Computational methods for Lexical Semantic Change Detection

Pierluigi Cassotti, pierluigi.cassotti@gu.se

Hello!

Pierluigi Cassotti

Researcher @ Change is Key!
Department of Philosophy, Linguistics and Theory of Science
University of Gothenburg (Sweden)
pierluigi.cassotti@gu.se

Outline

- Computational Modeling Lexical Semantics
  - Synchronic Modeling
  - Diachronic Modeling
- Human annotation of Lexical Semantic Change
- Diachronic Models of Language
  - Static Models and Alignment
  - Contextualized Models
  - Generative Models
- Hands-on

Outline

- Computational Modeling Lexical Semantics
  - Synchronic Modeling
  - Diachronic Modeling
- Human annotation of Lexical Semantic Change
- Diachronic Models of Language
  - **Static Models and Alignment**
  - Contextualized Models
  - Generative Models
- Hands-on

Models for Lexical Semantic Change Detection

Static vs Contextualized representations

Static Embedding

- Easily to be trained on specific (historical) corpus
- Produce one space for each period, spaces need to be aligned

Contextualized Embeddings

- Trained on large collection of documents
- Not easy to specialize on specific corpus (resources required, catastrophic forgetting)
- Usually use pretrained vectors

Static vs Contextualized representations

Historical Corpora

Training

Static Embedding

Big data - Billions of web pages

Pretraining

Contextualized Embeddings

Static vs Contextualized representations

Historical Corpora

Big data - Trillion of web pages

Training

Static Embedding

Inference

Pretraining

Contextualized Embeddings

Static vs Contextualized representations

Static Embedding

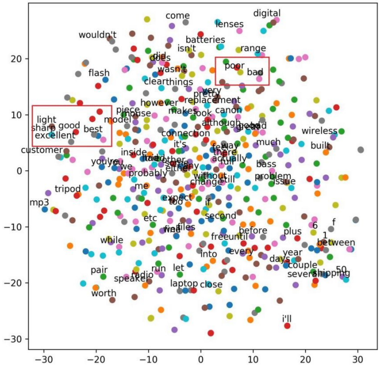
Collapse word semantics in a single point in the space, in order to compare semantics over time you need different vector spaces over time

Contextualized Embeddings

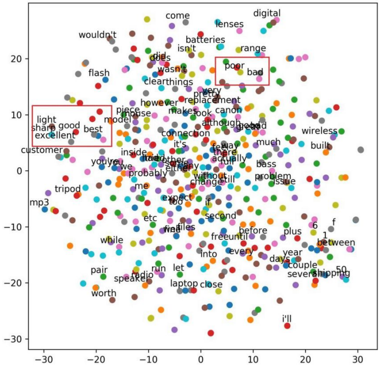
One vector for each usage of the word. You can then distinguish vectors computed for word usages coming from specific period

Static vs Contextualized representations

Static Embedding

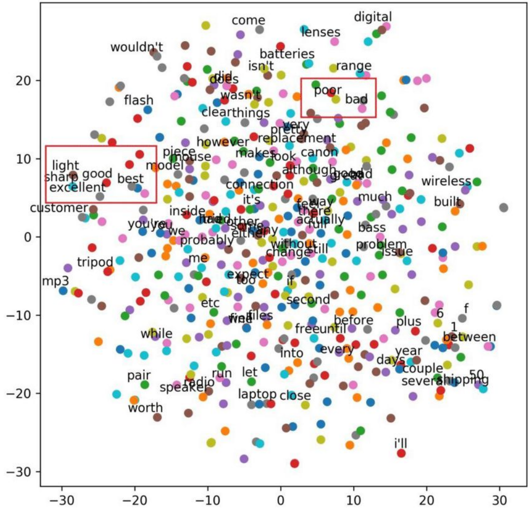


Time 1



Time 2

Contextualized Embedding



All at once

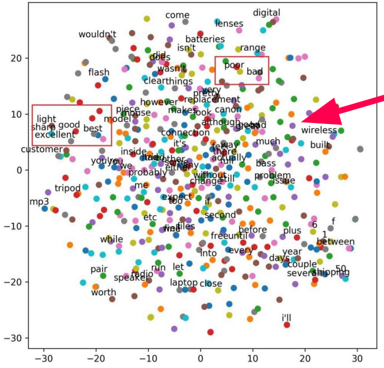
Static vs Contextualized representations

1836	If the sun's rays be parallel to any plane, that plane to which they are parallel, is called a plane of shade.
1836	its angles upon a given point A, in the plane, on which the ichnography is to be described;
1836	will be no difference between the shadow on the plane, and the side of the prism which projects that shadow;
1853	There are other kinds of planes besides the above; as the plough, for sinking a groove to receive a projecting tongue;

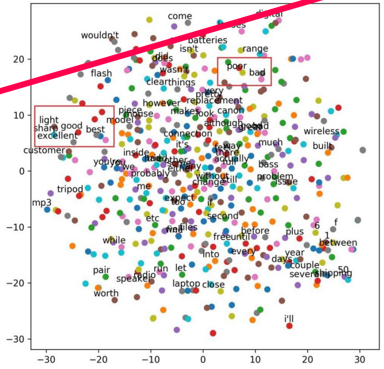
2003	Troy turned it to the right, and the plane turned to the right, just
1999	They had been making good progress, in spite of their greenness; next day Mr. Fulton was planning to stretch the silk over the planes;
1990	In the meantime, most of the troops and 25% of the supplies flying to Saudi Arabia are traveling on wide- body planes leased from commercial airlines.
2006	Reduction is only needed in patients near skeletal maturity whose fracture has more than 50-70 degrees of angulation in either the sagittal or coronal plane (Rab &; Grottkau, 2001).

Static vs Contextualized representations

Static Embedding



Time 1



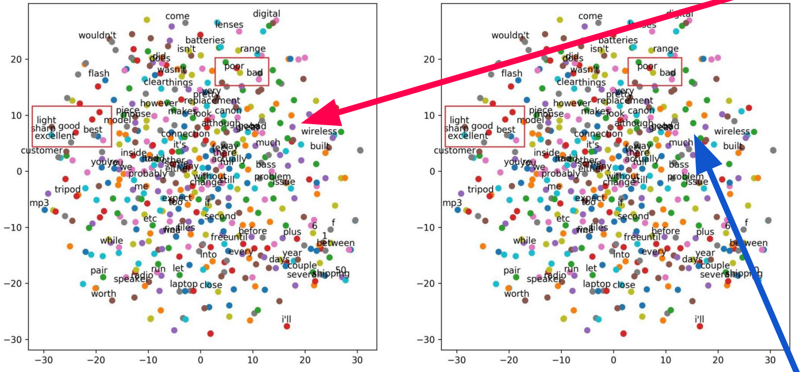
Time 2

1836	If the sun's rays be parallel to any plane, that plane to which they are parallel, is called a plane of shade.
1836	its angles upon a given point A, in the plane, on which the ichnography is to be described;
1836	will be no difference between the shadow on the plane, and the side of the prism which projects that shadow;
1853	There are other kinds of planes besides the above; as the plough, for sinking a groove to receive a projecting tongue;

2003	Troy turned it to the right, and the plane turned to the right, just
1999	They had been making good progress, in spite of their greenness; next day Mr. Fulton was planning to stretch the silk over the planes;
1990	In the meantime, most of the troops and 25% of the supplies flying to Saudi Arabia are traveling on wide- body planes leased from commercial airlines.
2006	Reduction is only needed in patients near skeletal maturity whose fracture has more than 50-70 degrees of angulation in either the sagittal or coronal plane (Rab &; Grottkau, 2001).

Static vs Contextualized representations

Static Embedding

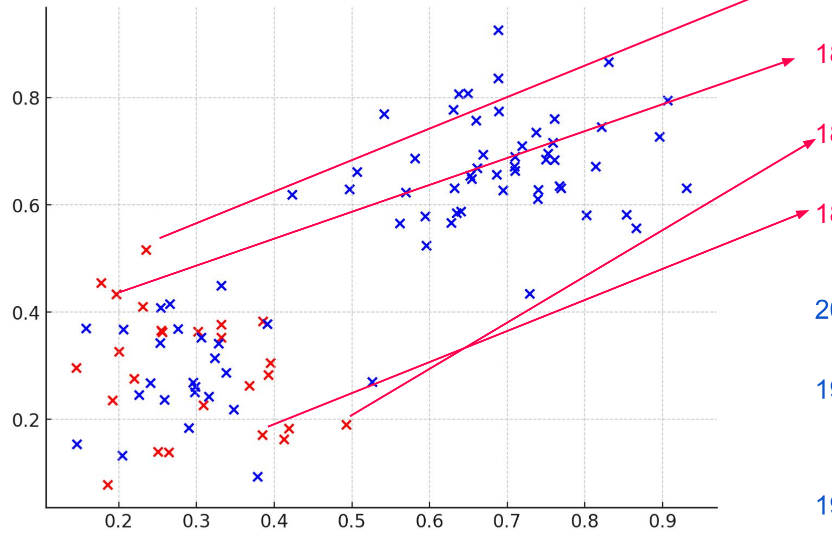


Time 1 Time 2

1836	If the sun's rays be parallel to any plane, that plane to which they are parallel, is called a plane of shade.
1836	its angles upon a given point A, in the plane, on which the ichnography is to be described;
1836	will be no difference between the shadow on the plane, and the side of the prism which projects that shadow;
1853	There are other kinds of planes besides the above; as the plough, for sinking a groove to receive a projecting tongue;

2003	Troy turned it to the right, and the plane turned to the right, just
1999	They had been making good progress, in spite of their greenness; next day Mr. Fulton was planning to stretch the silk over the planes;
1990	In the meantime, most of the troops and 25% of the supplies flying to Saudi Arabia are traveling on wide- body planes leased from commercial airlines.
2006	Reduction is only needed in patients near skeletal maturity whose fracture has more than 50-70 degrees of angulation in either the sagittal or coronal plane (Rab &; Grottkau, 2001).

Static vs Contextualized representations



1836 If the sun's rays be parallel to any plane, that plane to which they are parallel, is called a plane of shade.

1836 its angles upon a given point A, in the plane, on which the ichnography is to be described;

1836 will be no difference between the shadow on the plane, and the side of the prism which projects that shadow;

1853 There are other kinds of planes besides the above; as the plough, for sinking a groove to receive a projecting tongue;

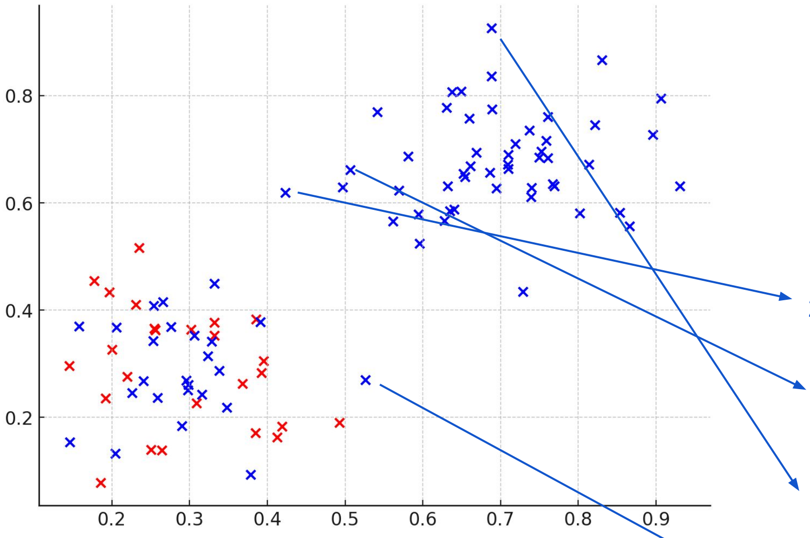
2003 Troy turned it to the right, and the plane turned to the right, just

1999 They had been making good progress, in spite of their greenness; next day Mr. Fulton was planning to stretch the silk over the planes;

1990 In the meantime, most of the troops and 25% of the supplies flying to Saudi Arabia are traveling on wide- body planes leased from commercial airlines.

2006 Reduction is only needed in patients near skeletal maturity whose fracture has more than 50-70 degrees of angulation in either the sagittal or coronal plane (Rab &; Grottkau, 2001).

