Word Sense Disambiguation

- Jurafsky, D. and Martin, J. H. (2019): Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Third Edition. Pearson: New Jersey: Chapter 18: https://web.stanford.edu/~jurafsky/slp3/18.pdf

- Navigli, R. (2009). Word sense disambiguation: A survey. ACM computing surveys (CSUR), 41(2), 1-69. https://dl.acm.org/doi/abs/10.1145/1459352.1459355

- Agirre, E., Edmonds, P. (2006): Word Sense Disambiguation: Algorithms and Applications (Text, Speech and Language Technology). Springer, Heidelberg

- Bevilacqua, M., Pasini, T., Raganato, A., & Navigli, R. (2021). Recent trends in word sense disambiguation: A survey. In Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21. International Joint Conference on Artificial Intelligence, Inc. https://helda.helsinki.fi/bitstream/handle/10138/333318/0593.pdf

PLAN OF THE LECTURE

- Lexical Ambiguity
- Word Sense Disambiguation
- Embeddings for Senses and Hypernyms
- Lexical Substitution

What do words mean?

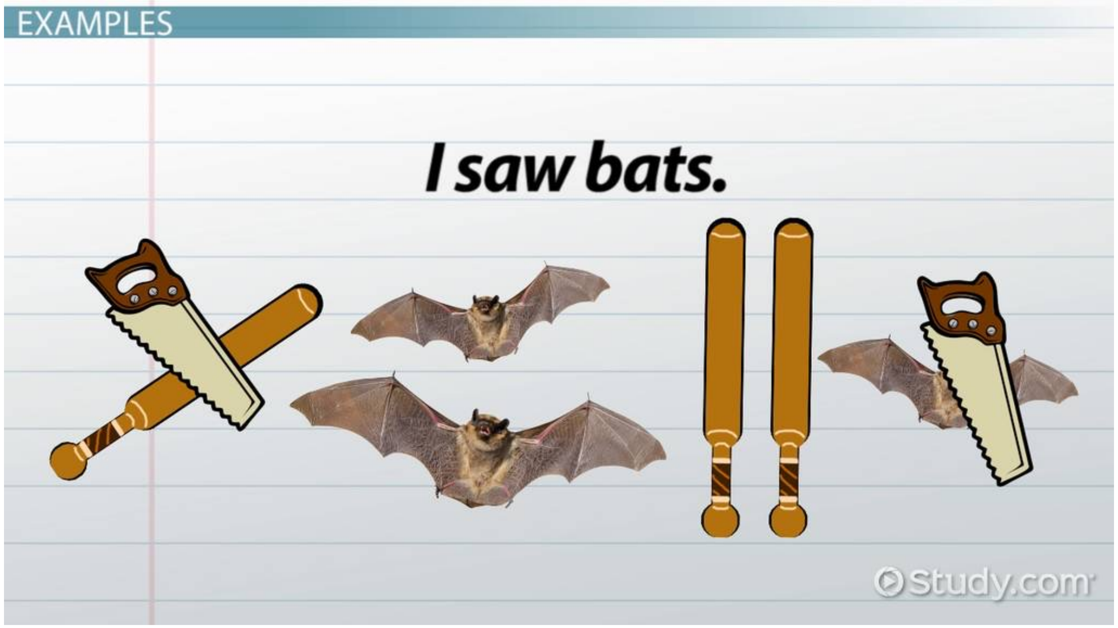
https://web.stanford.edu/~jurafsky/slp3/18.pdf

- A sense (or word sense) is a discrete representation of one aspect of the meaning of a word.
- Superscript represents different meanings.

mouse^1 : .... a *mouse* controlling a computer system in 1968.
mouse^2 : .... a quiet animal like a *mouse*
bank^1 : ...a *bank* can hold the investments in a custodial account ...
bank^2 : ...as agriculture burgeons on the east *bank*, the river ...

They rarely *serve* red meat, preferring to prepare seafood.
He *served* as U.S. ambassador to Norway in 1976 and 1977.

A Motivating Example



- Ambiguous words have multiple senses
- Word Sense Disambiguation assigns a single (out of multiple) sense in a given context
- Semantic disambiguities: assume that syntactic disambiguation has already been performed

http://study.com/academy/lesson/lexical-ambiguity-definition-examples.html

Word Sense Disambiguation (WSD)

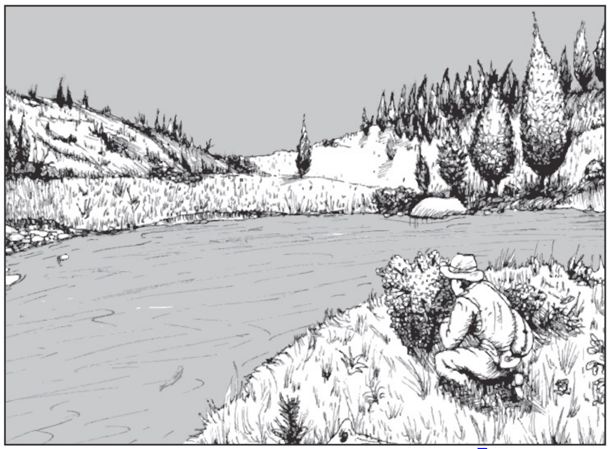


Image © dero

polysemous

synonymous

Concept Layer

Lexical Layer

He sat on the river $\boxed{\text{bank}}$ and counted his $\boxed{\text{dough}}$.

She went to the $\boxed{\text{bank}}$ and took out some $\boxed{\text{money}}$.

Types of Semantic Lexical Ambiguity

- **Homonymy**: two or more meanings happen to be expressed with the same string
  - withdrawing money from the **bank**
  - embark on a boat from the river **bank**

- **Polysemy**: the same string has different, but related senses, stemming from the same origin
  - the **bank** was robbed by Billy the Kid
  - the **bank** was constructed by a famous architect

PLAN OF THE LECTURE

- Lexical Ambiguity
- Word Sense Disambiguation
- Embeddings for Senses and Hypernyms
- Lexical Substitution

Approaches to WSD

- Knowledge Based Approaches (‘unsupervised’)
  – Rely on knowledge resources like WordNet, Thesaurus etc.
  – May use grammar rules for disambiguation.
  – May use hand coded rules for disambiguation.
- Machine Learning Based Approaches (‘supervised’)
  – Rely on corpus evidence.
  – Train a model using tagged or untagged corpus.
  – Probabilistic/Statistical models.
- Hybrid Approaches
  – Use corpus evidence as well as semantic relations from WordNet.
- Unsupervised, knowledge-free Approaches
  – Induce sense inventory
  – Disambiguate in context

WSD using Selectional Preferences and Arguments

Sense 1

This airline **serves** dinner in the evening flight.

- **serve** (Verb)
  - agent
  - object – edible

Sense 2

This airline **serves** the sector between Munich and Rome.

- **serve** (Verb)
  - agent
  - object – sector

Requires exhaustive enumeration of:

- Argument-structure of verbs.
- Selectional preferences of arguments.
- Description of properties of words such that meeting the selectional preference criteria can be decided.

  E.g. This flight serves the “region” between Paris and Warsaw.

  How do you decide if “region” is compatible with “sector”

Overlap-based Approaches

- Requires a machine readable dictionary: Oxford, Wiktionary, etc.
- Find the overlap between the features of different senses of an ambiguous word (sense bag) and the features of the words in its context.
- These features could be sense definitions, example sentences, etc.
- The sense which has the maximum overlap is selected as the contextually appropriate sense.

Concise Oxford English Dictionary

- **dance**
  - verb

    1 ▶ move rhythmically to music, typically following a set sequence of steps. ▶ perform (a particular dance or a role in a ballet).

    2 ▶ move in a quick and light or lively way. ▶ (of someone's eyes) sparkle with pleasure or excitement.
  - noun

    1 ▶ a series of steps and movements that match the speed and rhythm of a piece of music. ▶ an act or period of dancing.

    2 ▶ a social gathering at which people dance.

    3 ▶ (also **dance music**) music for dancing to, especially in a nightclub.

Lesk (1986) Algorithm

- Identify senses of words in context using definition overlap
- Can do this either between gloss and context or between all sense combinations
- Main problem: zero overlap for most contexts

```
function SimplifiedLesk(word, sentence) {
   bestSense = mostFrequentSense(word);
   maxOverlap = 0;
   context = allWords(sentence);

   foreach sense in allSenses(word) {
      signature=signature(sense);
      overlap = overlap(signature, context);
      if (overlap > maxOverlap) {
         maxOverlap = overlap;
         bestSense - sense
      }
   }
   return bestSense;
}
```

Dictionary functions

- `mostFrequentSense`: returns most frequent / first sense identifier from dictionary
- `allSenses`: returns all sense identifiers for a word from dictionary
- signature: returns set of words from sense definition in dictionary

Lesk, M. (1986). Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In SIGDOC '86: Proceedings of the 5th annual international conference on Systems documentation, pages 24-26, New York, NY, USA. ACM.

Simplified Lesk Algoritm

- Lesk algorithm relies on definitions of context words to disambiguate the senses of a target word
- Simplified Lesk:
  - Measure the overlap between the (sentence) context of the target, and the definition of its senses
  - If no overlap, use most frequent sense (MFS)

■ noun

1 ▶ a series of steps and movements that match the speed and rhythm of a piece of music. ▶ an act or period of dancing.

2 ▶ a social gathering at which people dance.

3 ▶ (also **dance music**) music for dancing to, especially in a nightclub.

1, 2 or 3 ?

"With the music, the dance started with slow movements."

WordNet – an Online Lexical Database: http://wordnet.princeton.edu

- High coverage lexical-semantic network built by psychologists

POS	Monosemous Words and Senses	Polysemous Words	Polysemous Senses
Noun	101863	15935	44449
Verb	6277	5252	18770
Adjective	16503	4976	14399
Adverb	3748	733	1832
Totals	**128391**	**26896**	**79450**

- Relations:
  - ISA-relation (hyponom - hypernym, taxonomic backbone)
  - Part-of (meronym - holonym)
  - Type-instance (e.g. Obama is an instance of President)
  - Opposite-of (antonym), mostly for adjectives
  - Derivative (pertainym), e.g. crime – criminal
  - Some semantic roles between verbs and nouns, e.g. AGENT, INSTRUMENT

BabelNet: a multilingual WordNet

https://babelnet.org/

python | English | Translate into... | 29 Noun | 0 Verb | 0 Adj. | 0 Adv.

EN **python** *noun*

Large Old World boas

bn:00065461n | Concept

A soothsaying spirit or a person who is possessed by such a spirit

bn:00065462n | Concept

(Greek mythology) dragon killed by Apollo at Delphi

bn:00065463n | Named Entity | mythical | Greek mythology | mythology

A general-purpose, multi-paradigm programming language for computers created by Guido van Rossum

bn:01713224n | Named Entity

computer language | computing | programming language | software | scripting language | language | lang | programming

Monty Python were a British surreal comedy troupe who created the sketch comedy television show Monty Python's Flying Circus, which fir...

Monty Python

bn:01157670n | Named Entity

• bn:01713224n • NOUN • Named Entity • Categories: High-level programming languages, Dutch inventions, Class-based programming languages, Cross-platform free software...

EN Python (programming language) • /usr/bin/python • /usr/local/bin/python • Python language • Python programming language

Python is a widely used general-purpose, high-level programming language. *Wikipedia*

More definitions

IS A	programming language • free software • scripting language
HAS PART	pandas
HAS KIND	Stackless Python
DESIGNER	Guido van Rossum
DEVELOPER	Python Software Foundation • Guido van Rossum
DIALECTS	Cython • Stackless Python
INFLUENCED BY	ALGOL 68 • alphabet • ruby
LICENSE	Python Software Foundation License

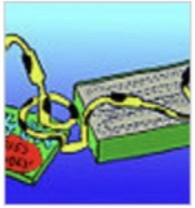
More relations

EXPLORE NETWORK









Synsets for the Noun "magazine"

WordNet Search - 3.0 - WordNet home page - Glossary - Help

Word to search for: magazine [Search WordNet]

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

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Synset

sample use

Noun

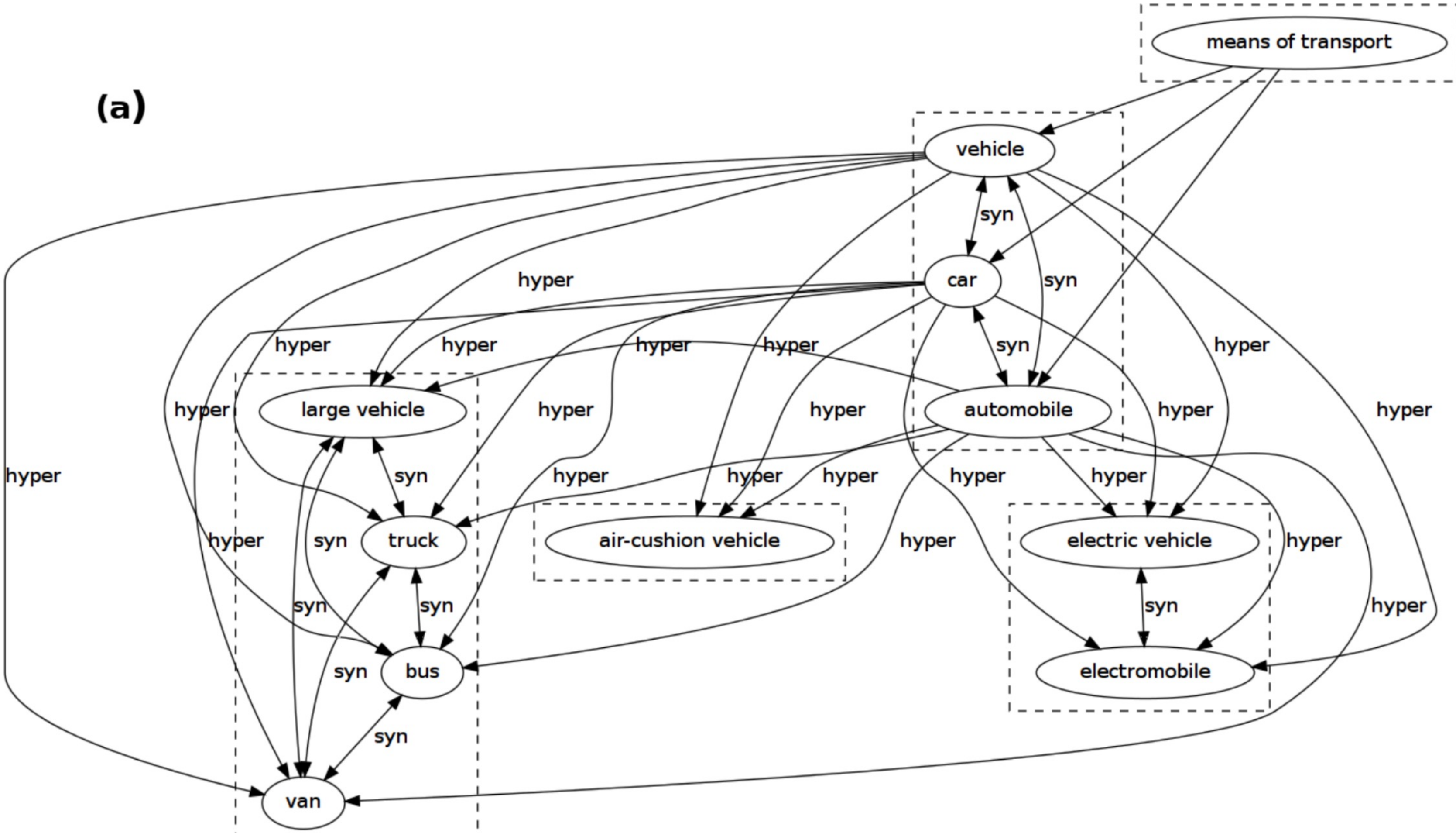
- (13)S: (n) **magazine#1**, mag#1 (a periodic publication containing pictures and stories and articles of interest to those who purchase it or subscribe to it) *"it takes several years before a magazine starts to break even or make money"*
- (2)S: (n) **magazine#2** (product consisting of a paperback periodic publication as a physical object) *"tripped over a pile of magazines"*
- (1)S: (n) **magazine#3**, magazine publisher#1 (a business firm that publishes magazines) *"he wo[rked] for a magazine"*
- S: (n) **magazine#4**, cartridge#2 (a light-tight supply chamber holding the film and supplying it for exposure as required)
- S: (n) **magazine#5**, powder store#1, powder magazine#1 (a storehouse (as a compartment on a warship) where weapons and ammunition are stored)
- S: (n) cartridge holder#1, cartridge clip#1, clip#1, **magazine#6** (a metal frame or container holding cartridges; can be inserted into an automatic gun)

