Contextualized Embeddings, Word Sense Disambiguation and Open Challenges

Jose Camacho Collados

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

gwc 2023 Donostia

Global WordNet Conference 2023
25 January 2023

About me

- Senior Lecturer at **Cardiff University** (Wales, UK)
  - **UKRI Future Leaders Fellow** (4+ years)
  - Co-founder and leader of the **Cardiff NLP group**

- Areas of expertise: **Semantics, resources, multilinguality, social media**
  - Co-author of “Embeddings in NLP” book
  - Program chair of *SEM-2023
  - Developer of TweetNLP (tweetnlp.org)

Cardiff NLP

- ➢ Very young group (2 years old)
- ➢ Growing fast (25+ lab members)
- ➢ **Website:** cardiffnlp.github.io
- ➢ Activities: hybrid seminars, workshops, hackathons, etc.
- ➢ **Twitter:** @Cardiff_NLP
- ➢ Open-source contributions

Cardiff NLP Workshop 2023

- **Dates:** TBA (Summer), 2 days (in-person)
- Especially targeted to **NLP PhD students** in Europe (but everyone is welcome)
- **Free registration**
- Mix of **invited speakers, tutorials** and **networking**
- Info from last year: https://www.cardiffnlpworkshop.org/

MSc in
Natural Language Processing

Starting date: **September 2023**

Length: **1 year**

Involvement from industry and Supercomputing Wales.

NLP-specific modules, interdisciplinary by nature.

Computational linguistics, Python programming, machine learning, cutting-edge NLP, etc.

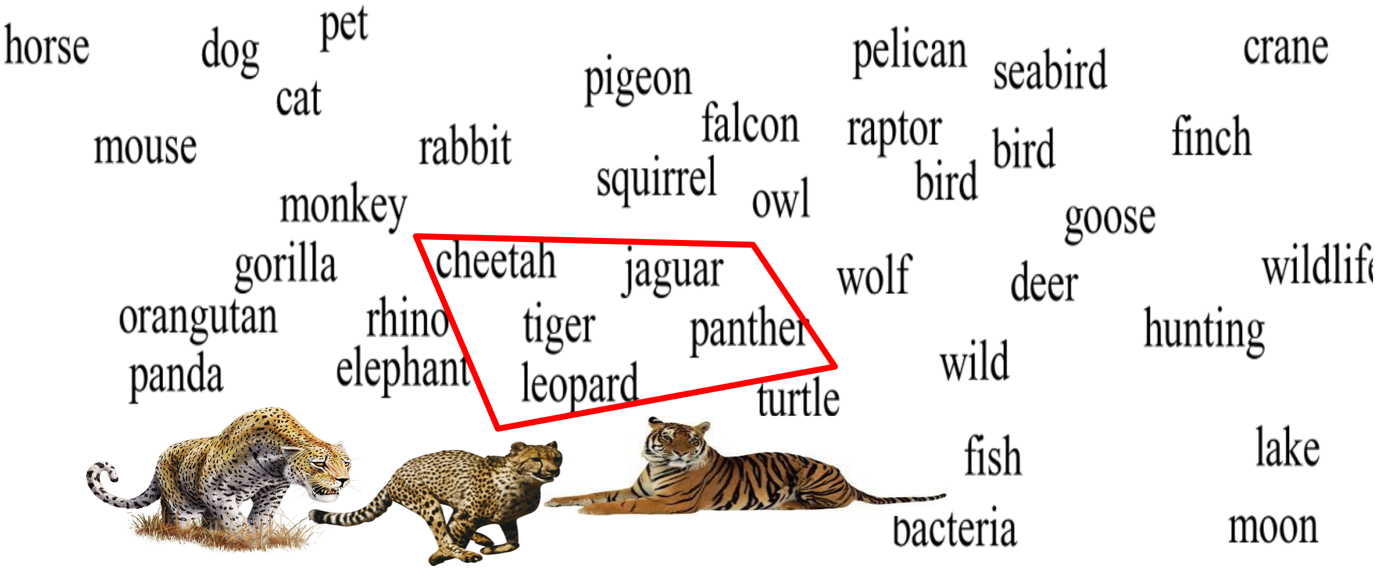
Outline

- ➢ From word to sense and contextualized embeddings:
  - ○ A “historical perspective”
- ➢ Contextualized embeddings and word sense disambiguation
- ➢ Open research questions

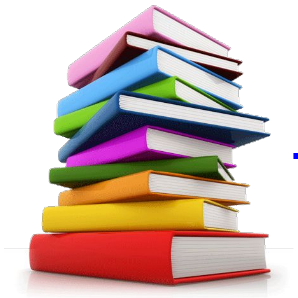
Embeddings in NLP

Word vector space models

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



Word embeddings: How to learn them



... **London** is the **capital** of **UK** …

London

[0.25, 0.32, -0.1 …. 0.1]

... Last night I **travelled** from **Cardiff** to **London**.

.

.

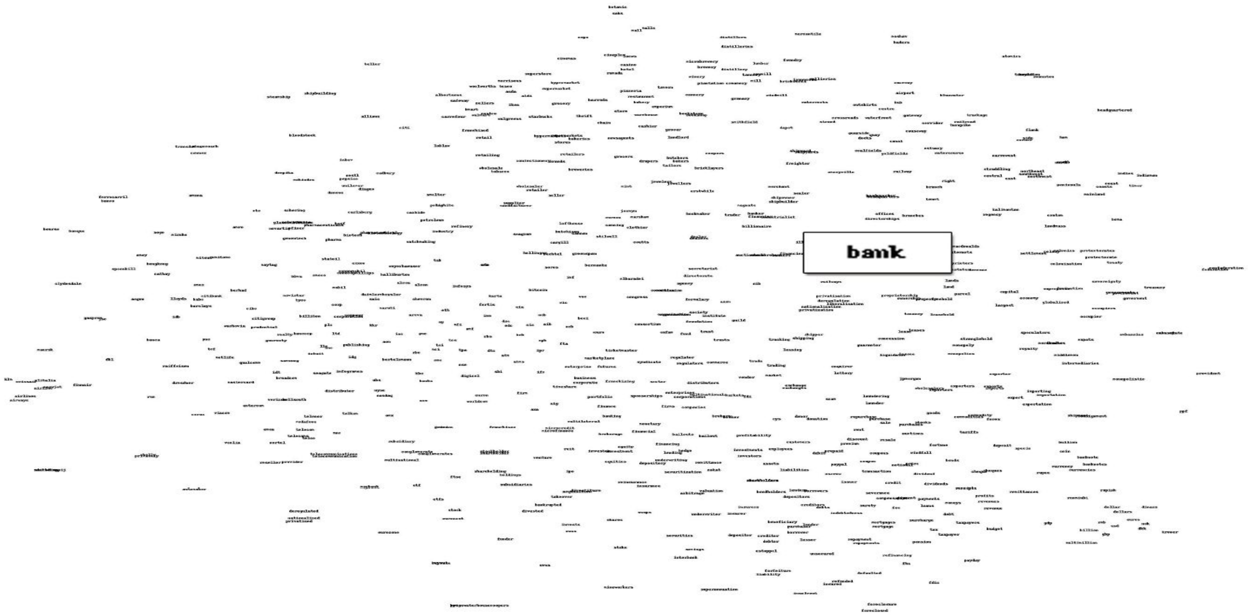
.

