A Hypothesis-Driven Framework for Detecting Lexical Semantic Change

Pierluigi Cassotti[1], Nina Tahmasebi[1]

[1]University of Gothenburg, Department of Philosophy, Linguistics and Theory of Science, Gothenburg, Sweden

Abstract

This paper introduces a hypothesis-driven framework aimed at detecting lexical semantic change, addressing the limitations of current computational methods that struggle with the dynamic and contextually modulated nature of word meanings. Traditional approaches, such as Word Sense Disambiguation (WSD), fail to capture the fluidity of senses, whereas Word Sense Induction (WSI), while more flexible, lacks the precision necessary to align with predefined semantic structures. Our approach systematically combines expert-defined sense hypotheses with advanced computational techniques, including generative models, encoding and prototyping methods, and targeted semantic analysis. Using words historically significant in scientific contexts—such as theory, gene, and force—we demonstrate the effectiveness of our method in tracing fine semantic changes and metaphorical extensions over time, highlighting its advantages over naive computational strategies.

Keywords

lexical semantic change, lexical semantics, diachronic, historical linguistics

1. Introduction

Polysemy, the phenomenon where a single word carries multiple meanings, has long intrigued researchers. Often, words reach a polysemic state, through a process of semantic change in which the (set of) senses of a word has been altered. Dictionaries serve as vital resources in this field, cataloging the various senses of words. However, they are not all-encompassing and the granularity of the recorded senses varies across dictionaries, reflecting the approaches of lexicographers, who are often categorized as "lumpers" or "splitters." Lumpers favor broader, more encompassing definitions, while splitters distinguish senses with subtle nuances.

This variability ties into contextual modulation [1], where a word's core meaning remains stable but shifts slightly depending on its context. Such shifts become more pronounced over time, as word meanings evolve in response to cultural and social changes. For instance, the Oxford English Dictionary [2] defines "phone" simply as a “telephone apparatus,” a broad enough definition to encompass its evolution from landline phones to public telephone booths to modern smartphones.

This dynamic nature of meaning poses significant challenges for computational modeling. Traditional approaches like Word Sense Disambiguation (WSD) [3] struggle because they assume fixed meanings, ignoring the fluid continuity of senses. In contrast, Word Sense Induction (WSI) is better suited, as it derives sense structures directly from data. However, WSI's open-ended nature makes it challenging to align derived senses with a predefined ground truth, especially when attempting to track meaning changes across centuries of a language's history.

Current computational models often fail to align with ground truth sense representations unless explicitly guided. One way to address this is by starting with predefined search hypotheses, which can simplify the modeling process and provide a clearer framework for tracking meaning shifts over time.

By establishing research hypotheses, we can predefine the organization and structure of word senses, guiding computational models toward a predetermined ground truth. However, this remains challenging with standard technologies, which require models capable of adapting to meaning representations without relying on specific senses.

In this paper, we present our hypothesis-driven theoretical framework for detecting meaning change (Section 3). We also demonstrate a practical implementation of this framework using recently developed computational models (Section 2). Furthermore, we provide a concrete example by comparing our approach to naive WSI methods (Section 4), highlighting the advantages of the hypothesis-driven approach.

2. Related Work

Detecting changes in word meaning typically involves two stages: first, representing the meaning of words in individual time periods, and second, verifying whether a change in meaning has occurred over time.

CLiC-it 2025: Eleventh Italian Conference on Computational Linguistics, September 24 — 26, 2025, Cagliari, Italy

pierluigi.cassotti@gu.se (P. Cassotti); nina.tahmasebi@gu.se (N. Tahmasebi)
0000-0001-7824-7167 (P. Cassotti); 0000-0003-1688-1845 (N. Tahmasebi)

2.1. Representation of Word Meanings

Representing word meanings in historical texts poses unique challenges for computational models [4]. These models must understand historical contexts, avoid reliance on lexicographic resources that may omit new or obsolete senses, and ideally capture subtle temporal shifts within a word's meaning, rather than just the addition or removal of senses. For example, the word "horse" once referred to the primary mode of transportation but no longer holds that role in our daily lives today.

To address these challenges, approaches to representing word meanings often use a greater degree of freedom that allow for nuanced representations. Models for word meaning representation can be viewed on a continuum. At one end, Word Sense Disambiguation (WSD) models assign all instances of a word's meaning to a single sense, offering limited flexibility. At the other end, contextualized models [5] treat each instance as a unique entity, providing greater freedom but often encoding extraneous information, such as syntactic or morphological variations, which may not be relevant for tracking meaning change. WSD-based models, while precise, are often too rigid to capture subtle variations within a sense.

In recent years, research has focused on developing balanced solutions—models that are nearly as flexible as contextualized approaches but prioritize semantic characteristics over other linguistic aspects. This enables more effective modeling of contextual modulation.

