Semantic Analysis

Mariana Romanyshyn
Grammarly, Inc.

Contents

1.

Word sense disambiguation

Words have meanings

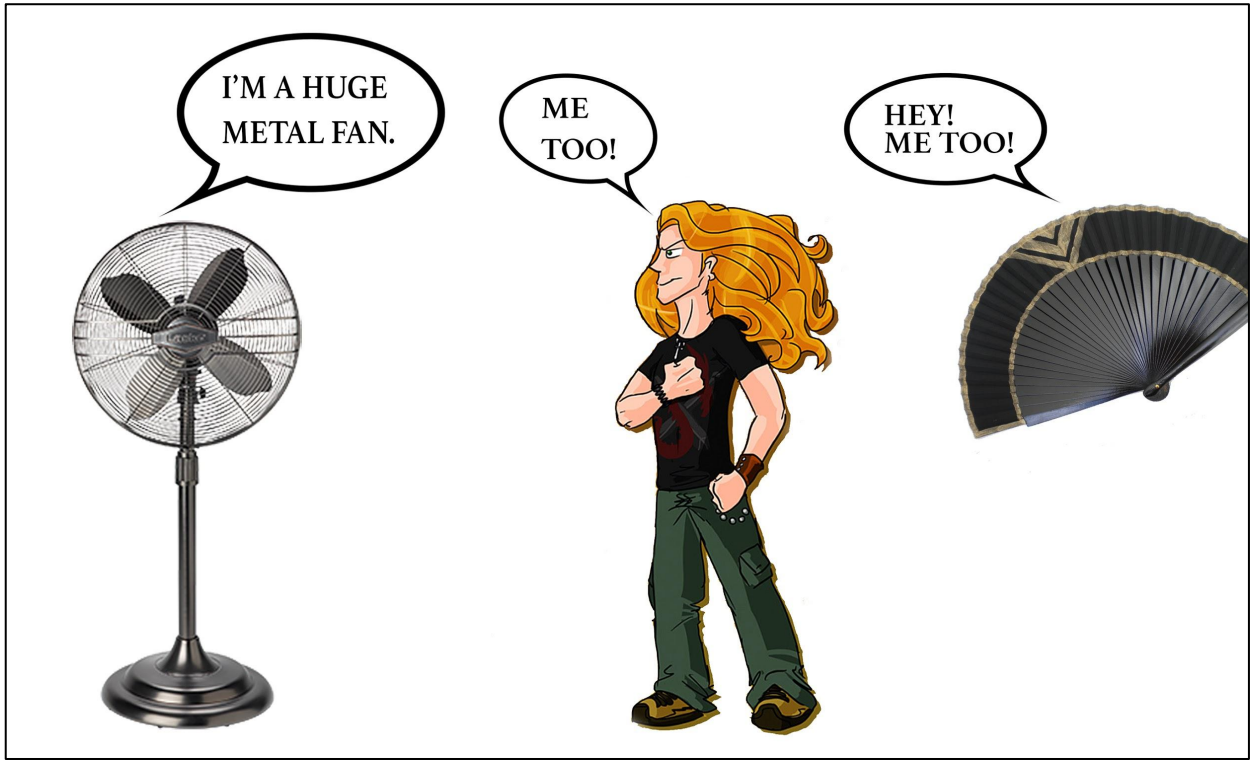


Image by Tetiana Turchyn

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.*

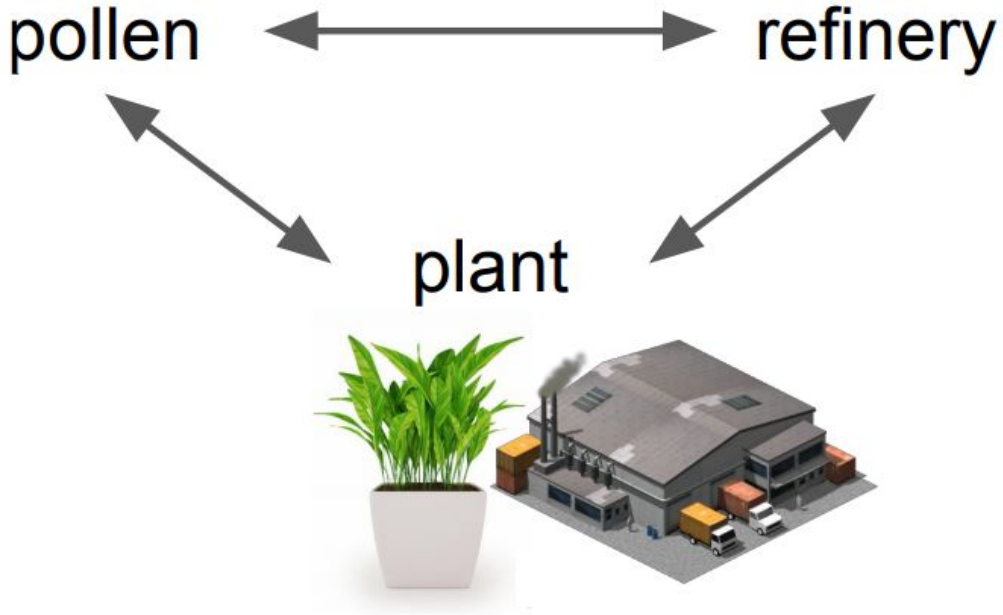
Is it just English?

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

1. стіна будинку
2. стіна урвища
3. мур
4. те, що відокремлює, роз'єднує

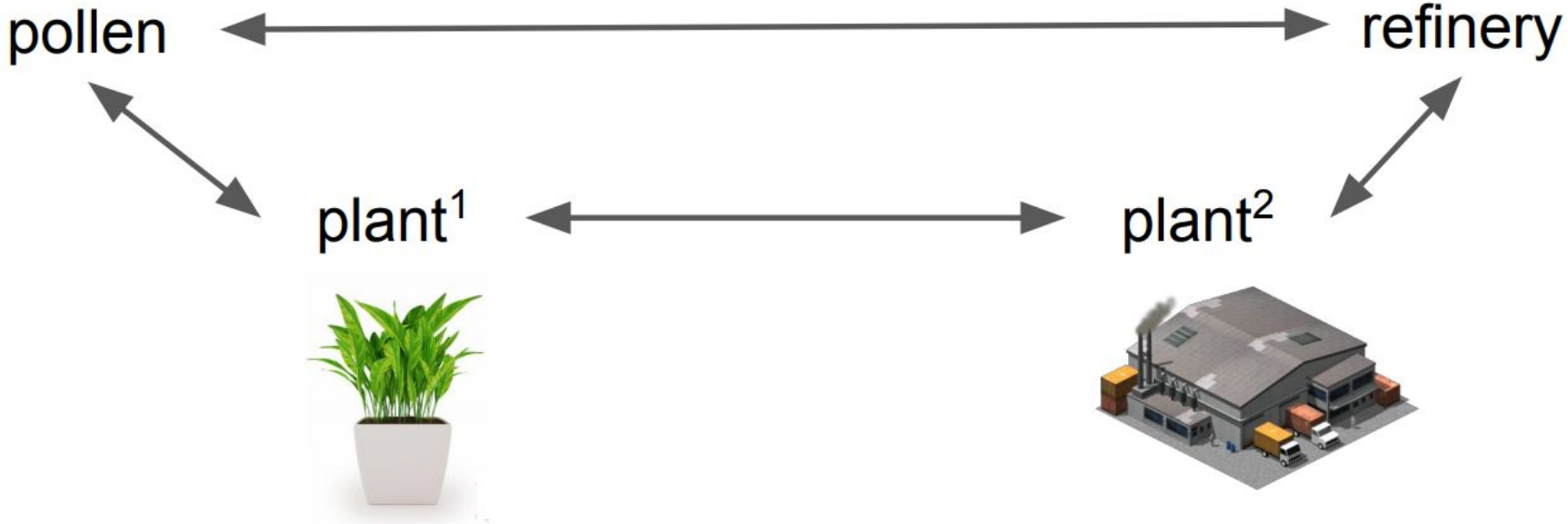
What does it mean for NLP?

Triangle inequality in word embeddings.



What does it mean for NLP?

Word embeddings => sense embeddings

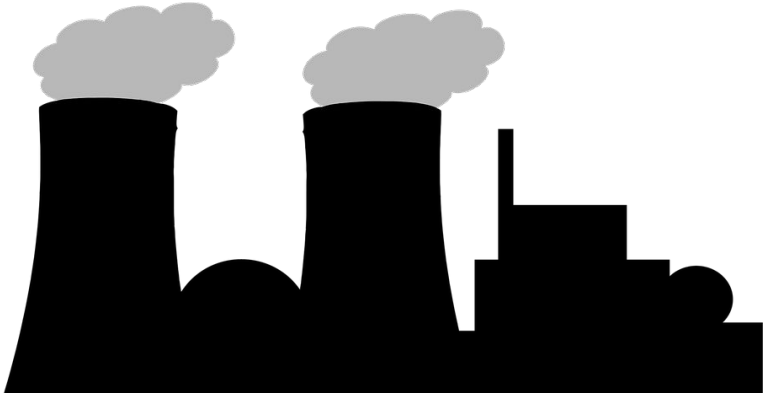


Personal assistants

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

Siri: Hmm...

?



Information retrieval

Where is **Paris** now and what is she doing?

?

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

Animate or inanimate?

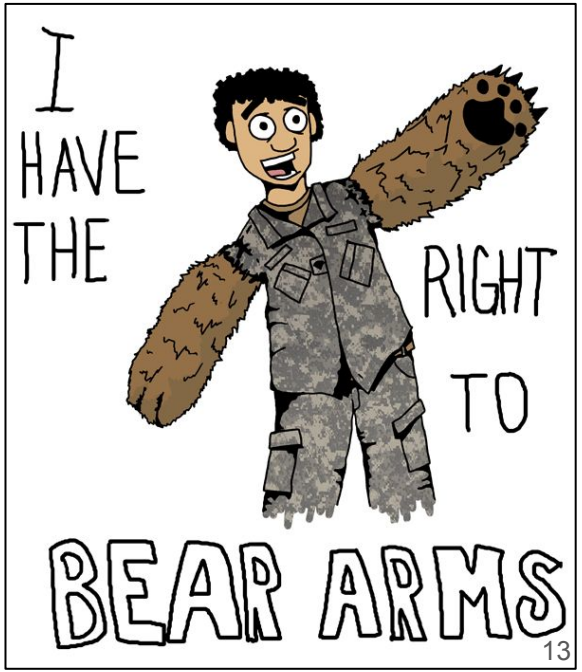
Then we have 2 laptop stands & 2 **mouses** left.

My cat catches **{mouses=>mice}** all the time.

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.



What is “sense”?

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

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**

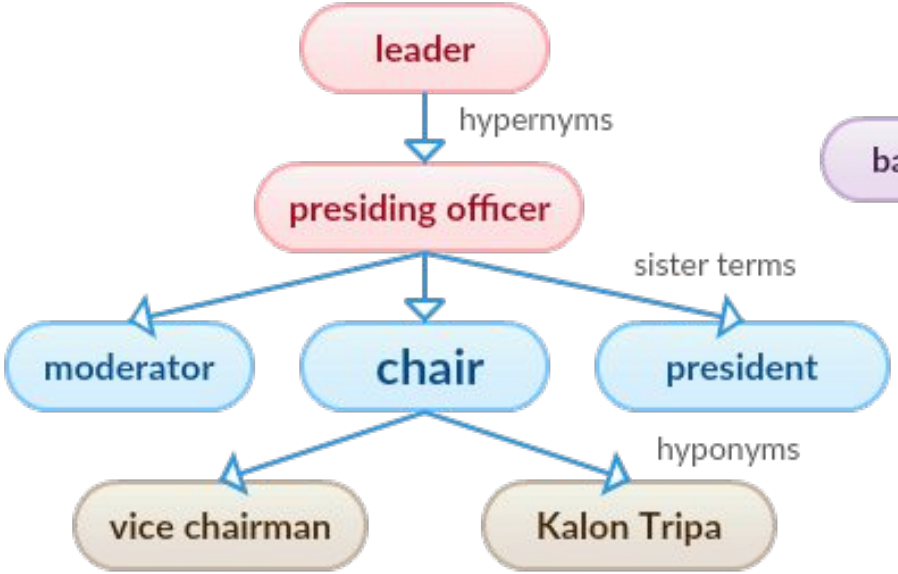
2 STRONG/BRAVE [countable usually singular] a man who has the qualities that people think a man should have, such as being brave, strong etc