Word2Vec, GloVe, fasttext…

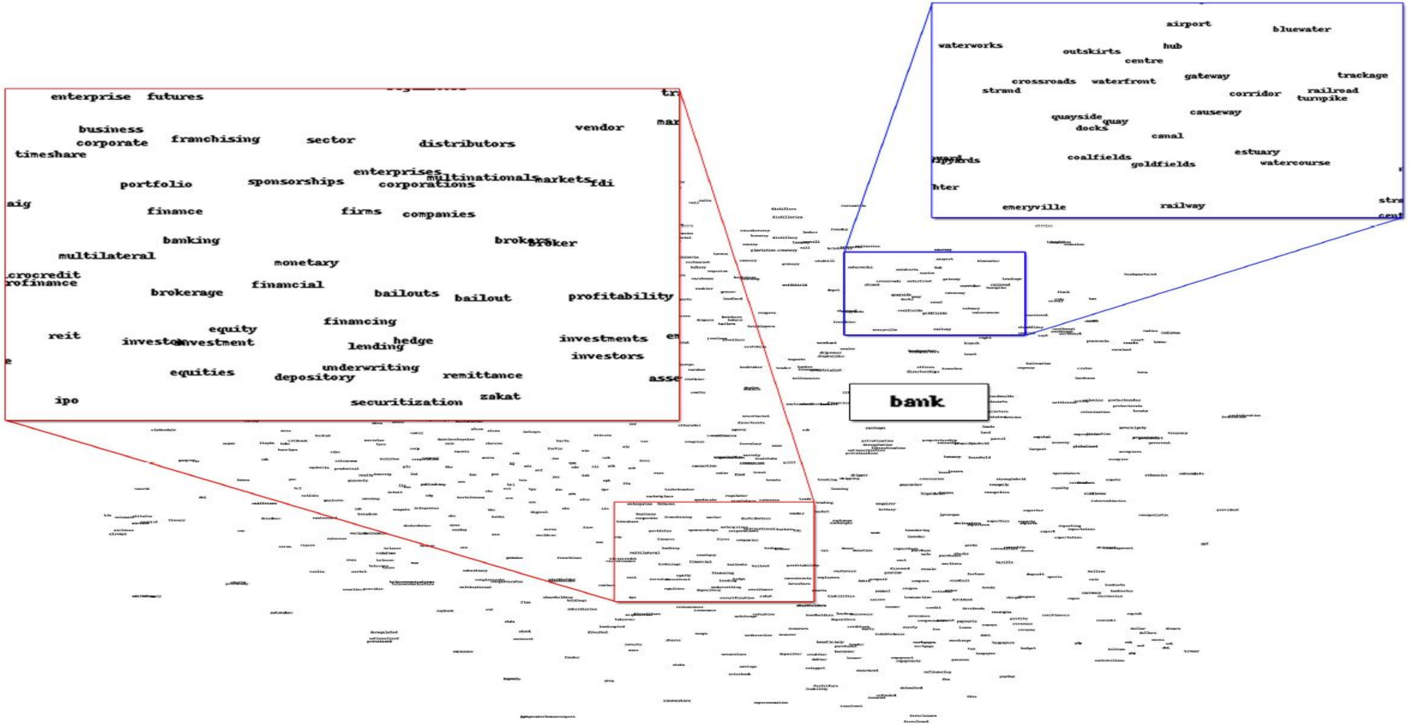
Limitations of word embeddings

- Word representations cannot capture ambiguity. For instance, ***bank***

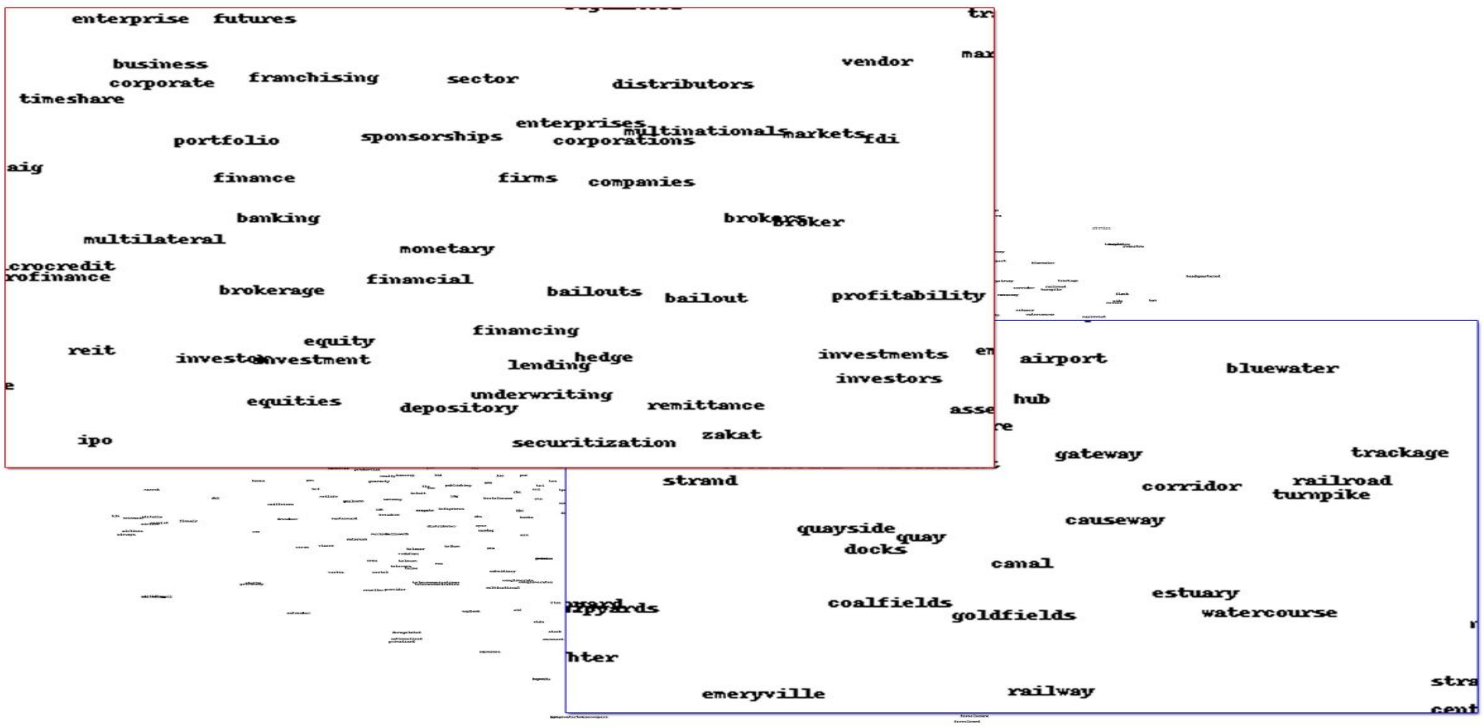
Problem:
word representations cannot capture ambiguity



Problem:
word representations cannot capture ambiguity

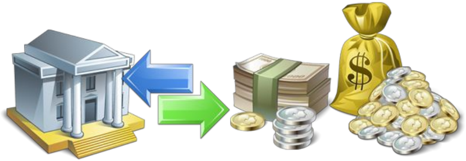


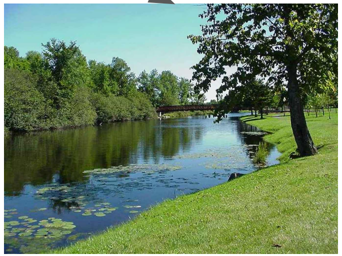
Problem:
word representations cannot capture ambiguity



Motivation: Model senses instead of words

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



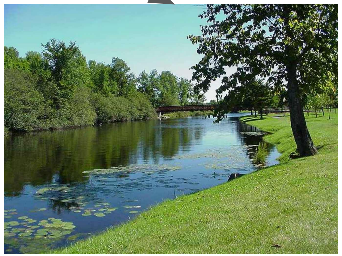


Motivation: Model senses instead of words

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

bank#1

...



bank#2

...

Motivation: Model senses instead of words

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

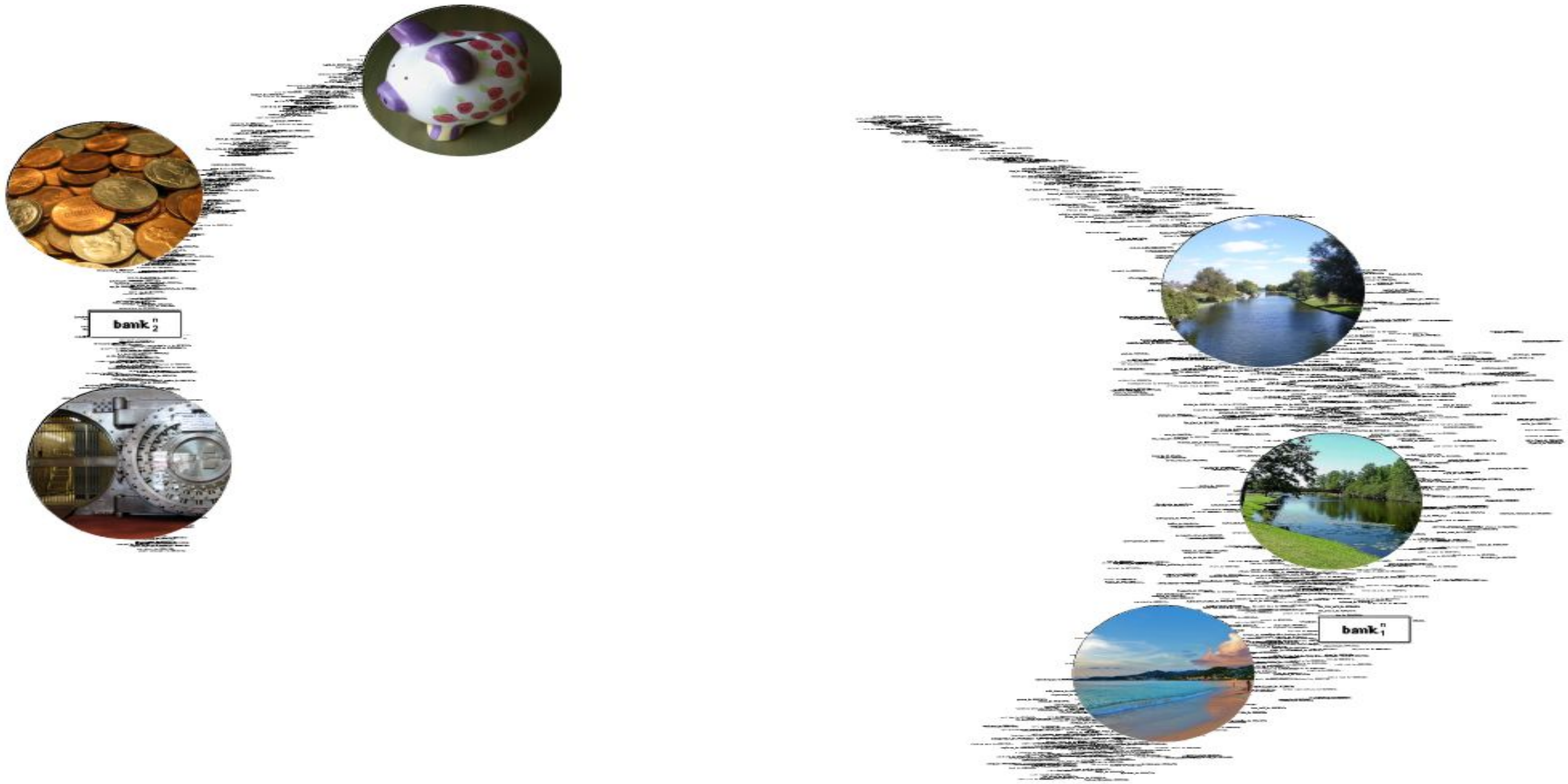
bank#1

...

