

Diachronic Lexical Semantics with Computational Models

Andrey Kutuzov

23 November, 2024



- 1 Introduction to computational modeling of semantic change
- 2 Current computational approaches to diachronic semantic change
 - Distributional hypothesis
 - Data defines everything
 - Inter-relation between methods and data
 - Research infrastructure
- 3 Open research questions
 - Detailed semantic change modeling
 - Onomasiological semantic change modeling

Extra-short summary of the talk

- ▶ Words change their meaning over time while preserving their form
- ▶ We call these changes **diachronic semantic shifts**
- ▶ It is possible to estimate these changes (semi-)automatically using diachronic corpora



Qualitative studies of semantic change

- ▶ Diachronic semantics, as defined in general linguistics
- ▶ Foundational research in, e.g., classification of semantic change [Bréal, 1899, Stern, 1931]
- ▶ inherently limited in the number of analyzed texts and target words

Qualitative studies of semantic change

- ▶ Diachronic semantics, as defined in general linguistics
- ▶ Foundational research in, e.g., classification of semantic change [Bréal, 1899, Stern, 1931]
- ▶ inherently limited in the number of analyzed texts and target words

Quantitative studies of semantic change

- ▶ A field in Natural Language Processing / Computational Linguistics
- ▶ emerged over the past two decades, starting from [Sagi et al., 2009]
 1. meaning defined computationally
 2. comparison over time is quantified
- ▶ opened vast new arenas of research
- ▶ one can now verify manually devised classifications and laws of diachronic semantics
- ▶ quantitative studies can complement and better inform theory-based qualitative approaches.

Aspects of change one can model

- ▶ **Nature of change**
 - ▶ ...using various taxonomies
[Bloomfield, 1933, Blank, 1997]

Aspects of change one can model

► Nature of change

- ...using various taxonomies
[Bloomfield, 1933, Blank, 1997]

► Source of change

- external factors (e.g, technological or socio-cultural, 'to broadcast')
- internal factors (e.g., the degree of prototypicality, [Dubossarsky et al., 2015])

Aspects of change one can model

- ▶ **Nature of change**
 - ▶ ...using various taxonomies
[Bloomfield, 1933, Blank, 1997]
- ▶ **Source of change**
 - ▶ external factors (e.g, technological or socio-cultural, 'to broadcast')
 - ▶ internal factors (e.g., the degree of prototypicality, [Dubossarsky et al., 2015])
- ▶ **Timescale (temporal granularity)**
 - ▶ short-term (weeks, months), long-term (decades, centuries)

Aspects of change one can model

► Nature of change

- ...using various taxonomies
[Bloomfield, 1933, Blank, 1997]

► Source of change

- external factors (e.g, technological or socio-cultural, 'to broadcast')
- internal factors (e.g., the degree of prototypicality, [Dubossarsky et al., 2015])

► Timescale (temporal granularity)

- short-term (weeks, months), long-term (decades, centuries)

► Important: level of analysis

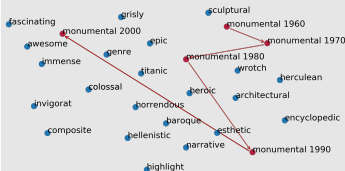
Level	Description
1. Detection	Word changed its meaning?
2. Descriptive categorization	Type of change?
3. <i>Mechanistic explanation</i>	How the change came to be?
4. Causal (internal)	Linguistic variables predicting change
5. Causal (external)	Non-linguistic variables predicting change

(**active research**, *early attempts*, future work)

Example applications

1. Detecting large-scale semantic change for lexicography and historical linguistics:

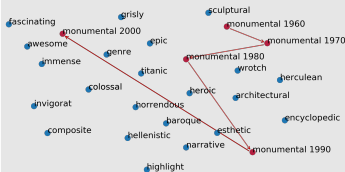
- ▶ comparing change intensity of verbs, nouns, and adjectives [Dubossarsky et al., 2016]
- ▶ testing whether evaluative adjectives in multiple languages are more prone to diachronic semantic change [Rodina et al., 2019]
- ▶ semantic change **discovery** [Zamora-Reina et al., 2022]



*Semantics of 'monumental':
from sculptures in the 1960s to
'awesome' in the 2000s*

[Rodina et al., 2019]

Example applications



*Semantics of 'monumental':
from sculptures in the 1960s to
'awesome' in the 2000s*
[Rodina et al., 2019]

1. Detecting large-scale semantic change for lexicography and historical linguistics:
 - ▶ comparing change intensity of verbs, nouns, and adjectives [Dubossarsky et al., 2016]
 - ▶ testing whether evaluative adjectives in multiple languages are more prone to diachronic semantic change [Rodina et al., 2019]
 - ▶ semantic change **discovery** [Zamora-Reina et al., 2022]
2. Conceptual (onomasiological) change for broader humanities and social sciences:
 - ▶ testing whether the concepts related to harm in psychology have broadened over the last four decades [Vylomova et al., 2019]
 - ▶ studying cultural evolution of racism via similarities between core and related concept terms [Sommerauer and Fokkens, 2019]
 - ▶ the concept of 'market' becoming increasingly abstract over time [Ohlsson et al., 2022]

- 1 Introduction to computational modeling of semantic change
- 2 Current computational approaches to diachronic semantic change
 - Distributional hypothesis
 - Data defines everything
 - Inter-relation between methods and data
 - Research infrastructure
- 3 Open research questions
 - Detailed semantic change modeling
 - Onomasiological semantic change modeling

Distributional hypothesis



Distributional hypothesis is at the core of modern computational models of semantic change.

Distributional hypothesis

Distributional hypothesis is at the core of modern computational models of semantic change.

