

Alexander Koplenig/Carolyn Müller-Spitzer/Sascha Wolfer
(Mannheim)

Wortschatzwandel im „Spiegel“

Abstract: Diachrone Wortschatzveränderungen werden in der Regel exemplarisch anhand bestimmter Phänomene oder Phänomenbereiche untersucht. Wir widmen uns der Frage, ob und wie Wandelprozesse auch auf globaler Ebene, also ohne sich auf bestimmte Wortschatzausschnitte festzulegen, messbar sind. Zur Untersuchung dieser Frage nutzen wir das Spiegel-Korpus, in dem alle Ausgaben der Wochenzeitschrift seit 1947 enthalten sind. Dabei gehen wir auf grundlegende Herausforderungen ein, die es dabei zu lösen gilt, wie die Verteilung sprachlicher Daten und die Folgen unterschiedlicher Subkorpusgrößen, d.h. im konkreten Fall die variierende Größe des Spiegelkorpus über die Zeit hinweg. Wir stellen ein Verfahren vor, mit dem wir in der Lage sind, flankiert von einem „Lackmustest“ zur Überprüfung der Ergebnisse, Wortschatzwandelprozesse bis auf die Mikroebene, d.h. zwischen zwei Monaten oder gar Wochen, quantitativ nachzuvollziehen.

1 Einleitung

Liest man eine Ausgabe des Nachrichtenmagazins „Der Spiegel“ aus den 50er- oder 60er-Jahren, so scheint sich die Sprache zumindest subtil von unserem aktuellen Sprachgebrauch zu unterscheiden. Ein kurzes Beispiel:

Kann ein Verleger oder ein Buchhändler die Niederschrift von Lebenserinnerungen, an der fremde Federn mitgearbeitet haben, überhaupt noch als Memoiren anbieten? (Memoiren: Solche erhöhten Wahrheiten, in: Der Spiegel 51/1952, S. 26¹)

Aus unserer Sprachintuition heraus würden wir sowohl den Titel des Artikels („Solche erhöhten Wahrheiten“), die „Niederschrift“ als auch die „fremden Federn“ als eher veraltet ansehen. Aber stimmt das überhaupt? Dies könnte man für diese einzelnen Wörter bzw. Phrasen durch Analysen in diachronen Korpora zu beantworten versuchen, auch wenn die Datenlage für das Deutsche in der Hinsicht immer noch zu wünschen übrig lässt. Aber selbst wenn man die Frage beantwor-

1 Online im Archiv unter www.spiegel.de/spiegel/print/d-21978329.html und <http://magazin.spiegel.de/EpubDelivery/spiegel/pdf/21978329> (Stand: 25.4.2018).

ten könnte, bleibt unklar: Inwiefern ist dies repräsentativ für die gesamte Sprache, die man im „Spiegel“ findet? Wie hat sich der Spiegelwortschatz über die Zeit hinweg verändert, wenn man ihn als Ganzes in den Blick nimmt?

Seit Kurzem haben wir das gesamte Spiegel-Archiv als Korpus am IDS verfügbar. Es umfasst alle 3.663 Ausgaben des „Spiegels“ von 1947 bis 2016, d.h. fast sieben Jahrzehnte deutscher Sprachgeschichte. Das Korpus besteht aus ca. 4,5 Millionen Wortformtypen und 284 Millionen Token.² Trotzdem ist das Spiegel-Korpus für uns eine sehr wertvolle Grundlage zur Untersuchung von Wortschatzwandel im Deutschen, da die verschiedenen Jahre inhaltlich vergleichbar sind (es handelt sich um die gleiche Quelle), ein nennenswerter historischer Zeitraum vollständig und in für diachrone Korpora relativ großer Textmenge abgedeckt ist und uns als Rohdaten zur Verfügung steht, was die Voraussetzung für jegliche umfassenden quantitativen Analysen darstellt. Insofern ist das Spiegel-Korpus für uns auch eine ideale Testgrundlage, um der anfänglich skizzierten Frage nachzugehen.

Das Ziel unseres Artikels ist es, Folgendes zu prüfen: Hält der impressionistische Eindruck, dass der Wortschatz im „Spiegel“ sich über die Jahre wandelt, einer quantitativen Überprüfung stand? Sind diese Wandelprozesse messbar, ohne sich auf bestimmte Wortschatzausschnitte festzulegen? Auf dem Weg hin zu diesem Ziel werden wir folgendermaßen vorgehen: Zunächst werden wir in Abschnitt 2 etablierte Wege zur Darstellung und Exploration von Wortschatzwandel am Spiegel-Korpus exemplifizieren, um einen Anschluss an die bisherige quantitativ-qualitative Forschung zumindest in Grundzügen zu gewährleisten. Daraufhin werden wir in Abschnitt 3 auf die Herausforderung der Ungleichverteilung sprachlicher Daten und in Abschnitt 4 auf die Zipf-Verteilung aus quantitativer Sicht eingehen. Im darauf folgenden Abschnitt werden wir ein Verfahren zur Messung von Wortschatzwandel vorstellen, um dann im fünften Abschnitt die Ergebnisse zum Wortschatzwandel im „Spiegel“ vorzustellen. Wir beschließen den Beitrag mit einem Fazit (Abschn. 7).

² Zum Vergleich: Das „Corpus of Historical American English“ (<https://corpus.byu.edu/coha/>; Stand: 28.4.2018), auf dessen Basis zahlreiche diachrone Untersuchungen in der Anglistik (vgl. z.B. Studien in Allan/Robinson (Hg.) 2012) durchgeführt wurden, umfasst 400 Millionen Token, spannt sich allerdings über einen breiteren historischen Zeitraum (1810–2000) und besteht aus vielseitigeren Quellen.

2 Etablierte Wege zur Darstellung und Exploration von Wortschatzwandel, exemplifiziert am Spiegel-Korpus

Diachrone Wortschatzveränderungen werden in der Regel exemplarisch anhand bestimmter Phänomene oder Phänomenbereiche untersucht. Zum Beispiel analysiert man, welche lexikalischen Innovationen im Zeitverlauf zu beobachten sind (Kerremans/Stegmayr/Schmid 2012; Würschinger et al. 2016; Müller-Spitzer/Wolfer/Kolpenig 2018). Unter einer quantitativen Perspektive wird dabei nicht nach neuen Wörtern im linguistischen Sinn gesucht, sondern entweder nach neuen Token oder neuen automatisch lemmatisierten Worttypen, die in einem neuen Jahr, einem neuen Monat oder einer neuen Ausgabe auftauchen. Die Herausforderung ist hierbei vor allem, die normale Varianz oder Fehllemmatisierungen z.B. aufgrund von Rechtschreibfehlern oder fehlerhafter automatischer Texterkennung (OCR) von echter lexikalischer Innovation zu trennen (vgl. z.B. Szmrecsanyi 2016). Dies funktioniert selten automatisch, sondern benötigt eine manuelle Auswahl. Um interaktiv explorieren zu können, wie viele echte lexikalische Innovationen neben Fehllemmatisierungen oder normaler Varianz mit einem rein automatischen Ansatz zu entdecken sind, haben wir für den „Spiegel“ ein Tool „Wortschatzwandel im Spiegel“ auf OWID^{plus} veröffentlicht.³ Mit diesem Tool kann man das gesamte Spiegelkorpus von 1947 bis 2016 nach neuen Wörtern oder Wörtern, die in ihrer Verwendung zurückgegangen sind, explorieren und visualisieren. Das Tool sucht wortartenbasiert nach (TreeTagger-basierten) Lemmata und kann über verschiedene zeitliche und häufigkeitsbasierte Parameter angepasst werden. So kann man selbst testen, wie weit man auf diesem Weg zur Entdeckung von Wandelprozessen kommt.

Eine andere Möglichkeit neben der Untersuchung lexikalischer Innovation ist, die Veränderung bestimmter Wortschatzgruppen oder Konstruktionen zu untersuchen (Perek 2016; Hilpert/Perek 2015). Auch diesen Ansatz haben wir beispielhaft durchgeführt, und zwar anhand des zeitlichen Verlaufs der Stichwörter des Wörterbuchs zum Protestdiskurs 1967/68 (vgl. Kämper 2013 und Protestdiskurs 1967/68 in OWID)⁴ im „Spiegel“ (vgl. Abb. 1).

³ <https://www.owid.de/plus/wvspiegel2018/> (Stand: 28.4.2018).

⁴ Protestdiskurs 1967/68 (2011), in: OWID – Online Wortschatz-Informationssystem Deutsch (<https://www.owid.de/wb/disk68/start.html>; Stand: 28.4.2018).