Static vs Contextualized representations



1836 If the sun's rays be parallel to any plane, that plane to which they are parallel, is called a plane of shade.

1836 its angles upon a given point A, in the plane, on which the ichnography is to be described;

1836 will be no difference between the shadow on the plane, and the side of the prism which projects that shadow;

1853 There are other kinds of planes besides the above; as the plough, for sinking a groove to receive a projecting tongue;

2003 Troy turned it to the right, and the plane turned to the right, just

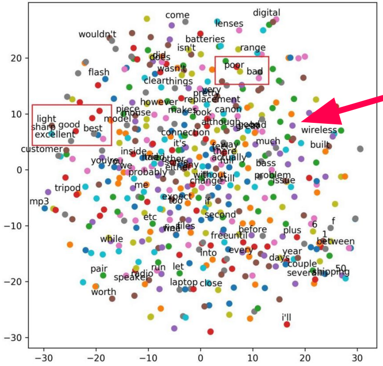
1999 They had been making good progress, in spite of their greenness; next day Mr. Fulton was planning to stretch the silk over the planes;

1990 In the meantime, most of the troops and 25% of the supplies flying to Saudi Arabia are traveling on wide- body planes leased from commercial airlines.

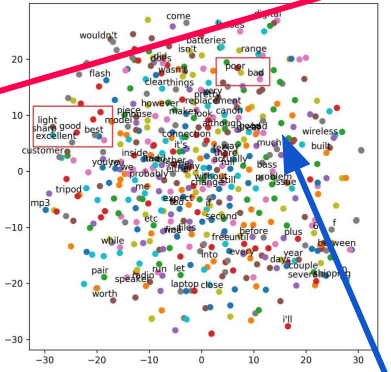
2006 Reduction is only needed in patients near skeletal maturity whose fracture has more than 50-70 degrees of angulation in either the sagittal or coronal plane (Rab &; Grottkau, 2001).

Static vs Contextualized representations

Static Embedding



Time 1

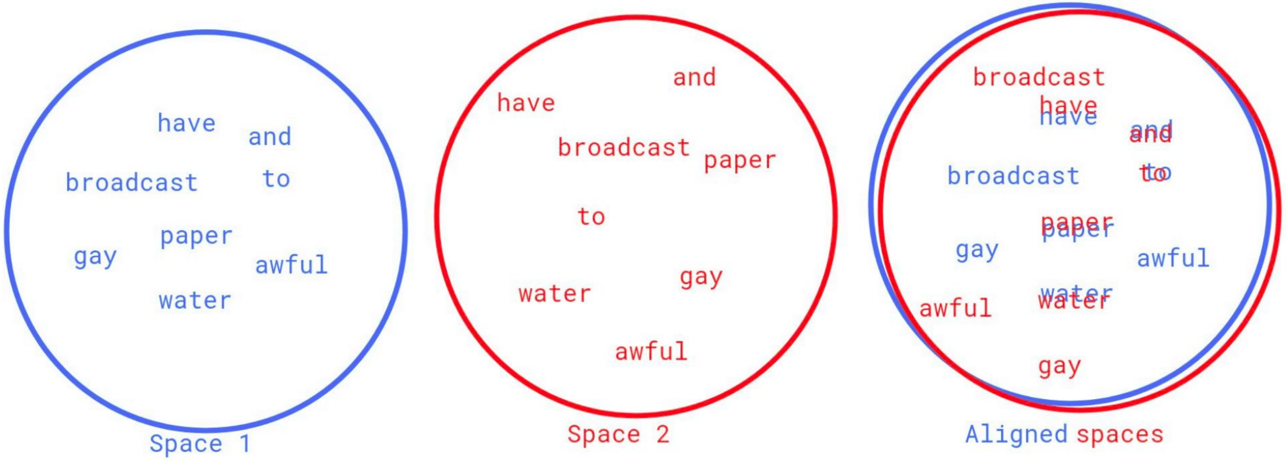


Time 2

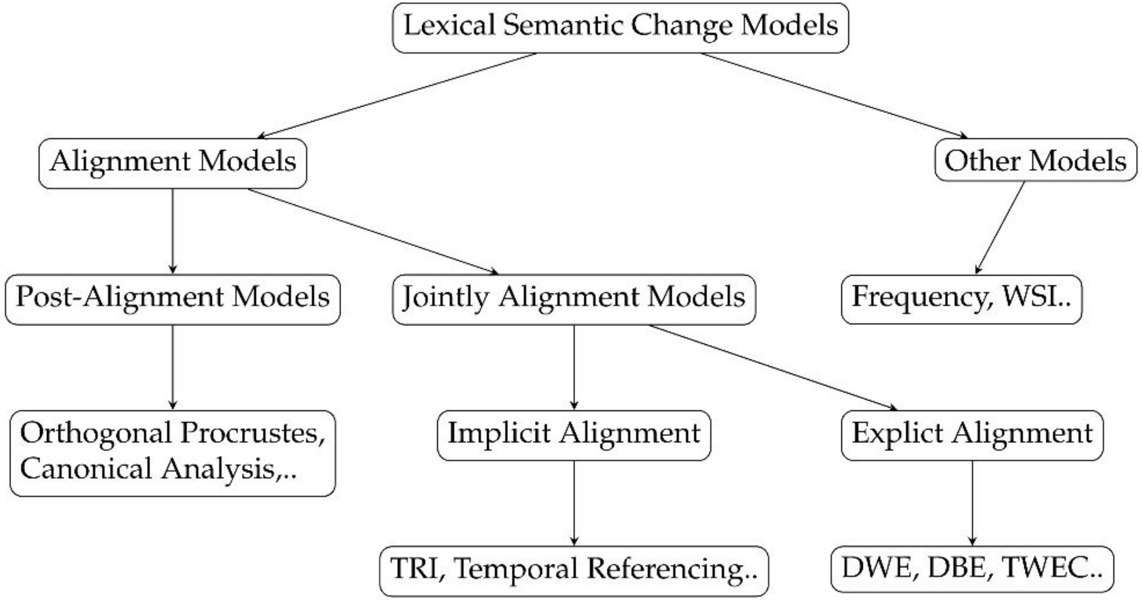
1836	If the sun's rays be parallel to any plane, that plane to which they are parallel, is called a plane of shade.
1836	its angles upon a given point A, in the plane, on which the ichnography is to be described;
1836	will be no difference between the shadow on the plane, and the side of the prism which projects that shadow;
1853	There are other kinds of planes besides the above; as the plough, for sinking a groove to receive a projecting tongue;

2003	Troy turned it to the right, and the plane turned to the right, just
1999	They had been making good progress, in spite of their greenness; next day Mr. Fulton was planning to stretch the silk over the planes;
1990	In the meantime, most of the troops and 25% of the supplies flying to Saudi Arabia are traveling on wide- body planes leased from commercial airlines.
2006	Reduction is only needed in patients near skeletal maturity whose fracture has more than 50-70 degrees of angulation in either the sagittal or coronal plane (Rab &; Grottkau, 2001).

Comparing vector spaces



Lexical Semantic Change Models



P. Cassotti, P. Basile, M. de Gemmis, and G. Semeraro, “Analyzing Gaussian distribution of semantic shifts in Lexical Semantic Change Models,” IJCoL Ital. J. Comput. Linguist., vol. 6, no. 6–2, pp. 23–36, 2020.

Alignment Models

Alignment approach

Post-alignment

- Post-alignment models first train static word embeddings for each time slice and then align them

Jointly alignment

- Jointly Alignment models train word embeddings and jointly align vectors across all time slices

- Jointly Alignment models can be distinguished in Explicit alignment models and Implicit alignment models.

Alignment approach

Post-alignment

- Post-alignment models first train static word embeddings for each time slice and then align them

Jointly alignment

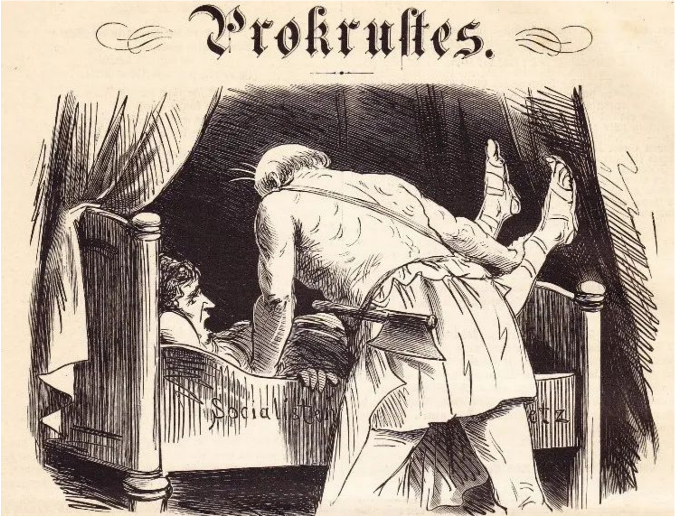
- Jointly Alignment models train word embeddings and jointly align vectors across all time slices

- Jointly Alignment models can be distinguished in Explicit alignment models and Implicit alignment models.

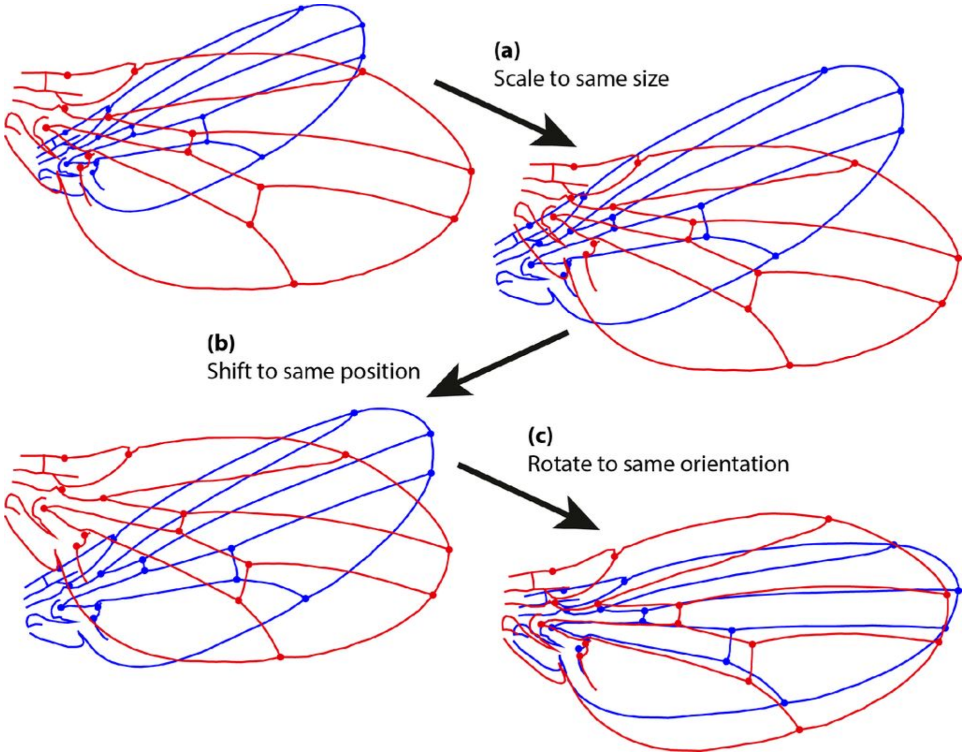
!! **Post-alignment and Explicit alignment rely on the assumption that only few words change their meaning**

Orthogonal Procrustes (OP)

Procrustes analysis is a form of statistical shape analysis used to analyse the distribution of a set of shapes. The name *Procrustes* (Greek: Προκρούστης) refers to a bandit from Greek mythology who made his victims fit his bed either by stretching their limbs or cutting them off



Orthogonal Procrustes (OP)



Orthogonal Procrustes (OP)

Embedding matrix time t

$$R = \arg\min_{Q^T Q = I} ||QW^t - W^{t+1}||_F$$

Rotation matrix

Embedding matrix time t+1

William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. [Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change](). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501, Berlin, Germany. Association for Computational Linguistics.