One such model is XL-LEXEME [6], a bi-encoder based on SBERT [7] with a Siamese architecture and an XLM-R [8] backbone. XL-LEXEME has been trained on the Word-in-Context (WiC) [9] task to predict whether a target word has the same meaning in two given sentences (1 for the same meaning, 0 for different meanings). This is done by generating two XL-LEXEME vector representations of the word's meaning in each sentence by aggregating subword embeddings from the entire sentence. These vectors are compared using cosine similarity, and a contrastive loss function encourages higher similarity for matching meanings and lower similarity otherwise.

However, XL-LEXEME's output—cosine similarity scores between sentence pairs—lacks the interpretability needed to fully understand the processes underlying meaning change.

Recently, we have seen novel methods for modeling meaning, namely *definition generation*, where for a given target word in context, the method generates a dictionary-like definition [10, 11]. Such definition generation models produce definitions that capture the intended word meaning but may deviate from ground-truth definitions for three main reasons. First, like humans, models may express the same concept using different words, requiring mappings to the underlying sense. Second, errors such as hallucinations can compromise performance. Third, a model may generate a definition that reflects *contextual modulation.* While this is not rewarded in the evaluation of the models (where generated definitions are evaluated against dictionary definitions), it is often a desirable outcome when we want to study meaning change.

Another way to use the potential of large language models (LLMs) is by using them as computational annotators. This involves prompting instructed LLMs to interpret the meaning of a word (by solving the WiC task) in a zero-shot setting, without requiring task-specific training. For example, in [12], we compared GPT-4 with contextualized models like BERT and XL-LEXEME on tasks such as Word-in-Context (WiC), Word Sense Induction (WSI), and Lexical Semantic Change Detection (LSCD). The results demonstrate that XL-LEXEME and zero-shot GPT-4 perform comparably across all tasks, despite GPT-4 having significantly more parameters (1,000 times larger) and higher computational costs.

2.2. Detection of changes

The process for detecting changes in word meaning over time *typically* follows a standard pipeline, c.f. [13]:

1. Collect the occurrences of a word w over time, denoted as $U_1, U_2, \ldots, U_T$, where U_k represents the instances in which the word w appears at time k.
2. Encode the uses of the word into vectors, resulting in the sequence $V_1, V_2, \ldots, V_T$, where V_k represents the vectors encoding the uses of the word w at time k.
3. Select a metric m for comparing the vectors, chosen from the following options [14]:
   - **Average Pairwise Distance (APD)**: Computes and averages distances between all pairs of vectors from two time points.
   - **Prototype Distance (PRT)**: Calculates the distance between centroids (prototypes) of two time points.
   - **Cluster-based Jensen-Shannon Distance (JSD)**: Clusters data irrespective of time, computes the frequency of senses for each time period separately, treats them as probability distributions, and calculates the distance between two time points via Jensen-Shannon distance of the probability distributions.
4. Compare the vectors using the metric m according to a specific strategy, e.g.
   a) Comparison with the first period: $(V_1, V_2), (V_1, V_3), \ldots, (V_1, V_T)$
   b) Comparison with the last period: $(V_1, V_T), (V_2, V_T), \ldots, (V_{T-1}, V_T)$

c) Comparison with the previous period: $(V_1, V_2), (V_2, V_3), \ldots, (V_{T-1}, V_T)$

d) Comparison within a window of size k: $(V_i, (V_{i-k}, V_{i+k})), (V_{i+k}, (V_i, V_{i+2k})), \ldots$

To tailor the pipeline to specific computational models, certain modifications can be introduced. For definition generation, an additional step can be inserted after step (1). First, generate definitions for each instance of word use. Then, in step 2, encode these definitions into vectors instead of the word uses themselves. For large language models (LLMs) as computational annotators, LLMs provide a semantic distance value for pairs of word uses directly. In this case, steps (1) and (2) are bypassed, and the Average Pairwise Distance (APD) is used to compute the average distances between pairs of time points.

2.3. Historical Word Usage Generation

The study of lexical semantic change requires large-scale, diachronic sense-annotated corpora, yet such resources are scarce due to the time, expertise, and cost involved in annotating historical texts. To overcome this barrier, Janus [15], a generative model fine-tuned on the Llama 3 8B architecture using 1,191,851 example sentences from the Oxford English Dictionary (OED), was developed. Janus generates historically accurate and sense-specific word usages for any given *word*, its *sense definition*, and a *target year* from 1700 onward. This capability enables the creation of extensive datasets for tasks such as word sense disambiguation and detecting semantic shifts over time.