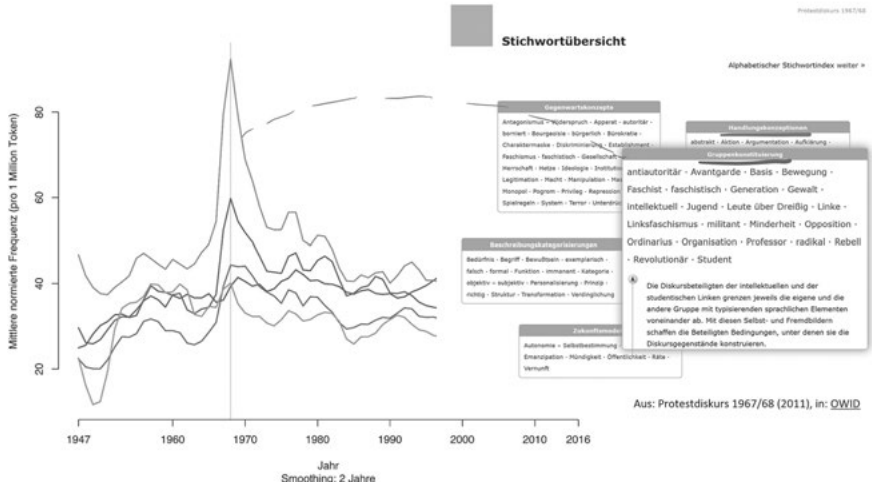


Abb. 1: Die Stichwörter des Protestdiskurses 1967/68 in ihrem zeitlichen Verlauf im Spiegel-Korpus

Bestimmte Stichwortgruppen des Protestdiskurses verändern sich deutlich in ihrer Frequenz Ende der 1960er-Jahre. Dies sind v.a. die Stichwörter zur Gruppenkonstituierung wie „antiautoritär“, „Avantgarde“, „intellektuell“ oder „militant“.⁵ Dies ist insofern nicht selbstverständlich, als für das Wörterbuch selbst nur Primärquellen als Korpusgrundlage dienten und keine pressesprachlichen Texte.

Trotzdem bleibt bei einer solchen Analyse offen: Sind diese Anstiege repräsentativ für einen Wandel als Ganzes? Stechen z.B. die späten 60er-Jahre des 20. Jahrhunderts als eine Zeit besonderen lexikalischen Wandels heraus? Und wandelt sich die Sprache auch in unauffälligen Bereichen, die nicht bestimmten Diskursen angehören und deshalb für eine qualitativ geleitete Auswahl von Wortschatzausschnitten auch kaum zum Tragen kommen?

So kommen wir zu unserer Ausgangsfrage zurück: Diese war, ob man die Unterschiede, die man bei weit auseinanderliegenden Ausgaben des „Spiegels“ oft zu beobachten glaubt, quantitativ messen kann, und zwar im Wortschatz des „Spiegels“ als Ganzem. Dies inkludiert auch die Frage, ab welchem Abstand von Ausgaben man einen Wandel messen kann, also ob auch Mikrowandelprozesse quantitativ messbar sind. Genauer wollen wir also fragen: Unterscheiden sich zwei Ausgaben des „Spiegels“ von Frühling und Herbst 1950 als Ganzes voneinander?

⁵ Die Stichwortgruppen sind online einzusehen unter https://www.owid.de/docs/disk68/stw_gruppen.jsp (Stand: 30.4.2018).

Unterscheiden sie sich weniger stark als zwei Ausgaben von 1950 und 1953? Unterscheiden sich diese wiederum messbar weniger stark als zwei Ausgaben von 1950 und 1970? Eine besondere Herausforderung bei solchen Analysen ist die Ungleichverteilung sprachlicher Daten, auf die wir im folgenden Abschnitt näher eingehen.

3 Die Ungleichverteilung sprachlicher Daten

Sprachliche Daten sind Zipf-verteilt, d.h., sehr wenige Typen kommen sehr häufig vor und sehr viele sehr selten (vgl. grundlegend Baayen 2001). Clauset und Kollegen (Clauset/Shalizi/Newman 2009; vgl. auch Engelberg 2015) haben diese Verteilung sehr anschaulich auf die Körpergröße der Bevölkerung der USA übertragen. Transferiert man deren Beispiel wiederum auf die Menschen in Deutschland, würden die ca. 80 Millionen Menschen hier folgende Körpergrößen aufweisen: Es gäbe einen Menschen so groß wie den Berliner Fernsehturm, 2.500 Menschen, die mit 5–6 Metern so groß sind wie eine durchschnittliche erwachsene männliche Giraffe, 150 Menschen, die so groß wie der größte jemals gemessene Mensch sind (ca. 2,70 m) und 50 Millionen Winzlinge, die ca. 17 Zentimeter groß sind.⁶ Dieses Beispiel macht plastisch deutlich, wie anders Zipf-verteilte Daten im Gegensatz zu normal verteilten Daten sind (Körpergrößen sind eigentlich normalverteilt). Diese extreme Verteilung der häufigen zu den seltenen Typen spiegelt sich in allen Sprachdaten, so auch im Spiegel-Korpus. Hier machen die 100 häufigsten Typen schon 47% aller Token aus, d.h. fast die Hälfte des Korpusumfangs besteht nur aus den 100 häufigsten Typen. Die 200 häufigsten Typen machen schon mehr als die Hälfte, nämlich 53% der Gesamttoken aus. Die Hapax legomena wiederum, d.h. die Typen, die nur einmal im gesamten Spiegel-Korpus vorkommen, machen 62% der Typen aus. Die häufigsten Wörter sind dabei zum großen Teil Funktionswörter und bestimmen damit vor allen Dingen, *wie* ein Sachverhalt versprachlicht wird. Die seltenen Wörter stehen mehr dafür, über was geschrieben wird. Untersucht man nun Wandelprozesse im Wortschatz, unterscheidet sich der Bereich der sehr häufigen Wörter von dem der selteneren sehr deutlich: Während der Anteil neuer Typen unter den 200 häufigsten Typen eines Jahres deutlich unter 10% liegt, ist der Anteil neuer Typen bei Hapax legomena kontinuierlich bei ca. 85%. Die Fluktuation ist im hochfrequenten Bereich also eher niedrig, während sie im nied-

⁶ Im Vortrag auf der Jahrestagung haben wir diese Verhältnisse in einem Video visualisiert. Dieses ist online verfügbar unter www.ids-mannheim.de/lexik/empirische-methoden/emp-arbeitspapiere/video-zur-zipf-verteilung.html (Stand: 5.11.2018).

rigfrequenten Bereich sehr hoch ist. Dabei ist der Wandel bei den niedrigfrequenten Typen schwerer zu interpretieren, da man nicht weiß, ob sie zufällig neu belegt sind, ob sie ein neues Thema anzeigen, welches aber womöglich sprachextern motiviert ist, oder tatsächlich sprachlichen Wandel indizieren. Möchte man also dem Sprachwandel im Kern auf die Spur kommen und nicht nur messen, welche unterschiedlichen Themen im „Spiegel“ behandelt werden, muss man sich unseres Erachtens auf die Analyse des hochfrequenten Bereichs konzentrieren.

Ein weiterer Aspekt der Ungleichverteilung soll noch kurz ausgeführt werden: Auch bei Mehrfachvorkommen ist die Verteilung der Typen meist nicht regelmäßig, sondern ist konzentriert auf einzelne textuelle Einheiten. So kommen im „Spiegel“ die 10% der Typen, die mehr als zehnmals im Spiegel-Korpus belegt sind, manchmal nur in einer Ausgabe oder sogar nur in einem Artikel vor (vgl. auch Lijffijt et al. 2014). So kommt das Wort „Krebsstammzelle“ zwar 20-mal im Spiegel-Korpus vor, aber nur in einem Artikel aus dem Juni 2007.⁷ Genauso kommt das Wort „Malvenblütentee“ zwar 14-mal vor, aber nur in einem Artikel im Juni 1961.⁸ Diese bekannte ‚Klumpung‘ sprachlicher Ereignisse wird auch unter dem Schlagwort der „Burstiness“ (Altmann/Cristadoro/Esposti 2012) diskutiert. Dieser Aspekt der Ungleichverteilung spricht dafür, die textuellen Einheiten bei der Analyse von Wortschatzwandel zu erhalten.

4 Die Zipfverteilung aus quantitativer Sicht

Abbildung 2 veranschaulicht nochmal eine typische Zipfverteilung anhand der tatsächlichen Verteilung des Spiegelwortschatzes. Die beiden Achsen sind jeweils logarithmiert, also bedeutet jeder längere Strich auf der Achse eine Verzehnfachung des jeweiligen Werts (Rang/Häufigkeit). Die Punkte im Diagramm stellen dabei die tatsächlich beobachteten Häufigkeiten für jeden Rang dar, während die gestrichelte Linie die auf den Daten basierende Zipf-Mandelbrot-Verteilung darstellt (die geschätzten Parameter finden sich unten links, die Verteilung und die Schätzung werden in Koplenig 2018 näher beschrieben). Auch in dieser Darstellung sieht man deutlich, dass es einige sehr häufige Wörter und eine große Anzahl von sehr seltenen Wörtern gibt. Zur Illustration haben wir, passend zum Titel der

⁷ „Saat des Bösen“, in: „Der Spiegel“ vom 11.6.2007, www.spiegel.de/spiegel/print/d-51878458.html (Stand: 30.4.2018).

⁸ „Suchten und fanden“, in: „Der Spiegel“ vom 21.6.1961, www.spiegel.de/spiegel/print/d-43364733.html (Stand: 30.4.2018).

diesjährigen Jahrestagung, noch die folgenden Lemmata in der Grafik abgetragen: „Neues“ „vom“ „heutigen“ „Deutsch“, „die“ „Jahrestagung“ „des“ „IDS“.

