I’m a huge metal fan!

Mariana Romanyshyn
Computational Linguist at Grammarly, Inc.

1.

The Matter of Meaning

Words have meanings

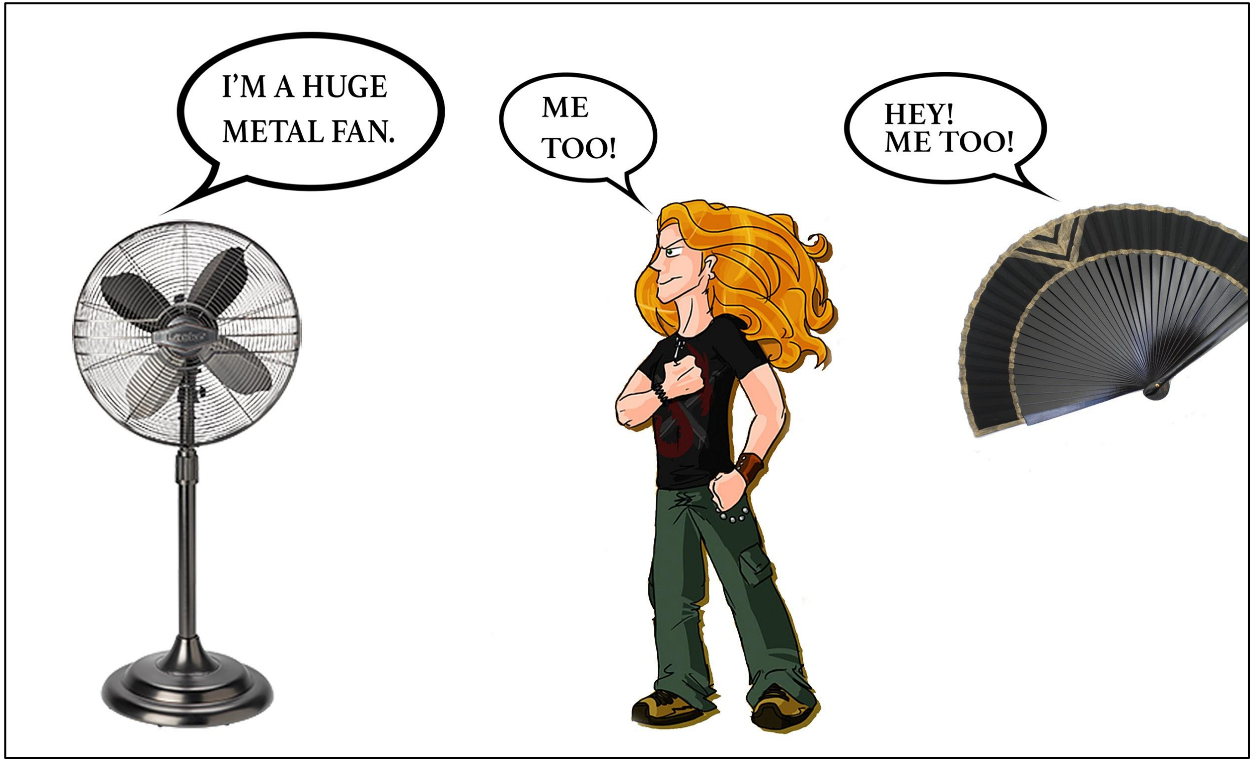


Image by Tetiana Turchyn

Homonymy vs. Polysemy

Homonymous **“bank”**

- a financial institution
- an area of land along the side of a river

Polysemous **“man”**

- the humanity
- male part of the humanity
- adult male part of the humanity

Homonymy vs. Polysemy

Homonymous **“bank”**

- a financial institution
- an area of land along the side of a river

Polysemous **“man”**

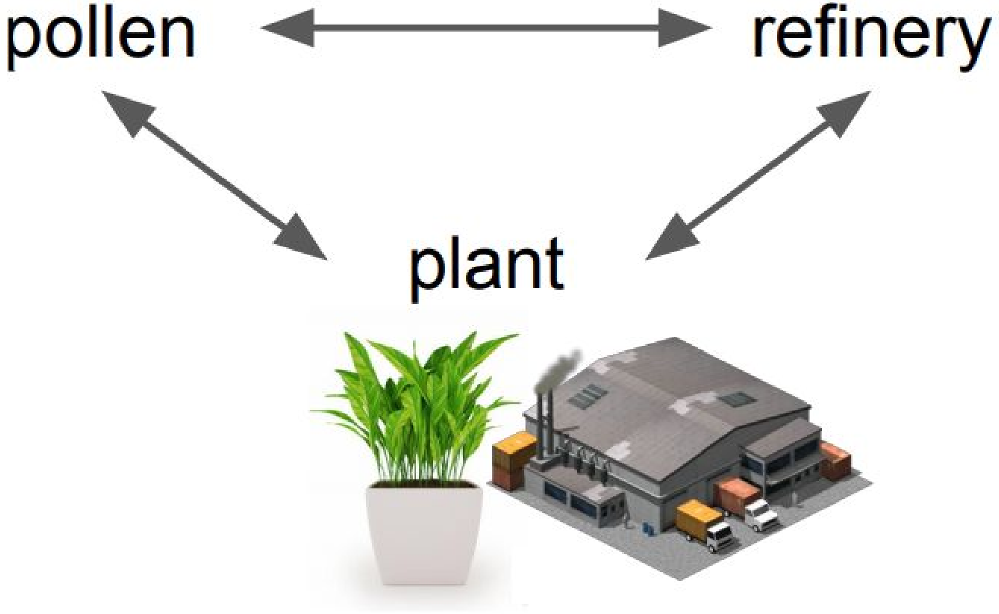
- the humanity
- male part of the humanity
- adult male part of the humanity
- ~~a person~~

Is it serious?

- ~40% of English words are polysemous
- most polysemous - verbs (~55% in WordNet)
- resources disagree
  - *“head”*, noun:
    - 11 meanings - Macmillan Dictionary
    - 16 meanings - Longman Dictionary
    - 33 meanings - WordNet
    - 34 meanings - Oxford Dictionary
- meanings overlap
  - *John works for the **newspaper** that you are reading.*

What does it mean for NLP?

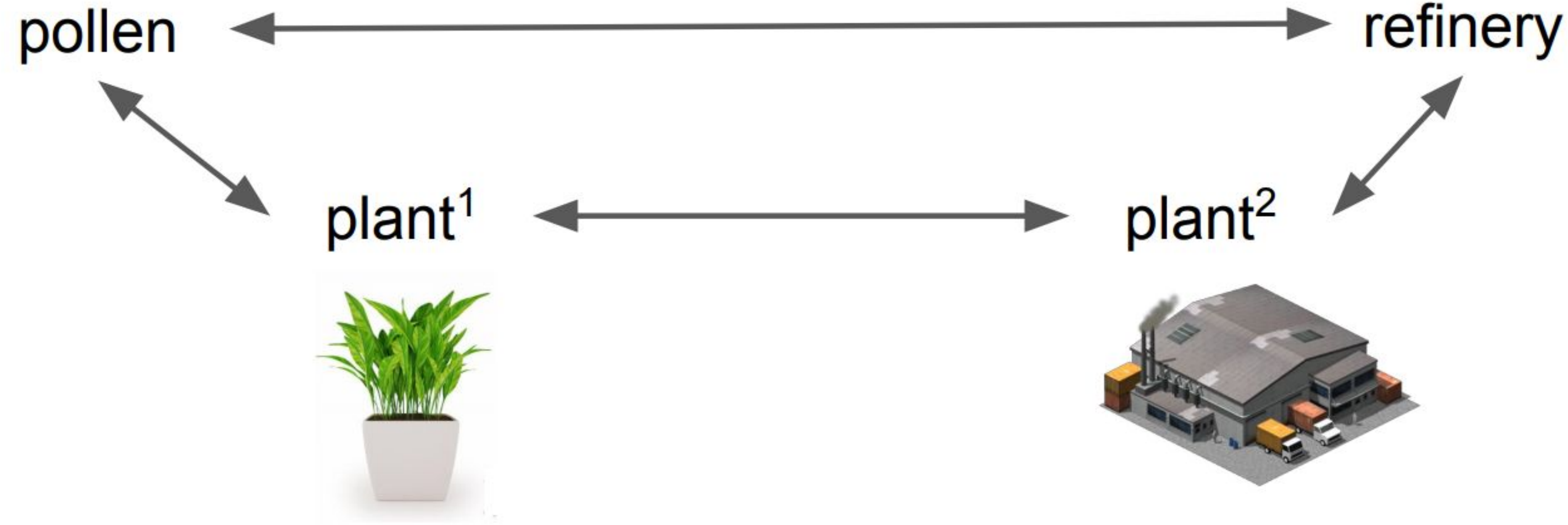
Triangle inequality in word embeddings.



Example from Neelakantan et al. (2014)

What does it mean for NLP?

Word embeddings => sense embeddings



Example from Neelakantan et al. (2014)

Is it just English?

... зробити так, щоби впала ***стіна****?*

- стіна будинку
- стіни айсбергів
- мур
- те, що відокремлює, роз'єднує

Can’t deep learning just figure it out?

Text classification/mining

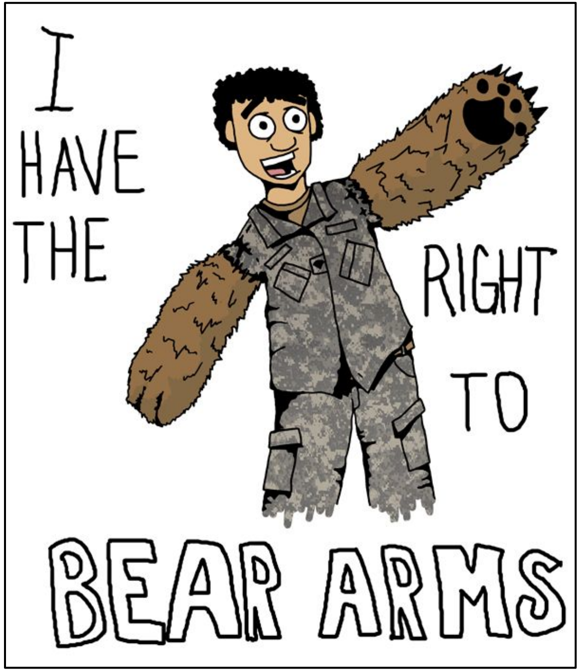
US **sells arms** to countries well-known for violating human rights.

Using recycled prosthesis, a hospital in Tanzania **sells arms** for around $500 each. There is also high demand for legs.

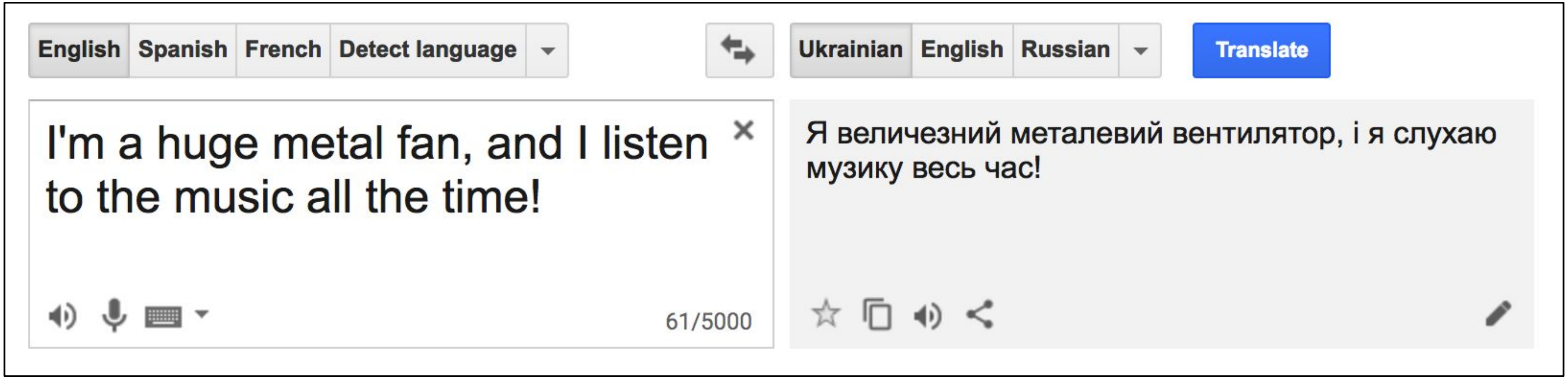
Text classification/mining

US **sells arms** to countries well-known for violating human rights.

Using recycled prosthesis, a hospital in Tanzania **sells arms** for around $500 each. There is also high demand for legs.