gloss

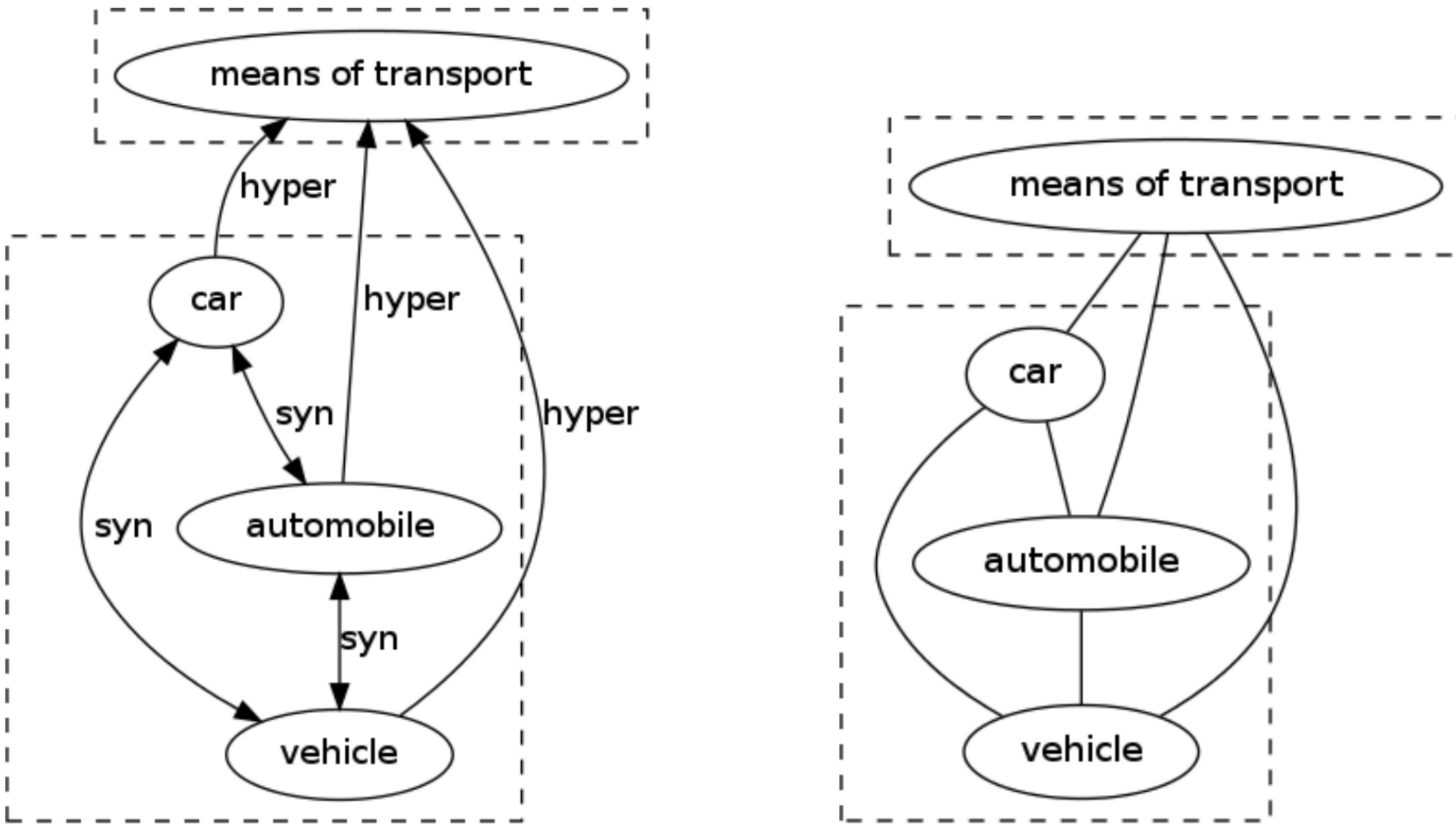
Lexical members

SemCor count

Synset: a set of synonyms



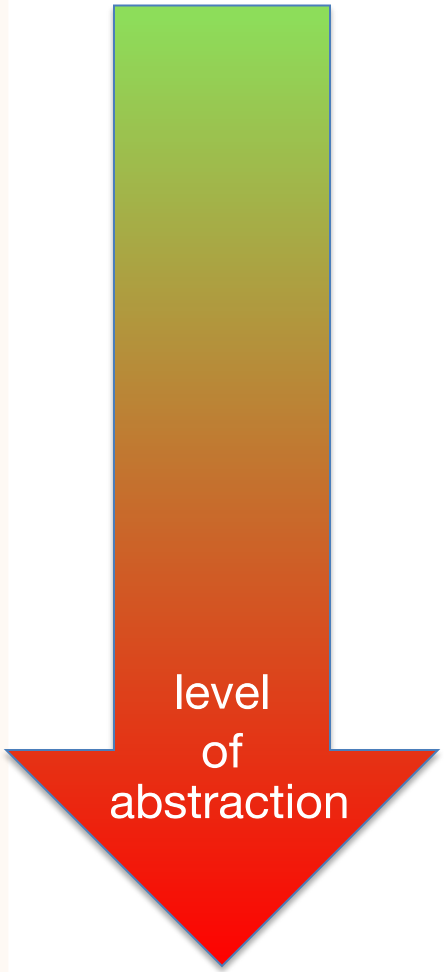
Synset: a set of synonyms



(a) **(b)**

WordNet Hypernym Chain

- (2)S: (n) **magazine#2** (product consisting of a paperback periodic publication as a physical object) *"tripped over a pile of magazines"*
  - *direct hypernym* / *inherited hypernym* / *sister term*
    - S: (n) product#2, production#3 (an artifact that has been created by someone or some process) *"they improve their product every year"; "they export most of their agricultural production"*
      - S: (n) creation#2 (an artifact that has been brought into existence by someone)
        - S: (n) artifact#1, artefact#1 (a man-made object taken as a whole)
          - S: (n) whole#2, unit#6 (an assemblage of parts that is regarded as a single entity) *"how big is that part compared to the whole?"; "the team is a unit"*
            - S: (n) object#1, physical object#1 (a tangible and visible entity; an entity that can cast a shadow) *"it was full of rackets, balls and other objects"*
              - S: (n) physical entity#1 (an entity that has physical existence)
                - S: (n) entity#1 (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))



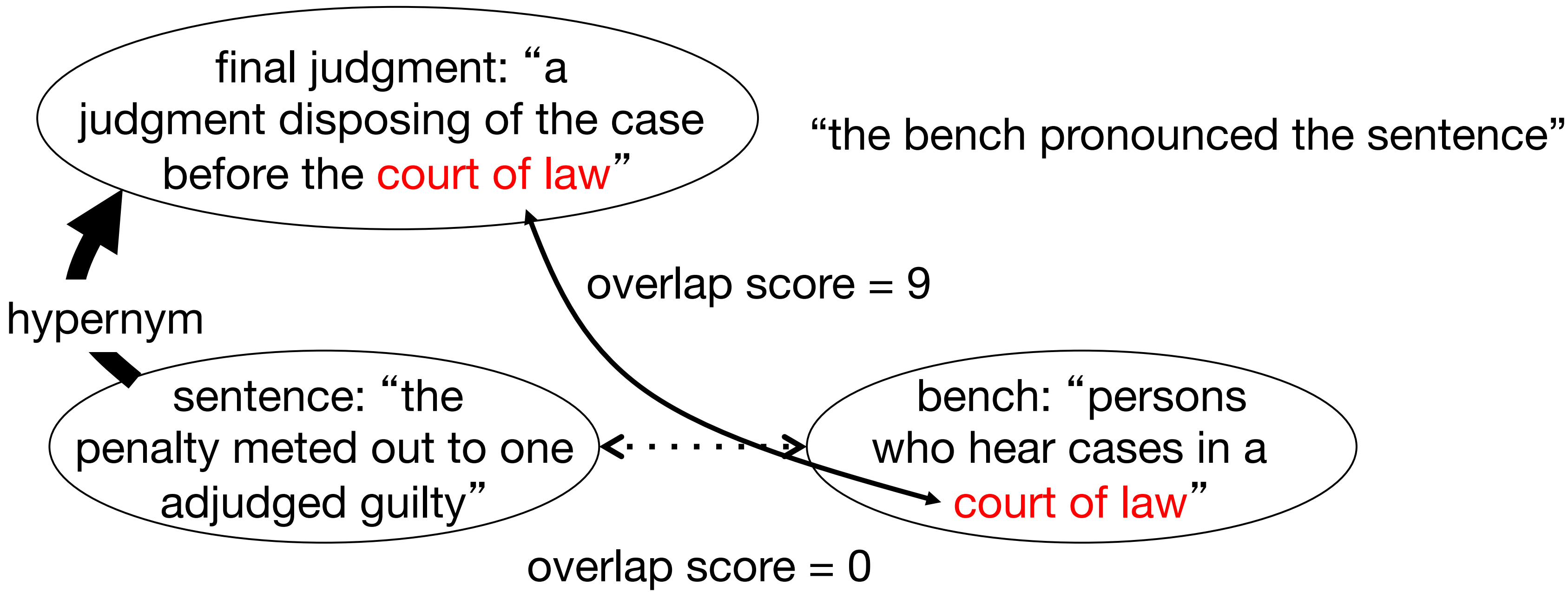
Gloss Overlaps ≈ Relatedness

- Lesk's (1986) idea: Related word senses are (often) defined *using the same words.* E.g:
  - bank(1): "a financial institution"
  - bank(2): "sloping land beside a body of water"
  - lake: "a body of water surrounded by land"
- Gloss overlaps = # content words common to two glosses ≈ relatedness
  - relatedness (bank(2), lake) = 3
  - relatedness (bank(1), lake) = 0

Problem: "lexical gap": Same or similar meaning can be expressed with a large variety of words from the vocabulary. For most pairs of glosses, overlap = 0 .

Extended Lesk (Banerjee and Pedersen, 2002)

- Utilize link structure of WordNet to pull in related glosses for overlap computation
- Addresses the overlap sparseness issue
- do this for one ambiguous word at-a-time
- Reweighting: For n-gram overlaps, add a score of n^2



(Banerjee and Pedersen, 2002). Extended Gloss Overlaps as a Measure of Semantic Relatedness. Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, pp. 805-810, August 9-15, 2003, Acapulco, Mexico.

SemEval: Shared Task for Semantic Evaluations

- Shared task initiative since 2001, currently preparing 2021 edition
- Increasing number of tasks and systems
- Core WSD tasks (for many languages):
  - lexical sample task: for a small set of ambiguous target words, a large number of labeled examples
  - all word task: every word in a short text is labeled with the appropriate sense
- Other tasks include:
  - (cross-lingual) lexical substitution
  - word sense induction
  - semantic role annotation
  - Sentiment analysis
  - temporal relation identification
  - semantic text similarity
  - …

e.g. Semeval-2 English lex. sample

Algorithm	Precision	Recall	F1
Sval-1st	0.402	0.401	0.401
Ext Lesk	0.351	0.342	0.346
Sval-2nd	0.293	0.293	0.293
Lesk	0.183	0.183	0.183
Random	0.141	0.141	0.141

Senseval/SemEval All-Word Task

Example annotation

Homeless
<head id="d00.s08.t01">people</head>
not
only
<head id="d00.s08.t04">lack</head>
safety
,
privacy
and
<head id="d00.s08.t09">shelter</head>
,
they
also
<head id="d00.s08.t13">lack</head>
the
elementary
<head id="d00.s08.t16">necessities</head>
of
nutrition
,
cleanliness
and
basic
health
care
.