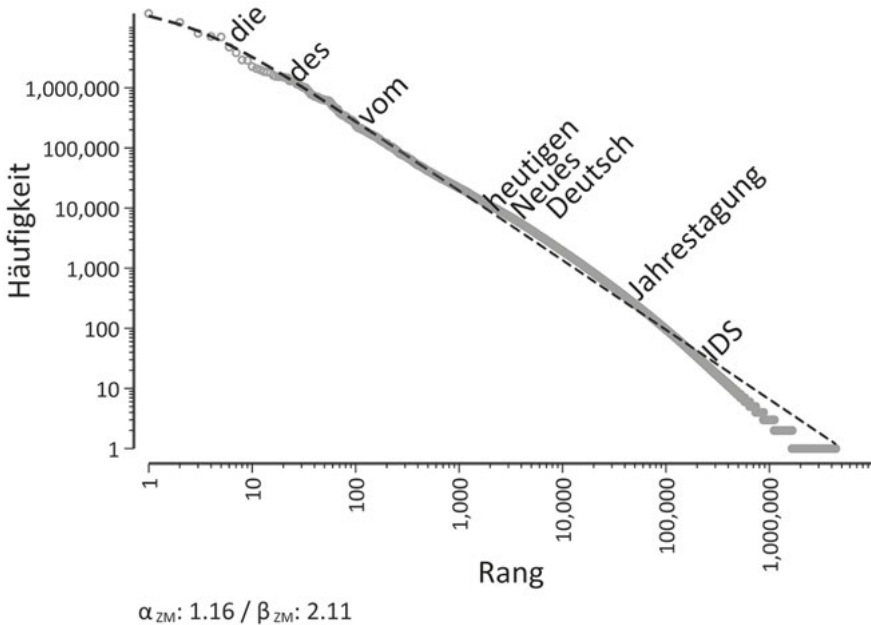


Abb. 2: Die Worthäufigkeitsverteilung des Spiegel-Korpus als Zipfverteilung

Aus theoretischer Sicht ist die Zipfverteilung insofern interessant, als es für die Entstehung der Verteilung auch nach Jahrzehnten der Forschung in diesem Bereich immer noch keine abschließende, allgemein akzeptierte Begründung gibt (Piantadosi 2014).⁹ Zusätzlich muss berücksichtigt werden, dass die Verteilung

⁹ Ein relativ aktueller und vielversprechender Ansatz führt die Entstehung auf das Zusammenspiel verschiedener Wortarten als sogenannten latenten Einflussfaktor zurück (Aitchison/Corradi/Latham 2016; Lestrade 2017). Zentral bei diesem Ansatz ist der Umstand, dass verschiedene Wortarten zum Teil sehr unterschiedliche Type-Token-Verteilungen aufweisen; man denke hier etwa an die Unterscheidung zwischen offenen und geschlossenen Wortarten. Zum Beispiel zeigt sich interessanterweise innerhalb der Wortart „Substantive“ selbst keine Zipfverteilung. Diese Beobachtungen verbinden Aitchison und KollegInnen dann in einem komplexen statistischen Modell mit den Unterschieden innerhalb der verschiedenen Wortarten hinsichtlich Syntax und Semantik und können belegen, dass sich dann unter Umständen „ganz von selbst“ eine Zipfverteilung ergibt. Auch hier gilt es jedoch noch abzuwarten, wie sich dieser Ansatz mittelfristig bewähren wird.

auch relevant für die kognitive Sprachverarbeitung ist. Dies demonstriert z.B. eine besonders prägnante Studie, die Piantadosi (2014) in seinem nach wie vor sehr lesenswerten Überblicksartikel zur Zipfverteilung erwähnt. Dabei bekamen 25 Versuchspersonen folgende Aufgabe: „Ein Raumschiff stürzt in der Wüste von Nevada ab. Acht Wesen steigen aus, ein Wug, ein Plit, ein Blicket, ein Flark, ein Warit, ein Jupe, ein Ralex und ein Timon. Beschreiben Sie mit mindestens 2.000 Worten, was als Nächstes passiert.“¹⁰ Piantadosi (2014) hat für jede Versuchsperson berechnet, wie häufig die jeweiligen Wesen in den Geschichten der Versuchspersonen genannt werden. Natürlich gibt es hier individuelle Unterschiede, zum Beispiel verwendet eine Versuchsperson das „Wug“ als Hauptcharakter, während eine andere Person sich eher für das „Blicket“ entscheidet. Wenn man dann aber die Verwendungshäufigkeiten über die Versuchspersonen hinweg aggregiert und von häufig nach selten anordnet, so ergibt sich wieder eine Art Zipfverteilung, d.h., dass das häufigste Wesen doppelt so oft genannt wird wie das zweithäufigste Wesen usw. Da es sich hier nur um eine Teststudie handelt, würde es sich sicherlich lohnen, eine größere Studie dieser Art durchzuführen, wie auch Piantadosi selbst anmerkt. Unter Umständen spiegeln sich in der Zipfverteilung allgemeine kognitive Prozesse. Dies ist auch deshalb plausibel, weil man weiß, dass die Häufigkeit des Auftretens einer sprachlichen Einheit Einfluss auf die kognitive Verarbeitung von Sprache hat, vgl. z.B. das folgende Zitat:

Die Aktivierungsneigung von Assoziationsmustern korreliert mit dem Grad der Verfestigung: häufig Verarbeitetes wird leichter wieder aktiviert als selten oder noch nie Verarbeitetes. Damit schließt sich die Entrenchmentrückkopplungsschleife, denn häufig Aktiviertes erfährt stärkeres Entrenchment, was sich wiederum in schnellerer und leichter Aktivierung niederschlägt. (Schmid 2018, S. 222; vgl. auch Bybee 2015; Behrens/Pfänder 2016; Ellis 2012)

Eine umfassende Erklärung der Zipf-Verteilung sollte dementsprechend die kognitive Relevanz miteinschließen.

Aus methodischer Sicht führt die Zipfverteilung häufig zu erheblichen Schwierigkeiten bei der Quantifizierung von empirischen Phänomenen. Oft vernachlässigt wird in diesem Kontext die Nicht-Stationarität von Worthäufigkeitsverteilungen. Hierunter versteht man den Umstand, dass die Häufigkeit eines Wortes von verschiedenen Faktoren abhängt. Zum Beispiel wird man in einem Bericht über die große Koalition eher das Wort „Politik“ lesen können als in einem Bericht über die Fußball-Bundesliga (vgl. die Ausführungen oben zu „Burstiness“). Auf

¹⁰ Unsere Übersetzung von „An alien space ship crashes in the Nevada desert. Eight creatures emerge, a Wug, a Plit, a Blicket, a Flark, a Warit, a Jupe, a Ralex, and a Timon. In at least 2000 words, describe what happens next“ (Piantadosi 2014, S. 16).

der einen Seite führt gerade diese Variabilität dazu, dass sich Texte thematisch so gut klassifizieren lassen, zum Beispiel im Rahmen von Topic-Modellen (Blei/Lafferty 2006), in der Latent-Semantic-Analysis (Deerwester et al. 1990) oder ganz allgemein im Rahmen des Information Retrieval (Manning/Raghavan/Schütze 2008). Auf der anderen Seite macht die Kohärenz von Texten die Wortschatzanalyse aber schwierig, wie Baayen zeigt (2001; Kap. 5).

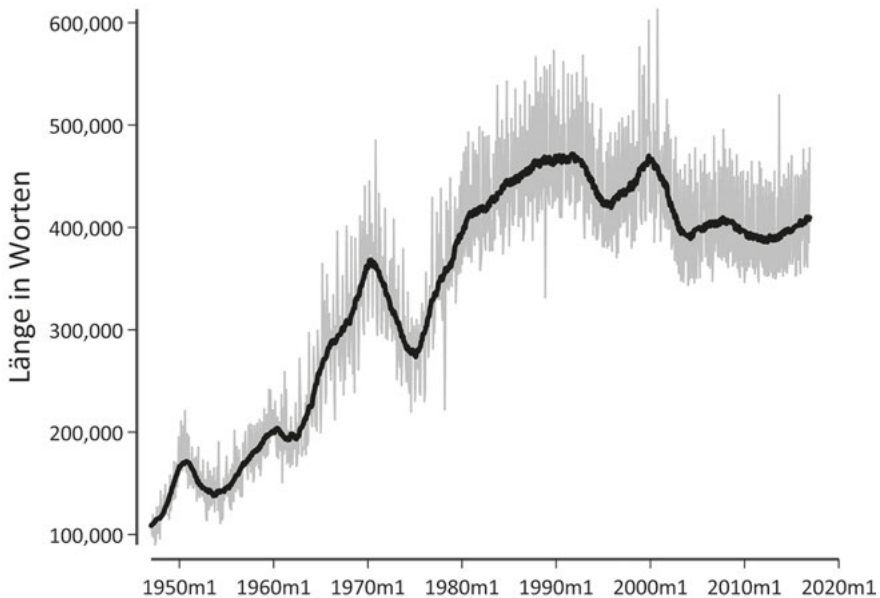


Abb. 3: Größe des Spiegelkorpus im Zeitverlauf

Unabhängig davon gilt ganz generell, dass Maßzahlen in der Korpuslinguistik im Prinzip ausnahmslos von der Korpusgröße abhängig sind (Tweedie/Baayen 1998). Für unsere Untersuchung ist das insofern wichtig, als die einzelnen Ausgaben des „Spiegels“ über die Zeit hinweg wesentlich länger geworden sind, wie Abbildung 3¹¹ zeigt. Dieser Umstand muss bei der Analyse des Spiegelwortschatzes berücksichtigt werden.