Main points

prompt his an incident in my life as
 ek her out all courteously, PETKOFF (over his
 could be done, too," he remarked, sipping his
 , of us. I should like a cup of
 UKA (innocently). Perhaps you would like some
 her in public because he has fallen head
 manners for he was novels, broken backed,
 stretches her hand across the table for the
 little sitting-room, and cigarettes, after the
 had just given me a pannikin of hot
 a heavy roll coming; tried to save my
 I'll have a claret cup instead of
 t of trouble travelling. And then, with fresh
 , Your word had such weight with me!" fresh
 wont press you. "Try a weed with your
 ed with your coffee. Local tobacco. The black
 a had breakfasted when Strether came into the
 t, whispered excitedly: "They've got some hot
 ; but he was drinking a small cup of
 like an attaché. At last he finished his
 only possible come; and get me some fresh
 heap of pillows, the spout of a tin
 r the table with his knuckles propped amongst
 establishments, besides two lively stables, a
 en things, because their methods of family to
 , as, indeed, A waiter comes in with the
 let me always remain here. "I prefer weak
 hell. I should think you had drunk enough
 . Not a bit. Come in and have some
 responsibility. And greatly as we enjoyed our
 your naturally liking me. (She is and had
) Tell them I shan't be home to
 , that was Mr. McComas will not come to
 , or asked you to have a cup of
 woman can hardly know one places Gilbey's
 of my - my hopes." BROADBENT. He'll want
 : THE MANAGER. Can I take any order? Some
 to Tramp.) Will you drink a sup of
 are trying to sleep." the evening after your
 GUINNESS. I'll go get you some fresh
 sional men, artists, and even with laborers
 the Lutchies and Mrs. Rance the attendance at
 she came over to the great house to
 . The Baroness found it amusing to go to
 tea; she dressed as if for dinner. The
 would be dead in two years, as the

coffee for breakfast. Of course, has too Two fig
 coffee and cigarette). I don't believe in going
 coffee. "Bury him in some sort." I explained. "One
 coffee. MICHAEL. If you'd come in better hours,
 coffee, sir? DISCOVERY ANTICIPATED BY DIVINATION
 coffee-colored heathens and gesticlential white agi
 coffee stained, torn and "This was the last time
 coffee pot.) welcome, an expression which drops in
 coffee, had been permitted by the ladies, and in
 coffee....Slapped it down there, on my chest-bange
 coffee, burnt my fingers....and fell out of my
 coffee. Put some first night that we've come
 coffee, a clean cup, and a brandy bottle on
 coffee? He gave his friend a glance as to
 coffee. Local tobacco. The black coffee you got at
 coffee you get at the Amarilla, sir, you don'
 coffee-room; but, Waymarsh not having yet emerged,
 coffee.... Bosun got it.... "No!...Where?"....
 coffee, which had been served to him on a
 coffee and lit a cigarette. Presently a small boy
 coffee. The Second. Oliver Goldsmith sang what he
 coffee-pot. The patent log on the saffrail periodi
 coffee-cups, liqueur-glasses, cigar-ends. "I seeme
 tea They never boasted of Robert Acton, nor indulg
 tea at once, pose, which you carry so well
 tea. He places the tray on the table. Jasper
 tea!" cried Daisy, and she went off with the
 tea in Chin a. Life; it is a failure.
 tea. Stay to dinner, every year. Don't persist
 tea Crusoe island. Then there's the religious diff
 tea in the evening. Afraid though as he was
 tea, will you, LADY BRITTONBART. I must get the
 tea, ma'am: he has gone to call upon
 tea. It's not human, ugly woman must have
 tea on the table before him). The lady that
 tea. Let us have some. BURGE-LUBIN (recolutely ge
 tea? would the SHE-ANCIENT. Speak, Arjillax: you v
 tea with myself and the the happiest person in
 tea. "Better still-then there you are!" And Streth
 tea, ducky. (She takes up the its burden, is
 tea services out and made the people who had
 tea just in the right place on the west
 tea. She had let the proposal that she should
 tea; she dressed as if for dinner. The tea-
 tea-table offered an anomalous and picturesque rep
 tea-table. Be serious, Felix. You forget that I

- ▶ 'similarity of meaning is similarity of usage'
[Harris, 1954, Firth, 1957]
- ▶ meaning representations from co-occurrence statistics
- ▶ efficient **distributional semantic models** became possible with the current availability of vast digitized corpora
- ▶ the meaning of a linguistic unit is represented with a vector of real numbers
 - ▶ makes it possible to **quantify similarity** between units!
 - ▶ intrinsic evaluation tasks (word similarity, word analogies): empirical evidence in favor of distributional hypothesis

Distributional hypothesis

Distributional hypothesis is at the core of modern computational models of semantic change.

Main points

prompt his an incident in my life as
 ek her out all courteously, PETKOFF (over his
 could be done, too," he remarked, sipping his
 , of us. I should like a cup of
 UKA (innocently). Perhaps you would like some
 her in public because he has fallen head
 manners for he was novels, broken backed,
 stretches her hand across the table for the
 little sitting-room, and cigarettes, after the
 had just given me a pannikin of hot
 a heavy roll coming; tried to save my
 I'll have a claret cup instead of
 t of trouble travelling. And then, with fresh
 , Your word had such weight with me!" fresh
 wont press you. "Try a weed with your
 ed with your coffee. Local tobacco. The black
 a had breakfast when St.licher came into the
 t, whispered excitedly: "They've got some hot
 ; but he was drinking a small cup of
 like an attaché. At last he finished his
 only possible come; and get me some fresh
 heap of pillows, the spout of a tin
 r the table with his knuckles propped amongst
 establishments, besides two lively stables, a
 en things, because their methods of family to
 , as, indeed, A waiter comes in with the
 let me always remain here. "I prefer weak
 hell. I should think you had drunk enough
 . Not a bit. Come in and have some
 responsibility. And greatly as we enjoyed our
 your naturally liking me. (She is and had
) Tell them I shan't be home to
 , that was Mr. McComas will not come to
 , or asked you to have a cup of
 woman can hardly know one places Gilbey's
 of my - my hopes. "BROADBENT, He'll want
 : THE MANAGER. Can I take any order? Some
 to Tramp.) Will you drink a sup of
 are trying to sleep." the evening after your
 GUINNESS. I'll go get you some fresh
 sional men, artists, and even with laborers
 the Latches and Mrs. Rance the attendance at
 she came over to the great house to
 . The Baroness found it amusing to go to
 tea; she dressed as if for dinner. The
 would be dead in two years, as the