- No training data, can only use lexical resources ('knowledge-based, unsupervised') and sense-labeled corpora
- All ambiguous words are marked and need to be assigned a sense
- Fine-grained scoring: upper bound is inter-annotator-agreement, which is at about 75%
- Coarse-grained scoring: inter-annotator-agreement at around 90%
- Top system performances: 65% (fine-grained), 82% (coarse-grained); MFS baseline: 78% (coarse-grained)

Senseval/SemEval Lexical Sample Task

Example annotation

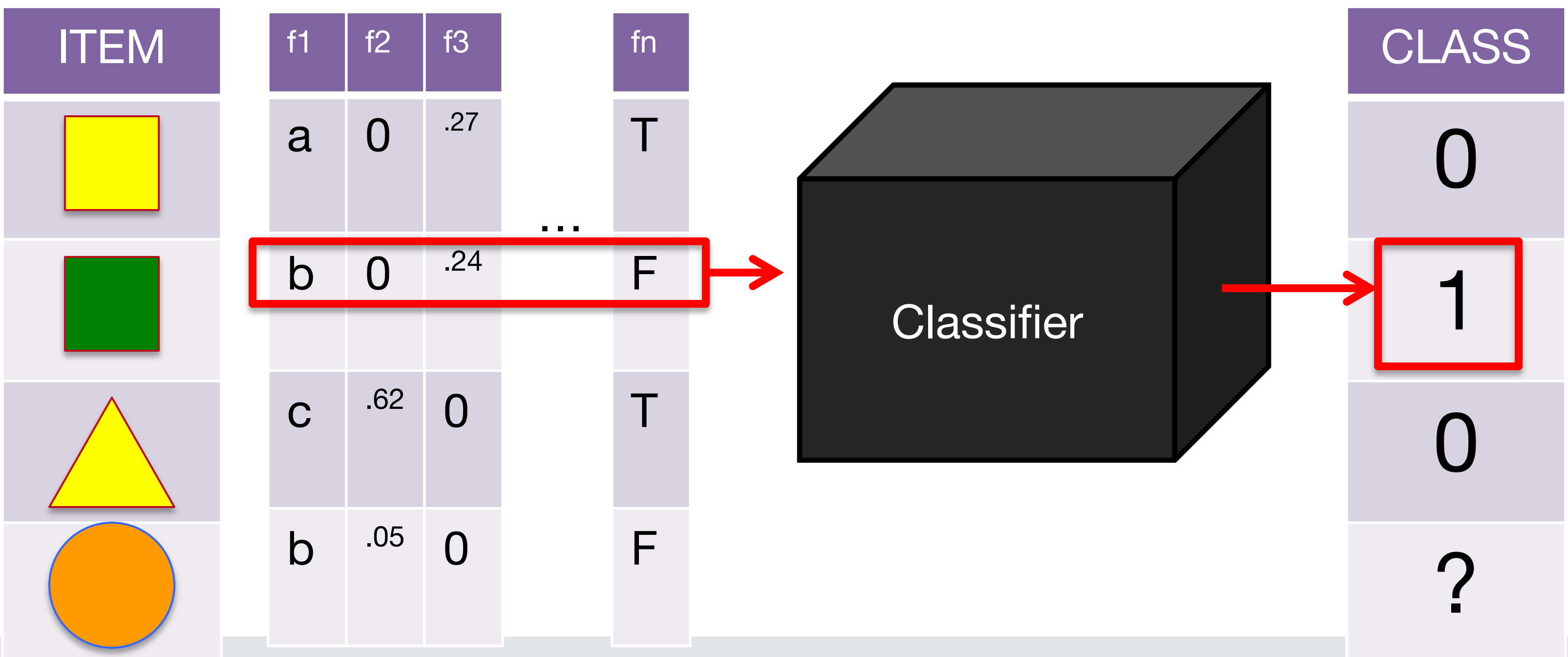
<instance id="11:0@2@wsj/14/wsj_1404@wsj" corpus="wsj">
<answer lemma="double" pos="v" on="1" wn="1" wn-version="2.1"/>
Groupe AG 's chairman said 0 the Belgian insurer is prepared *-1 to give up some of its independence to a white knight if * necessary * to repel a raider . Amid heavy buying of shares in Belgium 's largest insurer , Maurice Lippens also warned in an interview that a white knight , in *-1 buying out a raider , could leave speculators with big losses on their AG stock . Since the beginning of the year , the stock has nearly <head> doubled </head> , * giving AG a market value of about 105 billion Belgian francs -LRB- $ 2.7 billion *U* -RRB- . The most likely white knight would be Societe Generale de Belgique S.A. , which *T*-2 already owns 18 % of AG and which *T*-3 itself is controlled *-1 by Cie . Financiere de Suez , the acquisitive French financial conglomerate . But Mr. Lippens said 0 a rescue also could involve Asahi Mutual Life Insurance Co. , which *T*-1 owns 5 % of AG .
</instance>

<instance id="8:0@37@wsj/14/wsj_1432@wsj" corpus="wsj">
<answer lemma="double" pos="v" on="1" wn="1" wn-version="2.1"/>
We 'll coordinate on this end to places like Bangkok , Singapore and Manila . '' Asian traffic , which *T*-1 currently accounts for 65 % of Cathay 's business , is expected *-2 to continue as the carrier 's mainstay . Cathay has long stated its desire * to <head> double </head> its weekly flights into China to 14 , and it is applying *-1 to restart long-canceled flights into Vietnam . Further expansion into southern Europe is also possible , says 0 *T*-1 Mr. Bell , the spokesman . While a large number of Hong Kong companies have reincorporated offshore ahead of 1997 , such a move is n't an option for Cathay because it would jeopardize its landing rights in Hong Kong.</instance>

- Training and test data: can use a supervised system in a Machine learning setup
- Can also use knowledge-based systems
- Supervised systems show about 5-7% better performance in evaluations
- Performance: about 87% (coarse-grained)
- variations
  - ML learning algorithm
  - features computed on the context
  - features computed through analysis of large background corpora

Supervised Learning of a Classifier

- Classification: assigning (predefined) classes to items
- Supervised learning:
  - have a training set with items and their classes
  - train a ML classifier
  - can use the classifier on unseen items



Local Context Features for Word Sense Disambiguation

Standard features for WSD:

- word window
- lemma/baseform/stem window
- morphological information, e.g. gender, number, tense
- open class words in proximity, e.g. closest adjectives to target
- POS of target and context
- syntactic relations, e.g. headwords

Knowledge-based features (in hybrid systems):

- WordNet similarity with context - "Lesk"
- WordNet hypernym chains

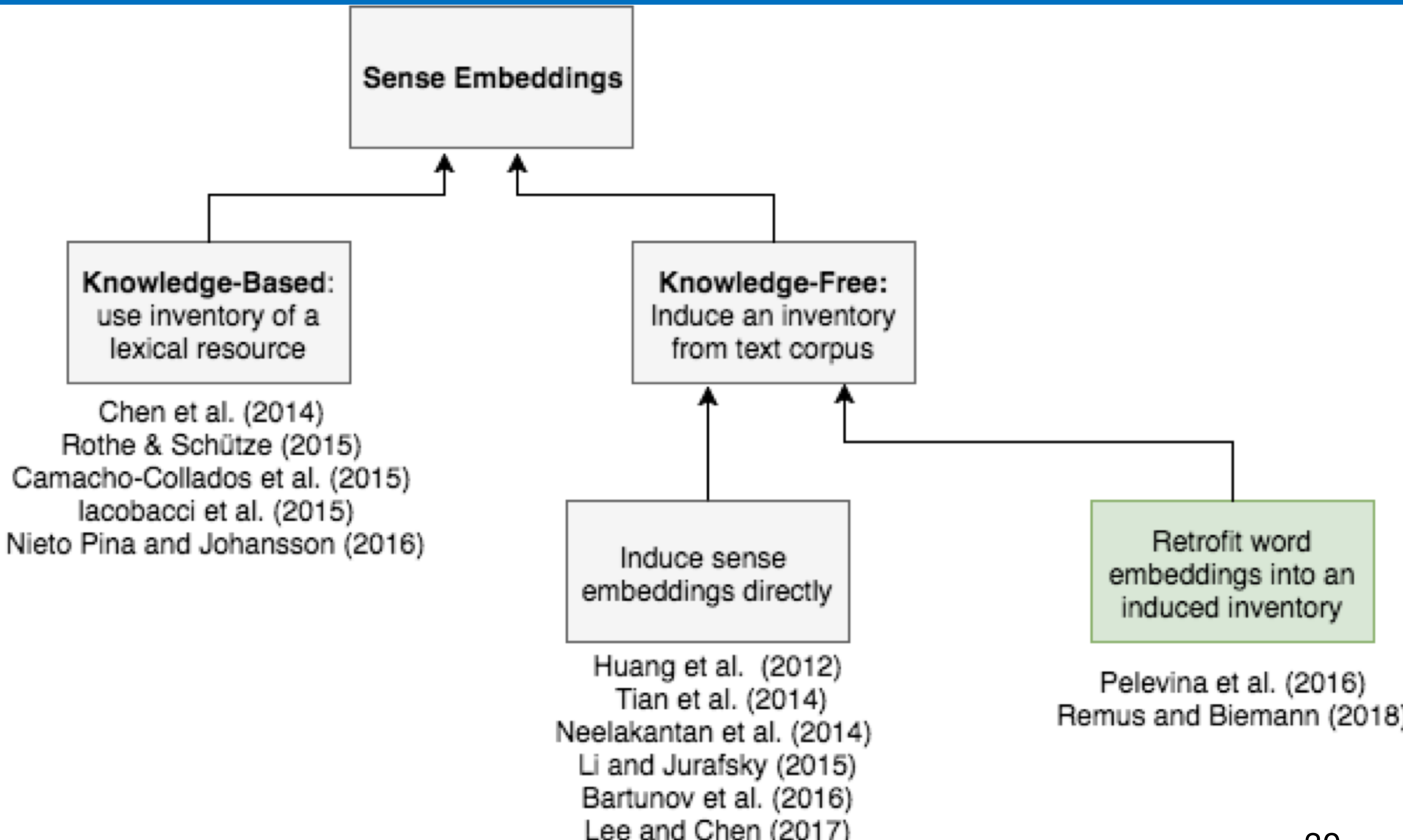
PLAN OF THE LECTURE

- Lexical Ambiguity
- Word Sense Disambiguation
- Embeddings for Senses and Hypernyms
- Lexical Substitution

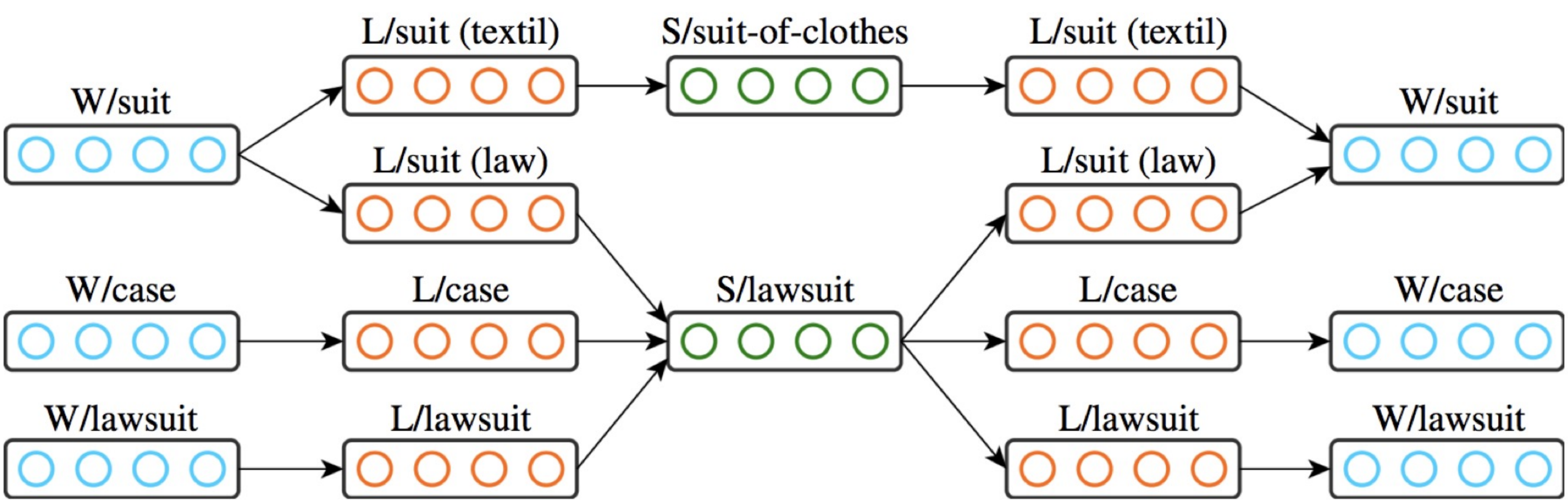
Word vs sense embeddings

Word vs sense embeddings

Sense embedding: various methods were proposed



AutoExtend: a knowledge-based model using WordNet



Source: Rothe, S., & Schuetze, H. (2015). Autoextend: Extending word embeddings to embeddings for synsets and lexemes. In EMNLP.

Multi-Sense Skip-gram: Neelakantan et al. (2015) model

- Step 1: The vector representation of the context is the **average of its context words**' vectors.
- Step 2: For every word type, **maintain clusters of its contexts**.
- Step 3: The sense of a word token is predicted as the cluster that is **closest to its context representation**.
- Step 4: After predicting the sense of a word token, perform a **gradient update on the embedding of that sense**.

- Note: Sense discrimination and learning embeddings are performed **jointly.**

Source: https://arxiv.org/pdf/1504.06654.pdf

Multi-Sense Skip-gram: Neelakantan et al. (2015) model

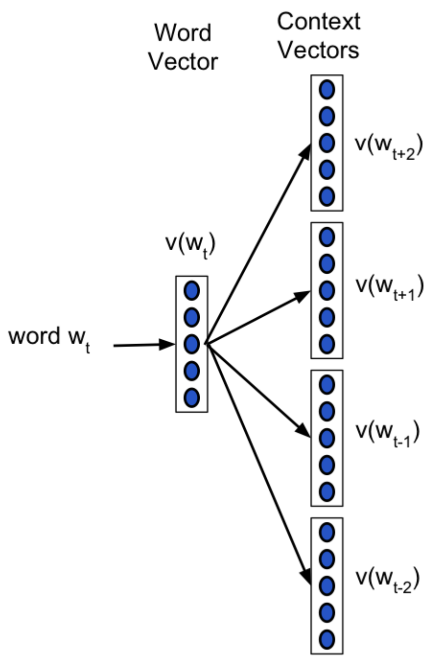


Figure 1: Architecture of the Skip-gram model with window size $R_t = 2$. Context c_t of word w_t consists of $w_{t-1}, w_{t-2}, w_{t+1}, w_{t+2}$.