Outline

- Computational Modeling Lexical Semantics
  - Synchronic Modeling
  - Diachronic Modeling

- Human annotation of Lexical Semantic Change

- Diachronic Models of Language
  - Static Models and Alignment
  - **Contextualized Models**
  - Generative Models

- Hands-on

Contextualized Models

TempoBERT

- Use time as additional context
- Exploit time masking

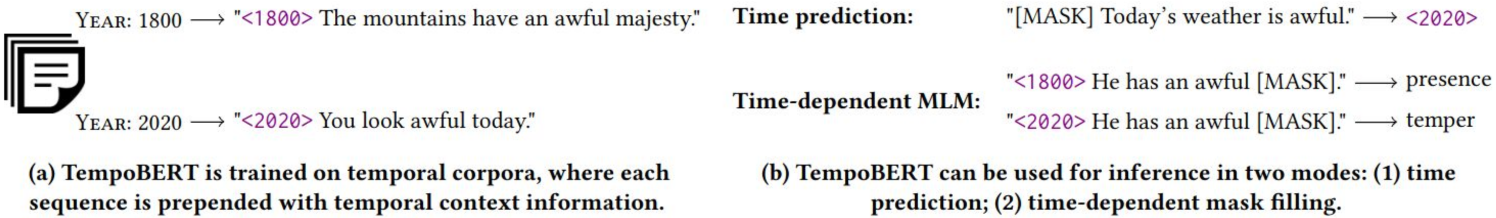


Figure 1: Example of TempoBERT's time masking for training and inference. The word 'awful' changed its meaning in the last two centuries from marvelous to disgusting.

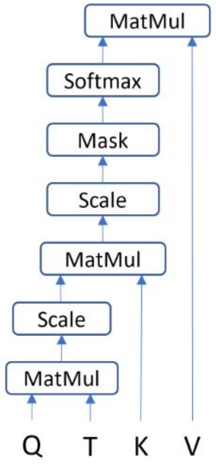
Rosin, Guy D., Ido Guy, and Kira Radinsky. "Time masking for temporal language models." *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. 2022.

Temporal Attention

- Extends self-attention to include time dimension

$$TemporalAttention(Q, K, V, T) = \mathrm{softmax}\left(\frac{Q\frac{T^{\intercal}T}{\|T\|}K^{\intercal}}{\sqrt{d_k}}\right)V$$

Time-specific weight matrix



Guy D. Rosin and Kira Radinsky. 2022. Temporal Attention for Language Models. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 1498–1508, Seattle, United States. Association for Computational Linguistics.

XLM-RoBERTa

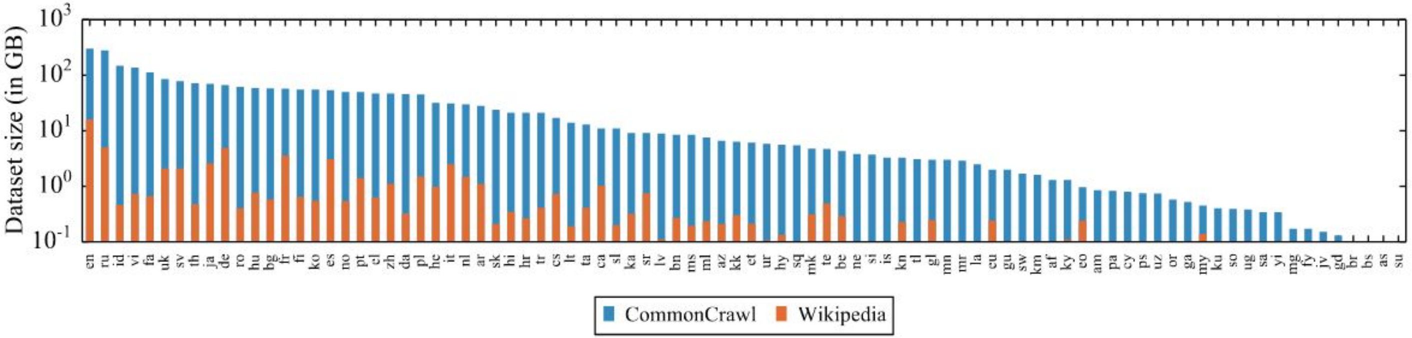


Figure 1: Amount of data in GiB (log-scale) for the 88 languages that appear in both the Wiki-100 corpus used for mBERT and XLM-100, and the CC-100 used for XLM-R. CC-100 increases the amount of data by several orders of magnitude, in particular for low-resource languages.

Sebastian Ruder, Anders Søgaard, and Ivan Vulić. 2019. [Unsupervised Cross-Lingual Representation Learning](). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, pages 31–38, Florence, Italy. Association for Computational Linguistics.

Gloss Reader

- Rely on XLM-RoBERTa and trained on a English Word Sense Disambiguation (WSD) dataset (SemCor)
- Zero-shot ability on other languages such as Russian

The bank can guarantee deposits will eventually cover future tuition costs because it invests in adjustable-rate mortgage securities.

Context Encoder

bank[1]	Gloss:	a financial institution that accepts deposits and channels the money into lending activities

bank[2]	Gloss:	sloping land (especially the slope beside a body of water)

Gloss Encoder

Rachinskiy, Maxim, and Nikolay Arefyev. "Zeroshot Crosslingual Transfer of a Gloss Language Model for Semantic Change Detection." *Computational linguistics and intellectual technologies: Papers from the annual conference Dialogue*. 2021.

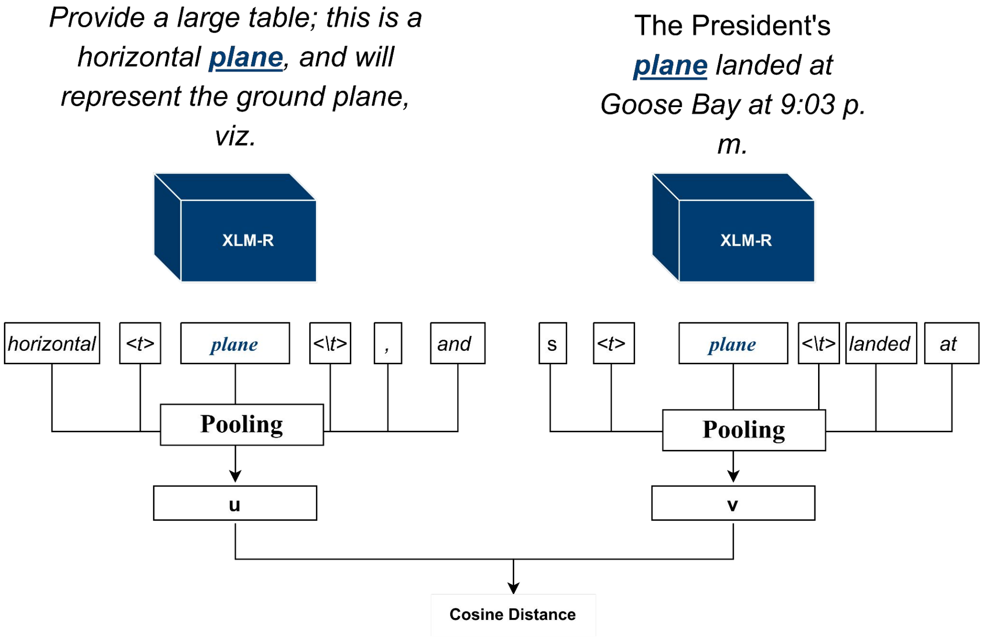
Deep Mistake

- Pretrained XLM-R fintuned on MCL-WiC task
- Not depends on fixed sense inventories

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 v̄ aljaspool aega ja ruumi.	Umberringi oli l̄ oputu t̄ 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

Arefyev, Nikolay, et al. "DeepMistake: Which Senses are Hard to Distinguish for a WordinContext Model." *Computational linguistics and intellectual technologies: Papers from the annual conference Dialogue*. 2021.

XL-LEXEME



Pierluigi Cassotti, Lucia Siciliani, Marco DeGemmis, Giovanni Semeraro, and Pierpaolo Basile. 2023. [XL-LEXEME: WiC Pretrained Model for Cross-Lingual LEXical sEMantic changE](). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1577–1585, Toronto, Canada. Association for Computational Linguistics.

XL-LEXEME

Dataset	Languages
WiC Pilehvar et al., (2019	**Monolingual** EN
XL-WiC (Raganato et al., 2020)	**Multilingual** EN, BG, ZH, HR, DA, NL, ET, FA, FR, DE, IT, JA, KO
MCL-WiC (Martelli et al., 2021)	**Multilingual** **EN, AR, FR, RU, ZH**
	Crosslingual **AR, FR, RU, ZH**
AM^{2}ICO (Liu et al., 2021)	**Crosslingual** EN, DE, RU, JA, KO, ZH,AR, IN, FI, TR, EU, KA, UR, BN, KK

Pierluigi Cassotti, Lucia Siciliani, Marco DeGemmis, Giovanni Semeraro, and Pierpaolo Basile. 2023. XL-LEXEME: WiC Pretrained Model for Cross-Lingual LEXical sEMantic changE. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1577–1585, Toronto, Canada. Association for Computational Linguistics.

XL-LEXEME

			EN	LA	DE	SV	ES	RU			NO		ZH	Avg_w
			$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_1 - C_2$	$C_2 - C_3$	$C_1 - C_3$	$C_1 - C_2$	$C_2 - C_3$	$C_1 - C_2$	$C_i - C_j$
form-based	APD	BERT	.563	-	.271	.270	.335	.518	.482	.416	.441	.466	.656	.449
		mBERT	.363	.102	.398	.389	.341	.368	.345	.386	.279	.488	.689	.371
		XLM-R	.444	.151	.264	.257	.386	.290	.287	.318	.195	.379	.500	.316
		XL-LEXEME	**.886***	**.231**	**.839***	**.812***	**.665***	**.796***	**.820***	**.863***	**.659**	**.640***	**.731***	**.751***
		SOTA: sup.	*.757*	*-.056*	*.877*	*.754*	*n.a.*	*.799*	*.833*	*.842*	*.757*	*.757*	*n.a.*	
		SOTA: uns.	*.706*	*.443*	*.731*	*.602*	*n.a.*	*.372*	*.480*	*.457*	*.389*	*.387*	*n.a.*	
	PRT	BERT	.457	-	.422	.158	.413	.400	.374	.347	.507	.444	**.712**	.406
		mBERT	.270	.380	.436	.193	.543	.391	.356	.423	.219	.438	.524	.395
		XLM-R	.411	.424	.369	.020	.505	.321	.443	.405	.387	.149	.558	.381
		XL-LEXEME	**.676**	**.506***	**.824**	**.696**	**.632**	**.704**	**.750**	**.727**	**.764***	**.519**	.699	**.693**
		SOTA: sup.	*.531*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	
		SOTA: uns.	*.467*	*.561*	*.755*	*.392*	*n.a.*	*.294*	*313*	*313*	*.378*	*.270*	*n.a.*	
sense-based	AP+JSD	BERT	.289	-	.469	- .090	.225	.069	.279	.094	**.314**	.011	.165	.179
		mBERT	.181	.277	.280	.023	.067	.017	.086	- .116	.035	- .090	.465	.077
		XLM-R	.278	**.398**	.224	-.076	.224	- .068	**.209**	**.130**	- .100	.030	.448	.142
		XL-LEXEME	**.493**	.033	**.499**	**.118**	**.392**	**.106**	.053	.117	.297	**.381**	**.308**	**.223**
		SOTA: sup.	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	
		SOTA: uns.	*.436*	*.481*	*.583*	*.343*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	
	WiDiD	BERT	.385	-	.355	.106	.383	.135	.102	.243	.233	.087	.533	.239
		mBERT	.323	- .039	.312	.195	.343	- .068	.160	.142	.241	.290	.338	.181
		XLM-R	.564	- .064	.499	.129	.459	**.268**	.216	.342	.226	.349	.382	.314
		XL-LEXEME	**.652**	**.236**	**.677**	**.475**	**.522**	.178	**.354**	**.364**	**.561**	**.457**	**.563**	**.422**
		SOTA: sup.	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	
		SOTA: uns.	*.651*	*-.096*	*.527*	*.499*	*.544*	*.273*	*.393*	*.407*	*n.a.*	*n.a.*	*n.a.*	

Periti, F., & Tahmasebi, N. (2024). A Systematic Comparison of Contextualized Word Embeddings for Lexical Semantic Change.