coffee for breakfast. Of course, has too Two fig
 coffee and cigarette). I don't believe in going
 coffee. "Bury him in some sort." I explained. "One
 coffee. MICHAEL. If you'd come in better hours,
 coffee, sir? DISCOVERY ANTICIPATED BY DIVINATION
 coffee-colored heathens and gesticulating white agi
 coffee stained, torn and "This was the last time
 coffee pot.) welcome, an expression which drops in
 coffee, had been permitted by the ladies, and in
 coffee....Slapped it down there, on my chest-bange
 coffee, burnt by fingers....and fell out of my
 coffee. Put some first night that we've come
 coffee, a clean cup, and a brandy bottle on
 coffee? He gave his friend a glance as to
 coffee. Local tobacco. The black coffee you got at
 coffee you get at the Amarilla, sir, you don't
 coffee-room; but, Weymarsh not having yet emerged,
 coffee.... Bosun got it...." "No!...Where?"....
 coffee, which had been served to him on a
 coffee and lit a cigarette. Presently a small boy
 coffee. The Second. Oliver Goldsmith sang what he
 coffee-pot. The patent log on the saffrail periodi
 coffee-cups, liqueur-glasses, cigar-ends. "I seeme
 tea They never boasted of Robert Acton, nor indulg
 tea at once, pose, which you carry so well
 tea. He places the tray on the table. Jasper
 tea!" cried Daisy, and she went off with the
 tea in Chin a. Life; it is a failure.
 tea. Stay to dinner, every year. Don't persist
 tea Crusoe island. Then there's the religious diff
 tea in the evening. Afraid though as he was
 tea, will you, LADY BRITOMART. I must get the
 tea, ma'am: he has gone to call upon
 tea. It's not human, ugly woman must have
 tea on the table before him). The lady that
 tea. Let us have some. BURGE-LUBIN (rejoicefully ge
 tea? would THE SHE-ANCIENT. Speak, Artilian: you v
 tea with myself and the the happiest person in
 tea. "Better still-then there you are!" And Streth
 tea, ducky. [She takes up the its burden, is
 tea services out and made the people who had
 tea just in the right place on the west
 tea. She had let the proposal that she should
 tea; she dressed as if for dinner. The tea-
 tea-table offered an anomalous and picturesque rep
 tea-table. Be serious, Felix. You forget that I

- ▶ 'similarity of meaning is similarity of usage'
[Harris, 1954, Firth, 1957]
- ▶ meaning representations from co-occurrence statistics
- ▶ efficient **distributional semantic models** became possible with the current availability of vast digitized corpora
- ▶ the meaning of a linguistic unit is represented with a vector of real numbers
 - ▶ makes it possible to **quantify similarity** between units!
 - ▶ intrinsic evaluation tasks (word similarity, word analogies): empirical evidence in favor of distributional hypothesis
- ▶ dominant approach in modern NLP, machine learning and AI.

More details on the history of distributional hypothesis in [Brunila and LaViolette, 2022].

Why data-driven?

- ▶ For most languages and time periods, we lack manually engineered historical ontologies or historical dictionaries.
- ▶ Hence, the need for **unsupervised and statistical methods** to infer meaning from corpora.
- ▶ The corpora are assumed to be representative samples of the languages.
- ▶ But how to analyze the data? That is, how to **represent** semantic entities in a machine-readable way?

Why data-driven?

- ▶ For most languages and time periods, we lack manually engineered historical ontologies or historical dictionaries.
- ▶ Hence, the need for **unsupervised and statistical methods** to infer meaning from corpora.
- ▶ The corpora are assumed to be representative samples of the languages.
- ▶ But how to analyze the data? That is, how to **represent** semantic entities in a machine-readable way?

Meaning representations

- ▶ In principle, a word can be represented by its corpus frequency only: useful in detecting the emergence of neologisms [Ryskina et al., 2020]
 - ▶ just calculate the difference between target word frequencies in two time-specific corpora.

Why data-driven?

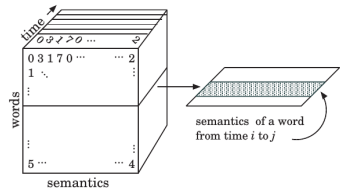
- ▶ For most languages and time periods, we lack manually engineered historical ontologies or historical dictionaries.
- ▶ Hence, the need for **unsupervised and statistical methods** to infer meaning from corpora.
- ▶ The corpora are assumed to be representative samples of the languages.
- ▶ But how to analyze the data? That is, how to **represent** semantic entities in a machine-readable way?

Meaning representations

- ▶ In principle, a word can be represented by its corpus frequency only: useful in detecting the emergence of neologisms [Ryskina et al., 2020]
 - ▶ just calculate the difference between target word frequencies in two time-specific corpora.
- ▶ But raw frequency is not enough to trace the semantic change of an existing word form.
- ▶ We need **word co-occurrence counts** in a corpus, to build time-dependent 'distributional profiles' of target words.

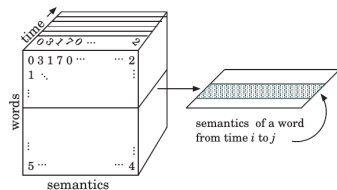
Diachronic semantics tensor

A sequence of distributional representations changing over time is a $Word \times SemanticVector \times Time$ tensor. The more different the time-specific vectors, the higher the degree of semantic change between the corresponding time bins. [Jurgens and Stevens, 2009]



Diachronic semantics tensor

A sequence of distributional representations changing over time is a $Word \times SemanticVector \times Time$ tensor. The more different the time-specific vectors, the higher the degree of semantic change between the corresponding time bins. [Jurgens and Stevens, 2009]



- ▶ Nowadays, we mostly use dense vectors, also known as '**word embeddings**', trained in an iterative unsupervised fashion with the objective of minimizing loss on the **language modeling (LM)** objective [Baroni et al., 2014].
- ▶ But also '**grammatical profiling**': comparing diachronic distributions of *morphological* and *syntactic* features [Kutuzov et al., 2021]
- ▶ **ensembling** grammatical profiles with LM-based embeddings improves the performance of semantic change detection [Giulianelli et al., 2022]