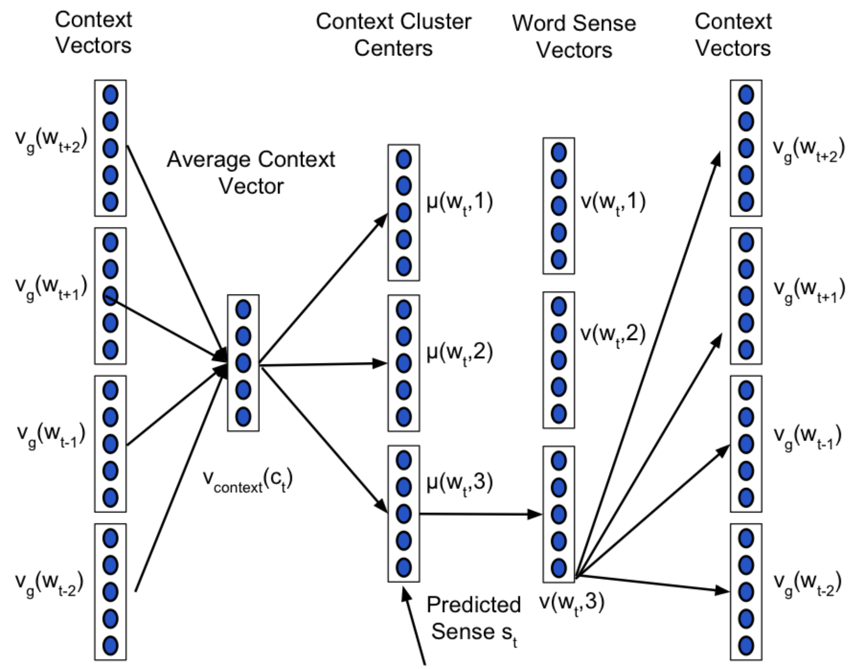


Figure 2: Architecture of Multi-Sense Skip-gram (MSSG) model with window size $R_t = 2$ and $K = 3$. Context c_t of word w_t consists of $w_{t-1}, w_{t-2}, w_{t+1}, w_{t+2}$. The sense is predicted by finding the cluster center of the context that is closest to the average of the context vectors.

Non-Parametric Multi-Sense Skip-gram: Neelakantan et al. (2015)

- Create a new cluster (sense) for a word type with probability proportional to the distance of its context to the nearest cluster (sense).
- The number of senses for a word is unknown and is learned during training.
- New context cluster and a sense vector are created online during training
  - when the word is observed with a context were the similarity between the vector representation of the context with every existing cluster center of the word is less than λ
  - λ is a hyperparameter

Non-Parametric Multi-Sense Skip-gram: Neelakantan et al. (2015)

- Nearest Neighbors of the word plant for different models:

Skip-gram	plants, flowering, weed, fungus, biomass
MS -SG	plants, tubers, soil, seed, biomass refinery, reactor, coal-fired, factory, smelter asteraceae, fabaceae, arecaceae, lamiaceae, ericaceae
NP MS -SG	plants, seeds, pollen, fungal, fungus factory, manufacturing, refinery, bottling, steel fabaceae, legume, asteraceae, apiaceae, flowering power, coal-fired, hydro-power, hydroelectric, refinery

Nearest neighbors of each sense of each word by cosine similarity

APPLE

Skip-gram	blackberry, macintosh, acorn, pear, plum
MSSG	pear, honey, pumpkin, potato, nut microsoft, activision, sony, retail, gamestop macintosh, pc, ibm, iigs, chipsets
NP-MSSG	apricot, blackberry, cabbage, blackberries, pear microsoft, ibm, wordperfect, amiga, trs-80

FOX

Skip-gram	abc, nbc, soapnet, espn, kttv
MSSG	beaver, wolf, moose, otter, swan nbc, espn, cbs, ctv, pbs dexter, myers, sawyer, kelly, griffith
NP-MSSG	rabbit, squirrel, wolf, badger, stoat cbs,abc, nbc, wnyw, abc-tv

NET

Skip-gram	profit, dividends, pegged, profits, nets
MSSG	snap, sideline, ball, game-trying, scoring negative, offset, constant, hence, potential pre-tax, billion, revenue, annualized, us$
NP-MSSG	negative, total, transfer, minimizes, loop pre-tax, taxable, per, billion, us$, income ball, yard, fouled, bounced, 50-yard wnet, tvontorio, cable, tv, tv-5

ROCK

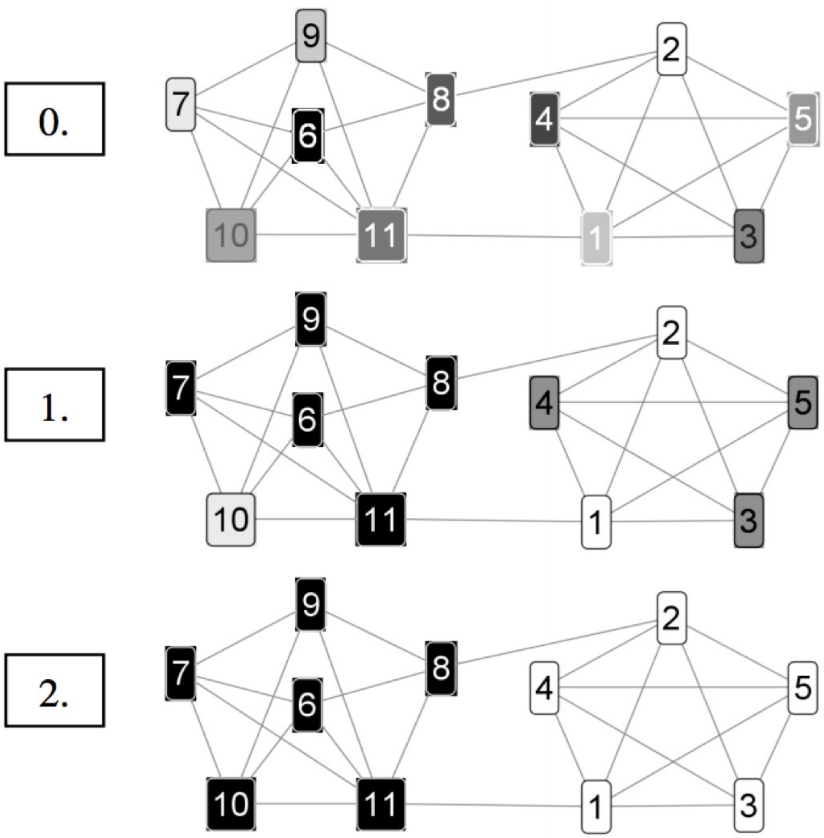
Skip-gram	glam, indie, punk, band, pop
MSSG	rocks, basalt, boulders, sand, quartzite alternative, progressive, roll, indie, blues-rock rocks, pine, rocky, butte, deer
NP-MSSG	granite, basalt, outcropping, rocks, quartzite alternative, indie, pop/rock, rock/metal, blues-rock

RUN

Skip-gram	running, ran, runs, afoul, amok
MSSG	running, stretch, ran, pinch-hit, runs operated , running, runs, operate, managed running, runs, operate, drivers, configure
NP-MSSG	two-run, walk-off, runs, three-runs, starts operated, runs, serviced, links, walk running, operating, ran, go, configure re-election, reelection, re-elect, unseat, term-limited helmed, longest-running, mtv, promoted, produced

SenseGram: from pre-trained word embeddings to sense embeddings

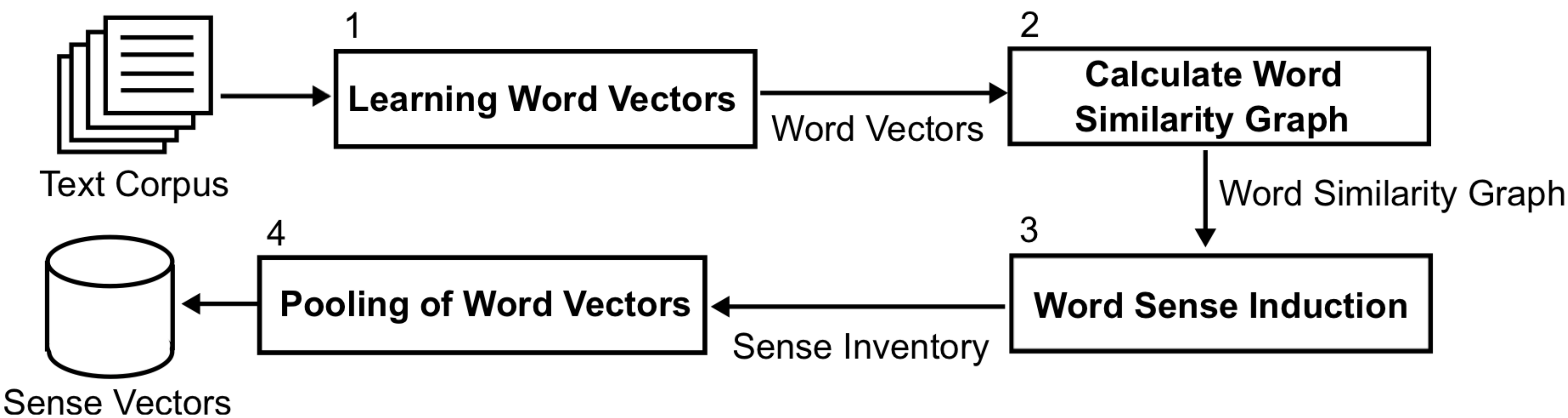
- Graph clustering
  - Chinese Whispers
  - (Biemann, 2006)




```
initialize:
 forall vi in V: class(vi)=i;
while changes:
 forall v in V, randomized order:
   class(v)=highest ranked class
            in neighborhood of v;
```

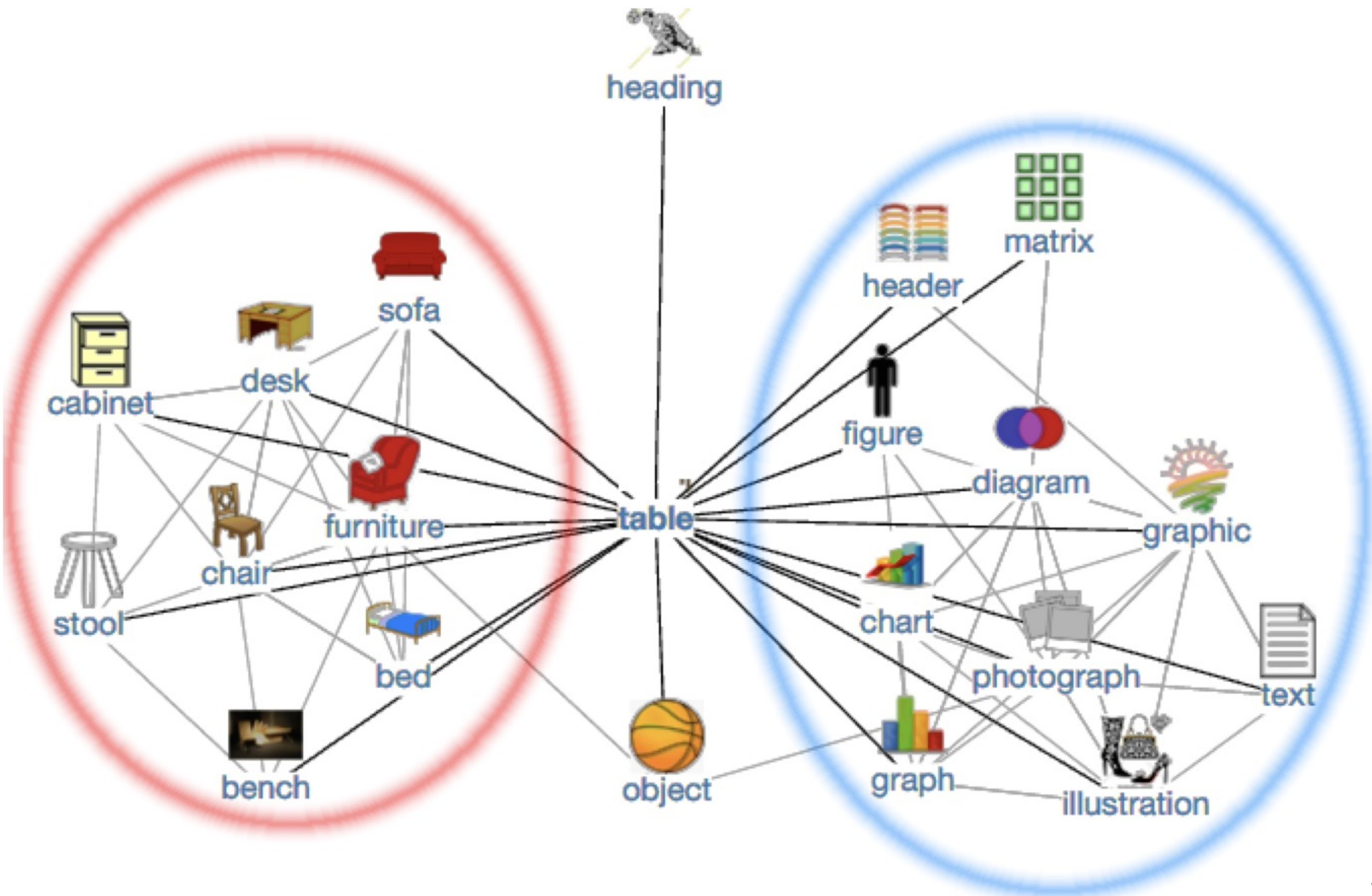
SenseGram: from pre-trained word embeddings to sense embeddings

- Sense embeddings using retrofitting:

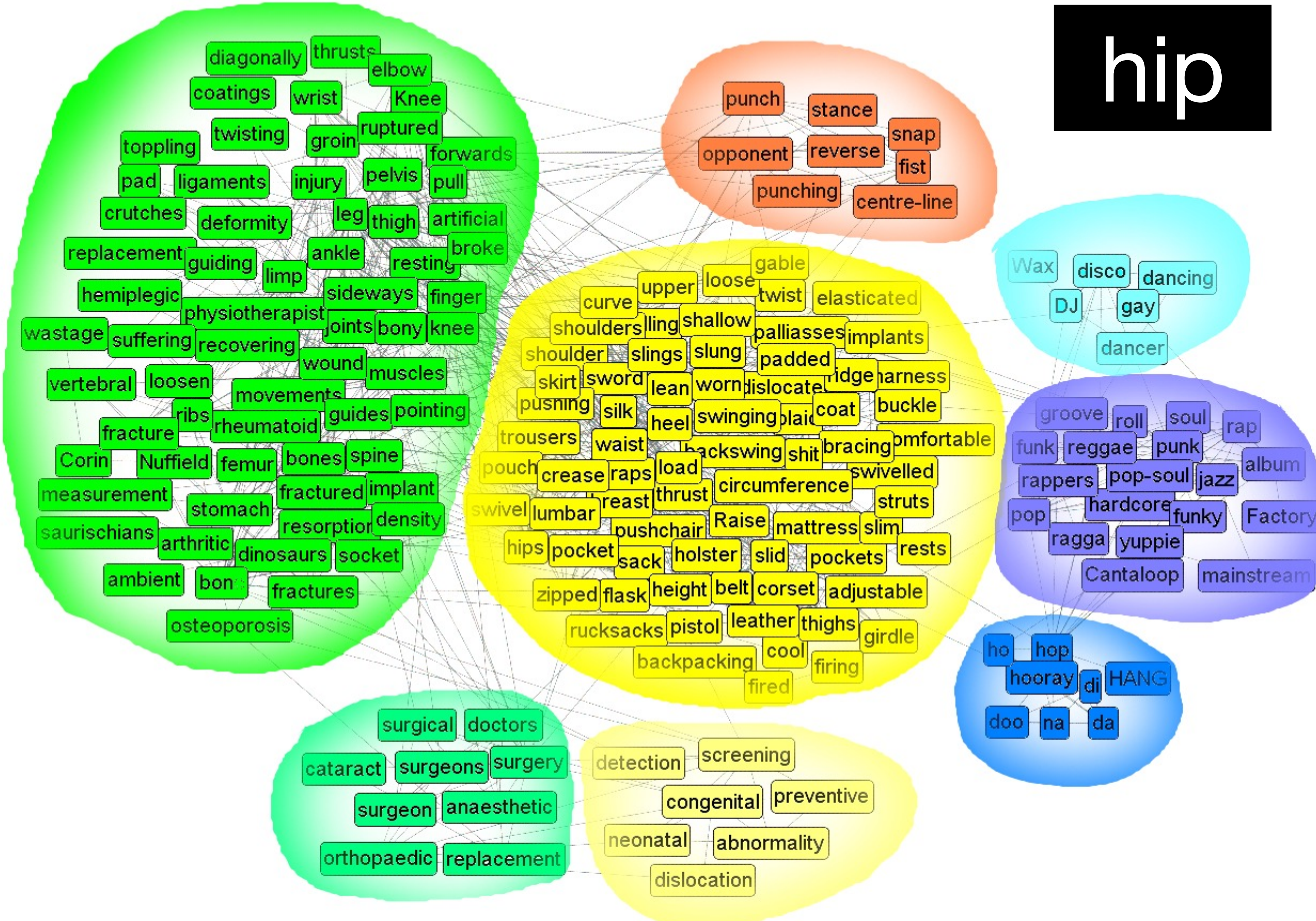


SenseGram: from pre-trained word embeddings to sense embeddings

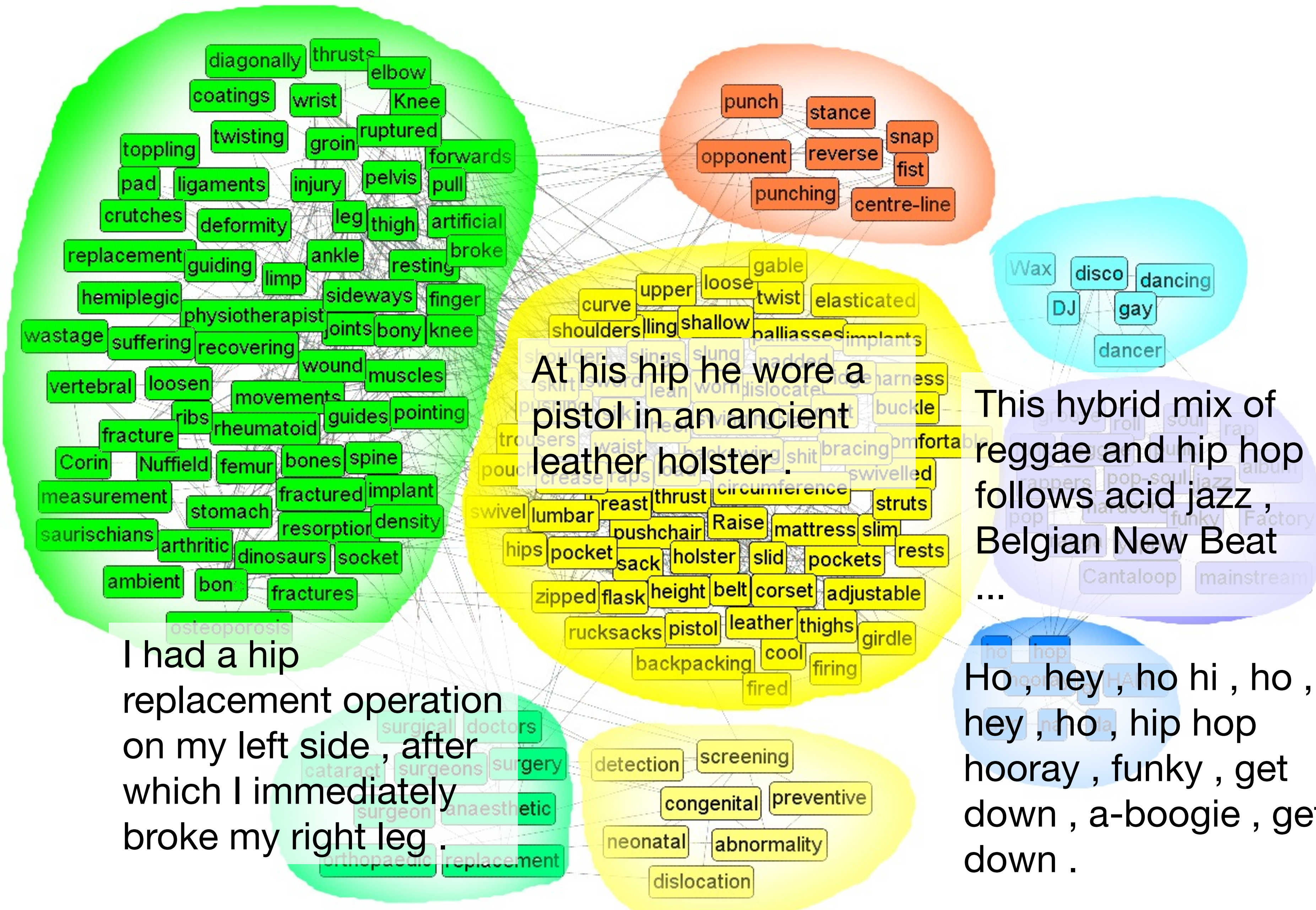
- Sense embeddings using retrofitting:



hip



diagonally
thrusts
elbow
coatings
wrist
Knee
toppling
twisting
groin
ruptured
forwards
pad
ligaments
injury
pelvis
pull
crutches
deformity
leg
thigh
artificial
replacement
guiding
limp
ankle
resting
broke
hemiplegic
sideways
finger
physiotherapist
wastage
suffering
recovering
joints
bony
knee
wound
muscles
vertebral
loosen
movements
ribs
rheumatoid
guides
pointing
fracture
Corin
Nuffield
femur
bones
spine
measurement
stomach
fractured
implant
saurischians
arthritic
resorption
density
dinosaurs
socket
ambient
fractures
osteoporosis
punch
stance
snap
opponent
reverse
fist
punching
centre-line
Wax
disco
dancing
DJ
gay
dancer
gable
upper
loose
twist
elasticated
curve
shoulders
shallow
palliasses
implants
shoulder
slings
slung
padded
skirt
sword
lean
worn
harness
pushing
silk
heel
swinging
coat
buckle
trousers
waist
shit
bracing
pouch
crease
raps
load
circumference
swivelled
thrust
struts
swivel
lumbar
pushchair
Raise
mattress
slim
hips
pocket
sack
holster
slid
pockets
rests
zipped
flask
height
belt
corset
adjustable
rucksacks
pistol
leather
thighs
girdle
backpacking
cool
firing
fired
groove
roll
soul
rap
funk
reggae
punk
album
rappers
pop-soul
jazz
pop
hardcore
funky
Factory
ragga
yuppie
Cantaloop
mainstream
ho
hop
hooray
di
HANG
doo
na
da
surgical
doctors
cataract
surgeons
surgery
surgeon
anaesthetic
orthopaedic
replacement
detection
screening
congenital
preventive
neonatal
abnormality
dislocation



At his hip he wore a pistol in an ancient leather holster .

This hybrid mix of reggae and hip hop follows acid jazz , Belgian New Beat ...

I had a hip replacement operation on my left side , after which I immediately broke my right leg .

Ho , hey , ho hi , ho , hey , ho , hip hop hooray , funky , get down , a-boogie , get down .

SenseGram: from pre-trained word embeddings to sense embeddings

- Neighbors of word and sense vectors:

Vector	**Nearest Neighbors**
table	tray, bottom, diagram, bucket, brackets, stack, basket, list, parenthesis, cup, saucer, pile, playfield, bracket, pot, drop-down, cue, plate
table#0	leftmost#0, column#1, tableau#1, indent#1, bracket#3, pointer#0, footer#1, cursor#1, diagram#0, grid#0
table#1	pile#1, stool#1, tray#0, basket#0, bowl#1, bucket#0, box#0, cage#0, saucer#3, mirror#1, pan#1, lid#0

JoBimText: Distributionally similar words to ‘mouse’

http://ltmaggie.informatik.uni-hamburg.de/jobimviz/

Jos

Jo	Score
mouse#NN	746
rat#NN	192
rodent#NN	122
monkey#NN	112
pig#NN	103
animal#NN	95
human#NN	94
rabbit#NN	91
keyboard#NN	91
cow#NN	83
hamster#NN	82
frog#NN	81
cat#NN	80
bird#NN	79

Bims

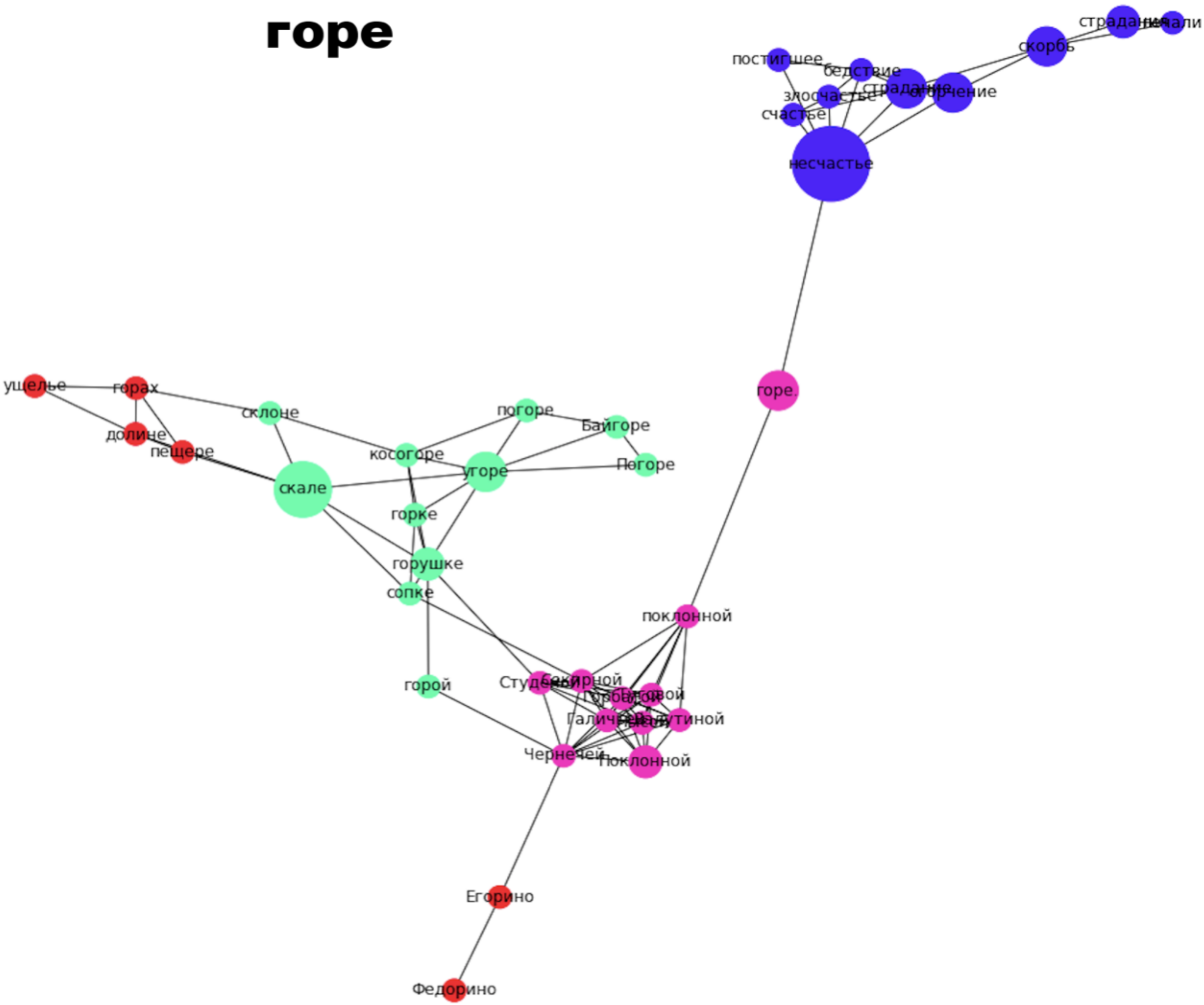
Bim	Score	Count
click#NN#-prep_of	14433.61	
a#DT#det	11612.08	
click#NN#-nn	9071.84	
the#DT#det	8613.77	
keyboard#NN#-conj_and	7548.80	
cat#NN#-conj_and	5417.09	
computer#NN#nn	4776.27	
keyboard#NN#conj...	4241.33	
button#NN#-nn	3987.05	
pad#NN#-nn	3320.76	
rat#NN#conj_and	2971.02	
rat#NN#-conj_and	2821.09	
click#VB#-dobj	2472.84	

CW

Sense 0 168: rat#NN · rodent#NN ...

Sense 1 32: keyboard#NN · joysti...