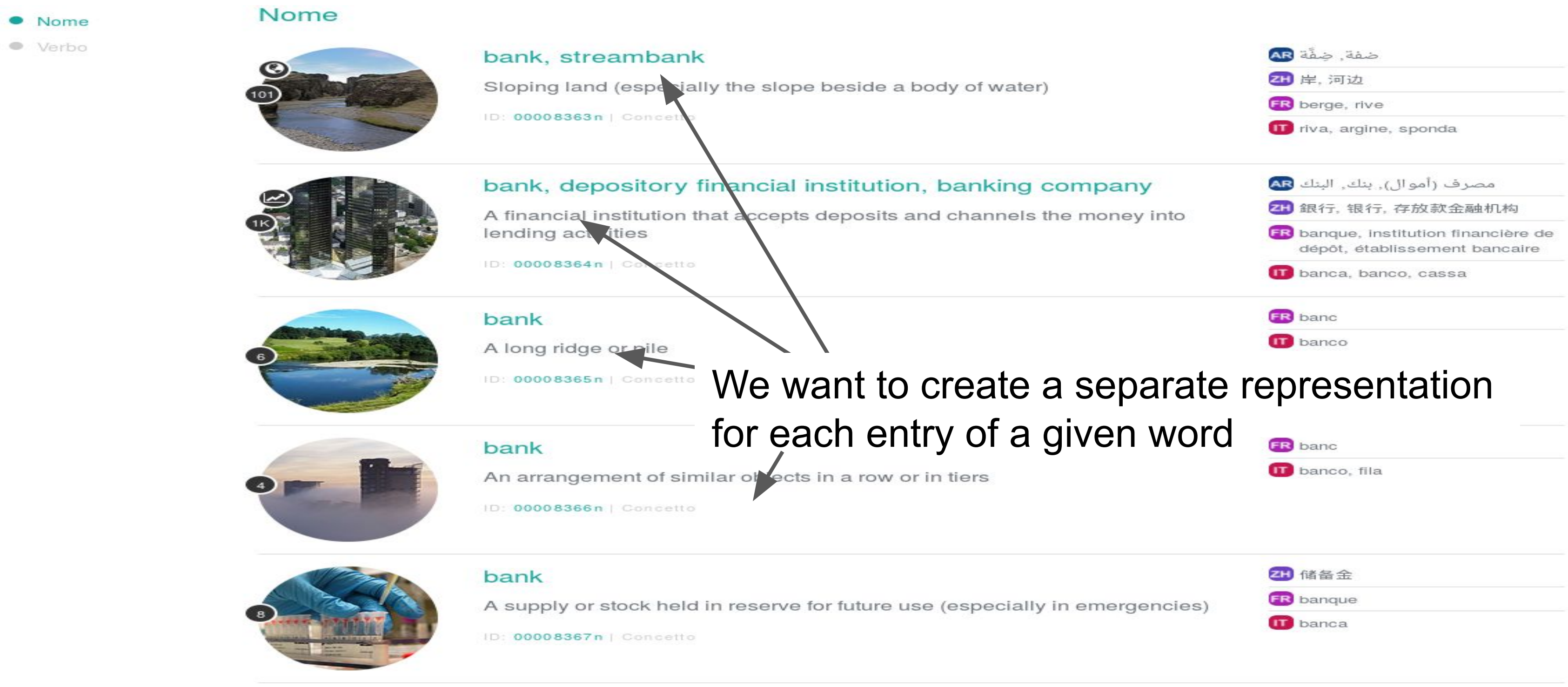
bank#2

...

Key goal: obtain sense representations



Key goal: obtain sense representations



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

Idea

Encyclopedic knowledge

Wikipedia
The Free Encyclopedia

Lexicographic knowledge

WordNet

Idea

(basis of my PhD thesis - NASARI)

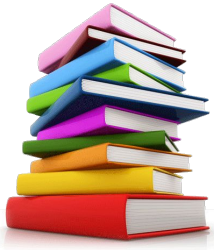
Encyclopedic knowledge

WIKIPEDIA
The Free Encyclopedia

BabelNet

Lexicographic knowledge

WordNet



Information from text corpora

Embedded sense representation

(Camacho-Collados et al. AIJ 2016)

Closest senses

Bank (financial institution)	
Closest senses	Cosine
Deposit account	0.99
Universal bank	0.99
British banking	0.98
German banking	0.98
Commercial bank	0.98
Banking in Israel	0.98
Financial institution	0.98
Community bank	0.97

Bank (geography)	
Closest senses	Cosine
Stream bed	0.98
Current (stream)	0.97
River engineering	0.97
Braided river	0.97
Fluvial terrace	0.97
Bar (river morphology)	0.97
River	0.97
Perennial stream	0.96

bank	
Closest senses	Cosine
Bank (financial institution)	0.86
Universal bank	0.86
British banking	0.86
German banking	0.85
Branch (banking)	0.85
McFadden Act	0.85
Four Northern Banks	0.84
State bank	0.84

Contextualized word embeddings

Contextualized word embeddings

ELMo

Peters et al.
(NAACL 2018)

Based on LSTMs

BERT

Devlin et al.
(NAACL 2019)

Based on Transformers

Contextualized word embeddings

ELMo

Peters et al.
(NAACL 2018)

Based on LSTMs

GPT-3
XLNet
RoBERTa
OPT
...

More successful nowadays

BERT

Devlin et al.
(NAACL 2019)

Based on Transformers

Contextualized word embeddings
ELMo/BERT

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

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

Important **improvements in many NLP tasks**.

Contextualized word embeddings ELMo/BERT (examples)

*She made a money transfer at the **bank**.*

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

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

Contextualized word embeddings ELMo/BERT (examples)

0.25, 0.32, -0.1 ….

She made a money transfer at the ***bank****.*

0.22, 0.30, -0.08 ….

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

-0.8, 0.01, 0.3 ….

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

Contextualized word embeddings ELMo/BERT (examples)

Similar vectors

0.25, 0.32, -0.1 ….

She made a money transfer at the ***bank****.*

0.22, 0.30, -0.08 ….

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

-0.8, 0.01, 0.3 ….

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

How well do these models capture “meaning”?

Good enough for many applications.

GLUE
Language understanding benchmark

Human baselines!

	Rank	Name	Model	URL	Score
	1	ERNIE Team - Baidu	ERNIE		90.1
	2	Microsoft D365 AI & MSR AI & GATECH	MT-DNN-SMART		89.9
	3	T5 Team - Google	T5		89.7
+	4	王玮	ALICE v2 large ensemble (Alibaba DAMO NLP)		89.5
	5	XLNet Team	XLNet (ensemble)		89.5
	6	ALBERT-Team Google Language	ALBERT (Ensemble)		89.4
	7	Microsoft D365 AI & UMD	FreeLB-RoBERTa (ensemble)		88.8
	8	Facebook AI	RoBERTa		88.5
	9	Junjie Yang	HIRE-RoBERTa		88.3
+	10	Microsoft D365 AI & MSR AI	MT-DNN-ensemble		87.6
	11	GLUE Human Baselines	GLUE Human Baselines		87.1

How well do these models capture "meaning"?

Good enough for many applications.

Room for improvement. For example, in SuperGLUE:

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

- ➢ *Word-in-Context (WiC) Challenge*: BERT ~69% **vs** Humans ~80%

How well do these models capture "meaning"?

Good enough for many applications.

Room for improvement. For example, in SuperGLUE:

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

  requires commonsense reasoning

- *Word-in-Context (WiC) Challenge*: BERT ~69% **vs** Humans ~80%

  requires abstracting the notion of sense

How well do these models capture "meaning"?

Good enough for many applications.

Room for improvement. For example, in SuperGLUE:

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

  requires commonsense reasoning

- ➢ *Word-in-Context (WiC) Challenge*: BERT ~69% **vs** Humans ~80%

  requires abstracting the notion of sense

Word-in-Context (WiC) Challenge

(Pilehvar and Camacho-Collados, NAACL 2019)

Task: Identify the most suitable meaning of a word in context

Framed as binary classification (True/False)

Examples:

*There's a lot of trash on the **bed** of the river*
*I keep a glass of water next to my **bed** when I sleep* ⇨ False

*He cashed a check at the **bank***
*The **bank** is on the corner of Nassau and Witherspoon* ⇨ True

Word-in-Context (WiC) Challenge

(Pilehvar and Camacho-Collados, NAACL 2019)

Despite smashing most benchmarks,
GPT-3 performance in WiC (few-shot) below 50%!

WiC is a notable weak spot with few-shot performance at 49.4% (at random chance). We tried a number of different phrasings and formulations for WiC (which involves determining if a word is being used with the same meaning in two sentences), none of which was able to achieve strong performance. This hints at a phenomenon that will become clearer

	WiC Accuracy
Fine-tuned SOTA	**76.1**
Fine-tuned BERT-Large	69.6
GPT-3 Few-Shot	49.4

Original GPT-3 paper (Brown et al., NeurIPS 2020)

WiC Challenge

But still…

WiC leaderboard

System	Implementation	Acc.
SenseBERT	Levine et al. (2019)	72.1
KnowBERT-W+W	Peters et al (2019)	70.9
RoBERTa	Liu et al. (2019)	69.9
BERT-large	SuperGLUE baseline	69.6
LMMS-WSD (BERT+)	Loureiro and Jorge (2019)	67.7
Ensemble (BERT+USE+ELMo)	Garí Soler et al. (2019)	66.7
BERT-large	WiC baseline	65.5

BERT is everywhere!

WiC challenge (now also multilingual!)

(Raganato and Pasini et al., EMNLP 2020)

12 languages

Lang	Target	Context-1	Context-2	Label
EN	Beat	We beat the competition	Agassi beat Becker in the tennis championship.	True
DA	Tro	Jeg tror p˚a det, min mor fortalte.	Maria troede ikke sine egne øjne.	True
ET	Ruum	Uhel hetkel olin ̈v aljaspool aega ja ruumi.	Umberringi oli ̃l oputu ̈t uhi ruum.	True
FR	Causticité	Sa causticité lui a fait bien des ennemis.	La causticité des acides.	False
KO	틀림	틀림이 있는지 없는지 세어 보시오.	그 아이 하는 짓에 틀림이 있다면 모두 이 어미 죄이지요.	False
ZH	發	建築師希望發大火燒掉城市的三分之一。	如果南美洲氣壓偏低，則印度可能發乾旱	True
FA	صرف	صرف غذا نیم ساعت طول کشید	معلم صرف افعال ماضی عربی را آموزش داد	False

https://pilehvar.github.io/xlwic/

WiC challenge (now also multilingual!)