¹¹ Die graue Linie zeigt dabei die auf Monatebene aggregierten Rohdaten, während die schwarze Linie einen sogenannten „Smoother“ darstellt (Beckett 2013), welcher für jeden Datenpunkt einen lokalen Mittelwert (hier jeweils ± 12 Monate) berechnet und damit ein Glättungsverfahren darstellt, welches zufällige Schwankungen aus den Daten herausrechnet, damit sich zeitliche

5 Die Messung von Wortschatzwandel

Möchte man sich ganz generell dem Thema Wortschatzwandel aus quantitativer Sicht nähern, so kann man beispielsweise einzelne Wörter bzw. Lemmata auswählen und deren Häufigkeitsverlauf über die Zeit hinweg analysieren (Michel et al. 2010; Davies 2012; siehe auch Abschn. 2). Während dies für einzelne Phänomene und Bereiche ein durchaus vielversprechendes Verfahren darstellt, hat man gleichzeitig das Problem, dass – wie bereits ausgeführt – immer nur ein bestimmter, nach subjektiven Kriterien ausgewählter Wortschatzausschnitt untersucht wird. Möchte man dagegen wissen, wie sich der Wortschatz in bestimmten Zeiträumen ‚global‘ geändert hat, muss man anders vorgehen. Eine Möglichkeit ist, zwei Korpora (in diesem Fall zwei diachrone Korpusausschnitte), wie in Kopenig (2017b) dargestellt, zu vergleichen: Man verrechnet mittels eines geeigneten statistischen Verfahrens die Ähnlichkeit der Worthäufigkeitsverteilung zu zwei Zeitpunkten, zum Beispiel indem man zwei verschiedene Spiegelausgaben miteinander vergleicht. Ein niedriger Wert steht dann (zumeist) für hohe Ähnlichkeit, also für ein niedriges Maß an Wortschatzwandel, ein hoher Wert dagegen für hohe Unähnlichkeit, also für stark ausgeprägten Wortschatzwandel. Dabei handelt es sich also um eine globale Maßzahl, welche Veränderungen auf verschiedenen Ebenen subsumiert (neue Wörter, obsoleete Wörter, unterschiedliche Verwendungshäufigkeiten) und dann eine tiefere Analyse auf der Ebene von einzelnen Lemmata oder auch Kollokationen ermöglicht (Kilgarriff 2001; Kopenig 2015). Altmann/Dias/Gerlach (2017) präsentieren basierend auf dem Konzept der „generalisierten Entropien“ allerdings ein völlig neues Messverfahren zum Vergleich zweier Korpora (vgl. auch Gerlach/Font-Clos/Altmann 2016; Dias et al. 2018). Dieses verspricht differenziertere Einsichten in die Ähnlichkeit zweier Korpora im Bereich verschiedener Häufigkeitsklassen als das eben genannte Verfahren, da das Verfahren einen Parameter α beinhaltet, den man frei variieren kann. Wenn man α erhöht (bzw. reduziert), erhöht (bzw. reduziert) man den Einfluss der häufigsten Wörter bei der Ähnlichkeitsberechnung. Der Parameter ist deshalb so interessant, weil das Verfahren damit in gewisser Sicht eine elegante „quantitative Lupe“ darstellt: Wählt man einen niedrigen Parameter, so kann man zum Beispiel den Fokus darauf richten, *worüber* im „Spiegel“ berichtet wird, während ein hoher Parameter die Aufmerksamkeit darauf richtet, *wie* über etwas im „Spiegel“ berichtet wird. Gleichzeitig zeigen Altmann und Kollegen (Altmann/Dias/Gerlach 2017) jedoch auch, dass die Zipfverteilung einen erheblichen Einfluss auf solche informationstheore-

Trends besser visuell erkennen lassen. Wenn nicht anders erwähnt, sind alle im Folgenden dargestellten Zeitreihen nach diesem Verfahren geglättet.

tischen Maßzahlen hat. Dabei belegen die Autoren theoretisch und empirisch, wie groß zwei zu vergleichende Korpora im Allgemeinen sein müssen, um überhaupt präzise Ähnlichkeiten ausrechnen zu können. Auf diesen Umstand wurde in der Literatur bisher noch kaum Rücksicht genommen. Setzt man für α etwa den Wert 1 an, was der Standard-Jensen-Shannon-Divergenz entspricht, so benötigt man zur unverzerrten Abschätzung der Werte Korpora, die mindestens 1.000 Millionen Wortformtoken groß sein müssen.¹² Um Ähnlichkeiten auf Basis eines niedrigeren Parameters zu schätzen, benötigt man schnell unrealistisch große Textsammlungen; für $\alpha = 0,6$ beispielsweise benötigt man bereits 100 Trillionen Wortformtoken.

Wie anfangs gesagt, wollen wir den Spiegelwortschatz als Ganzes in den Blick nehmen. Da wir dafür Subkorpora für diachrone Zeitabschnitte, z.B. für einzelne Monate, bilden müssen, müssen wir auch hierfür den α -Parameter entsprechend einstellen. Für unsere Untersuchung setzen wir deshalb α bei 1,5 an. Hierfür benötigt man circa 100.000 Wortformtoken zur Ähnlichkeitsschätzung, eine Größenordnung, die für die auf Monatsebene aggregierten Spiegeldaten gut erfüllt ist (siehe Abb. 3). Über das Heaps'sche Gesetz lässt sich dann ausrechnen (Altmann/Dias/Gerlach 2017), dass 99% der Unähnlichkeit zwischen zwei Korpora bzw. Texten auf den circa 10.000 häufigsten Wortformtypen beruht. Die Ähnlichkeit berechnet sich über die folgende Formel:

$$\tilde{D}_\alpha(p, q) = \frac{\frac{1}{1-\alpha} \left(\sum_i \left(\frac{p+q}{2} \right)_i^\alpha - 1 \right) - \frac{1}{2} \left(\frac{1}{1-\alpha} (\sum_i p_i^\alpha - 1) \right) - \frac{1}{2} \left(\left(\frac{1}{1-\alpha} (\sum_i q_i^\alpha - 1) \right) \right)}{\frac{2^{1-\alpha}}{2} \left(\left(\frac{1}{1-\alpha} (\sum_i p_i^\alpha - 1) \right) + \left(\frac{1}{1-\alpha} (\sum_i q_i^\alpha - 1) \right) + \frac{2}{1-\alpha} \right)}$$

Dabei wird die Ähnlichkeit zwischen zwei Korpora p und q berechnet, indem für jeden Wortformtyp i die relative Tokenhäufigkeit p_i mit der relativen Tokenhäufigkeit q_i verglichen wird. Im Prinzip werden hier also prozentuale Wortfrequenzen miteinander verglichen. Für unsere Untersuchung haben wir dazu für alle Spiegelausgaben eines Monats die jeweiligen Worthäufigkeitsverteilungen aggregiert. Darauf aufbauend haben wir für alle insgesamt 322.806 Monatspaare die Ähnlichkeit berechnet (zum Beispiel wird der Januar 2010 mit dem April 1960 verglichen, usw.). Dennoch gilt auch hier mit Verweis auf Tweedie/Baayen (1998), dass

¹² Im Umkehrschluss folgt aus dem Ergebnis von Altmann und Kollegen eigentlich, dass viele veröffentlichte Papiere und Studien in diesem Bereich aufgrund variierender Korpusgrößen überprüft werden müssten.

ausgeschlossen werden muss, dass die gefundenen Unähnlichkeiten systematisch durch die wachsenden Ausgabenlängen verzerrt werden. Um zu testen, ob wir tatsächlich auch Sprachwandel oder nur die unterschiedlichen Ausgabenlängen messen, werfen wir in einem nächsten Schritt alle 361.596 uns vorliegenden im „Spiegel“ veröffentlichten Texte in eine (virtuelle) Urne. Aus dieser Urne ziehen wir dann per Zufall die einzelnen Texte und machen daraus einen Text, der circa 284 Millionen Wörter lang ist. Aus diesem Text schneiden wir anschließend sequenziell „Ausgaben“, die in ihrer Länge genau den originalen Ausgaben entsprechen, und wiederholen die Ähnlichkeitsmessung. Dies stellt dann quasi einen Lackmustrast im Sinne eines Prüfsteins für die Reliabilität der Methode dar: Durch die Zerstörung des diachronen Signals können wir prüfen, ob unsere Ergebnisse durch die variierenden Korpusgrößen systematisch beeinflusst werden.

6 Ergebnisse zum Wortschatzwandel im „Spiegel“

Abbildung 4 visualisiert den Wortschatzwandel für die Spiegeldaten. Das Diagramm auf der linken Seite zeigt dabei, dass im Schnitt die durchschnittliche Unähnlichkeit mit der zeitlichen Distanz (in Monaten) steigt. Ein Spearman-Korrelationskoeffizient in Höhe von $r = 0,79$ bestätigt diesen visuellen Eindruck. Die schwarze Linie stellt hier (und in den folgenden Abbildungen) einen nicht-parametrischen Smoother da, welcher den tendenziellen Zusammenhang in den Daten visualisiert.