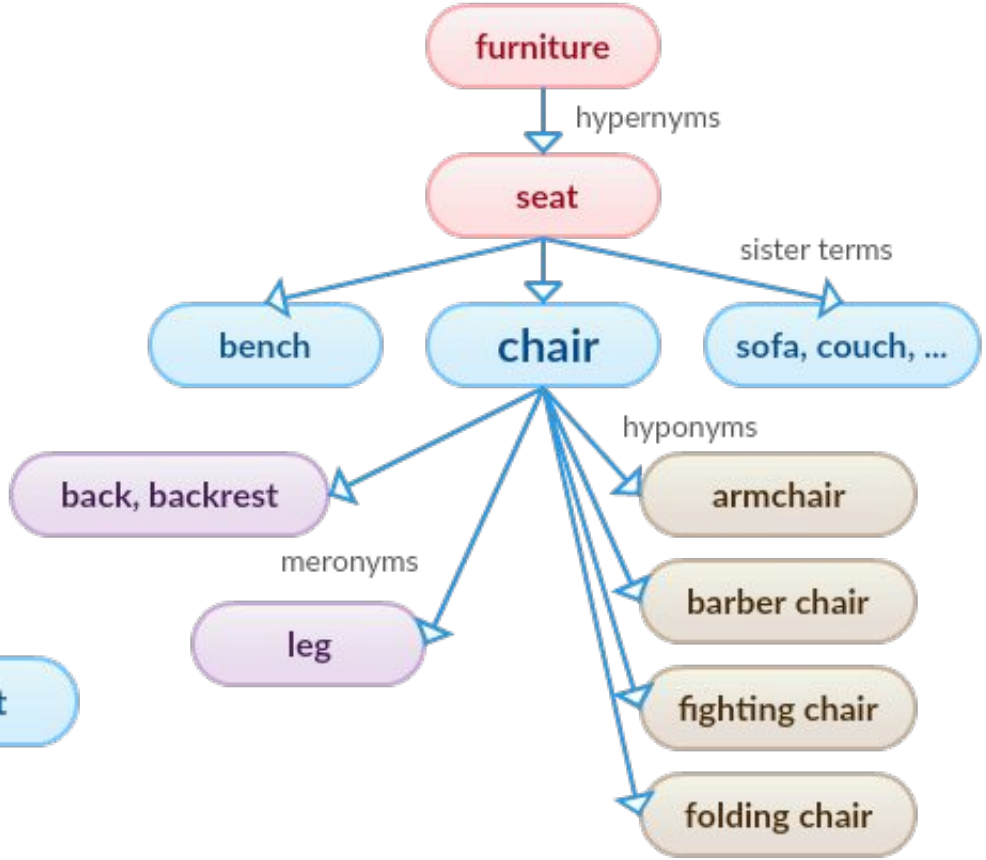
3 PERSON [countable] a person, either male or female – used especially in formal situations or in the past

4 PEOPLE [uncountable] people as a group

Example from www.ldoceonline.com

Ontologies





Example of relations in WordNet

Wikipedia, Wikidata, DBpedia

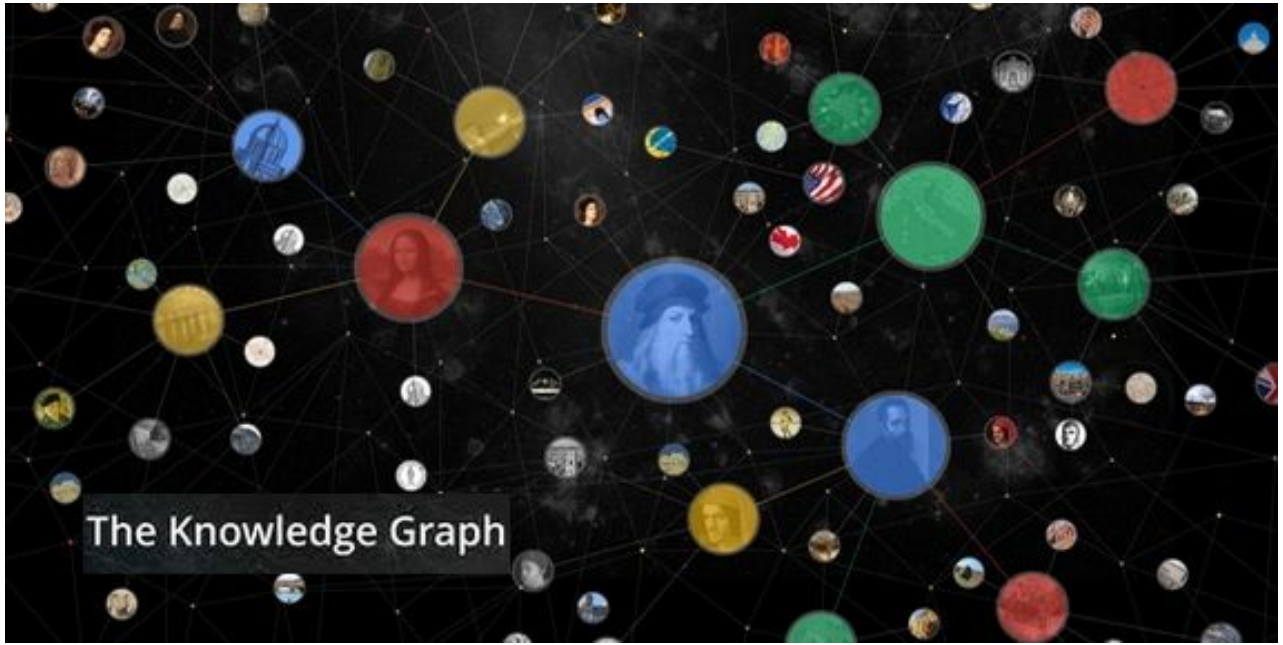
Finance [edit]

- 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

Knowledge Graph

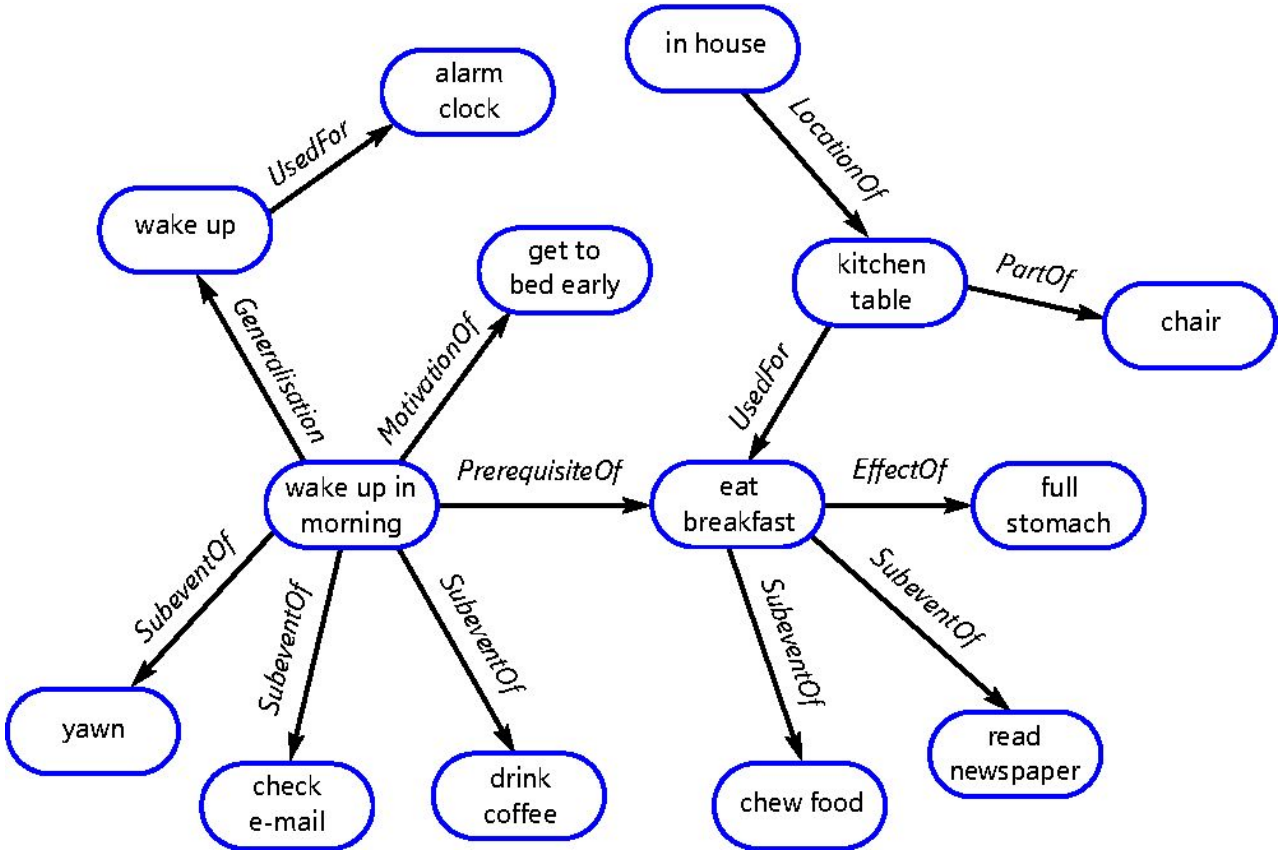


570 million entities

18 billion facts

Used in Google infoboxes, Google Assistant, and Google Home

ConceptNet



8 mln nodes

21 mln edges

83 languages

36 relation types

Represents human knowledge about the world.

Example from conceptnet.io

BabelNet

Noun

fan, mechanical fan, ventilator

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

ID: 00033599n | Concept

fan, rooter, sports fan

An enthusiastic devotee of sports

ID: 00033600n | Concept

15.8 mln nodes

380 mln edges

284 languages

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

How to determine the sense?

You shall know a word by the company it keeps.

John Rupert Firth, 1957

1. 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

- for context, use lemmas and filter stopwords
- 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(docs) / C(docs with word *i*))
- …

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?*

Results

Pros:

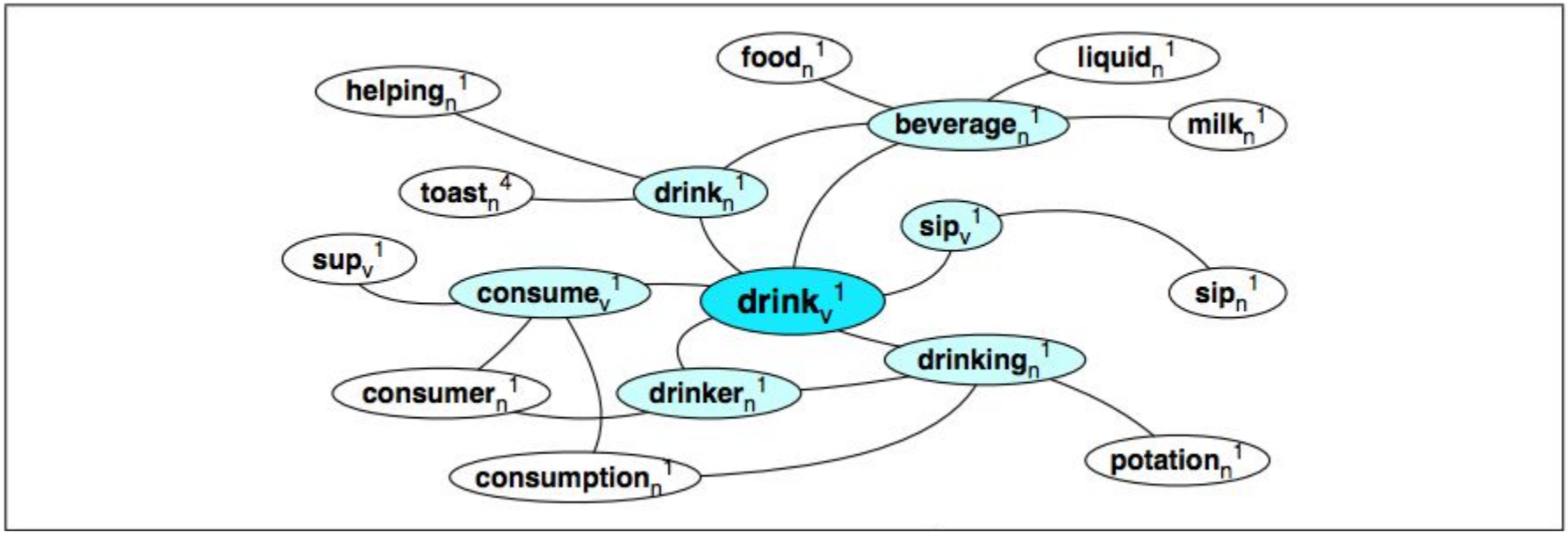
- good for partially annotating corpora
  - can be continued in a semi-supervised fashion
- links to the KB can be preserved
- unreasonably effective

Cons:

- some senses are poorly covered
- mapping (e.g., WordNet to Wikipedia) is a tricky task

2. Graph-Based

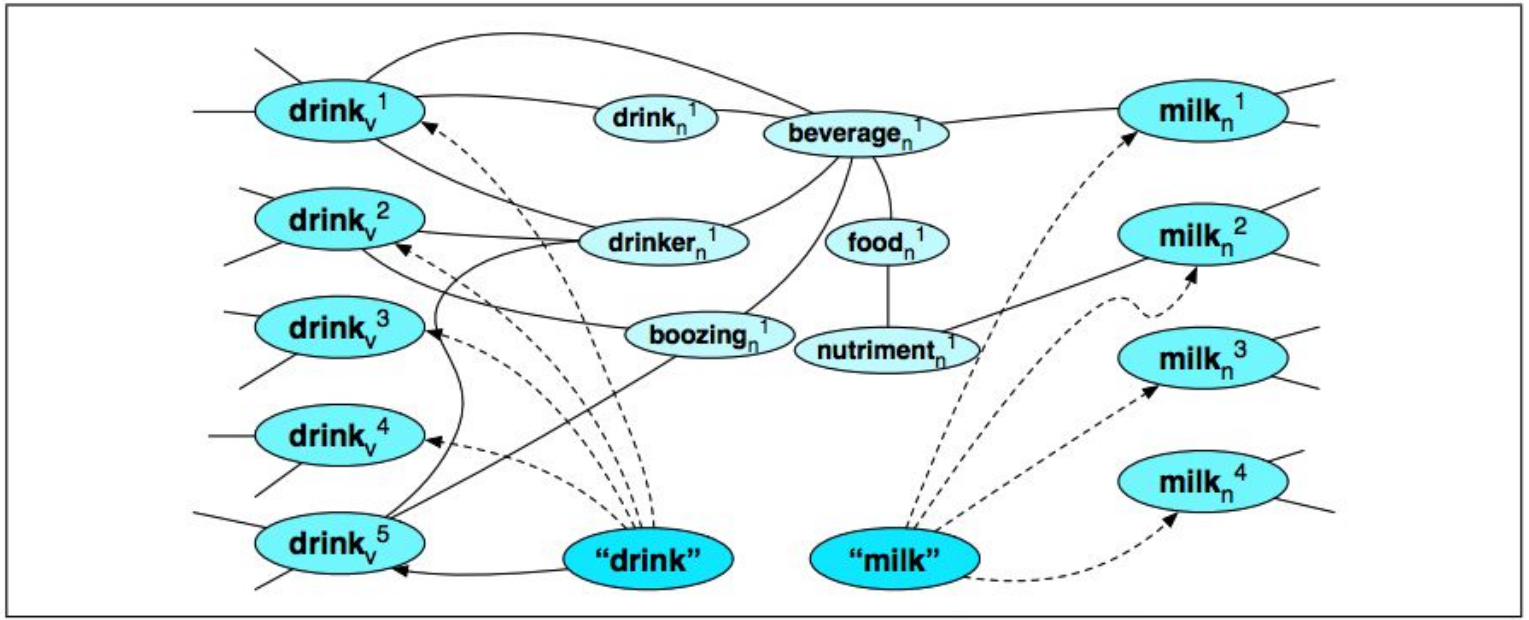
Which sense is the closest to context words?