Machine translation

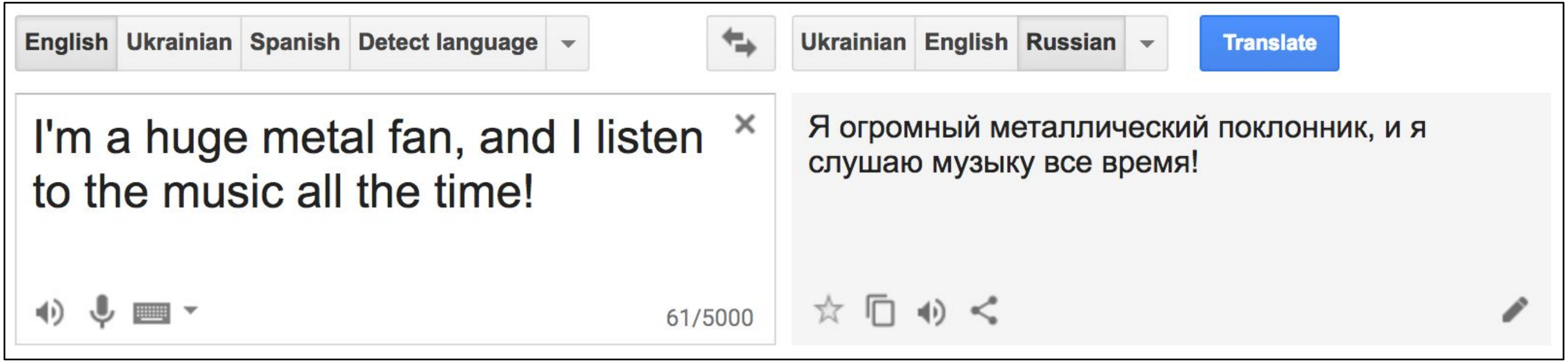


I'm a huge metal fan, and I listen to the music all the time!

Я величезний металевий вентилятор, і я слухаю музику весь час!

Example from Google Translate

Machine translation



I'm a huge metal fan, and I listen to the music all the time!

Я огромный металлический поклонник, и я слушаю музыку все время!

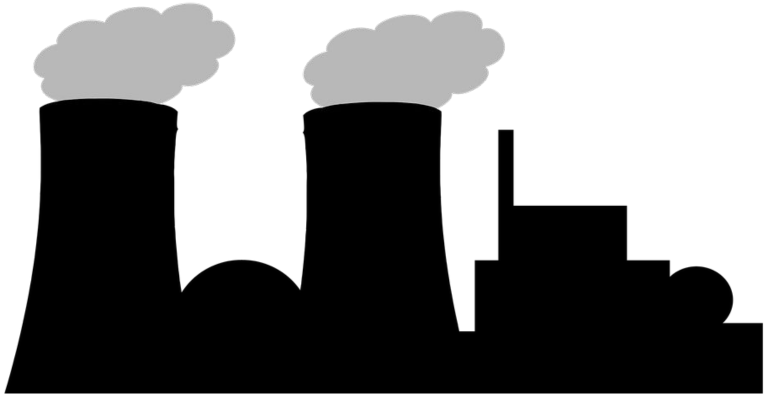
Example from Google Translate

Personal assistants

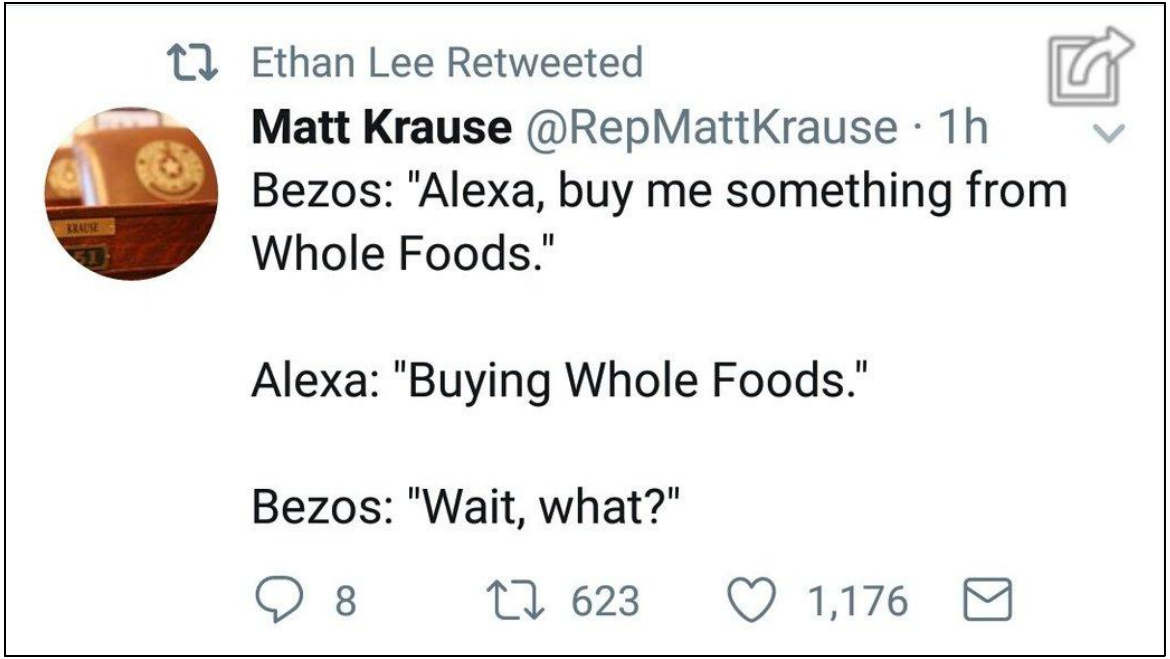
You: I need to buy a big **plant** for my mom. She likes gardening!

Siri: Hmm...

?



Personal assistants



Ethan Lee Retweeted

Matt Krause @RepMattKrause · 1h

Bezos: "Alexa, buy me something from Whole Foods."

Alexa: "Buying Whole Foods."

Bezos: "Wait, what?"

Sentiment analysis

Interest rates are very **high**.

These socks are a little **high**.

This area is **rich** in natural resources.

These comments are a bit **rich** coming from someone with no money worries.

Sentiment analysis

Interest rates are very **high**.

These socks are a little **high**. (= smelly)

This area is **rich** in natural resources.

These comments are a bit **rich** coming from someone with no money worries.

Sentiment analysis

Interest rates are very **high**.

These socks are a little **high**. (= smelly)

This area is **rich** in natural resources.

These comments are a bit **rich** coming from someone with no money worries.

Error correction

Abstract or concrete?

Man is rapidly destroying the earth.

Do you recognize **man** in the grey suit?

Error correction

Abstract or concrete?

Man is rapidly destroying the earth.

Do you recognize **the** **man** in the grey suit?

Error correction

Countable or uncountable?

This is a minor but moving **work** of literature.

Employees may take a **work** home if they wish.

Error correction

Countable or uncountable?

This is a minor but moving **work** of literature.

Employees may take ~~a~~ **work** home if they wish.

Error correction

Standard vs. non-standard

I believe women should be paid the same as **men**.

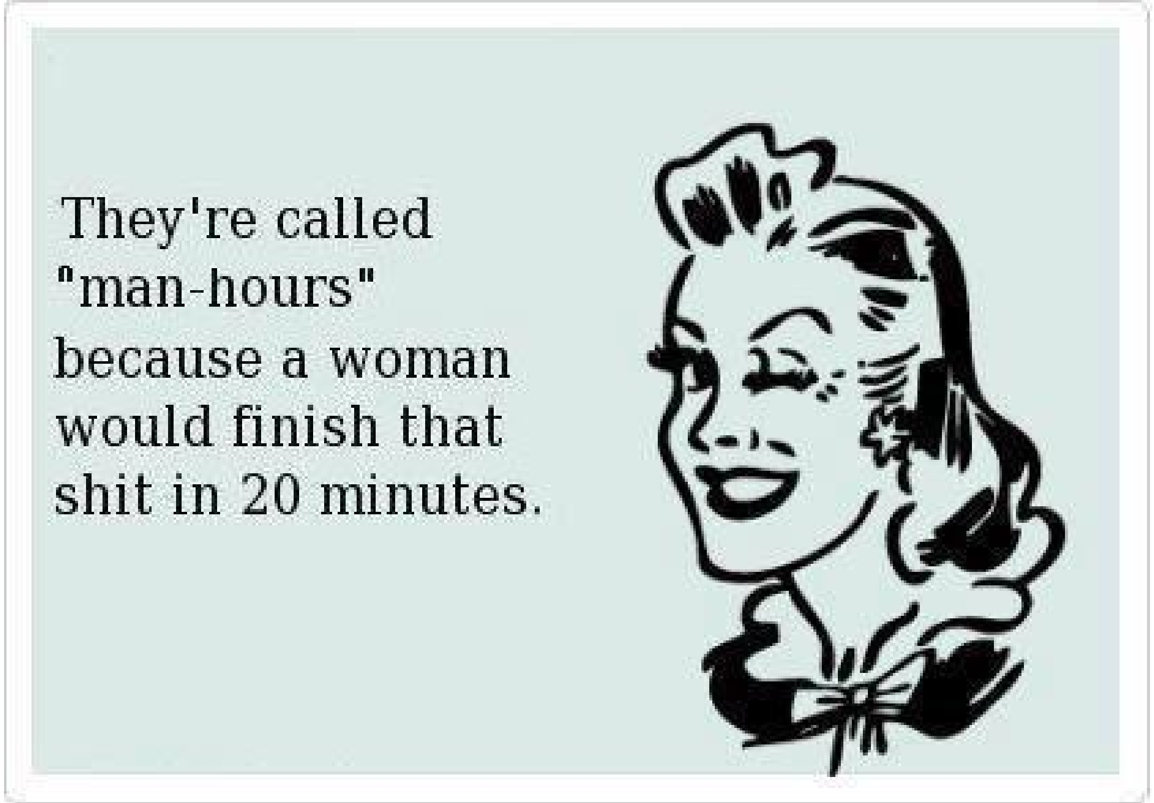
All **men** are equal in the sight of the law.

Error correction

Standard vs. non-standard

I believe women should be paid the same as **men**.

All **{men=>people}** are equal in the sight of the law.



They're called "man-hours" because a woman would finish that shit in 20 minutes.

Error correction

Animate or inanimate?

The software learns **models** from large quantities of data.

How to learn a **model** to flip her hair.

The **chair** was placed in the museum. He's part of the exhibit now.

The **chair** was awarded for a poem. He’s famous now.

Error correction

Animate or inanimate?

The software learns **models** from large quantities of data.

How to **{learn=>teach}** a **model** to flip her hair.

The **chair** was placed in the museum. He's part of the exhibit now.

The **chair** was awarded for a poem. He’s famous now.

Error correction

Animate or inanimate?

The software learns **models** from large quantities of data.

How to **{learn=>teach}** a **model** to flip her hair.

The **chair** was placed in the museum. **{He=>It}**'s part of the exhibit now.

The **chair** was awarded for a poem. He’s famous now.

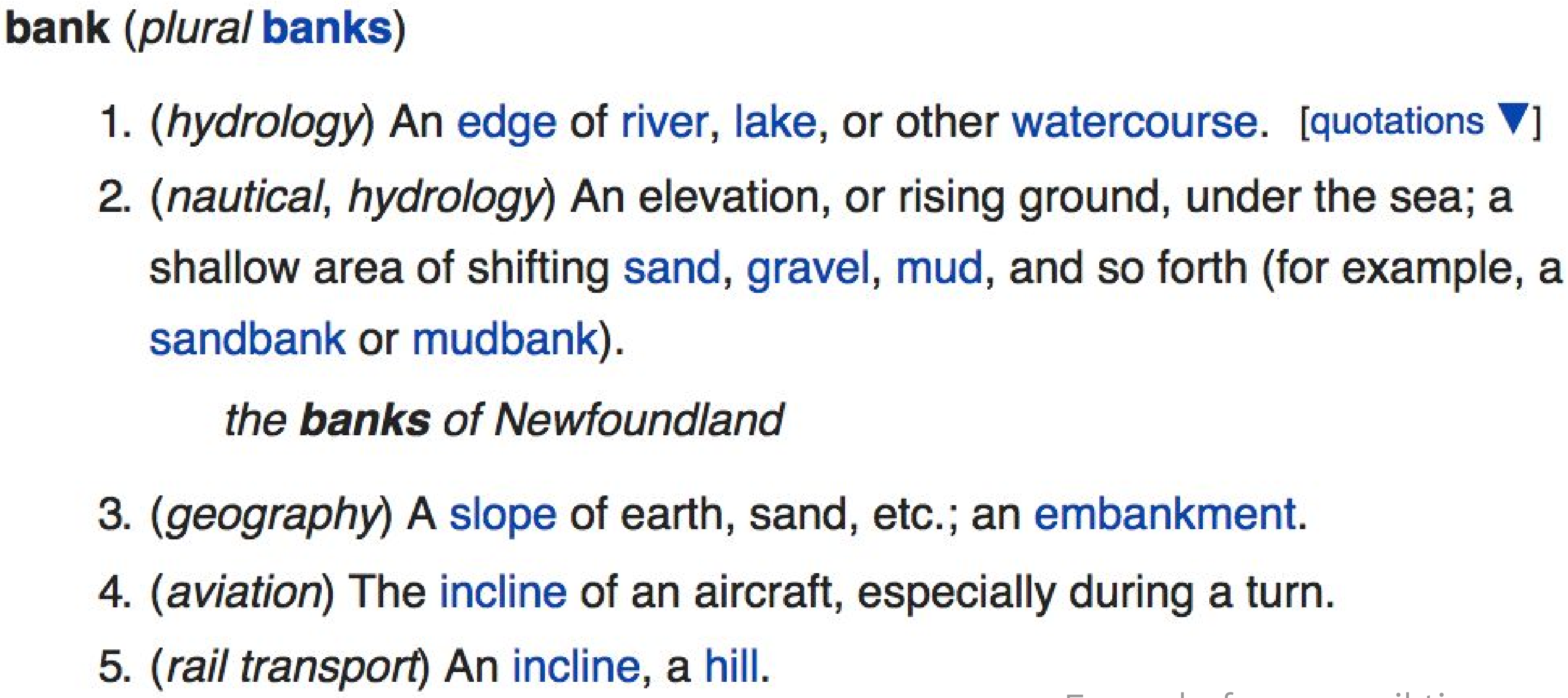
What is “sense” than?