Janus produces sentences that reflect the intended meaning of a word in a specific historical context. Its performance was compared to baseline models, including GPT-3.5, GPT-4o, and Llama 3 Instruct variants, across three key metrics: (i) context variability, which measures the diversity of generated sentences to ensure varied expressions of the same sense; (ii) temporal accuracy, which assesses how well the language aligns with the specified historical period (e.g., avoiding "airplane" before 1903); and (iii) semantic accuracy, which evaluates how closely the generated sentences match the provided sense definition. Janus outperforms baselines in context variability and temporal accuracy, producing diverse sentences with a root mean squared error (RMSE) of 54.75 years for historical alignment (in line with the baseline). Qualitative analysis highlights Janus's ability to emulate temporal linguistic shifts, such as the declining use of archaic pronouns like "thee" and the evolving meaning of "awful" from impressive to negative.

3. Hypothesis-Driven LSCD

To investigate the historical evolution of word senses, we propose a hypothesis-driven methodology. For instance, a research hypothesis might posit that the word **gene** began to be used metaphorically shortly after its establishment in the biological sciences during the 1950s, reflecting its profound influence on modern thought. Our goal is to trace the evolution of **gene** across the 20th century and identify its earliest occurrences in various senses, a task traditionally performed by experts manually examining thousands of concordances.

A conventional word sense disambiguation (WSD) system, often based on resources like WordNet [16], is limited in this context. WordNet, for example, provides only a single definition for **gene**:

> *(genetics) a segment of DNA that is involved in producing a polypeptide chain; it can include regions preceding and following the coding DNA as well as introns between the exons; it is considered a unit of heredity.*

Instead, OED contains a second sense:

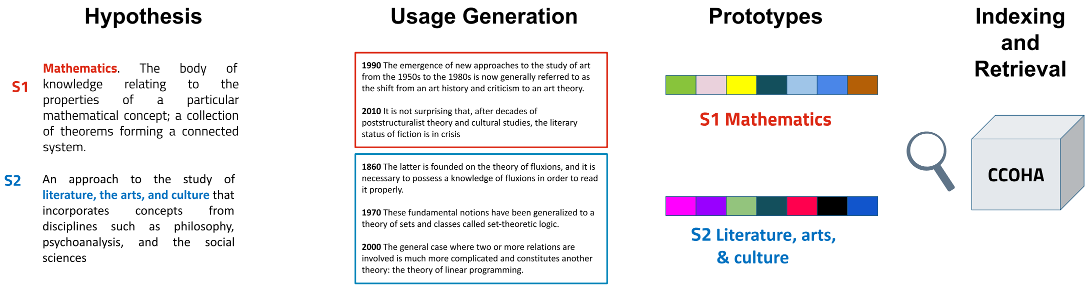
> *In figurative and extended use, esp. with reference to qualities regarded as deeply ingrained or (often humorously) as inherited. Often in plural.*

Such systems struggle with historical texts due to (i) their incompatibility with archaic language and (ii) their incomplete coverage of senses, particularly metaphorical or emerging uses. Large language model (LLM)-based models, on the other hand, offer improved sense identification but are computationally expensive and environmentally unsustainable for analyzing thousands of word occurrences in large historical corpora.

3.1. Our Approach

We propose a scalable, hypothesis-driven framework comprising three components: an encoder C, a prototyper P, and a comparison function F. This framework systematically analyzes word sense evolution by combining expert-defined sense definitions with computational techniques.

1. **Definition of Senses**: Let $S = \{s_1, s_2, \ldots, s_N\}$ represent a set of N sense definitions for the target word (e.g., *gene*), crafted to align with the research hypotheses. For each sense s_i, we use a generative model (e.g., Janus) to produce a collection of synthetic examples, $E_i = \{e_{i_1}, e_{i_2}, \ldots, e_{i_m}\}$, representing the word's usage in that sense across the target time period.



Mathematics. The body of knowledge relating to the properties of a particular mathematical concept; a collection of theorems forming a connected system.

An approach to the study of **literature, the arts, and culture** that incorporates concepts from disciplines such as philosophy, psychoanalysis, and the social sciences

1990 The emergence of new approaches to the study of art from the 1950s to the 1980s is now generally referred to as the shift from an art history and criticism to an art theory.

2010 It is not surprising that, after decades of poststructuralist theory and cultural studies, the literary status of fiction is in crisis

1860 The latter is founded on the theory of fluxions, and it is necessary to possess a knowledge of fluxions in order to read it properly.

1970 These fundamental notions have been generalized to a theory of sets and classes called set-theoretic logic.

2000 The general case where two or more relations are involved is much more complicated and constitutes another theory: the theory of linear programming.

Figure 1: The computational pipeline: (1) synthetic usages are generated using Janus; (2) their XL-LEXEME embeddings are average into prototypes; (3) the prototypes are used to retrieve sentences that are most similar to each prototype. Once relevant corpus data is retrieved, we apply the traditional LSC framework to them.

