Anchors in Embedding Space: A Simple Concept Tracking Approach to Support Conceptual History Research

Jetske Adams[1], Martha Larson[1], Jaap Verheul[2], Michael Boyden[2]
[1]Centre for Language Studies, Radboud University
[2]Radboud Institute for Culture and History, Radboud University

Abstract

We introduce a simple concept tracking approach to support conceptual history research. Building on the existing practices of conceptual historians, we use dictionaries to identify "anchors", which represent primary dimensions of meaning of a concept. Then, we create a plot showing how a key concept has evolved over time in a historical corpus in relation to these dimensions. We demonstrate the approach by plotting the change of several key concepts in the COHA corpus.

1 Introduction

Conceptual history is the study of the abstract, sociopolitical concepts that are used to describe and understand history. The purpose of our work is to complement the computational methods that are available for research in conceptual history by introducing an approach specifically designed to be easily used by conceptual historians.

Conceptual historians are interested in the evolution over time of "key concepts" that have social or political relevance. Our approach follows the work of Reinhart Koselleck, a pioneer of conceptual history. While key concepts are necessarily expressed as words, not all words are concepts in the sense of Koselleck. In his introduction to *Geschichtliche Grundbegriffe*, a collaborative multivolume lexicon of key concepts during the period 1750-1850, Koselleck states, "a word becomes a concept when a single word is needed that contains—and is indispensable for articulating—the full range of meanings derived from a given sociopolitical context" (Koselleck and Richter, 2011, p. 19). Analysis of sociopolitical concepts aims to bring out the coexistence of meaning layers ("temporal strata") in a given concept (Koselleck, 2004).

One approach used by Koselleck is to analyze how definitions in dictionaries change over the years, e.g., in Koselleck and Richter (2006). A definition of a word in a dictionary explicitly expresses primary dimensions of meaning. Here, we do not analyze historical dictionaries, but rather investigate changes in the relative importance of primary dimensions of meaning derived from a contemporary dictionary. As such, our approach represents a way to base concept tracking on known, and explicitly expressed, dimensions of meaning, without relying on the existence and availability of multiple historical dictionaries. Note that dictionaries are not the only source that conceptual historians use to identify dimensions of meaning that are interesting for investigation. However, due to their importance and easy accessibility, we focus on them here.

Our approach uses two primary dimensions of meaning represented by words we refer to as "anchors". We use the contextual word embeddings of the two anchors to create a plot that visualizes how the meaning of a key concept has changed over time within a diachronic corpus of historical texts. We argue that this simple concept tracking approach is useful for conceptual history research. First, it yields a concise plot that is easily interpretable for historians. Second, contextual embeddings do not necessarily have to be trained on the specific data being analyzed, which in the case of conceptual history research might be quite limited.

2 Background and related work

2.1 Concepts in conceptual history

Key concepts, in the sense of Koselleck, display specific properties that can be analyzed on a formal level: they are abstract, freestanding terms that often function as political catchwords that can be mobilized by various ideologies and factions (Koselleck, 2002). Often, they also display a metahistorical quality (e.g., the concept "progress" captures a linear understanding of time). Because of their function in steering public debate, concepts of this sort are inherently ambiguous. Thus, although a word such as "bank" is polysemous, it does not

Proceedings of the 4th Workshop on Computational Approaches to Historical Language Change, pages 87–92
December 6, 2023

count as a key concept because its meanings can be distinguished on the basis of its immediate context of use and because it does not carry forward political discourse in the way that, for instance, "justice" or "democracy" do. The meanings of a polysemous word are often assumed by linguists to be "well behaved", whereas conceptual historians are interested in the "ungovernability" of meaning.

Although conceptual history should be understood as distinct from semantic history, the approach as it emerged has remained, as De Bolla et al. (2019, p. 70) have noted, "at base a semantically motivated field of inquiry". The computational study of language, however, allows us to disentangle conceptual and semantic history in ways not possible before. These methodological advances should make it possible to uncover deeper conceptual structures of the sort theorized by Koselleck (De Bolla et al., 2019).

2.2 Computational concept tracking

In contrast to our approach that uses contextual word embeddings, many of the computational approaches to tracking concept change split a text corpus into (possibly overlapping) time windows and train (or fine-tune) a static word embedding model on the data in each window. Then, given a target concept, for each time window, they determine the neighbors of the embedding of the target concept, i.e., they calculate the embeddings of the words that are semantically closest to the target concept in the time window. Changes in the identity of the closest neighbors and in their degree of closeness are then analyzed over time. This information is summarized with a neighborhood change measure (Hamilton et al., 2016a) or a meaning stability score (Azarbonyad et al., 2017), or it is visualized as a plot tracing the distance of the target concept to individual neighbors (Viola and Verheul, 2020), as a series of graphs centered on the target concept, one for each time window (Martinez-Ortiz et al., 2016; Verheul et al., 2022), as t-SNE embeddings (Hamilton et al., 2016b), or as a complex graph (Haase et al., 2021).