- senses = domains?
- senses = sentiments?
- senses = animate/inanimate?
- senses = jargon/standard?
- senses = countable/uncountable?
- senses = senses?

2.

Resources

Dictionaries



bank (*plural* **banks**)

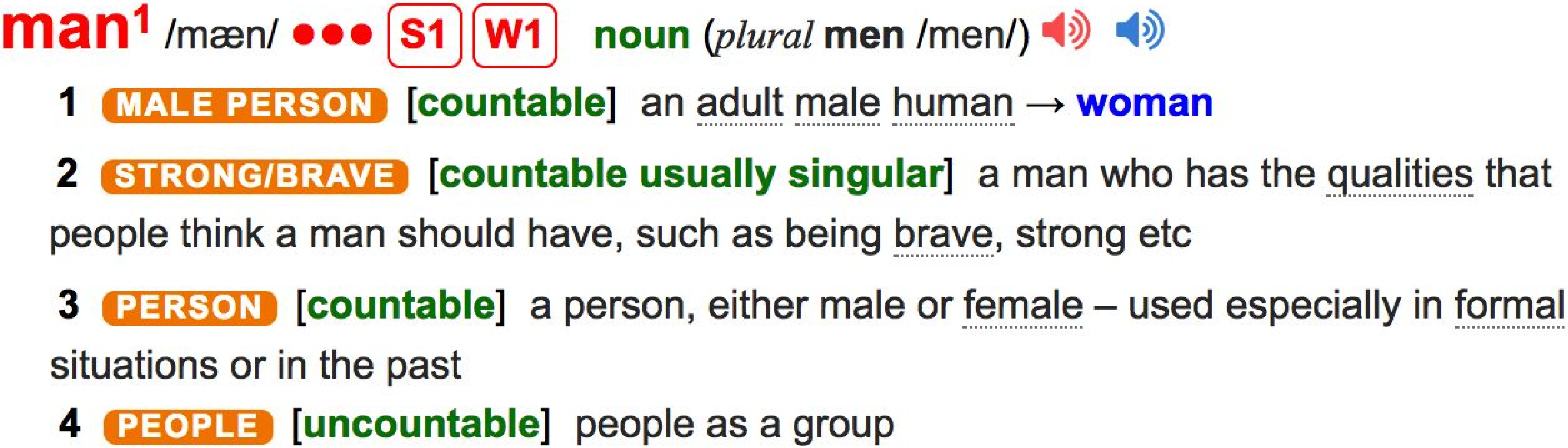
1. (*hydrology*) An edge of river, lake, or other watercourse. [quotations ▼]
2. (*nautical*, *hydrology*) An elevation, or rising ground, under the sea; a shallow area of shifting sand, gravel, mud, and so forth (for example, a sandbank or mudbank).

   *the **banks** of Newfoundland*

3. (*geography*) A slope of earth, sand, etc.; an embankment.
4. (*aviation*) The incline of an aircraft, especially during a turn.
5. (*rail transport*) An incline, a hill.

Example from en.wiktionary.org

Dictionaries



man^1 /mæn/ ●●● S1 W1 **noun** (*plural* **men** /men/)

1 MALE PERSON [**countable**] an adult male human → **woman**

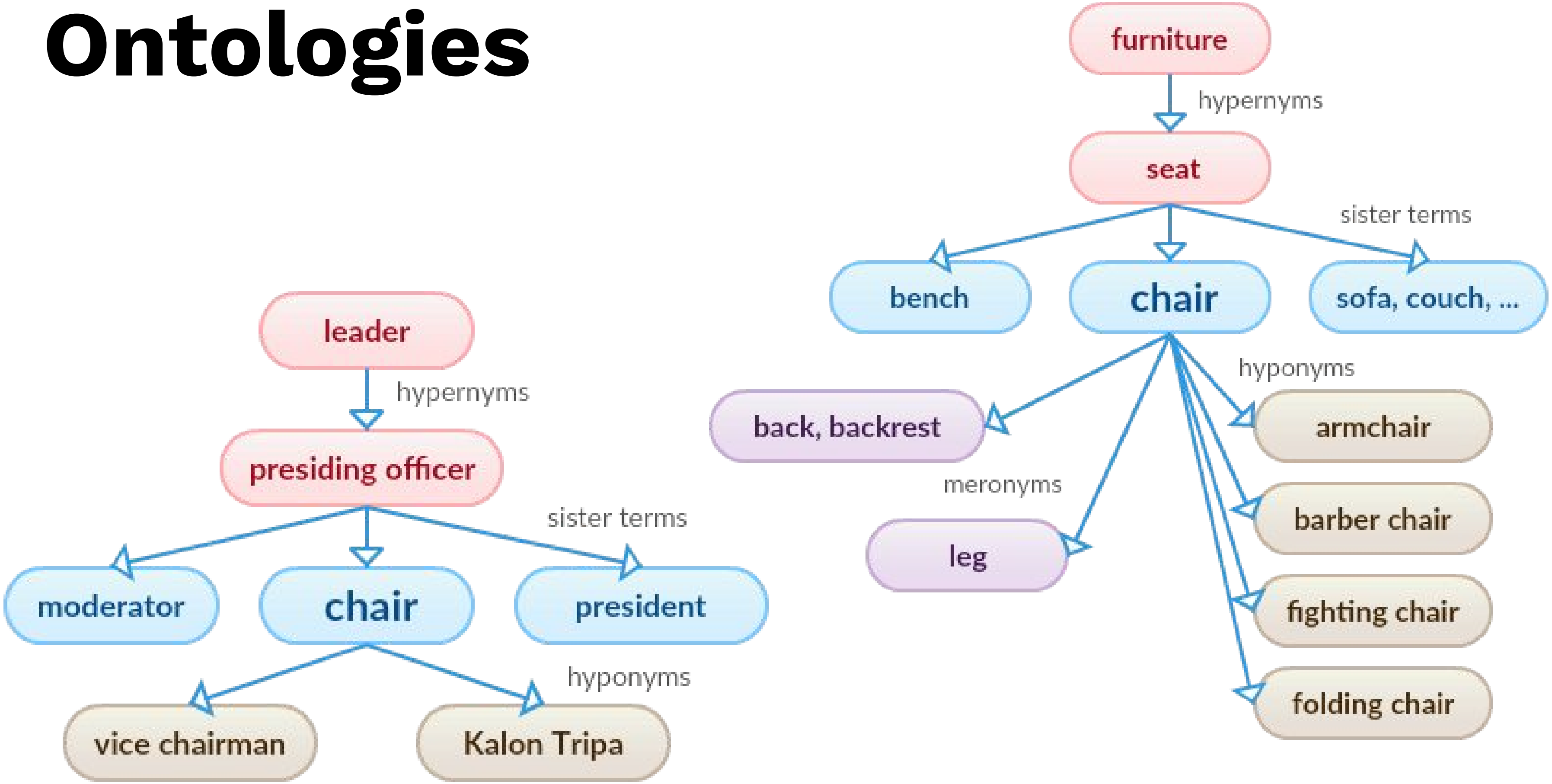
2 STRONG/BRAVE [**countable usually singular**] a man who has the qualities that people think a man should have, such as being brave, strong etc

3 PERSON [**countable**] a person, either male or female – used especially in formal situations or in the past

4 PEOPLE [**uncountable**] people as a group

Example from www.ldoceonline.com

Ontologies

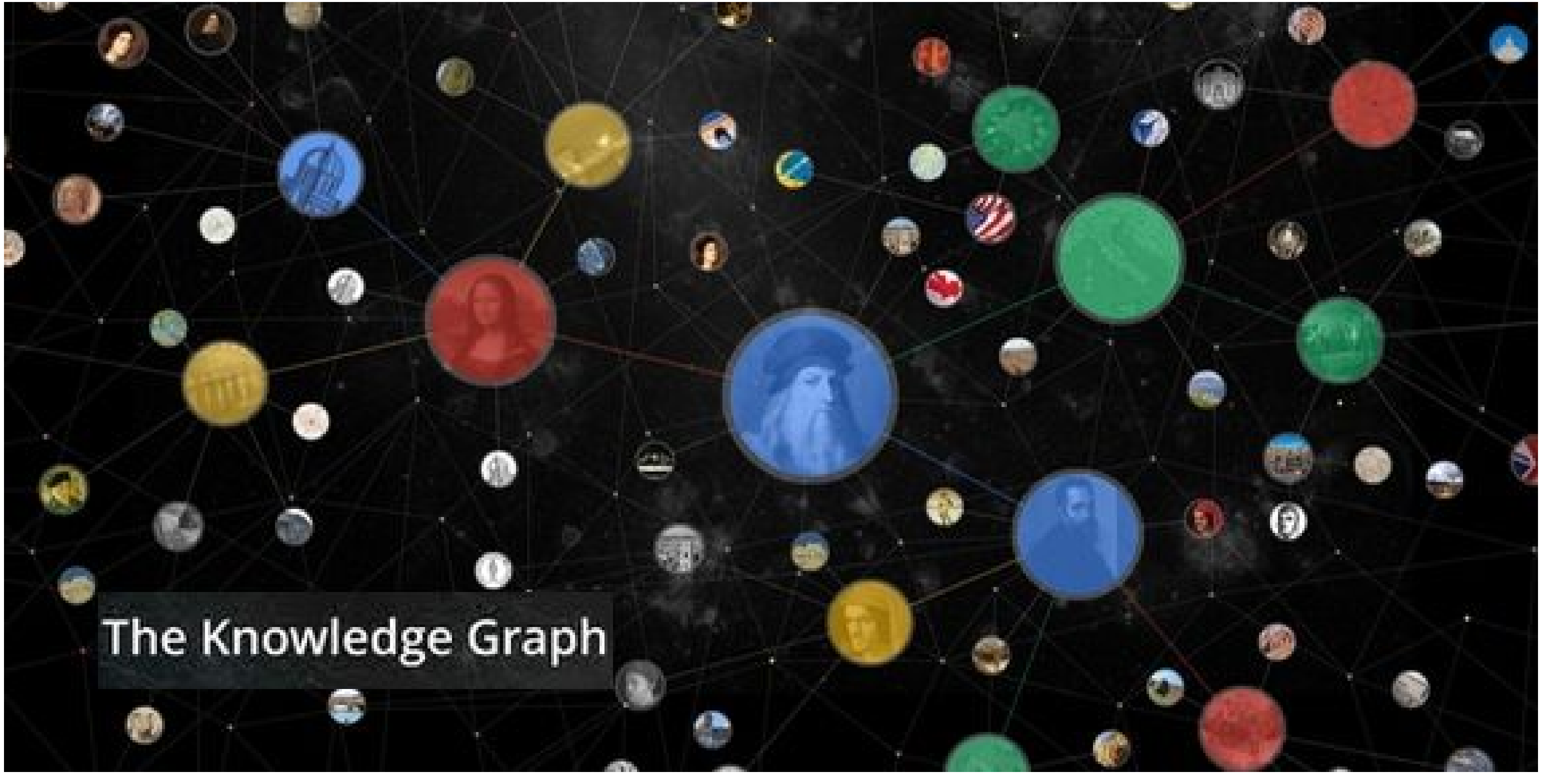


furniture
hypernyms
seat
sister terms
bench
chair
sofa, couch, ...
hyponyms
back, backrest
meronyms
leg
armchair
barber chair
fighting chair
folding chair

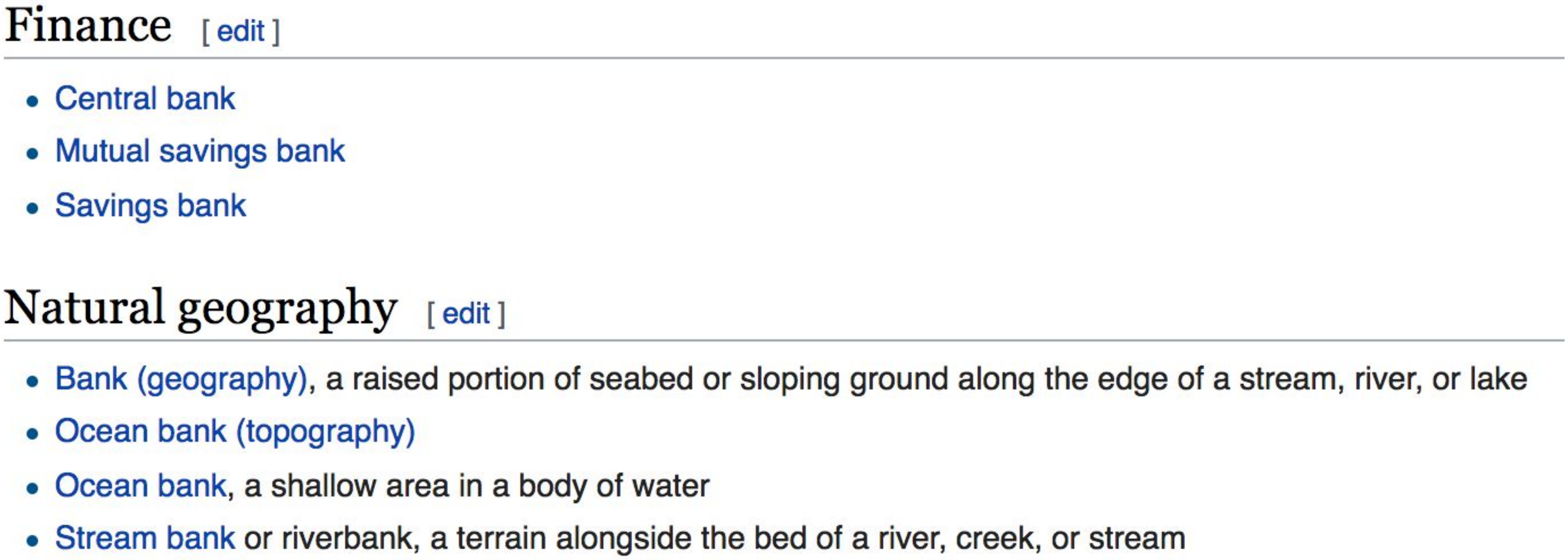
leader
hypernyms
presiding officer
sister terms
moderator
chair
president
hyponyms
vice chairman
Kalon Tripa