XL-LEXEME

		EN	DE	SV	ES	RU			NO		ZH	Avg$_w$
		C_1-C_2	C_1-C_2	C_1-C_2	C_1-C_2	C_1-C_2	C_2-C_3	C_1-C_3	C_1-C_2	C_2-C_3	C_1-C_2	C_i-C_j
WiC	BERT	.503	.350	.221	.319	.314	.344	.350	.429	.406	.516	.358
	mBERT	.332	.344	.284	.289	.280	.273	.293	.283	.333	.413	.301
	XLM-R	.352	.289	.255	.288	.212	.250	.251	.317	.261	.392	.272
	XL-LEXEME	**.626**	**.628**	**.631**	**.547**	**.549**	**.558**	**.564**	**.484**	**.521**	**.630**	**.568**
	GPT-4.0	.606	-	-	-	-	-	-	-	-	-	-
	Agreement	*.633*	*.666*	*.672*	*.531*	*.531*	*.567*	*.564*	*.761*	*.667*	*.602*	*.593*
WSI	BERT	.136 / .700	.047 / .662	.023 / .596	.189 / .695	- / -	- / -	- / -	.251 / .771	.247 / .758	.279 / .759	.166 / .702
	mBERT	.067 / .644	.054 / .679	.024 / .648	.228 / .700	- / -	- / -	- / -	.241 / .759	.159 / .753	.172 / .713	.146 / .696
	XLM-R	.068 / .737	.024 / .725	.031 / .680	.164 / .755	- / -	- / -	- / -	.179 / .775	.183 / .715	.279 / .806	.133 / .743
	XL-LEXEME	.273 / .834	**.300 / .788**	**.249 / .766**	**.400 / .820**	- / -	- / -	- / -	**.337 / .806**	**.304 / .808**	**.448 / .836**	**.339 / .810**
	GPT-4.0	**.340 / .877**	- / -	- / -	- / -	- / -	- / -	- / -	- / -	- / -	- / -	- / -
GCD	BERT	.425	.116	.148	.284	.487	.452	.469	.571	.521	**.808**	.422
	mBERT	.120	.205	.234	.394	.372	.325	.408	.290	.454	.737	.357
	XLM-R	.219	.069	.143	.464	.284	.301	.375	.395	.345	.557	.324
	XL-LEXEME	.801	**.799**	**.721**	**.655**	**.780**	**.824**	**.851**	**.620**	**.567**	.716	**.754**
	GPT-4.0	**.818**	-	-	-	-	-	-	-	-	-	-

Periti, F., & Tahmasebi, N. (2024). A Systematic Comparison of Contextualized Word Embeddings for Lexical Semantic Change.

Outline

- Computational Modeling Lexical Semantics
  - Synchronic Modeling
  - Diachronic Modeling

- Human annotation of Lexical Semantic Change

- Diachronic Models of Language
  - Static Models and Alignment
  - Contextualized Models
  - **Generative Models**

- Hands-on

Generative Models

Definition Generation

FLAN T5

Mario Giulianelli, Iris Luden, Raquel Fernandez, and Andrey Kutuzov. 2023. Interpretable Word Sense Representations via Definition Generation: The Case of Semantic Change Analysis. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3130–3148, Toronto, Canada. Association for Computational Linguistics.

Definition Generation

Model	Test	WordNet BLEU	WordNet ROUGE-L	WordNet BERT-F1	Oxford BLEU	Oxford ROUGE-L	Oxford BERT-F1
Huang et al. (2021)	*Unknown*	32.72	-	-	**26.52**	-	-
Flan-T5 XL	Zero-shot (task shift)	2.70	12.72	86.72	2.88	16.20	86.52
Flan-T5 XL	In-distribution	11.49	28.96	88.90	16.61	36.27	89.40
Flan-T5 XL	Hard domain shift	29.55	48.17	91.39	8.37	25.06	87.56
Flan-T5 XL	Soft domain shift	**32.81**	**52.21**	**92.16**	18.69	**38.72**	**89.75**

Table 3: Results of the definition generation experiments.

Mario Giulianelli, Iris Luden, Raquel Fernandez, and Andrey Kutuzov. 2023. Interpretable Word Sense Representations via Definition Generation: The Case of Semantic Change Analysis. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3130–3148, Toronto, Canada. Association for Computational Linguistics.

Definition Generation

Usage example	Target word	Generated definition
'about half of the soldiers in our rifle platoons were **draftees** whom we had trained for about six weeks'	**draftee**	'A PERSON WHO IS BEING ENLISTED IN THE ARMED FORCES'

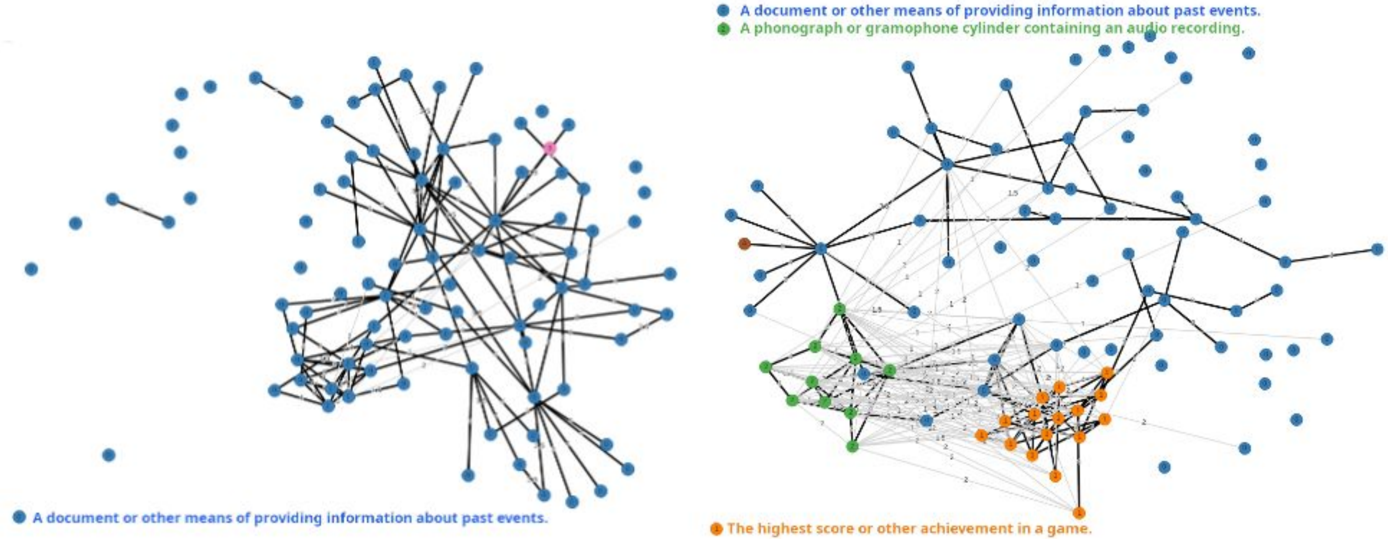
Table 1: An example of a definition generated by our fine-tuned Flan-T5 XL. The model is prompted with the usage example, post-fixed with the phrase *'What is the definition of draftee?'*

Method	**Cosine**	**SacreBLEU**	**METEOR**
Token embeddings	0.141	-	-
Sentence embeddings	0.114	-	-
Generated definitions			
Flan-T5 XL Zero-shot	0.188	0.041	0.083
Flan-T5 XXL Zero-shot	0.206	0.045	0.092
Flan-T5 base FT	0.221	0.078	0.077
Flan-T5 XL FT	**0.264**	**0.108**	**0.117**

Table 4: Correlations with pairwise similarity judgements by humans. 'FT' stands for 'fine-tuned model'.

Mario Giulianelli, Iris Luden, Raquel Fernandez, and Andrey Kutuzov. 2023. Interpretable Word Sense Representations via Definition Generation: The Case of Semantic Change Analysis. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3130–3148, Toronto, Canada. Association for Computational Linguistics.

Definition Generation



Mario Giulianelli, Iris Luden, Raquel Fernandez, and Andrey Kutuzov. 2023. [Interpretable Word Sense Representations via Definition Generation: The Case of Semantic Change Analysis](). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3130–3148, Toronto, Canada. Association for Computational Linguistics.

Substitution-based

	GEMS	SE Eng	SE Ger	SE Lat	SE Swe	Average	Average (weighted)
Number of words	96*	37	40	48	31		
Static Embedding Methods							
Pömsl and Lyapin (2020)	-	0.422	**0.725**	0.412	0.547	-	-
Montariol et al. (2021) [static]	0.347	0.321	0.712	0.372	**0.631**	0.477	0.452
Contextual Embedding Methods							
Martinc et al. (2020b)	0.510	0.313	0.436	0.467	-0.026	0.340	0.394
Montariol et al. (2021) [contextual]	0.352	0.437	0.561	0.488	0.321	0.432	0.422
Scaled JSD	**0.535**	**0.547**	0.563	**0.533**	0.310	**0.498**	**0.514**

Dallas Card. 2023. Substitution-based Semantic Change Detection using Contextual Embeddings. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 590–602, Toronto, Canada. Association for Computational Linguistics.

Substitution-based

Word	SE rating	SE rank	Scaled JSD	Scaled JSD rank	Corpus A substitutes (1810–1860)	Corpus B substitutes (1960–2010)
plane	0.88	1	0.97	1	plane line planes point surface lines	plane aircraft planes jet airplane car
graft	0.55	4	0.97	2	tree plant stock vine fruit wood	corruption bribery fraud crime violence
tip	0.68	2	0.85	7	tipped tip covered end filled tips give	tip tips end tipped edge point top ends
gas	0.16	23	0.72	14	gas gases vapor air fire water	gas gasoline oil gases fuel water air
head	0.30	10	0.68	16	head face hand heads hands eyes	head face heads hand body hands eyes
bit	0.31	9	0.51	23	bit piece sort little pieces bits kind	bit little lot touch tad piece bits pieces
fiction	0.02	35	0.41	27	fiction history literature art poetry	fiction fact fantasy story stories novels
tree	0.07	33	0.22	33	trees tree plants branches plant wood	trees tree plants woods branches bushes
ounce	0.28	11	0.08	37	ounce inch pounds hour acre dollars	ounce pounds inch inches cups pieces

Dallas Card. 2023. Substitution-based Semantic Change Detection using Contextual Embeddings. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 590–602, Toronto, Canada. Association for Computational Linguistics.

Substitution-based

	T1	T2
	remember that it be only such line as be nearer the ground **plane** than the eye that be draw under the horizon line	as his **plane** cross north carolina and head south over the atlantic it pick up a small convoy of escort military craft that try to make radio contact but fail
BERT	there, be, where, here, and	planes, over, out, boats, aircraft
XLM-R	line, rather, and, more, level	planes, crew, men, vehicles, team
LLaMa 2	level,surface,flat plane,horizontal plane	aircraft,airplane,jet,plane model,propeller-driven vehicle

Table 6: Generated substitutions for usages of **plane** extracted by SemEval 2020 Task 1 English.