2. **Prototype Generation**: For each sense s_i, the encoder C transforms the synthetic examples E_i into a set of vector representations $V_i = \{C(e_{i_1}), C(e_{i_2}), \ldots, C(e_{i_m})\}$, where C : text $\rightarrow \mathbb{R}^d$ maps text to a d-dimensional vector space. The prototyper P aggregates these vectors into a single prototype vector $p_i = P(V_i)$, which encapsulates the semantic characteristics of sense s_i.
3. **Corpus Analysis**: Let $U = \{u_1, u_2, \ldots, u_K\}$ denote the set of actual occurrences of the target word in the historical corpus. Each occurrence u_j is encoded into a vector $v_j = C(u_j)$. The comparison function $F : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ measures the similarity between each corpus vector v_j and each prototype vector p_i. For each sense s_i and time period t, we **identify the most relevant corpus occurrences** by ranking $F(v_j, p_i)$.
4. **Analysis and Interpretation** This approach enables experts to examine the highest-ranked sentences for each sense and time period, facilitating the identification of when a particular sense, such as a metaphorical use of *gene*, first emerged. By leveraging encoded representations and prototype-based comparisons, our method provides a scalable and systematic alternative to manual concordance analysis, while maintaining interpretability for domain experts.

4. Use case

In this section, we outline a comprehensive pipeline for analyzing semantic shifts in three words relevant to the history of science: *theory*, *gene*, and *force*. Our approach combines exploratory analysis using traditional Lexical Semantic Change Detection (LSCD) methods (outlined in Section 2.2) with targeted, hypothesis-driven investigations based on our novel framework.[1]

For our analysis, we sourced historical data from the Clean Corpus of Historical American English (CCOHA) [17]. To generate usage embedding representations, we utilized XL-LEXEME because of its role as the state-of-the-art model for LSCD. Sense clusters were induced from these embeddings using agglomerative clustering with a distance threshold of 0.5.

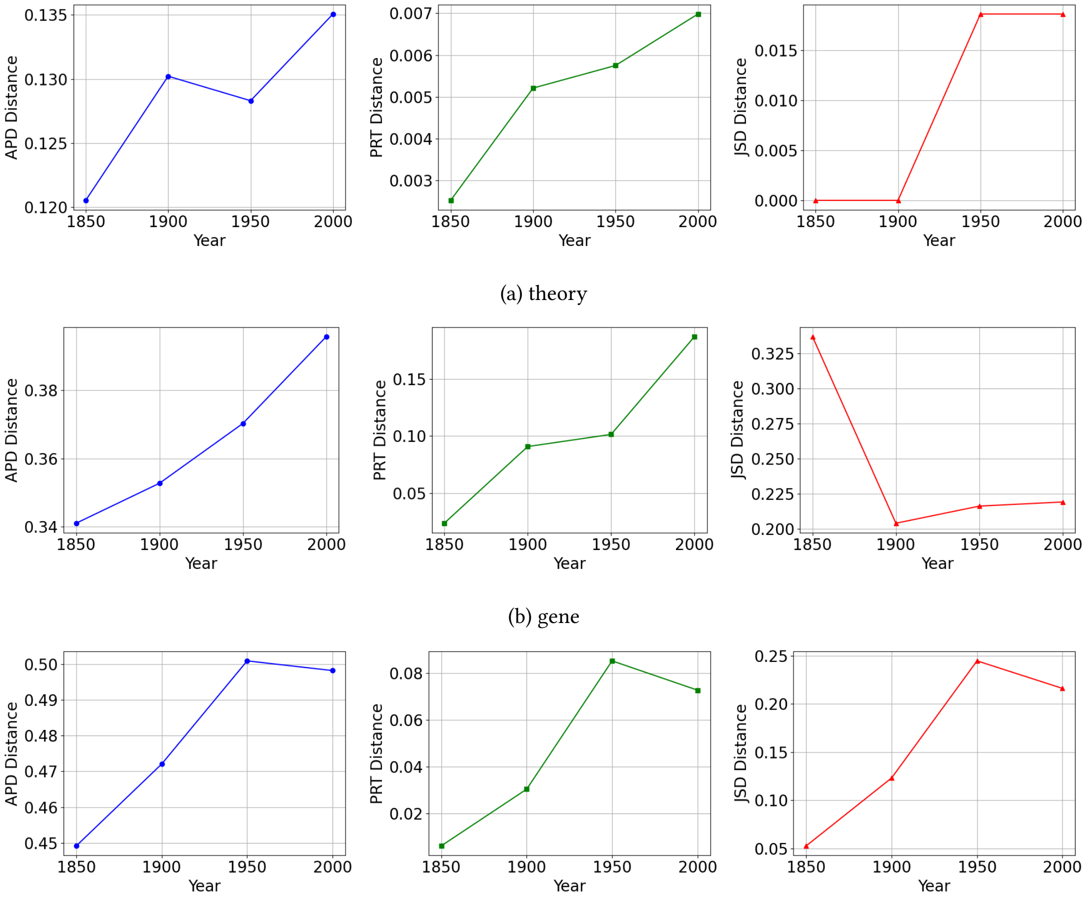
To assign semantic content to clusters and monitor semantic change, we employ LLama-Dictionary to generate context-specific definitions by selecting one representative instance for each cluster. Additionally, we use Janus to create synthetic usage examples based on predefined hypotheses and sense definitions. These examples are used to construct prototype vectors and enabling hypothesis-driven queries of the corpus.

