



Institut für
deutsche Philologie

04-DH-I1

Hausarbeit Praxis digitaler Objekte 1

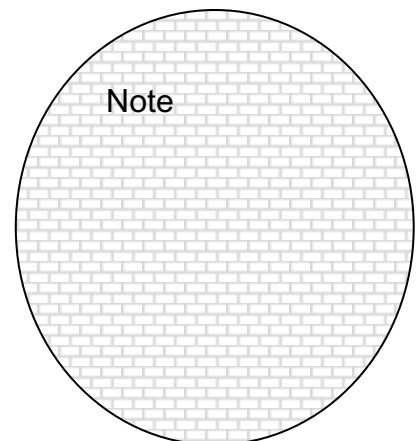
(04-DH-I-1)

Dozent:	Prof. Dr. Fotis Jannidis
Semester:	25SS
Nachname:	Lu
Vorname:	Huanyu
Matrikelnummer:	2649797
Geburtsdatum:	02.10.1999

_____ vom Korrektor auszufüllen _____

Datum	Unterschrift des Korrektors

Note



Gliederung

Einleitung	1
1. Daten	2
2. Aufbau und Methoden des Verfahrens	3
3. Qualitative Analyse	5
4. Diskussion	9
5. Evaluation	10
Fazit	12
Literatur und Quelle	14
Abbildung	16

Einleitung

Von den frühesten menschlichen Gesellschaften bis heute hat die Sprache tiefgreifende Veränderungen durchlaufen. Sie hat sich ständig weiterentwickelt, um an neue Zwecke und Bedürfnisse anzupassen¹. Die Faktoren, die diese Veränderungen verursacht haben, sind auch vielfältig. Soziale, historische, kulturelle und technologische Entwicklungen beeinflussen die menschliche Kommunikation und beschleunigen die sprachlichen Veränderungen². Die industrielle Revolution des 18. und 19. Jahrhunderts gilt als einer der wichtigsten Wendepunkte der Geschichte. Sie revolutionierte nicht nur die Produktionsmethoden, sondern führte auch zu neuen sozialen Strukturen und Wirtschaftsmodellen, z.B. die Gründung und Entwicklung des Unternehmenssystems. Diese gesellschaftlichen Veränderungen können auch Spuren in der Sprache hinterlassen, wie Bedeutungswandel. Die Analyse solcher Veränderungen in der Bedeutung trägt dazu bei, die sprachliche Entwicklung und ihren Zusammenhang mit gesellschaftlichen Faktoren zu beleuchten. In der linguistischen Forschung wird semantischer Bedeutungswandel typischerweise in sechs Arten unterteilt, nämlich broadening, narrowing, amelioration, pejoration, metaphor, und metonymy. Frühere Forschungsanalysen basierten in erster Linie auf manuellen Methoden. Mit dem Entstehen der Computerlinguistik haben sich maschinellen Lernens und der Verarbeitung natürlicher Sprache zu effektiven Werkzeugen für die Analyse großer Mengen von Textdaten entwickelt³. Vor diesem Forschungshintergrund untersucht diese Arbeit anhand des BERT-Wort-Embedding-Modells historische amerikanische Zeitungen aus dem späten 18. bis späten 19. Jahrhundert. Auf Grundlage der Ergebnisse werden semantische Bedeutungswandlungen während dieses historischen Zeitraums untersucht. Konkret gesagt, hat sich die ausgewählten kommerziellen Wörter im Laufe des

¹ Jader Martins Camboim de Sá, Marcos Da Silveira, Cédric Pruski: Survey in Characterization of Semantic Change. 2024. Online: <https://arxiv.org/html/2402.19088v2> (Stand:30.09.2025).

² Asri, Wahyu Kurniati; Rhamadanty, Winda Ayu Utami; Burhamzah, Muftihaturrahmah; Alamsyah: Analyzing Semantic Shifts in English and German by Exploring Historical Influences and Societal Dynamics. 2024. Online: https://www.researchgate.net/publication/381722204_Analyzing_semantic_shifts_in_English_and_German_by_exploring_historical_influences_and_societal_dynamics_S2 (Stand:30.09.2025).

³ Siehe 1

19. Jahrhunderts semantisch verändert? Wenn ja. Hängen sie mit Wandel im sozio-historischen Kontext zusammen? Der Rest dieses Artikels ist wie folgt aufgebaut. Kapitel 1 beschreibt die Datenquellen und die Erstellung des Korpus. Kapitel 2 beschreibt den konkreten Prozess der Modellumsetzung. In Kapitel 3 werden die quantitativen Ergebnisse präsentiert und analysiert. Kapitel 4 diskutiert diese Ergebnisse weiter im historischen Kontext. Kapitel 5 evaluiert die Validität der angewandten Methode. Das letzte Kapitel fasst die Forschungsergebnisse zusammen und erklärt die Einschränkung des gesamten Forschungsprozesses.

1. Daten

Diese Arbeit verwendet Zeitungstexte als Datenquelle des Korpus. Die Aktualität dieses Textgenres ist ein wichtiger Vorteil für die Analyse des semantischen Wandels. Der darin enthaltene Inhalt spiegelt die laufenden gesellschaftlichen Veränderungen wider. Daher sind Zeitungstexte im Vergleich zu anderen Textgattungen sensibler, die sofortige sprachliche Wirkung des sozialen Wandels zu dokumentieren. Unter Berücksichtigung der Auswirkungen des amerikanischen Bürgerkriegs und der industriellen Revolution auf die amerikanische Wirtschaftsgeschichte vergleicht diese Arbeit semantische Wandel in diesen Zeiträumen 1750–1800 und 1850–1900.