A common position, to our knowledge first expressed by Kenter et al. (2015), is that algorithms that track semantic change over time should be "ad hoc" in the sense that they should generate words that are similar to the concept being tracked on the fly from the data, and no input should be required by the user. In this paper, we argue that in the case of conceptual history it is useful to take the opposite starting point. Specifically, concepts should be tracked with respect to known dimensions of meaning that are derived from explicit knowledge. Here, we focus on knowledge captured by lexicographers in the form of dictionary.

The closest related work to our own is, to our knowledge, work by Martinc et al. (2020) on diachronic semantic shift, which also uses contextual word embeddings. However, this work uses the ad hoc approach, deriving semantic neighbors from the data rather than using a dictionary or other knowledge sources. Further, Martinc et al. (2020) plot individual word similarities, whereas our plots visualize the relative movement between two different primary meanings.

3 Our anchor-based approach

For each key concept to be tracked, we retrieve its dictionary definition and look at the different meanings (sub-definitions) there. We select the two major meanings and, for each, choose an "anchor", a term that captures that meaning (i.e., a keyword from the sub-definition). In the online Miriam-Webster dictionary that we used, such words are often bolded. If more than two major meanings are present, we choose the meanings that are most related to the research interests of the conceptual historian. Once we have two anchors, we plot the difference over time between the similarity of the key concept to one anchor and its similarity to the other. The data and the details of our implementation are described in this section.

3.1 Data

Our study uses the Corpus of Historical American English (COHA) (Davies, 2010), specifically, the data between 1900 and 2000. Pre-processing steps included removal of irrelevant characters such as the article number at the beginning of texts, removal of punctuation marks except for apostrophes, removal of numerical characters, splitting of the texts into sentences, and conversion to lowercase.

3.2 Implementation

The English BERT 'bert-base-uncased' (Devlin et al., 2019) was used as the model to acquire contextual embeddings. It was pre-trained on BookCorpus and English Wikipedia (a total of 3.3×10^9 running words). Since COHA is not domain-specific, we did not fine-tune the model for this study.

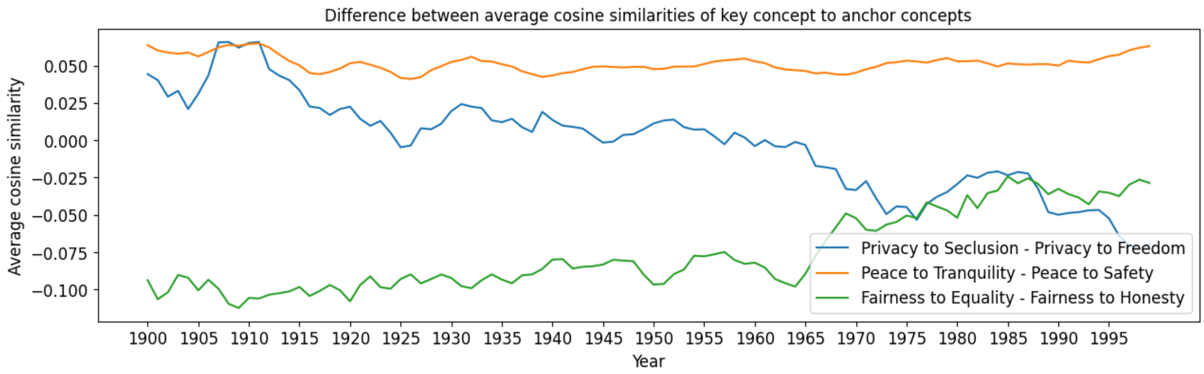

Figure 1: "Anchors" plot, which shows the difference in average cosine similarity per year between the key concept and the anchor concepts. The graph was smoothed by averaging the results over a period of 5 years.

Key Concept	Anchor Concepts	Mentions/year	Median std
Privacy	Seclusion - Freedom	27	0.078
Peace	Tranquility - Safety	455	0.041
Fairness	Equality - Honesty	164	0.066

Table 1: Information about the data for all three key concepts. Mentions per year gives the mean annual number of mentions of the key concept in the corpus. The median standard deviation is obtained from the difference values (similarity to anchor concept - similarity to other anchor concept) of all years.