(Raganato and Pasini et al., EMNLP 2020)

12 languages

Lang	Target	Context-1	Context-2	Label
EN	Beat	We <u>beat</u> the competition	Agassi <u>beat</u> Becker in the tennis championship.	True
DA	Tro	Jeg <u>tror</u> p˚a det, min mor fortalte.	Maria <u>troede</u> ikke sine egne øjne.	True
ET	Ruum	Uhel hetkel olin v̈ aljaspool aega ja <u>ruumi</u>.	Umberringi oli ̃ oputu ̈ uhi <u>ruum</u>.	True
FR	Causticité	Sa <u>causticité</u> lui a fait bien des ennemis.	La <u>causticité</u> des acides.	False
KO	틀림	<u>틀림이</u> 있는지 없는지 세어 보시오.	그 아이 하는 짓에 <u>틀림이</u> 있다면 모두 이 어미 죄이지요.	False
ZH	發	建築師希望<u>發</u>大火燒掉城市的三分之一。	如果南美洲氣壓偏低，則印度可能<u>發</u>乾旱	True
FA	صرف	<u>صرف</u> غذا نیم ساعت طول کشید	معلم <u>صرف</u> افعال ماضی عربی را آموزش داد	False

https://pilehvar.github.io/xlwic/

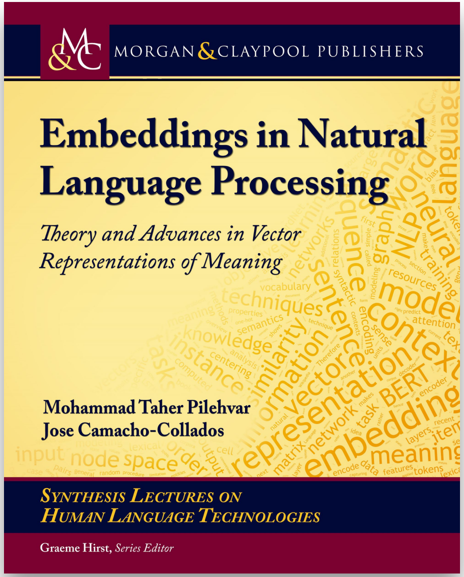
Other WiC extensions: **MCL-WiC** (Martelli et al., SemEval 2021); **AM2iCo** (Liu et al., EMNLP 2021); **WiC-TSV** (Breit et al., EACL 2021); **TempoWiC** (Loureiro et al., COLING 2022)

For more information on meaning representations (embeddings):

- ❖ Blog post on *"How to Represent Meaning in Natural Language Processing? Word, Sense and Contextualized Embeddings: Word, Sense and Contextualized Embeddings"*
- ❖ From Word to Sense Embeddings: A Survey on Vector Representations of Meaning (JAIR, Dec 2018)
- ❖ Book on "Embeddings in Natural Language Processing",

including COLING-20 Tutorial and ESSLLI-21 course:

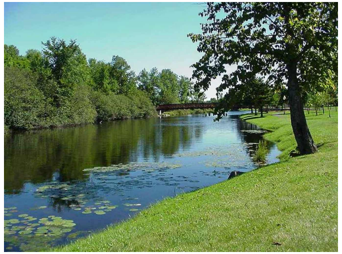
https://sites.google.com/view/embeddings-in-nlp



Word Sense Disambiguation (WSD)

Word Sense Disambiguation

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

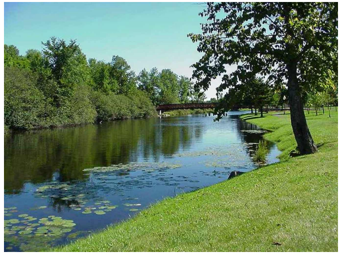


?

- S: (n) **bank** (sloping land (especially the slope beside a body of water)) *"they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the currents"*

Word Sense Disambiguation

He withdrew money from the bank.

X 

Word Sense Disambiguation

He withdrew money from the bank.

✓ 

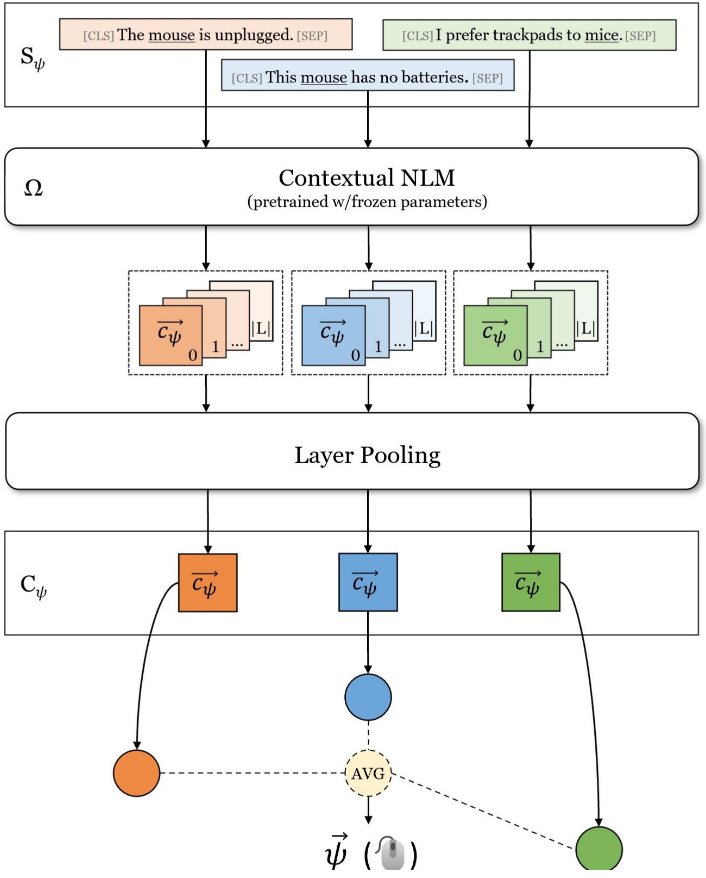
- S: (n) depository financial institution, **bank**, banking concern, banking company (a financial institution that accepts deposits and channels the money into lending activities) *"he cashed a check at the bank"; "that bank holds the mortgage on my home"*

WSD and language models

Common approach -> **“contextualized sense embeddings”**. For each sense:

- ➔ Gather all sentences in a sense-annotated corpus (e.g. SemCor)
- ➔ Get contextualized embeddings for each sentence
- ➔ Average all contextualized embeddings

At test time, nearest neighbour (1NN) between contextualized embedding and the computed representation for each candidate sense.



Compute *contextualized sense embeddings*

0.25, 0.32, -0.1 ….

*She made a money transfer at the **bank**.*

0.22, 0.30, -0.08 ….

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

0.1, 0.15, -0.02 ….

*The **bank** transaction was successful.*

Compute *contextualized sense embeddings*

AVERAGE

…

0.25, 0.32, -0.1 ….

She made a money transfer at the ***bank****.*

0.22, 0.30, -0.08 ….

…

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

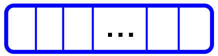
…

0.1, 0.15, -0.02 ….

The ***bank*** *transaction was successful.*

Disambiguating with contextualized embeddings

Contextualized embedding



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

Disambiguating with contextualized embeddings

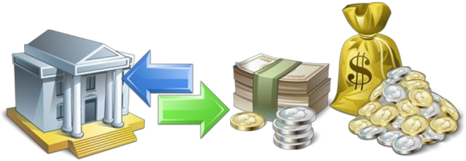
Contextualized embedding

[| | | ... | |]

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

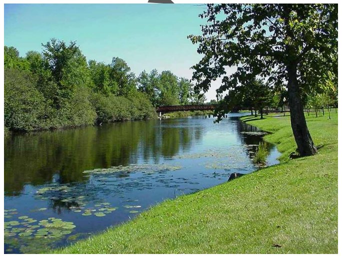
Similarity: 0.95

Similarity: 0.3



bank#1

[| | | ... | |]



bank#2

[| | | ... | |]

Disambiguating with contextualized embeddings

Contextualized embedding

[| | | … | |]

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

Similarity: 0.95

Similarity: 0.3

bank#1

[| | | … | |]

bank#2

[| | | … | |]

F1 WSD performance
(unified WordNet WSD benchmark)

Type	System	SE2		SE3		SE07		SE13		SE15		ALL	
		FN	CS	FN	CS	FN	CS	FN	CS	FN	CS	FN	CS
KB	Lesk_{ext}+emb	63.0	74.9	63.7	75.5	**56.7**	71.6	66.2	77.4	64.6	73.9	63.7	75.3
KB	Babelfy†	67.0	78.4	63.5	77.5	51.6	68.8	66.4	77.0	**70.3**	79.1	65.5	77.3
KB	TM	**69.0**	-	**66.9**	-	55.6	-	65.3	-	69.6	-	66.9	-
KB	UKB	68.8	81.2	66.1	78.1	53.0*	70.8	**68.8**	79.1	**70.3**	77.4	**67.3***	78.7*
Supervised SVM	IMS	70.9	81.5	69.3	80.8	61.3	74.3	65.3	77.4	69.5	75.7	68.4	79.1
Supervised SVM	IMS+emb	72.2	82.8	70.4	81.5	62.6	75.8	65.9	76.9	71.5	76.7	69.6	79.8
Supervised 1NN	Context2vec	71.8	82.6	69.1	80.5	61.3	74.5	65.6	78.0	71.9	76.6	69.0	79.7
Supervised 1NN	ELMo	71.6	82.8	69.6	80.9	62.2	74.7	66.2	77.7	71.3	77.0	69.0	79.6
Supervised 1NN	BERT-Base	75.5	**84.9**	71.5	81.4	65.1	78.9	69.8	82.1	73.4	78.1	72.2	82.0
Supervised 1NN	BERT-Large	**76.3**	84.8	**73.2**	**82.9**	**66.2**	**80.0**	**71.7**	**83.1**	**74.1**	**79.1**	**73.5**	**82.8**
Hybrid	Seq2Seq Att+Lex+PoS	70.1	-	68.5	-	63.1*	-	66.5	-	69.2	-	68.6*	-
Hybrid	Sense Compr. Ens.	79.7	-	77.8	-	73.4	-	78.7	-	82.6	-	**79.0**	-
Hybrid	LMMS 1024	75.4	-	74.0	-	66.4	-	72.7	-	75.3	-	73.8	-
Hybrid	LMMS 2048	76.3	84.5	75.6	85.1	68.1	81.3	75.1	86.4	77.0	80.8	75.4	84.4
Hybrid	EWISE	73.8	-	71.1	-	67.3*	-	69.4	-	74.5	-	71.8*	-
Hybrid	KnowBert† WN+WK	76.4	85.6	76.0	85.1	71.4	82.6	73.1	83.8	75.4	80.2	75.1	84.1
Hybrid	GlossBERT	77.7	-	75.2	-	72.5*	-	76.1	-	80.4	-	77.0*	-
Hybrid	BEM	79.4	-	77.4	-	74.5*	-	79.7	-	**81.7**	-	79.0*	-
Hybrid	EWISER†	**80.8**	-	**79.0**	-	**75.2**	-	**80.7**	-	81.8*	-	80.1*	-
-	*MFS Baseline*	65.6	77.4	66.0	77.8	54.5	70.6	63.8	74.8	67.1	75.3	64.8	76.2