Die für die Korpusanalyse verwendeten Daten stammen aus der Datenbank „Chronicling America: Historic American Newspapers“ der Library of Congress. In Bezug auf die Verfügbarkeit von Daten erklärt die Library of Congress deutlich, dass die Zeitungsmaterialien in dieser Datenbank entweder öffentlich zugänglich sind oder keinen bekannten urheberrechtlichen Beschränkungen unterliegen⁴. Zeitungen, die vor 95 Jahren in den USA veröffentlicht wurden, sind komplett öffentlich zugänglich. Daher ist die Verwendung für Forschungszwecke völlig legal. Bei der Datensammlung führte diese Arbeit eine Kombination aus Website-Recherchen und offiziellen OCR-Daten durch, um Textdateien für die

⁴ Library of Congress: Rights and Access. Online: <https://www.loc.gov/collections/chronicling-america/about-this-collection/rights-and-access/> (Stand:30.09.2025).

erforderlichen Zeiträume zu erhalten. Angesichts der begrenzten Menge an Texten aus den Jahren 1750–1800 wurden so viele Daten wie möglich aus dieser Zeitraum gesammelt. Für den datenreicheren Zeitraum von 1850-1900 wurden Zeitungstexte aus drei zufällig ausgewählten Ordnern entnommen. Dieser Weg reduzierte den Rechenaufwand und stellte gleichzeitig die Repräsentativität der Texte sicher. Während des Korpuserstellungsprozesses wurden die Originaltexte auch einer Vorbearbeitung zugeführt. Dazu gehörte das Entfernen von durch die OCR-Erkennung verursachten Wort- und Zeilenumbrüchen sowie die Umwandlung des gesamten Textes in Kleinbuchstaben, um die Auswirkungen der Groß- und Kleinbuchstaben auf die Modellanalyse zu vermeiden.

Bei der Auswahl der Zielwörter hat diese Arbeit einen gezielten Weg verfolgt. Genauer gesagt wurden die Begriffe „bank“, „market“ und „company“ als Zielwörter für die Forschung des semantischen Bedeutungswandels ausgewählt. Einerseits werden diese Begriffe häufig im wirtschaftlichen Bereich verwendet. Andererseits sind es Wörter, die einen semantischen Wandel durchlaufen haben. Beispielsweise hat sich der Begriff „Bank“ allmählich von seiner ursprünglichen geografischen Bedeutung „Flussufer oder Damm“ auf die Bedeutung „Finanzinstitut“ erweitert. Zu Beginn der Forschungsphase wurde der Begriff „manufacture“ auch als Zielwort aufgenommen. Seine Bedeutung hatte sich von der ursprünglichen Bedeutung „handwerkliche Herstellung“ zu „Massenproduktion“ verändert⁵. Allerdings ist seine Anzahl im Textkorpus zu gering, um ausreichend repräsentativ zu sein. Deshalb wurde es aus der endgültigen Forschung ausgeschlossen. Um die Korrektheit der Erkennung semantischen Bedeutungswandels sicherzustellen, wurde in dieser Arbeit „Winter“ als Vergleichsgruppe festgelegt. Durch den Vergleich der quantitativen Ergebnisse der Zielwörter, mit denen der stabilen Wörter lassen sich echten semantischen Bedeutungswandels genauer nachweisen.

2. Aufbau und Methoden des Verfahrens

Der gesamte Prozess umfasst das Auffinden von Sätzen, die Generierung von

⁵ Siehe 2.

Einbettungen, den Vergleich semantischen Bedeutungswandels und die Evaluierung der Wirksamkeit der Methode.

Zu Beginn werden die Textdateien innerhalb des Korpus durchlaufen und diejenigen herausgefiltert, die die Zielwörter enthalten. Für jedes Vorkommen des Zielwortes wird der Kontext anhand eines Fensters von 50 Wörtern vor und nach dem Wort extrahiert. Dieses Fenster bietet semantische Informationen über das Zielwort in einem spezifischen Kontext. Basierend auf diesen extrahierten Daten wurden Tokenisierung und Embedding mithilfe des vortrainierten BertTokenizer und BertModel aus der Hugging Face Transformers durchgeführt, wodurch letztendlich Vektordarstellungen der Wortbedeutungen generiert wurden. Jeder in diesem Prozess generierte Vektor kann als eine semantische Darstellung des Zielworts innerhalb bestimmter Kontexte betrachtet werden. Dies bildet die Grundlage für die nachfolgende Analyse des semantischen Bedeutungswandels. Basierend auf diesen Vektoren wurden quantitative Analysen anhand verschiedener Messmethoden durchgeführt. Im Prozess der semantischen Wandelbewertung wurden die Embedding-Vektoren derselben Zeitperiode zusätzlich einer Mittelwertbildung und einer auf k-means-Clustering (k=5) basierenden Aggregation unterzogen. In dieser Arbeit wurden zunächst die formbasierten Methoden Inverted similarity over word prototype (PRT) und Average pairwise distance (APD) ausgewählt, um Veränderungen in der Gesamtheit zu messen. Diese Ansätze untersuchen semantische Wandel durch die Polysemie von Zielwörtern oder Veränderungen des Vektorraums die dominante Bedeutung⁶. Dann wurde die sinnbasierte Methode Cosine distance between cluster distributions (CDCD) verwendet, um Veränderungen in der semantischen Vielfalt der Zielwörter zu evaluieren, beispielsweise ob sich die vorherrschende semantische Bedeutung verschoben hat.

Darüber hinaus wurden in dieser Arbeit mehrere Bewertungsmethoden, wie z. B. Validierung anhand von Testdatenbanken, vergleichende Analyse bekannter

⁶ Periti, Francesco; Montanell, Stefano: Lexical Semantic Change through Large Language Models: a Survey. ACM Comput. Surv. 56, 11, Article 282 (June 2024), S.7. Online: <https://dl.acm.org/doi/pdf/10.1145/3672393> (Stand: 30.09.2025).

