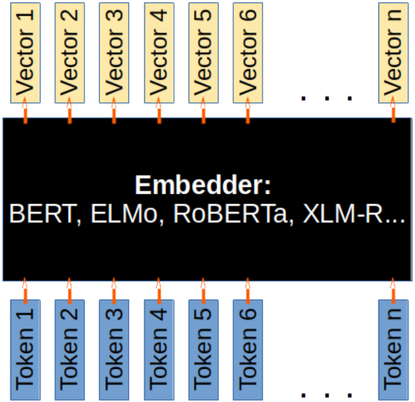
Computational approaches to semantic change detection
Day 3
Learned representations for semantic change detection part 2: contextualized embeddings and pre-trained language models

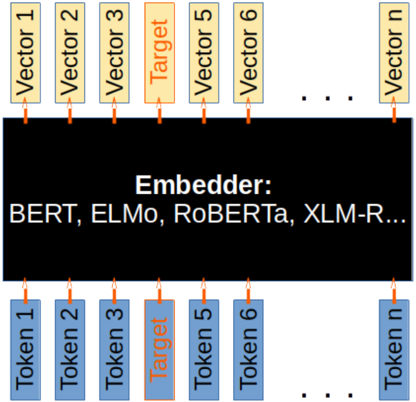
Andrey Kutuzov, Lidia Pivovarova

University of Oslo, University of Helsinki

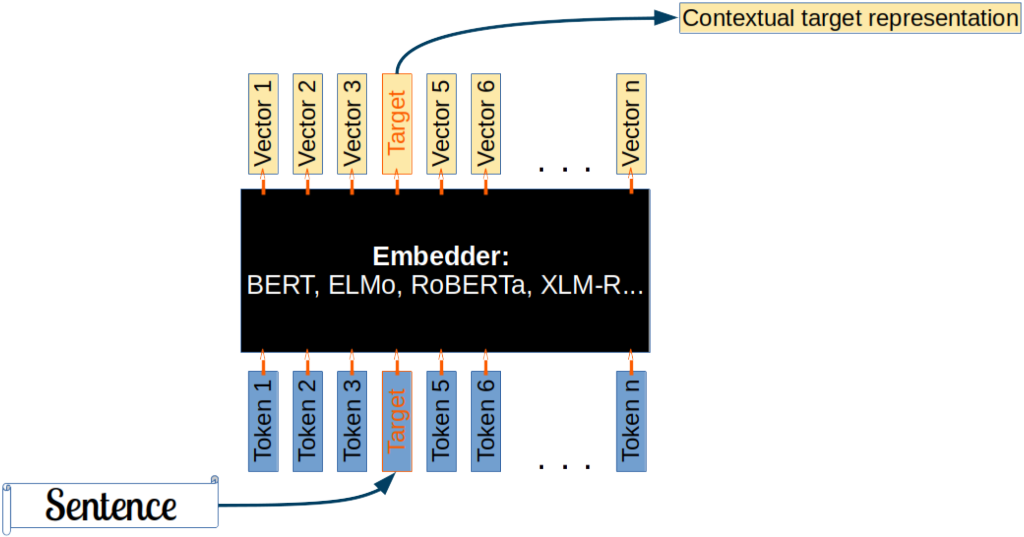
Contextualized Embeddings



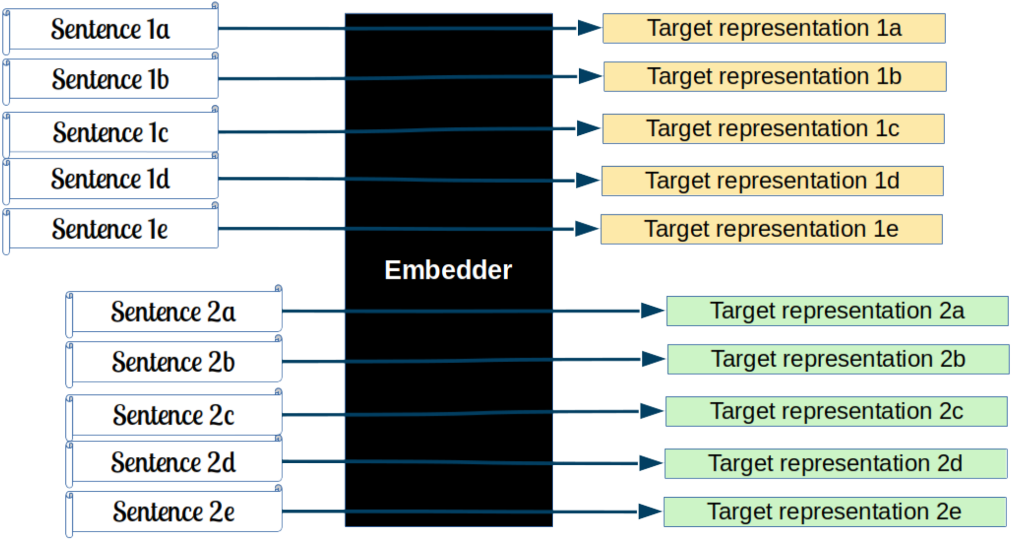
Contextualized Embeddings



Contextualized Embeddings



Contextualized Embeddings



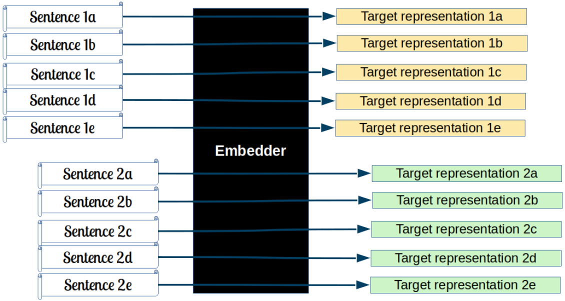
Contextualized Embeddings

- While static embeddings allows for **one word representation per corpus**
- contextualized embeddings output a separate vector for each mention in a corpus,
- thus preserving much more information, accounting for polysemy and enabling interpretability.

Contextualized Embeddings

- While static embeddings allows for **one word representation per corpus**
- contextualized embeddings output a separate vector for each mention in a corpus,
- thus preserving much more information, accounting for polysemy and enabling interpretability.