FN=
Fine-grained

CS=
Coarse-grained

WSD performance *without candidates* (unified WordNet WSD benchmark)

(Loureiro et al., AIJ 2022)

Model	Sensekeys			Synsets		
	F1	P@5	MRR	F1	P@5	MRR
ARES	61.4	84.7	71.8	**60.7**†	**86.5**†	**71.8**†
LMMS_{1024} [61]	52.2	66.9	59.0	29.4†	53.9†	40.7†
LMMS_{2048} [61]	34.8	60.3	46.3	32.5†	58.9†	44.5†
$\text{LMMS-SP}_{\text{BERT-L}}$	60.8	86.7	72.2	51.0	81.7	64.3
$\text{LMMS-SP}_{\text{XLNet-L}}$	60.1	87.3	71.9	51.7	82.7	65.1
$\text{LMMS-SP}_{\text{RoBERTa-L}}$	62.2	86.9	73.1	50.2	80.1	63.3
$\text{LMMS-SP}_{\text{ALBERT-XXL}}$	**62.9**	**87.6**	**73.7**	52.7	81.9	65.5

In this setting, nearest neighbours if performed over **all senses/synsets in WordNet!**

Qualitative analysis

(20 words with human-interpretable senses)

CoarseWSD-20 Dataset (Loureiro and Rezaee et al., Computational Linguistics 2021)

Word	F2R	Ent.	Senses	Frequency
apple	1.6	0.96	apple_inc	1466/634
			apple	892/398
arm	2.8	0.83	arm_architecture	311/121
			arm	112/43
bank	23.1	0.28	bank	1061/433
			bank_(geography)	46/22
bass	2.9	0.67	bass_guitar	2356/1005
			bass_(voice_type)	609/298
			double_bass	208/88
bow	1.0	0.87	bow_ship	266/117
			bow_and_arrow	185/72
			bow_(music)	72/26
chair	1.4	0.91	chairman	156/88
			chair	115/42
club	0.9	0.85	club	186/108
			nightclub	148/73
			club_(weapon)	54/21
crane	1.3	0.99	crane_(machine)	211/81
			crane_(bird)	161/76
deck	8.4	0.37	deck_(ship)	152/92
			deck_(building)	18/7
digit	2.2	0.74	numerical_digit	47/33
			digit_(anatomy)	21/9
hood	1.6	0.88	hood_(comics)	105/47
			hood_(vehicle)	42/13
			hood_(headgear)	24/22

Word	F2R	Ent.	Senses	Frequency
java	1.4	0.96	java	2641/1180
			java_(programm._lang.)	1863/749
mole	0.4	0.93	mole_(animal)	148/77
			mole_(espionage)	120/44
			mole_(unit)	108/42
			mole_sauce	53/23
			mole_(architecture)	51/20
pitcher	355.7	0.04	pitcher	6403/2806
			pitcher_(container)	18/13
pound	6.2	0.48	pound_mass	160/87
			pound_(currency)	26/10
seal	0.5	0.87	pinniped	305/131
			seal_(musician)	267/106
			seal_(emblem)	265/114
			seal_(mechanical)	38/12
spring	0.9	0.91	spring_(hidrology)	516/236
			spring_(season)	389/148
			spring_(device)	159/73
square	1.1	0.83	square	264/103
			square_(company)	167/62
			town_square	56/29
			square_number	21/13
trunk	1.3	0.85	trunk_(botany)	93/47
			trunk_(automobile)	36/16
			trunk_(anatomy)	35/14
yard	5.3	0.62	yard	121/61
			yard_(sailing)	23/11

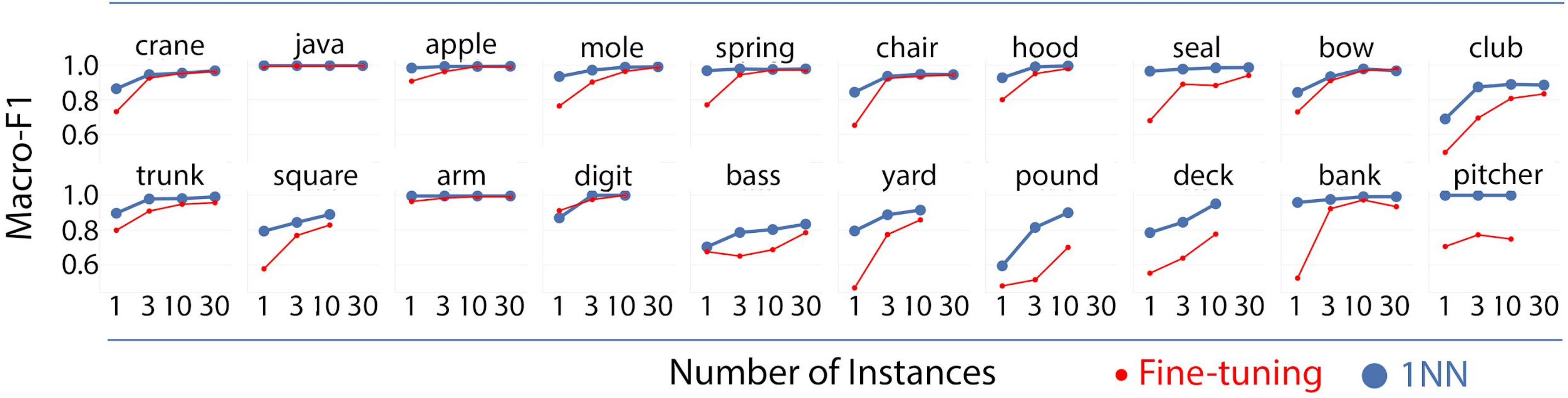
WSD Results (CoarseWSD-20)

Word	Micro F1						Macro F1					
	Static emb.		1NN		Fine-tune		Static emb.		1NN		Fine-tune	
	FTX-B	FTX-C	BRT-B	BRT-L	BRT-B	BRT-L	FTX-B	FTX-C	BRT-B	BRT-L	BRT-B	BRT-L
crane	91.7	94.9	93.6	96.8	97.5	98.1	91.7	94.8	93.5	96.7	97.5	98.1
java	98.8	99.4	99.6	99.6	99.7	99.7	98.7	99.4	99.7	99.6	99.7	99.7
apple	96.5	98.4	99.0	99.2	99.6	99.6	96.2	98.1	99.0	99.1	99.6	99.6
mole	87.4	93.2	97.1	98.5	98.9	98.9	84.4	91.0	97.6	99.0	98.9	99.2
spring	91.9	94.5	97.4	97.8	98.0	98.3	91.1	94.9	97.4	97.8	97.8	98.1
chair	81.5	88.5	96.2	96.2	96.7	96.2	79.5	86.5	94.7	94.7	96.1	95.5
hood	80.5	89.0	98.8	100	98.0	99.6	70.5	83.2	98.5	100	97.8	99.6
seal	88.7	95.0	96.4	98.1	99.0	99.0	72.7	92.6	97.3	98.5	98.9	98.6
bow	89.8	95.8	96.3	95.3	97.5	98.5	83.3	93.7	97.0	95.7	97.5	98.6
club	79.2	80.7	81.2	85.1	85.2	84.7	73.2	80.5	84.6	88.7	84.3	84.1
trunk	84.4	90.9	96.1	98.7	97.8	98.3	76.0	85.9	97.9	99.3	97.6	98.0
square	87.0	90.3	95.2	96.1	95.8	95.7	67.7	76.3	92.5	94.7	92.2	91.4
arm	94.5	98.2	99.4	99.4	99.4	99.4	92.5	98.0	99.6	99.6	99.2	99.2
digit	92.9	100	100	100	99.2	100	83.3	100	100	100	98.8	100
bass	93.9	94.2	80.7	84.5	95.5	95.8	80.2	81.3	79.1	84.0	87.5	87.6
yard	86.1	94.4	76.4	88.9	98.6	99.5	54.5	81.8	86.1	93.4	97.2	99.1
pound	87.6	87.6	86.6	89.7	94.9	94.9	48.9	53.3	92.5	94.3	84.4	83.9
deck	91.9	93.9	89.9	91.9	96.6	95.3	56.1	57.1	88.0	95.7	83.4	78.0
bank	96.9	98.0	99.6	99.8	99.6	99.3	68.2	79.5	95.5	97.7	97.9	95.6
pitcher	99.6	99.7	99.9	99.9	100	100	61.5	69.2	99.9	100	97.3	97.3
AVG	90.0	93.8	94.0	95.8	97.4	97.5	76.5	84.9	94.5	96.4	95.2	95.1

WSD Results (CoarseWSD-20)