Informationen und Shuffling-Tests, angewandt, um die Gültigkeit der vorgeschlagenen Methode zu überprüfen.

3. Qualitative Analyse

Zunächst wurde eine Analyse der Häufigkeit des Vorkommens des Zielworts innerhalb des Korpus durchgeführt, einschließlich der Häufigkeit von Dokumenten, die das Zielwort innerhalb des gesamten Korpus enthalten, und der Anzahl der Vorkommen des Zielworts innerhalb jedes Dokuments, das es enthält. Basierend auf den Ergebnissen der Textfilterung und Embedding enthielten innerhalb des C1-Korpus aus den Jahren 1750–1800 (bestehend aus 8.532 Dokumenten) 2.450 Dokumente mit dem Zielbegriff „company“. Im C2-Korpus aus den Jahren 1850–1900, der 12.180 Dokumente umfasst, enthielten 10.250 Dokumente das Zielwort. Dies bedeutet, dass der Anteil der Texte mit diesem Zielwort in diesen beiden Zeiträumen von fast 29 % auf 84 % gestiegen ist, während die durchschnittliche Häufigkeit pro Dokument von 1,72 auf 5,24 gestiegen ist (Abb.1). Das Zielwort „Markt“ hat eine ähnliche Situation gezeigt. Die Anzahl der gefundenen Dateien in C1 war 2.057, während es in C2 10.324 waren. Die durchschnittliche Anzahl der Vorkommen pro Datei ist auch von 1,46 auf 5,23 gestiegen (Abb.2). Die Textabdeckung und die durchschnittliche Häufigkeit des Vorkommens des Zielworts „Bank“ zeigten ebenfalls einen Anstieg (Abb.3). Um zu überprüfen, dass dieser steigende Trend nicht auf Eigenschaften des Korpus selbst zurückzuführen ist, wurde in dieser Arbeit „Winter“ als Kontrollgruppe eingeführt. Die Analyse zeigt, dass sowohl die Textabdeckung als auch die durchschnittliche Häufigkeit des Vorkommens von „Winter“ zugenommen haben. Die Dokumentabdeckung stieg von 13,4 % auf 63,8 %, während die durchschnittliche Häufigkeit von 1,37 auf 2,49 anstieg (Abb. 4). Insgesamt haben sowohl der Anteil der Texte mit Zielwörter als auch die Vorkommenshäufigkeit der Zielwörter erhöht. Dies spiegelt den möglichen Einfluss von Unterschieden in der Textanzahl und Textlänge im Zeitungskorpus wider⁷. Selbst wenn die tatsächliche Häufigkeit der Wortverwendung unverändert

⁷ Tahmasebi, Nina; Dubossarsky, Haim: Computational modeling of semantic change. S.13.

bleibt, steigt die absolute Anzahl der Vorkommen aufgrund der längeren Text entsprechend an. Allerdings stieg die durchschnittliche Häufigkeit des Vorkommens sowohl für „Company“ als auch für „Markt“ um über 200 Prozent. Ein solch deutlicher Anstieg kann nicht mehr allein auf die zunehmende Textlänge innerhalb des Korpus zurückgeführt werden. Er könnte in gewissem Maße auf einen starken Anstieg der Verwendungshäufigkeit dieser beiden Zielwörter hindeuten. Beispielsweise sind „Company“ und „Markt“ zu häufigen Themen im gesellschaftlichen Diskurs geworden, und es ist auch möglich, dass sie zusätzliche neue semantische Bedeutungen erhalten haben.

Während der quantitativen Analysephase gibt es mehrere Methoden zur Messung semantischer Wandel an. Diese Arbeit verwendet zunächst eine formbasierte Approaches. Dieser Weg ermöglicht die Messung des semantischen Wandels durch die Berücksichtigung von hochrangigen Attributen des Zielworts, wie z. B. Polysemie und dominante Bedeutung⁸.

Inverted Similarity over Word Prototype (PRT) ist eine Methode, die auf dem Kosinusabstand zwischen den Prototypen der Vektoren basiert. Bei diesem Verfahren werden alle Tokens des Zielworts zu einer einzigen semantischen Einheit, dem Prototyp, aggregiert. Dieser Prototypvektor kann als approximation of the dominant sense innerhalb eines bestimmten Zeitraums bezeichnet werden. Durch die Messung des Abstands zwischen Prototypen lässt sich erkennen, ob sich die dominante Semantik eines Zielworts verändert hat⁹. Je größer der PRT-Wert, desto größer ist der Abstand zwischen der dominanten Semantik der beiden Zeiträume innerhalb desselben Vektorraums. Diese Distanz ist auch ein deutliches Zeichen für semantische Bedeutungswandel. Die Ergebnisse zeigen, dass Winter (PRT = 1,024) den niedrigsten Wert besitzt, was auf eine relative Stabilität zwischen den beiden Zeiträumen hindeutet. Im Gegensatz dazu weisen alle anderen Zielwörter, die zum kommerziellen Wortschatz gehören, höhere

Online: <https://www.changeiskey.org/publication/2023-routledge-handbook/2304.06337.pdf> (Stand:30.09.2025).

⁸ Periti, Francesco; Montanell, Stefano: Lexical Semantic Change through Large Language Models: a Survey. ACM Comput. Surv. 56, 11, Article 282 (June 2024), S.7. Online: <https://dl.acm.org/doi/pdf/10.1145/3672393> (Stand:30.09.2025).

⁹ Ebd. S.7.

PRT-Werte auf als „winter“. Unter diesen steht „Markt“ (1,131) an der höchsten Position, gefolgt von „Unternehmen“ (1,127) und „Bank“ (1,121). Dies deutet darauf hin, dass sich die vorherrschenden Bedeutungen dieser drei Begriffe stärker verändert haben.