Example from Navigli and Lapata (2010)

Graph-Based

Which sense of the context word to choose?



Example from Navigli and Lapata (2010)

Path similarity

*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')
```

Path similarity

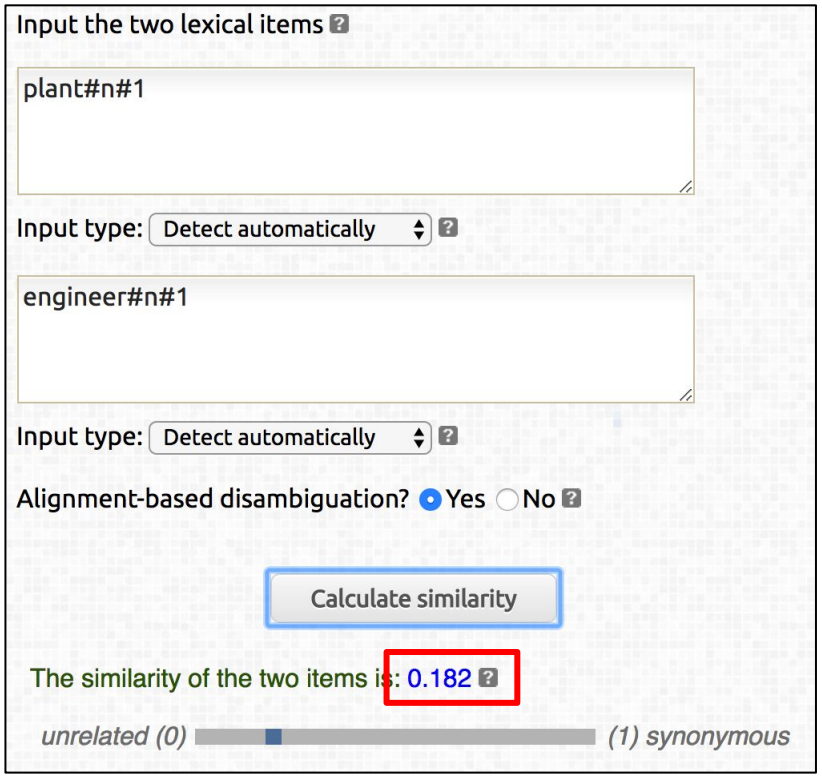
*Simon works at an industrial **plant** as an engineer.*

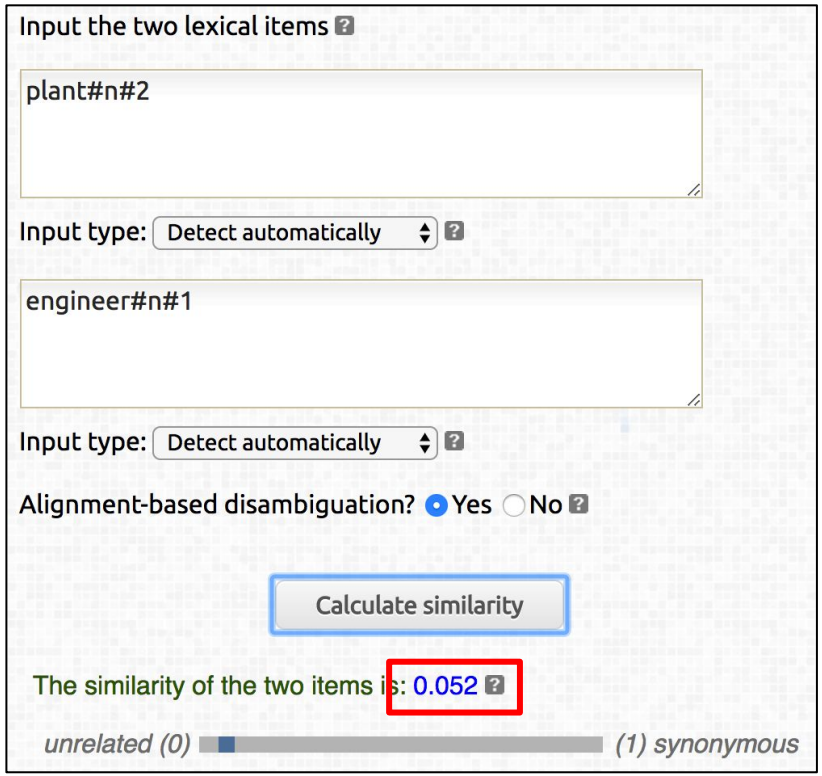
```
>>> plant_1 = wn.synset('plant.n.01')
>>> plant_1.definition()
u'buildings for carrying on industrial labor'