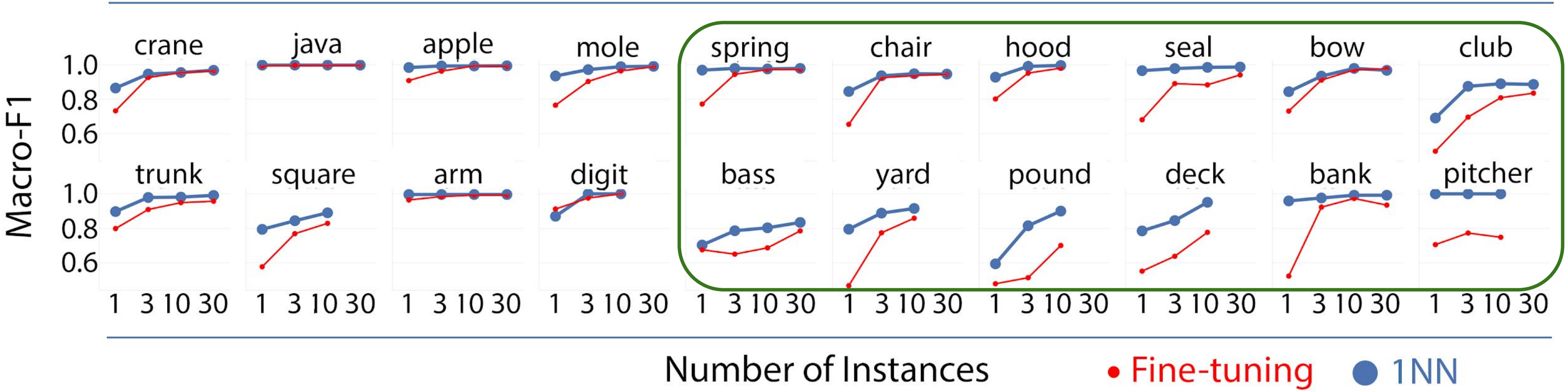
Word	Micro F1						Macro F1					
	Static emb.		1NN		Fine-tune		Static emb.		1NN		Fine-tune	
	FTX-B	FTX-C	BRT-B	BRT-L	BRT-B	BRT-L	FTX-B	FTX-C	BRT-B	BRT-L	BRT-B	BRT-L
crane	91.7	94.9	93.6	96.8	97.5	98.1	91.7	94.8	93.5	96.7	97.5	98.1
java	98.8	99.4	99.6	99.6	99.7	99.7	98.7	99.4	99.7	99.6	99.7	99.7
apple	96.5	98.4	99.0	99.2	99.6	99.6	96.2	98.1	99.0	99.1	99.6	99.6
mole	87.4	93.2	97.1	98.5	98.9	98.9	84.4	91.0	97.6	99.0	98.9	99.2
spring	91.9	94.5	97.4	97.8	98.0	98.3	91.1	94.9	97.4	97.8	97.8	98.1
chair	81.5	88.5	96.2	96.2	96.7	96.2	79.5	86.5	94.7	94.7	96.1	95.5
hood	80.5	89.0	98.8	100	98.0	99.6	70.5	83.2	98.5	100	97.8	99.6
seal	88.7	95.0	96.4	98.1	99.0	99.0	72.7	92.6	97.3	98.5	98.9	98.6
bow	89.8	95.8	96.3	95.3	97.5	98.5	83.3	93.7	97.0	95.7	97.5	98.6
club	79.2	80.7	81.2	85.1	85.2	84.7	73.2	80.5	84.6	88.7	84.3	84.1
trunk	84.4	90.9	96.1	98.7	97.8	98.3	76.0	85.9	97.9	99.3	97.6	98.0
square	87.0	90.3	95.2	96.1	95.8	95.7	67.7	76.3	92.5	94.7	92.2	91.4
arm	94.5	98.2	99.4	99.4	99.4	99.4	92.5	98.0	99.6	99.6	99.2	99.2
digit	92.9	100	100	100	99.2	100	83.3	100	100	100	98.8	100
bass	93.9	94.2	80.7	84.5	95.5	95.8	80.2	81.3	79.1	84.0	87.5	87.6
yard	86.1	94.4	76.4	88.9	98.6	99.5	54.5	81.8	86.1	93.4	97.2	99.1
pound	87.6	87.6	86.6	89.7	94.9	94.9	48.9	53.3	92.5	94.3	84.4	83.9
deck	91.9	93.9	89.9	91.9	96.6	95.3	56.1	57.1	88.0	95.7	83.4	78.0
bank	96.9	98.0	99.6	99.8	99.6	99.3	68.2	79.5	95.5	97.7	97.9	95.6
pitcher	99.6	99.7	99.9	99.9	100	100	61.5	69.2	99.9	100	97.3	97.3
AVG	90.0	93.8	94.0	95.8	97.4	97.5	76.5	84.9	94.5	96.4	95.2	95.1

Fine-tuning vs. 1NN (few-shot)



1NN: Method based on contextualized embeddings nearest neighbour

Fine-tuning vs. 1NN (few-shot)



1NN (contextualized embeddings) more robust
with low number of training instances

Many good news!

A simple 1NN based on **contextualized embeddings** method performs remarkably well (**over 90%**) in most settings.

It is **more robust than a fine-tuning approach** that is more computationally-demanding, needs one model per word, etc.

Only a **handful of annotated examples are needed** to achieve this performance (generally <= 3).

So, is it lexical ambiguity not a problem anymore in NLP? Is it WSD solved?

So, is it lexical ambiguity not a problem anymore in NLP? Is it WSD solved?

Certainly not!

So, it is lexical ambiguity not a problem anymore in NLP? Is it WSD solved?

Certainly not!

Many **challenges** remain, for example:

1) Lack of **sense-annotated corpora** (especially low-resource languages)
2) **Understand** how language models work (and take the most of them)
3) **Verbs**, what about them?
4) Dynamic nature of meaning (**meaning shift**, etc.)
5) **Multimodality** (images?)

(1) Lack of sense-annotated corpora

Existing manually sense-annotated corpora cover a small fraction of all senses

-> **Annotating senses is a hard and time-consuming task!**

This causes the so-called knowledge-acquisition bottleneck

For example, **SemCor covers 16.1% only!**

Solutions to lack of sense-annotated corpora

Extensions through **definitions and/or graph propagation**:

-> EWISER (Bevilacqua and Navigli, 2020), Scarlini et al. (2020), Blevins and Zettlemoyer (2020), LMMS (Loureiro et al. 2019), Vial et al. (2019), GlossBERT (Huang et al. 2019), etc.

Problem:

The initial annotations were still very limited and propagation methods cannot address all the problems (sparsity)

Solutions to lack of sense-annotated corpora

Extensions through **definitions and/or graph propagation**:

-> EWISER (Bevilacqua and Navigli, 2020), Scarlini et al. (2020), Blevins and Zettlemoyer (2020), LMMS (Loureiro et al. 2019), Vial et al. (2019), GlossBERT (Huang et al. 2019), etc.

Problem:

The initial annotations were still very limited and propagation methods cannot address all the problems (sparsity)

Solution:

Use unambiguous words!

Unambiguous Sense Annotations (UWA)

(Loureiro and Camacho-Collados, EMNLP 2020)

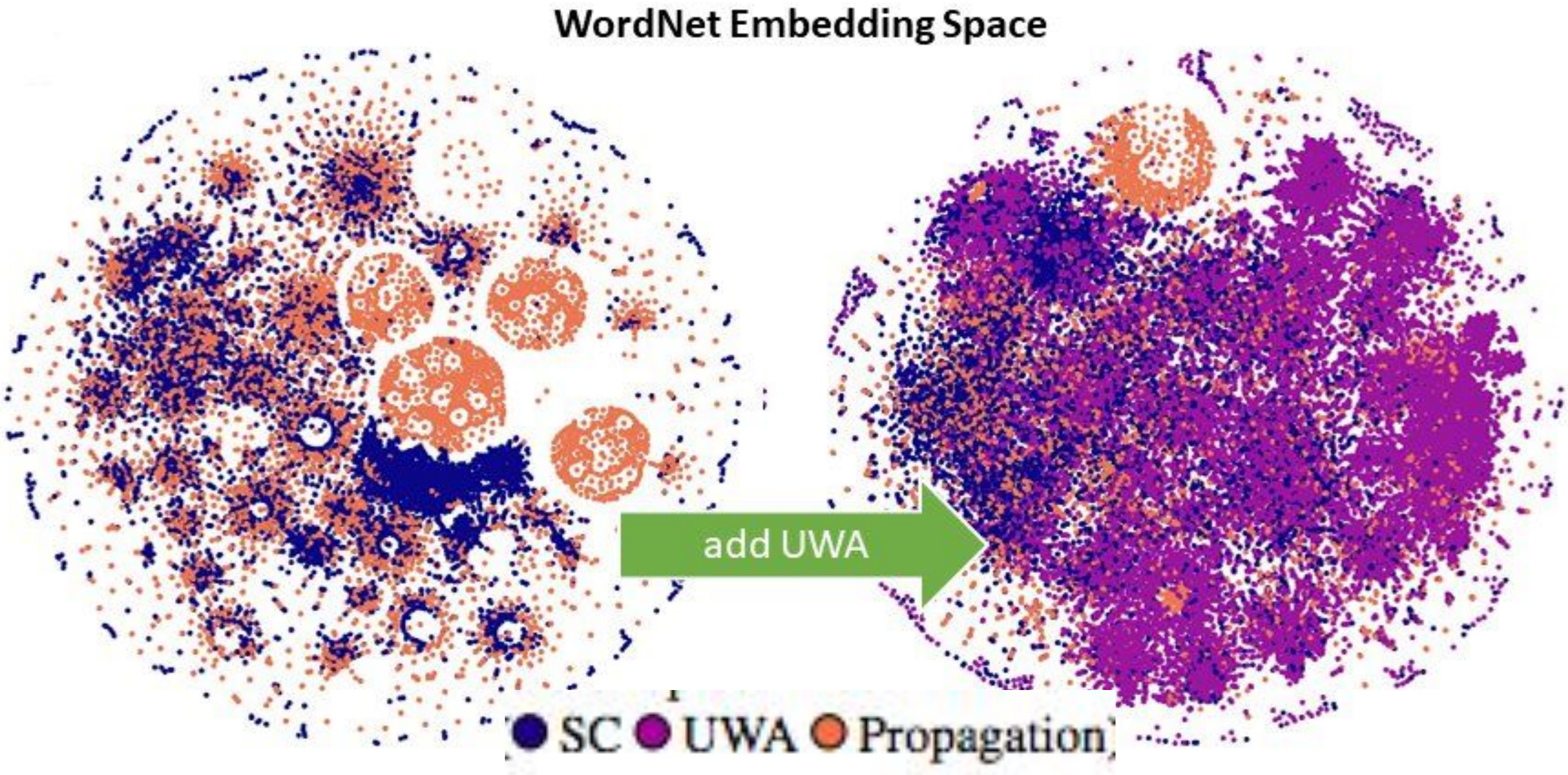
Unambiguous words amount to almost 80% of all words in WordNet! 🤯