Im Gegensatz zu PRT führt Average Pairwise Distance (APD) keine Clustering- oder Mittelwertberechnungen durch. Es wird direkt auf alle Wort-Embeddings in zwei Korpora bezogen. Konkret berechnet diese Methode die Varianz zwischen allen Elementpaaren aus zwei Mengen kontextueller Darstellungen anhand des Euklidischer Abstand. Diese Varianz ermöglicht die Bewertung der Verteilungsunterschiede von Zielwörtern innerhalb des semantischen Raums und letztlich die Messung semantischer Bedeutungswandel. Wenn beispielsweise ein Zielwort neue semantische Bedeutungen erhält, unterscheidet sich seine Einbettungsverteilung von früheren Versionen, wodurch sich der APD-Wert ändert. Je höher der APD-Wert, desto ausgeprägter ist der Unterschied bei der Verteilung des Zielworts zwischen den beiden Zeiträumen. Diese Veränderung könnte als Signal für das Auftreten neuer Bedeutungen und Verwendungen oder für den Niedergang älterer Bedeutungen und Verwendungen dienen. Die APD-Ergebnisse dieser Arbeit zeigen einen ähnlichen Trend zu PRT. Die APD-Werte für Markt (13,5566) und Company (13,2987) sind höher als die für Bank (12,0051) und Winter (11,2459). Diese Daten zeigen, dass Markt und Company innerhalb des Korpus von 1850 bis 1900 deutlichere Verschiebungen oder Verbreitungen aufwiesen. Solche veränderten semantischen Verteilungen sind auch Symbol für semantische Bedeutungswandel.

Es ist zu beachten, dass diese typenbasierten Methoden zwar eine Analyse der allgemeinen semantischen Bedeutungswandel ermöglichen. Es ist jedoch schwierig, genau zu bestimmen, wie diese Veränderungen ablaufen, ob es sich um semantische Verbreitung, semantische Verschiebung oder andere Arten von Veränderungen handelt. Um dieser Einschränkung zu begegnen, brauchen weitere Methoden eingeführt zu werden.

Im Gegensatz zu typbasierten Methoden berücksichtigen sinnbasierte Methoden den kontextuellen Gebrauch von Zielwörtern. Bei der Analyse werden alle

Bedeutungen des Zielworts berücksichtigt. Dadurch lassen sich semantische Bedeutungswandel bei mehrdeutigen Wörtern besser erkennen. Cosine distance between cluster distributions (CDCD) ist eine sinnbasierte Approaches. Dieser Methode behandelt Clusterverteilungen als Vektoren. Sie bewertet semantische Verschiebungen anhand der Kosinusdistanz zwischen Clusterverteilungen. Konkret werden innerhalb derselben Funktion alle Einbettungsvektoren aus zwei Zeiträumen geclustert und die Anzahl der Einbettungsvektoren innerhalb jedes Clusters gezählt. Anhand dieses Ergebnisses lässt sich das Verteilungsverhältnis der Einbettungsvektoren in beiden Korpora ermitteln. Wenn beispielsweise 80 von 100 Verwendungsbeispielen zur Bedeutung A und 20 zur Bedeutung B gehören, beträgt die Verteilungswahrscheinlichkeit $[0,8, 0,2]$. Daher kann CDCD beurteilen, ob die häufigen verwendeten Bedeutungen des Zielworts eine Verschiebung erfahren haben. Nach den Ergebnissen der Forschung ist der CDCD-Wert für „winter“ mit 0,1288 am niedrigsten. Die Werte für die anderen drei Zielwörter aus dem kommerziellen Bereich waren deutlich höher als für „winter“ (Abb.5). Der Wert für „company“ beträgt 0,61 und ist besonders hoch im Vergleich zu den anderen Zielwörtern. Dies zeigt, dass sich die semantischen Verteilungsanteile dieser drei Wörter in gewissem Maße verändert haben, wobei die Veränderung bei „company“ am ausgeprägtesten ist. Dies bedeutet, dass die Verwendungshäufigkeit einiger Bedeutungen der Zielwörter zugenommen hat, wobei einige sogar dazu tendieren, sich zu neuen dominanten Bedeutungen zu wandeln.

Die quantitative Analyse dieser drei Funktionen zeigt, dass die semantischen Bedeutungswandel für die vier Zielwörter weitgehend übereinstimmen. Die Werte für „Winter“ blieben durchweg am niedrigsten. Die drei Zielwörter aus dem kommerziellen Wortschatz weisen deutlich höhere Werte. Diese Erkenntnis bestätigt, dass diese drei Wörter im 19. Jahrhundert ein stärkerer semantischer Bedeutungswandel durchlaufen haben. Bei genauerer Betrachtung zeigen sich jedoch Unterschiede in den Einzelheiten. Der CDCD-Wert für „company“ ist am höchsten. Dies deutet darauf hin, dass die semantische Bedeutung, die den größten Anteil während dieses Zeitraums aufweist, die bedeutendste

Veränderung erfahren hat. Dies zeigt ein mögliches Phänomen auf, bei dem eine Bedeutung eine andere verdrängt hat und zur neuen dominanten Bedeutung geworden ist. Im Vergleich dazu sind die PRT- und APD-Werte für „Markt“ am höchsten, was auf erhebliche Verschiebungen in der Polysemie und der dominanten Wortbedeutung dieses Begriffs hinweist. Die PRT- und APD-Werte für „Bank“ sind relativ niedrig, doch ihr CDCD-Wert ist höher als der von „Markt“. Dies deutet darauf hin, dass sich der semantische Vektorraum von „Bank“ nicht erheblich verändert hat, sondern dass sich die Häufigkeit seiner semantischen Verwendung verschoben hat.