How to use these embeddings, specifically?



Form-based vs. sense-based methods

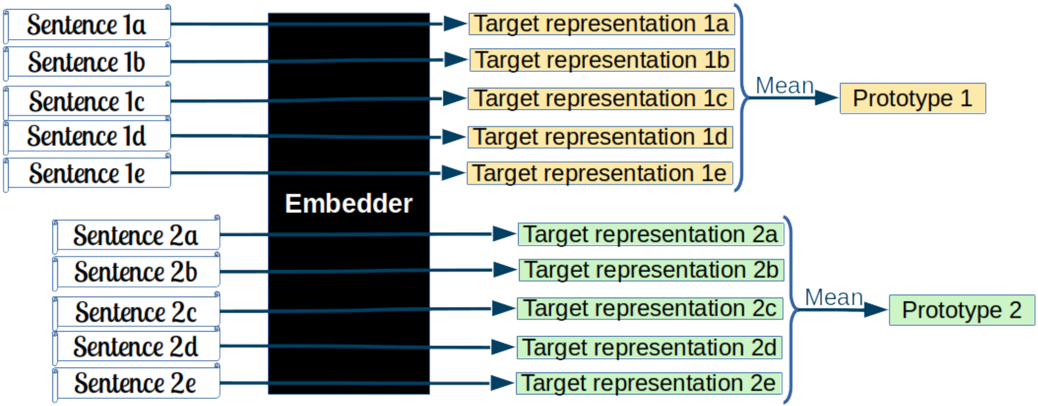
Methods using contextualized embeddings

Form-based: the meaning representation concerns the high-level properties of a target word, e.g. degree of polysemy or a dominant sense

Sense-based: the meaning representation concerns the low-level properties of a target word, i.e. different word usages (multiple meanings)

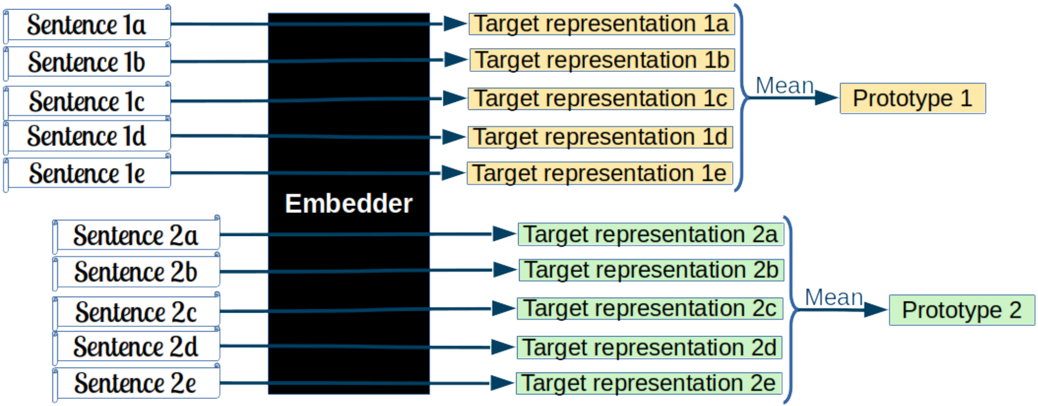
[Montanelli and Periti, 2023, Giulianelli et al., 2020]

Averaging



Degree of change: cosine distance between prototypes [Martinc et al., 2020a].

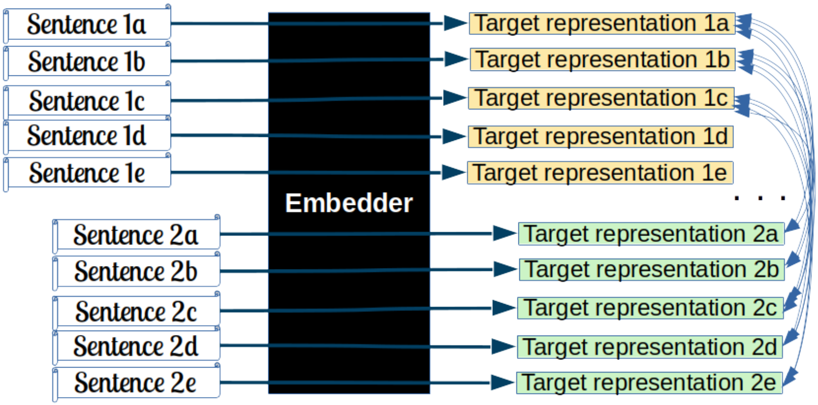
Averaging



Degree of change: cosine distance between prototypes [Martinc et al., 2020a].

$$PRT = 1/cosine\_similarity(prototype1, prototype2)$$ [Kutuzov and Giulianelli, 2020]

Distances



Degree of change: averaged pairwise distance (APD) [Giulianelli et al., 2020].

Form-based methods

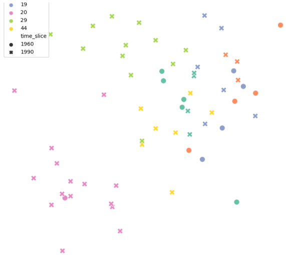
Form-based methods are straightforward to compute but not easily interpretable – an advantage of nuanced representation is lost as all word mentions in the corpus are collapsed into a single representation.

Form-based methods

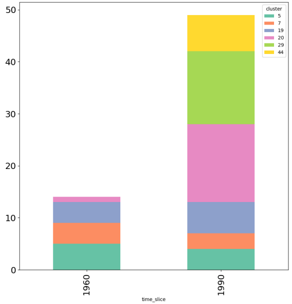
Form-based methods are straightforward to compute but not easily interpretable – an advantage of nuanced representation is lost as all word mentions in the corpus are collapsed into a single representation.