Example of relations in WordNet

Knowledge Graph



Wikipedia, Wikidata, DBpedia



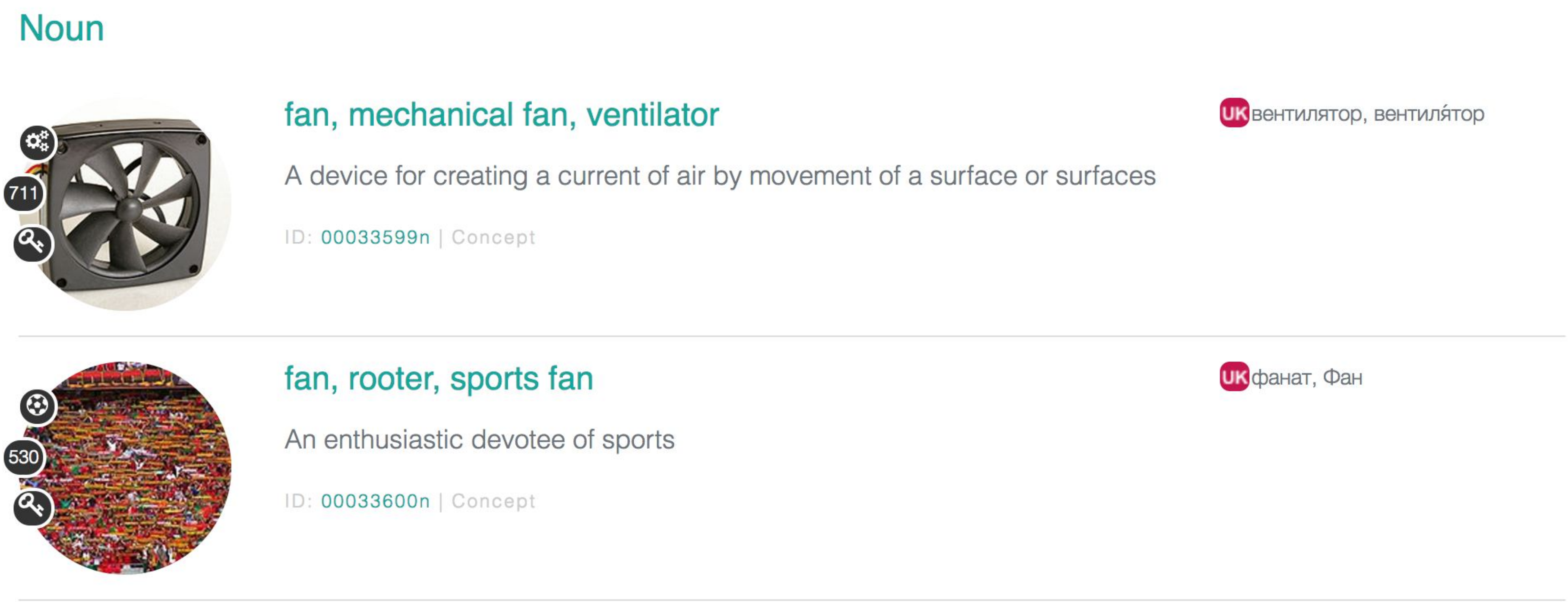
Finance [edit]

- Central bank
- Mutual savings bank
- Savings bank

Natural geography [edit]

- Bank (geography), a raised portion of seabed or sloping ground along the edge of a stream, river, or lake
- Ocean bank (topography)
- Ocean bank, a shallow area in a body of water
- Stream bank or riverbank, a terrain alongside the bed of a river, creek, or stream

BabelNet



Noun

fan, mechanical fan, ventilator

UK вентилятор, вентиля́тор

A device for creating a current of air by movement of a surface or surfaces

ID: 00033599n | Concept

fan, rooter, sports fan

UK фанат, Фан

An enthusiastic devotee of sports

ID: 00033600n | Concept

Example from babelnet.org

Corpora: SemCor

```
<wf>The</wf>
<wf lemma="model" wnsn="3">model</wf>
<wf lemma="quite" wnsn="1">quite</wf>
<wf lemma="plainly" wnsn="1">plainly</wf>
<wf lemma="think" wnsn="1">thought</wf>
<wf lemma="person" wnsn="1">Michelangelo</wf>
<wf lemma="crazy" wnsn="1">crazy</wf>
<wf>;</wf>
```

http://web.eecs.umich.edu/~mihalcea/downloads.html

Corpora: Wikipedia

Beverly Johnson (born October 13, 1952) is an **[American|"United States"] [model|"Model (person)"]**, **[actress|"Actress"]**, **[singer|"Singer"]**, and **[businesswoman|"Businesswoman"]**.

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

3.

Supervised word-sense disambiguation

If you have a corpus...

Features:

- collocations
- bag of words

Containing:

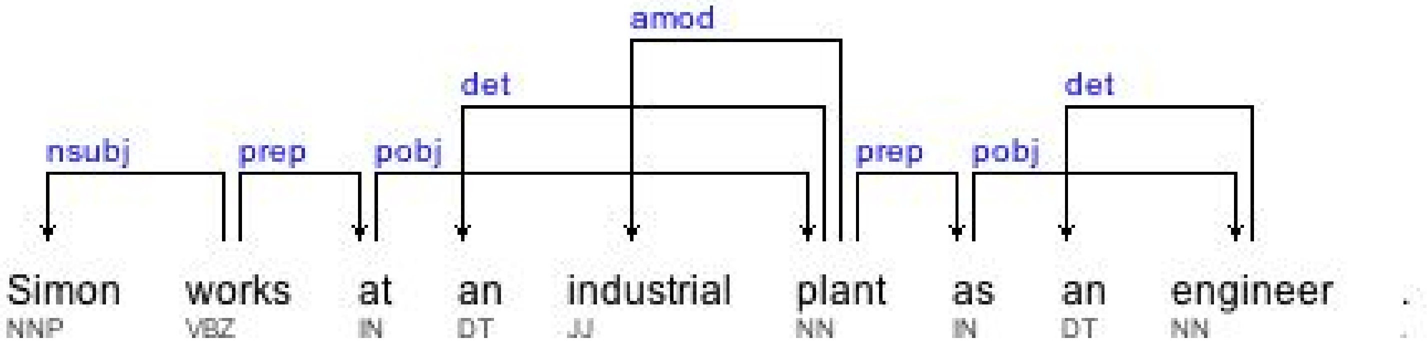
- word
- lemma
- part of speech
- dependencies

Collocations

Simon works at an industrial ***plant.n.1*** *as an engineer.*

Ngrams: [industrial plant, plant as, an industrial plant,...]

Syngrams: [works:prep_at:plant, work:prep:as, plant:amod:industrial,...]



Parse tree by nlp.stanford.edu:8080/parser/

Bag of words

*Simon works at an industrial **plant** as an engineer.*

plant: [soil, assembly, root, industrial, contraband, agent, work…]
[0, 0, 0, 1, 0, 0, 1…]

Idea

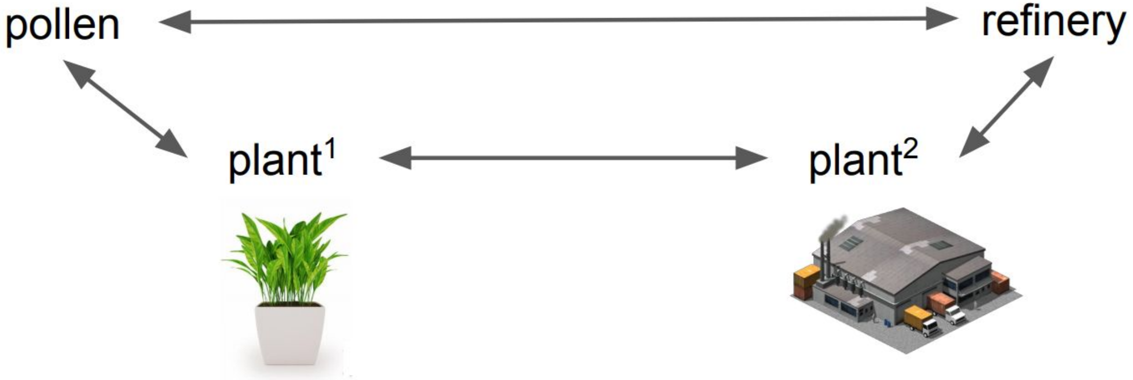
- use a predefined set of context words for each word
- useful for homonyms, to detect the general topic

Results

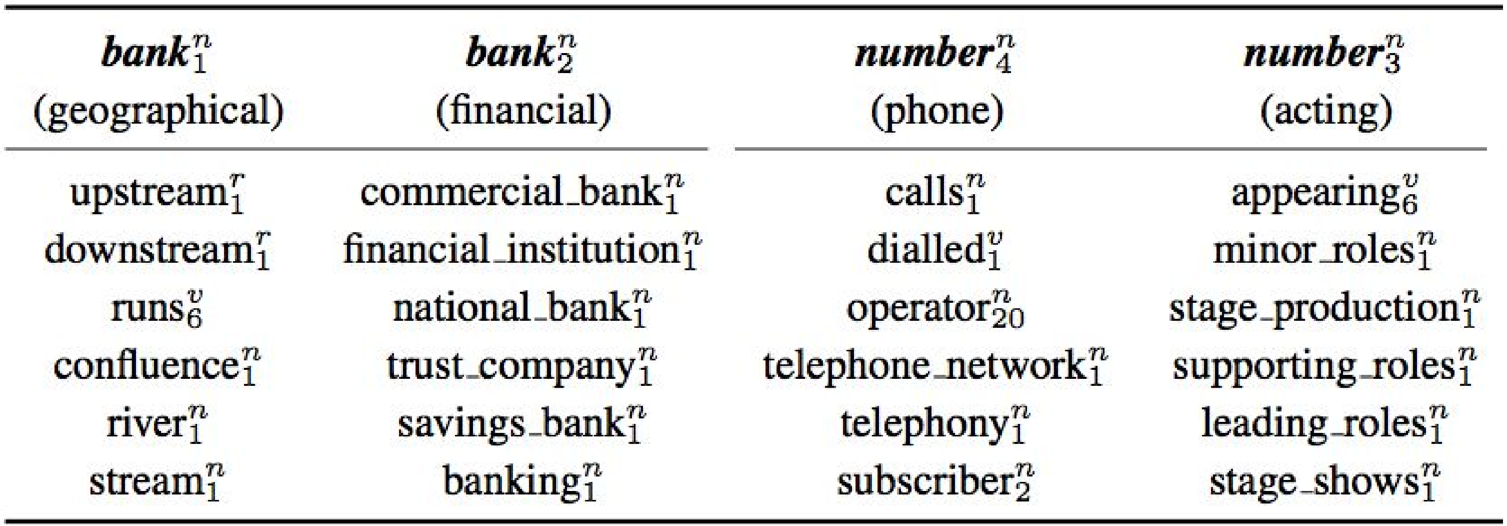
1. Annotate corpora

I need to buy a big ***plant.n.1*** *for my mom. She likes gardening!*
Simon works at an industrial ***plant.n.2*** *as an engineer.*