4. Diskussion

Die quantitative Analyse zeigt, dass im Vergleich zum relativ stabilen Begriff „Winter“ die drei Begriffe „company“, „Bank“ und „Markt“ alle in unterschiedlichem Maße eine semantische Verschiebung erfahren haben (Abb.5). Durch die Verbindung von historischem Kontext und spezifischem semantischem Inhalt kann man die semantischen Bedeutungswandel dieser drei Handelsbegriffe genauer erklären.

Von den drei Zielbegriffen erfuhr „company“ die deutlichste Verschiebung in der Verteilung der Bedeutungen. Der Streudiagramm zeigt, dass die kommerzielle Bedeutung von „company“ nach und nach zur häufigen auftretenden Bedeutung geworden ist. Diese Entwicklung entsprach weitgehend dem Entwicklungsverlauf des amerikanischen Unternehmenssystems. Im 18. und 19. Jahrhundert kam es zu einem raschen Wirtschaftswachstum. In dieser Zeit tauchten zahlreiche Begriffe auf, die sich auf die kommerzielle organisatorische Bedeutung von „company“ bezogen, wie beispielsweise „subsidiary company“, „proprietary company“ und „stock-company“¹⁰.

Die semantischen Bedeutungswandel von „Markt“ weist andere Merkmale auf als die von „company“. Dieser Zielbegriff zeigt die bedeutendsten Verschiebungen in den vorherrschenden Bedeutungen und Verteilungsvariationen innerhalb des

¹⁰ Oxford English Dictionary: company. Online: https://www.oed.com/dictionary/company_n?tab=meaning_and_use&hide-all-quotations=true#8794106 (Stand:30.09.2025).

semantischen Raums. Dies deutet darauf hin, dass „Markt“ in diesem Zeitraum die ausgeprägtesten Veränderungen in seiner Polysemie durchlaufen hat. Eine mögliche Form ist das Aufkommen abstrakter, wirtschaftlichbezogener Bedeutungen¹¹. Dabei ist zu beachten, dass die alte Bedeutung dieses Zielbegriffs im Sprachgebrauch nicht durch die neue ersetzt wurde. Für den „Markt“ besteht die Bedeutung physischer Handelsplätze zusammen mit der abstrakten wirtschaftlichen Tätigkeit. Dies entspricht der historischen Realität, in der abstrakte Handelsplätze und physische Märkte gleichzeitig existierten¹².

Unter den drei Zielbegriffen aus dem Bereich Handel und Wirtschaft weist „Bank“ die kleinste Änderung in Hinblick auf Polysemie oder Verschiebungen innerhalb des Vektorraums seine dominante Bedeutung auf. Dies entspricht der historischen Tatsache, die im Oxford English Dictionary festgehalten ist, dass die Bedeutung von „Bank“ als Wirtschaftsinstitution bereits vor dem 18. Jahrhundert existierte¹³. Daher ist die Veränderung im Vektorraum der Semantik des Zielworts innerhalb des in diesem Artikel untersuchten historischen Zeitraums relativ gering. Dennoch lässt sich beobachten, dass die Bedeutung von „Finanzinstitute“ gegenüber anderen Bedeutungen zunehmend häufiger verwendet wird und sich zu einer vorherrschenden Interpretation zu wandeln beginnt. Diese Veränderung spiegelt auch die zunehmende Verbreitung von Banken als Finanzinstitute im breiteren gesellschaftlichen Diskurs wider.

5. Evaluation

Um sicherzustellen, dass diese Arbeit richtige und ziemlich komplette Signale der semantischen Bedeutungswandel erkennen kann, ist die Evaluation der Methode ein wichtiger Teil. Aber im Bereich der semantischen Bedeutungswandel gibt es im Moment keinen sogenannten Standarddatensatz oder Goldstandard für die Evaluation. Daher wurden im Rahmen dieser Arbeit zwei Arten von

¹¹ John Regan, Semantic change and knowledge transmission: Some case studies in the distribution of words in ECCO, *Digital Scholarship in the Humanities*, Volume 37, Issue 4, December 2022. Online: <https://academic.oup.com/dsh/article/37/4/1141/6510715> (Stand:30.09.2025).

¹² Ebd.

¹³ Oxford English Dictionary: bank. Online: https://www.oed.com/dictionary/bank_n3?tab=meaning_and_use#1208979140 (Stand:30.09.2025).

Validierungsversuchen konzipiert, um die Wirksamkeit und Sicherheit des Analyseverfahrens zu bewerten.

Zunächst wurde die Methode anhand eines kleinen Testkorpus validiert, der durch Zufallsstichproben aus einer umfassenden Korpus gebildet wurde, wobei „Bank“, „Lane“ und „One“ als Zielwörter verwendet wurden. „Lane“ ist ein semantisch stabiles Wort, das in den Word Usage Graphs (WUGs) der Universität Stuttgart markiert ist. „One“ ist ebenfalls ein semantisch stabiles Wort, basierend auf einer allgemeinen Erkenntnis. Durch die Ergebnisse dieser drei Wörter im Testkorpus können in gewissem Maße beurteilen, ob das Modell die unterschiedlichen Bedeutungen der Zielwörter in verschiedenen Kontexten unterscheiden kann. Zunächst werden die nach der Einbettung generierten Vektoren einer visuellen Prüfung unterzogen. Die manuelle Stichprobenentnahme aus Streudiagrammen, die aus dem Kontext von „bank“ generiert wurden, bestätigt, dass Vektoren, die die Bedeutung „Flussufer“ repräsentieren, und solche, die die Bedeutung „Finanzinstitute“ repräsentieren, räumlich klar voneinander getrennt sind (Abb.6). Darüber hinaus zeigt ein Vergleich der Verteilungen der drei Zielwörter, dass sich der Vektorzentrum für „bank“ innerhalb zwischen c1 und c2 deutlich verschoben hat. Die Vektorverteilungen für „lane“ und „one“ erscheinen visuell einheitlich und überschneiden sich stark (Abb. 7,8). Dies entspricht ihren stabilen Eigenschaften. Dies zeigt, dass das Modell zwischen den verschiedenen Bedeutungen des Zielworts innerhalb des Kontexts unterscheiden kann. Aus den quantitativen Ergebnissen ist zu erkennen, dass „bank“ höhere APD-, CDCD- und PRT-Werte aufweist als die beiden anderen Wörter. Dies deutet darauf hin, dass „bank“ deutlichere semantische Verschiebungen erfahren hat. Es ist jedoch auffällig, dass „lane“, ein semantisch stabiles Wort aus WUGs, höhere Werte als „one“ aufweist. Eine mögliche Erklärung ist, dass die Forschung des WUGs-Projekts zu diesem Zielwort die Zeiträume 1821–1860 und 1960–2008 umfasst. Dies deckt jedoch nicht den frühen Korpus dieser Arbeit ab. Folglich kann nicht völlig ausgeschlossen werden, dass dieses Zielwort im späten 18. Jahrhundert semantische Bedeutungswandel erfahren hat. Um den Einfluss der Anzahl der