Stay tuned, we'll elaborate on interpretability problem on Friday

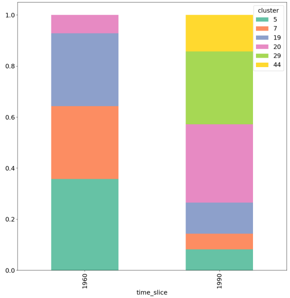
Clustering



1. Clustering

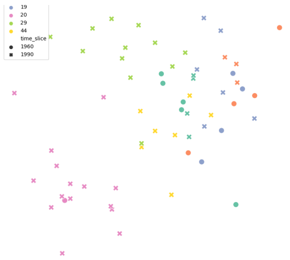


2. Cluster distributions

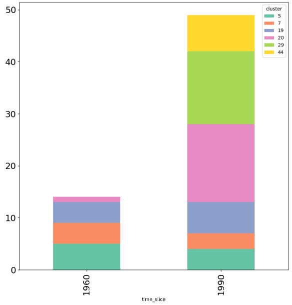


3. Divergence

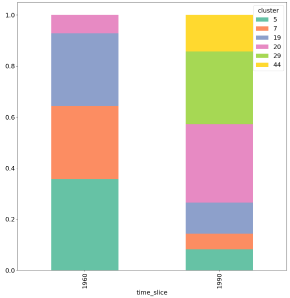
Clustering



1. Clustering



2. Cluster distributions



3. Divergence

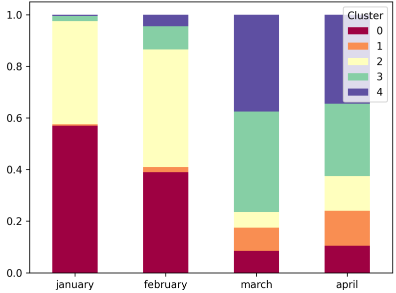
Jensen-Shannon divergence [Martinc et al., 2020b, Giulianelli et al., 2020]

$$JSD(P_w^1, P_w^2) = 1/2(KL(P_w^1||M) + KL(P_w^2||M))$$

$$M = (P_w^1 + P_w^2)/2$$

$$KL(P||M) = \sum_{x \in X} P(x) log(\frac{P(x)}{M(x)})$$

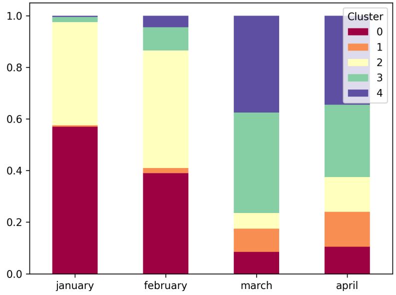
Clustering: interpretation



2020 news: cluster distributions per month for word *strain* and top-10 tf-idf keywords for each cluster [Montariol et al., 2021].

#	Keywords
0	strain coronavirus, new strain, city wuhan, novel strain, strain virus, chinese city
1	strain health, strain resources, stream, network infrastructure, international resources, likely strain
2	new strain, acute respiratory, 2019 ncov, respiratory syndrome, severe acute, identified humans
3	financial strain, feeling strain, strain coronavirus, economic strain, signs strain, strain said
4	ease strain, putting strain, strain health, reduce strain, care system, strain hospitals

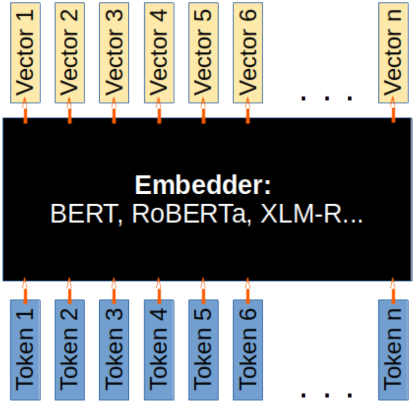
Clustering: interpretation



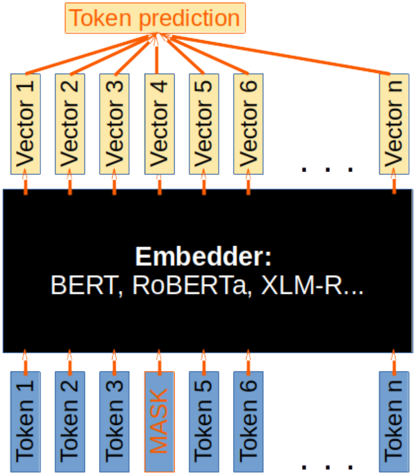
#	Keywords
0	strain coronavirus, new strain, city wuhan, novel strain, strain virus, chinese city
1	strain health, strain resources, stream, network infrastructure, international resources, likely strain
2	new strain, acute respiratory, 2019 ncov, respiratory syndrome, severe acute, identified humans
3	financial strain, feeling strain, strain coronavirus, economic strain, signs strain, strain said
4	ease strain, putting strain, strain health, reduce strain, care system, strain hospitals

But: correspondence between clusters and *senses* is unclear; number of clusters, clustering method affect results; top-performing method is not necessary the most interpretable.