Idea: We can annotate unambiguous words for free (with some caveats) and this should help propagation methods

We construct UWA, a corpus with unambiguous sense annotations (WordNet)

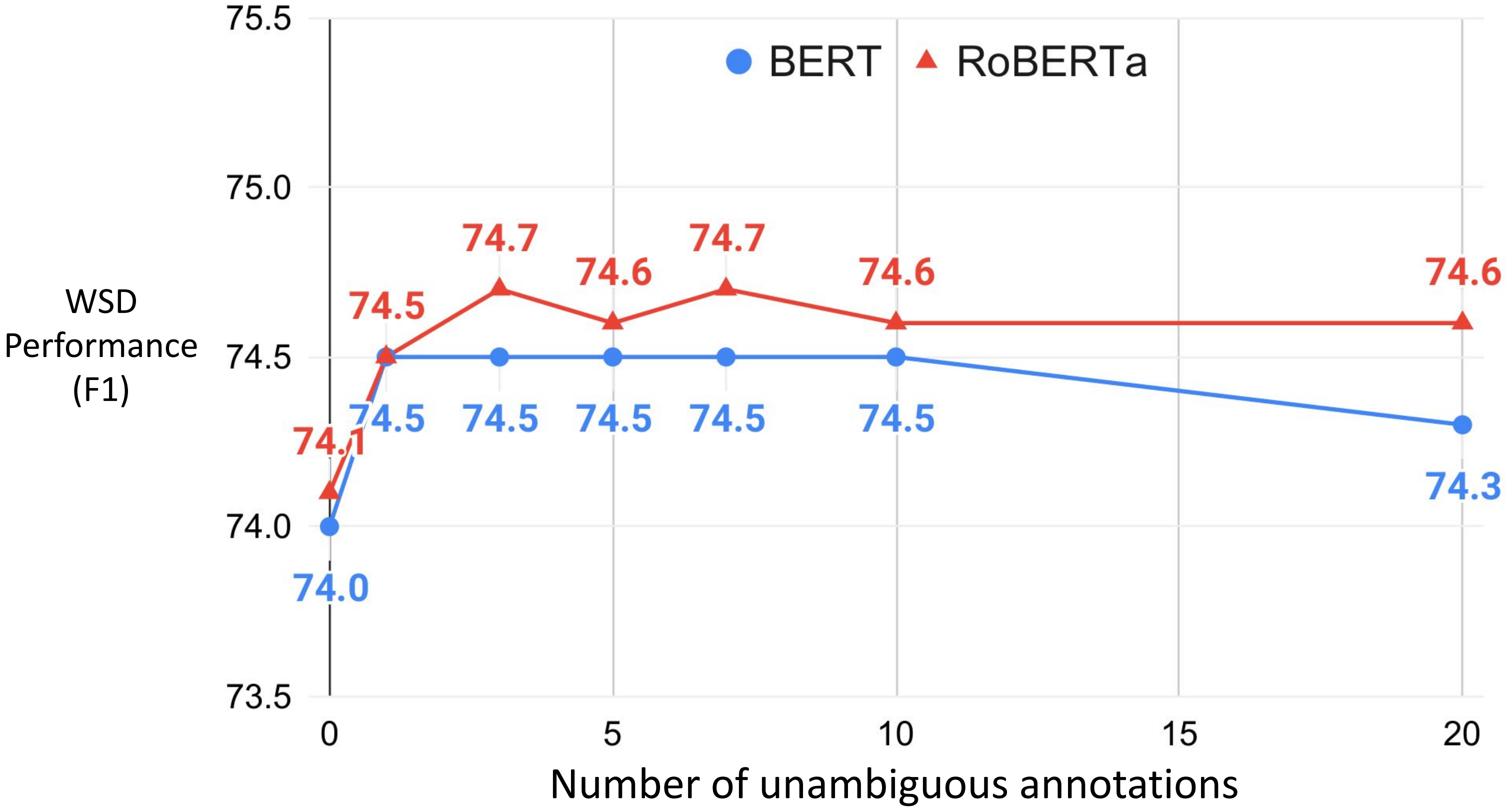
Unambiguous word annotations (UWA)

(Loureiro and Camacho-Collados, EMNLP 2020)



Unambiguous Sense Annotations (UWA)

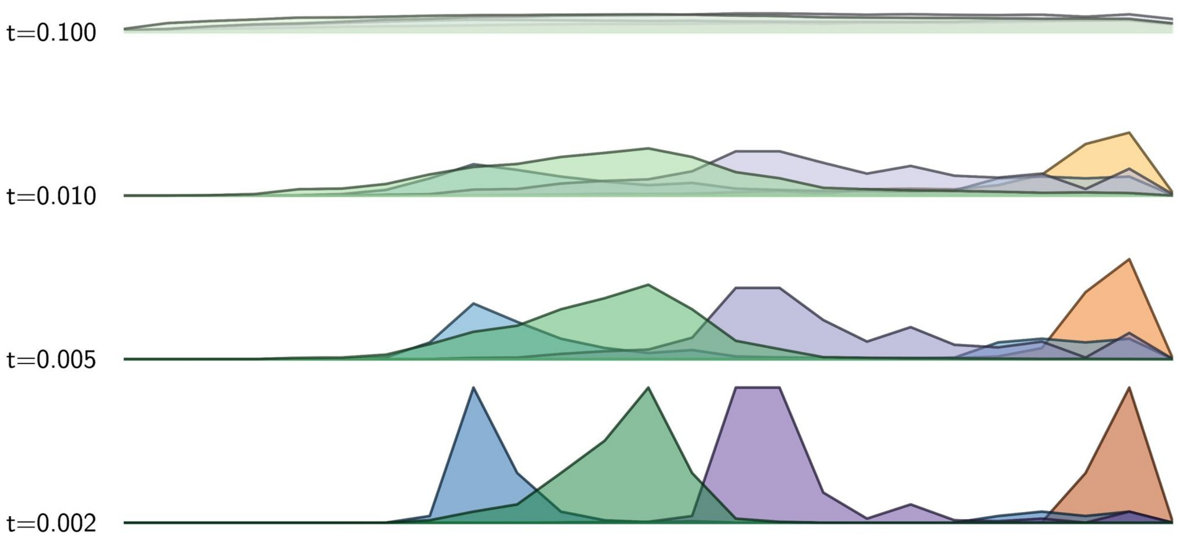
(Loureiro and Camacho-Collados, EMNLP 2020)



(2) Understanding LMs: Layer probing analysis

(Loureiro et al., AIJ 2022)

	INIT	-24	-23	-22	-21	-20	-19	-18	-17	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1
BERT	53	58	62	63	65	67	68	68	69	70	71	71	71	71	72	72	71	72	73	72	73	74	**75**	**75**	72
XLNet	51	57	65	67	68	70	71	72	**73**	**73**	72	72	72	72	71	71	71	71	71	71	72	72	72	72	68
RoBERTa	53	57	63	66	67	69	71	72	73	73	74	74	74	74	**75**	**75**	**75**	74	**75**	74	74	74	73	74	71
ALBERT	54	65	67	68	69	70	70	**71**	**71**	**71**	**71**	**71**	**71**	**71**	**71**	70	70	69	69	69	69	69	69	69	64



(3) Verbs: How to model them?

	Type	System	Nouns		Verbs		Adjectives		Adverbs	
			FN	CS	FN	CS	FN	CS	FN	CS
KB		UKB*	71.2	80.5	50.7	69.2	75.0	82.7	77.7	91.3
		Lesk_{ext}+emb	69.8	79.0	51.2	69.2	51.7	62.4	80.6	92.8
		Babelfy†	68.6	78.9	49.9	67.6	73.2	82.1	79.8	91.6
Supervised	1NN	Context2vec	71.0	80.5	57.6	72.9	75.2	83.1	82.7	92.5
		ELMo	70.9	80.0	57.3	73.5	77.4	85.4	82.4	92.8
		BERT-Base	74.0	83.0	61.7	75.3	77.7	84.9	85.8	93.9
		BERT-Large	75.1	83.7	63.2	76.6	79.5	85.4	85.3	94.2
	SVM	IMS	70.4	79.4	56.1	72.5	75.6	84.1	82.9	93.1
		IMS+emb	71.9	80.5	56.9	73.1	75.9	83.8	84.7	93.4
Hybrid		LMMS_{2048}	78.0	86.2	64.0	76.5	80.7	86.7	83.5	92.8
		KnowBert† WN+WK	77.0	85.0	66.4	78.8	78.3	86.1	84.7	93.9
	-	*MFS Baseline*	67.6	77.0	49.6	67.2	73.1	82.0	80.5	92.9

(3) Fine-granularity of verbs (and not only)