Francesco Periti, Pierluigi Cassotti, Haim Dubossarsky, Nina Tahmasebi. 2024. Analyzing Semantic Change through Lexical Replacements. ACL 2024

Beyond Binary Lexical Semantic Change Detection

Using Synchronic Definitions and Semantic Relations to Classify Semantic Change Types

Pierluigi Cassotti[1], Stefano De Pascale[2], Nina Tahmasebi[1]
[1]University of Gothenburg
[2]VUB/FWO/KU Leuven

CHANGE IS KEY.ORG

Change is Key!
changeiskey.org

ACL 2024
Bangkok, Thailand

Synchronic Relations as a product of Semantic Change

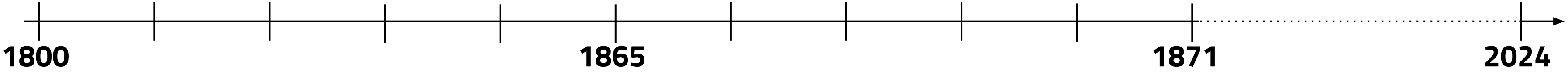
Horticulture

graft

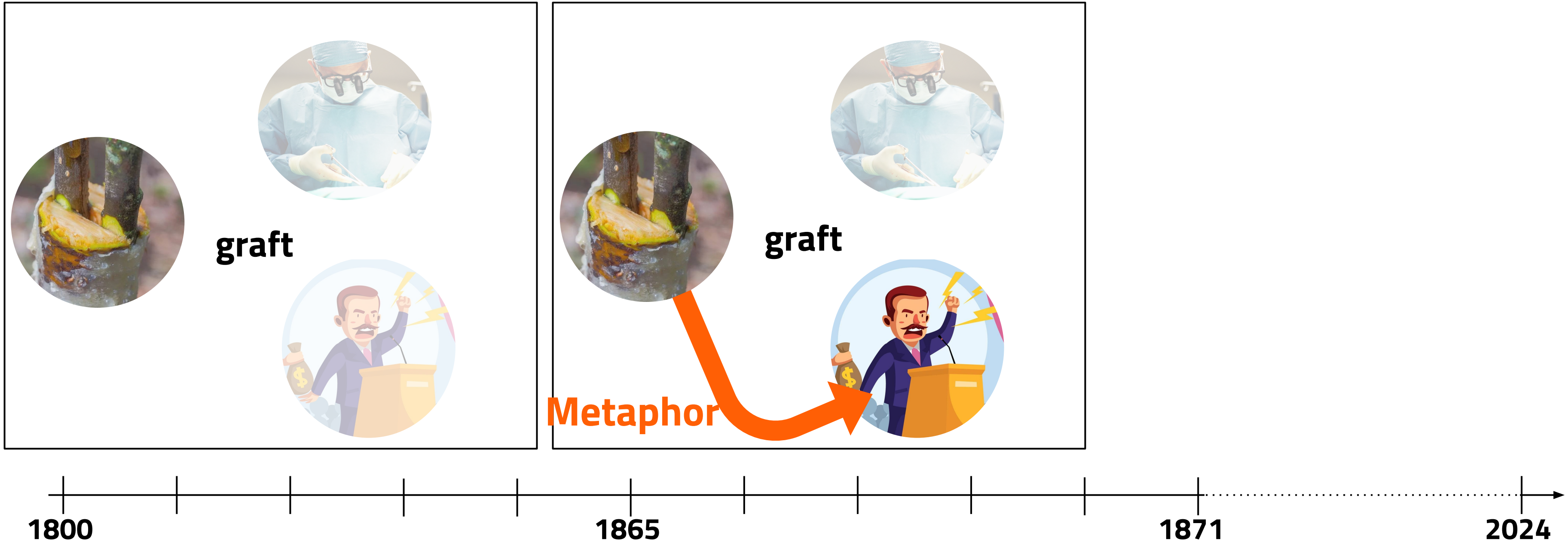
Medicine

Corruption

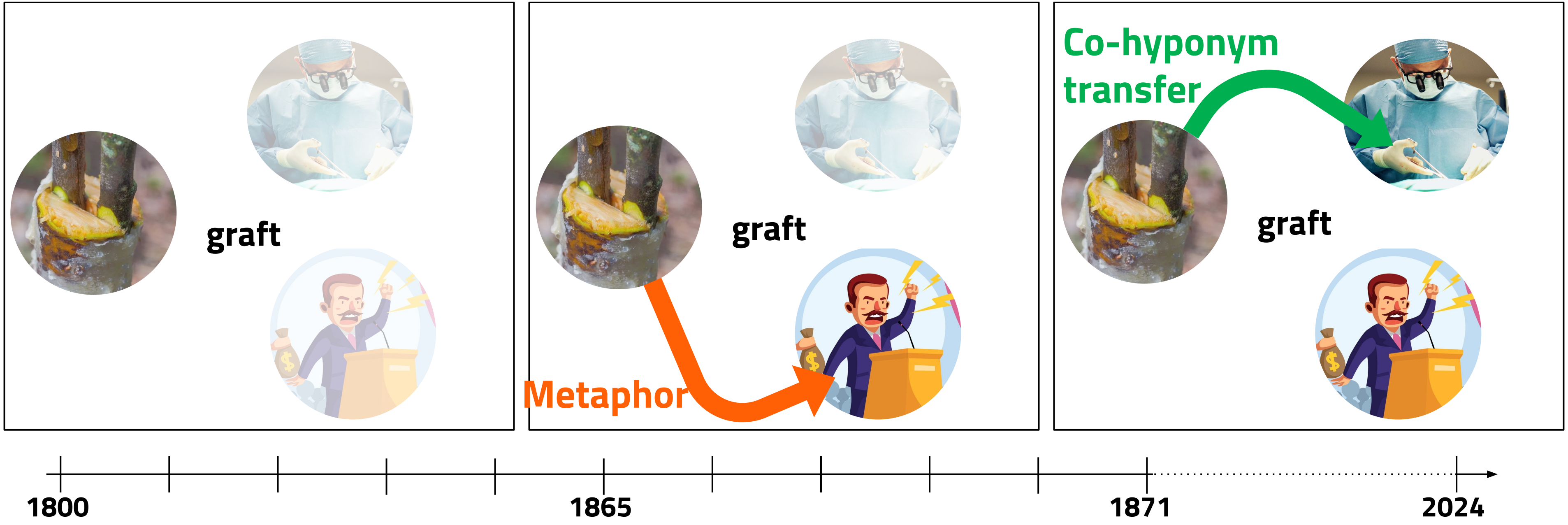
Synchronic Relations as a product of Semantic Change



Synchronic Relations as a product of Semantic Change



Synchronic Relations as a product of Semantic Change



Synchronic Relations as a product of Semantic Change

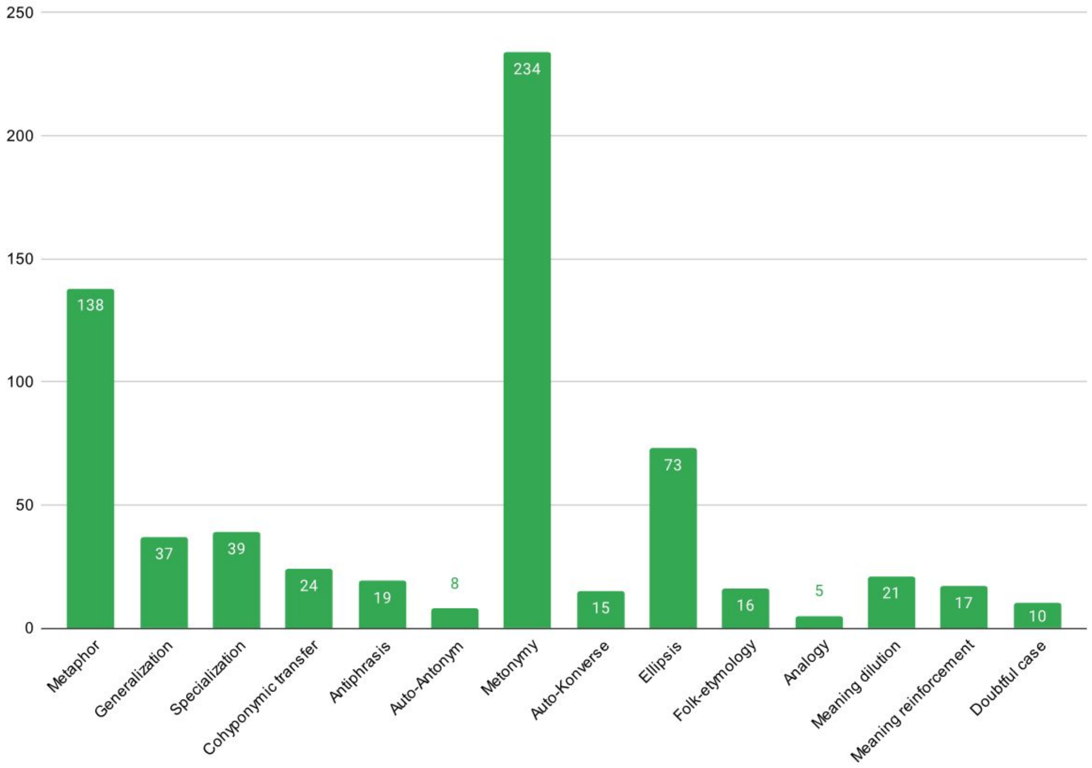
Horticulture

graft

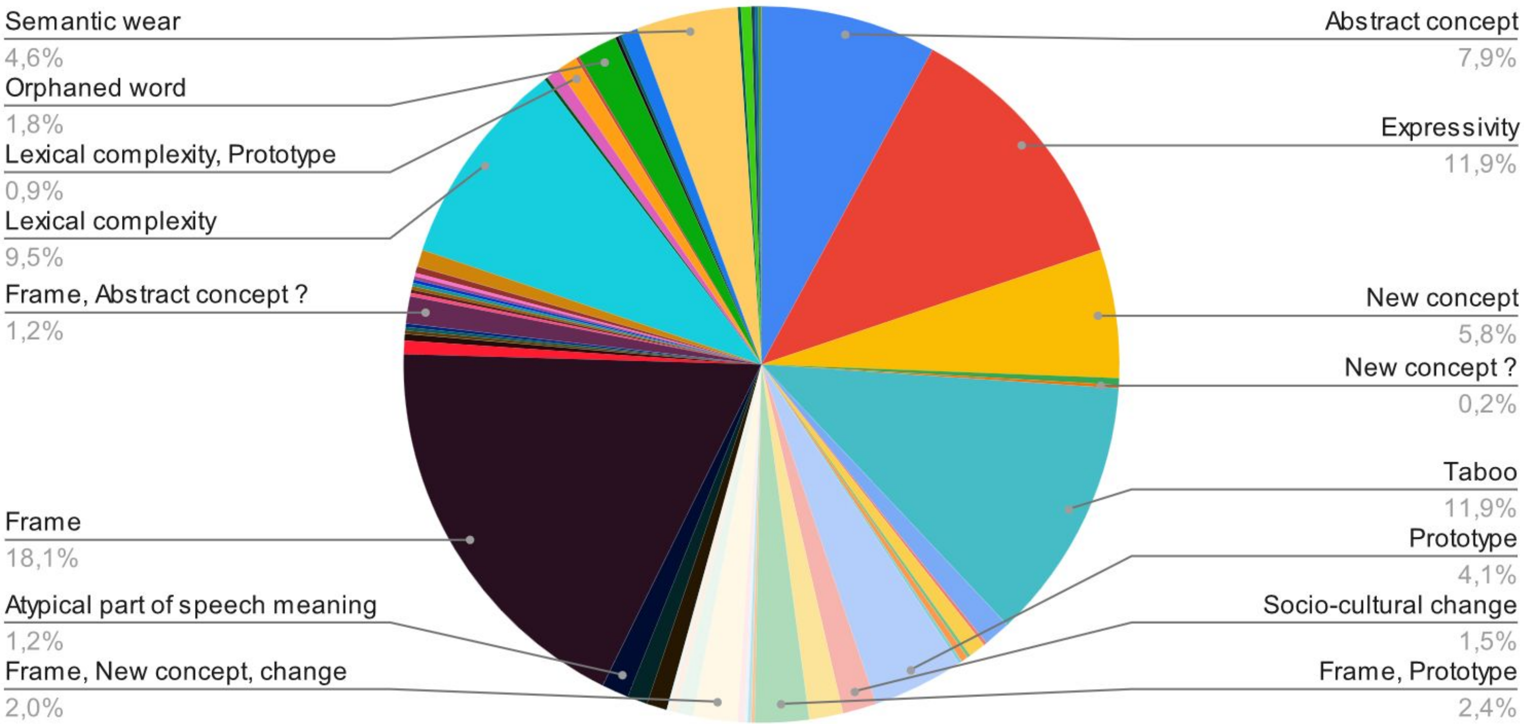
Medicine

Corruption

Blank's Taxonomy



Blank's Taxonomy



Blank's Taxonomy

Word	Old Meaning	New meaning	Cause	Type
**adripare:vlt*	am Ufer ankommen	ankommen	prototype / frame	generalization
necare:lt	töten	ertränken	socio-cultural change	specialization
**ratta*	Ratte	Maus	referential vagueness	co-hyponymous transfer
sacer:lt	heilig , geheiligt	verflucht	taboo	auto-antonymy

Definitions manually curated by an Historical Linguist aided by ChatGPT

- We translate Blank's glosses into English using Google Translate's API

- Following this, we prompt ChatGPT API, providing some examples we initially set up to generate definitions that are more akin to those found in a dictionary.

- Finally, a linguist with expertise in historical linguistics and semantics manually reviewed and refined the collection of generated translations and definitions, replacing or modifying them as necessary.