>>> 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
```

Align, Disambiguate, and Walk





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

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 algorithm

- *[done once]* create semantic signatures for BabelNet concepts
  - assign weights to edges; higher weights in more densely connected areas

Babelfy algorithm

- *[done once]* create semantic signatures for BabelNet concepts
  - assign weights to edges; higher weights in more densely connected areas
- extract all linkable fragments from the text
- list possible meanings for the fragments
- create a graph-based semantic interpretation of the whole text by linking the possible meanings of the fragments

Babelfy algorithm

- *[done once]* create semantic signatures for BabelNet concepts
  - assign weights to edges; higher weights in more densely connected areas
- extract all linkable fragments from the text
- list possible meanings for the fragments
- create a graph-based semantic interpretation of the whole text by linking the possible meanings of the fragments
- extract a dense subgraph of this representation
  - connect meanings if they are in each other's signature
- select the best candidate meaning for each fragment

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

The teacher and the pupils entered the classroom .

teacher
A person whose occupation is teaching

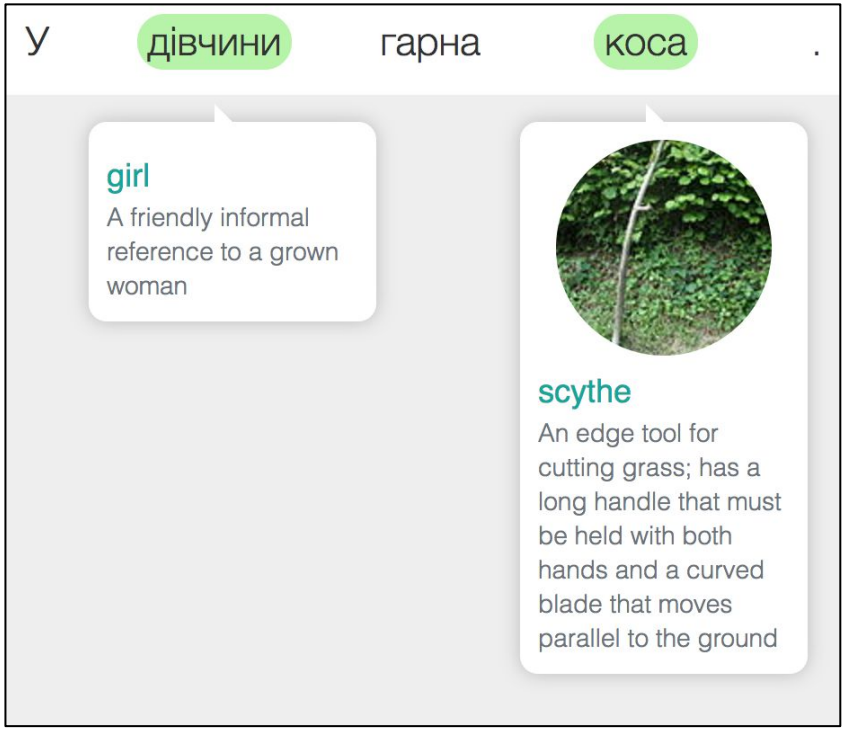
pupil
The contractile aperture in the center of the iris of the eye; resembles a large black dot

enroll
Register formally as a participant or member

classroom
A room in a school where lessons take place

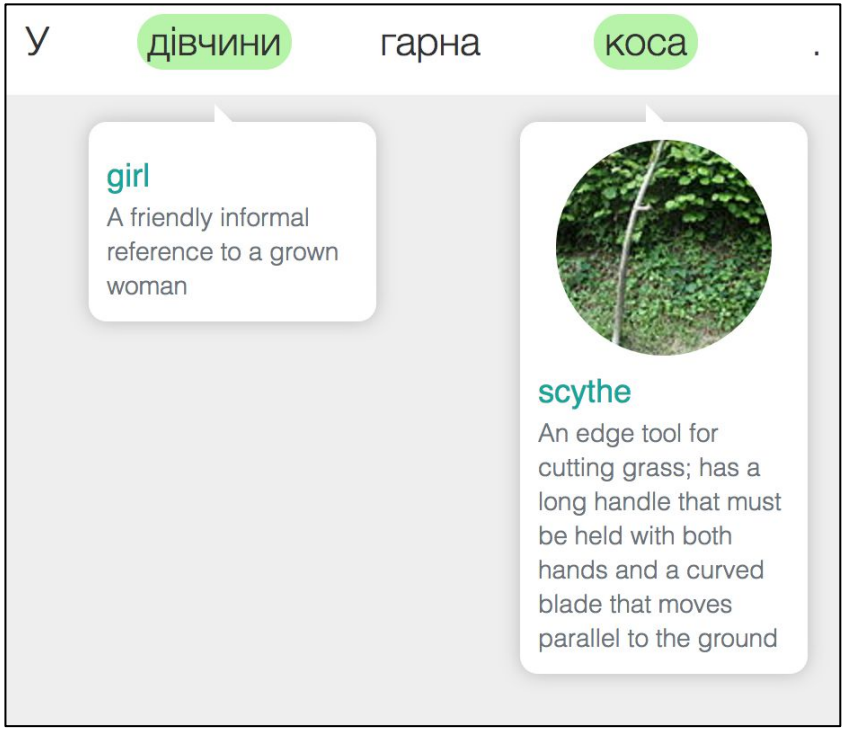
Example from babelfy.org

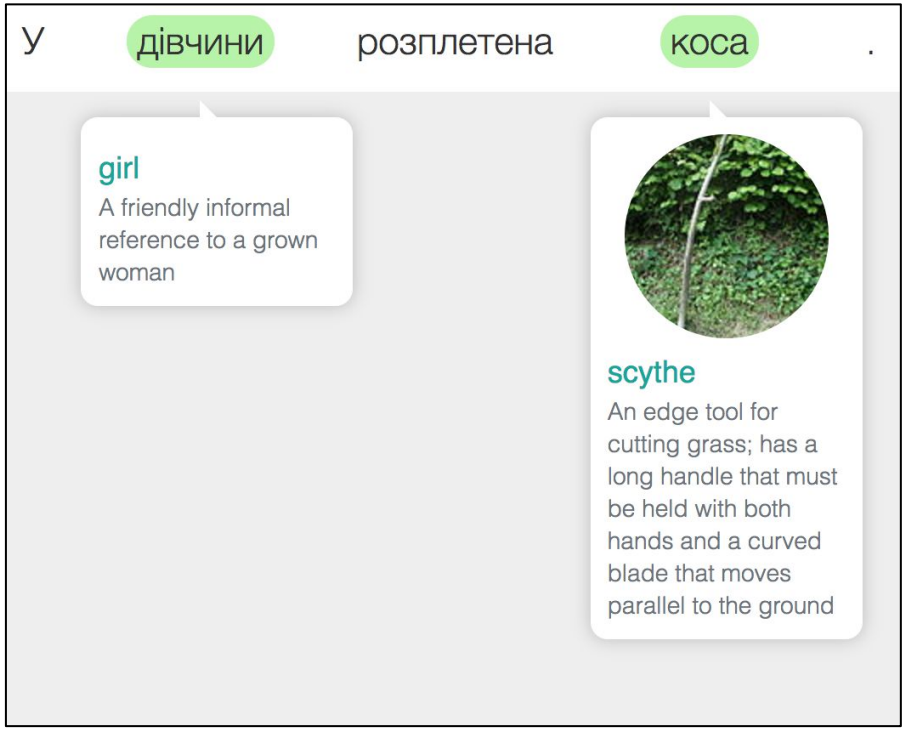
Babelfy



Example from babelfy.org

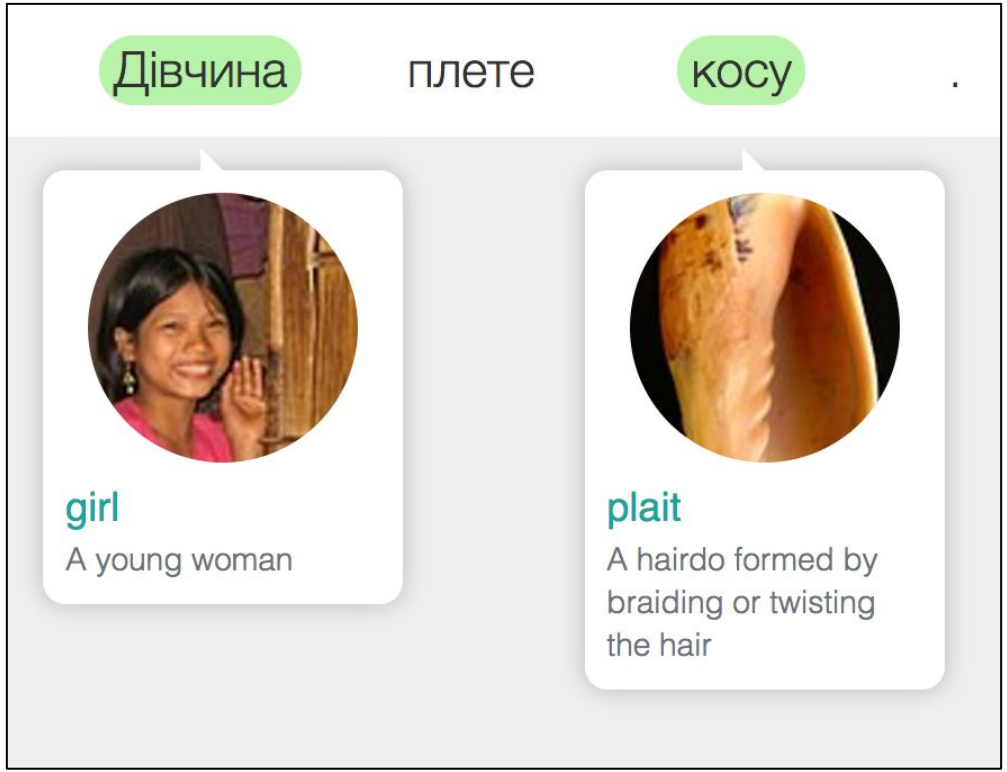
Babelfy





Example from babelfy.org

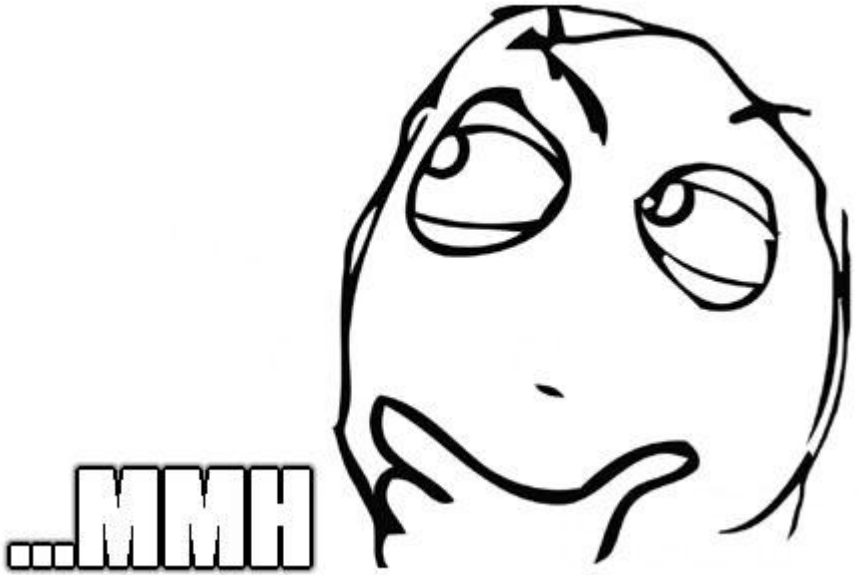
Babelfy



Example from babelfy.org

What else can we do?

- build a classifier
- build sense embeddings

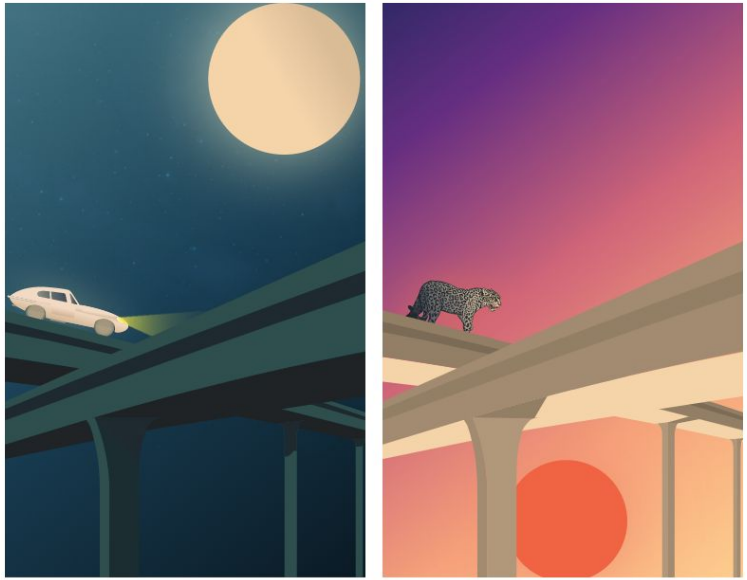


3. Classification

- For a specific word
  - *ngrams, syngrams, morphological features*
  - *overlap of current context with a predefined set of content words (extracted from a dictionary)*

- For a category

OpenAI: Type-based Neural Entity Disambiguation



The man saw a **Jaguar** speed on the highway.

Jaguar Cars 0.70

jaguar 0.12

The prey saw the **jaguar** cross the jungle.

Jaguar Cars 0.03

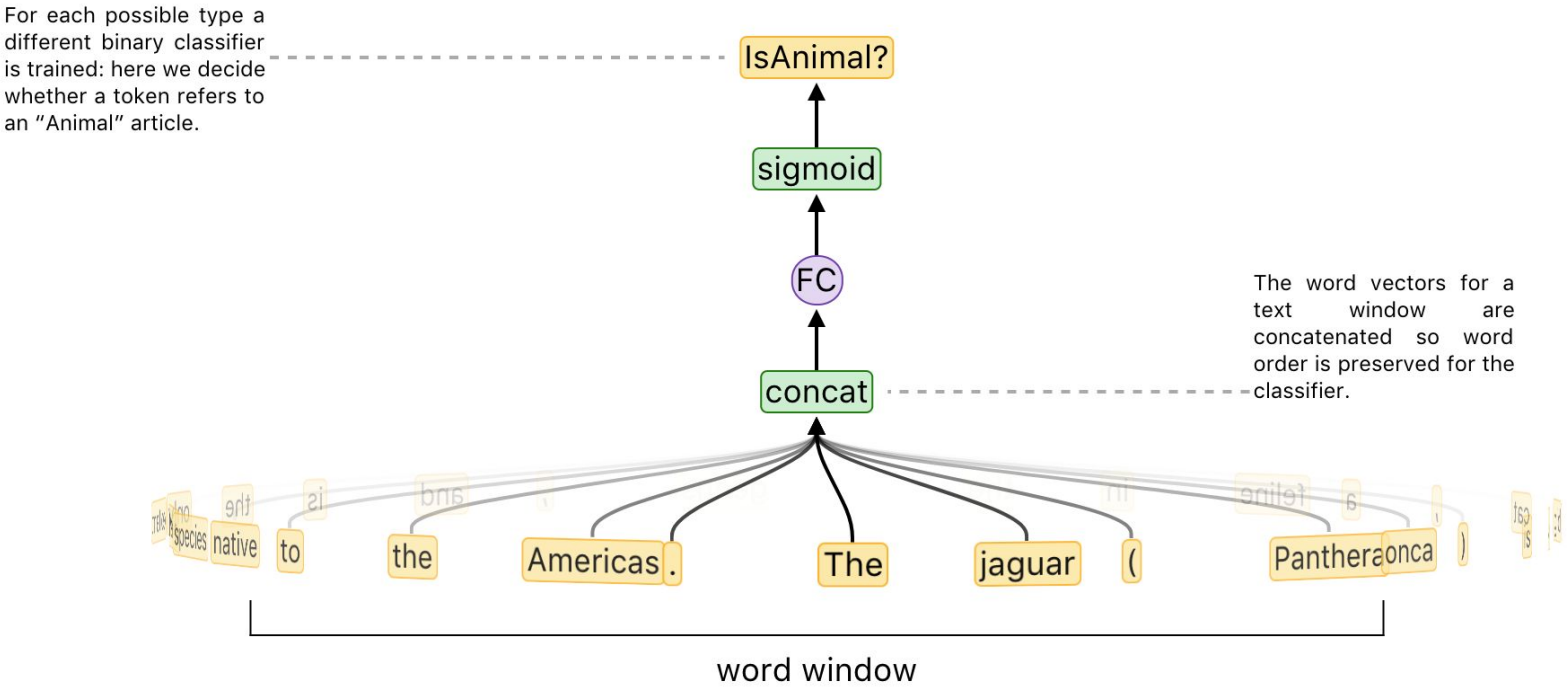
jaguar 0.89

Example from blog.openai.com

OpenAI: Type-based Neural Entity Disambiguation

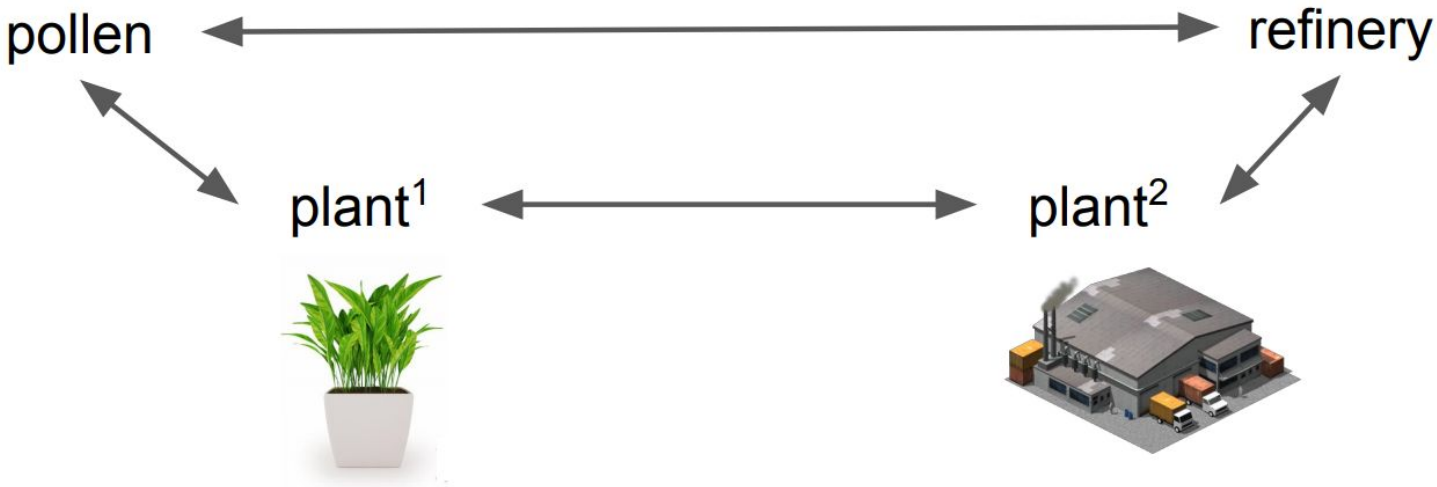
- Extract entities from Wikipedia
- Extract the set of categories for each entity from Wikipedia
- Pick a list of ~100 categories to be your “type” system
- Generate train data: word contexts mapped to 100-dimensional vectors
- Train: 400M tokens, bidirectional LSTM, F1 - 0.91

OpenAI: Type-based Neural Entity Disambiguation



Example from blog.openai.com

4. Sense Embeddings



How to get sense embeddings?

- Train on a sense-annotated corpus
  - *also: add related words from the KB to the context vector*
- Derive from word embeddings using ontologies

SensEmbed vectors

- use *BabelNet* as a sense inventory
- use a dump of the *English Wikipedia* for corpus
- do disambiguation with *Babelfy*
- train *word2vec CBOW* to obtain sense embeddings (window = 5, size = 400)

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)

How to get sense embeddings?

- Train on a sense-annotated corpus
  - *also: add related words from the KB to the context vector*
- Derive from word embeddings using ontologies

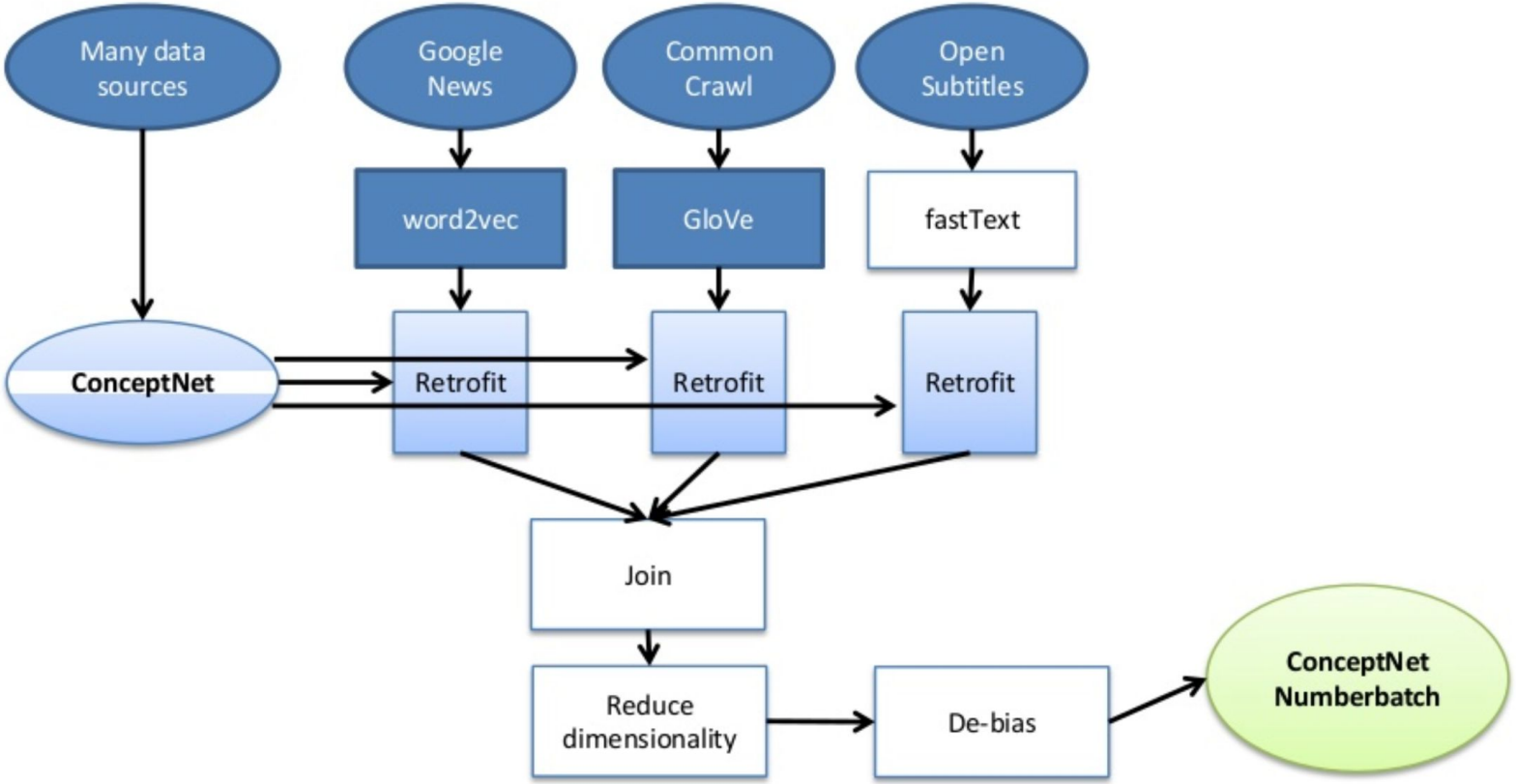
Retrofitting