Das Diagramm auf der rechten Seite zeigt das Ergebnis des Lackmustrasts. Dieser verdeutlicht auf den ersten Blick, dass der Unterschied tatsächlich auf Sprachwandelprozesse rückführbar zu sein scheint. Zwar ergibt sich auch hier eine positive Spearman-Korrelation, jedoch nur in Höhe von $r = 0,30$. Zusätzlich zeigt der visuelle Vergleich der beiden Diagramme eindeutig einen qualitativen Unterschied für die tatsächlichen Daten (links), verglichen mit den Daten, für die das diachrone Signal zerstört wurde (rechts). Auf Basis dieser Ergebnisse kann man nun einen Schritt weiter gehen und die Forschungsfrage stellen, ob sich hier ein Trend über die Zeit hinweg ausmachen lässt, um so zu prüfen, ob sich die Rate des Wortschatzwandels über die Zeit hinweg verändert oder ob sie konstant bleibt (Nettle 1999). Um diese Frage empirisch evaluieren zu können, kann man jede Monatsausgabe mit der Monatsausgabe aus dem Monat davor vergleichen (Koplenig 2016); beispielsweise vergleicht man die aggregierte Häufigkeitsverteilung aus den vier im Februar 2010 erschienenen Ausgaben mit jenen aus dem Januar 2010. Empirisch lassen sich die berechneten Werte dann als Zeitreihe visualisieren. Abbildung 5 visualisiert das Ergebnis.

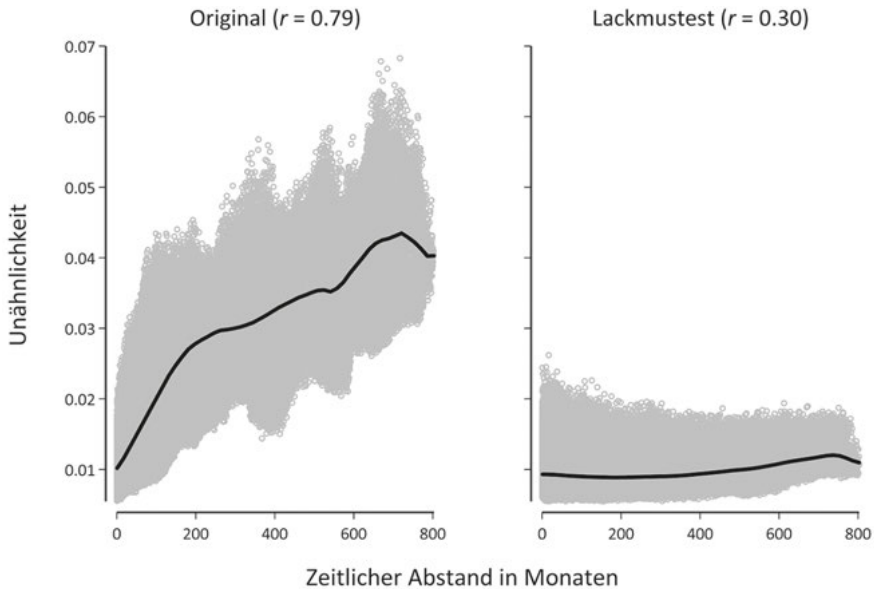


Abb. 4: Unähnlichkeit des Spiegelwortschatzes als Funktion des zeitlichen Abstands in Monaten

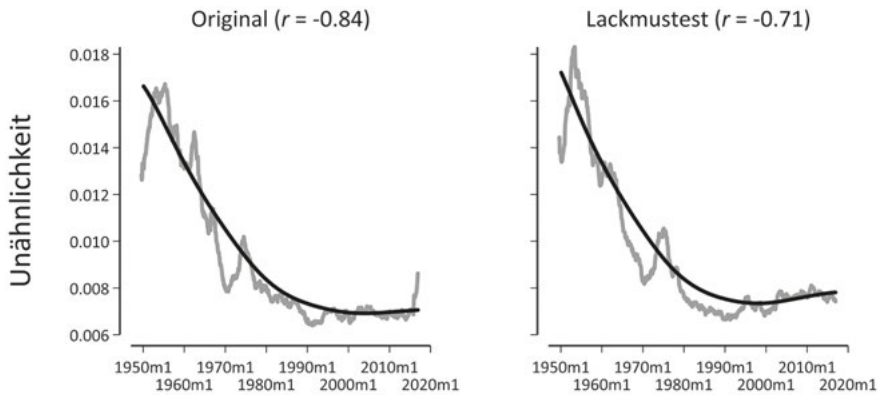


Abb. 5: Die Rate des monatlichen Wortschatzwandels im „Spiegel“ über die Zeit hinweg

Betrachtet man zunächst nur den Verlauf auf der linken Seite, also für die tatsächlichen Daten, so zeigt sich ein vermeintlich spektakuläres Ergebnis. Es scheint sich zu zeigen, dass die Rate des Sprachwandels im „Spiegel“ über die Zeit hinweg abgenommen hat. Gleichzeitig deutet sich auch an, dass dieser Abnahmeprozess über die Zeit hinweg zum Erliegen gekommen ist. Dies würde bedeuten, dass sich

zwei Spiegel-Ausgaben von 1960 und 1961 deutlicher voneinander unterscheiden als zwei Ausgaben von 2009 und 2010. Jedoch zieht der Lackmustest im Diagramm auf der rechten Seite diese Interpretation stark in Zweifel, da sich auch hier ein ähnliches, beinahe analoges Bild ergibt. Dies ist insofern interessant, als wir doch eigentlich im Rahmen des Lackmustests sämtliche diachronen Informationen aus den Daten getilgt haben. Dies trifft jedoch nicht ganz zu, erhalten geblieben ist die über die Zeit hinweg variierende Korpusgröße, welche für die Ergebnisse verantwortlich zu sein scheint. Um dies zu überprüfen, haben wir in Abbildung 6 zusätzlich zu den Wortschatzwandelraten die Ausgabenlängen im zeitlichen Verlauf abgebildet (siehe Abb. 6). Aus Gründen der besseren Illustration haben wir dabei den Verlauf der Ausgabenlänge umgedreht (von hoch zu niedrig).

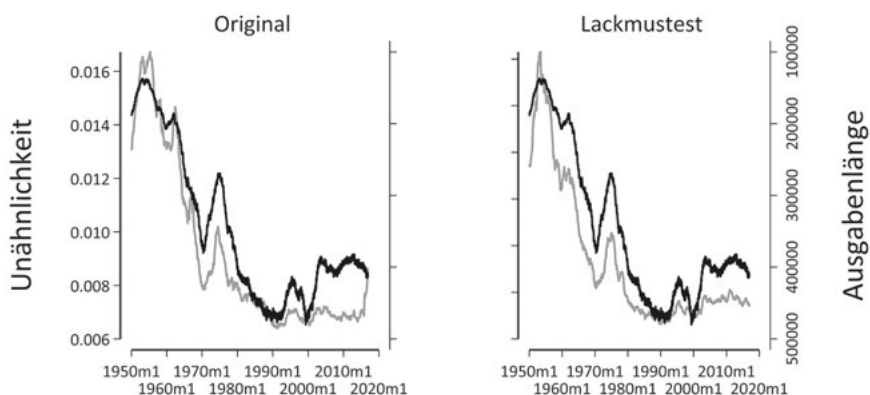


Abb. 6: Wortschatzwandelraten und Korpusgrößen im Zeitverlauf

Man sieht deutliche Ähnlichkeiten zwischen den grauen und den schwarzen Verläufen. Hieraus lässt sich schließen, dass, obschon bei der verwendeten Methode eigentlich nur die relativen Häufigkeiten für die ca. 10.000 häufigsten Typen miteinander verglichen werden, die variierende Korpusgröße die Ergebnisse systematisch und stark beeinflusst. An diesem Punkt stellt sich eine zentrale Frage: Kann man überhaupt Korpora mit unterschiedlichen Größen sinnvoll miteinander vergleichen? Dieser aus quantitativ-empirischer Sicht in höchstem Maße relevanten Frage möchten wir mittelfristig auf den Grund gehen. Zunächst haben wir jedoch versucht, das Problem in diesem Kontext zu lösen, indem wir für jedes Jahr per Zufall aus den Daten mittels eines sogenannten Binomialsplits Zufallsstichproben der Größe 100.000 Wortformtokens erzeugt haben (Piantadosi 2014; Kopenig 2018). Da dieser Ansatz das Problem aber augenscheinlich auch nicht lösen konnte, folgern wir daraus, dass hier eine Verbindung mit der oben angesprochenen text-

lichen Kohärenz bzw. der „Nicht-Stationarität“ von Worthäufigkeitsverteilungen zusammenhängt. Aus diesem Grund haben wir für jeden Monat alle im „Spiegel“ erschienenen Texte in einer zufälligen Reihenfolge angeordnet und hiervon dann die ersten 100.000 Wörter behalten. Somit vergleichen wir für jeden Monat eine Stichprobengröße in Höhe von 100.000 Wörtern unter Beibehaltung der textlichen Kohärenz. Für diesen Datensatz haben wir dann die weiter oben skizzierte Methode wiederholt; Abbildung 7 zeigt das Ergebnis analog zu Abbildung 4.