ChatGPT Prompting

Words	Old Meaning	New Meaning	Cause	Association Model	Notes	Type	Old Meaning (EN)	New Meaning (EN)	Old Meaning (EN Definition)	New Meaning (EN Definition)
afferrare:it	ergreifen	verstehen	abstraktes Konzept	SimiDES	KörperGeist	Metaphor	grab	understand	to quickly take…	to understand …
affreusement:fr	scheußlich	sehr, stark (Adv.)	Expressivität	SimiDES	Intensität	Metaphor	awful	very, strong (adv.)	extremely bad…	very much or in a…
agenda:sp	Notizbuch	NotebookComputer	neues Konzept	SimiDES	Gegenstand Gegenstand	Metaphor	Notebook	Notebook computer	a book of paper for…	a small, light computer
águila:sp	Adler	schlauer Mensch, Fuchs	Expressivität	SimiDES	TierMensch	Metaphor	Adler	Clever person, fox	a large, strong bird…	having or showing…
aile:fr; ala:it,sp,pt	Flügel	Flügel eines Gebäudes	neues Konzept	SimiDES	Tier Gegenstand	Metaphor	wing	Wing of a building	the flat part of the…	a part of a large…

Generate Old Meaning (EN Definition) and New Meaning (EN Definition) for the following entries, providing descriptive, dictionary-like definitions as in the examples above.

Blank's Taxonomy

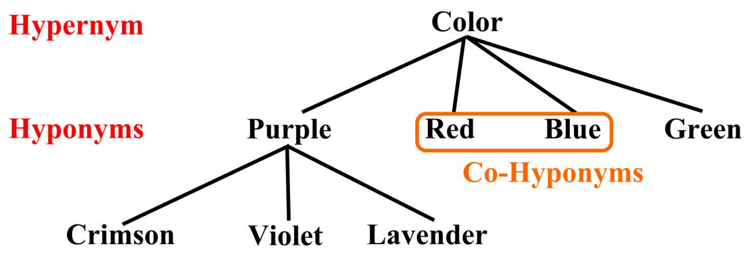
Word	Old Meaning	New meaning	Cause	Type
**adripare:vlt*	am Ufer ankommen	ankommen	prototype / frame	generalization
necare:lt	töten	ertränken	socio-cultural change	specialization
**ratta*	Ratte	Maus	referential vagueness	co-hyponymous transfer
sacer:lt	heilig , geheiligt	verflucht	taboo	auto-antonymy

Blank's Taxonomy + Definitions = Lexical Semantic Change Cause-Type-Definitions Benchmark

Word	Old Meaning	New meaning	Cause	Type
**adripare:vlt*	am Ufer ankommen "arrive at the bank/shore" 'arrive at the bank of a river or the shore of a lake or sea'	ankommen "arrive" 'to reach a place, especially at the end of a journey'	prototype / frame	generalization
necare:lt	töten "kill" 'to cause the death of a living thing, typically involving an act of violence or an intention to harm.'	ertränken "drown" 'to cause to die by submersion in liquid, especially by forcing the head under the water.'	socio-cultural change	specialization
**ratta*	Ratte "rat" 'a small rodent, larger than a mouse, that has a long tail and is considered to be harmful'	Maus "mouse" 'a small mammal with short fur, a pointed face, and a long tail'	referential vagueness	co-hyponymous transfer
sacer:lt	heilig , geheiligt "sacred" 'considered to be holy and deserving respect, especially because of a connection with a god'	verflucht "cursed" 'experiencing bad luck caused by a magic curse'	taboo	auto-antonymy

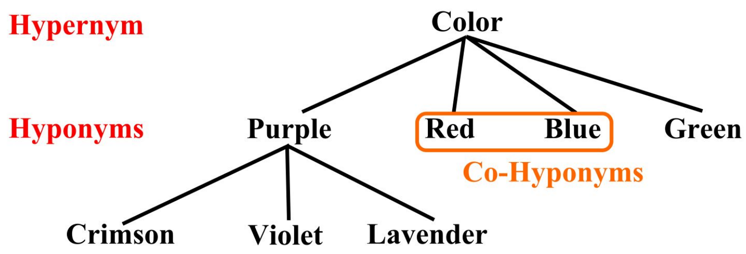
WordNet

- **Antonymy**: Opposite meanings, e.g. *hot* and *cold*
- **Hyponymy**: Sense more specific instance of a general category, e.g. *sparrow* is a hyponym of *bird*
- **Hypernymy**: A sense that is a general category of more specific instances, e.g. *bird* is a hypernym of *sparrow*
- **Co-hyponyms:** Senses that share the same hypernym, e.g. *cat* and *dog* are hyponym of “animal”

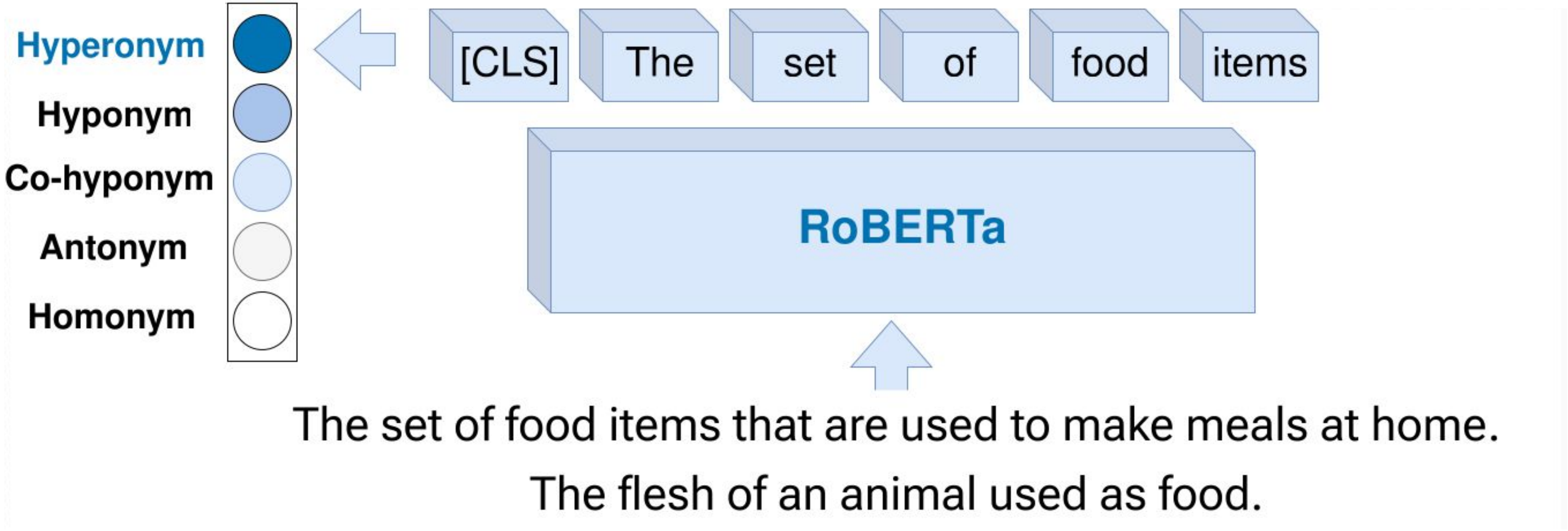


WordNet

- **Antonymy**: Opposite meanings, e.g. *hot* and *cold*
- **Hyponymy**: Sense more specific instance of a general category, e.g. *sparrow* is a hyponym of *bird*
- **Hypernymy**: A sense that is a general category of more specific instances, e.g. *bird* is a hypernym of *sparrow*
- **Co-hyponyms:** Senses that share the same hypernym, e.g. *cat* and *dog* are hyponym of “animal”
- **Homonymy:** We random sample pairs of senses to emulate homonymy

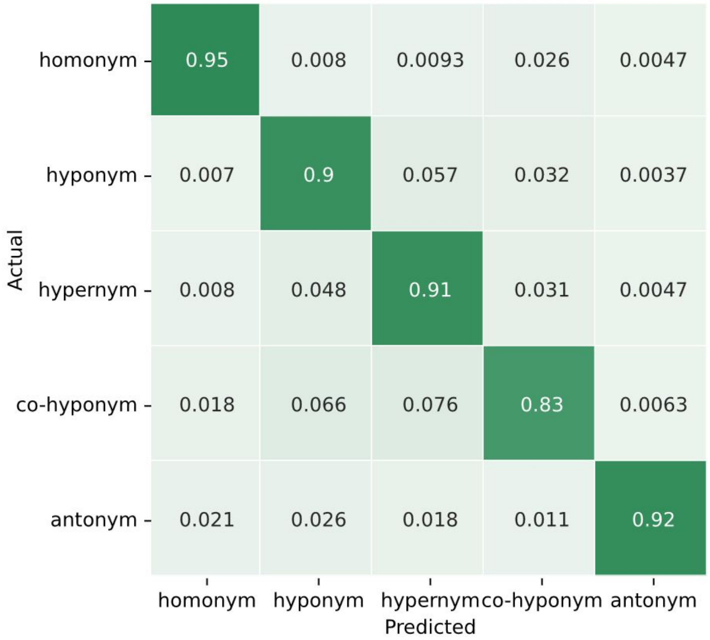


Classification



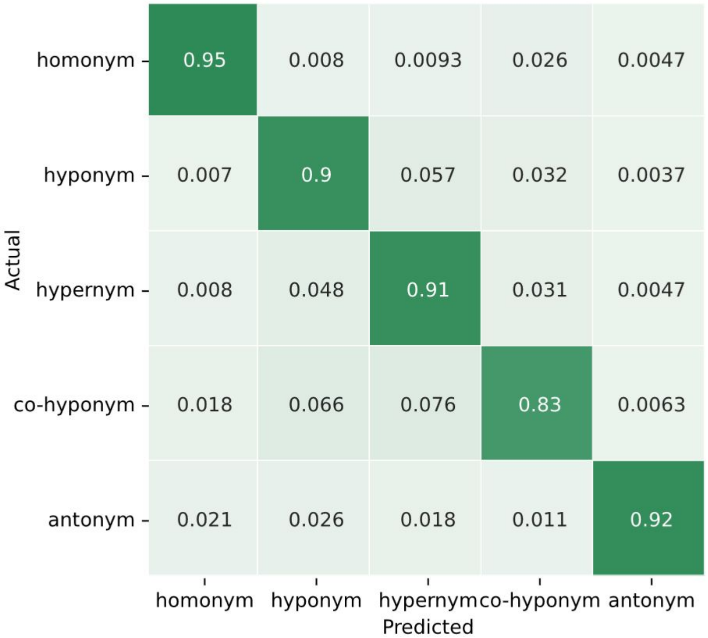
Results

WordNet Test set (Synchronic)

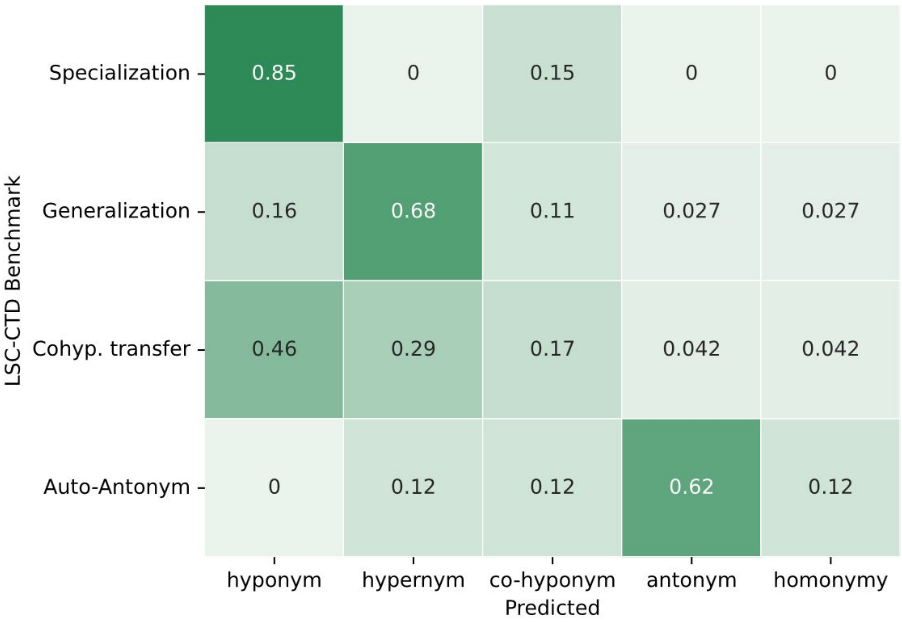


Results

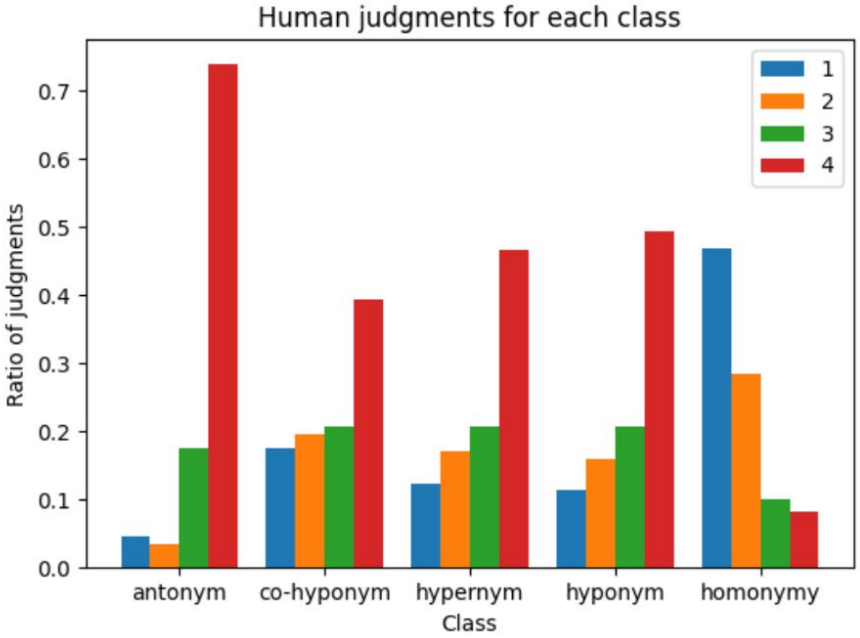
WordNet Test set (Synchronic)