2. Build sense embeddings



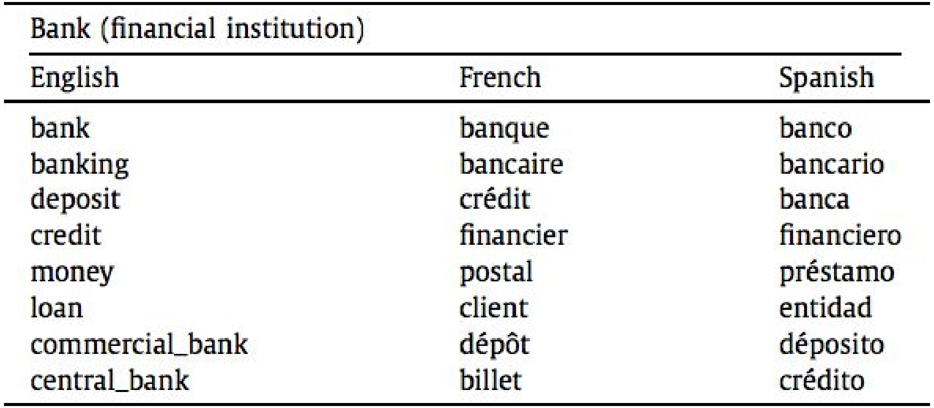
SensEmbed vectors



bank$_1^n$ (geographical)	***bank***$_2^n$ (financial)	***number***$_4^n$ (phone)	***number***$_3^n$ (acting)
upstream$_1^r$	commercial_bank$_1^n$	calls$_1^n$	appearing$_6^v$
downstream$_1^r$	financial_institution$_1^n$	dialled$_1^v$	minor_roles$_1^n$
runs$_6^v$	national_bank$_1^n$	operator$_{20}^n$	stage_production$_1^n$
confluence$_1^n$	trust_company$_1^n$	telephone_network$_1^n$	supporting_roles$_1^n$
river$_1^n$	savings_bank$_1^n$	telephony$_1^n$	leading_roles$_1^n$
stream$_1^n$	banking$_1^n$	subscriber$_2^n$	stage_shows$_1^n$

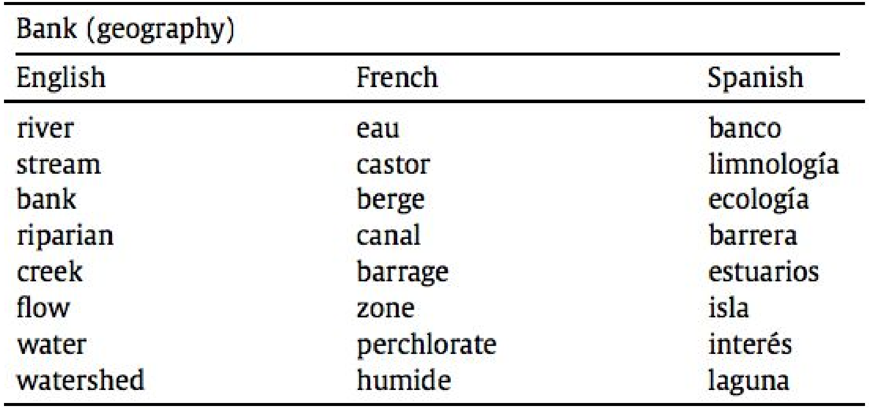
Example from Iacobacci et al. (2015)

Nasari vectors



Bank (financial institution)

English	French	Spanish
bank	banque	banco
banking	bancaire	bancario
deposit	crédit	banca
credit	financier	financiero
money	postal	préstamo
loan	client	entidad
commercial_bank	dépôt	déposito
central_bank	billet	crédito



Bank (geography)

English	French	Spanish
river	eau	banco
stream	castor	limnología
bank	berge	ecología
riparian	canal	barrera
creek	barrage	estuarios
flow	zone	isla
water	perchlorate	interés
watershed	humide	laguna

Example from Camacho-Collados (2016)

A couple of questions…

1. Where do I get annotated data…
2. Where do I get these bags of words…

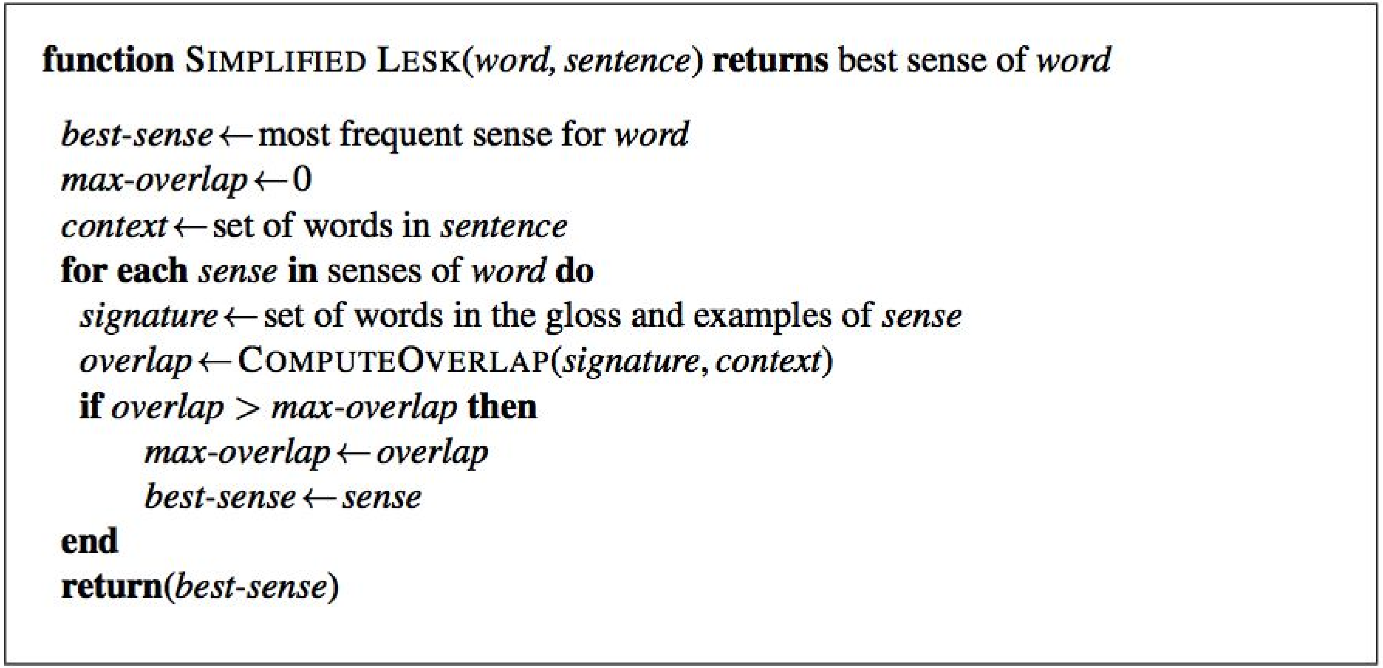
…for each word and each sense that I need in my task?

4.

Linguistically-motivated word-sense disambiguation

Lesk

With which sense **signature** does your **context** overlap the most?

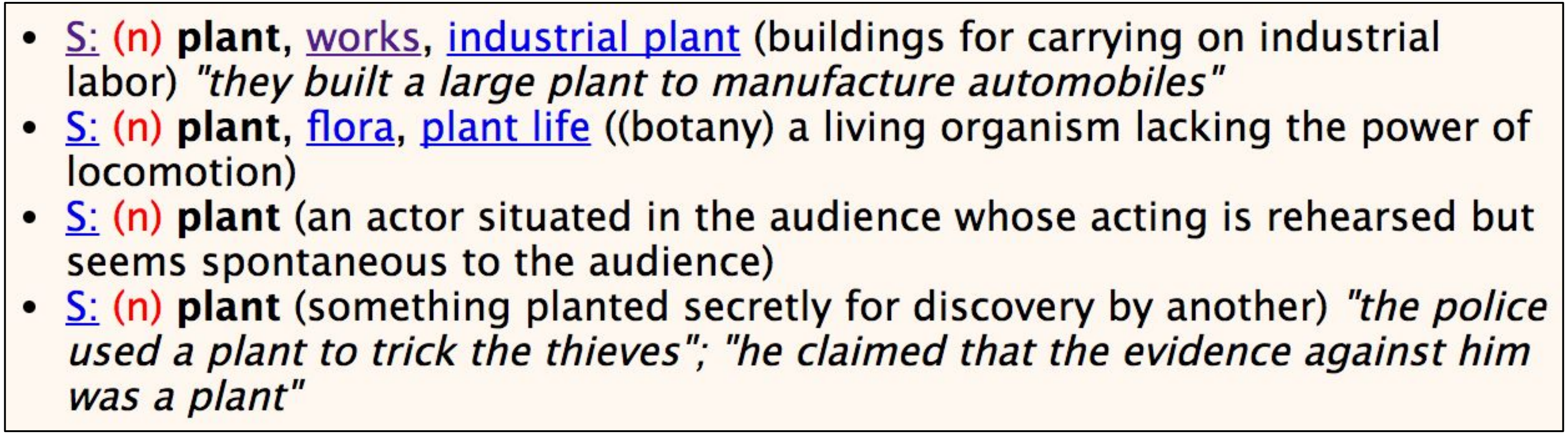


```
function SIMPLIFIED LESK(word, sentence) returns best sense of word

 best-sense ← most frequent sense for word
 max-overlap ← 0
 context ← set of words in sentence
 for each sense in senses of word do
   signature ← set of words in the gloss and examples of sense
   overlap ← COMPUTEOVERLAP(signature, context)
   if overlap > max-overlap then
         max-overlap ← overlap
         best-sense ← sense
 end
 return(best-sense)
```

Lesk

Simon works at an industrial ***plant*** *as an engineer.*



- S: (n) **plant**, works, industrial plant (buildings for carrying on industrial labor) *"they built a large plant to manufacture automobiles"*
- S: (n) **plant**, flora, plant life ((botany) a living organism lacking the power of locomotion)
- S: (n) **plant** (an actor situated in the audience whose acting is rehearsed but seems spontaneous to the audience)
- S: (n) **plant** (something planted secretly for discovery by another) *"the police used a plant to trick the thieves"; "he claimed that the evidence against him was a plant"*

Example from WordNet

Lesk

How to find context words?

- filter functional words
- take lemmas
- for *signature* of each sense, use
  - examples
  - definitions
  - related terms
  - synonyms, hyponyms, hypernyms, holonyms, meronyms…
  - sentences from corpora, etc.

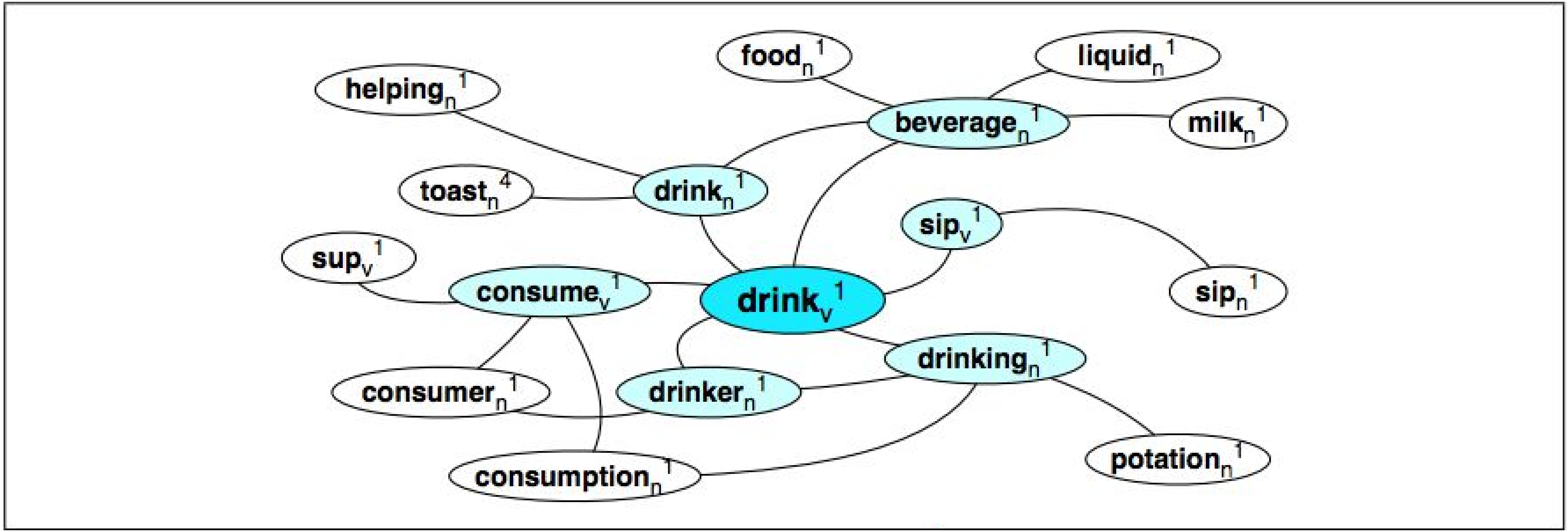
Lesk

How to compute overlap?