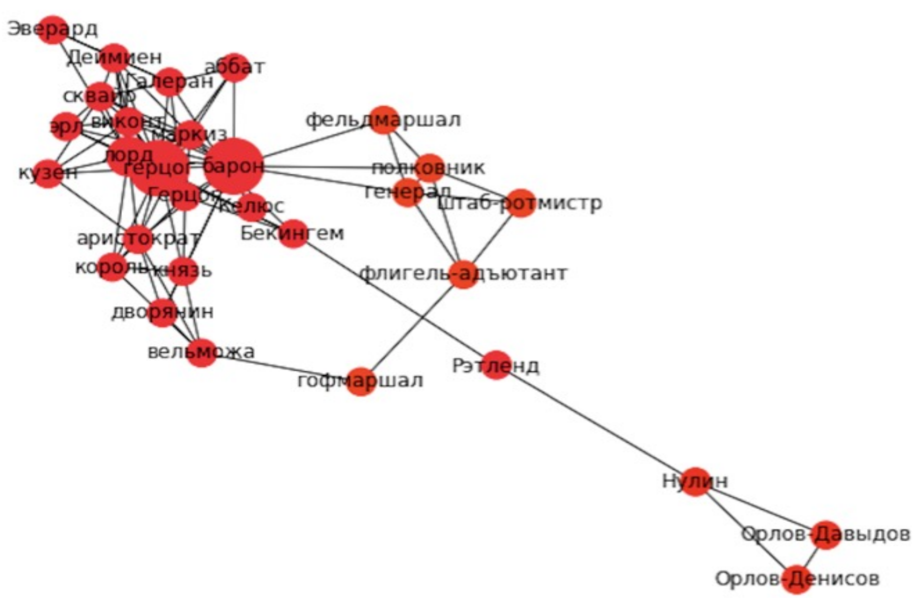
горе


печали
скорбь
постигшее
бедствие
огорчение
счастье
несчастье
горе.
ущелье
горах
склоне
погоре
Байгоре
долине
пещере
косогоре
угоре
Погоре
скале
горке
горушке
сопке
поклонной
горой
Поклонной
Егорино
Федорино

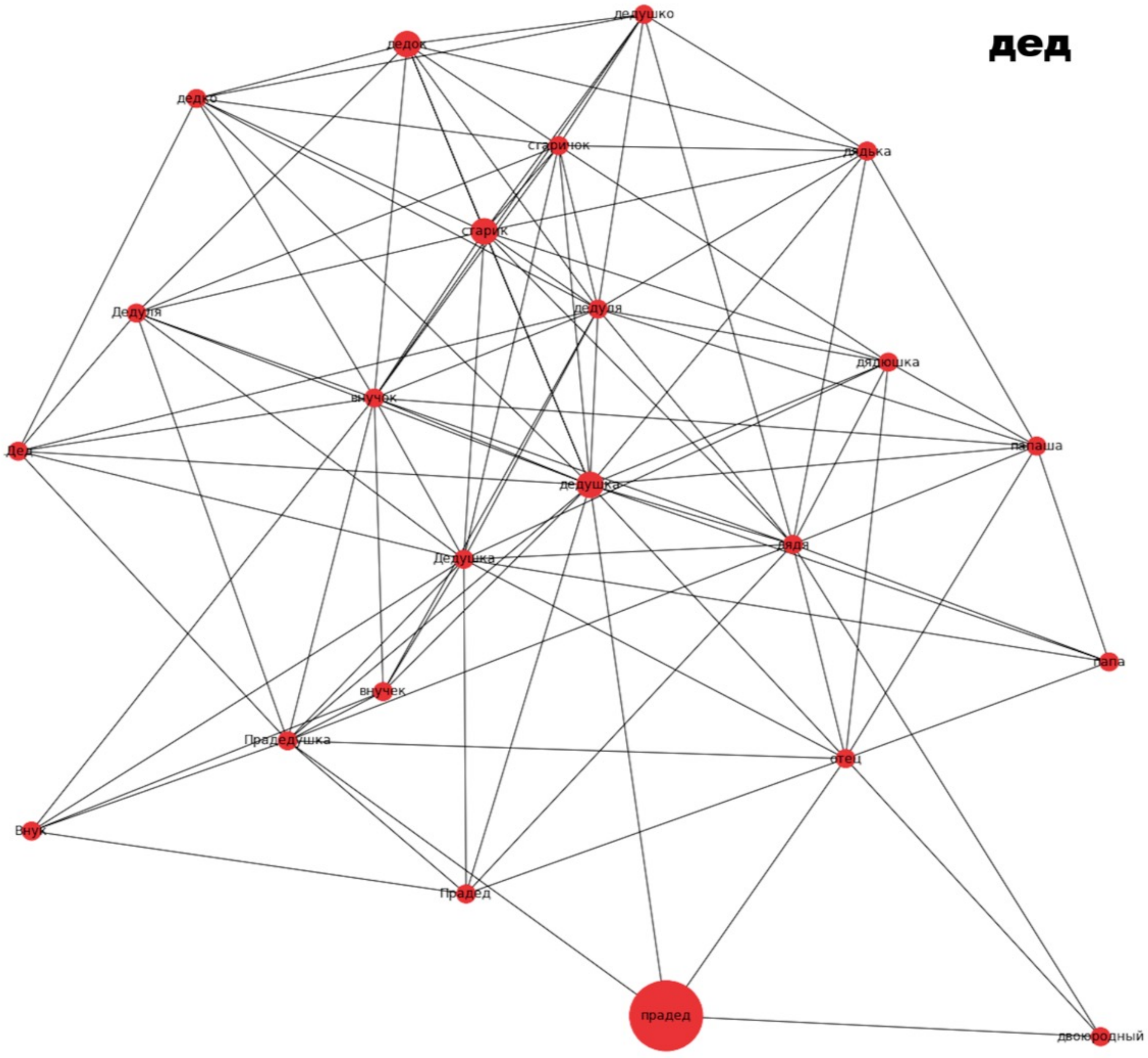
граф


графы
Графы
графам
графами
графах
графа
графини


подграф
орграф

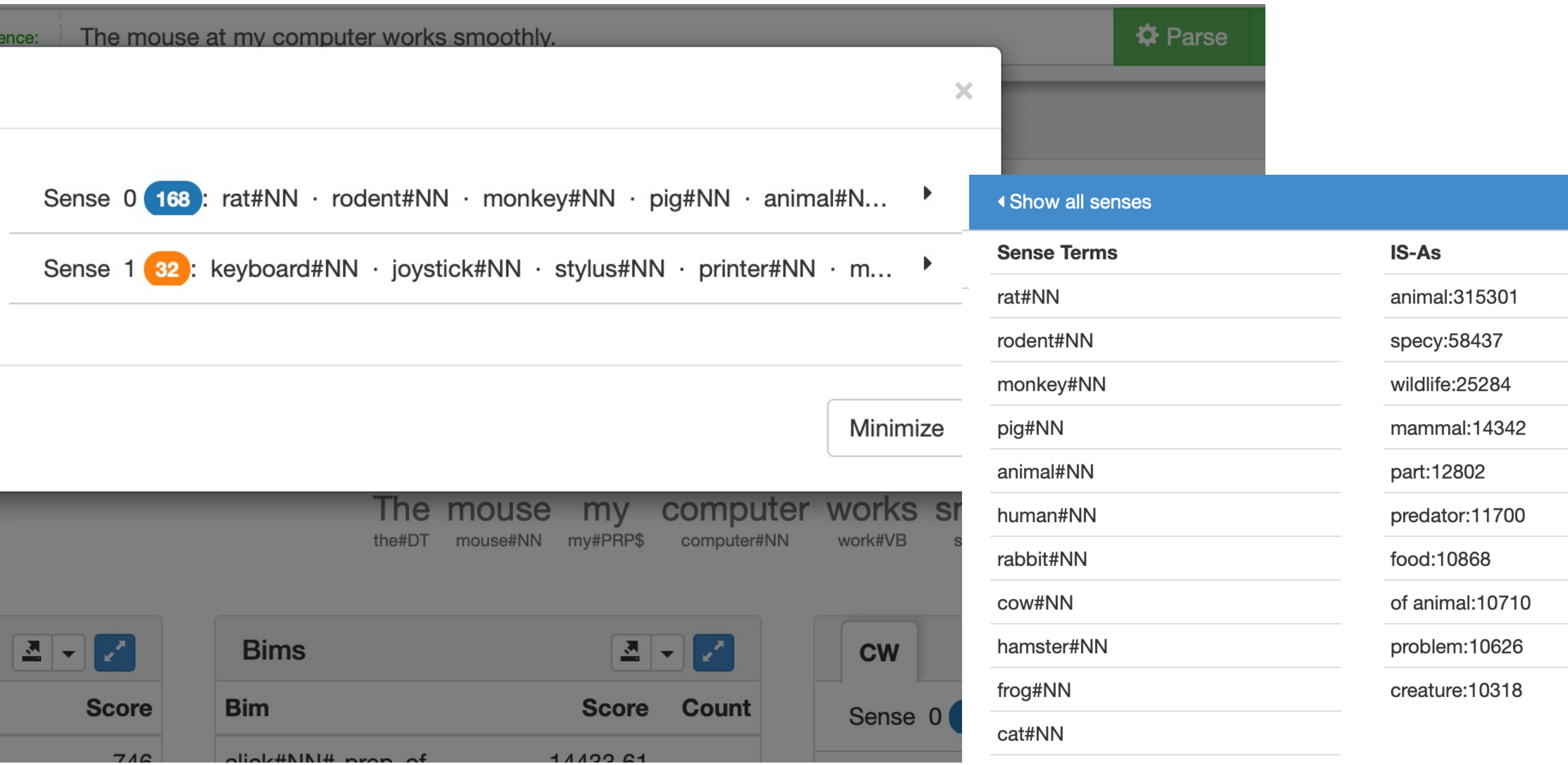

Эверард
Деймиен
аббат
Галеран
сквайр
эрл
виконт
маркиз
лорд
кузен
герцог
барон
Герцог
Келлос
Бекингем
аристократ
король
князь
дворянин
вельможа
фельдмаршал
полковник
генерал
штаб-ротмистр
флигель-адъютант
гофмаршал
Рэтленд
Нулин
Орлов-Давыдов
Орлов-Денисов

дед

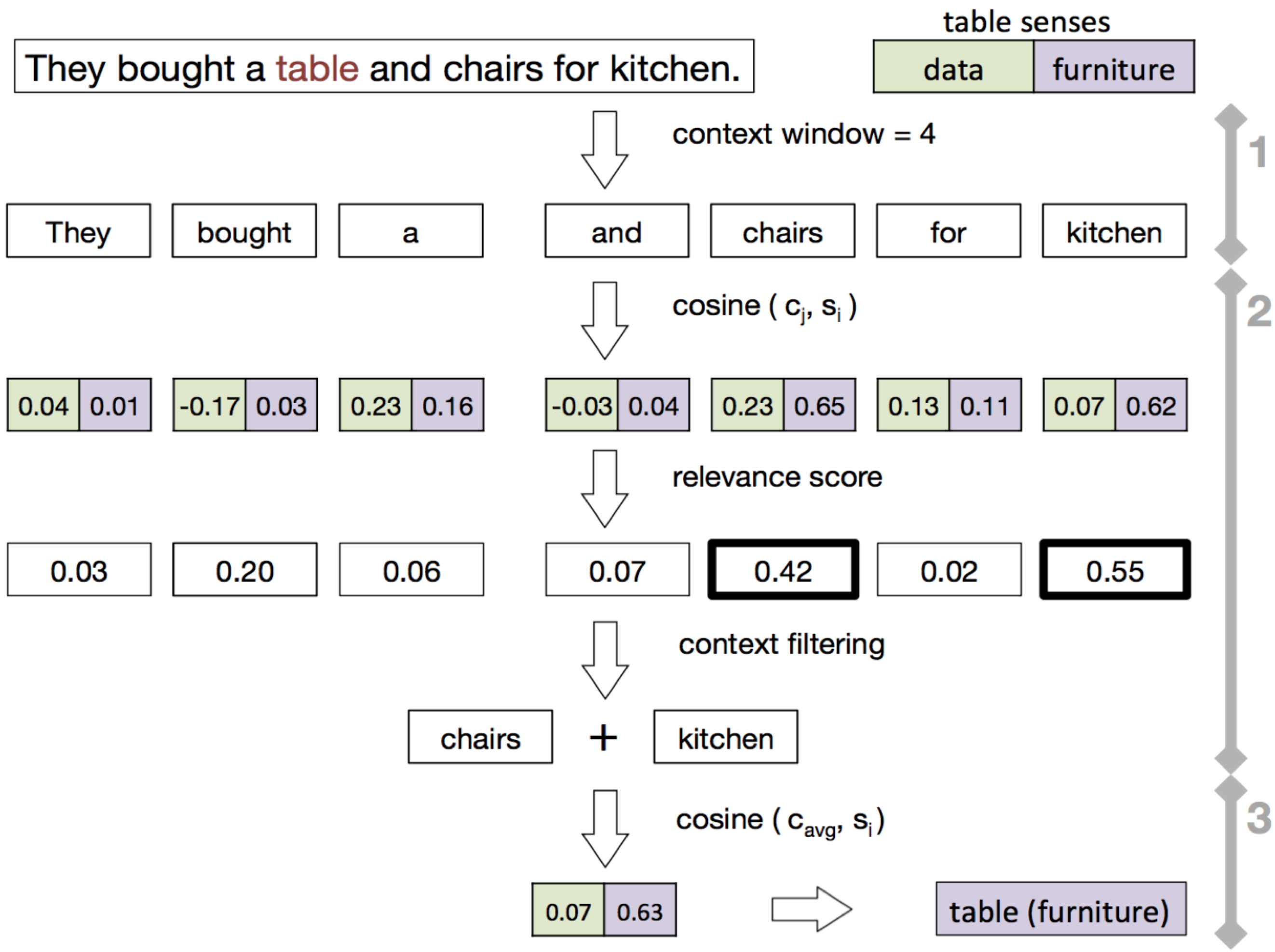

дедушко
дедок
дедко
старичок
дядька
старик
Дедуля
дедуля
дядюшка
внучок
Дед
папаша
дедушка
Дедушка
дядя
папа
внучек
Прадедушка
отец
Внук
Прадед
прадед
двоюродный

JoBimText: online demo

http://ltmaggie.informatik.uni-hamburg.de/jobimviz/



SenseGram: word sense disambiguation



Application of sense representations: humor detection and generation?

