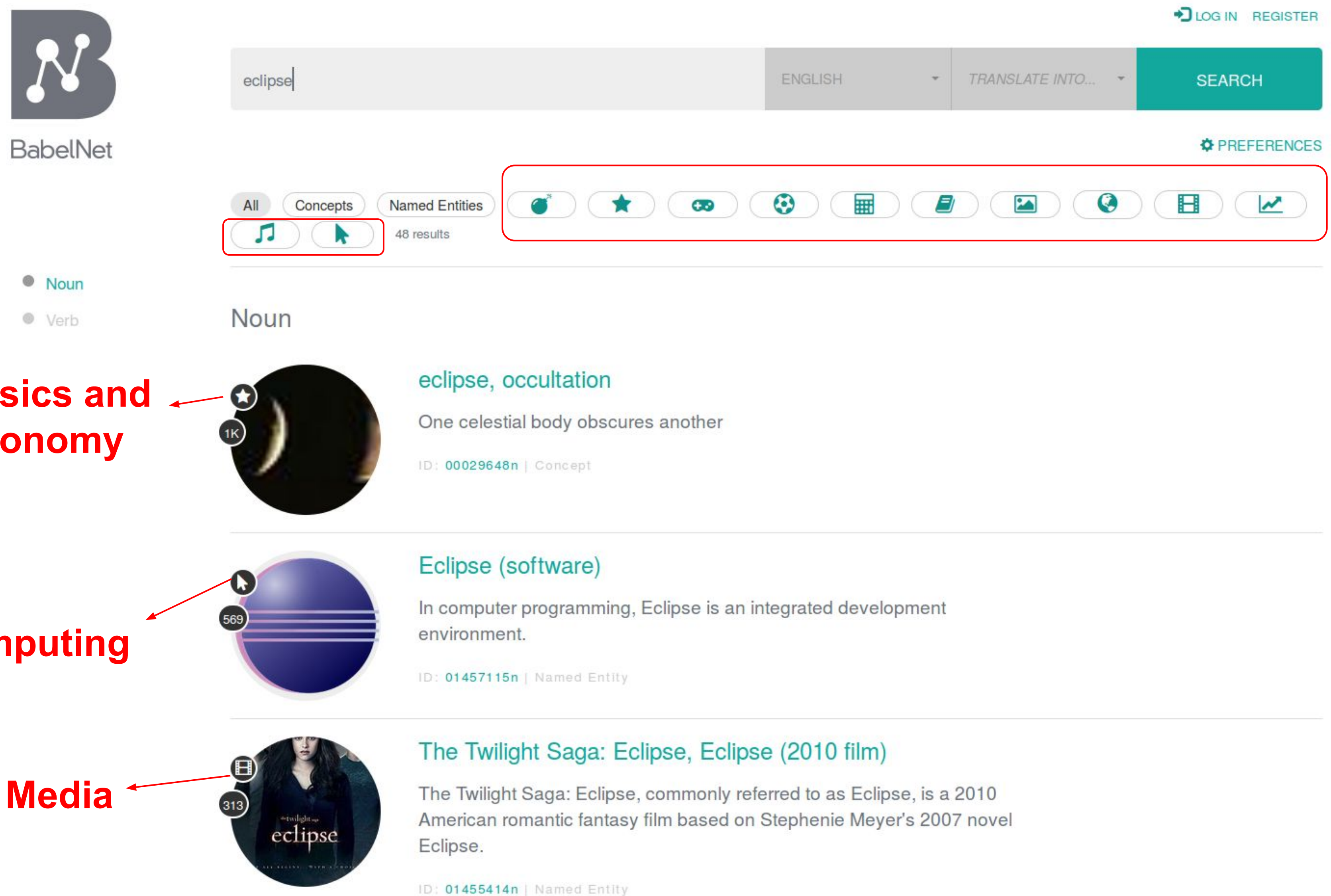
Unified resource with information about domains of knowledge

*BabelDomains available for **BabelNet, Wikipedia and WordNet** available at*

http://lcl.uniroma1.it/babeldomains

Already integrated into BabelNet (online interface and API)

BabelDomains



Physics and astronomy

Computing

Media

Domain filtering for supervised distributional hypernym discovery

(Espinosa-Anke et al., EMNLP 2016;
Camacho-Collados and Navigli, EACL 2017)

Apple

is a

Fruit

Task: Given a term, predict its hypernym(s)

Model: Distributional supervised system based on the transformation matrix of Mikolov et al. (2013).