- number of overlapping words
- weighed by the number of occurrences
- weighed by –log(P(w))
- weighed by IDF score: log(C(doc) / C(d_i))
- weighed by ontological distance

Graph-Based

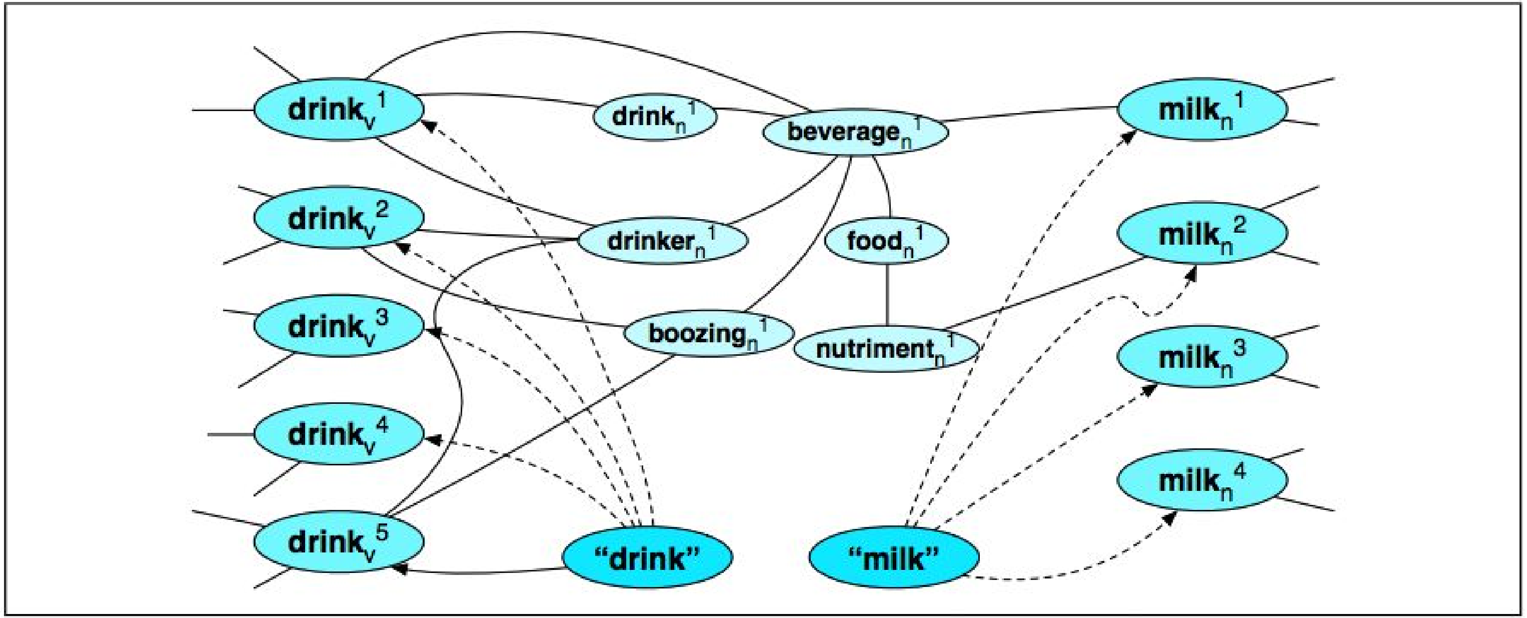
Which sense is the closest to context words?



Example from Navigli and Lapata (2010)

Graph-Based

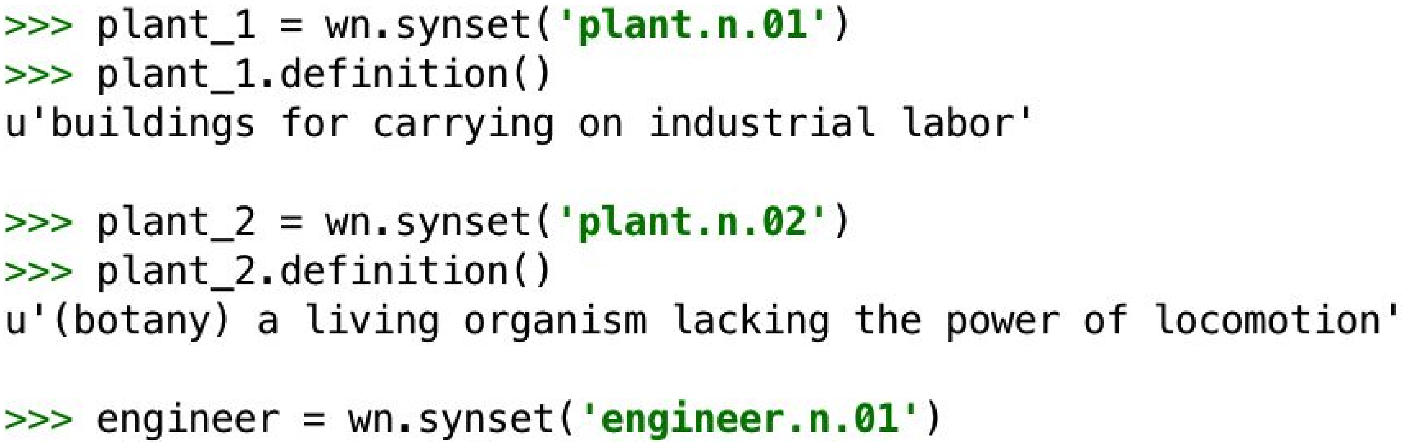
Which sense of the context word to choose?



Example from Navigli and Lapata (2010)

Graph-Based

Simon works at an industrial ***plant*** *as an <u>engineer</u>.*



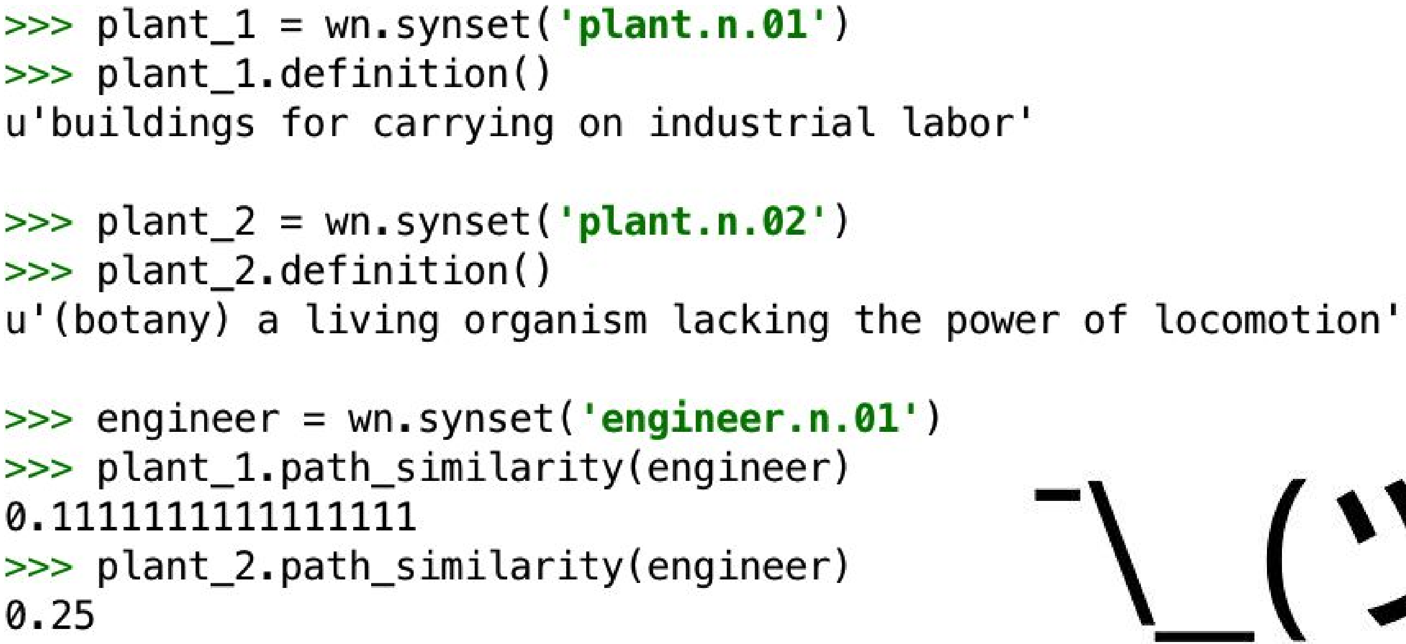
```
>>> plant_1 = wn.synset('plant.n.01')
>>> plant_1.definition()
u'buildings for carrying on industrial labor'

>>> plant_2 = wn.synset('plant.n.02')
>>> plant_2.definition()
u'(botany) a living organism lacking the power of locomotion'

>>> engineer = wn.synset('engineer.n.01')
```

Graph-Based

*Simon works at an industrial **plant** as an engineer.*



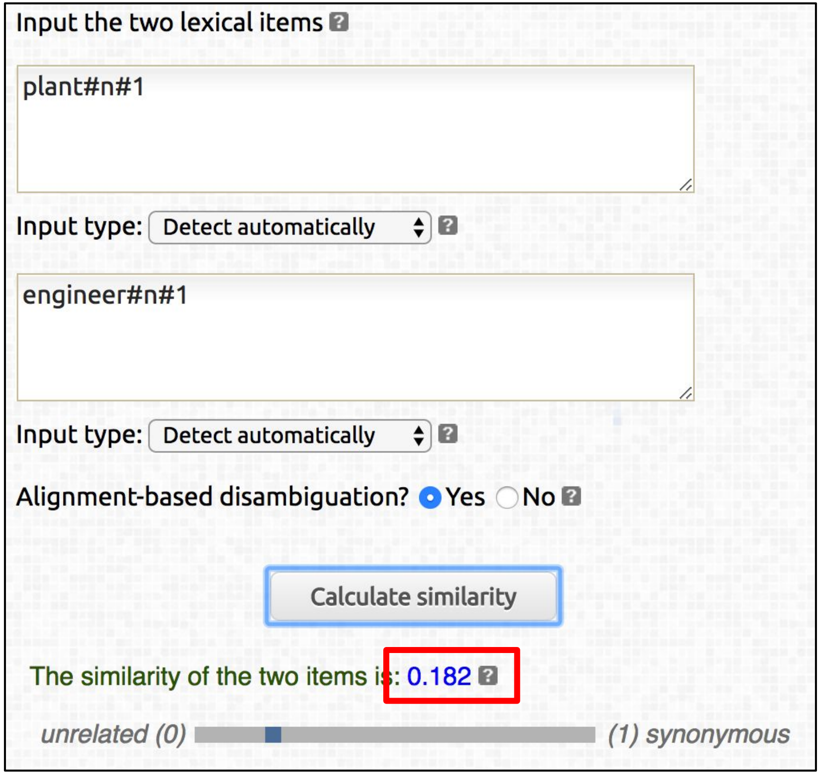
```
>>> plant_1 = wn.synset('plant.n.01')
>>> plant_1.definition()
u'buildings for carrying on industrial labor'

>>> plant_2 = wn.synset('plant.n.02')
>>> plant_2.definition()
u'(botany) a living organism lacking the power of locomotion'

>>> engineer = wn.synset('engineer.n.01')
>>> plant_1.path_similarity(engineer)
0.1111111111111111
>>> plant_2.path_similarity(engineer)
0.25
```

¯_(ツ)_/¯

Graph-Based



Input the two lexical items

plant#n#1

Input type: Detect automatically

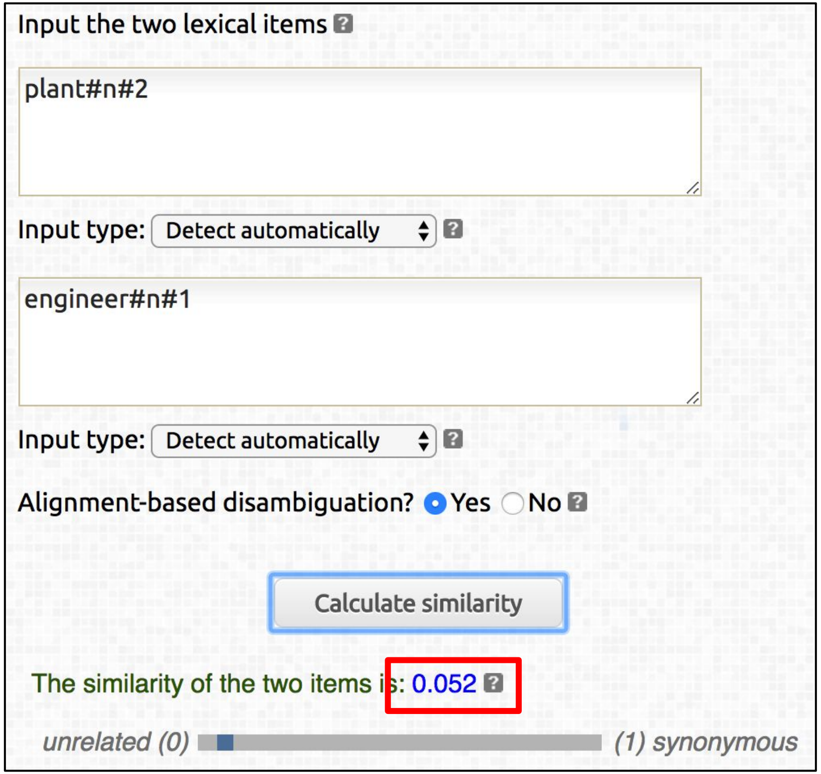
engineer#n#1

Input type: Detect automatically

Alignment-based disambiguation? ● Yes ○ No

The similarity of the two items is: 0.182

unrelated (0) (1) synonymous



Input the two lexical items

plant#n#2

Input type: Detect automatically

engineer#n#1

Input type: Detect automatically

Alignment-based disambiguation? ● Yes ○ No

The similarity of the two items is: 0.052

unrelated (0) (1) synonymous

Demo: http://lcl.uniroma1.it/adw/

Impact

Pros:

- good for partially annotating corpora
  - can be continued in a semi-supervised fashion
- good for bag-of-words feature set
- unreasonably effective: ~0.7% prec and ~0.7% recall

Cons:

- some senses are poorly covered
- mapping e.g. WordNet and Wikipedia is a tricky task

Important linguistic hypothesis

One sense per discourse!