LSC-CTD Benchmark (Diachronic)



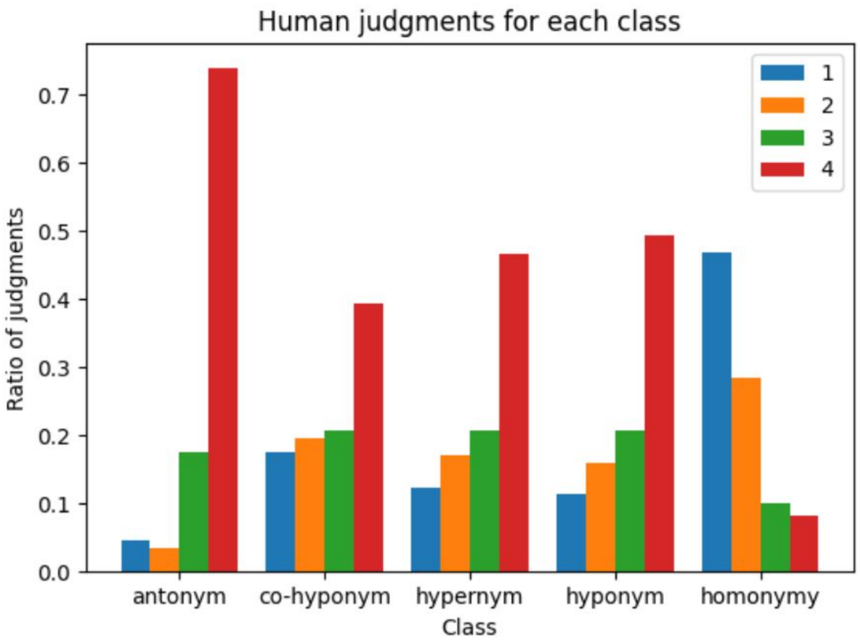
Results



Model	Correlation
Definitions + SacreBLEU	0.108
Definitions + METEOR	0.117
Definitions + Cosine similarity	0.264
Definitions + Homonym	0.472
XL-LEXEME	0.623
Definitions + Homonym + XL-LEXEME	**0.646**

Table 3: Spearman correlation of human judgments vs model predictions, Definitions generated using the method of Giulianelli et al. (2023)

Results

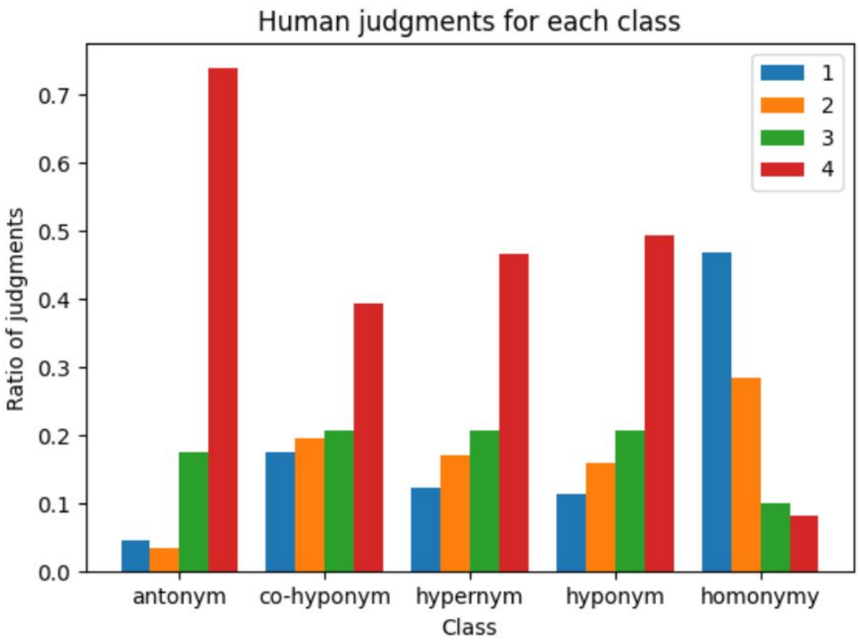


Model	Correlation
Definitions + SacreBLEU	0.108
Definitions + METEOR	0.117
Definitions + Cosine similarity	0.264
Definitions + Homonym	0.472
XL-LEXEME	0.623
Definitions + Homonym + XL-LEXEME	**0.646**

Table 3: Spearman correlation of human judgments vs model predictions, Definitions generated using the method of Giulianelli et al. (2023)

$$\rho(u_1, u_2) = \begin{cases} \cos(u_1, u_2), & \text{if } u_1, u_2 \textit{ Related.} \\ 0, & \text{otherwise.} \end{cases}$$

Results



Model	Correlation
Definitions + SacreBLEU	0.108
Definitions + METEOR	0.117
Definitions + Cosine similarity	0.264
Definitions + Homonym	0.472
XL-LEXEME	0.623
Definitions + Homonym + XL-LEXEME	**0.646**

Table 3: Spearman correlation of human judgments vs model predictions, Definitions generated using the method of Giulianelli et al. (2023)

Model	Accuracy
Definitions + Homonym	**0.783**
XL-LEXEME + 0.5 threshold	0.761
XL-LEXEME + Opt. threshold	0.848

Table 4: Binary task SemEval-2020 Task 1 (EN)

Conclusion

In this work, we have showed that:

- ***definitions*** of word senses can be used to detect ***semantic change type***;
- we can ***classify the type of semantic change*** by training on synchronic sense relationships using sense definitions; and that
- type information can ***improve*** models for both graded Word-In-Context (WiC) as well as ***semantic change detection***.

Links

hf.co/ChangeIsKey/change-type-classifier

GitHub github.com/ChangeIsKey/change-type-classification

zenodo zenodo.org/records/11471318

pierluigi.cassotti@gu.se

CHANGE IS KEY.ORG

Change is Key!
changeiskey.org

Francesco Periti[1] Pierluigi Cassotti[2] Haim Dubossarsky[3] Nina Tahmasebi[2]

Analyzing **Semantic Change** through **Lexical Replacements**

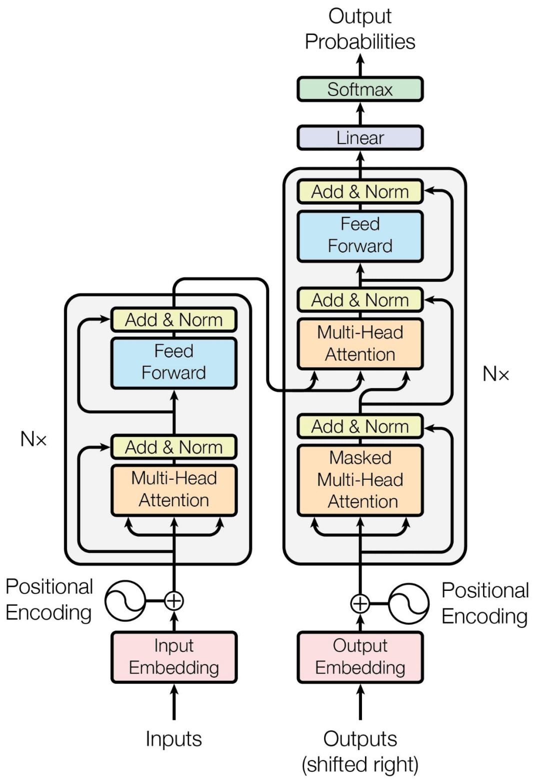
[1] University of Milan, Italy [2] University of Gothenburg, Sweden [3] Queen Mary University of London, England

CHANGE IS KEY.ORG

ACL 2024 Bangkok, Thailand

Introduction

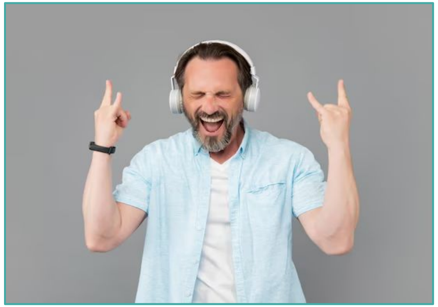
Contextualization



The advancement of LLMs: `Contextualization`

When words are used in contexts similar to those encountered *during training*, LLMs can easily differentiate, in a computational way, between word meanings.

Sitting on a **rock**



Listening to **rock**

Introduction
Language is always changing

Lexical Semantic Change
Words changing in meaning over time.

E.g., `gay` has changed its meaning from *happy* to *homosexual*

Lexical Replacement
Words replaced by other words over time.

E.g., `happy` has replaced the word `gay` in some of his contexts

Can LLMs contextualize word meanings that have not been encountered during training?

The heart is sportive, light, and **happy**, life seems a long glad summer's day

The heart is sportive, light, and **gay**, life seems a long glad summer's day

1854, found via https://discovery.nationalarchives.gov.uk

`gay`
homosexual ← *happy*

CHANGE
IS KEY .ORG

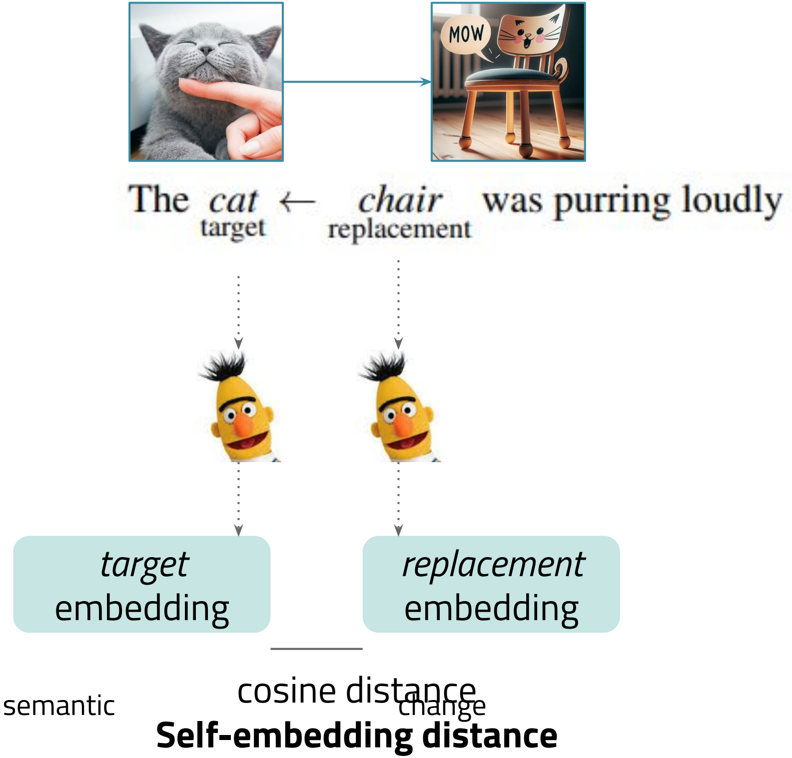
The replacement schema
Self-embedding distance

synonyms (e.g. *sadness* ← *unhappiness*)
- Expectation: similar embeddings
- Emulation: the absence of any semantic change

antonyms (e.g. *hot* ← *cold*)
- Expectation: slightly less similar embeddings
- Emulation: a *contronym* change