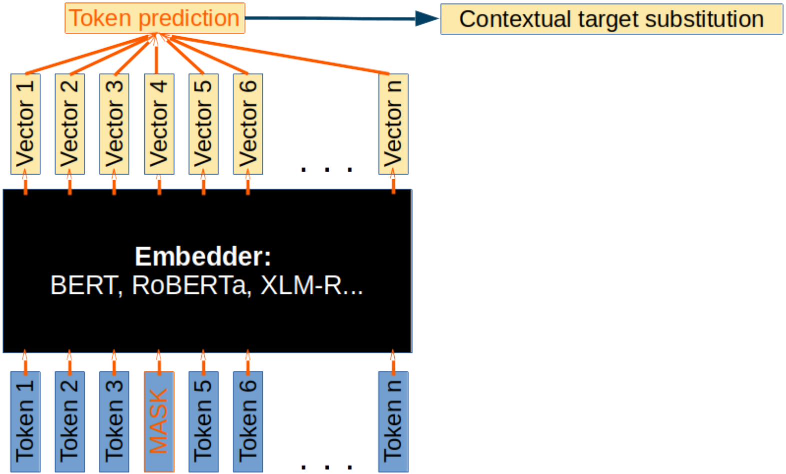
Substitutes



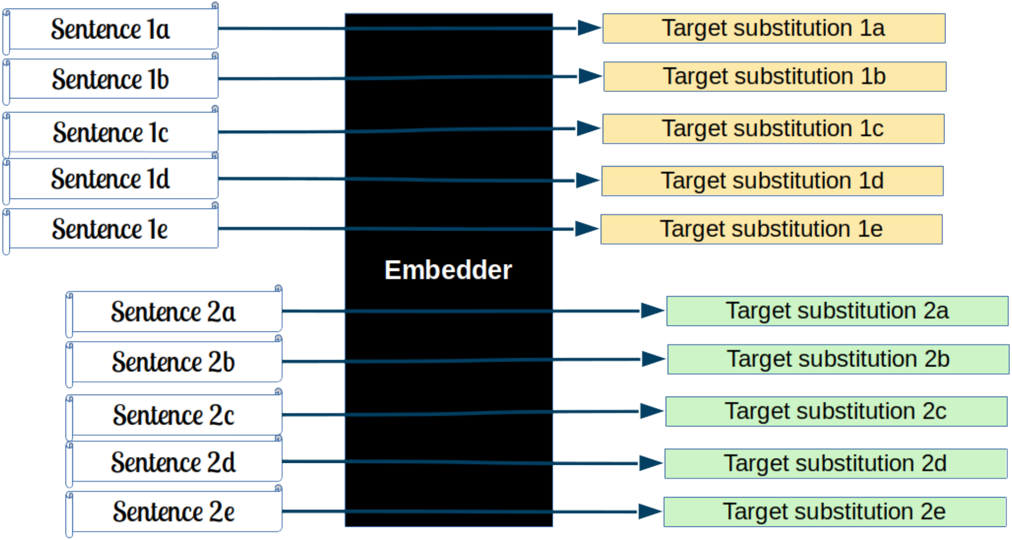
Substitutes



Substitutes



Substitutes



Substitutes

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

Table 2: Example terms from the SE English dataset, showing the most common substitutes from our approach.

[Card, 2023]

Substitutes

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

Table 2: Example terms from the SE English dataset, showing the most common substitutes from our approach.

[Card, 2023]

Same issues: correspondence between substitutions and *senses*, hyperparameters...

1. Contextualized Embeddings
2. Form-based methods
3. Sense-based methods
4. **Supervised methods**
5. Time-awareness
6. Performance

Supervised methods

- Most methods for semantic shift detection are unsupervised
  - the main reason is the lack of training data due to difficulty in manual annotation for this task

Supervised methods

- Most methods for semantic shift detection are unsupervised
  - the main reason is the lack of training data due to difficulty in manual annotation for this task
- When training data exists LLMs show their potential

Direct supervision

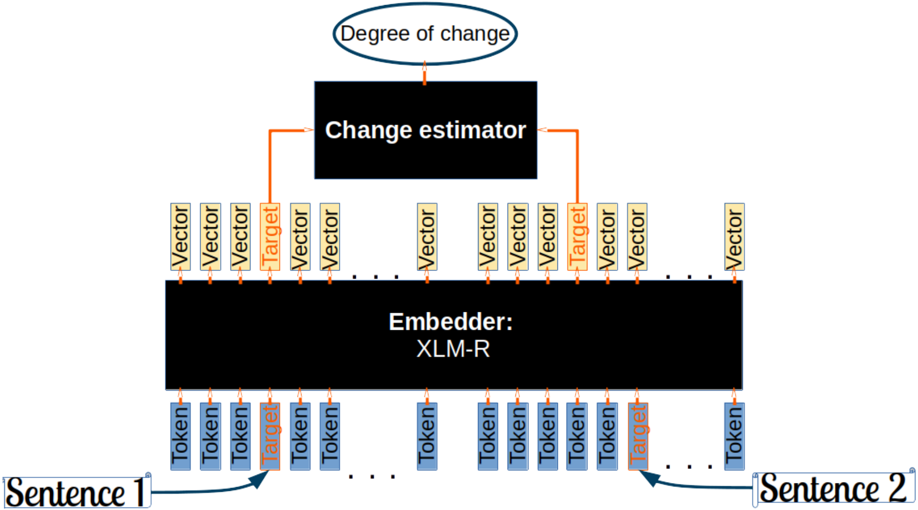
RuShiftEval—the only shared task (so far) that included a training set: the same language (Russian) and the same time periods (before/after 1917 and before/after 1991) [Kutuzov and Pivovarova, 2021]. Three best-performing solutions used the training set to explicitly emulate the annotation process.

Direct supervision

RuShiftEval—the only shared task (so far) that included a training set: the same language (Russian) and the same time periods (before/after 1917 and before/after 1991) [Kutuzov and Pivovarova, 2021]. Three best-performing solutions used the training set to explicitly emulate the annotation process.

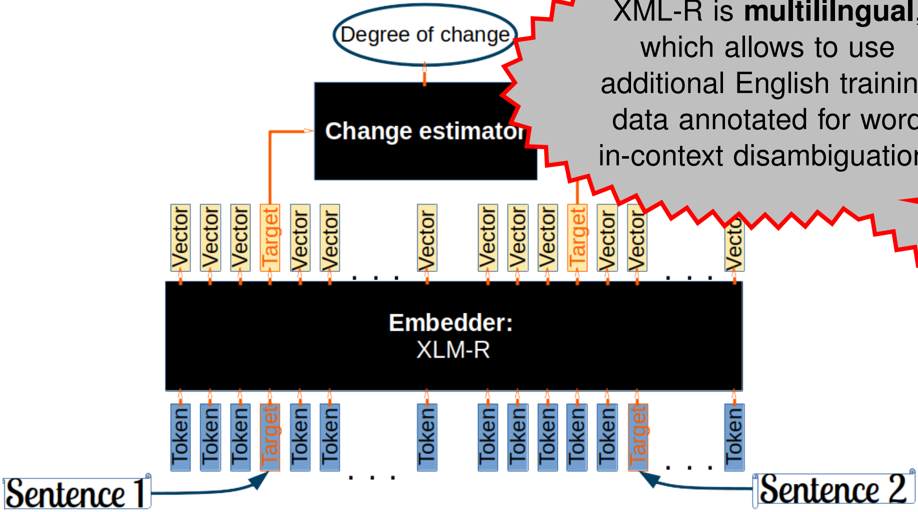
	Team	RuShiftEval1	RuShiftEval2	RuShiftEval3	Mean	Type
1	**GlossReader**	0.781	**0.803**	**0.822**	**0.802**	token
2	**DeepMistake**	**0.798**	0.773	0.803	0.791	token
3	vanyatko	0.678	0.746	0.737	0.720	token
4	aryzhova	0.469	0.450	0.453	0.457	token
5	Discovery	0.455	0.410	0.494	0.453	token
6	UWB	0.362	0.354	0.533	0.417	type
7	dschlechtweg	0.419	0.373	0.383	0.392	type
8	jenskaiser	0.430	0.310	0.406	0.382	token
9	SBX-HY	0.388	0.281	0.439	0.369	type
	Baseline	0.314	0.302	0.381	0.332	type