To represent key concepts, we used contextual embeddings of the key concept's occurrences. The two sentences before and after the sentence in which the target occurred were added as context. This text window was chosen because it gives enough leeway even if the key concept word occurs in the first sentence of the text or if a surrounding sentence is very short. In case the context was longer than 512 tokens (infrequent) the remainder was left out. The final embeddings of the key concepts were obtained by extracting the hidden states from the last (12^{th}) layer of the model. These hidden states yield embeddings of 768 dimensions.

For the two anchors, the steps described above were also followed to obtain embeddings. Then, the embeddings of all occurrences of the anchor word between 1900 and 2000 were averaged to obtain the two final anchor embeddings. Averaging of contextual embeddings is also used in Martinc et al. (2020). The cosine similarity between the embedding of each occurrence of a key concept in the text and both anchor embeddings was computed.

Next, for each anchor, the cosine similarities were averaged per year. We calculated the difference between the average similarity of the key concept to one anchor and the average similarity of the key concept to the other anchor and plotted this difference over time. We also calculated the standard deviation of the differences for each year.

3.3 Statistical testing

We used the Mann-Kendall test to identify monotonic trends in time series, either upwards or downwards. This test is frequently used for hydrometeorological time series (Wang et al., 2020). The trend is deemed statistically significant when the p-value is lower than 0.05. For this test to be effective it is not necessary that the trend is linear or that the data is normally distributed. Because the Mann-Kendall test can only deal with one score for a time period, we use the average difference in cosine similarity per year. We also apply a second statistical test, Spearman's rank correlation, since it has been used in the literature (Hamilton et al., 2016b).This test has the same significance threshold and gives a correlation coefficient that reflects the direction of the trend.

4 Tracking key concepts: Three examples

We illustrate our approach with the key concepts "privacy" (anchors: "seclusion"/"freedom"), "peace" (anchors: "tranquility"/"safety") and "fairness" (anchors: "equality"/"honesty"). Results are shown in table 2 and figure 1. The trends for both "privacy" and "fairness" are statistically significant, but "peace" has no overall trend. Spearman's correlation, used by Hamilton et al. (2016b), yielded the same significance result.

Key Concept	Anchor Concepts	Slope$_{MK}$	p-value$_{MK}$	Correlation$_{SP}$	p-value$_{SP}$
Privacy	Seclusion - Freedom	-9.97$\times 10^{-4}$	<.001	-0.767	<.001
Peace	Tranquility - Safety	-0.05$\times 10^{-4}$	.856	-0.051	.614
Fairness	Equality - Honesty	8.21$\times 10^{-4}$	<.001	0.707	<.001

Table 2: Results for all three key concepts. For the Mann-Kendall test, the slope and p-value are given. For the Spearman's rank correlation test, the correlation coefficient and p-value are given.

Sentence	Year	S_C Seclusion	S_C Freedom
"it wiould be better to have it out with the railway representative in the **privacy** of the council room"	1917	0.60	0.46
"our **privacy** is under attack not just from government but also from corporations and even ourselves"	1998	0.47	0.64

Table 3: Example sentences from COHA for key concept "privacy" given with the cosine similarity (S_C) of the "privacy" embedding to each anchor embedding. Both examples are from newspaper text.

"Fairness" consistently leans towards "honesty" rather than to "equality" in terms of similarity, although the difference becomes smaller over time. ”Peace” remains closer to "tranquility" than to "safety". During the 1900s, "privacy" was more similar to "seclusion" than to "freedom", but this reversed around the 1960s. The two sentences in table 3 illustrate the difference.

Table 4 highlights two important points concerning our approach using the example "peace". First (top two rows), our plot does not reflect the case in which both anchor concepts' cosine similarity to the key concept move in the same direction. We advise historians not to abandon plots of the cosine between key concepts and individual terms, but to use them alongside our difference plots. Second (bottom two rows), before World War II, a slight but significant trend was found of "peace" towards "safety", followed by a small reversal. The Mann-Kendall test is not suited for detecting such changes without choosing a point to split up the data.

5 Connecting to conceptual history

In this section, we present an example illustrating how our anchor-based approach might connect to existing conceptual history research. Specifically, we look at work by Boyden et al. (2022) on how "climate" has emerged as a key concept. Before the rise of climate science, "climate" was undifferentiated from geography and weather associated with places. In early modern geography, "climate" was roughly identical to geodetic position. However, over the years "climate" has become *globalized*, i.e., associated with future weather conditions of the entire planet. We can explain the shift from "local" to "global" in terms of the difference between meteorology and climate science. The latter deals with weather patterns averaged over long periods of time and on a planetary scale.