Type-based models for semantic change detection

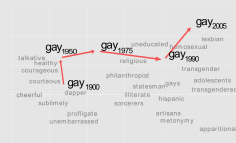
- ▶ Distributional semantic models with **one representation per word type**
 - ▶ Word2Vec [Mikolov et al., 2013] and FastText [Bojanowski et al., 2017]

Type-based models for semantic change detection

- ▶ Distributional semantic models with **one representation per word type**
 - ▶ Word2Vec [Mikolov et al., 2013] and FastText [Bojanowski et al., 2017]
- ▶ Algorithm:
 1. produce 2 vector representations for 'X' in 2 different time periods
 2. compute the **cosine similarity** between these 2 vectors: **change score**
 3. the representations are very similar → the meaning of 'X' did not change
 4. the representations are very dissimilar → we observe a semantic shift

Type-based models for semantic change detection

- Distributional semantic models with **one representation per word type**
 - Word2Vec [Mikolov et al., 2013] and FastText [Bojanowski et al., 2017]
- Algorithm:
 1. produce 2 vector representations for 'X' in 2 different time periods
 2. compute the **cosine similarity** between these 2 vectors: **change score**
 3. the representations are very similar → the meaning of 'X' did not change
 4. the representations are very dissimilar → we observe a semantic shift



[Kulkarni et al., 2015]

Type-based models for semantic change detection

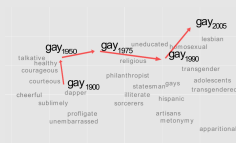
- Distributional semantic models with **one representation per word type**

- Word2Vec [Mikolov et al., 2013] and FastText [Bojanowski et al., 2017]

- Algorithm:

1. produce 2 vector representations for 'X' in 2 different time periods
2. compute the **cosine similarity** between these 2 vectors: **change score**
3. the representations are very similar → the meaning of 'X' did not change
4. the representations are very dissimilar → we observe a semantic shift

[Kim et al., 2014] first used Word2Vec for semantic change; then [Shoemark et al., 2019] and others.



[Kulkarni et al., 2015]

Data defines everything



Type-based models for semantic change detection

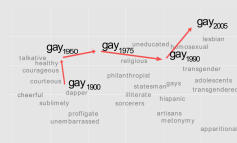
- ▶ Distributional semantic models with **one representation per word type**

- ▶ Word2Vec [Mikolov et al., 2013] and FastText [Bojanowski et al., 2017]

- ▶ Algorithm:

1. produce 2 vector representations for 'X' in 2 different time periods
2. compute the **cosine similarity** between these 2 vectors: **change score**
3. the representations are very similar → the meaning of 'X' did not change
4. the representations are very dissimilar → we observe a semantic shift

[Kim et al., 2014] first used Word2Vec for semantic change; then [Shoemark et al., 2019] and others.



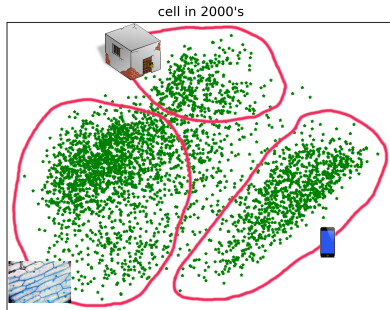
[Kulkarni et al., 2015]

Limitations of type-based models

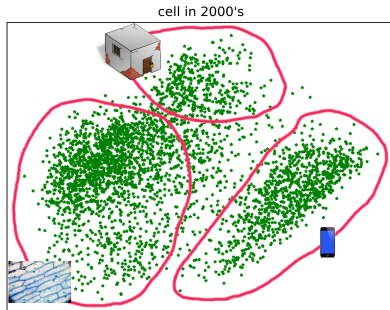
- ▶ **all the senses of ambiguous words are merged into one vector**
 - ▶ difficult to model less dominant senses
- ▶ **The problem of alignment**
 - ▶ making embedding spaces from different time periods **comparable**
 - ▶ orthogonal linear transformations, etc, but still a problem.

Token-based models for semantic change detection ('contextualized models')

- ▶ Large language models (LLMs) with **one representation per word token**
 - ▶ ELMo [Peters et al., 2018], BERT [Devlin et al., 2019], XLM-R [Conneau et al., 2020]...
- ▶ After training, the model produces **token embeddings**: one and the same target word will have different representations in different contexts... and in different **senses**!
- ▶ diachronic corpora are now used to **infer** word representations, not to **train** them.
- ▶ **No need for alignment:**
 - ▶ embeddings for tokens from time-specific corpora come from one LLM: comparable by design.



Token embeddings for the word 'cell' in three different senses in a corpus from 2000s. One dot = one usage in context.



Token embeddings for the word 'cell' in three different senses in a corpus from 2000s. One dot = one usage in context.

Typical methods

- ▶ Pretrained contextualized models are now the *de facto* standard [Montanelli and Periti, 2023]
- ▶ Change score is produced from **usage matrices** in various ways:
 1. clustering into 'senses' and comparing their time-specific distributions [Martinc et al., 2020]
 2. average pairwise cosine distances (APD) between token embeddings [Kutuzov and Giulianelli, 2020]
 3. LLM fine-tuned on Word-in-Context (WiC) task: XL-LEXEME [Cassotti et al., 2023]



- ▶ Different combinations of methods and datasets yield different answers
- ▶ We observe corpora-specific properties, not language-specific properties. And corpora **are** different!

- ▶ Different **combinations of methods and datasets** yield different answers
- ▶ We observe **corpora-specific properties**, not language-specific properties. And corpora **are** different!
- ▶ **Time span**:
 - ▶ Twitter corpora covering one or a few years or Latin corpora covering millennia?
- ▶ **Gaps between time periods**:
 - ▶ Monthly data bins or century-long gaps and century-long bins?
- ▶ **The number of time bins**:
 - ▶ Two time periods or hundreds of (daily, monthly, or yearly) time periods?
- ▶ **The size of the corpora**:
 - ▶ Millions of words? Billions of words? Thousands of words?
- ▶ **Genre of the texts**:
 - ▶ Information-dense like newspapers or sparse like fiction books?

The suitability of certain methods to different datasets is still an open question.