Direct supervision



[Arefyev et al., 2021]

Direct supervision



XML-R is **multililngual**, which allows to use additional English training data annotated for word in-context disambiguation.

[Arefyev et al., 2021]

Distant Supervision

[Hu et al., 2019]

1. Use Oxford English dictionary for **explicit** sense representation

Distant Supervision

[Hu et al., 2019]

1. Use Oxford English dictionary for **explicit** sense representation
2. For each word, for each sense: take example sentences from the dictionary, run them through BERT and thus obtain *sense representation*;

Distant Supervision

[Hu et al., 2019]

1. Use Oxford English dictionary for **explicit** sense representation
2. For each word, for each sense: take example sentences from the dictionary, run them through BERT and thus obtain *sense representation*;
3. For each word mention within a corpus: run a sentence through BERT, get representation and compute cosine similarity to all sense representations;

Distant Supervision

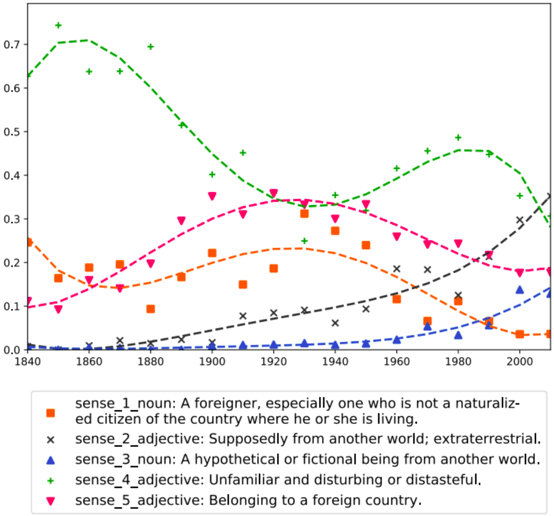
[Hu et al., 2019]

1. Use Oxford English dictionary for **explicit** sense representation
2. For each word, for each sense: take example sentences from the dictionary, run them through BERT and thus obtain *sense representation*;
3. For each word mention within a corpus: run a sentence through BERT, get representation and compute cosine similarity to all sense representations;
4. Choose the sense with the highest similarity.

Distant Supervision

[Hu et al., 2019]

1. Use Oxford English dictionary for **explicit** sense representation
2. For each word, for each sense: take example sentences from the dictionary, run them through BERT and thus obtain *sense representation*;
3. For each word mention within a corpus: run a sentence through BERT, get representation and compute cosine similarity to all sense representations;
4. Choose the sense with the highest similarity.



(b) *alien*

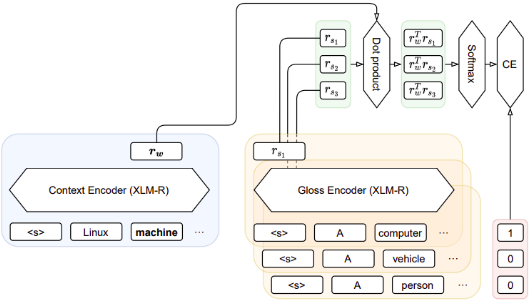
Distant Supervision

[Rachinskiy and Arefyev, 2022]
"Finetuning on some task that requires understanding word senses and at the same time ignoring word forms shall help to get rid of grammatical bias in the contextualized embeddings"

Distant Supervision

[Rachinskiy and Arefyev, 2022]
"Finetuning on some task that requires understanding word senses and at the same time ignoring word forms shall help to get rid of grammatical bias in the contextualized embeddings"

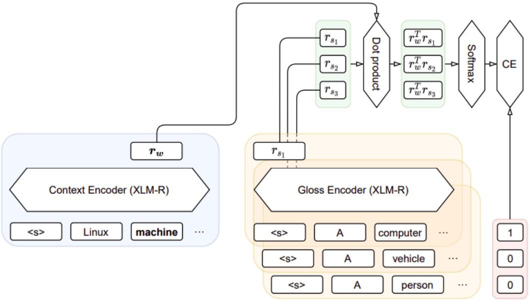
1. **Fine-tune** a model for sense disambiguation (English)

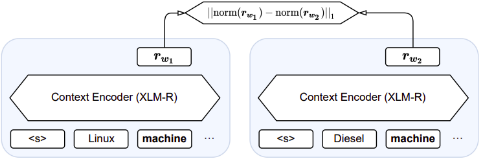


Distant Supervision

[Rachinskiy and Arefyev, 2022]
"Finetuning on some task that requires *understanding word senses* *and at the same time* *ignoring word forms* *shall help to get rid of grammatical bias in the contextualized embeddings"*

1. **Fine-tune** a model for sense disambiguation (English)
2. Use a fine-tuned model to compute similarity (Spanish)





Time-awareness

The methods we discussed so far are **time-oblivious** [Montanelli and Periti, 2023]

- No explicit information on time direction is used: corpora can be switched without changing of the model outputs
- Information on time periods is presented implicitly: in majority of the methods contextualized embeddings are trained or finetuned on historical corpora, which improves performance

Time-awareness

The methods we discussed so far are **time-oblivious** [Montanelli and Periti, 2023]

- No explicit information on time direction is used: corpora can be switched without changing of the model outputs
- Information on time periods is presented implicitly: in majority of the methods contextualized embeddings are trained or finetuned on historical corpora, which improves performance
- Time-oblivious methods can be re-used to track difference across other dimensions (topic, genre, etc.)