4.1. LSCD Metrics

To evaluate lexical semantic change, we employed three distinct metrics—APD, PRT, and JSD—to quantify shifts in the meanings of the words *theory*, *gene*, and *force* over time, as depicted in Figure 2. These metrics were applied to vector representations generated by XL-LEXEME. For each word, we calculated the three metrics with respect to the first time point (e.g., $\langle V_1, V_t \rangle$).

APD The APD metric computes the average cosine distance between all pairs of vectors representing word uses from two time periods. Figure 2(a–c) illustrates that APD values for *theory* show moderate fluctuations, indicating subtle shifts in usage, while *gene* exhibits a sharp increase in APD around the 1900s, reflecting the emergence of its biological sense. Similarly, *force* displays

[1] This pipeline and these results were presented first in a keynote for the workshop Large Language Models for the History, Philosophy, and Sociology of Science.



(a) theory

(b) gene

(c) force

Figure 2: Semantic Change Over Time for *theory* (a), *theory* (b), and *theory* (c) using APD, PRT, and Jensen-Shannon Distance (JSD). Each subplot illustrates how the meaning of a word shifts across historical time points.

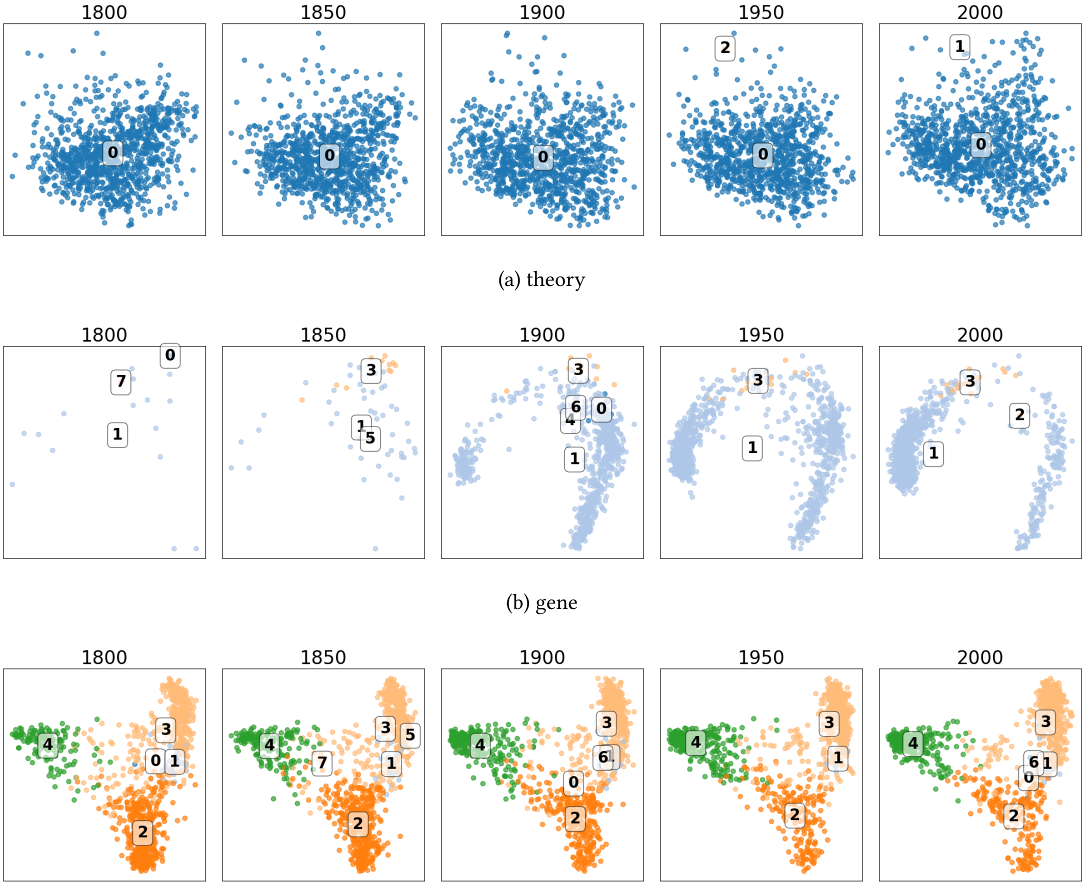
varying APD trends, with peaks corresponding to the 1950s.