Figure 2 shows a graph of the key concept "climate". We see "climate" moving further from "local" and closer to "global" over time. The Mann-Kendall test gave an increasing trend (slope = 5.42×10^{-4}, p < 0.001) so the shift was significant. In this time span, "climate" appeared an average of 48 times per year in the corpus, with a median standard deviation of the subtracted average cosine similarities of all years of 0.045. Particularly in the last quarter of the 20^{th} Century, our analysis shows "climate" became increasingly associated with "global". However, before that time it was already evolving away from "local" and towards "global". These observations are consistent with the ideas and insights of Boyden et al. (2022). We note that Boyden et al. (2022) point to the Oxford English Dictionary as a source of support for older "local" meanings of "climate", but the choice of the anchors here is also based on other considerations, such the rise of climate science, mentioned above. Finally, we emphasize that our anchor-based approach is not intended to replace concept graphs, such as those also used by Boyden et al. (2022), but rather complements them.

6 Conclusion and outlook

In this study we have proposed a "Definitions as Anchors" approach to tracking the evolution of "key concepts", i.e., abstract sociopolitical concepts, which makes use of "anchors" drawn from dictionary definitions. Our approach maps key con-

Key Concept	Anchor Concept(s)	Years	Slope$_{MK}$	p-value$_{MK}$
Peace	Tranquility	1900-2000	-2.32$\times 10^{-4}$	<0.001
Peace	Safety	1900-2000	-2.10$\times 10^{-4}$	<0.001
Peace	Tranquility - Safety	1900-1945	-3.66$\times 10^{-4}$	<0.001
Peace	Tranquility - Safety	1945-2000	1.12$\times 10^{-4}$	0.032

Table 4: Further points about our approach demonstrated by key concept "peace"

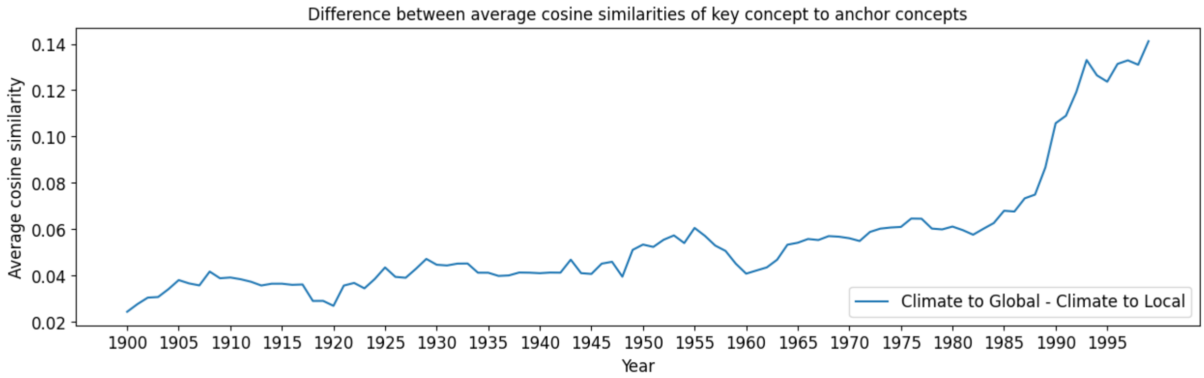

Figure 2: "Anchors" plot for the key concept "climate", with anchors "global" and "local". As above, the graph was smoothed by averaging the results over a period of 5 years.

cepts to a relative position in semantic space, much like approaches that build semantic graphs for individual time windows, e.g., Martinez-Ortiz et al. (2016). Instead of being positioned with respect to a larger number of ad hoc neighbors, key concepts are traced with respect to two pre-defined anchors, dramatically simplifying the interpretation and allowing straightforward calculation of the statistical significance of trends.

We have argued for pre-defined anchors because it builds on conceptual historians' established practices. However, we also note that using pre-defined anchors may help to address our concern that the neighbors of a key concept within a time window are determined more by the dominant topics in that time window, rather than by an actual shift in the semantics of the key concept. The importance of this concern should be investigated in future work.

Future work should also investigate the advantage that contextual word embeddings offer in leveraging more training data that non-contextual embeddings. Our word embedding model was pre-trained on the order of 10^9 running words. In contrast, if the COHA collection is split into year-length windows and a static word embedding model is built on each window, i.e., the approach of Martinez-Ortiz et al. (2016), each model is trained on only on the order of 10^6 words, three orders of magnitude fewer words.