Cluster (n-Cluster) auf die Ergebnisse zu überprüfen, wurden bei dieser Evaluation drei verschiedene Einstellungen getestet, nämlich $n=3$, 5 und 7. Die Ergebnisse zeigten weitgehend den gleichen Trend. Dies belegt, dass die Anzahl der Cluster einen geringen Einfluss auf die Forschungsergebnisse hat. Letztendlich wurde $n=5$ als Einstellung ausgewählt.

Um weiter zu evaluieren, ob die Forschungsmethode semantische Bedeutungswandel, die auf zeitliche Faktoren zurückzuführen sind, effektiv messen kann, wurde in dieser Arbeit auch ein Kontrollversuch durchgeführt. Das Shuffling-Experiment ist eine typische Methode bei Kontrollversuch. Dabei werden alle Dateien aus zwei Testkorpora gemischt und dann zufällig zwei neue Korpusse extrahiert, die nicht von zeitlichen Faktoren beeinflusst sind. Die Ergebnisse der vollständigen Analyse mit „Bank“ als Zielbegriff zeigten, dass die Vektorverteilungen der beiden Teilkorpora nach der Vermischung sehr ähnlich waren, was einen starken Kontrast zu den ursprünglichen Testergebnissen darstellt (Abb.9). Dies bedeutet auch, dass in Korpora, die nicht von zeitlichen Faktoren beeinflusst sind, keine erkennbaren semantischen Bedeutungswandel erkennbar sind. Aus quantitativer Sicht zeigten alle Werte dieses Zielwort nach dem Mischen einen deutlichen Abfall, beispielsweise sank der CDCD-Wert von 0,8706 auf 0,2852 (Abb.10). Diese Ergebnisse bestätigen, dass die in dieser Arbeit gefundenen semantischen Bedeutungswandel auf die chronologische Abfolge des Korpus zurückzuführen sind und nicht auf andere Störfaktoren. Diese beiden genannten Methoden zeigen, dass die Methoden und Verfahren dieser Studie in der Lage sind, semantische Bedeutungswandel des Zielwortes zu erkennen, und diese Änderungen durch den Zeitfaktor verursacht werden.

Fazit

Diese Arbeit führt eine Analyse der semantischen Bedeutungswandel ausgewählter Begriffe aus der Wirtschafts- und Geschäftswelt im Englisch des 19. Jahrhunderts durch. Die Ergebnisse, die auf drei Messmethoden (PRT, APD und CDCD) basieren, zeigen, dass die Wirtschaftsbegriffe „market“, „company“ und „bank“ im 19. Jahrhundert einen deutlichen Anstieg der

Verwendungshäufigkeit sowie semantische Bedeutungsveränderungen erfahren haben. Die semantischen Verschiebungen dieser drei Zielwörter können bis zu einem gewissen Grad als Symbol für den Einfluss neuer sozialer Strukturen und Geschäftsmodelle auf die Sprache angesehen werden.

Allerdings müssen einige Einschränkungen berücksichtigt werden. Der in dieser Untersuchung verwendete Korpus enthielt nur eine Textart, nämlich Zeitungsartikel, und eine begrenzte Auswahl an Zielwörtern. Diese Faktoren beeinflussen möglicherweise die Repräsentativität der Versuchsergebnisse. Daher sollten zukünftige Untersuchungen eine Erweiterung des Korpus oder die Einbeziehung vergleichender Analysen verschiedener Textgattungen in Betracht ziehen.

Literatur und Quelle

Asri, Wahyu Kurniati; Rhamadanty, Winda Ayu Utami; Burhamzah, Muftihaturrahmah; Alamsyah: Analyzing Semantic Shifts in English and German by Exploring Historical Influences and Societal Dynamics. 2024. Online: https://www.researchgate.net/publication/381722204_Analyzing_semantic_shifts_in_English_and_German_by_exploring_historical_influences_and_societal_dynamics S2 (Stand:30.09.2025).

Jader Martins Camboim de Sá, Marcos Da Silveira, Cédric Pruski: Survey in Characterization of Semantic Change. 2024. Online: <https://arxiv.org/html/2402.19088v2> (Stand:30.09.2025).

John Regan, Semantic change and knowledge transmission: Some case studies in the distribution of words in ECCO, Digital Scholarship in the Humanities, Volume 37, Issue 4, December 2022. Online: <https://academic.oup.com/dsh/article/37/4/1141/6510715> (Stand:30.09.2025).

Library of Congress: Rights and Access. Online: <https://www.loc.gov/collections/chronicling-america/about-this-collection/rights-and-access/> (Stand:30.09.2025).