PRT The PRT metric measures the cosine distance between centroid vectors (prototypes) of word uses at different time points. For each word, prototypes were generated by averaging the XL-LEXEME embeddings for all occurrences within a time period. Figure 2(a–c) shows that PRT distances for *gene* increase significantly post-1950, while for *theory* and *force*, PRT reveals more stable transitions.

JSD The JSD metric involves clustering word use embeddings (using agglomerative clustering with a distance threshold of 0.5, as shown in Figure **??**) and treating the frequency of senses as probability distributions. JSD then quantifies semantic change by computing the distance between distributions of two time periods. Figure 2(a–c) indicates that JSD captures pronounced shifts for *gene* and *force*, while for *theory* values remain relatively low. This is because only one cluster is mainly present across all time points for *theory*, with two small clusters appearing only in the final two periods.

4.2. Labeling Clusters with Definitions

We employed LLama-Dictionary to generate context-specific definitions for the words *force*, *theory*, and *gene*. For each word, sense clusters were induced in Section 4. A representative instance from each cluster was selected, and LLama-Dictionary generated a definition reflecting the word's meaning in that context. These definitions, presented in Table 1, provide a structured representation


(a) theory

(b) gene

(c) force

Figure 3: PCA Visualization of semantic clusters for the words *theory*, *gene*, and *force*, derived from CCOHA data. Each cluster represents distinct semantic interpretations or senses.

of the senses for each word.

For the word *force*, Table 1 lists seven distinct senses, ranging from physical influences (e.g., *an influence tending to change the motion of a body*) to military contexts (e.g., *a military unit engaged in a particular operation or mission*) and coercive actions (e.g., *to cause (something) to perform an action against its will or inclinations*). These definitions highlight the word's polysemy, capturing both concrete and abstract uses across historical contexts.

The word *theory* has three identified senses in Table 1: a speculative belief (*a belief that is based on speculation rather than adequate evidence*), a fashion-related sense (*a fashion theory, a style of fashion design*), and a narrative account (*a narrative account of a phenomenon, event or chain of events*). These definitions reflect the word's evolution from abstract intellectual constructs to more specific, domain-related meanings.

For *gene*, Table 1 identifies seven senses, including its modern biological meaning (e.g., *a distinct sequence of nucleotides forming part of a chromosome*) and clusters containing instances with OCR errors (e.g., *to go* or *a set of generations*).

4.3. Hypotesis-Driven Investigation

In our hypothesis-driven investigation, we conducted an in-depth semantic analysis of the lexical items *theory*, *force*, and *gene*. In particular, we selected word sense definitions from the OED that do not appear to emerge through the traditional pipeline. For theory, which appeared to have only one dominant sense in previous analyses, we identified two sub-senses: one relating to the arts and another to mathematics. For force, we chose the specific sense associated with physics, while for gene,

Word	Cluster Definition
Force	**0** A body of water or air moving under the influence of a force; **1** To cause (something) to perform an action against its will or inclination; **2** An influence tending to change the motion of a body or produce motion or stress in a stationary body; **3** To put out a runner by requiring him to run; **4** A military unit engaged in a particular operation or mission; **5** To advance or mature by natural or inevitable progression; **6** To cause (a result) by the exertion of force; **7** An army.
Theory	**0** A belief that is based on speculation rather than adequate evidence as to its truth; **1** A fashion theory, a style of fashion design; **2** A narrative account of a phenomenon, event or chain of events.
Gene	**0** To go; **1** A distinct sequence of nucleotides forming part of a chromosome, the order of which determines the order of monomers in a polypeptide or nucleic acid molecule which a cell (or virus) may synthesize; **2** A unit of heredity which is transferred from a parent to offspring and is held to determine some characteristic of the offspring; **3** A set of genetic instructions; **4** A set or class; **5** A name, especially a shortened name; **6** A set of people descended from a common ancestor; **7** A set of generations.

Table 1

Definitions of semantic clusters for the words *force*, *theory*, and *gene*. For each cluster, one representative instance was selected, and LLama-Dictionary was used to produce a context-specific definition reflecting the word's meaning in that instance.

we focused on the metaphorical sense referring to inherited traits. Table 2 illustrates representative sentences from historical periods for each targeted sense, along with corresponding similarity scores.

For *theory*, we identified clear semantic distinctions between its mathematical and arts-related conceptualizations. The mathematical sense consistently emphasizes structured systems of knowledge or deduction, notably stable across historical contexts with high similarity scores (ranging from 0.9632 in 1850 to 0.9835 in 1950). Conversely, the artistic sense of *theory* reflects broader cultural and philosophical applications, maintaining moderate similarity scores (around 0.96) but allowing variations tied to aesthetics and criticism.

The physical sense of *force* remains remarkably stable and contextually consistent, as evidenced by similarity scores consistently exceeding 0.96 across time periods.