Time-awareness

The methods we discussed so far are **time-oblivious** [Montanelli and Periti, 2023]

- No explicit information on time direction is used: corpora can be switched without changing of the model outputs
- Information on time periods is presented implicitly: in majority of the methods contextualized embeddings are trained or finetuned on historical corpora, which improves performance
- Time-oblivious methods can be re-used to track difference across other dimensions (topic, genre, etc.)
- Still, semantic shift detection could benefit from explicit knowledge of a time period

Time-aware Models

- Time-referencing: target word mentions in corpora are replaced with *target_new* and *target_old*, respectively [Zhou and Li, 2020]

Time-aware Models

- Time-referencing: target word mentions in corpora are replaced with *target_new* and *target_old*, respectively [Zhou and Li, 2020]
- Sentence-time classification: a model is fine-tuned to predict time period [Pömsl and Lyapin, 2020]

Time-aware Models

- Time-referencing: target word mentions in corpora are replaced with *target_new* and *target_old*, respectively [Zhou and Li, 2020]
- Sentence-time classification: a model is fine-tuned to predict time period [Pömsl and Lyapin, 2020]
- Temporal attention [Rosin and Radinsky, 2022]:

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

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

Time-aware Models

- Time-referencing: target word mentions in corpora are replaced with *target_new* and *target_old*, respectively [Zhou and Li, 2020]
- Sentence-time classification: a model is fine-tuned to predict time period [Pömsl and Lyapin, 2020]
- Temporal attention [Rosin and Radinsky, 2022]:

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

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

- Adjusting input embeddings [Hofmann et al., 2021]

$$e = e' + o_t,$$

where e - token embedding at time point t, e' - time-agnostic representation, o_t - time-point embedding

Performance [Montanelli and Periti, 2023]

Ref.	SemEval English	SemEval German	SemEval Latin	SemEval Swedish	GEMS English	LivFC English	COHA English	LSCD Spanish	DURel German	SURel German	RSE Russian			NOR Norwegian	
	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_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
[Teodorescu et al., 2022]	-	-	-	-	-	-	-	ensemble APD 0.573	-	-	-	-	-	-	-
[Zhou and Li, 2020]	form-based CD 0.392	form-based CD 0.392	form-based CD 0.392	form-based CD 0.392	-	-	-	-	-	-	-	-	-	-	-
[Montariol et al., 2021]	sense-based AP + WD 0.456	sense-based AP + JSD 0.583	form-based CD 0.496	sense-based K-Means + WD 0.332	**sense-based** **AP + JSD** **0.510**	-	-	-	sense-based AP + JSD 0.712	-	-	-	-	-	-
[Periti et al., 2022]	sense-based AP + JSD 0.514*	-	sense-based APP + JSD 0.512*	-	-	-	-	-	-	-	-	-	-	-	-
[Pömsl and Lyapin, 2020]	ensemble APD 0.246	ensemble APD 0.725	ensemble APD 0.463	ensemble APD 0.546	-	-	-	-	**ensemble** **APD** **0.802**	**ensemble** **APD** **0.723**	-	-	-	-	-
[Rachinskiy and Arefyev, 2021]	-	-	-	-	-	-	-	-	-	-	ensemble APD 0.781	ensemble APD 0.803	ensemble APD 0.822	-	-
[Rachinskiy and Arefyev, 2022]	-	-	-	-	-	-	-	**sense** **APDP** **0.745**	-	-	-	-	-	-	-
[Rodina et al., 2021]	-	-	-	-	-	-	-	-	-	-	form-based PRT 0.557	sense-based AP + JSD 0.406	-	-	-
[Rosin et al., 2022]	form-based CD 0.467	-	form-based CD 0.512	-	-	**form-based** **TD** **0.620**	-	-	-	-	-	-	-	-	-
[Rosin and Radinsky, 2022]	**form-based** **CD** **0.627**	**form-based** **CD** **0.763**	**form-based** **CD** **0.565**	-	-	-	-	-	-	-	-	-	-	-	-
[Rother et al., 2020]	sense-based HDBSCAN 0.512	sense-based GMMs 0.605	sense-based GMMs 0.321	sense-based HDBSCAN 0.308	-	-	-	-	-	-	-	-	-	-	-
[Ryzhova et al., 2021]	-	-	-	-	-	-	-	-	-	-	ensemble regression 0.480*	ensemble regression 0.487*	ensemble regression 0.560*	-	
[Kudisov and Arefyev, 2022]	-	-	-	-	-	-	-	form-based APD 0.637	-	-	-	-	-	-	-
[Kutuzov, 2020]	form-based APD 0.605	form-based PRT 0.740	form-based PRT 0.561	**form-based** **APD** **0.610**	sense-based AP + JSD 0.456*	-	-	-	-	-	-	-	-	-	-
[Laicher et al., 2021]	form-based APD 0.571*	form-based CD 0.755*	-	form-based APD 0.602*	-	-	-	-	-	-	-	-	-	-	-
[Liu et al., 2021]	form-based CD 0.341	form-based CD 0.512	form-based CD 0.304	form-based CD 0.304	form-based CD 0.286	form-based CD 0.561	-	-	-	-	-	-	-	-	-
[Martinc et al., 2020c]	ensemble AP + JSD 0.361	ensemble AP + JSD 0.642	form-based CD 0.496	ensemble AP + JSD 0.343	-	-	-	-	-	-	-	-	-	-	-
[Giulianelli et al., 2020]	-	-	-	-	form-based APD 0.285*	-	-	-	-	-	-	-	-	-	-
[Giulianelli et al., 2022]	form-based APD 0.514	ensemble PRT 0.354	ensemble PRT 0.572	ensemble APD 0.397	-	-	-	-	-	-	ensemble APD + PRT 0.376	form-based APD 0.480	form-based APD 0.457	**ensemble** **APD + PRT** **0.394**	**ensemble** **APD** **0.503**
[Hu et al., 2019]	-	-	-	-	-	-	**sense-based** **MNS** **0.428***	-	-	-	-	-	-	-	-