*I bought a **plant** yesterday and put it in my small tank with some inch long baby cichlids.Lost 3 fish over night i never lose fish. i dont see any nibbles on the **plant** though.. any advice?*

5.

Unsupervised word-sense disambiguation

Word sense induction

Idea:

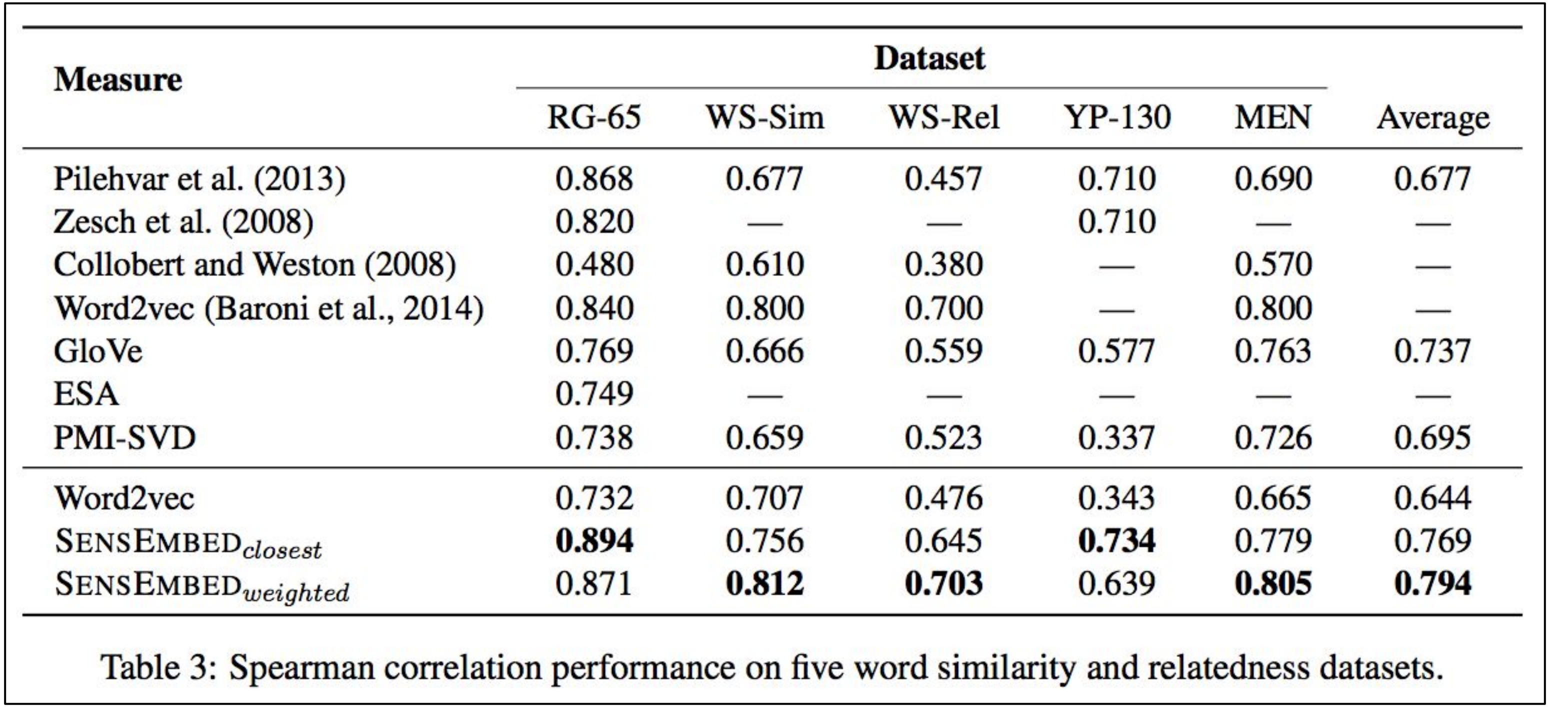
- for each word occurrence, compute a context vector
- cluster these context vectors
- compute the sense vector in each cluster
- map sense vectors to senses

The number of clusters should be predefined. Or not.

6.

To conclude

Quality

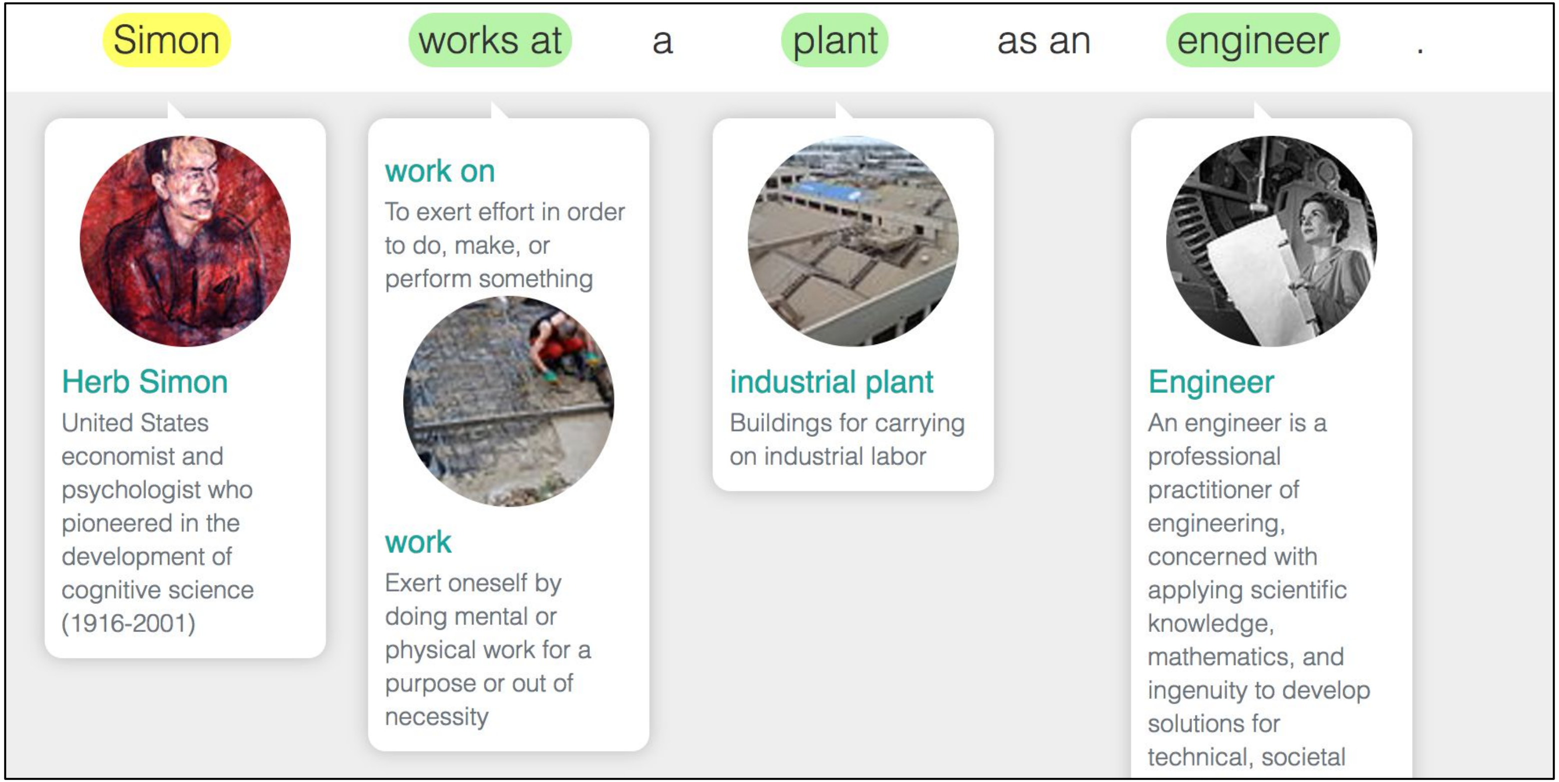


Measure	Dataset					
	RG-65	WS-Sim	WS-Rel	YP-130	MEN	Average
Pilehvar et al. (2013)	0.868	0.677	0.457	0.710	0.690	0.677
Zesch et al. (2008)	0.820	—	—	0.710	—	—
Collobert and Weston (2008)	0.480	0.610	0.380	—	0.570	—
Word2vec (Baroni et al., 2014)	0.840	0.800	0.700	—	0.800	—
GloVe	0.769	0.666	0.559	0.577	0.763	0.737
ESA	0.749	—	—	—	—	—
PMI-SVD	0.738	0.659	0.523	0.337	0.726	0.695
Word2vec	0.732	0.707	0.476	0.343	0.665	0.644
SENSEMBED$_{closest}$	**0.894**	0.756	0.645	**0.734**	0.779	0.769
SENSEMBED$_{weighted}$	0.871	**0.812**	**0.703**	0.639	**0.805**	**0.794**

Table 3: Spearman correlation performance on five word similarity and relatedness datasets.

Example from Iacobacci et al. (2015)

Babelfy



Simon works at a plant as an engineer .

Herb Simon
United States economist and psychologist who pioneered in the development of cognitive science (1916-2001)

work on
To exert effort in order to do, make, or perform something

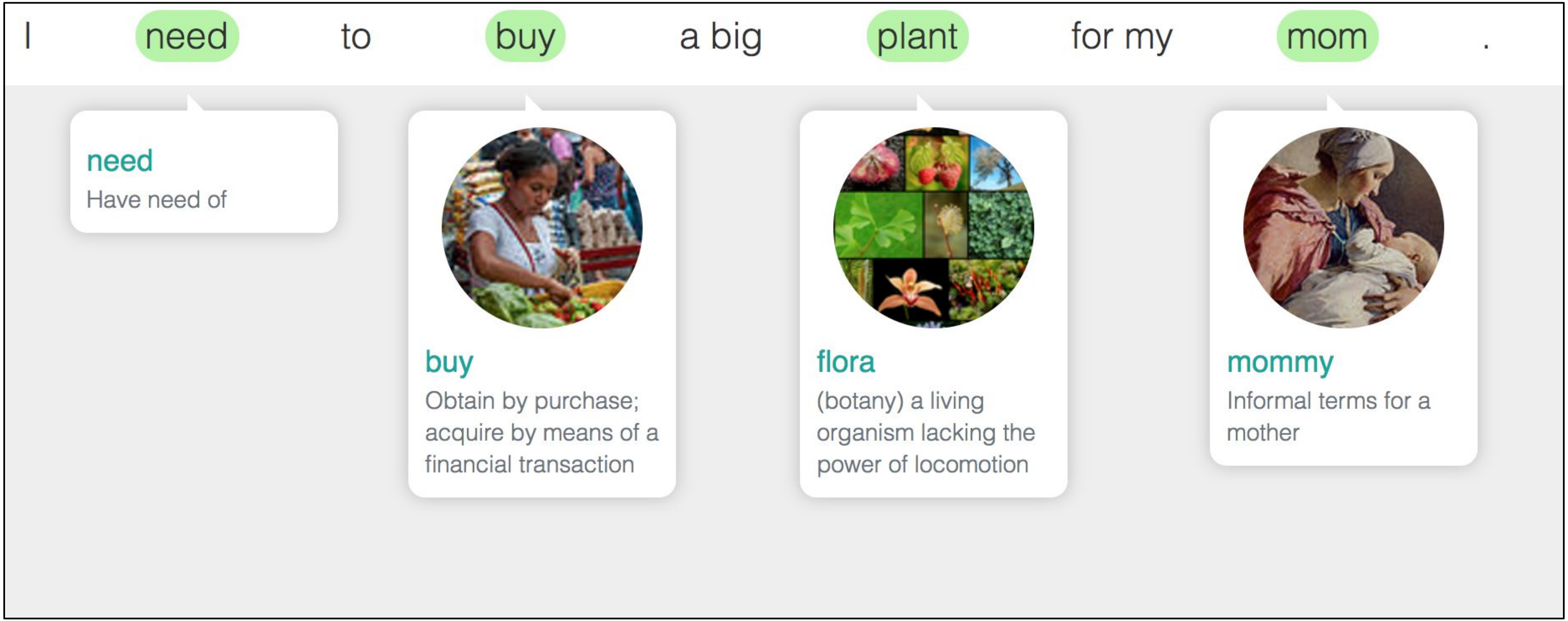
work
Exert oneself by doing mental or physical work for a purpose or out of necessity

industrial plant
Buildings for carrying on industrial labor

Engineer
An engineer is a professional practitioner of engineering, concerned with applying scientific knowledge, mathematics, and ingenuity to develop solutions for technical, societal

Example from babelfy.org

Babelfy



I need to buy a big plant for my mom .