Idea: Training data filtered by domain of knowledge

Domain filtering for supervised distributional hypernym discovery

Domain-filtered training data

Non-filtered training data

	art			biology			education			geography			health		
Train	MRR	MAP	R-P	MRR	MAP	R-P	MRR	MAP	R-P	MRR	MAP	R-P	MRR	MAP	R-P
5k	0.12	0.12	0.12	0.63	0.63	0.59	0.00	0.00	0.00	0.08	0.07	0.07	0.08	0.08	0.07
15k	0.21	0.20	0.18	**0.84**	0.72	0.79	0.22	0.22	0.21	0.15	0.14	0.14	0.08	0.07	0.07
25k	**0.29**	0.27	**0.26**	**0.84**	**0.83**	**0.81**	0.33	0.32	0.30	**0.23**	**0.22**	**0.21**	0.09	0.09	0.08
25k+K^d_{1k}	**0.29**	**0.28**	**0.26**	**0.84**	0.80	0.79	0.32	0.29	0.27	0.22	**0.22**	**0.21**	0.09	0.09	0.08
25k+K^d_{25k}	0.26	0.24	0.22	0.70	0.63	0.56	**0.38**	**0.36**	**0.33**	0.15	0.13	0.12	0.11	0.11	0.10
25k+K^r_{50k}	0.28	0.26	0.24	0.82	0.77	0.72	0.36	0.33	0.30	0.17	0.16	0.16	**0.12**	0.11	0.10
100k$^r_{wd}$	0.00	0.00	0.00	**0.84**	0.81	0.77	0.00	0.00	0.00	0.01	0.01	0.01	0.07	0.06	0.06
100k$^r_{kbu}$	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	**0.12**	0.12	0.11
Baseline	0.13	0.12	0.10	0.58	0.57	**0.57**	0.10	0.10	0.09	0.12	0.09	0.05	0.07	**0.13**	**0.14**

Results on the hypernym discovery task for five domains

Conclusion: Filtering training data by domains prove to be clearly beneficial

Word Sense Disambiguation

<u>Kobe</u>, which is one of Japan's largest cities, [...]

?

Word Sense Disambiguation

Kobe, which is one of Japan's largest cities, [...]

X

Word Sense Disambiguation

Kobe, which is one of Japan's largest cities, [...]

Word Sense Disambiguation

(Camacho-Collados et al., AIJ 2016)

Basic idea

Select the sense which is semantically closer to the semantic representation of the whole document (**global context**).

$$\hat{d}(s) = \underset{d \in D}{\operatorname{argmax}}\, WO(\vec{\text{NASARI}}_{lex}(s), \vec{v}_{lex}(d))$$

Word Sense Disambiguation on textual definitions

(Camacho-Collados et al., LREC 2016; LREV 2018)

Combination of a graph-based disambiguation system (Babelfy) with NASARI to **disambiguate** the concepts and named entities of **over 35M definitions** in **256 languages**.

Sense-annotated corpus freely available at

http://lcl.uniroma1.it/disambiguated-glosses/

Context-rich WSD

castling *(chess)*

*Interchanging the positions of the **king** and a **rook**.*

Context-rich WSD

castling *(chess)*

*Interchanging the positions of the **king** and a **rook**.*

***Castling** is a move in the game of **chess** involving a player's **king** and either of the player's original **rooks**.*

*A move in which the **king** moves two **squares** towards a **rook**, and the **rook** moves to the other side of the **king.***

Context-rich WSD

castling *(chess)*

*Interchanging the positions of the **king** and a **rook**.*

***Castling** is a move in the game of **chess** involving a player's **king** and either of the player's original **rooks**.*

*A move in which the **king** moves two **squares** towards a **rook**, and the **rook** moves to the other side of the **king.***