Performance

- Exhaustive comparison of all existing methods is hard to perform
- Previous results do not show a single best-performing method

Performance

- Exhaustive comparison of all existing methods is hard to perform
- Previous results do not show a single best-performing method
  - Results depends on corpus size, time distance, noise in the data and other properties
  - May also depend on specific languages

Performance

- Exhaustive comparison of all existing methods is hard to perform
- Previous results do not show a single best-performing method
  - Results depends on corpus size, time distance, noise in the data and other properties
  - May also depend on specific languages
- Sense-based methods perform worse than form-based in a fully unsupervised setting [Schlechtweg et al., 2020],

Performance

- Exhaustive comparison of all existing methods is hard to perform
- Previous results do not show a single best-performing method
  - Results depends on corpus size, time distance, noise in the data and other properties
  - May also depend on specific languages
- Sense-based methods perform worse than form-based in a fully unsupervised setting [Schlechtweg et al., 2020],
- Contextualized models are over-sensitive to grammatical and stylistical properties [Laicher et al., 2021]

Performance

- Exhaustive comparison of all existing methods is hard to perform
- Previous results do not show a single best-performing method
  - Results depends on corpus size, time distance, noise in the data and other properties
  - May also depend on specific languages
- Sense-based methods perform worse than form-based in a fully unsupervised setting [Schlechtweg et al., 2020],
- Contextualized models are over-sensitive to grammatical and stylistical properties [Laicher et al., 2021]
- Language models have large potential for supervised training, including distant supervision [Kutuzov and Pivovarova, 2021, Zamora-Reina et al., 2022]

Performance

- Exhaustive comparison of all existing methods is hard to perform
- Previous results do not show a single best-performing method
  - Results depends on corpus size, time distance, noise in the data and other properties
  - May also depend on specific languages
- Sense-based methods perform worse than form-based in a fully unsupervised setting [Schlechtweg et al., 2020],
- Contextualized models are over-sensitive to grammatical and stylistical properties [Laicher et al., 2021]
- Language models have large potential for supervised training, including distant supervision [Kutuzov and Pivovarova, 2021, Zamora-Reina et al., 2022]
- Methods based on contextualized embeddings are more interpretable

Performance

- Exhaustive comparison of all existing methods is hard to perform
- Previous results do not show a single best-performing method
  - Results depends on corpus size, time distance, noise in the data and other properties
  - May also depend on specific languages
- Sense-based methods perform worse than form-based in a fully unsupervised setting [Schlechtweg et al., 2020],
- Contextualized models are over-sensitive to grammatical and stylistical properties [Laicher et al., 2021]
- Language models have large potential for supervised training, including distant supervision [Kutuzov and Pivovarova, 2021, Zamora-Reina et al., 2022]
- Methods based on contextualized embeddings are more interpretable
  - Though most interpretable methods are not necessary best-performing

References I

Arefyev, N., Homskiy, D., Fedoseev, M., Davletov, A., Protasov, V., and Panchenko, A. (2021).
Deepmistake: Which senses are hard to distinguish for a wordincontext model.
In *Computational linguistics and intellectual technologies: Papers from the annual conference Dialogue*.

Arefyev, N. and Zhikov, V. (2020).
BOS at SemEval-2020 task 1: Word sense induction via lexical substitution for lexical semantic change detection.
In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 171–179, Barcelona (online). International Committee for Computational Linguistics.

References II

Beck, C. (2020).
DiaSense at SemEval-2020 task 1: Modeling sense change via pre-trained BERT embeddings.
In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 50–58, Barcelona (online). International Committee for Computational Linguistics.

Card, D. (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.

References III

Cuba Gyllensten, A., Gogoulou, E., Ekgren, A., and Sahlgren, M. (2020).
SenseCluster at SemEval-2020 task 1: Unsupervised lexical semantic change detection.
In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 112–118, Barcelona (online). International Committee for Computational Linguistics.

Giulianelli, M., Del Tredici, M., and Fernández, R. (2020).
Analysing lexical semantic change with contextualised word representations.
In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3960–3973, Online. Association for Computational Linguistics.

References IV

Giulianelli, M., Kutuzov, A., and Pivovarova, L. (2022).
Do not fire the linguist: Grammatical profiles help language models detect semantic change.
In *Proceedings of the 3rd Workshop on Computational Approaches to Historical Language Change*, pages 54–67, Dublin, Ireland. Association for Computational Linguistics.

Hofmann, V., Pierrehumbert, J., and Schütze, H. (2021).
Dynamic contextualized word embeddings.
In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6970–6984, Online. Association for Computational Linguistics.

References V

Hu, R., Li, S., and Liang, S. (2019).
Diachronic sense modeling with deep contextualized word embeddings: An ecological view.
In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3899–3908, Florence, Italy. Association for Computational Linguistics.

Kanjirangat, V., Mitrovic, S., Antonucci, A., and Rinaldi, F. (2020).
SST-BERT at SemEval-2020 task 1: Semantic shift tracing by clustering in BERT-based embedding spaces.
In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 214–221, Barcelona (online). International Committee for Computational Linguistics.

References VI

Karnysheva, A. and Schwarz, P. (2020).
TUE at SemEval-2020 task 1: Detecting semantic change by clustering contextual word embeddings.
In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 232–238, Barcelona (online). International Committee for Computational Linguistics.

Kashleva, K., Shein, A., Tukhtina, E., and Vydrina, S. (2022).
HSE at LSCDiscovery in Spanish: Clustering and profiling for lexical semantic change discovery.
In *Proceedings of the 3rd Workshop on Computational Approaches to Historical Language Change*, pages 193–197, Dublin, Ireland. Association for Computational Linguistics.

References VII

Keidar, D., Opedal, A., Jin, Z., and Sachan, M. (2022).
Slangvolution: A causal analysis of semantic change and frequency dynamics in slang.
In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1422–1442, Dublin, Ireland. Association for Computational Linguistics.