Diachronic corpora

The most fundamental resource for semantic change modeling.

Corpus	Size, words	Reference
Helsinki Corpus	10^6	[Rissanen, 1994]
New York Times Annotated Corpus	$\approx 2 \times 10^9$	[Sandhaus, 2008]
Google Books Ngrams	$\approx 100 \times 10^9$	[Michel et al., 2011]
English Gigaword	$\approx 4 \times 10^9$	[Parker et al., 2011]
Corpus of Historical American English (COHA)	400×10^6	[Davies, 2012]
Amazon Movie Reviews	$\approx 9 \times 10^8$	[McAuley and Leskovec, 2013]
Twitter (multilingual)	$\approx 14 \times 10^9$	[Banda et al., 2020]
HPLT web corpora (multilingual)	$\approx 7 \times 10^{12}$	[de Gibert et al., 2024]
...	...	...

Many other diachronic corpora exist for specific languages, see
<https://www.clarin.eu/resource-families/historical-corpora>.

NLP is based on empirically validating the proposed approaches using test sets (benchmarks).

NLP is based on empirically validating the proposed approaches using test sets (benchmarks).

Benchmarks for computational semantic change detection

- ▶ 'Traditional' diachronic linguistics

[Traugott and Dasher, 2001, Dobrushina and Daniel, 2016]: too few target words and examples to evaluate automatic systems.

NLP is based on empirically validating the proposed approaches using test sets (benchmarks).

Benchmarks for computational semantic change detection

- ▶ ‘Traditional’ diachronic linguistics
[Traugott and Dasher, 2001, Dobrushina and Daniel, 2016]: too few target words and examples to evaluate automatic systems.
- ▶ DatSemShift database [Zalizniak, 2018]:
 - ▶ not clear how to use it for evaluation of data-driven systems

NLP is based on empirically validating the proposed approaches using test sets (benchmarks).

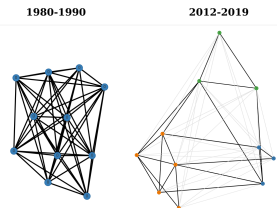
Benchmarks for computational semantic change detection

- ▶ ‘Traditional’ diachronic linguistics
[Traugott and Dasher, 2001, Dobrushina and Daniel, 2016]: too few target words and examples to evaluate automatic systems.
- ▶ DatSemShift database [Zalizniak, 2018]:
 - ▶ not clear how to use it for evaluation of data-driven systems
- ▶ **DWUGs**: ‘diachronic word usage graphs’ [Schlechtweg et al., 2021]
 - ▶ **graded** contextual word meaning annotation
 - ▶ human annotators estimate **semantic similarity of pairs of target word usages on a graded scale**
 - ▶ per-word **graphs**: word usages as nodes, annotators’ judgments as weights on the edges
 - ▶ graph clustered into communities, representing (diachronic) senses.
 - ▶ **change score** ← changes in cluster distributions between time periods

NLP is based on **empirically validating the proposed approaches using test sets (benchmarks)**.

Benchmarks for computational semantic change detection

- ▶ ‘Traditional’ diachronic linguistics
[Traugott and Dasher, 2001, Dobrushina and Daniel, 2016]: too few target words and examples to evaluate automatic systems.
- ▶ DatSemShift database [Zalizniak, 2018]:
 - ▶ not clear how to use it for evaluation of data-driven systems
- ▶ **DWUGs**: ‘diachronic word usage graphs’ [Schlechtweg et al., 2021]
 - ▶ **graded** contextual word meaning annotation
 - ▶ human annotators estimate **semantic similarity of pairs of target word usages on a graded scale**
 - ▶ per-word **graphs**: word usages as nodes, annotators’ judgments as weights on the edges
 - ▶ graph clustered into communities, representing (diachronic) senses.
 - ▶ **change score** $\leftarrow$ changes in cluster distributions between time periods



Excerpts from the test sets for SemEval 2020 (target words and change scores)

English (1810:1860 – 1960:2010)

- ▶ plane_nn 0.88
- ▶ tip_vb 0.68
- ▶ prop_nn 0.62
- ▶ ...
- ▶ fiction_nn 0.02
- ▶ chairman_nn 0.00
- ▶ risk_nn 0.00

Swedish (1790:1830 – 1895:1903)

- ▶ medium 0.60
- ▶ krita 0.44
- ▶ motiv 0.35
- ▶ ...
- ▶ uträtta 0.00
- ▶ vaktmästare 0.00
- ▶ vegetation 0.00

[Schlechtweg et al., 2020]

- ▶ Performance is evaluated with **Spearman rank correlation between gold and predicted scores**
- ▶ Test sets for Russian [Kutuzov and Pivovarova, 2021b], Spanish [Zamora-Reina et al., 2022], Norwegian [Kutuzov et al., 2022], Chinese [Chen et al., 2023], German [Schlechtweg et al., 2024]...

Shared tasks in NLP

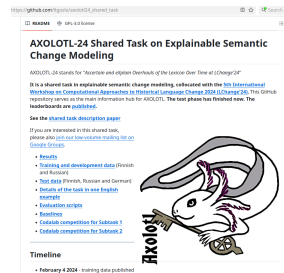
- ▶ Organizers publish the task and a test dataset without the correct answers
- ▶ Participants prepare their systems and submit the predictions for the test set
- ▶ Systems are ranked by their performance: **state-of-the-art** is established.
- ▶ Best models and the datasets are published openly, to the benefit of the wide community.

Shared tasks in NLP

- ▶ Organizers publish the task and a test dataset without the correct answers
- ▶ Participants prepare their systems and submit the predictions for the test set
- ▶ Systems are ranked by their performance: **state-of-the-art** is established.
- ▶ Best models and the datasets are published openly, to the benefit of the wide community.