```
''' Retrofit word vectors to a lexicon '''
def retrofit(wordVecs, lexicon, numIters):
  newWordVecs = deepcopy(wordVecs)
  wvVocab = set(newWordVecs.keys())
  loopVocab = wvVocab.intersection(set(lexicon.keys()))
  for it in range(numIters):
    # loop through every node also in ontology (else just use data estimate)
    for word in loopVocab:
      wordNeighbours = set(lexicon[word]).intersection(wvVocab)
      numNeighbours = len(wordNeighbours)
      #no neighbours, pass - use data estimate
      if numNeighbours == 0:
        continue
      # the weight of the data estimate if the number of neighbours
      newVec = numNeighbours * wordVecs[word]
      # loop over neighbours and add to new vector (currently with weight 1)
      for ppWord in wordNeighbours:
        newVec += newWordVecs[ppWord]
      newWordVecs[word] = newVec/(2*numNeighbours)
  return newWordVecs
```

Retrofit - pull the words closer to their relations in the KB.

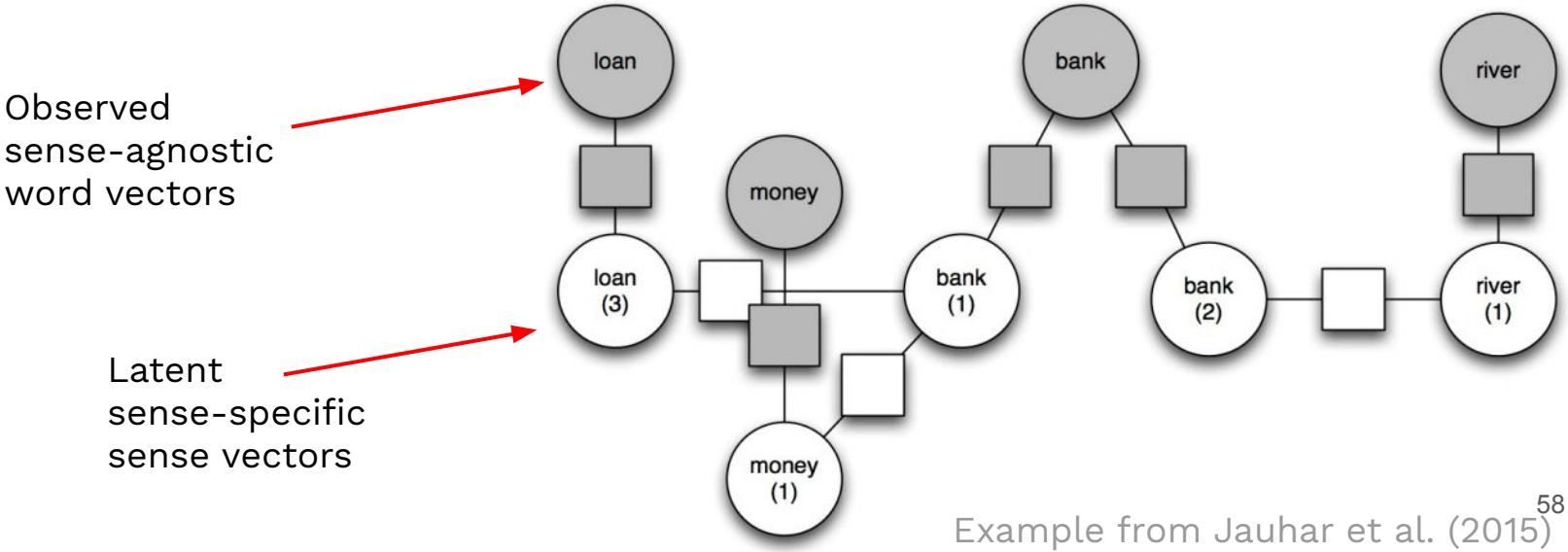
Example from Faruqui et al. (2015)

ConceptNet Numberbatch



Sense vectors by retrofitting

- Assign each sense the vector of the word.
- Pull vectors of senses closer to vectors of words they relate to in the KB.



Example from Jauhar et al. (2015)

5. 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

Problem: how many clusters is enough?

Evaluation

- Intrinsic:
  - word similarity
  - word relatedness
  - analogy relations
  - synonym selection
- External:
  - sentiment analysis
  - textual entaiment
  - question answering

2.

Semantic role labeling

Text mining

Bloomberg

Cantor Fitzgerald Sued by Partners Who Moved to Reorient

China Lawsuit

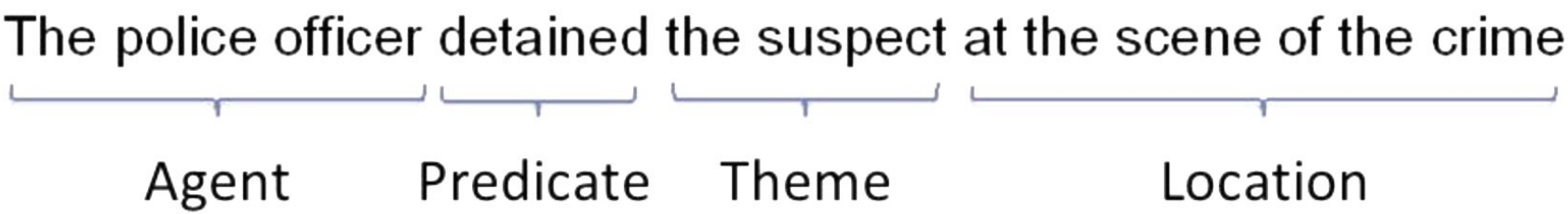
In 2011 Cantor filed a lawsuit in China against Boyer, Ainslie and other traders who left its Hong Kong office, accusing them of breaching their employment agreements and causing a 29 percent drop in average monthly revenue at the branch. Two years later, Cantor officials settled their claims against the former executives, according to filings with the Hong Kong Stock Exchange. The terms weren't made public.

Sheryl Lee, a Cantor spokeswoman, said today by phone that the company has a policy of not commenting on litigation.

Who did what to whom

- Causer (agent/force)
- Instrument
- Result
- Patient
- Theme

- Source, path, goal/recipient, location
- Experiencer
- Stimulus
- Beneficiary
- Time, manner, reason...



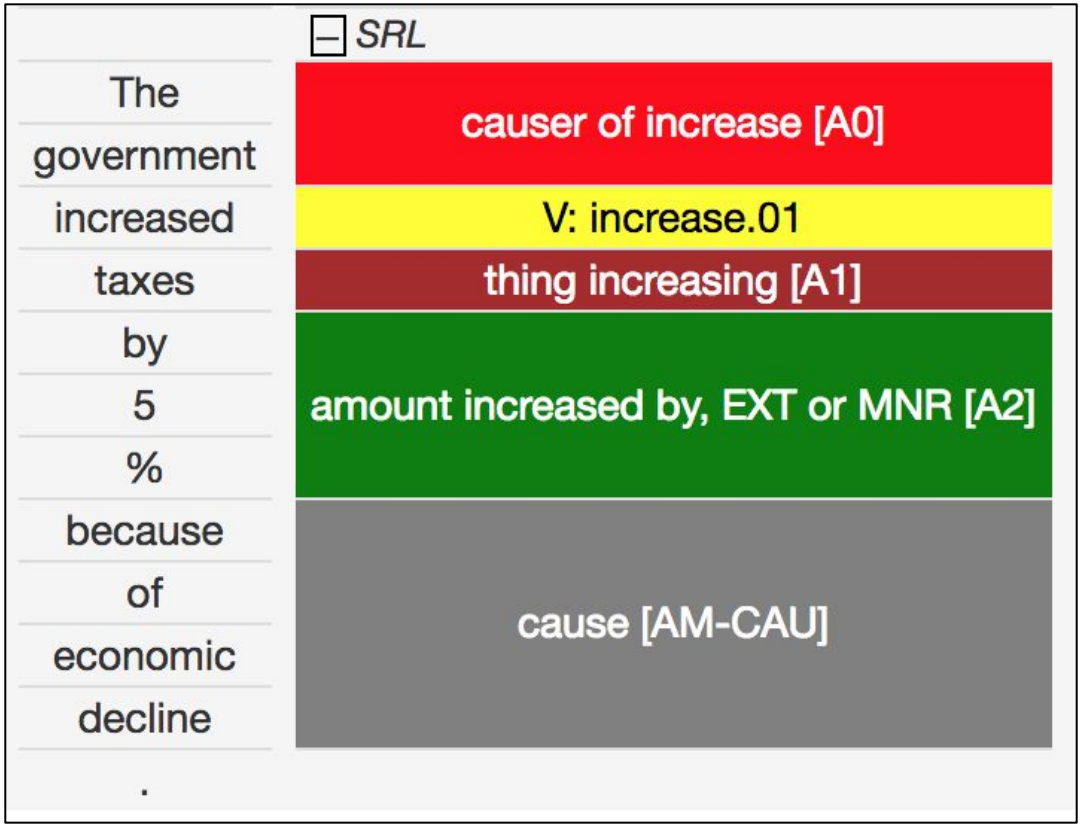
PropBank

- *increase.01 “go up incrementally”*
  - *Arg0: causer of increase*
  - *Arg1: thing increasing*
  - *Arg2: amount increased by*
  - *Arg3: start point*
  - *Arg4: end point*

TMP	when?
LOC	where?
DIR	where to/from?
MNR	how?
PRP/CAU	why?
REC	
ADV	miscellaneous

- ***The government*** *increased* ***taxes*** *by* ***5%****.*
- ***Taxes*** *increased.*

PropBank



Parsed by cogcomp.org/page/demo_view/srl

FrameNet

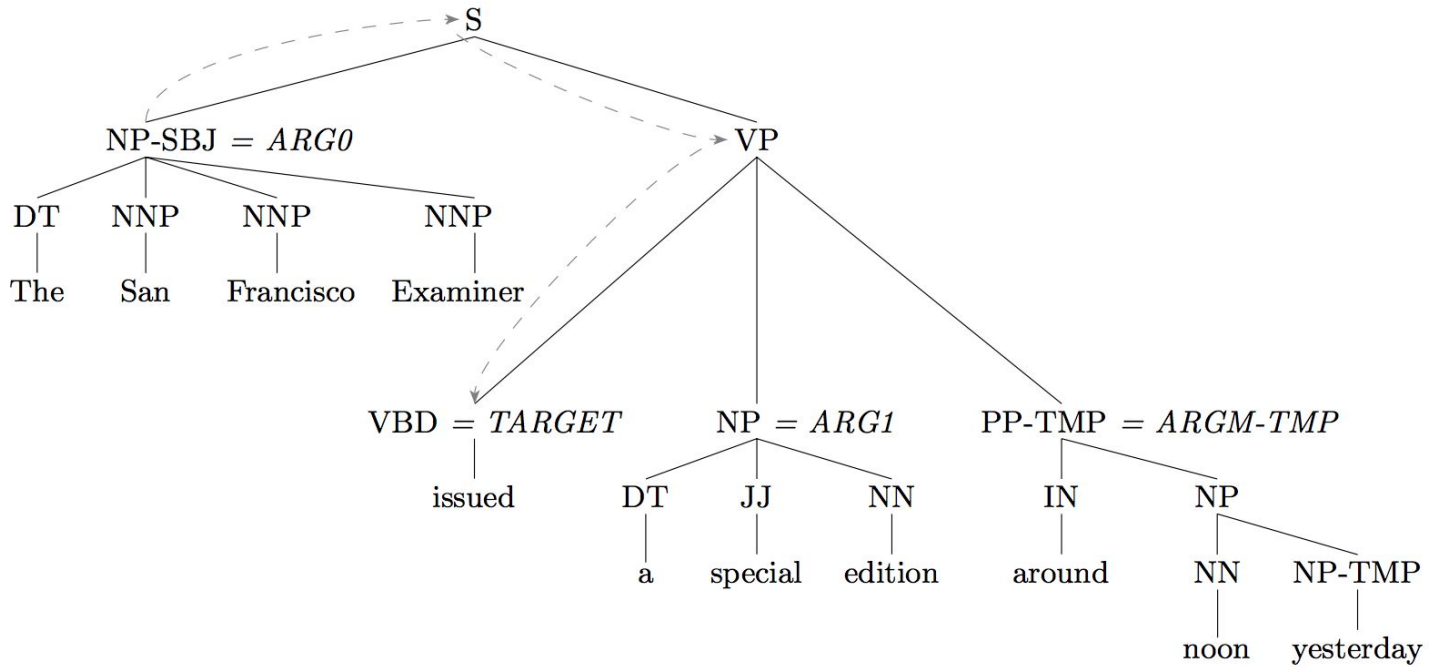
- Abandonment**:**
  - *abandon, abandoned, abandonment, leave, forget*
- An ***Agent*** leaves behind a ***Theme*** effectively rendering it no longer within their control or as one's property...
- examples:
  - ***Carolyn*** *abandoned* ***her car*** *and jumped on a bus.*
  - *Abandonment of* ***a child*** *is considered to be a serious crime in many jurisdictions.*
  - *Perhaps* ***he*** *left* ***the key*** *in the ignition.*

Basic SRL algorithm

- parse the sentence
- find all predicates *(mapped to PropBank or FrameNet)*
- for each predicate in the sentence
  - for each node in the parse tree
    - assign the semantic role (if any) for the predicate

Logistic regression, SVM, Perceptron, CRF, etc.

Basic SRL example



What features would you use?

Image from Jurafsky, Martin (2017)

Basic SRL features

- predicate
  - lemma, POS, active/passive voice, etc.
  - syntactic frame
    - e.g., VP → VBD NP PP
- node
  - label
  - headword lemma, headword POS, etc.
  - linear position, distance to predicate
- path from the constituent to the predicate
  - e.g., NP↑S↓VP↓VBD
- more: ngrams, dependencies, etc.

Basic SRL notes

- Evaluation:
  - precision, recall, f-measure
- Also possible two steps:
  - identification
    - to prune unlikely constituents
    - to avoid nested constituents
  - classification
- To avoid two ARG0s:
  - return probability distribution of classes for nodes
  - do reranking

3.

Semantic parsing

Question answering

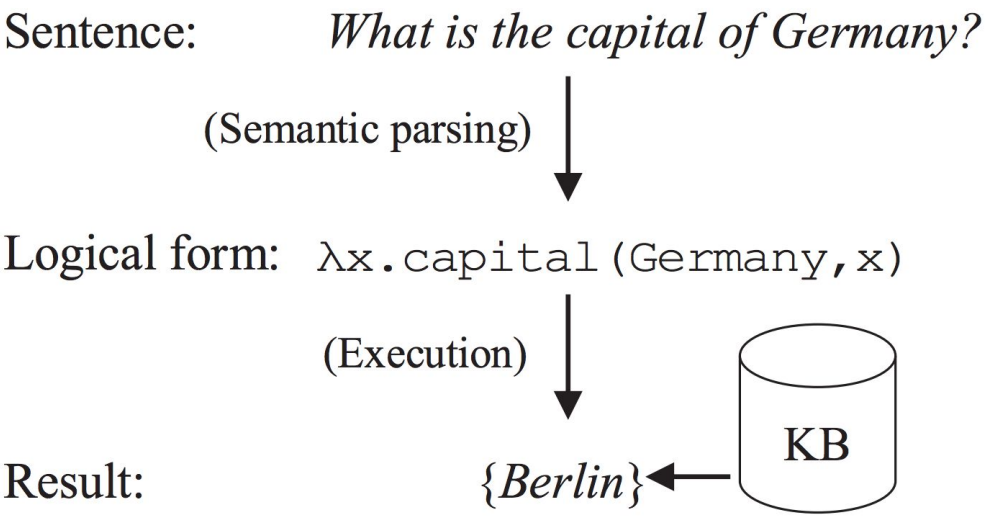
- IR-based
  - search across a collection of documents or web
- knowledge-based
  - search a knowledge base
- hybrid (e.g., IBM’s Watson)

IR-based question answering

- formulate the query
  - *e.g., process morphology, use synonyms or paraphrasing*
- find relevant documents
  - *e.g., indexed with tf-idf or using a search engine*
- rank text paragraphs from the database by relevance
  - *using the answer type*
  - *using NEs, keywords, keyword ngrams, similarity*
- do span labelling
  - *NER or patterns*
  - *ML: NEs, keywords, is appositive, novel, followed by comma...*
  - *DL: compute query vector and compare it with tokens' vectors to detect start and end of the span*

Knowledge-based question answering

- build a semantic representation of the query
- query a knowledge base



Language as a logical structure

Using lambda calculus:

- Constants
  - *entity* Texas, *property* size(e, r)
- Logical connectors - $\wedge$, $\vee$, $\neg$,
- Quantification - $\forall$, $\exists$
- Lambda expressions
  - λx.borders(x, texas)
- Additional functions
  - count, argmax, argmin, etc.

Language as a logical structure

1. What states border Texas?

$$\lambda x.\text{state}(x) \wedge \text{borders}(x, \text{texas})$$

Language as a logical structure

1. What states border Texas?

   $\lambda x.\text{state}(x) \wedge \text{borders}(x, \text{texas})$

2. What is the largest state bordering Texas?

   $\text{argmax}(\lambda x.\text{state}(x) \wedge \text{borders}(x, \text{texas}), \lambda x.\text{size}(x))$

Language as a logical structure

1. What states border Texas?

   $\lambda x.\text{state}(x) \wedge \text{borders}(x, \text{texas})$

2. What is the largest state bordering Texas?

   $\text{argmax}(\lambda x.\text{state}(x) \wedge \text{borders}(x, \text{texas}), \lambda x.\text{size}(x))$

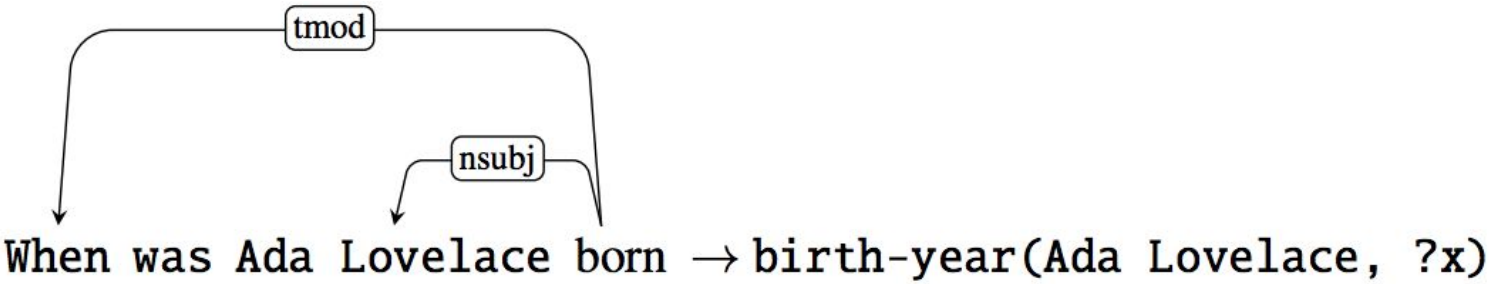
3. What states border the state that borders the most states?

   $\lambda x.\text{state}(x) \wedge \text{borders}(x, \text{argmax}(\lambda y.\text{state}(y), \lambda y.\text{count}(\lambda z.\text{state}(z) \wedge \text{borders}(y, z))))$

Semantic parsing

- parse a sentence
- split the parse tree into tuples by predicates
- classify each predicate to a KB relation
- map arguments

Data: QA datasets.



Semantic parsing variation problem

Problems:

- each KB has its own set of relations
- the same relation can be expressed with different words

s_1	*What is the population of Berlin?*
s_2	*How many people live in Berlin?*
lf_1	λx.population(Berlin,x)
lf_2	count(λx.person(x)$\wedge$live(x,Berlin))

Semantic parsing variation problem

Paraphrasing:

- with a dictionary
  - *daughter => female child*
- with a paraphrase database
  - *created from => made out of*
- with templates
  - *How many people live in => What is the population*

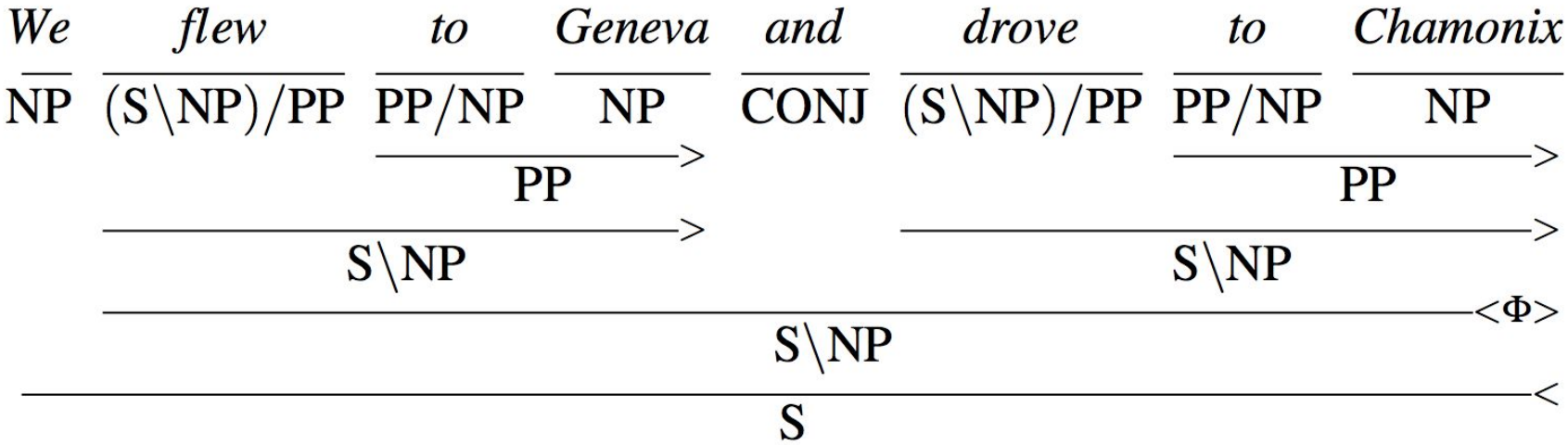
Semantic parsing variation problem

WikiAnswers

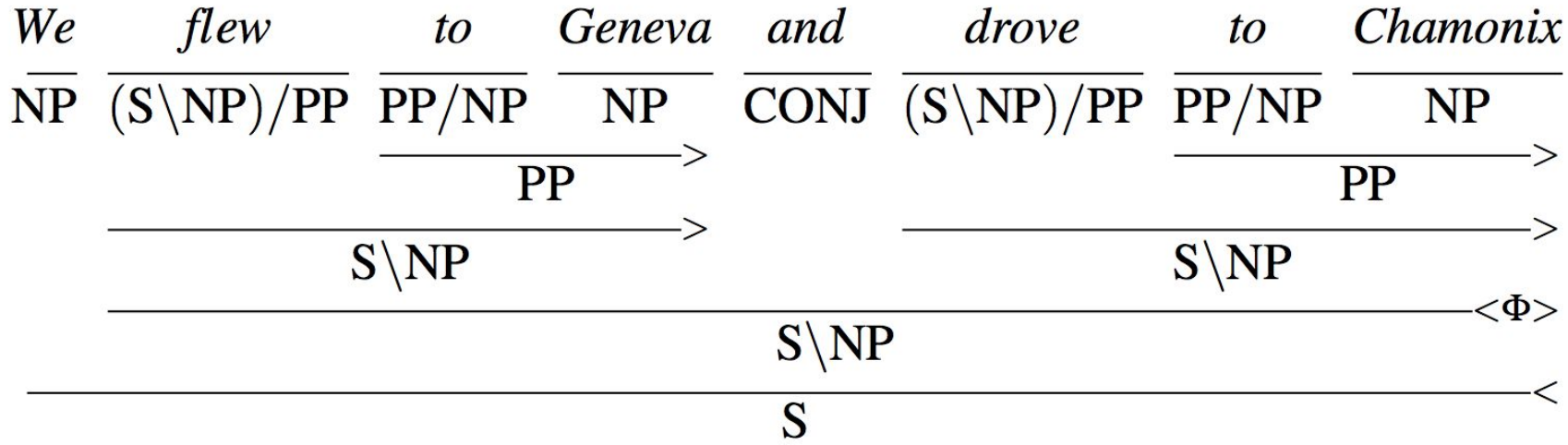
How many people live in chembakolli? *How many people is in chembakolli?* *How many people live in chembakolli india?* *How many people live there chembakolli?* *How many people live there in chembakolli?* *What is the population of Chembakolli india?*
What currency is used on St Lucia? *What is st lucia money?* *What is the money used in st lucia?* *What kind of money did st lucia have?* *What money do st Lucia use?* *Which money is used in St Lucia?*

Taken from Chen et al. (2016)

Combinatory Categorial Grammar

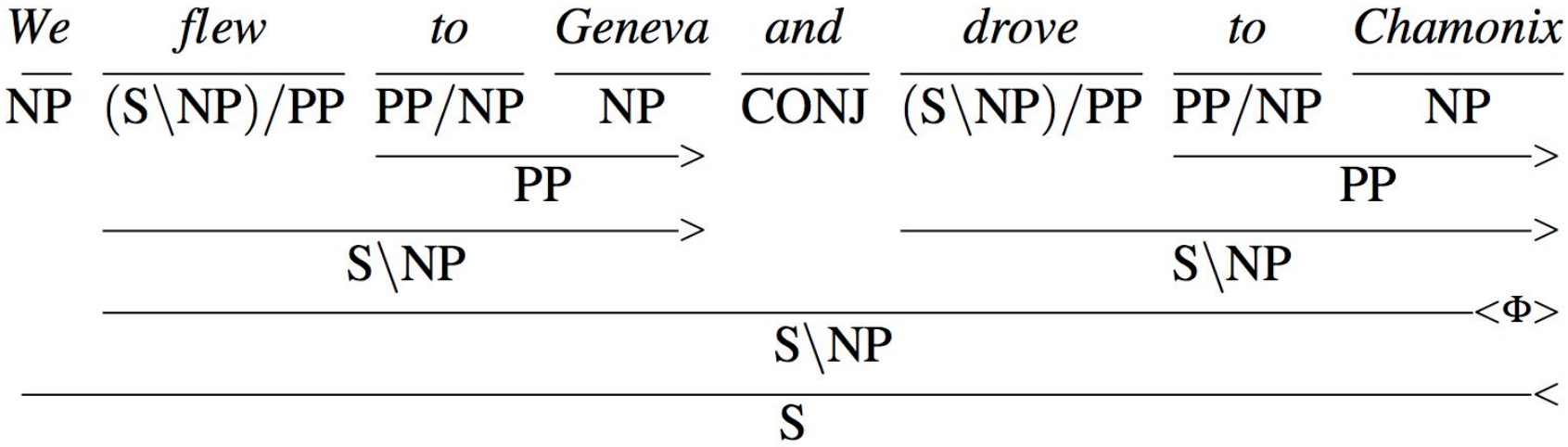


Combinatory Categorial Grammar



flew(we, to(Geneva)) $\wedge$ *drove(we, to(Chamonix))*

Combinatory Categorial Grammar



More nice examples from NLTK:
https://github.com/nltk/nltk/blob/develop/nltk/test/ccg_semantics.doctest

Combinatory Categorial Grammar

- Categories
  - atomic elements: *NP, ADJ, S*
  - functions: *(S\NP)/ADJ*
- Lexicon
  - Tom: NP
  - cancel: (S\NP)/NP
  - give: ((S\NP)/NP)/NP
  - be: (S\NP)/ADJ
- Rules
  - X/Y Y ⇒ X
  - Y X\Y ⇒ X
  - X CONJ X ⇒ X

CCG parsing

- tagging
  - a sequence labeling task (425 categories)
- A* algorithm
  - a modification of the probabilistic CKY
  - instead of filling in the whole chart, rerank elements by probability and proceed with the most probable ones
- Evaluation
  - Parseval
  - Recovery of word-word dependencies

CCG Bank