Verb

- S: (v) **run** (move fast by using one's feet, with one foot off the ground at any given time) *"Don't run--you'll be out of breath"; "The children ran to the store"*
- S: (v) scat, **run**, scarper, turn tail, lam, run away, hightail it, bunk, head for the hills, take to the woods, escape, fly the coop, break away (flee; take to one's heels; cut and run) *"If you see this man, run!"; "The burglars escaped before the police showed up"*
- S: (v) **run**, go, pass, lead, extend (stretch out over a distance, space, time, or scope; run or extend between two points or beyond a certain point) *"Service runs all the way to Cranbury"; "His knowledge doesn't go very far"; "My memory extends back to my fourth year of life"; "The facts extend beyond a consideration of her personal assets"*
- S: (v) operate, **run** (direct or control; projects, businesses, etc.) *"She is running a relief operation in the Sudan"*
- S: (v) **run**, go (have a particular form) *"the story or argument runs as follows"; "as the saying goes..."*
- S: (v) **run**, flow, feed, course (move along, of liquids) *"Water flowed into the cave"; "the Missouri feeds into the Mississippi"*
- S: (v) function, work, operate, go, **run** (perform as expected when applied) *"The washing machine won't go unless it's plugged in"; "Does this old car still run well?"; "This old radio doesn't work anymore"*
- S: (v) range, **run** (change or be different within limits) *"Estimates for the losses in the earthquake range as high as $2 billion"; "Interest rates run from 5 to 10 percent"; "The instruments ranged from tuba to cymbals"; "My students range from very bright to dull"*
- S: (v) campaign, **run** (run, stand, or compete for an office or a position) *"Who's running for treasurer this year?"*
- S: (v) play, **run** (cause to emit recorded audio or video) *"They ran the tapes over and over again"; "I'll play you my favorite record"; "He never tires of playing that video"*
- S: (v) **run** (move about freely and without restraint, or act as if running around in an uncontrolled way) *"who are these people running around in the building?"; "She runs around telling everyone of her troubles"; "let the dogs run free"*
- S: (v) tend, be given, lean, incline, **run** (have a tendency or disposition to do or be something; be inclined) *"She tends to be nervous before her lectures"; "These dresses run small"; "He inclined to corpulence"*
- S: (v) **run** (be operating, running or functioning) *"The car is still running--turn it off!"*
- S: (v) **run** (change from one state to another) *"run amok"; "run rogue"; "run riot"*
- S: (v) **run** (cause to perform) *"run a subject"; "run a process"*
- S: (v) **run** (be affected by; be subjected to) *"run a temperature"; "run a risk"*
- S: (v) prevail, persist, die hard, **run**, endure (continue to exist) *"These stories die hard"; "The legend of Elvis endures"*
- S: (v) **run** (occur persistently) *"Musical talent runs in the family"*
- S: (v) **run**, execute (carry out a process or program, as on a computer or a machine) *"Run the dishwasher"; "run a new program on the Mac"; "the computer executed the instruction"*
- S: (v) carry, **run** (include as the content; broadcast or publicize) *"We ran the ad three times"; "This paper carries a restaurant review"; "All major networks carried the press conference"*
- S: (v) **run** (carry out) *"run an errand"*
- S: (v) guide, **run**, draw, pass (pass over, across, or through) *"He ran his eyes over her body"; "She ran her fingers along the carved figurine"; "He drew her hair through his fingers"*
- S: (v) **run**, lead (cause something to pass or lead somewhere) *"Run the wire behind the cabinet"*
- S: (v) **run** (make without a miss)
- S: (v) **run**, black market (deal in illegally, such as arms or liquor)
- S: (v) **run** (cause an animal to move fast) *"run the dogs"*
- S: (v) **run**, bleed (be diffused) *"These dyes and colors are guaranteed not to run"*
- S: (v) **run** (sail before the wind)
- S: (v) **run** (cover by running; run a certain distance) *"She ran 10 miles that day"*
- S: (v) **run**, run for (extend or continue for a certain period of time) *"The film runs 5 hours"*
- S: (v) **run** (set animals loose to graze)
- S: (v) **run**, consort (keep company) *"the heifers run with the bulls to produce offspring"*
- S: (v) **run** (run with the ball; in such sports as football)
- S: (v) **run** (travel rapidly, by any (unspecified) means) *"Run to the store!"; "She always runs to Italy, because she has a lover there"*
- S: (v) ply, **run** (travel a route regularly) *"Ships ply the waters near the coast"*
- S: (v) hunt, **run**, hunt down, track down (pursue for food or sport (as of wild animals)) *"Goering often hunted wild boars in Poland"; "The dogs are running deer"; "The Duke hunted in these woods"*
- S: (v) race, **run** (compete in a race) *"he is running the Marathon this year"; "let's race and see who gets there first"*
- S: (v) move, go, **run** (progress by being changed) *"The speech has to go through several more drafts"; "run through your presentation before the meeting"*
- S: (v) melt, **run**, melt down (reduce or cause to be reduced from a solid to a liquid state, usually by heating) *"melt butter"; "melt down gold"; "The wax melted in the sun"*
- S: (v) ladder, **run** (come unraveled or undone as if by snagging) *"Her nylons were running"*
- S: (v) **run**, unravel (become undone) *"the sweater unraveled"*

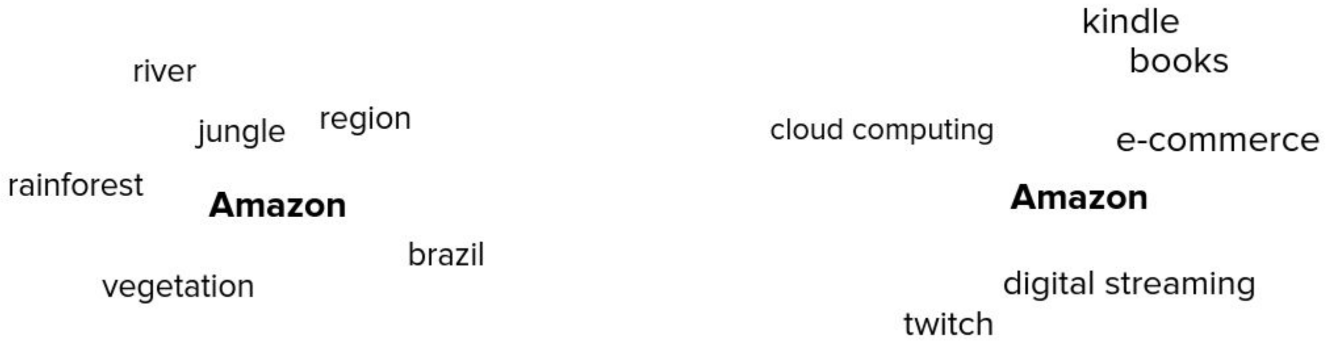
Example:

The verb *"run"* has
41 senses in WordNet!

See Kallini and Fellbaum (GWC 2023)!

(4) Language is dynamic

Meaning changes over time



More problematic in social media!

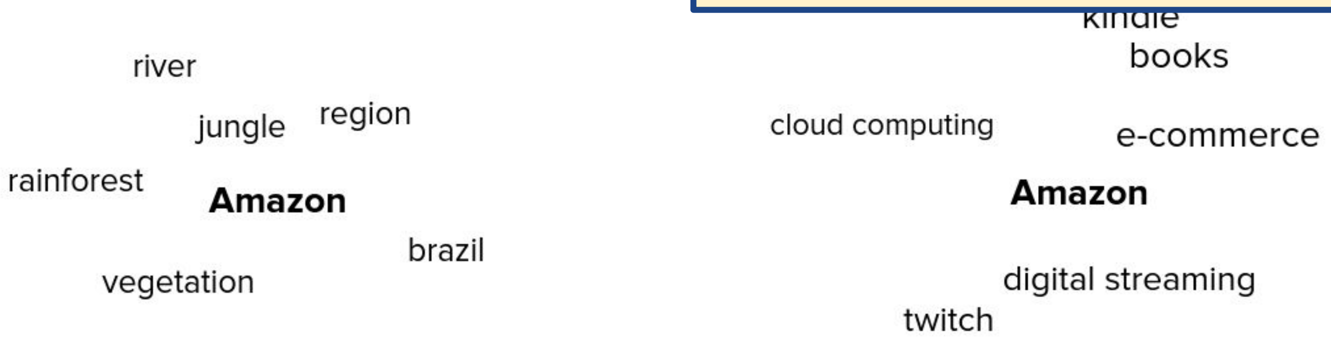
👉 [EMNLP-2022 EvoNLP workshop] (including TempoWiC shared task)

(4) Language is dynamic

More questions:

- What about Entity Linking?
- What is the ideal sense inventory? *WordNet? Wikidata? BabelNet? Public social media accounts?*

Meaning changes over time



More problematic in social media!

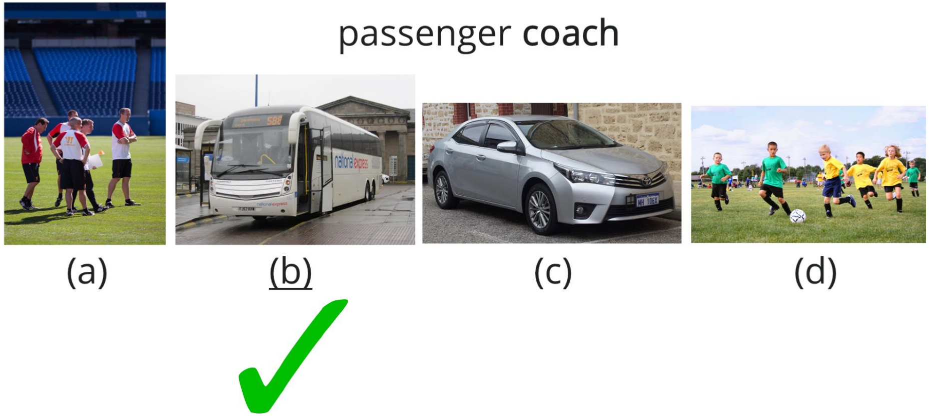
👉 [EMNLP-2022 EvoNLP workshop]() (including TempoWiC shared task)

(5) Multimodality (images?)

Visual Word Sense Disambiguation task (SemEval 2023)

Open for submissions until Jan 31st!

Data in English, Farsi and Italian



Conclusion

Language models represent a powerful tool to deal with lexical ambiguity, but many **challenges remain**.

Often fine-tuning not necessary: **contextualized embeddings are flexible and robust**.

But… is WSD still relevant in the language model era?

Conclusion

Language models represent a powerful tool to deal with lexical ambiguity, but many **challenges remain**.

Often fine-tuning not necessary: **contextualized embeddings are flexible and robust**.

But… is WSD still relevant in the language model era?

Yes! Added *interpretability*, *extra-info* from resources, *multilinguality* for free, needed for *retrieval*…

¡Gracias!

Thank you!

Eskerrik asko!

Questions? 🤔

✉ camachocolladosj@cardiff.ac.uk

@CamachoCollados

josecamachocollados.com