*Manœuvre du jeu d'**échecs***

***Rošáda** je zvláštní tah v šachu, při kterém táhne zároveň **král** a **věž**.*

*Spielzug im **Schach**, bei dem **König** und **Turm** einer Farbe bewegt werden*

*El **enroque** es un movimiento especial en el juego de **ajedrez** que involucra al **rey** y a una de las **torres** del jugador.*

***Rokade** er et spesialtrekk i **sjakk**.*

***Rok** İngilizce'de kaleye **rook** denmektedir.*

Το ροκέ είναι μια ειδική **κίνηση** στο **σκάκι** που συμμετέχουν ο βασιλιάς και ένας από τους δυο **πύργους**.

Context-rich WSD

castling *(chess)*

*Interchanging the positions of the **king** and a **rook**.*

***Castling** is a move in the game of **chess** involving a ... r's **king** and either of the player ... s.*

*... which the **king** moves two ... ards a **rook**, and the **rook** ... other side of the **king.***

*Manc ... d'**éche** ...*

***Rošáda** je zvláštní tah v šachu, při kterém táhne zároveň **král** a **věž**.*

*Spielzug im **Schach**, bei dem **König** und **Turm** einer Farbe bewegt werden*

*El **enroque** es un movimiento especial en el juego de **ajedrez** que involucra al **rey** y a una de las **torres** del jugador.*

***Rokade** er et spesialtrekk i **sjakk**.*

***Rok** İngilizce'de kaleye **rook** denmektedir.*

Το ροκέ είναι μια ειδική **κίνηση** στο **σκάκι** που συμμετέχουν ο βασιλιάς και ένας από τους δυο **πύργους**.

Context-rich WSD exploiting parallel corpora

(Delli Bovi et al., ACL 2017)

Applying the same method to provide **high-quality sense annotations from parallel corpora** (Europarl): 120M+ sense annotations for 21 languages.

http://lcl.uniroma1.it/eurosense/

Extrinsic evaluation: Improved performance of a standard supervised WSD system using this automatically sense-annotated corpora.

Towards a seamless integration of senses in downstream NLP applications

(Pilehvar et al., ACL 2017)

Question: What if we apply WSD and inject sense embeddings to a standard neural classifier?

Problems:

Towards a seamless integration of senses in downstream NLP applications

(Pilehvar et al., ACL 2017)

Question: What if we apply WSD and inject sense embeddings to a standard neural classifier?

Problems:

- WSD is not perfect

Towards a seamless integration of senses in downstream NLP applications

(Pilehvar et al., ACL 2017)

Question: What if we apply WSD and inject sense embeddings to a standard neural classifier?

Problems:

- WSD is not perfect **-> Solution:** High-confidence disambiguation

Towards a seamless integration of senses in downstream NLP applications

(Pilehvar et al., ACL 2017)

Question: What if we apply WSD and inject sense embeddings to a standard neural classifier?

Problems:

- WSD is not perfect **-> Solution:** High-confidence disambiguation

- WordNet lacks coverage

Towards a seamless integration of senses in downstream NLP applications

(Pilehvar et al., ACL 2017)

Question: What if we apply WSD and inject sense embeddings to a standard neural classifier?

Problems:

- WSD is not perfect **-> Solution:** High-confidence disambiguation

- WordNet lacks coverage **-> Solution:** Use of Wikipedia

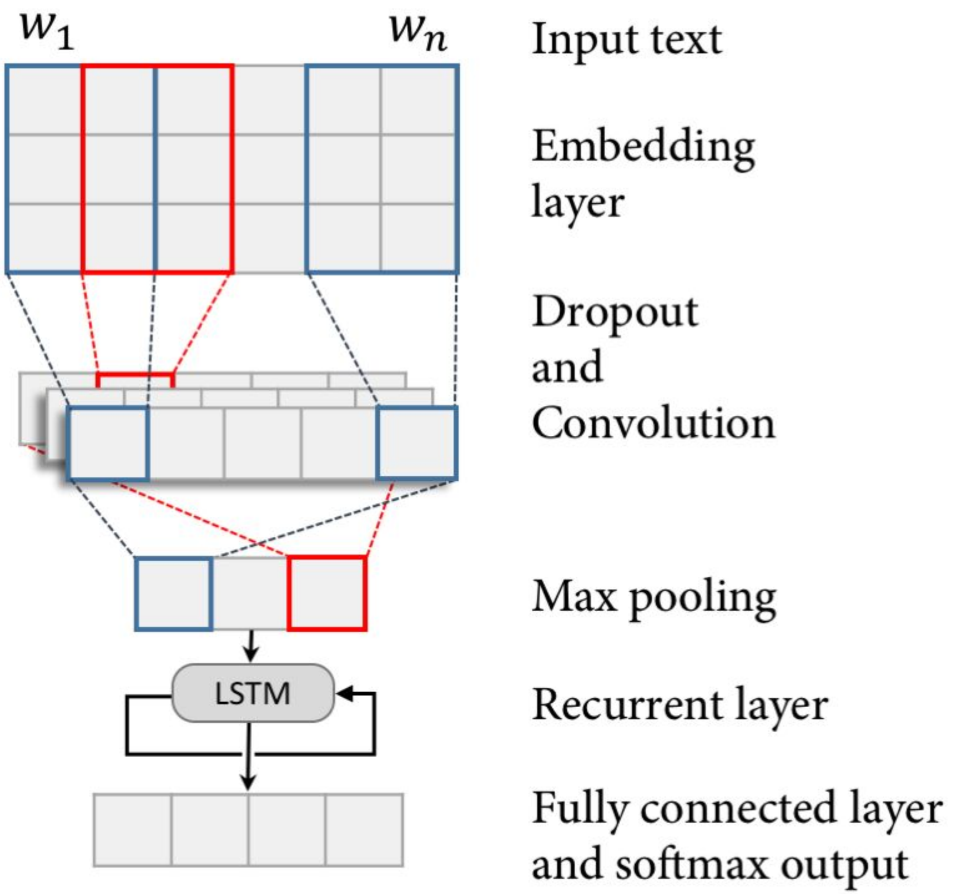
Tasks: Topic categorization and sentiment analysis (polarity detection)

Topic categorization: Given a text, assign it a topic (e.g. politics, sports, etc.).

Polarity detection: Predict the sentiment of the sentence/review as either positive or negative.

Classification model

Standard CNN classifier inspired by Kim (2014)

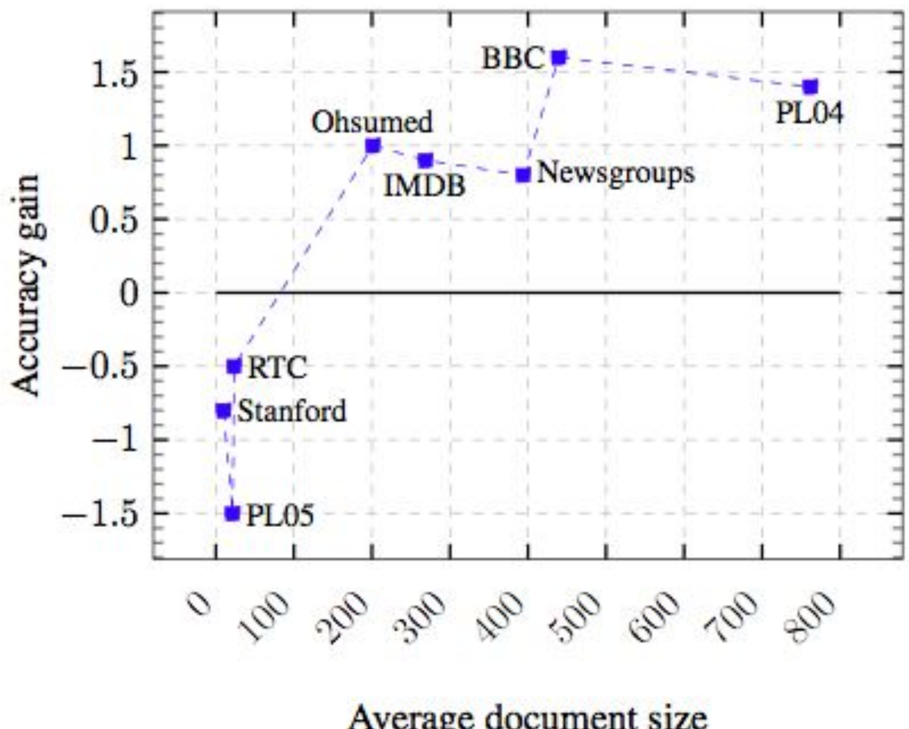


Sense-based vs. word-based: Conclusions

Sense-based **better** than word-based... when the **input text is large enough**