Affine transformation for prediction of hypernymy relations (Fu et al., 2014)

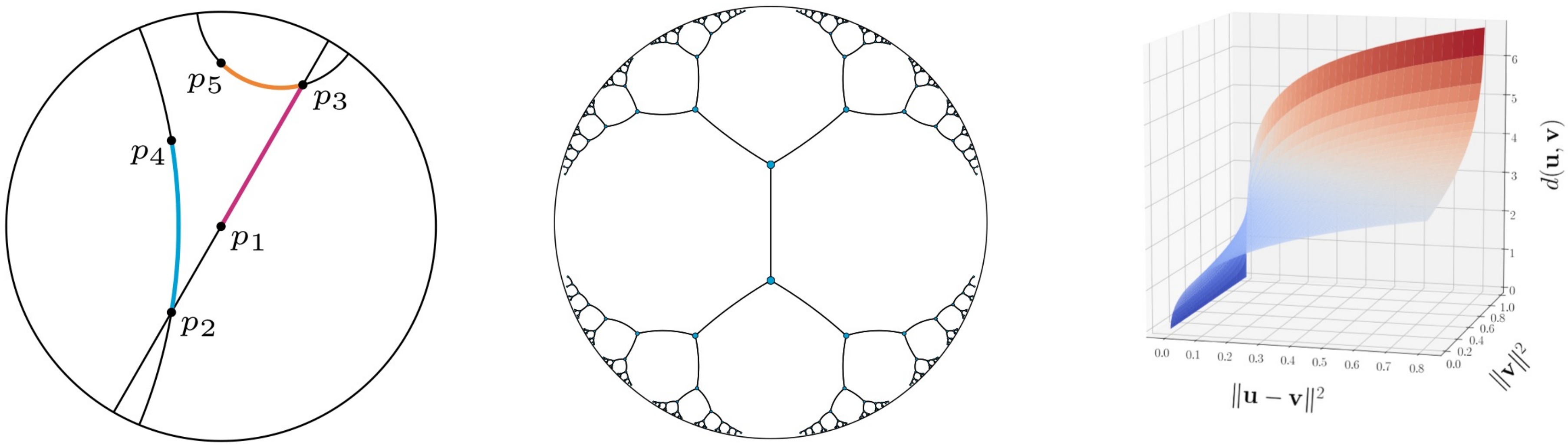
- Hypernyms: cat → animal, dog → animal, banana → fruit, apple → fruit, …
- Learn a linear projection from more specific word (hyponym) to more generic word (hypernym) using:

$$\mathbf{\Phi}^* = \arg\min_{\mathbf{\Phi}} \frac{1}{|\mathcal{P}|} \sum_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{P}} \|\boldsymbol{x}\mathbf{\Phi} - \boldsymbol{y}\|^2$$

- *P* – is a set of training hyponym-hypernym pairs

Hyperbolic (Poincaré) embeddings

Source: https://arxiv.org/pdf/1705.08039.pdf



(a) Geodesics of the Poincaré disk (b) Embedding of a tree in $\mathcal{B}^2$ (c) Growth of Poincaré distance

Figure 1: (a) Due to the negative curvature of $\mathcal{B}$, the distance of points increases exponentially (relative to their Euclidean distance) the closer they are to the boundary. (c) Growth of the Poincaré distance $d(\boldsymbol{u}, \boldsymbol{v})$ relative to the Euclidean distance and the norm of $\boldsymbol{v}$ (for fixed $\|\boldsymbol{u}\| = 0.9$). (b) Embedding of a regular tree in $\mathcal{B}^2$ such that all connected nodes are spaced equally far apart (i.e., all black line segments have identical hyperbolic length).

Hyperbolic (Poincaré) embeddings

Source: https://arxiv.org/pdf/1705.08039.pdf

$$\mathcal{B}^d = \{\boldsymbol{x} \in \mathbb{R}^d \mid \|\boldsymbol{x}\| < 1\}$$

- Distance on a ball between two points:

$$d(\boldsymbol{u}, \boldsymbol{v}) = \operatorname{arcosh}\left(1 + 2\frac{\|\boldsymbol{u} - \boldsymbol{v}\|^2}{(1 - \|\boldsymbol{u}\|^2)(1 - \|\boldsymbol{v}\|^2)}\right)$$

- Loss

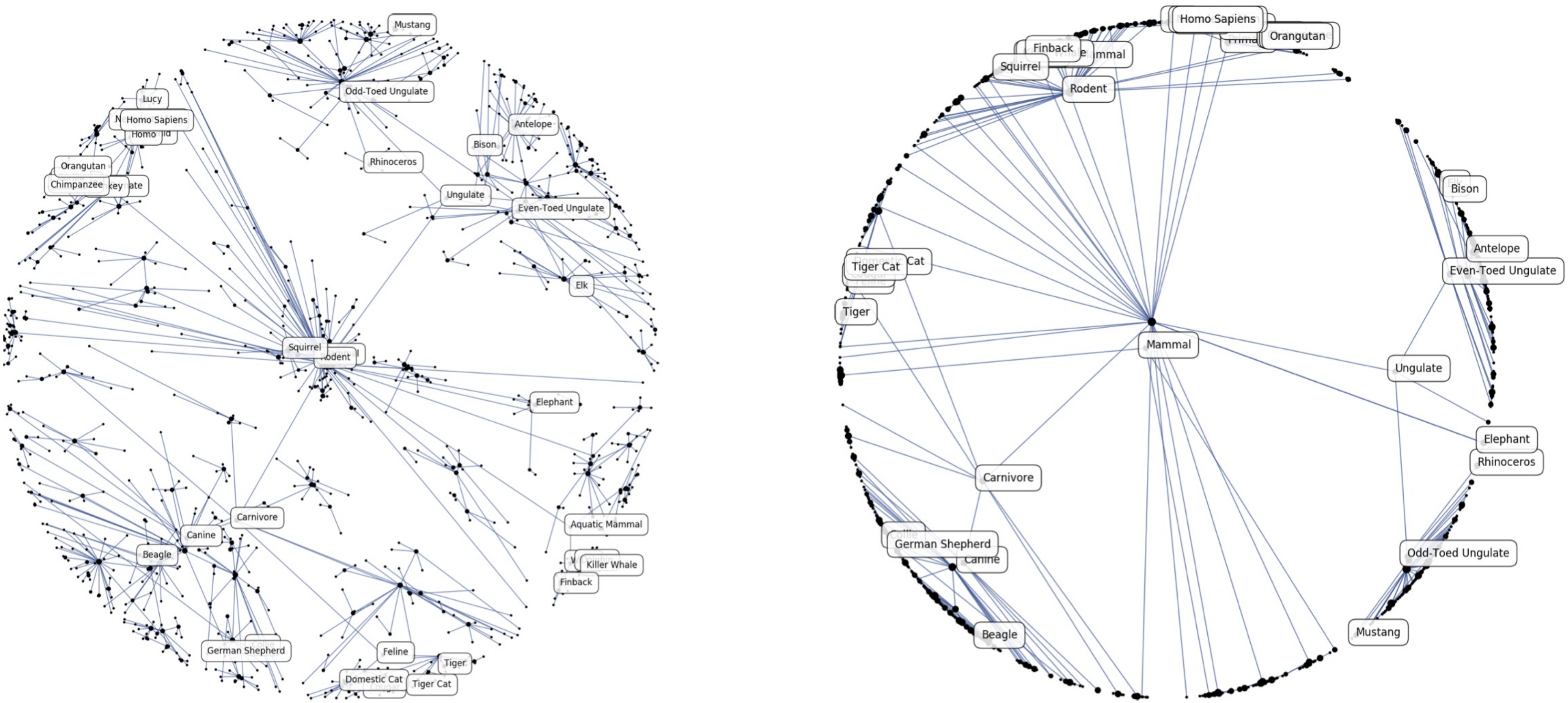
$$\mathcal{L}(\Theta) = \sum_{(u,v)\in\mathcal{D}} \log \frac{e^{-d(\boldsymbol{u},\boldsymbol{v})}}{\sum_{\boldsymbol{v}'\in\mathcal{N}(u)} e^{-d(\boldsymbol{u},\boldsymbol{v}')}}$$

  - set of negative examples for *u:* $\mathcal{N}(u) = \{v \mid (u,v) \notin \mathcal{D}\} \cup \{u\}$
  - 10 negative samples per 1 positive

$$\Theta' \leftarrow \arg\min_{\Theta} \mathcal{L}(\Theta) \qquad \text{s.t. } \forall\, \boldsymbol{\theta}_i \in \Theta : \|\boldsymbol{\theta}_i\| < 1$$

Trained on WordNet relations

- Two-dimensional Poincaré embeddings of transitive closure of the WordNet mammals subtree.



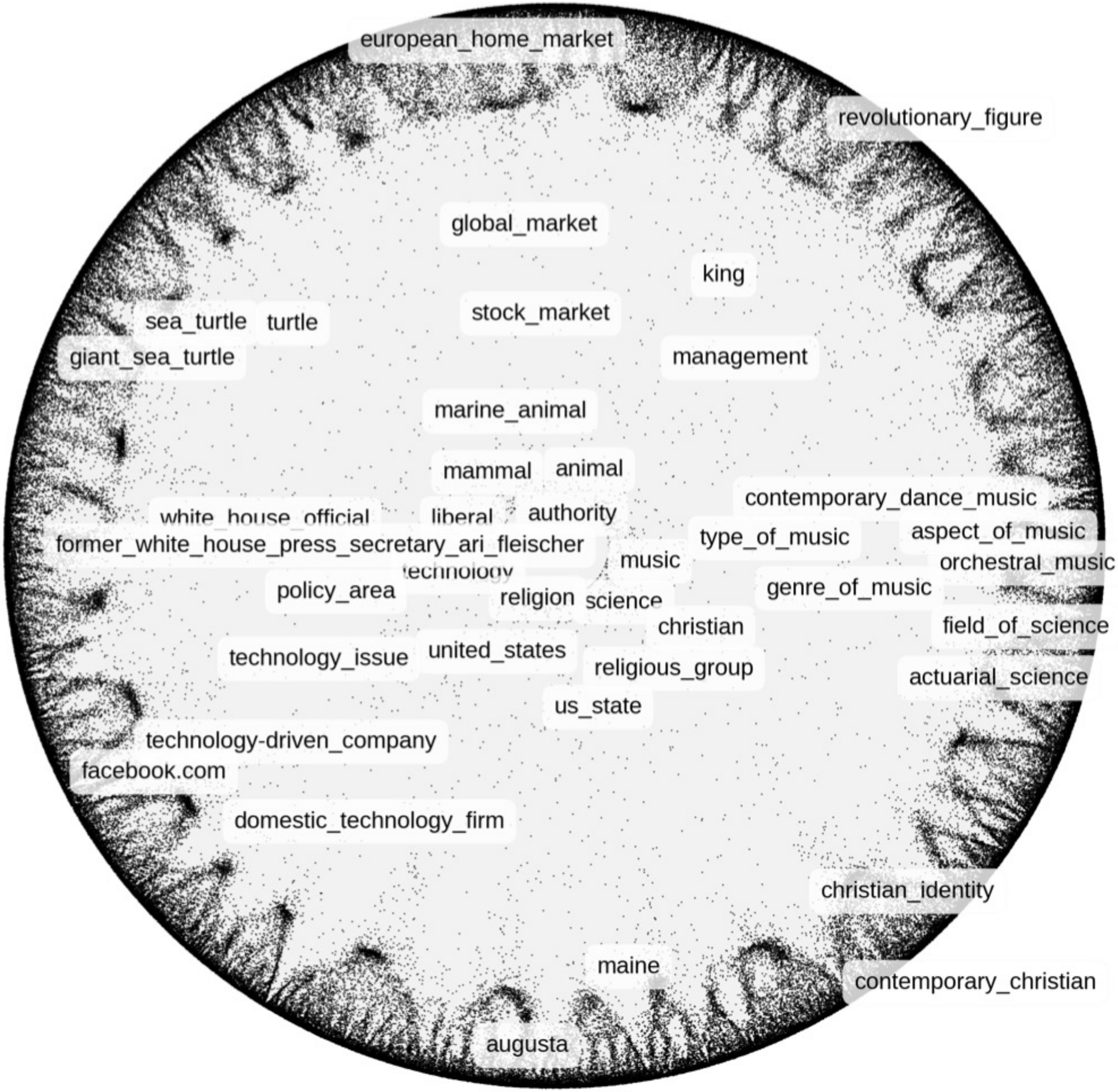
(a) Intermediate embedding after 20 epochs (b) Embedding after convergence

Hyperbolic (Poincaré) embeddings: Hearst patterns

Pattern
X which is a (example | class | kind | ...) of Y
X (and | or) (any | some) other Y
X which is called Y
X is `JJS` (most)? Y
X a special case of Y
X is an Y that
X is a !(member | part | given) Y
!(features | properties) Y such as X_1, X_2, ...
(Unlike | like) (most | all | any | other) Y, X
Y including X_1, X_2, ...

Source: https://arxiv.org/pdf/1902.00913.pdf

S



Source: https://arxiv.org/pdf/1902.00913.pdf

PLAN OF THE LECTURE

- Lexical Ambiguity
- Word Sense Disambiguation
- Embeddings for Senses and Hypernyms
- Lexical Substitution

Lexical Sample Substitution

- Two steps: For a range of target words,
  - Disambiguate word sense
  - Provide ranked list of substitutions per word sense
- Like lexical sample WSD, one classifier per word needs to be trained
- Features: Same as in WSD

Data: e.g. obtained by crowdsourcing.

Find Substitutable Words

In the sentence below, what words or phrases could replace the **bolded** word without changing the meaning?
Please use the singular form, even if the bolded word is plural.

Example:
In most countries **children** are required by law to attend school.

You might enter:
kid
youngster
pupil
young person

Try to enter single words or short phrases like "water bottle" or "post office." You are encouraged to use the target word in short phrases, e.g. "railway line" for "The **line** ends at the Amtrak station".
Avoid descriptive phrases, e.g. "a container you drink out of," or "a place you mail things from" unless you absolutely can't find a better substitution.
Further, tell us how easy or difficult it is to assign one of several possible meanings for the **bolded** word in the sentence.

Your sentence is: After the first two series of Auf Wiedersehen , Pet , Nail found himself typecast and had no more major **breaks** until the detective series , Spender , which he co – wrote .

Enter *one term* per box. You don't need to fill in all the boxes -- only add terms that can substitute for the target word *without changing the meaning*.

Substitution (use singular) :

Substitution (use singular) :

Substitution (use singular) :

Substitution (use singular) :

Substitution (use singular) :

Finding the meaning of the **bolded** word in this sentence is
- EASY: This sentence is a good example for illustrating the meaning of the bolded word
- MEDIUM: I could find the meaning, but this sentence is not great for illustrating it
- HARD: There might be several possible interpretations for the bolded word in this sentence
- IMPOSSIBLE: The bolded word is not a noun or the meaning is impossible to determine

Example Output of Lexical Substitution System

Darmstadt is a <target= “city" lemma= “city" sense= “city" confidence= “1.0" substitutions= “[town, 89] [metropolis, 50] [municipality, 40] [metropolitan area, 17] [urban area, 14] [village, 14] [urban, 13] [community, 12] [megalopolis, 12] [township, 10]"> in the Bundesland (federal <target= “state" lemma= “state" sense= “state@@3" confidence= “0.6666667" substitutions= “[government, 7] [province, 2]">) of Hesse in Germany , located in the southern <target= “part" lemma= “part" sense= “part@@1" confidence= “1.0" substitutions= “[portion, 21] [section, 21] [area, 17] [region, 15] [piece, 14] [component, 13] [segment, 11] [side, 8] [division, 6] [element, 4] [unit, 4]"> of the Rhine Main <target= “Area" lemma= “Area" sense= “area" confidence= “1.0" substitutions= “[region, 65] [zone, 24] [district, 22] [location, 21] [place, 19] [section, 17] [territory, 16] [field, 14] [part, 14] [vicinity, 14]"> .