```
{S[dcl]
  {S[dcl]
    {NP {N {N/N  Mr.}
           {N  Vinken}}}
    {S[dcl]\NP {(S[dcl]\NP)/NP  is}
               {NP {NP {N  chairman}}
                   {NP\NP {(NP\NP)/NP  of}
                          {NP {NP {N {N/N  Elsevier}
                                     {N  N.V.}}}
                              {NP[conj] {,  ,}
                                        {NP {NP[nb]/N  the}
                                            {N {N/N  Dutch}
                                               {N {N/N  publishing}
                                                  {N  group}}}}}}}}}}
  {.  .}}
```

Abstract Meaning Representation

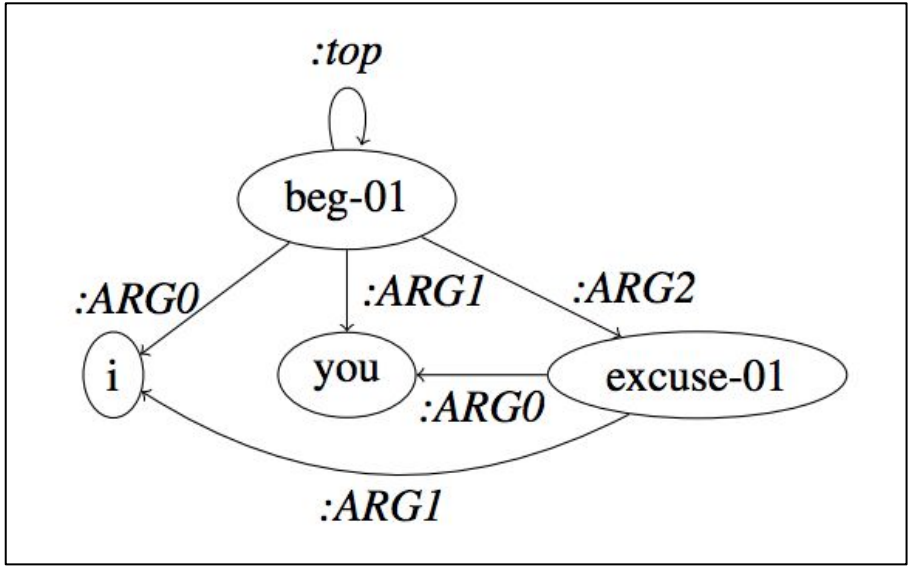
The boy desires the girl to believe him.
The boy wants the girl to believe him.
The boy desires to be believed by the girl.
The boy has a desire to be believed by the girl.
The boy's desire is for the girl to believe him.
The boy is desirous of the girl believing him.