Applying the same methodology to *gene*, specifically focusing on its metaphorical sense, clarified the earlier observed anomaly. Early instances from the 1800s were OCR errors (e.g., "genie rose," "genie really"). Genuine metaphorical usage of "gene" emerged gradually, with similarity values steadily increasing until the metaphorical sense became clearly established around the 2000s.

The hypothesis-driven investigation provides significant precision and interpretability advantages over the traditional lexical semantic change detection pipeline. By explicitly defining and targeting specific subsenses, such as distinguishing between the mathematical and artistic senses of *theory*, identifying the metaphorical usage of *gene*, and isolating the physical meaning of *force*, our method captures semantic differences that previously remained hidden within broader senses. Moreover, by directly analyzing real corpus sentences from the CCOHA dataset, experts gain improved control over the interpretation and validation of results.

5. Conclusion

In this work, we introduced a hypothesis-driven framework for detecting lexical semantic change. By integrating expert-defined sense definitions with SOTA computational models like XL-LEXEME and Janus, our framework systematically traces the evolution of word meanings across historical corpora. Starting with a word and its senses (or only the ones that we want to study), we utilize the strength of LLMs to allow for easy investigation into relevant corpus data. The method is not limited in terms of data it can be applied to, thus the user can choose the data of interest, and limit to the relevant senses. We envision that the researcher can also define senses of interest, rather than using those listed in dictionaries, for example by adding connotational information. This would allow for the investigation of when word sense e.g., became more positive in meaning.

The proposed hypothesis-driven framework offers a robust methodology for accurately detecting and analyzing lexical semantic changes in historical texts. By integrating predefined hypotheses, generative language models, and vector encoding techniques, our approach not only results interpretable for domain experts but also systematically scales to large historical corpora. The case studies on words like "theory," "gene," and "force" illustrate the framework's capability to reveal significant shifts in meaning, particularly those reflective of cultural and scientific developments.

Acknowledgments

This work has in part been funded by the research program Change is Key! supported by Riksbankens Jubileumsfond (under reference number M21-0021). The computational resources were provided by the National

Concept	Year	Most Similar Sentence (Similarity)
Theory (Mathematics) — The body of knowledge relating to the properties of a particular mathematical concept; a collection of theorems forming a connected system.		
	1800	...Fourier ’s large work , entitled , **Theory** of Universal Unity. (0.9761)
	1850	...the real object of the law is the mental image , the **theory** of the thing. (0.9632)
	1900	...a strictly consistent deduction from the **theory**... (0.9714)
	1950	...to place the **theory** of abstraction in a perspective unchallenged... (0.9835)
	2000	..., 2000) , is bio-informational **theory** (Lang , 1979 , 1985). (0.9669)
Theory (Arts) — An approach to the study of literature, the arts, and culture that incorporates concepts from disciplines such as philosophy, psychoanalysis, and the social sciences.		
	1800	...to accommodate himself to his **theory** frequently involves him in a dialect... (0.9587)
	1850	...error of his **theory** of poetry , and is the source of his one conspicuous failure... (0.9665)
	1900	...a knowledge of aesthetic history and philosophy , **theory** and practice... (0.9663)
	1950	...grammar is a **theory** of language , and a works. (0.9597)
	2000	...snake oil of art criticism and elixir of **theory**. (0.9712)
Force — Used in various senses developed from the older popular uses, and corresponding to modern scientific uses of Latin *vis*. The cause of any one of the classes of physical phenomena, e.g., of motion, heat, electricity, etc., conceived as consisting in principle or power inherent in, or coexisting with, matter.		
	1800	...the **force** d e , which it exerts upon D B. (0.9688)
	1850	...as a mechanical **force** , and as an agent in effecting chemical changes... (0.9828)
	1900	...It is the **force** of a body in motion. (0.9821)
	1950	...flowed a the **force** of gravity. (0.9823)
	2000	...the nuclear **force** is a short-range force , acting mainly over the distance... (0.9668)
Gene — In figurative and extended use, esp. with reference to qualities regarded as deeply ingrained or (often humorously) as inherited. Often in plural.		
	1800	...evinced in a more familiar way , by the **gene** ’. (0.8829)
	1850	...some people complained of a certain ’**gene**’ in him... (0.9280)
	1900	...started life with the very best of mental **genes**? (0.9335)
	1950	Apparently Johnny got all the family ’s **genes** for music... (0.9531)
	2000	...lack of the self-awareness **gene** , spearheads the awkwardness. (0.9665)

Table 2
Most similar usages by concept and year, with similarity scores.

Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through grant agreement no. 2022-06725.

References

[1] C. S. Armendariz, M. Purver, M. Ulčar, S. Pollak, N. Ljubešić, M. Granroth-Wilding, CoSimLex: A Resource for Evaluating Graded Word Similarity in Context, in: Proc. of LREC, ELRA, Marseille, France, 2020, pp. 5878–5886.