Semantic change modeling shared tasks

- ▶ English, German, Latin, Swedish [Schlechtweg et al., 2020]
- ▶ Italian [Basile et al., 2020], Russian [Kutuzov and Pivovarova, 2021a]
- ▶ Spanish [Zamora-Reina et al., 2022]
- ▶ AXOLOTL: **explainable** semantic change for Finnish, Russian, German [Fedorova et al., 2024]



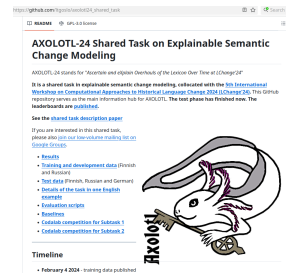
Shared tasks in NLP

- ▶ Organizers publish the task and a test dataset without the correct answers
- ▶ Participants prepare their systems and submit the predictions for the test set
- ▶ Systems are ranked by their performance: **state-of-the-art** is established.
- ▶ Best models and the datasets are published openly, to the benefit of the wide community.

Semantic change modeling shared tasks

- ▶ English, German, Latin, Swedish [Schlechtweg et al., 2020]
- ▶ Italian [Basile et al., 2020], Russian [Kutuzov and Pivovarova, 2021a]
- ▶ Spanish [Zamora-Reina et al., 2022]
- ▶ AXOLOTL: **explainable** semantic change for Finnish, Russian, German [Fedorova et al., 2024]

Starting from 2021, **the leaderboards are dominated by token-based approaches.** XL-LEXEME [Cassotti et al., 2023] is current state-of-the-art for most languages.



- 1 Introduction to computational modeling of semantic change
- 2 Current computational approaches to diachronic semantic change
 - Distributional hypothesis
 - Data defines everything
 - Inter-relation between methods and data
 - Research infrastructure
- 3 Open research questions
 - Detailed semantic change modeling
 - Onomasiological semantic change modeling

Future work

1. **Sub-classification of types of semantic shifts** (e.g., broadening, narrowing, metaphorization)
[Mitra et al., 2014, Tahmasebi and Risse, 2017, Giulianelli et al., 2023].
2. **Identifying the source of a change** (e.g., linguistic or extralinguistic causes).
 - ▶ closely linked to the division between linguistic drifts and cultural shifts [Hamilton et al., 2016]
 - ▶ lack of manually annotated datasets of linguistically motivated and culturally motivated semantic shifts
3. **Identifying groups of words that change their meaning together**
[Dubossarsky et al., 2016, Rodina et al., 2019]
 - ▶ linked to 'laws of semantic change' and co-lexification processes

More future work

- ▶ Most recent studies are **semasiological** (anchored on form):
 - ▶ *'How the meaning of the term controller has changed from the 15th century to today?'*
- ▶ This is natural for distributional representations.
- ▶ More work on **onomasiological** change is needed (anchored on meaning)
 - ▶ *'What you call a handheld peripheral device designed to provide input to a computer or a gaming console, and what this object was called in the 60s?'*
- ▶ Analogous to looking up words associated with a certain target concept, rather than looking up the meaning of a certain target word.
- ▶ More difficult! But see [Tahmasebi et al., 2012, Yao et al., 2018, Szymanski, 2017].
- ▶ This is where DatSemShift can be of great use.

That's it!



Thanks you!
Questions are welcome



Banda, J. M., Tekumalla, R., Wang, G., Yu, J., Liu, T., Ding, Y., and Chowell, G. (2020). A twitter dataset of 150+ million tweets related to covid-19 for open research.

Technical report, Panacea lab.

This dataset will be updated bi-weekly at least with additional tweets, look at the github repo for these updates.



Baroni, M., Dinu, G., and Kruszewski, G. (2014).

Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors.

In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, volume 1, pages 238–247, Baltimore, USA.

References II

-  Basile, P., Caputo, A., Caselli, T., Cassotti, P., and Varvara, R. (2020).
DIACR-Ita @ EVALITA2020: Overview of the EVALITA2020 Diachronic Lexical Semantics
(DIACR-Ita) Task.
Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and
Speech Tools for Italian. Final Workshop (EVALITA 2020) CEUR Workshop Proceedings
(CEUR-WS.org).
-  Blank, A. (1997).
Prinzipien des lexikalischen Bedeutungswandels am Beispiel der romanischen Sprachen.
Max Niemeyer Verlag, Berlin, Boston.
-  Bloomfield, L. (1933).
Language.
Allen & Unwin.

References III



Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017).
Enriching word vectors with subword information.
Transactions of the Association of Computational Linguistics, 5:135–146.



Bréal, M. (1899).
Essai de sémantique (2nd ed.).
Hachette, Paris.



Brunila, M. and LaViolette, J. (2022).
What company do words keep? revisiting the distributional semantics of J.R. firth & zellig
Harris.
In Carpuat, M., de Marneffe, M.-C., and Meza Ruiz, I. V., editors, Proceedings of the 2022
Conference of the North American Chapter of the Association for Computational
Linguistics: Human Language Technologies, pages 4403–4417, Seattle, United States.
Association for Computational Linguistics.



Cassotti, P., Siciliani, L., DeGemmis, M., Semeraro, G., and Basile, P. (2023). XL-LEXEME: WiC pretrained model for cross-lingual LEXical sEMantic changE. In Rogers, A., Boyd-Graber, J., and Okazaki, N., editors, 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.



Chen, J., Chersoni, E., Schlechtweg, D., Prokic, J., and Huang, C.-R. (2023). ChiWUG: A graph-based evaluation dataset for Chinese lexical semantic change detection. In Tahmasebi, N., Montariol, S., Dubossarsky, H., Kutuzov, A., Hengchen, S., Alfter, D., Periti, F., and Cassotti, P., editors, Proceedings of the 4th Workshop on Computational Approaches to Historical Language Change, pages 93–99, Singapore. Association for Computational Linguistics.



Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., and Stoyanov, V. (2020).

Unsupervised cross-lingual representation learning at scale.

In Jurafsky, D., Chai, J., Schluter, N., and Tetreault, J., editors, Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 8440–8451, Online. Association for Computational Linguistics.



Davies, M. (2012).

Expanding horizons in historical linguistics with the 400-million word Corpus of Historical American English.

Corpora, 7(2):121–157.