```
(w / desire-01
   :ARG0 (b / boy)
   :ARG1 (b2 / believe-01
              :ARG0 (g / girl)
              :ARG1 b))
```

Abstract Meaning Representation

Each sentence is a ***rooted***, ***directed***, ***acyclic*** graph with labels on edges (relations) and leaves (concepts). Functional words are omitted.

```
(b / beg-01
    :ARG0 (i / i
    :ARG1 (y / you)
    :ARG2 (e / excuse-01
        :ARG0 y
        :ARG1 i))
```



AMR examples

his boat
his own boat

```
(b / boat
   :poss (h / he))
```

The boy must not go.
It is obligatory that the boy not go.

```
(o / obligate-01
   :ARG2 (g / go-02
            :ARG0 (b / boy)
            :polarity -))
```

AMR examples

He described the mission as a failure.
As he described it, the mission was a failure.
His description of the mission: failure.

```
(d / describe-01
   :ARG0 (h / he)
   :ARG1 (m / mission)
   :ARG2 (f / fail-01))
```

AMR for question answering

Which state borders with Kansas?

```
(b / border-01
   :ARG0 (s / state
            :name (n / name :op1 (a / amr-unknown)))
   :ARG1 (s2 / state
            :name (n2 / name :op1 “Kansas”)))
```

Does Texas border with Kansas?

```
(b / border-01
   :ARG0 (s / state
            :name (n / name :op1 “Texas”))
   :ARG1 (s2 / state
            :name (n2 / name :op1 “Kansas”))
   :polarity (a / amr-unknown))
```

AMR examples

About 14,000 people fled their homes at the weekend after a local tsunami warning was issued, the UN said on its Web site.

```
(s / say-01
  :ARG0 (g / organization
          :name (n / name :op1 "UN"))
  :ARG1 (f / flee-01
          :ARG0 (p / person
                  :quant (a / about :op1 14000))
          :ARG1 (h / home :poss p)
          :time (w / weekend)
          :time (a2 / after
                  :op1 (w2 / warn-01
                         :ARG1 (t / tsunami)
                         :location (l / local))))
  :medium (s2 / site
            :poss g
            :mod (w3 / web)))
```

AMR parsing

- Data - AMR Bank
  - Little Prince :)
  - 39,260 sentences of newswire, discussion forum and other web logs, television transcripts
  - 6,452 sentences of Bio AMR (cancer-related PubMed articles)
  - need alignment

- Parsing algorithms:
  - graph-based
  - transition-based

AMR alignment

JAMR Aligner (2014):

- align graph fragments to spans of words
- uses a set of handcrafted rules in a specific order:
  - dates
  - named entities
  - predicate
  - single concepts
  - fuzzy single concepts
  - etc.
- precision: 79%, recall: 63%

AMR alignment

JAMR Aligner (2014)

IAEA accepted North Korea 's proposal in November.

```
(a / accept-01
    :ARG0 (o / organization
        :ARG0 (n / name
            :op1 "IAEA"))
    :ARG1 (t2 / thing
        :ARG1-of (p / propose-01
            :ARG0 (c / country
                :name (n2 / name
                    :op1 "North"
                    :op2 "Korea"))))
    :time (d / date-entity
            :month 11))
```

AMR alignment

ISI Aligner (2014):

- convert the graph to a linear string
- remove stop words from the sentence and auxiliary tags from the graph
- stem to the first 4 letters both graph and sentence
- use word alignment model

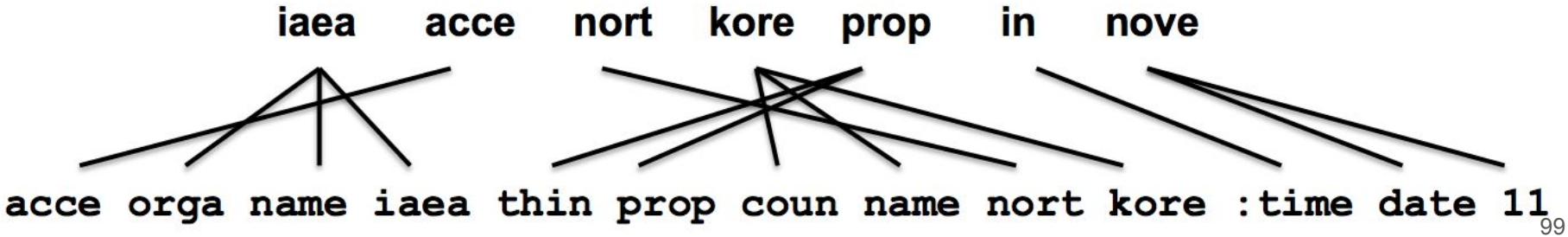
IAEA accepted North Korea proposal in November

```
accept organization name IAEA thing propose-01
country name North Korea :time date-entity 11
```

AMR alignment

ISI Aligner (2014):

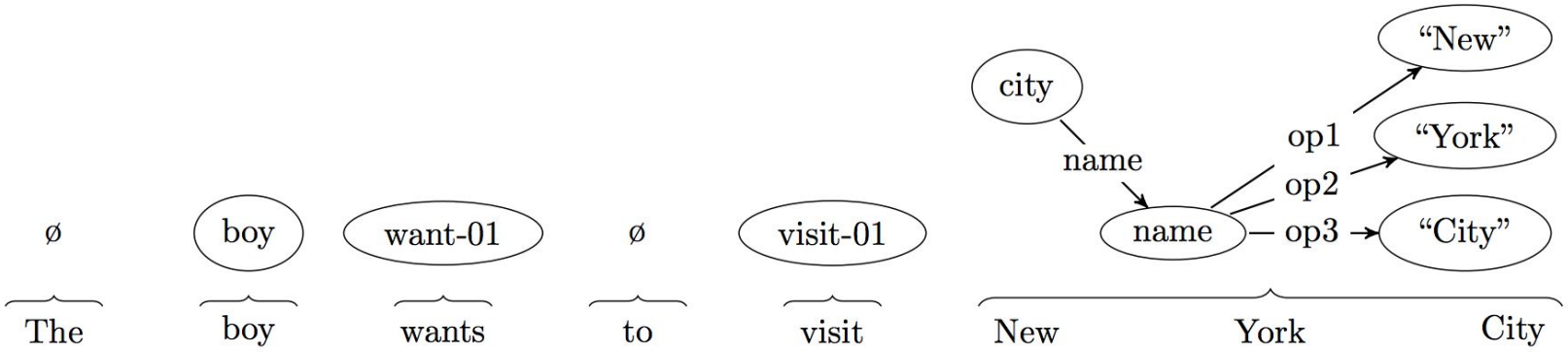
- convert the graph to a linear string
- remove stop words from the sentence and auxiliary tags from the graph
- stem to the first 4 letters both graph and sentence
- use word alignment model
- precision: 79%, recall: 72%



AMR: graph-based parsing

Algorithm

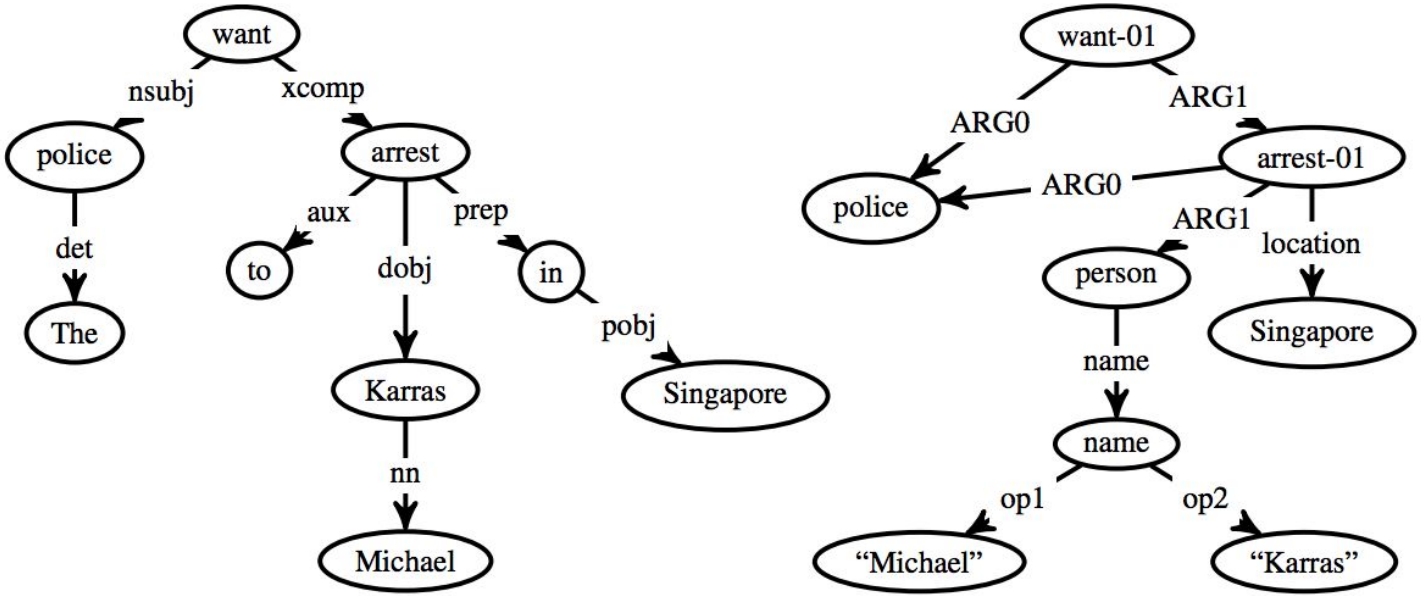
- identify concepts
  - *a sequence labeling task*
- identify relations (maximum spanning connected subgraph)
  - *similar to graph-based dependency parsing*



Example from Flanigan et al (2014)

AMR: transition-based parsing

Idea: convert dependency tree into AMR graph.



Example from Wang et al (2015)

AMR: transition-based parsing

Setup:

- σ - a buffer that stores indices of the nodes which have not been processed; $\sigma 0$ - top of the buffer
- β - a buffer $[\beta 0, \beta 1, \ldots, \beta j\ ]$ and each element βi of β indicates the edge $(\sigma 0, \beta i)$ which has not been processed in the partial graph; $\beta 0$ - top of the buffer
- G - span graph that stores the partial parses

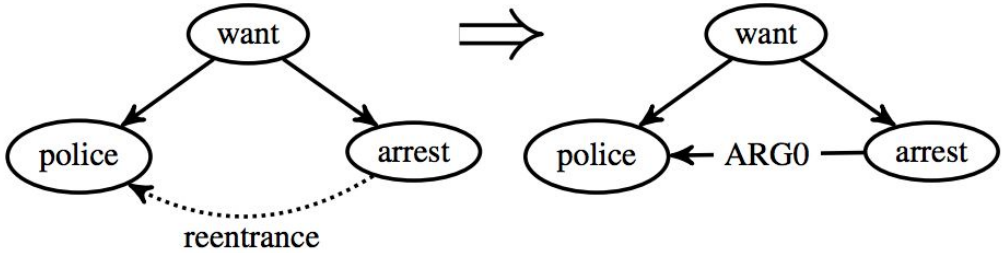
AMR: transition-based parsing

Actions:

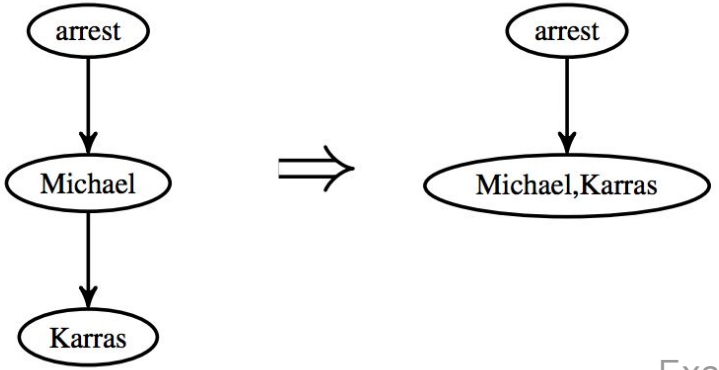
- NEXT-EDGE-l_r - attach edge *(σ0, β0)* and move to next node
- SWAP-l_r - swap nodes and attach with edge
- REATTACH$_k$-l_r - delete edge and reattach to already processed node
- REPLACE-HEAD - replace node with another node
- REENTRANCE$_k$-l_r - attach edge to already processed node
- MERGE - merge two nodes
- NEXT-NODE-l_c - label with concept and move to next word
- DELETE-NODE - delete a word

AMR: transition-based parsing

REENTRANCE$_k$-l$_r$



MERGE



Example from Wang et al (2015)

AMR evaluation: smatch