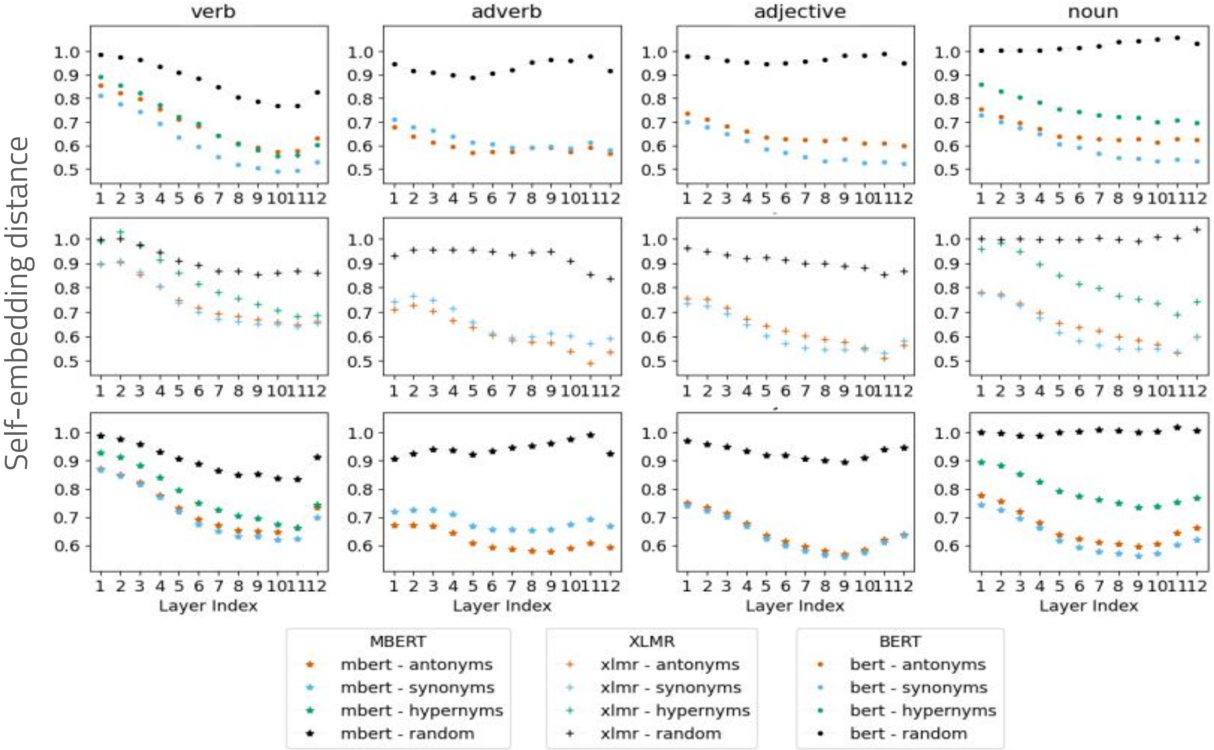
hypernyms (e.g. *animal* ← *bird*)
- Expectation: slightly less similar embeddings
- Emulation: a *broadening* change

random (e.g. *sadness* ← *eld*)
- Expectation: very dissimilar embeddings
- Emulation: strong
  (i.e., the emergence of a homonymic sense)



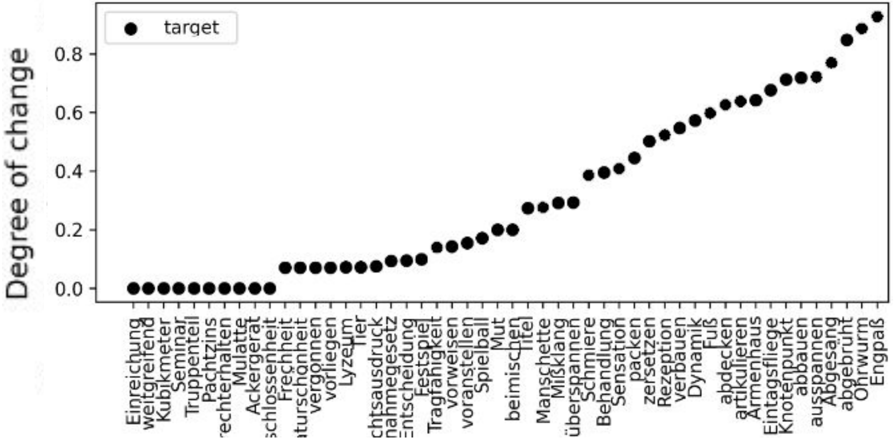
The replacement schema

Self-embedding distance



CHANGE IS KEY .ORG

Modeling Semantic Change
Graded Change Detection



Graded Change Detection consists in ranking a set of targets word according to their degree of semantic change between two time periods.

E.g., A noisy ***fly*** sat on my shoulder

Lexical replacements

`fly ← bug,`
`fly ← beetle,`
…
`fly ← butterfly`

degree of change

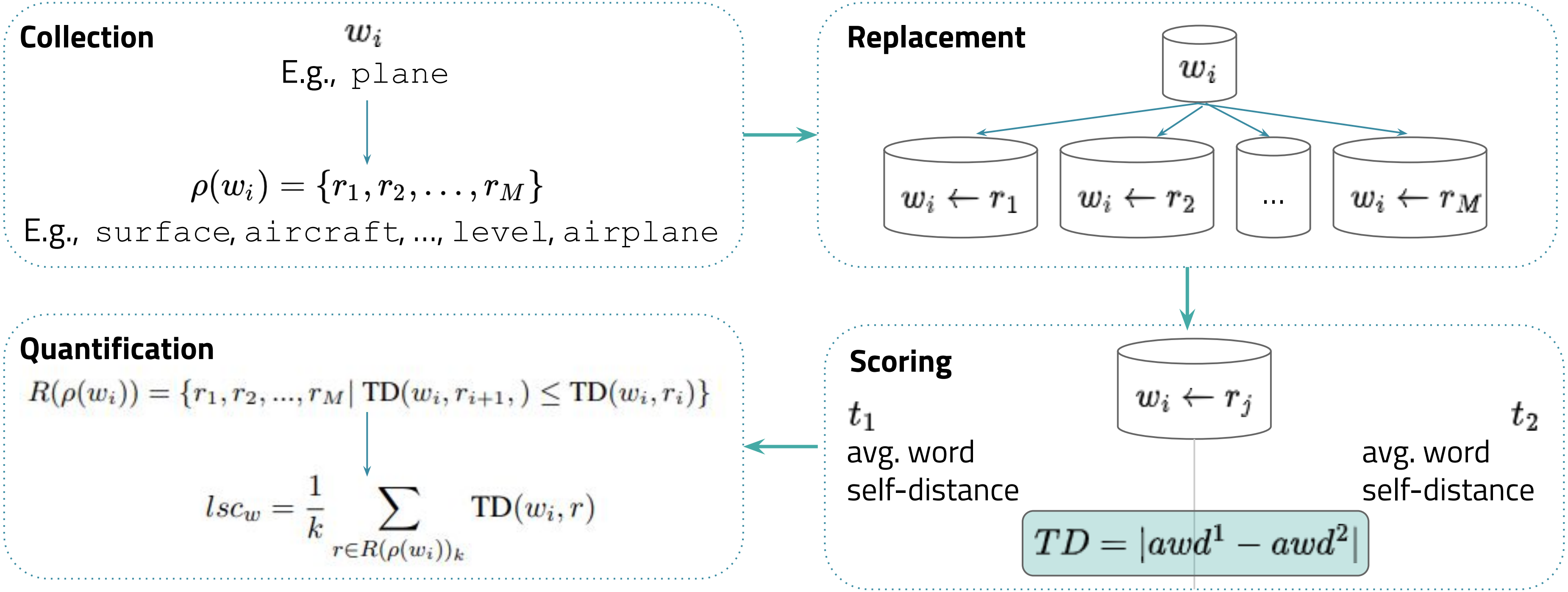
Lexical substitutes

`fly ← bug,`
`fly ← beetle,`
…
`fly ← butterfly`

degree of change

Modeling Semantic Change

A novel approach based on *lexical replacements*

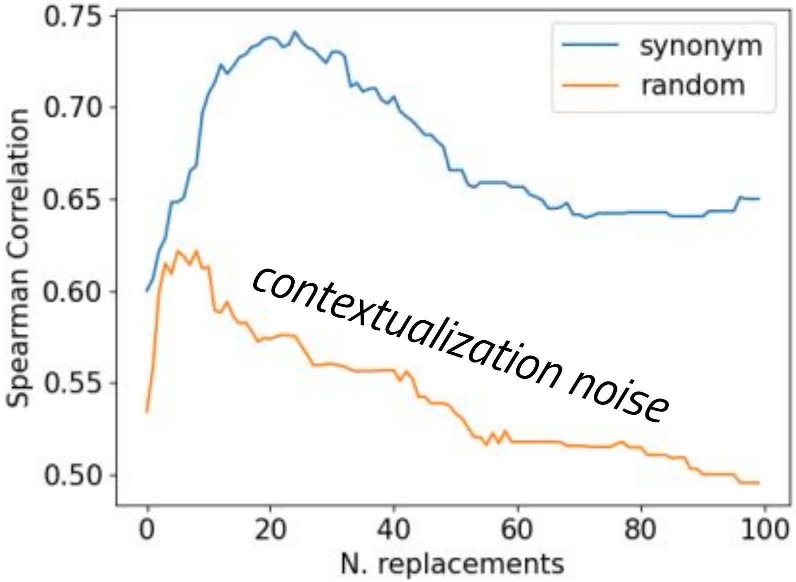


Modeling Semantic Change

A novel approach based on *lexical replacements*

Lexical replacements

	Model	Spearman Correlation
	Rosin and Radinsky	0.629
	Kutuzov and Giulianelli	0.605
	Laicher et al.	0.571
	Periti et al.	0.512
	Cassotti et al. (XL-LEXEME)	0.757
Synonym Replacement	Replacement Min. Corr.	0.600
	Replacement Max. Corr.	0.741
	Replacement Avg. Corr.	0.674
Random Replacement	Replacement Min. Corr.	0.495
	Replacement Max. Corr.	0.622
	Replacement Avg. Corr.	0.542



$$lsc_w = \frac{1}{k} \sum_{r \in R(\rho(w_i))_k} \mathrm{TD}(w_i, r)$$

Modeling Semantic Change

lexical replacements vs. *lexical substitutes*

Lexical replacements

	Model	Spearman Correlation
	Rosin and Radinsky	0.629
	Kutuzov and Giulianelli	0.605
	Laicher et al.	0.571
	Periti et al.	0.512
	Cassotti et al. (XL-LEXEME)	0.757
Synonym Replacement	Replacement Min. Corr.	0.600
	Replacement Max. Corr.	0.741
	Replacement Avg. Corr.	0.674
Random Replacement	Replacement Min. Corr.	0.495
	Replacement Max. Corr.	0.622
	Replacement Avg. Corr.	0.542

Lexical substitutes

Model	Spearman Correlation
Arefyev and Zhikov (2020)	0.299
Card (2023)	0.547
LLaMa 2 7B	**0.731**
BERT	*0.450*

Conclusion

In our work...

- We analyzed semantic change though lexical replacements
- We proposed a new *interpretable* method to model semantic change
- We compared the use of *lexical replacements* and *lexical substitutes*

Takeaways

- BERT, mBERT, and XLM-R struggle to contextualize word meanings that they did not encounter during training
- Semantic change can be modeled as contextualization noise
- Our method obtain state-of-the-art results while being interpretable

Outline

- Computational Modeling Lexical Semantics
    - Synchronic Modeling
    - Diachronic Modeling
- Human annotation of Lexical Semantic Change
- Diachronic Models of Language
    - Static Models and Alignment
    - Contextualized Models
    - Generative Models
- **Hands-on**

Thank for your attention!

Change Is Key!

https://www.changeiskey.org/

pierluigi.cassotti@gu.se