-  de Gibert, O., Nail, G., Arefyev, N., Bañón, M., van der Linde, J., Ji, S., Zaragoza-Bernabeu, J., Aulamo, M., Ramírez-Sánchez, G., Kutuzov, A., Pyysalo, S., Oepen, S., and Tiedemann, J. (2024).
A new massive multilingual dataset for high-performance language technologies.
In Calzolari, N., Kan, M.-Y., Hoste, V., Lenci, A., Sakti, S., and Xue, N., editors, Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), pages 1116–1128, Torino, Italia. ELRA and ICCL.
-  Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019).
BERT: Pre-training of deep bidirectional transformers for language understanding.
In Burstein, J., Doran, C., and Solorio, T., editors, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

References VII

 Dobrushina, N. and Daniel, M. (2016).

Two centuries in twenty words.

NRU HSE, Moscow, Russian.

 Dubossarsky, H., Tsvetkov, Y., Dyer, C., and Grossman, E. (2015).

A bottom up approach to category mapping and meaning change.

In Proceedings of the NetWordS 2015 Word Knowledge and Word Usage, pages 66–70,
Pisa, Italy.

 Dubossarsky, H., Weinshall, D., and Grossman, E. (2016).

Verbs change more than nouns: A bottom-up computational approach to semantic change.

Lingue e linguaggio, 15(1):7–28.

-  Fedorova, M., Mickus, T., Partanen, N., Siewert, J., Spaziani, E., and Kutuzov, A. (2024). AXOLOTL'24 shared task on multilingual explainable semantic change modeling. In Tahmasebi, N., Montariol, S., Kutuzov, A., Alfter, D., Periti, F., Cassotti, P., and Huebscher, N., editors, Proceedings of the 5th Workshop on Computational Approaches to Historical Language Change, pages 72–91, Bangkok, Thailand. Association for Computational Linguistics.
-  Firth, J. (1957). A synopsis of linguistic theory, 1930-1955. Blackwell.



Giulianelli, M., Kutuzov, A., and Pivovarova, L. (2022).

Do not fire the linguist: Grammatical profiles help language models detect semantic change.

In Tahmasebi, N., Montariol, S., Kutuzov, A., Hengchen, S., Dubossarsky, H., and Borin, L., editors, Proceedings of the 3rd Workshop on Computational Approaches to Historical Language Change, pages 54–67, Dublin, Ireland. Association for Computational Linguistics.



Giulianelli, M., Luden, I., Fernandez, R., and Kutuzov, A. (2023).

Interpretable word sense representations via definition generation: The case of semantic change analysis.

In Rogers, A., Boyd-Graber, J., and Okazaki, N., editors, 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.

References X



Hamilton, W., Leskovec, J., and Jurafsky, D. (2016).
Cultural shift or linguistic drift? Comparing two computational measures of semantic change.
In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pages 2116–2121. Association for Computational Linguistics.



Harris, Z. S. (1954).
Distributional structure.
Word, 10(2-3):146–162.



Jurgens, D. and Stevens, K. (2009).
Event detection in blogs using temporal random indexing.
In Orasan, C., Hasler, L., and Forăscu, C., editors, Proceedings of the Workshop on Events in Emerging Text Types, pages 9–16, Borovets, Bulgaria. Association for Computational Linguistics.

References XI



Kim, Y., Chiu, Y.-I., Hanaki, K., Hegde, D., and Petrov, S. (2014).
Temporal analysis of language through neural language models.
In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, page 61, Baltimore, USA.



Kulkarni, V., Al-Rfou, R., Perozzi, B., and Skiena, S. (2015).
Statistically significant detection of linguistic change.
In Proceedings of the 24th International Conference on World Wide Web, pages 625–635, Florence, Italy.



Kutuzov, A. and Giulianelli, M. (2020).
UiO-UvA at SemEval-2020 task 1: Contextualised embeddings for lexical semantic change detection.
In Herbelot, A., Zhu, X., Palmer, A., Schneider, N., May, J., and Shutova, E., editors, Proceedings of the Fourteenth Workshop on Semantic Evaluation, pages 126–134, Barcelona (online). International Committee for Computational Linguistics.

 Kutuzov, A. and Pivovarova, L. (2021a).

RuShiftEval: A shared task on semantic shift detection for Russian.

Computational linguistics and intellectual technologies: Papers from the annual conference Dialogue, pages 533–546.

 Kutuzov, A. and Pivovarova, L. (2021b).

Three-part diachronic semantic change dataset for Russian.

In Tahmasebi, N., Jatowt, A., Xu, Y., Hengchen, S., Montariol, S., and Dubossarsky, H., editors, Proceedings of the 2nd International Workshop on Computational Approaches to Historical Language Change 2021, pages 7–13, Online. Association for Computational Linguistics.

References XIII



Kutuzov, A., Pivovarova, L., and Giulianelli, M. (2021).
Grammatical profiling for semantic change detection.
In Bisazza, A. and Abend, O., editors, Proceedings of the 25th Conference on Computational Natural Language Learning, pages 423–434, Online. Association for Computational Linguistics.



Kutuzov, A., Touileb, S., Mæhlum, P., Enstad, T., and Wittemann, A. (2022).
NorDiaChange: Diachronic semantic change dataset for Norwegian.
In Calzolari, N., Béchet, F., Blache, P., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Isahara, H., Maegaard, B., Mariani, J., Mazo, H., Odijk, J., and Piperidis, S., editors, Proceedings of the Thirteenth Language Resources and Evaluation Conference, pages 2563–2572, Marseille, France. European Language Resources Association.



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

 McAuley, J. J. and Leskovec, J. (2013).

From amateurs to connoisseurs: Modeling the evolution of user expertise through online reviews.

In Proceedings of the 22Nd International Conference on World Wide Web, WWW '13, pages 897–908. ACM.

 Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., and Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. Science, 331(6014):176–182.

 Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. Advances in Neural Information Processing Systems 26, pages 3111–3119.



Mitra, S., Mitra, R., Riedl, M., Biemann, C., Mukherjee, A., and Goyal, P. (2014). That's sick dude!: Automatic identification of word sense change across different timescales.

In Toutanova, K. and Wu, H., editors, Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1020–1029, Baltimore, Maryland. Association for Computational Linguistics.



Montanelli, S. and Periti, F. (2023).

A survey on contextualised semantic shift detection.

arXiv preprint arXiv:2304.01666.