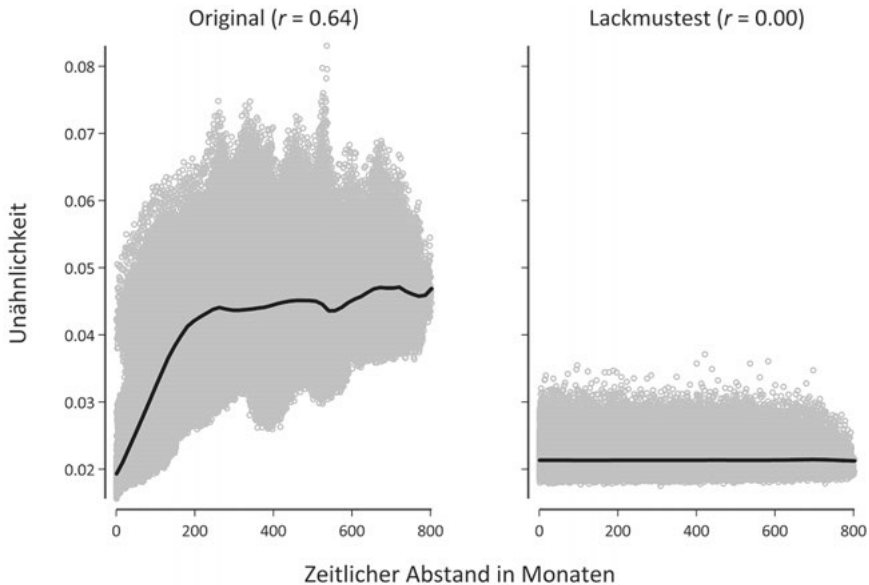


Abb. 7: Unähnlichkeit als Funktion des zeitlichen Abstands in Monaten (Stichproben-Version)

Im Vergleich zu Letzterer fällt besonders auf, dass der Lackmustest nun wirklich überhaupt nicht mehr anschlägt (Korrelationskoeffizient $r = 0,00$); somit können wir zumindest für diese Stichprobe davon ausgehen, dass es sich hier um einen unverzerrten Effekt handelt. Von der Form her deutet sich hier (linkes Diagramm) ein logarithmischer Zusammenhang zwischen der geschätzten Unähnlichkeit und der zeitlichen Distanz an. In der Tat erklärt ein einfaches Regressionsmodell mit der Unähnlichkeit als dem zu erklärenden Merkmal und der logarithmierten Distanz als Prädiktor mehr als 50% der Varianz der Unähnlichkeit. Dies legt den Schluss nahe, dass der Einfluss der zeitlichen Distanz zwischen zwei Ausgaben auf die gemessene Unähnlichkeit mit der Distanz fällt. Inhaltlich bedeutet dies, dass beispielsweise zwischen einer zeitlichen Distanz von 10 und 20 Monaten ein

größerer Unterschied in der Unähnlichkeit besteht als zwischen zum Beispiel 100 und 200 Monaten. Eventuell impliziert dies, dass anfangs noch thematische Unterschiede einen Einfluss auf die gemessene Unähnlichkeit haben, während man – auf lange Sicht – quasi nicht nur den Wandel in der Lexik, sondern auf breiterer Ebene Sprachwandel misst. Dies ist sicherlich ein Bereich, den wir zukünftig noch genauer unter die ‚quantitative‘ Lupe nehmen wollen.

Abbildung 8 visualisiert die Ergebnisse für die Stichprobendaten analog zu Abbildung 5 hinsichtlich der Frage, ob sich für die Spiegeldaten ein Trend über die Zeit hinweg erkennen lässt. Wiederum zeigt unsere Prüfmethode im Diagramm auf der rechten Seite, dass keine zu links vergleichbare Tendenz mehr erkennbar ist. Damit können wir davon ausgehen, dass es sich links auch tatsächlich um den unverzerrten Effekt handelt.

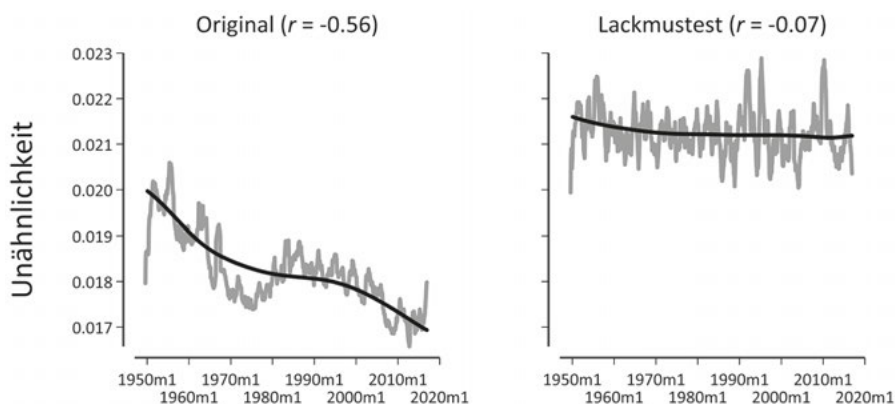


Abb. 8: Die Rate des monatlichen Wortschatzwandels über die Zeit hinweg (Stichproben-Version)

Diesen Eindruck bestätigt auch Abbildung 9 im Vergleich zu Abbildung 6, die variierenden Ausgabenlängen beeinflussen das Ergebnis nun nicht mehr. Empirisch zeigt sich also (rechtes Diagramm der Abb. 9), dass die Unähnlichkeitsrate über die Zeit hinweg abgenommen hat. Hieran ließe sich eine genauere Folgestudie anschließen, welche unter Verbindung von quantitativen Methoden und qualitativen Elementen die Identifizierung der „linguistischen“ Gründe für diese Abnahme zum Ziel haben könnte.

Abschließend möchten wir noch versuchen, die in der Einleitung aufgeworfene Frage zu beantworten, ob es einen Unterschied auf der Mikroebene gibt, d.h., ob eine Ausgabe, die zeitlich einen Monat von der heutigen Ausgabe entfernt ist, ähnlicher als eine Ausgabe ist, welche zwei, drei oder vier Monate entfernt ist.

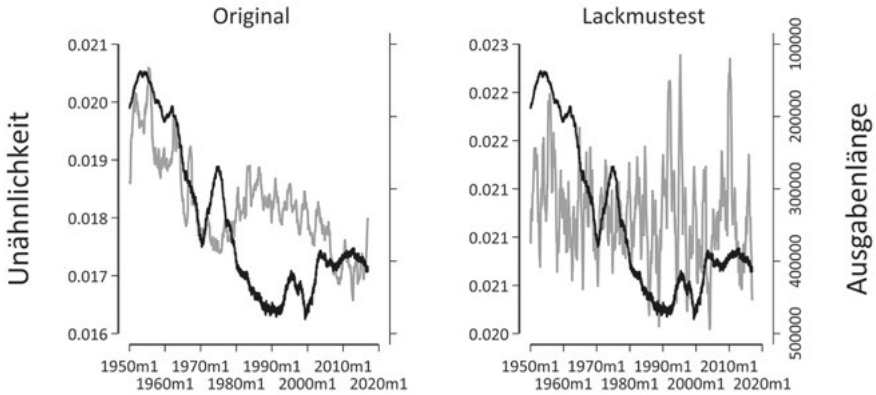


Abb. 9: Wortschatzwandelraten und Korpusgrößen im Zeitverlauf (Stichproben-Version)

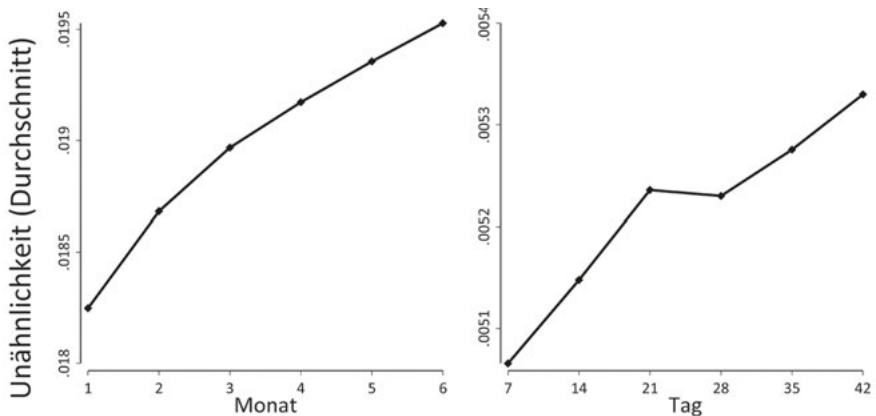


Abb. 10: Wortschatzwandel auf der Mikroebene

Das linke Diagramm in Abbildung 10 zeigt auf der Ordinatenachse die durchschnittliche Unähnlichkeit. Zum Beispiel ist unter „1“ auf der Abszissenachse der Durchschnittswert für alle (in diesem Fall 802) „Spiegel“-Paare mit einem Abstand von einem Monat abgetragen. Es zeigt sich hier auch auf Monatebene ein deutlicher Zusammenhang: Je weiter zwei Ausgaben voneinander entfernt sind, desto unähnlicher sind sie sich. Für das Diagramm auf der rechten Seite sind wir noch einen Schritt weiter gegangen und haben tatsächliche Einzelausgaben miteinander verglichen. Dazu haben wir einen α -Wert in Höhe von 2 angesetzt (d.h., relevant für die Unähnlichkeitsberechnung sind hier die 200 häufigsten Wortform-

typen). Auch hier zeigt sich wieder ein offensichtlicher, wenn auch kein perfekter Zusammenhang, was darauf zurückführbar ist, dass es viel Rauschen in den Daten gibt. Auf der Ebene von einzelnen Ausgaben handelt es sich natürlich um noch feinkörnigere Unterschiede verglichen mit der Ebene von Monaten, illustriert wird dies an den Größenordnungen der Skalen der beiden Ordinatenachsen.