Oxford English Dictionary: bank. Online: https://www.oed.com/dictionary/bank_n3?tab=meaning_and_use#1208979140 (Stand:30.09.2025).

Oxford English Dictionary: company. Online: https://www.oed.com/dictionary/company_n?tab=meaning_and_use&hide-all-quotations=true#8794106 (Stand:30.09.2025).

Periti, Francesco; Montanell, Stefano: Lexical Semantic Change through Large

Language Models: a Survey. ACM Comput. Surv. 56, 11, Article 282 (June 2024).

Online: <https://dl.acm.org/doi/pdf/10.1145/3672393> (Stand:30.09.2025).

Tahmasebi, Nina; Dubossarsky, Haim: Computational modeling of semantic change.

Abbildung

Abb. 1: Ergebnisse für Textabdeckung und Wortfrequenz für „company“

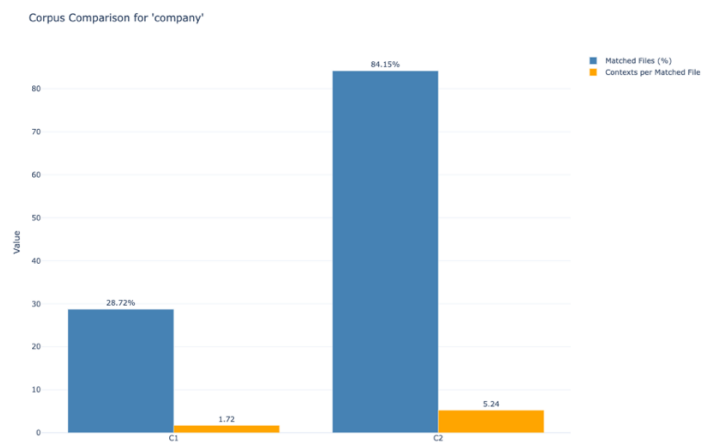


Abb.2: Ergebnisse für Textabdeckung und Wortfrequenz für „market“

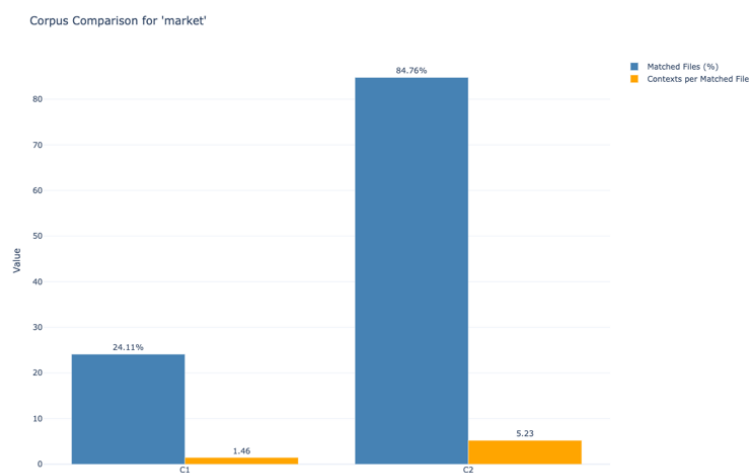


Abb.3: Ergebnisse für Textabdeckung und Wortfrequenz für „bank“

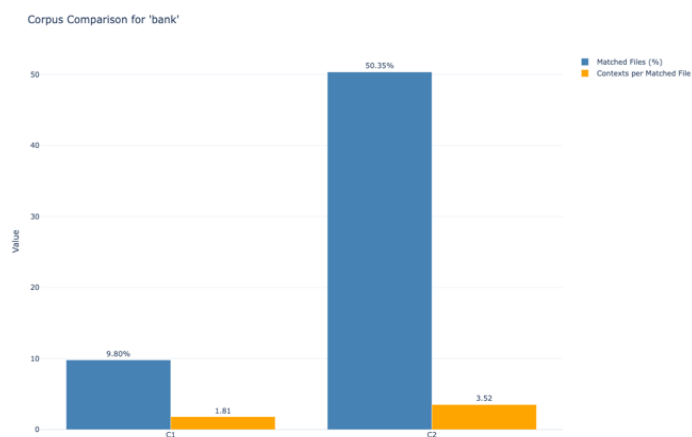


Abb.4: Ergebnisse für Textabdeckung und Wortfrequenz für „winter“

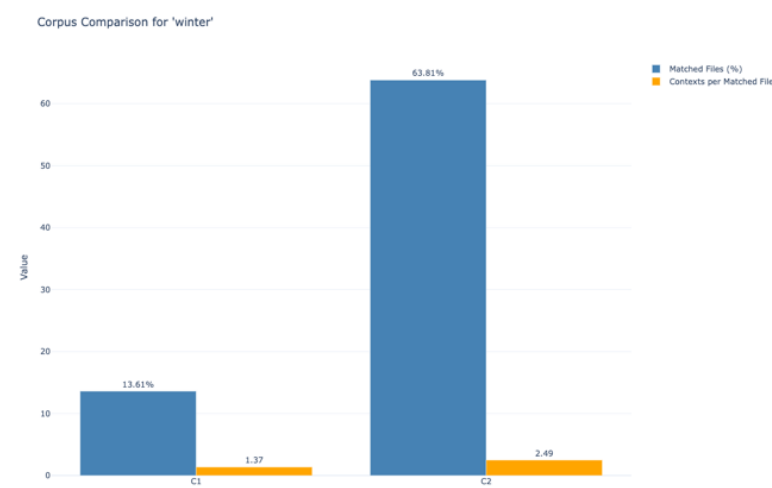


Abb.5: Tabelle mit den quantitativen Ergebnissen der Zielwörter

	PRT	APD	CDCD
market	1.1305	13.5566	0.3299
company	1.1266	13.2987	0.6083
bank	1.1209	12.0051	0.3774
winter	1.0242	11.2459	0.1288

Abb.6: Streudiagramm der Einbettungsergebnisse für das Zielwort „Bank“ im Testkorpus.