Kudisov, A. and Arefyev, N. (2022).
BOS at LSCDiscovery: Lexical substitution for interpretable lexical semantic change detection.
In *Proceedings of the 3rd Workshop on Computational Approaches to Historical Language Change*, pages 165–172, Dublin, Ireland. Association for Computational Linguistics.

References VIII

- Kutuzov, A. (2020).
  Distributional word embeddings in modeling diachronic semantic change.
  PhD thesis, University of Oslo.

- Kutuzov, A. and Giulianelli, M. (2020).
  UiO-UvA at SemEval-2020 task 1: Contextualised embeddings for lexical semantic change detection.
  In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 126–134, Barcelona (online). International Committee for Computational Linguistics.

- Kutuzov, A. and Pivovarova, L. (2021).
  RuShiftEval: a shared task on semantic shift detection for Russian.
  Computational linguistics and intellectual technologies: Papers from the annual conference Dialogue.

References IX

Kutuzov, A., Velldal, E., and Øvrelid, L. (2022).
Contextualized embeddings for semantic change detection: Lessons learned.
In *Northern European Journal of Language Technology, Volume 8*, Copenhagen, Denmark. Northern European Association of Language Technology.

Laicher, S., Kurtyigit, S., Schlechtweg, D., Kuhn, J., and Schulte im Walde, S. (2021).
Explaining and improving BERT performance on lexical semantic change detection.
In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop*, pages 192–202, Online. Association for Computational Linguistics.

References X

Liu, Y., Medlar, A., and Glowacka, D. (2021).
Statistically significant detection of semantic shifts using contextual word embeddings.

In *Proceedings of the 2nd Workshop on Evaluation and Comparison of NLP Systems*, pages 104–113, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Martinc, M., Kralj Novak, P., and Pollak, S. (2020a).
Leveraging contextual embeddings for detecting diachronic semantic shift.
In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4811–4819, Marseille, France. European Language Resources Association.

Martinc, M., Montariol, S., Zosa, E., and Pivovarova, L. (2020b).
Capturing evolution in word usage: just add more clusters?
In *Companion Proceedings of the Web Conference 2020*, pages 343–349.

References XI

Martinc, M., Montariol, S., Zosa, E., and Pivovarova, L. (2020c).
Discovery team at SemEval-2020 task 1: Context-sensitive embeddings not always better than static for semantic change detection.
In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 67–73, Barcelona (online). International Committee for Computational Linguistics.

Montanelli, S. and Periti, F. (2023).
A survey on contextualised semantic shift detection.
arXiv preprint arXiv:2304.01666.

Montariol, S., Martinc, M., and Pivovarova, L. (2021).
Scalable and interpretable semantic change detection.
In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4642–4652, Online. Association for Computational Linguistics.

References XII

Periti, F., Ferrara, A., Montanelli, S., and Ruskov, M. (2022).
What is done is done: an incremental approach to semantic shift detection.
In *Proceedings of the 3rd Workshop on Computational Approaches to Historical Language Change*, pages 33–43, Dublin, Ireland. Association for Computational Linguistics.

Pömsl, M. and Lyapin, R. (2020).
CIRCE at SemEval-2020 task 1: Ensembling context-free and context-dependent word representations.
In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 180–186, Barcelona (online). International Committee for Computational Linguistics.

References XIII

Rachinskiy, M. and Arefyev, N. (2021).
Zeroshot crosslingual transfer of a gloss language model for semantic change detection.
In *Computational linguistics and intellectual technologies: Papers from the annual conference Dialogue*, volume 20, pages 578–586.

Rachinskiy, M. and Arefyev, N. (2022).
GlossReader at LSCDiscovery: Train to select a proper gloss in English – discover lexical semantic change in Spanish.
In *Proceedings of the 3rd Workshop on Computational Approaches to Historical Language Change*, pages 198–203, Dublin, Ireland. Association for Computational Linguistics.

References XIV

Rodina, J., Trofimova, Y., Kutuzov, A., and Artemova, E. (2021).
ELMo and BERT in semantic change detection for Russian.
In *Analysis of Images, Social Networks and Texts: 9th International Conference, AIST 2020, Skolkovo, Moscow, Russia, October 15–16, 2020, Revised Selected Papers 9*, pages 175–186. Springer.

Rosin, G. D., Guy, I., and Radinsky, K. (2022).
Time masking for temporal language models.
In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pages 833–841.

Rosin, G. D. and Radinsky, K. (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.

References XV

Rother, D., Haider, T., and Eger, S. (2020).
CMCE at SemEval-2020 task 1: Clustering on manifolds of contextualized embeddings to detect historical meaning shifts.
In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 187–193, Barcelona (online). International Committee for Computational Linguistics.

Ryzhova, A., Ryzhova, D., and Sochenkov, I. (2021).
Detection of semantic changes in Russian nouns with distributional models and grammatical features.
Computational linguistics and intellectual technologies: Papers from the annual conference Dialogue.

References XVI

- Schlechtweg, D., McGillivray, B., Hengchen, S., Dubossarsky, H., and Tahmasebi, N. (2020).
  SemEval-2020 task 1: Unsupervised lexical semantic change detection.
  In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 1–23, Barcelona (online). International Committee for Computational Linguistics.

- Teodorescu, D., von der Ohe, S., and Kondrak, G. (2022).
  UAlberta at LSCDiscovery: Lexical semantic change detection via word sense disambiguation.
  In *Proceedings of the 3rd Workshop on Computational Approaches to Historical Language Change*, pages 180–186, Dublin, Ireland. Association for Computational Linguistics.

References XVII

Zamora-Reina, F. D., Bravo-Marquez, F., and Schlechtweg, D. (2022).
LSCDiscovery: A shared task on semantic change discovery and detection in Spanish.
In *Proceedings of the 3rd Workshop on Computational Approaches to Historical Language Change*, pages 149–164, Dublin, Ireland. Association for Computational Linguistics.

Zhou, J. and Li, J. (2020).
TemporalTeller at SemEval-2020 task 1: Unsupervised lexical semantic change detection with temporal referencing.
In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 222–231, Barcelona (online). International Committee for Computational Linguistics.