```
(w / want-01
   :ARG0 (b / boy)
   :ARG1 (b2 / believe-01
             :ARG0 (g / girl)
             :ARG1 b))
```

```
instance(w, want-01)          /* w is an instance of wanting */
instance(b, boy)              /* b is an instance of boy */
instance(b2, believe-01)      /* b2 is an instance of believing */
instance(g, girl)             /* g is an instance of girl */
ARG0(w, b)                    /* b is the wanter in w */
ARG1(w, b2)                   /* b2 is the wantee in w */
ARG0(b2, g)                   /* g is the believer in b2 */
ARG1(b2, b)                   /* b is the believee in b2 */
```

Minimal Recursive Semantics

The jungle lion was chasing a small giraffe.

$$
\begin{array}{l}
\langle\, h_1, e_3, \\
\quad \left| \begin{array}{l}
h_4{:}\_\text{the\_q}(x_6, h_7, h_5), \\
h_8{:}\text{compound}(e_{10}, x_6, x_9), \\
h_{11}{:}\text{udef\_q}(x_9, h_{12}, h_{13}), \\
h_{14}{:}\_\text{jungle\_n\_1}(x_9), \\
h_8{:}\_\text{lion\_n\_1}(x_6), \\
h_2{:}\_\text{chase\_v\_1}(e_3, x_6, x_{15}), \\
h_{16}{:}\_\text{a\_q}(x_{15}, h_{18}, h_{17}), \\
h_{19}{:}\_\text{small\_a\_1}(e_{20}, x_{15}), \\
h_{19}{:}\_\text{giraffe\_n\_1}(x_{15})
\end{array} \right| \\
\quad \{\, h_1 =_q h_2, h_7 =_q h_8, h_{12} =_q h_{14}, h_{18} =_q h_{19} \,\} \,\rangle
\end{array}
$$

4.

Textual entailment

Textual entailment

We say that text *T* entails hypothesis *H* (i.e., *T* ⇒ *H*) if the meaning of *H* can be inferred from the meaning of *T*, as would typically be interpreted by people.

- Positive TE (text entails hypothesis):
  - *A girl swings high in the air.*
  - *A girl is on a swing.*
- Negative TE (text contradicts hypothesis):
  - *A girl swings high in the air.*
  - *A girl is laying in the pool.*
- Non-TE (text does not entail nor contradict):
  - *A girl swings high in the air.*
  - *A girl is looking for her mother.*

Textual entailment

Text: The purchase of Houston-based LexCorp by BMI for $2Bn prompted widespread sell-offs by traders as they sought to minimize exposure. LexCorp had been an employee-owned concern since 2008.

Hyp 1**:** BMI acquired an American company.
Hyp 2**:** BMI bought employee-owned LexCorp for $3.4Bn.
Hyp 3**:** BMI is an employee-owned concern.

Textual entailment

Text: The purchase of Houston-based LexCorp by BMI for $2Bn prompted widespread sell-offs by traders as they sought to minimize exposure. LexCorp had been an employee-owned concern since 2008.

Hyp 1**:** BMI acquired an American company.
Hyp 2**:** BMI bought employee-owned LexCorp for $3.4Bn.
Hyp 3**:** BMI is an employee-owned concern.

- NER: *LexCorp, BMI, $2Bn*
- WordNet: *purchase (n.) => purchase (v.), acquire, buy*
- Knowledge base: *Houston-based => American*
- Syntactic analysis + semantic role labeling / semantic parsing

Textual entailment

Text: The purchase of Houston-based LexCorp by BMI for $2Bn prompted widespread sell-offs by traders as they sought to minimize exposure. LexCorp had been an employee-owned concern since 2008.

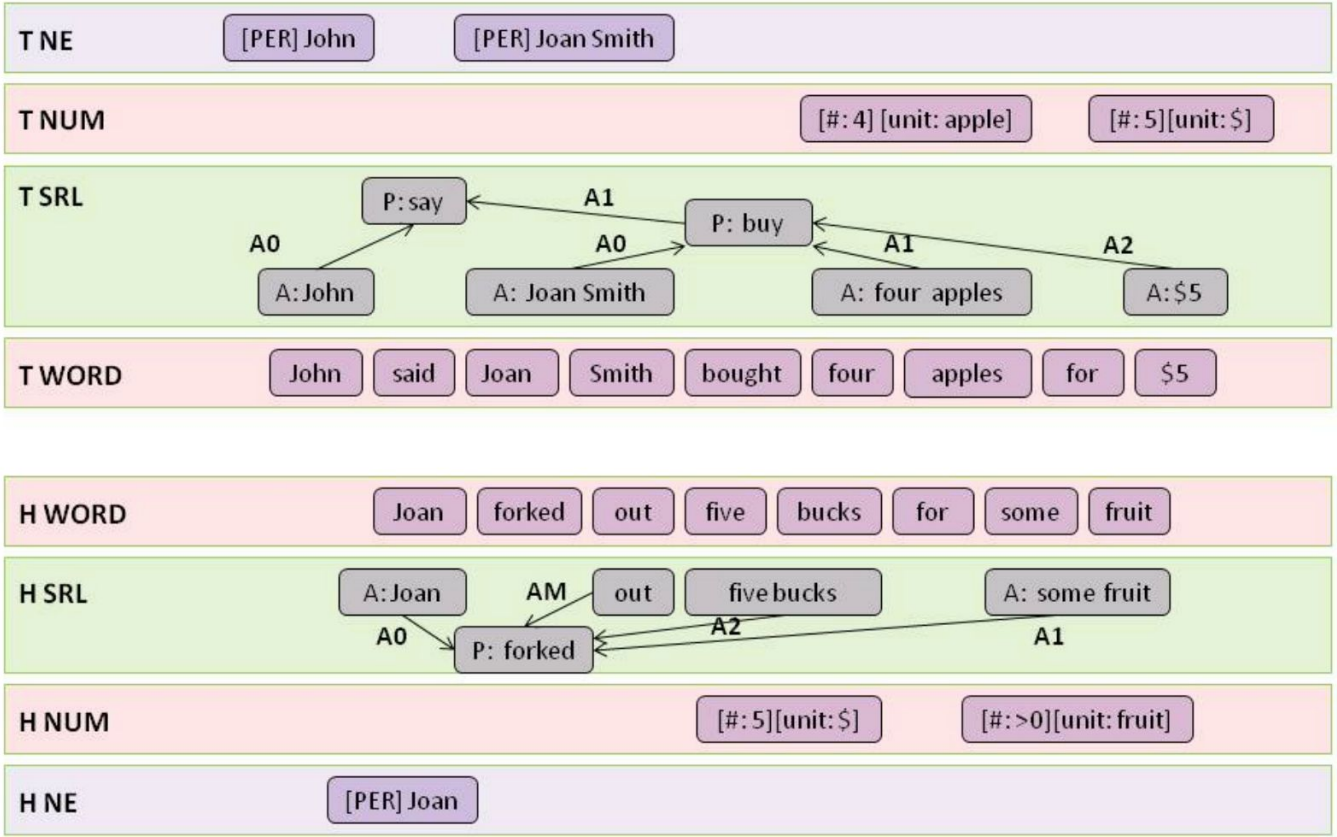
Hyp 1: BMI acquired an American company.
Hyp 2: BMI bought employee-owned LexCorp for $3.4Bn.
Hyp 3: BMI is an employee-owned concern.

- Normalization of quantities: *$2Bn => 2,000,000,000*
- Fuzzy match of NE: *BMI => Business Machine Inc.*
- Coreference resolution
- Scope of negation

Textual entailment pipeline

- Text and hypothesis preprocessing
- Align syntactic/semantic representations of *T* and *H*
  - only a small portion of *T* is relevant to *H*
- Features: compare the constituents
  - match of NEs and numerical measures
  - lexico-semantic relations (synonyms, antonyms, hypernyms, meronyms, entailment)
  - word similarity
  - noun-verb relations (*“is a winner” => “won”*)
  - idioms (*“kick the bucket” => “die”*)
  - paraphrases (*“has been unable to” => “could not”*)
  - syntactic mapping (active => passive)

Textual entailment pipeline



Textual entailment: TEASE

- Extract entailment paraphrases from web using a dependency parser
  - both lexical and grammatical
- Apply iteratively to transform *T* to *H*

X write *Y*	*X* who write *Y*	*X* produce *Y*
	X publish *Y*	*X* pen *Y*
	X compose *Y*	*X* create *Y*
	read *Y* by *X*	*X*'s *Y*
	Y attributed to *X*	*X* complete *Y*
	perform *Y* by *X*	*X* book of *Y*
	X writer of *Y*	*X* say in *Y*
	selected *Y* of *X*	*X* work include *Y*

References

- Nematzadeh, Meylan and Griffiths (2017), The Unreasonable Effectiveness of Counting Words Near Other Words
- Navigli and Lapata (2010), An Experimental Study of Graph Connectivity for Unsupervised Word Sense Disambiguation
- Raiman and Raiman (2018), Discovering Types for Entity Disambiguation
- Moro, Raganato, and Navigli (2015), Babelfy: Entity Linking meets Word Sense Disambiguation
- Robert Speer (2018), ConceptNet 5.5: An Open Multilingual Graph of General Knowledge
- 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
- Faruqui et al. (2015), Retrofitting Word Vectors to Semantic Lexicons

References

- Jauhar, Dyer et al. (2015), Ontologically Grounded Multi-sense Representation Learning for Semantic Vector Space Models
- Athiwaratkun and Wilson (2017), Multimodal Word Distributions
- Chen et al. (2016), Sentence Rewriting for Semantic Parsing
- Hockenmaier and Steedman (2007), CCGbank
- CCG Software: http://groups.inf.ed.ac.uk/ccg/software.html
- AMR: https://amr.isi.edu/index.html
- Chen and Palmer (2017), Unsupervised AMR-Dependency Parse Alignment
- Wang et al. (2015), A Transition-based Algorithm for AMR Parsing
- Flanigan et al. (2014), A Discriminative Graph-Based Parser for AMR
- Alicia Ageno (2014), Textual Entailment
- Mark Sammons (2012), Recognizing Textual Entailment
- Jurafsky and Martin (2017), Speech and Language Processing, Chapters 14, 18 and 23