Sense-based vs. word-based:

Sense-based **better** than word-based… when the **input text is large enough**:



Why
does the input text size matter?

- Word sense disambiguation works better in larger texts (Moro et al. 2014; Raganato et al. 2017)

- Disambiguation increases sparsity

Contextualized word embeddings ELMo/BERT

Peters et al.
(NAACL 2018)

Devlin et al.
(NAACL 2019)

Contextualized word embeddings
ELMo/BERT

New AI fake text generator may be too dangerous to release, say creators

The Elon Musk-backed nonprofit company OpenAI declines to release research publicly for fear of misuse

Contextualized word embeddings
ELMo/BERT

As word embeddings, learned by leveraging language models on **massive amounts of text corpora**.

Contextualized word embeddings
ELMo/BERT

As word embeddings, learned by leveraging language models on **massive amounts of text corpora**.

New: each word vector depends on the context. It is **dynamic**.

Contextualized word embeddings
ELMo/BERT

As word embeddings, learned by leveraging language models on **massive amounts of text corpora**.

New: each word vector depends on the context. It is **dynamic**.

Important **improvements in many NLP tasks**.

Contextualized word embeddings ELMo/BERT (examples)

*He withdrew money from the **bank**.*

*The **bank** remained closed yesterday.*

*We found a nice spot by the **bank** of the river.*

Contextualized word embeddings ELMo/BERT (examples)

0.25, 0.32, -0.1 ….

He withdrew money from the ***bank****.*

0.22, 0.30, -0.08 ….

The ***bank*** *remained closed yesterday.*

-0.8, 0.01, 0.3 ….

We found a nice spot by the ***bank*** *of the river.*

Contextualized word embeddings ELMo/BERT (examples)

Similar vectors

0.25, 0.32, -0.1 ….

He withdrew money from the ***bank****.*

0.22, 0.30, -0.08 ….

The ***bank*** *remained closed yesterday.*

-0.8, 0.01, 0.3 ….

We found a nice spot by the ***bank*** *of the river.*

How well do these models capture “meaning”?

Good enough for many applications.

Room for improvement. No noticeable improvements in:

- ➢ *Winograd Schema Challenge*: BERT ~65% **vs** Humans ~95%

- ➢ *Word-in-Context Challenge*: BERT ~65% **vs** Humans ~85%

How well do these models capture “meaning”?

Good enough for many applications.

Room for improvement. No noticeable improvements in:

- *Winograd Schema Challenge*: BERT ~65% **vs** Humans ~95%

  requires commonsense reasoning

- *Word-in-Context Challenge*: BERT ~65% **vs** Humans ~85%

  requires abstracting the notion of sense

For more information on meaning representations (embeddings):

- ❖ ACL 2016 Tutorial on "**Semantic representations of word senses and concepts**": http://josecamachocollados.com/slides/Slides_ACL16Tutorial_SemanticRepresentation.pdf

- ❖ EACL 2017 workshop on "**Sense, Concept and Entity Representations and their Applications**": https://sites.google.com/site/senseworkshop2017/

- ❖ NAACL 2018 Tutorial on "**Interplay between lexical resources and NLP**": https://bitbucket.org/luisespinosa/lr-nlp/

- ❖ **"From Word to Sense Embeddings: A Survey on Vector Representations of Meaning"** (JAIR 2018): https://www.jair.org/index.php/jair/article/view/11259

Thank you!

Questions please!

camachocolladosj@cardiff.ac.uk
@CamachoCollados
josecamachocollados.com