Im Ergebnis lässt sich nun Folgendes schließen: Wenn man heute eine Spiegelausgabe liest und diese mit einer Ausgabe von vor einem Monat vergleicht, so ist der Wortschatz ein wenig ‚anders‘, vergleicht man die heutige Ausgabe mit der Ausgabe von vor zwei Monaten, so unterscheiden sich diese, im Schnitt, noch ein wenig stärker. Dieses Ergebnis zeigt sich auch für einzelne Ausgaben. Nach unserem Kenntnisstand ist dies das erste Mal, dass Sprachwandel auf einem so feinkörnigen Level quantitativ sichtbar gemacht wird. Zwar gilt es in diesem Kontext zu betonen, dass dieses Ergebnis bei der Lektüre qualitativ nicht ‚spürbar‘ ist, dennoch ist es quantitativ auf einem so feinkörnigen Level messbar, was aus unserer Sicht ein durchaus interessantes Ergebnis ist.

7 Fazit

Die Analyse des Spiegelwortschatzes hat aus methodischer Sicht einmal mehr gezeigt, wie weitreichend das Problem unterschiedlich großer Korpora in Verbindung mit der Zipfverteilung von Worthäufigkeiten im Bereich der quantitativen Linguistik tatsächlich ist. Allein durch variierende Ausgabenlängen wurden unsere initialen Ergebnisse deutlich verzerrt, wie unser Lackmustest eindeutig zutage gefördert hat.

Führt man sich ein berühmtes Zitat von William Labov vor Augen, so wird klar, warum das Problem besonders relevant für die diachrone Korpuslinguistik ist: „Historische Dokumente überleben zufällig, nicht absichtlich, und die zur Verfügung stehende Auswahl ist das Produkt einer nicht vorhersagbaren Verkettung historischer Umstände“ (Labov 1994).¹³ Man kann hier auch einen Schritt weiter gehen und argumentieren, dass diachrone Dokumente nicht nur zufällig überleben, sondern dass die ‚Überlebensrate‘ mit der Zeit variiert, oder anders ausgedrückt, je neuer ein Dokument, desto größer die Wahrscheinlichkeit, dass es überlebt.¹⁴ Für die in diesem Beitrag vorgestellte Analyse des Spiegelwortschatzes

¹³ Unsere Übersetzung von: „Historical documents survive by chance, not by design, and the selection that is available is the product of an unpredictable series of historical accidents“.

¹⁴ Vielleicht bedarf es hier eines generellen Umdenkens: In vielen Situationen gibt es einfach keine andere Möglichkeit, als mit den Daten zu arbeiten, die, aus welchen Gründen auch immer,

stellt sich dieses Problem natürlich nicht, da wir hier alle Daten zur Verfügung haben. Wenn man aber an diachrone Studien über größere Zeitspannen denkt, so wird klar, dass man hier, quasi aus prinzipiellen Gründen, weniger Material zur Verfügung hat, je weiter man in die Vergangenheit zurückgeht. Noch problematischer wird es, wenn man zum Beispiel an die Google-Books-Ngram-Korpora denkt (Michel et al. 2010), wo man generell nur aggregierte Worthäufigkeitslisten zur Verfügung hat. Lassen sich hiermit überhaupt sinnvoll Wandelprozesse abschätzen? Die hier vorgestellten Ergebnisse werfen jedenfalls Zweifel auf.

Gleichzeitig gilt, dass das Problem unterschiedlicher Korpusgrößen natürlich nicht nur für die diachrone, sondern auch für die synchrone Korpuslinguistik besteht. So stellt sich hier die Frage, was unsere Ergebnisse ganz generell für den quantitativen Vergleich von Korpora bzw. Texten mit unterschiedlichen Größen bedeutet. Genauso werfen die Ausführungen von Altmann und Gerlach zur notwendigen Größe von Korpora bei der Analyse von sprachlichen Phänomenen im niedrigfrequenten Bereich viele noch zu klärende Fragen auf. Aus diesem Grund glauben wir, dass weitere Forschung in diesem Bereich äußerst wichtig ist.

Gleichzeitig möchten wir Folgendes festhalten: Bei all den Schwierigkeiten, die Sprachdaten bei der quantitativen Analyse bereiten, bieten sie auch besondere Chancen. Man kann sozusagen mit den Daten ‚arbeiten‘, sie manipulieren, indem man sie neu anordnet oder die vorhandene Information maskiert, um gezielt bestimmte Effekte ‚auszuschalten‘ (siehe z.B. Koplenig et al. 2017). Weitere Möglichkeiten sind der in diesem Beitrag vorgestellte Lackmustest sowie das daran anschließende Stichprobenverfahren, welches es uns ermöglicht hat, die Abhängigkeit von der Ausgabenlänge auszuschließen.¹⁵

Abschließend möchten wir Folgendes betonen: Natürlich sind auch wir der Ansicht, dass eine empirische Überprüfung von Erkenntnissen, die auf introspektiven/subjektiven Urteilen beruhen, für die Sprachwissenschaften, sowohl aus methodologischer als auch aus erkenntnistheoretischer Sicht, essenziell ist (Wasow/Arnold 2005; Pullum 2007; Sternefeld/Richter 2012). Gleichzeitig glauben wir aber auch, dass wir hierfür bessere Methoden benötigen, welche den Besonderheiten von Sprachdaten Rechnung tragen (Koplenig 2017a). Deshalb denken wir, dass es wichtig ist, sich in der linguistischen Forschung mehr als bisher mit den Fall-

zur Verfügung stehen. Vielleicht muss man deshalb diachrone Korpuslinguistik mehr als archäologische Linguistik verstehen, um mit geeigneten Methoden etwas Interessantes über die Vergangenheit zu lernen (vgl. auch Koplenig 2016, Kap. 2).

¹⁵ Dies ist jedoch nicht völlig unproblematisch. In gewisser Weise stellt es ja auch eine Form von Sprachwandel dar, dass die Spiegelausgaben über die Zeit hinweg länger geworden sind. Durch die Stichprobenziehung verliert man natürlich auch Information. Auch hier sehen wir einen evidenten Forschungsbedarf.

stricken zu beschäftigen, die die quantitative Analyse von Sprachdaten mit sich bringt. Auch wenn die Implikationen dieser Fallstricke, z.B. die Schwierigkeiten, Korpora unterschiedlicher Größe miteinander zu vergleichen oder Phänomene im niedrigfrequenten Bereich gut messen zu können, unbequem sind, führt – denken wir – kein Weg daran vorbei, dieses Problemfeld methodisch näher zu ergründen. Unsere Analyse des Wortschatzwandels im „Spiegel“ hat deutlich gezeigt, dass man sonst Gefahr läuft, auf den ersten Blick interessante Ergebnisse nicht kritisch genug zu hinterfragen und sie gelten zu lassen, obwohl sie nicht das messen, was impliziert ist. In einem sehr empfehlenswerten Beitrag mit dem passenden Titel „Statistische Annahmen als empirische Verpflichtungen“ zeigen Berk/Freedman (2003),¹⁶ dass die Annahmen, welche statistischen Modellen zugrunde liegen, eben kein ‚Kann‘, sondern ein ‚Muss‘ sind.

Literatur

- Aitchison, Laurence/Corradi, Nicola/Latham, Peter E. (2016): Zipf's Law arises naturally when there are underlying, unobserved variables. In: Sporns, Olaf (Hg.): *PLOS Computational Biology* 12, 12. DOI: 10.1371/journal.pcbi.1005110.
- Allan, Kathryn/Robinson, Justyna A. (Hg.) (2012): *Current methods in historical semantics.* (= *Topics in English Linguistics (TiEL)* 73). Berlin/Boston.
- Altmann, Eduardo G./Cristadoro, Giampaolo/Esposti, Mirko Degli (2012): On the origin of long-range correlations in texts. In: *Proceedings of the National Academy of Sciences* 109, 29, S. 11582–11587. DOI: 10.1073/pnas.1117723109.
- Altmann, Eduardo G./Dias, Laércio/Gerlach, Martin (2017): Generalized entropies and the similarity of texts. In: *Journal of Statistical Mechanics. Theory and Experiment* 2017, 1.
- Baayen, R. Harald (2001): *Word frequency distributions.* Dordrecht.
- Beckett, Sean (2013): *Introduction to time series using Stata.* College Station.
- Behrens, Heike/Pfänder, Stefan (2016): *Experience counts: Frequency effects in language.* (= *Linguae & litterae* 54). Berlin/Boston.
- Berk, Richard A./Freedman, David A. (2003): *Statistical assumptions as empirical commitments.* In: Messinger, Sheldon L./Blomberg, Thomas G./Cohen, Stanley (Hg.): *Law, punishment, and social control. Essays in honor of Sheldon Messinger.* 2. Aufl. New York. Internet: <http://stat.berkeley.edu/~census/berk2.pdf> (Stand: 15.6.2015).
- Blei, David M./Lafferty, John D (2006): *Dynamic topic models.* In: Cohen, William W./Moore, Andrew (Hg.): *Proceedings of the Twenty-Third International Conference on Machine Learning, June 25–29, 2006, Pittsburgh.* New York, S. 113–120.
- Bybee, Joan (2015): *Language change.* Cambridge.
- Clauset, Aaron/Shalizi, Cosma Rohilla/Newman, Mark E.J. (2009): *Power-law distributions in empirical data.* In: *SIAM Review* 51, 4, S. 661–703. DOI: 10.1137/070710111.