need

Have need of

buy

Obtain by purchase; acquire by means of a financial transaction

flora

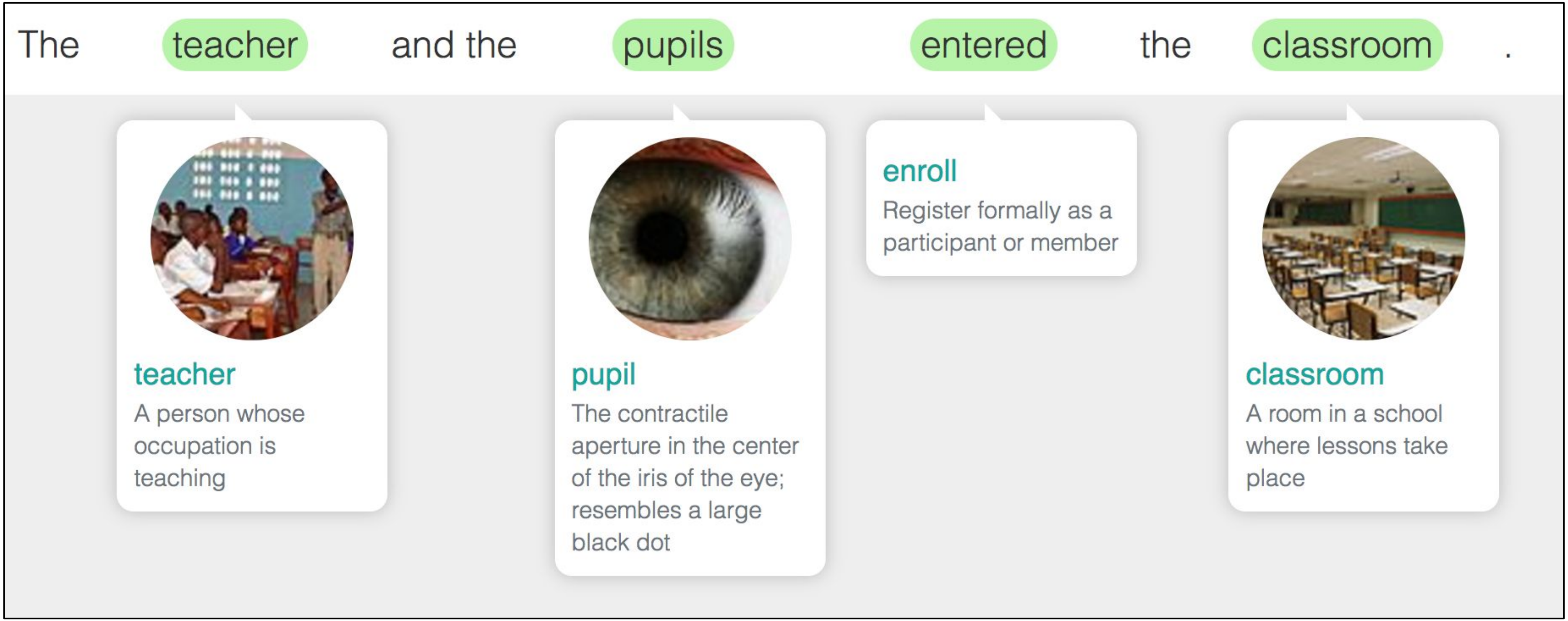
(botany) a living organism lacking the power of locomotion

mommy

Informal terms for a mother

Example from babelfy.org

Babelfy



The teacher and the pupils entered the classroom .

teacher
A person whose occupation is teaching

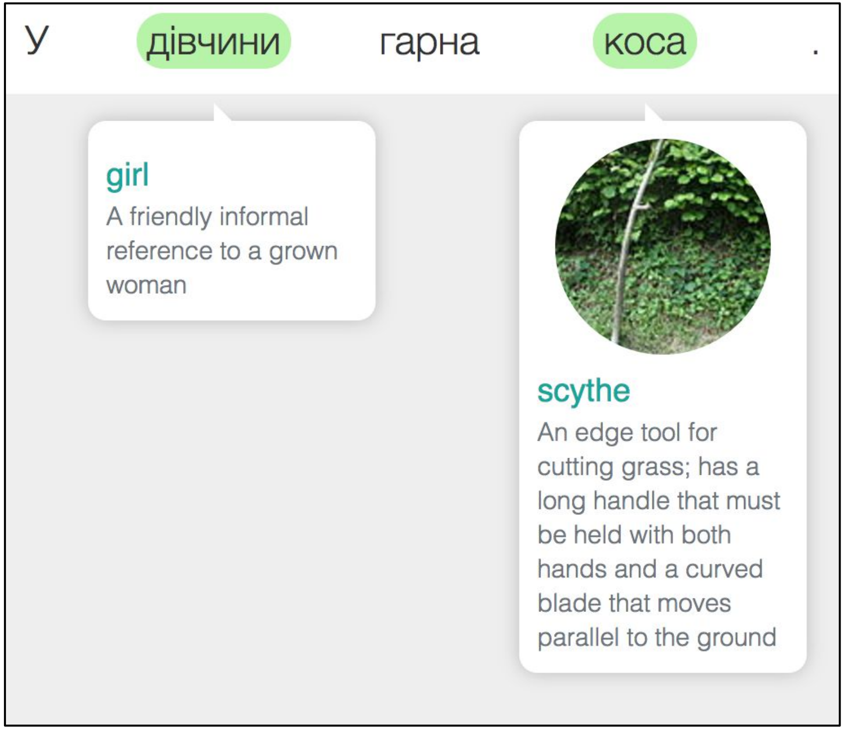
pupil
The contractile aperture in the center of the iris of the eye; resembles a large black dot

enroll
Register formally as a participant or member

classroom
A room in a school where lessons take place

Example from babelfy.org

Babelfy



У дівчини гарна коса .

girl

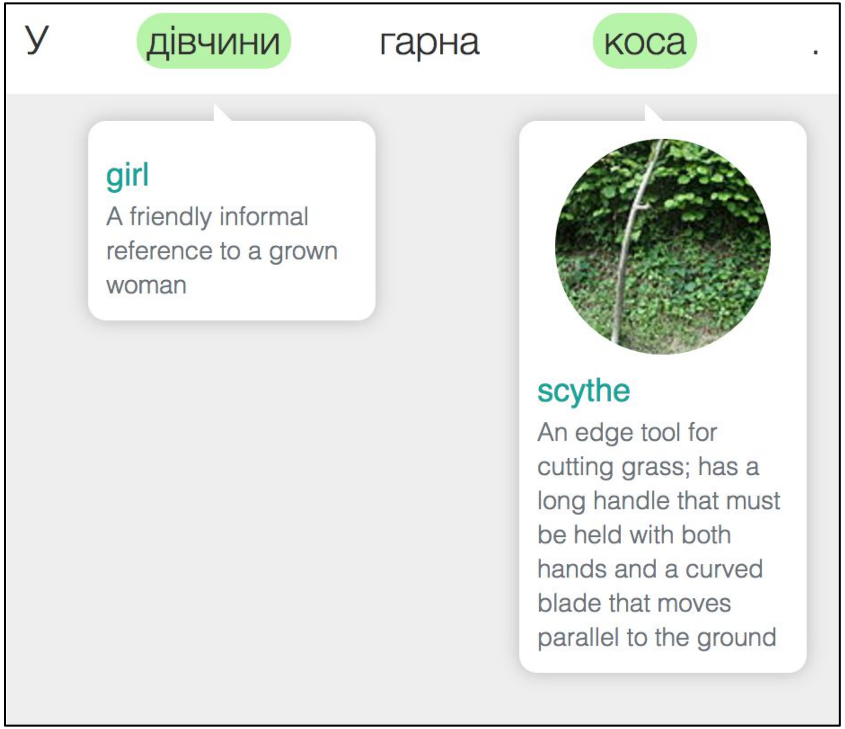
A friendly informal reference to a grown woman

scythe

An edge tool for cutting grass; has a long handle that must be held with both hands and a curved blade that moves parallel to the ground

Example from babelfy.org

Babelfy



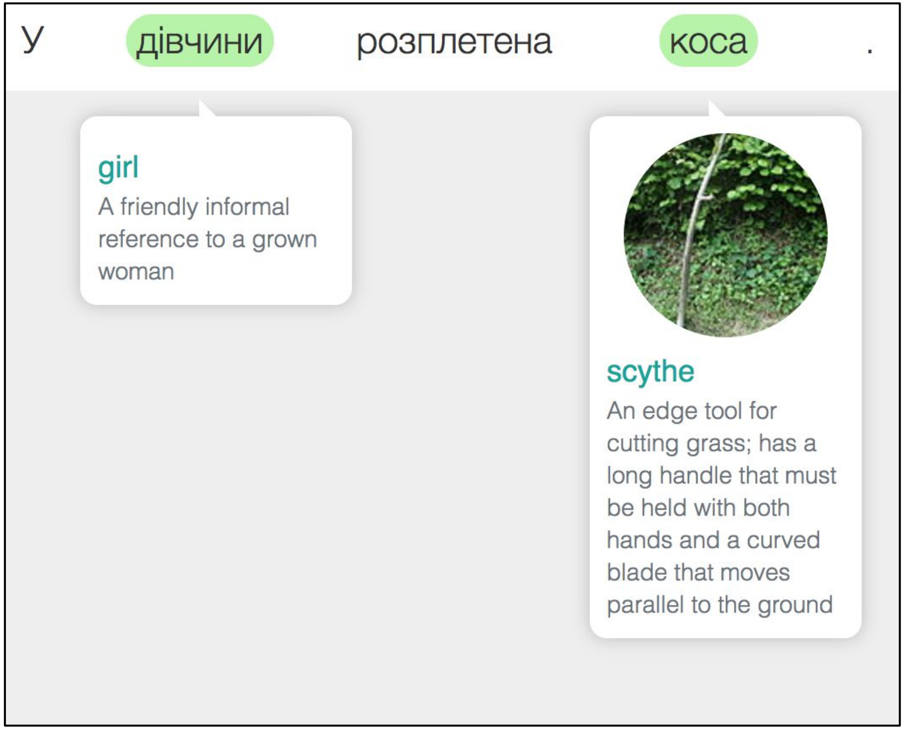
У дівчини гарна коса .

girl

A friendly informal reference to a grown woman

scythe

An edge tool for cutting grass; has a long handle that must be held with both hands and a curved blade that moves parallel to the ground



У дівчини розплетена коса .

girl

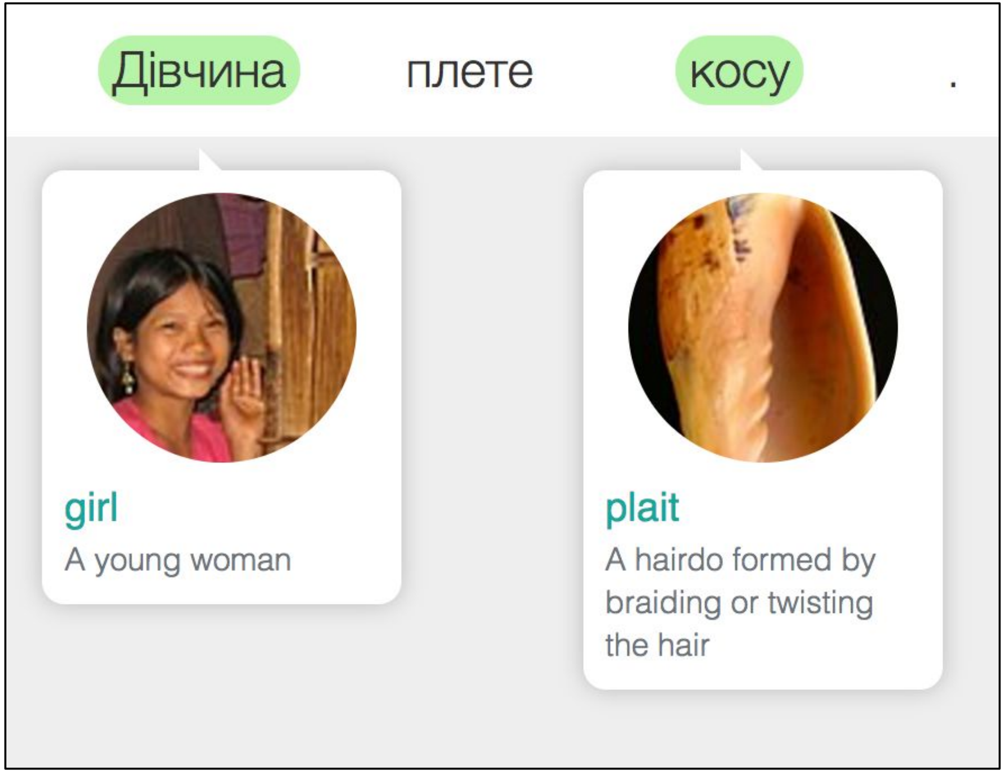
A friendly informal reference to a grown woman

scythe

An edge tool for cutting grass; has a long handle that must be held with both hands and a curved blade that moves parallel to the ground

Example from babelfy.org

Babelfy



Дівчина плете косу .

girl

A young woman

plait

A hairdo formed by braiding or twisting the hair

Example from babelfy.org

Thank.v.01 you!

Any questions.n.01?

References

- Neelakantan et al. (2014), Efficient Non-parametric Estimation of Multiple Embeddings per Word in Vector Space
- Iacobacci et al. (2015), SENSEMBED: Learning Sense Embeddings for Word and Relational Similarity
- Camacho-Collados et al. (2016), Nasari: Integrating explicit knowledge and corpus statistics for a multilingual representation of concepts and entities
- Navigli and Lapata (2010), An Experimental Study of Graph Connectivity for Unsupervised Word Sense Disambiguation
- Athiwaratkun and Wilson (2017), Multimodal Word Distributions
- Abigail See (2017), Four deep learning trends from ACL 2017