Ohlsson, C., Wählstrand Skärström, V., and Björck, H. (2022).

The market as a concept in swedish parliamentary records from 1867 to 1970: A mixed methods study.

In Digital Parliamentary Data in Action (DiPaDA 2022) workshop, Uppsala, Sweden, March 15, 2022, pages 22–34. CEUR-WS. org.



Parker, R., Graff, D., Kong, J., Chen, K., and Maeda, K. (2011).
English Gigaword Fifth Edition LDC2011T07.
Technical report, Technical Report. Linguistic Data Consortium, Philadelphia.



Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., and Zettlemoyer, L. (2018).
Deep contextualized word representations.
In Walker, M., Ji, H., and Stent, A., editors, Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pages 2227–2237, New Orleans, Louisiana. Association for Computational Linguistics.



Rissanen, M. (1994).
The Helsinki corpus of English texts.
In Corpora across the centuries, pages 73–79. Brill.



Rodina, J., Bakshandaeva, D., Fomin, V., Kutuzov, A., Touileb, S., and Velldal, E. (2019). Measuring diachronic evolution of evaluative adjectives with word embeddings: the case for English, Norwegian, and Russian.

In Tahmasebi, N., Borin, L., Jatowt, A., and Xu, Y., editors, Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change, pages 202–209, Florence, Italy. Association for Computational Linguistics.



Ryskina, M., Rabinovich, E., Berg-Kirkpatrick, T., Mortensen, D. R., and Tsvetkov, Y. (2020).

Where new words are born: Distributional semantic analysis of neologisms and their semantic neighborhoods.

Proceedings of the Society for Computation in Linguistics, 3(1):43–52.

References XVIII



Sagi, E., Kaufmann, S., and Clark, B. (2009).

Semantic density analysis: Comparing word meaning across time and phonetic space.
In Proceedings of the Workshop on Geometrical Models of Natural Language Semantics,
pages 104–111. Association for Computational Linguistics.



Sandhaus, E. (2008).

The New York Times annotated corpus overview.
Linguistic Data Consortium, Philadelphia, 6(12):e26752.



Schlechtweg, D., Cassotti, P., Noble, B., Alfter, D., Schulte Im Walde, S., and Tahmasebi, N. (2024).

More DWUGs: Extending and evaluating word usage graph datasets in multiple languages.
In Al-Onaizan, Y., Bansal, M., and Chen, Y.-N., editors, Proceedings of the 2024
Conference on Empirical Methods in Natural Language Processing, pages 14379–14393,
Miami, Florida, USA. Association for Computational Linguistics.

References XIX



Schlechtweg, D., McGillivray, B., Hengchen, S., Dubossarsky, H., and Tahmasebi, N. (2020).

SemEval-2020 task 1: Unsupervised lexical semantic change detection.

In Herbelot, A., Zhu, X., Palmer, A., Schneider, N., May, J., and Shutova, E., editors, Proceedings of the Fourteenth Workshop on Semantic Evaluation, pages 1–23, Barcelona (online). International Committee for Computational Linguistics.



Schlechtweg, D., Tahmasebi, N., Hengchen, S., Dubossarsky, H., and McGillivray, B. (2021).

DWUG: A large resource of diachronic word usage graphs in four languages.

In Moens, M.-F., Huang, X., Specia, L., and Yih, S. W.-t., editors, Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 7079–7091, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.



Shoemark, P., Liza, F. F., Nguyen, D., Hale, S., and McGillivray, B. (2019).
Room to Glo: A systematic comparison of semantic change detection approaches with
word embeddings.

In Inui, K., Jiang, J., Ng, V., and Wan, X., editors, Proceedings of the 2019 Conference on
Empirical Methods in Natural Language Processing and the 9th International Joint
Conference on Natural Language Processing (EMNLP-IJCNLP), pages 66–76, Hong Kong,
China. Association for Computational Linguistics.



Sommerauer, P. and Fokkens, A. (2019).
Conceptual change and distributional semantic models: An exploratory study on pitfalls
and possibilities.

In Proceedings of the 1st International Workshop on Computational Approaches to
Historical Language Change, pages 223–233.

References XXI



Stern, G. (1931).

Meaning and change of meaning; with special reference to the English language.

Wettergren & Kerbers, Göteborg.



Szymanski, T. (2017).

Temporal word analogies: Identifying lexical replacement with diachronic word embeddings.

In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 448–453. Association for Computational Linguistics.



Tahmasebi, N., Gossen, G., Kanhabua, N., Holzmann, H., and Risse, T. (2012).

NEER: An unsupervised method for named entity evolution recognition.

In Kay, M. and Boitet, C., editors, Proceedings of COLING 2012, pages 2553–2568, Mumbai, India. The COLING 2012 Organizing Committee.



Tahmasebi, N. and Risse, T. (2017).

Finding individual word sense changes and their delay in appearance.

In Mitkov, R. and Angelova, G., editors, Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017, pages 741–749, Varna, Bulgaria. INCOMA Ltd.



Traugott, E. and Dasher, R. (2001).

Regularity in semantic change.

Cambridge University Press.



Vylomova, E., Murphy, S., and Haslam, N. (2019).

Evaluation of semantic change of harm-related concepts in psychology.

In Tahmasebi, N., Borin, L., Jatowt, A., and Xu, Y., editors, Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change, pages 29–34, Florence, Italy. Association for Computational Linguistics.

-  Yao, Z., Sun, Y., Ding, W., Rao, N., and Xiong, H. (2018).
Dynamic word embeddings for evolving semantic discovery.
In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM '18, pages 673–681. ACM.
-  Zalizniak, A. (2018).
The catalogue of semantic shifts: 20 years later.
Russian Journal of Linguistics, 22(4):770–787.
-  Zamora-Reina, F. D., Bravo-Marquez, F., and Schlechtweg, D. (2022).
LSCDiscovery: A shared task on semantic change discovery and detection in Spanish.
In Tahmasebi, N., Montariol, S., Kutuzov, A., Hengchen, S., Dubossarsky, H., and Borin, L., editors, Proceedings of the 3rd Workshop on Computational Approaches to Historical Language Change, pages 149–164, Dublin, Ireland. Association for Computational Linguistics.