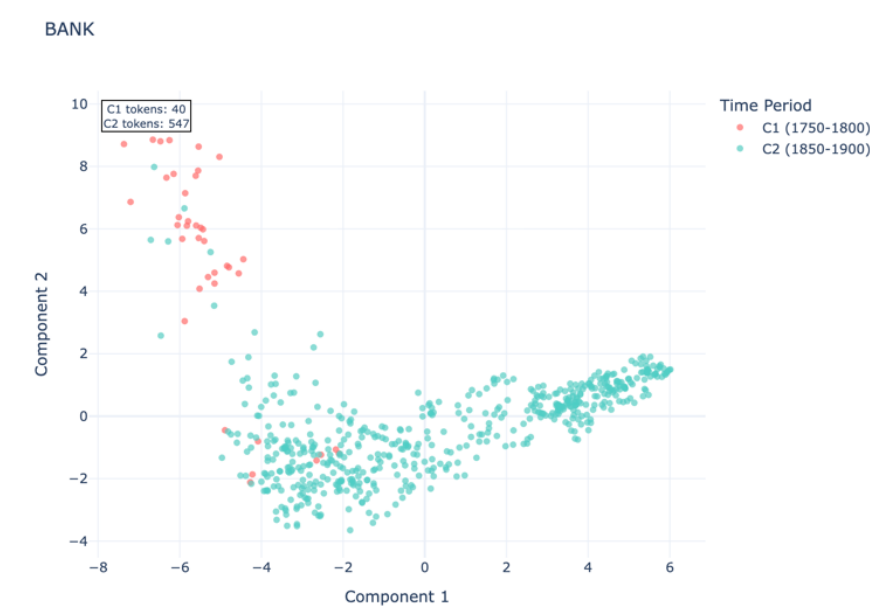


Abb.7: Streudiagramm der Einbettungsergebnisse für das Zielwort „One“ im Testkorpus.



Abb.8: Streudiagramm der Einbettungsergebnisse für das Zielwort „Lane“ im Testkorpus.

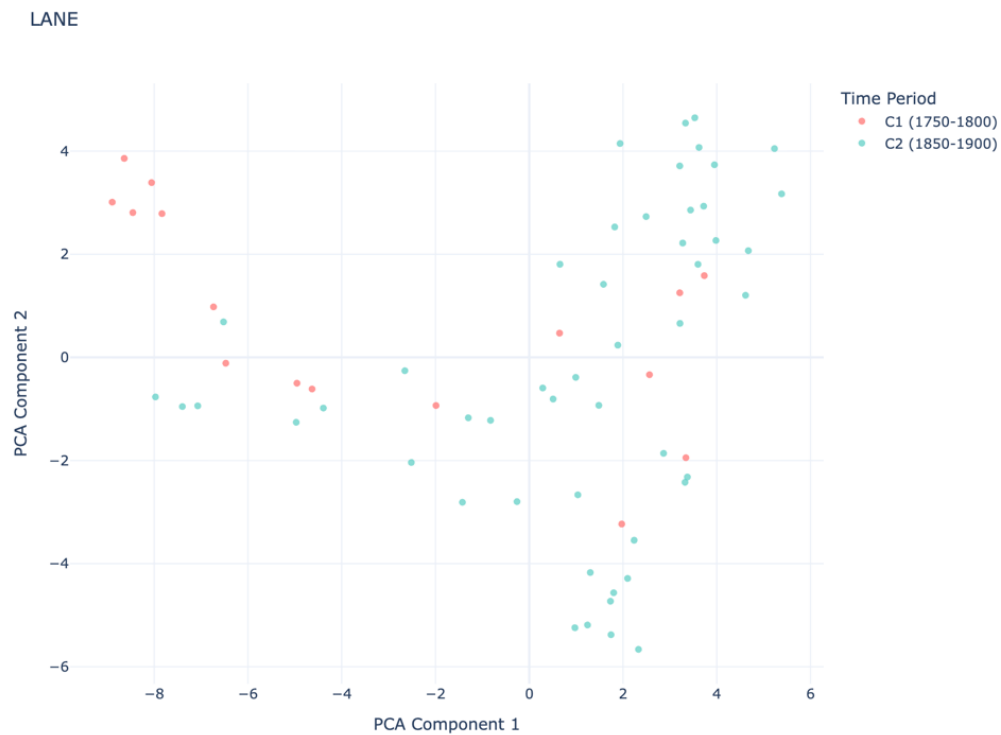


Abb.9: Streudiagramm der Einbettungsergebnisse für das Zielwort „Bank“ im Shuffle-Test.



Abb.10: Quantitative Ergebnisse der Tests zur Evaluierung der Methode.

Evaluation					
	PRT	APD	CDCD(n_cluster =5)	CDCD(n_cluster =3)	CDCD(n_cluster =7)
bank	1.2922	13.3367	0.8706	0.7873	0.8780
one	1.0139	12.9648	0.0332	0.0037	0.0703
lane	1.1329	12.7808	0.4145	0.2921	0.5108
bank(misch)	1.0115	10.5398	0.2852	0.0323	0.3143

Ich versichere, dass ich die Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Sämtliche wörtlichen oder sinngemäßen Übernahmen und Zitate sind kenntlich gemacht und nachgewiesen.

Ferner versichere ich, dass das Thema dieser Arbeit nicht identisch ist mit dem Thema einer von mir bereits für eine andere Prüfung eingereichten Arbeit.

Ich erkläre weiterhin, dass ich die Arbeit nicht bereits an einer anderen Hochschule als Prüfungsleistung eingereicht habe.

Datum, Unterschrift