In sum, our study enriches conceptual history with an approach that can statistically confirm monotonic changes of abstract sociopolitical concepts over time in a diachronic text corpus. It contributes to the practical understanding of how and over what time periods conceptual shifts occur.

7 Limitations

We present a simple concept tracking approach, which we have designed to be easy for conceptual historians to interpret and also relatively robust to variation (for example changes of topic) that is not relevant to underlying conceptual change. We have not, however, demonstrated experimentally that our approach has either of these properties. We have not compared non-contextual embeddings to show the advantage of contextual embeddings.

Further, as noted in section 3.2, the 'bert-base-uncased' model was not further trained or fine-tuned. Although COHA is broad in topic and genre (i.e., not domain specific) and fine-tuning may be inconvenient for historians, we do find that future work should test a model pre-trained on COHA, such as histBERT (Qiu and Xu, 2022).

Also, the stability of the anchor concepts requires additional evaluation. Finally, our statistical tests analyze monotonic trends. Future work should consider trends that change direction and also change point detection.

References

Hosein Azarbonyad, Mostafa Dehghani, Kaspar Beelen, Alexandra Arkut, Maarten Marx, and Jaap Kamps. 2017. Words are malleable: Computing semantic shifts in political and media discourse. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, CIKM '17, page 1509–1518, New York, NY, USA. Association for Computing Machinery.

Michael Boyden, Ali Basirat, and Karl Berglund. 2022. Digital conceptual history and the emergence of a globalized climate imaginary. *Contributions to the History of Concepts*, 17(2):95 – 122.

Mark Davies. 2010. The corpus of historical American English: 400 million words, 1810-2009. `http://corpus.byu.edu/coha/`. Accessed: 2023-08-31.

Peter De Bolla, Ewan Jones, Paul Nulty, Gabriel Recchia, and John Regan. 2019. Distributional concept analysis: A computational model for history of concepts. *Contributions to the History of Concepts*, 14(1):66 – 92.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *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.

Christian Haase, Saba Anwar, Seid Muhie Yimam, Alexander Friedrich, and Chris Biemann. 2021. SCoT: Sense clustering over time: a tool for the analysis of lexical change. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 198–204, Online. Association for Computational Linguistics.

William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016a. 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, Austin, Texas. Association for Computational Linguistics.

William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016b. Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501, Berlin, Germany. Association for Computational Linguistics.

Tom Kenter, Melvin Wevers, Pim Huijnen, and Maarten de Rijke. 2015. Ad hoc monitoring of vocabulary shifts over time. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, CIKM '15, page 1191–1200, New York, NY, USA. Association for Computing Machinery.

Reinhart Koselleck. 2002. *"Progress" and "Decline": An Appendix to the History of Two Concepts*, pages 218–235. Stanford University Press, Redwood City.

Reinhart Koselleck. 2004. *History, Histories, and Formal Time Structures*, pages 93–114. Columbia UP, New York.

Reinhart Koselleck and Michaela Richter. 2011. Introduction and prefaces to the Geschichtliche Grundbegriffe: (basic concepts in history: A historical dictionary of political and social language in Germany). *Contributions to the History of Concepts*, 6(1):1–37.

Reinhart Koselleck and Michaela W. Richter. 2006. Crisis. *Journal of the History of Ideas*, 67(2):357–400.

Matej Martinc, Petra Kralj Novak, and Senja Pollak. 2020. 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.

Carlos Martinez-Ortiz, Tom Kenter, Melvin Wevers, Pim Huijnen, Jaap Verheul, and Joris van Eijnatten. 2016. Design and implementation of ShiCo: Visualising shifting concepts over time. In *Proceedings of the 3rd International Workshop on Computational History (HistoInformatics2016) (CEUR Workshop Proceedings 1632)*.

Wenjun Qiu and Yang Xu. 2022. Histbert: A pre-trained language model for diachronic lexical semantic analysis.

Jaap Verheul, Hannu Salmi, Martin Riedl, Asko Nivala, Lorella Viola, Jana Keck, and Emily Bell. 2022. Using word vector models to trace conceptual change over time and space in historical newspapers, 1840-1914. *Digital Humanities Quarterly*, 16(2).

Lorella Viola and Jaap Verheul. 2020. One hundred years of migration discourse in The Times: A discourse-historical word vector space approach to the construction of meaning. *Frontiers in Artificial Intelligence*, 3:64.

Fan Wang, Wei Shao, Haijun Yu, Guangyuan Kan, Xiaoyan He, Dawei Zhang, Minglei Ren, and Gang Wang. 2020. Re-evaluation of the power of the Mann-Kendall test for detecting monotonic trends in hydrometeorological time series. *Frontiers in Earth Science*, 8.