The sandy <target= “soils" lemma= “soil" sense= “soil@@1" confidence= “1.0" substitutions= “[earth, 26] [dirt, 23] [ground, 8] [loam, 6] [land, 3] [topsoil, 2]"> in the Darmstadt <target= “area" lemma= “area" sense= “area" confidence= “1.0" substitutions= “[region, 65] [zone, 24] [district, 22] [location, 21] [place, 19] [section, 17] [territory, 16] [field, 14] [part, 14] [vicinity, 14]"> , ill-suited for agriculture in <target= “times" lemma= “time" sense= “time@@1" confidence= “0.5" substitutions= “[instance, 99] [occasion, 95] [period, 82] [moment, 60] [era, 50] [age, 24] [event, 23] [point, 22] [occurrence, 17] [duration, 16]"> before industrial fertilisation , [2] prevented any larger <target= “settlement" lemma= “settlement" sense= “settlement@@1" confidence= “1.0" substitutions= “[colony, 21] [community, 19] [village, 12] [town, 8] [hamlet, 6] [establishment, 5] [habitation, 5]"> from developing , until the <target= “city" lemma= “city" sense= “city" confidence= “1.0" substitutions= “[town, 89] [metropolis, 50] [municipality, 40] [metropolitan area, 17] [urban area, 14] [village, 14] [urban, 13] [community, 12] [megalopolis, 12] [township, 10]"> became the <target= “seat" lemma= “seat" sense= “seat@@1" confidence= “1.0" substitutions= “[position, 21] [post, 19] [place, 10] [spot, 10] [elected post, 7] [station, 5] [rank, 3] [chair position, 2] [seat of government, 2]"> of the Landgraves of Hessen-Darmstadt in the 16th <target= “century" lemma= “century" sense= “century" confidence= “1.0" substitutions= “[era, 13] [hundred, 9] [hundred year period, 9] [century period, 8] [epoch, 8] [period, 8] [age, 7] [generation, 5] [100 year period, 4] [100 years, 4] [years, 4]">

Download: https://www.inf.uni-hamburg.de/en/inst/ab/lt/resources/software/twsi-substituter.html

Supervised All-Words Lexical Substitution

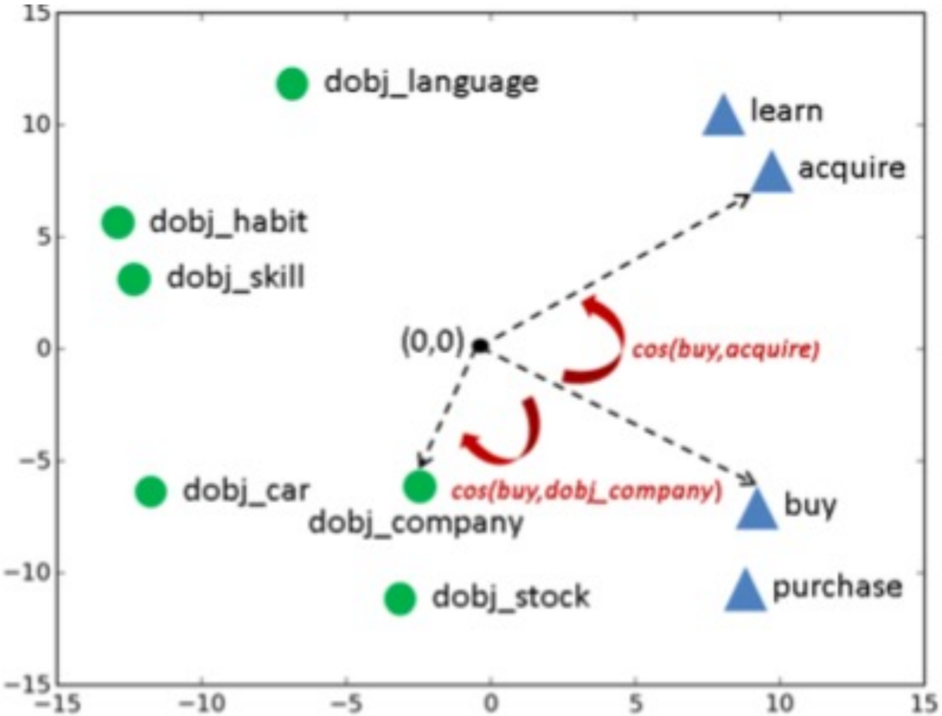
Motivation for supervised all-words approach:

- Supervised WSD/LexSub is more accurate than unsupervised approaches
  - but needs enough training to train one model per target
  - We'd like to be able to substitute all words, and leverage supervision
- We want one single model that generalizes over all words
  - cannot use features that directly use target or context words

→ We can only construct a supervised all-word system if we have delexicalized features.

Syntactic Embedding Features (Melamud et al., 2015)

A good lexical substitute for a target word instance, under a given context, needs to be both:
- semantically similar to the target word and
- compatible with the given context



Include as features:
- target-substitute similarity
- context-substitute similarity
- unsupervised combinations (Add, Mul, BalAdd, BalMul)

Melamud et al. A simple word embedding model for lexical substitution. VSM Workshop. Denver, CO, USA.

Syntactic Embedding Features (Melamud et al., 2015)

A good lexical substitute for a target word instance, under a given context, needs to be both:
- semantically similar to the target word and
- compatible with the given context

Add	$\frac{cos(s,t)+\sum_{c\in C} cos(s,c)}{\|C\|+1}$
BalAdd	$\frac{\|C\|\cdot cos(s,t)+\sum_{c\in C} cos(s,c)}{2\cdot\|C\|}$
Mult	$\sqrt[\|C\|+1]{pcos(s,t)\cdot\prod_{c\in C} pcos(s,c)}$
BalMult	$\sqrt[2\cdot\|C\|]{pcos(s,t)^{\|C\|}\cdot\prod_{c\in C} pcos(s,c)}$

Table 1: The different substitutability measures considered in our model for a lexical substitute s of the target word t in sentential context C. C is represented by the set of the target word's context elements in the context sentence, where c denotes an individual context element. cos is the vector Cosine function applied to the vector representations of the words or contexts, and $pcos(v,v') = \frac{cos(v,v')+1}{2}$ is used to avoid negative values in Mult and BalMult.

Melamud et al. A simple word embedding model for lexical substitution. VSM Workshop. Denver, CO, USA.

Neural Language Models for Lexical Substitution

https://github.com/Samsung/LexSubGen

- Use a pre-trained language model to generate substitutes {s}
  - BERT, ELMo, RoBERTa, XLNet, …
- Mix the contextual distribution P(s|C) with the similarity of the target word to the substitutes P(s|T):

$$P(s|C,T) \propto \frac{P(s|C)P(s|T)}{P(s)^{\beta}}$$

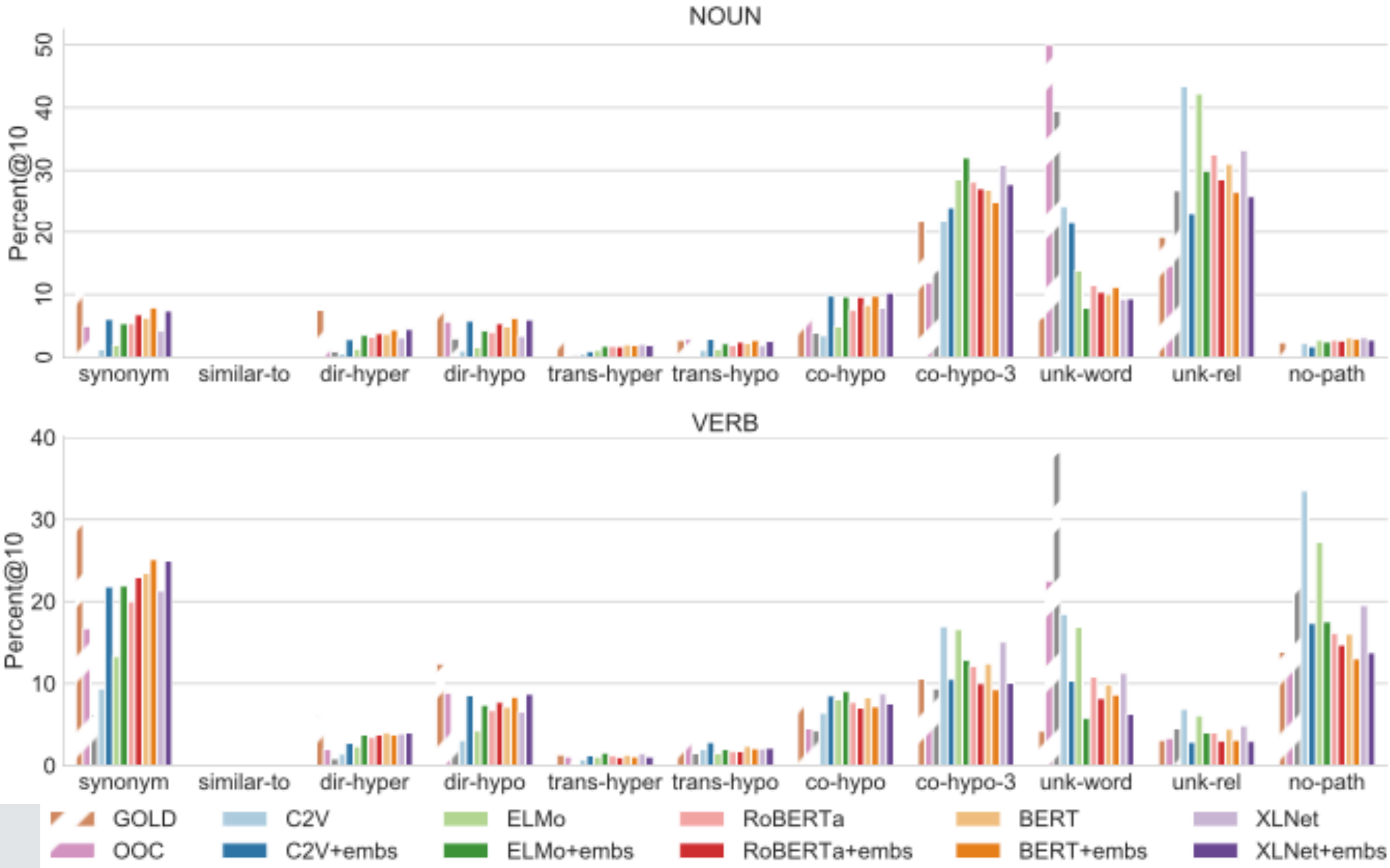
- To align the orders of distributions, use temperature softmax

$$P(s|T) \propto \exp\left(\frac{\langle emb_s, emb_T \rangle}{\mathcal{T}}\right)$$

Model	SemEval 2007			CoInCo		
	GAP	P@1	P@3	GAP	P@1	P@3
Transfer Learning (Hintz and Biemann, 2016)	51.9	-	-	-	-	-
PIC (Roller and Erk, 2016)	52.4	19.7	14.8	48.3	18.2	13.8
Supervised Learning (Szarvas et al., 2013b)	55.0	-	-	-	-	-
Substitute vector (Melamud et al., 2015a)	55.1	-	-	50.2	-	-
context2vec (Melamud et al., 2016)	56.0	-	-	47.9	-	-
BERT for lexical substitution (Zhou et al., 2019)[6]	**60.5**	**51.1**	-	**57.6**	**56.3**	-
XLNet+embs	**59.6**	**49.5**	**34.9**	**55.6**	**51.5**	**39.9**
XLNet+embs (w/o target exclusion)	59.6	0.4	26.0	53.5	2.5	30.0
XLNet+embs (w/o lemmatization)	59.2	38.3	27.5	53.2	34.5	27.1
XLNet+embs ($\mathcal{T} = 1.0$)	52.6	34.8	24.6	49.4	40.5	30.4

Arefyev et al. (2020): Always Keep your Target in Mind: Studying Semantics and Improving Performance of Neural Lexical Substitution. COLING.

What Types of Relations Lexical Subtition Generates?



Arefyev et al. (2020): Always Keep your Target in Mind: Studying Semantics and Improving Performance of Neural Lexical Substitution. COLING.

What Types of Relations Lexical Subtition Generates?

We were not able to travel in the weather , and there was no ***phone*** .										
GOLD	**telephone (5)**									
OOC	phone	**telephone**	phones	cellphone	fone	videophone	handset	telephones	p990i	cell-phone
XLNet	electricity	internet	phone	power	**telephone**	car	water	communication	radio	tv
XLNet+embs	phone	**telephone**	phones	cellphone	internet	radio	electricity	iphone	car	computer
What happened to the big , new garbage ***can*** at Church and Chambers Streets ?										
GOLD	**bin (4)**	**disposal (1)**	**container (1)**							
OOC	can	could	should	would	will	must	might	to	may	ll
XLNet	can	dump	**bin**	truck	**disposal**	pit	heap	pile	**container**	stand
XLNet+embs	can	could	will	**bin**	cannot	dump	may	truck	**disposal**	stand

Types of semantic relations: synonym co-hyponym co-hyponym 3 target direct hypernym transitive hypernym direct hyponym transitive hyponym unknown word unknown relation

Figure 3: Visualization of top 10 substitutes provided by annotators (gold), baseline method (OOC), and two presented models: XLNet and XLNet+embs. The target word is written in bold in the sentence. For gold substitutes, weights are given in brackets. Each word is colored according to WordNet relation between the substitute and the target word (substitutes from models without post-processing are shown).

Arefyev et al. (2020): Always Keep your Target in Mind: Studying Semantics and Improving Performance of Neural Lexical Substitution. COLING.

Summary on Word Senses and Meaning

- Formalization is less clear than in morphology and syntax: e.g. the **number of senses is not fixed**
  - Contextualized word embeddings to rescue?
  - Sense-inventory–based WSD vs. Lexical Substitution
- Extensive semantic resources (WordNet, BabelNet), still lack a connection to text
- Meaning has many facets and manifests itself in context
- More semantic tasks:
  - Semantic Role Labeling
  - Frame-Semantic Parsing
  - Word-in-Contex (WiC) + multi and cross-language setups.