[2] O. E. D. OED, Oxford english dictionary, Simpson, Ja & Weiner, Esc 3 (1989).

[3] R. Navigli, Word Sense Disambiguation: A Survey, ACM Comput. Surv. 41 (2009). URL: https://doi.org/10.1145/1459352.1459355. doi:10.1145/1459352.1459355.

[4] N. Tahmasebi, L. Borin, A. Jatowt, Survey of Computational Approaches to Lexical Semantic Change Detection, Language Science Press, Berlin, 2021, pp. 1–91. doi:10.5281/zenodo.5040302.

[5] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, CoRR abs/1810.04805 (2018). URL: http://arxiv.org/abs/1810.04805. arXiv:1810.04805.

[6] P. Cassotti, L. Siciliani, M. DeGemmis, G. Semeraro, P. Basile, 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), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 1577–1585. URL: https://aclanthology.org/2023.acl-short.135. doi:10.18653/v1/2023.acl-short.135.

[7] N. Reimers, I. Gurevych, Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China,

2019, pp. 3982–3992. URL: https://aclanthology.org/D19-1410. doi:10.18653/v1/D19-1410.

[8] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised Cross-lingual Representation Learning at Scale, in: D. Jurafsky, J. Chai, N. Schluter, J. R. Tetreault (Eds.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020, Association for Computational Linguistics, 2020, pp. 8440–8451. URL: https://doi.org/10.18653/v1/2020.acl-main.747. doi:10.18653/v1/2020.acl-main.747.

[9] M. T. Pilehvar, J. Camacho-Collados, WiC: the Word-in-Context Dataset for Evaluating Context-Sensitive Meaning Representations, in: J. Burstein, C. Doran, T. Solorio (Eds.), Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers), Association for Computational Linguistics, 2019, pp. 1267–1273. URL: https://doi.org/10.18653/v1/n19-1128. doi:10.18653/v1/n19-1128.

[10] M. Fedorova, A. Kutuzov, Y. Scherrer, Definition generation for lexical semantic change detection, in: L.-W. Ku, A. Martins, V. Srikumar (Eds.), Findings of the Association for Computational Linguistics ACL 2024, Association for Computational Linguistics, Bangkok, Thailand and virtual meeting, 2024, pp. 5712–5724. URL: https://aclanthology.org/2024.findings-acl.339.

[11] F. Periti, D. Alfter, N. Tahmasebi, Automatically generated definitions and their utility for modeling word meaning, in: Y. Al-Onaizan, M. Bansal, Y.-N. Chen (Eds.), Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Miami, Florida, USA, 2024, pp. 14008–14026. URL: https://aclanthology.org/2024.emnlp-main.776/. doi:10.18653/v1/2024.emnlp-main.776.

[12] F. Periti, N. Tahmasebi, A systematic comparison of contextualized word embeddings for lexical semantic change, in: K. Duh, H. Gomez, S. Bethard (Eds.), Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), Association for Computational Linguistics, Mexico City, Mexico, 2024, pp. 4262–4282. URL: https://aclanthology.org/2024.naacl-long.240.

[13] F. Periti, N. Tahmasebi, Towards a complete solution to lexical semantic change: an extension to multiple time periods and diachronic word sense induction, in: N. Tahmasebi, S. Montariol, A. Kutuzov, D. Alfter, F. Periti, P. Cassotti, N. Huebscher (Eds.), Proceedings of the 5th Workshop on Computational Approaches to Historical Language Change, Association for Computational Linguistics, Bangkok, Thailand, 2024, pp. 108–119. URL: https://aclanthology.org/2024.lchange-1.10/. doi:10.18653/v1/2024.lchange-1.10.

[14] M. Giulianelli, M. Del Tredici, R. Fernández, Analysing lexical semantic change with contextualised word representations, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 3960–3973. URL: https://www.aclweb.org/anthology/2020.acl-main.365. doi:10.18653/v1/2020.acl-main.365.

[15] P. Cassotti, N. Tahmasebi, Sense-specific historical word usage generation, Transactions of the Association for Computational Linguistics (2025).

[16] G. A. Miller, WORDNET: a lexical database for english, in: Speech and Natural Language: Proceedings of a Workshop Held at Harriman, New York, USA, February 23-26, 1992, Morgan Kaufmann, 1992. URL: https://aclanthology.org/H92-1116/.

[17] R. Alatrash, D. Schlechtweg, J. Kuhn, S. S. im Walde, CCOHA: Clean Corpus of Historical American English, in: N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis (Eds.), Proceedings of The 12th Language Resources and Evaluation Conference, LREC 2020, European Language Resources Association, Marseille, France, 2020, pp. 6958–6966. URL: https://www.aclweb.org/anthology/2020.lrec-1.859/.

Declaration on Generative AI

During the preparation of this work, the author(s) did not use any generative AI tools or services.