16 Unsere Übersetzung von: „Statistical assumptions as empirical commitments“.

- Davies, Mark (2012): Expanding horizons in historical linguistics with the 400-million word corpus of historical American English. In: *Corpora* 7, 2, S. 121–157. DOI: 10.3366/cor.2012.0024.
- Deerwester, Scott et al. (1990): Indexing by latent semantic analysis. In: *Journal of the American Society for Information Science* 41, 6, S. 391–407.
- Dias, Laércio et al. (2018): Using text analysis to quantify the similarity and evolution of scientific disciplines. In: *Royal Society Open Science* 5. DOI: 10.1098/rsos.171545.
- Ellis, Nick C. (2012): What can we count in language, and what counts in language acquisition, cognition, and use? In: Gries, Stefan Th./Divjak, Dagmar (Hg.): *Frequency effects in language learning and processing*. Bd. 1. (= *Trends in Linguistics. Studies and Monographs (TiLSM)* 244.1). Berlin/Boston, S. 7–34.
- Engelberg, Stefan (2015): Quantitative Verteilungen im Wortschatz. Zu lexikologischen und lexikografischen Aspekten eines dynamischen Lexikons. In: Eichinger, Ludwig M. (Hg.): *Sprachwissenschaft im Fokus. Positionsbestimmungen und Perspektiven*. (= *Jahrbuch des Instituts für Deutsche Sprache* 2014). Berlin u.a., S. 205–230. Internet: <https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/3481> (Stand: 1.6.2017).
- Engelberg, Stefan/Lobin, Henning/Steyer, Kathrin (Hg.) (2018): *Wortschätze. Dynamik, Muster, Komplexität*. (= *Jahrbuch des Instituts für Deutsche Sprache* 2017). Berlin/Boston.
- Gerlach, Martin/Font-Clos, Francesc/Altmann, Eduardo G. (2016): Similarity of symbol frequency distributions with heavy tails. In: *Physical Review X* 6, 2. DOI: 10.1103/PhysRevX.6.021009. Internet: <https://link.aps.org/doi/10.1103/PhysRevX.6.021009> (Stand: 8.1.2018).
- Hilpert, Martin/Perek, Florent (2015): Meaning change in a petri dish: Constructions, semantic vector spaces, and motion charts. In: *Linguistics Vanguard* 1, 1, S. 339–350.
- Kämper, Heidrun (2013): *Wörterbuch zum Demokratiediskurs 1967/68*. Berlin/Boston: DOI: 10.1524/9783050065113. Internet: <https://degruyter.com/view/product/219708> (Stand: 2.5.2018).
- Kerremans, Daphné/Stegmayr, Susanne/Schmid, Hans-Jörg (2012): The NeoCrawler. Identifying and retrieving neologisms from the internet and monitoring ongoing change. In: Allan/Robinson (Hg.), S. 59–96.
- Kilgarriff, Adam (2001): Comparing corpora. In: *International Journal of Corpus Linguistics* 6, 1, S. 97–133. DOI: 10.1075/ijcl.6.1.05kil.
- Koplenig, Alexander (2015): The impact of lacking metadata for the measurement of cultural and linguistic change using the Google Ngram data sets. Reconstructing the composition of the German corpus in times of WWII. In: *Digital Scholarship in the Humanities* 32, 1, S. 169–188. DOI: 10.1093/lc/fqv037.
- Koplenig, Alexander (2016): *Analyzing lexical change in diachronic corpora*. Diss. Mannheim. Internet: <https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/4890> (Stand: 19.9.2018).
- Koplenig, Alexander (2017a): Against statistical significance testing in corpus linguistics. In: *Corpus Linguistics and Linguistic Theory* 0/0. DOI: 10.1515/cllt-2016-0036. Internet: <https://degruyter.com/view/j/cllt.ahead-of-print/cllt-2016-0036/cllt-2016-0036.xml?format=INT> (Stand: 8.1.2018).
- Koplenig, Alexander (2017b): A data-driven method to identify (correlated) changes in chronological corpora. In: *Journal of Quantitative Linguistics* 24, 4, S. 289–318. DOI: 10.1080/09296174.2017.1311447.
- Koplenig, Alexander (2018): Using the parameters of the Zipf–Mandelbrot law to measure diachronic lexical, syntactical and stylistic changes. A large-scale corpus analysis. In: *Corpus Linguistics and Linguistic Theory* 14, 1, S. 1–34. DOI: 10.1515/cllt-2014-0049.

- Kopenig, Alexander et al. (2017): The statistical trade-off between word order and word structure. Large-scale evidence for the principle of least effort. In: *Plos One* 12, 3. DOI: 10.1371/journal.pone.0173614.
- Labov, William (1994): *Principles of linguistic change*. (= *Language in Society* 20). Oxford/Cambridge, MA.
- Lestrade, Sander (2017): Unzipping Zipf's law. In: *Plos One* 12, 8. DOI: 10.1371/journal.pone.0181987.
- Lijffijt, Jeffrey et al. (2014): Significance testing of word frequencies in corpora. In: *Digital Scholarship in the Humanities* 31, 2, S. 374–397. DOI: 10.1093/lc/fqu064. Internet: <http://dsh.oxfordjournals.org/cgi/doi/10.1093/lc/fqu064> (Stand: 22.4.2015).
- Manning, Christopher D./Raghavan, Prabhakar/Schütze, Hinrich (2008): *Introduction to information retrieval*. New York.
- Michel, Jean-Baptiste et al. (2010): Quantitative analysis of culture using millions of digitized books. Supporting Online Material II. In: *Science* 331, 14. Internet: <http://sciencemag.org/content/331/6014/176/suppl/DC1> (Stand: 5.4.2014).
- Müller-Spitzer, Carolin/Wolfer, Sascha/Kopenig, Alexander (2018): Quantitative Analyse lexikalischer Daten. Methodenreflexion am Beispiel von Wandel und Sequenzialität. In: Engelberg/Lobin/Steyer (Hg.), S. 245–266.
- Nettle, Daniel (1999): Is the rate of linguistic change constant? In: *Lingua* 108, 2–3, S. 119–136. DOI: 10.1016/S0024-3841(98)00047-3.
- Perek, Florent (2016): Using distributional semantics to study syntactic productivity in diachrony. A case study. In: *Linguistics* 54, 1, S. 149–188.
- Piantadosi, Steven T. (2014): Zipf's word frequency law in natural language. A critical review and future directions. In: *Psychonomic Bulletin & Review* 21, 5, S. 1112–1130. DOI: 10.3758/s13423-014-0585-6. Internet: <http://link.springer.com/10.3758/s13423-014-0585-6> (Stand: 2.5.2014).
- Pullum, Geoffrey K. (2007): Ungrammaticality, rarity, and corpus use. In: *Corpus Linguistics and Linguistic Theory* 3, S. 33–47.
- Schmid, Hans-Jörg (2018): Ein integratives soziokognitives Modell des dynamischen Lexikons. In: Engelberg/Lobin/Steyer (Hg.), S. 215–232.
- Sternefeld, Wolfgang/Richter, Frank (2012): Wo stehen wir in der Grammatiktheorie? Bemerkungen anlässlich eines Buchs von Stefan Müller. In: *Zeitschrift für Sprachwissenschaft* 31, 2, S. 263–291. DOI: 10.1515/zfs-2012-0010. Internet: <https://degruyter.com/view/j/zfs-2012-31.issue-2/zfs-2012-0010/zfs-2012-0010.xml> (Stand: 19.4.2018).
- Szmrecsanyi, Benedikt (2016): About text frequencies in historical linguistics. Disentangling environmental and grammatical change. In: *Corpus Linguistics and Linguistic Theory* 12, 1, S. 153–171. DOI: 10.1515/cllt-2015-0068. Internet: <http://degruyter.com/view/j/cllt.2016.12.issue-1/cllt-2015-0068/cllt-2015-0068.xml> (Stand: 6.6.2016).
- Tweedie, Fiona J./Baayen, Harald R. (1998): How variable may a constant be? Measures of lexical richness in perspective. In: *Computers and the Humanities* 32, 5, S. 323–352.
- Wasow, Thomas/Arnold, Jennifer (2005): Intuitions in linguistic argumentation. In: *Lingua* 114, 11, S. 1481–1496.
- Würschinger, Quirin et al. (2016): Using the web and social media as corpora for monitoring the spread of neologisms. The case of *rapefugee*, *rapeugee*, and *rapugee*. In: *Proceedings of the 10th Web as Corpus Workshop (WAC-X)*. Annual meeting of the Association for Computational Linguistics (ACL). Berlin, S. 35–43. Internet: <http://aclweb.org/anthology/W16-2605> (Stand: 19.9.2018).