

Marc Kupietz/Cyril Belica/Harald Längen/Rainer Perkuhn
(Mannheim)

Zwischen Empirie und Ästhetik – Ansätze zur korpuslinguistischen Untersuchung und Bewertung von Sprachwandel

Abstract

Der Beitrag beschäftigt sich mit der Frage, wie und inwieweit korpusbasierte Ansätze zur Untersuchung und Bewertung von Sprachwandel beitragen können. Die Bewertung von Sprachwandel erscheint in dieser Hinsicht interessant, da sie erstens von größerem öffentlichen Interesse ist, zweitens nicht zu den Kernthemen der Sprachwissenschaft zählt und drittens sowohl die geisteswissenschaftlichen Aspekte der Sprachwissenschaft berührt als auch die empirischen, die eher für die so genannten harten Wissenschaften typisch sind. Letzteres trifft bei der Frage nach Sprachverfall (gutem vs. schlechtem Deutsch diachron) vermutlich unbestrittener zu als bei der Frage nach richtigem vs. falschem Deutsch, da zu ihrer Beantwortung offensichtlich einerseits empirische, messbare Kriterien herangezogen werden müssen, andererseits aber auch weitere Kriterien notwendig sind und es außerdem einer Entscheidung zur Einordnung und Gewichtung der verschiedenartigen Kriterien sowie einer Begründung dieser Entscheidung bedarf.

Zur Annäherung an die Fragestellung werden zunächst gängige, leicht operationalisierbare Hypothesen zu Symptomen eines potenziellen Verfalls des Deutschen auf verschiedenen DEREKO-basierten Korpora überprüft und im Hinblick auf ihre Verallgemeinerbarkeit und Tragweite diskutiert. Im zweiten Teil werden weitere empirische Ansätze zur Untersuchung von Wandel, Variation und Dynamik skizziert, die zur Diskussion spezieller Aspekte von Sprachverfall beitragen könnten. Im Schlussteil werden die vorgestellten Ansätze in den Gesamtkontext einer sprachwissenschaftlichen Untersuchung von Sprachverfall gestellt und vor dem Hintergrund seines gesellschaftlichen Diskurses reflektiert.

1. Einleitung

Die Frage nach Sprachverfall berührt, obgleich sie kein Kernthema der Sprachwissenschaft ist, ein recht breites Spektrum von Aspekten der Linguistik zwischen Empirie und Ästhetik: Einerseits erfordern belastbare Antworten auf die Frage fundierte empirische Belege, andererseits können auch ästhetische Kriterien bei der Beantwortung eine Rolle spielen, die – wenn überhaupt – nur schwer empirisch operationalisierbar sind. Außerdem sind Entscheidungen zur Einordnung und Gewichtung der verschiedenartigen Kriterien und Begründungen dieser Entscheidungen für eine sprachwissenschaftliche Beantwortung notwendig.

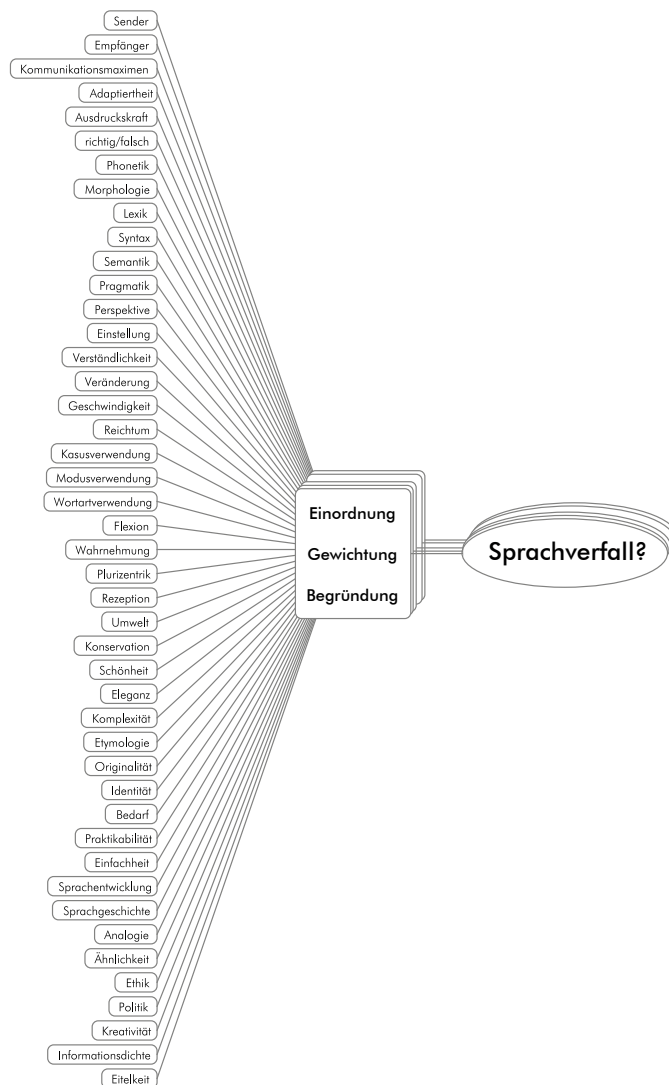


Abb. 1: Bestandteile der Untersuchung von Sprachverfall:

- (1.) Bereiche von Einflussfaktoren (unvollständig und unsortiert),
- (2.) Gewichtungen,
- (3.) verschiedene Konzepte von Sprachverfall

In Abbildung 1 haben wir versucht, die potenziellen methodologischen Bestandteile einer Untersuchung und Bewertung von Sprachwandel zu skizzieren. In der linken Spalte ist eine Reihe von möglichen Bereichen von Kriterien oder Einflussfaktoren, unsortiert und sicherlich unvollständig, aufgezählt. Wie in der mittleren Spalte dargestellt, müssen diese Einflussfaktoren

sortiert, eingeordnet und gewichtet werden. Wie dies zu erfolgen hat, kann dabei nicht aus empirischen Beobachtungen abgeleitet werden. Es müssen vielmehr Entscheidungen getroffen und begründet werden. Eine für diese Fragestellung typische Schwierigkeit besteht dabei darin, dass die potenziellen Einflussbereiche sehr heterogen und z.T. inkommensurabel sind. So ist es z.B. unklar, wie sich Faktoren aus den Bereichen Ökonomie, Verständlichkeit und Kasusverwendung vergleichbar messen oder miteinander verknüpfen lassen.

Typisch für die Fragestellung ist außerdem, dass sicherlich viele unterschiedliche Konzepte von Sprachverfall (rechts in Abb. 1) existieren können, die unterschiedliche Gewichtungen unterschiedlicher Einflussfaktoren mit jeweils unterschiedlichen Begründungen erfordern.

Ausgehend von der Diskussion in Medien und Populärliteratur und der Beobachtung, dass der Sprachverfall-Diskurs offensichtlich eine gesellschaftliche Relevanz hat, sollen im Folgenden zunächst einige der einer empirischen Untersuchung zugänglichen potenziellen Einflussfaktoren näher untersucht werden und später auch Ansätze zur Untersuchung der Wahrnehmung von Sprachverfall, Sprachverfall als Phänomen der Textrezeption und zur Gewichtung von Einflussfaktoren punktuell vorgestellt werden.

2. Kurzstudien zu Hypothesen der Manifestation von Sprachverfall

Von den zahlreichen Hypothesen dazu, worin sich ein Verfall manifestiert bzw. manifestieren könnte, haben wir die folgenden sechs zur näheren Untersuchung ausgewählt: zunehmende Verwendung von (1.) Nomina, (2.) schwachen Präteritumsformen bei starken Verben, (3.) Adverbien als Adjektive, (4.) „wegen“ mit Dativ, (5.) analytischem Konjunktiv und (6.) Anglizismen.

Unsere Auswahlkriterien waren dabei unsere ungefähre Einschätzung der Relevanz im aktuellen Diskurs und eine möglichst leichte Operationalisierbarkeit. Als Inspirationsquelle haben wir vor allem Sick (2004, 2005) verwendet.

2.1 Sprachausschnitt und Datengrundlage

Als zu untersuchenden Sprachausschnitt haben wir Zeitungen aus Deutschland, Österreich und der Schweiz (aufgeschlüsselt nach Ressort: Kultur, Lokales, Politik, Sport, Wirtschaft, NA/sonstige) sowie Plenarprotokolle aus dem Deutschen Bundestag und den Landtagen jeweils aus dem Zeitraum 1995 bis 2012 gewählt. Als Korpusgrundlage hat uns dabei das DEREKO-Release 2013-I (IDS 2013; Kupietz et al. 2010) mit den durch Connexor MPT erstellten Wortartannotationen gedient (siehe Belica et al.

2011). Die auf den Wortartannotationen beruhenden Ergebnisse haben wir, wenn möglich, anhand von RFTagger-Annotationen (Schmid/Laws 2008) ansatzweise auf Plausibilität überprüft.

Der Grund für den relativ kleinen Sprachausschnitt und das relativ kurze Zeitintervall war vor allem, dass für diese Partition in DEREKO ausreichende und homogene Daten vorliegen. Außerdem sollten Überschneidungen mit dem Akademie-Projekt zur Lage der Deutschen Sprache (vgl. Eichinger in diesem Band) vermieden werden. In jedem Fall sollte der gewählte Sprachausschnitt weitergehende Rückschlüsse zulassen, da davon auszugehen ist, dass sich gravierende Änderungen im Sprachgebrauch auch in Zeitungen und Parlamentsdebatten manifestieren. Hinsichtlich des Akademie-Projekts war der Zeitraum bis 2012 mit hoher Auflösung (ein Jahr vs. eine Dekade) auch besonders interessant, um etwa zu überprüfen, inwieweit sich dort beobachtete Tendenzen fortsetzen oder nicht.

Bezüglich der Plenarprotokolle ist darauf hinzuweisen, dass in diesen potenziell für die Fragestellung relevante Faktoren bzw. Verunreinigungen miteinander vermischt werden: gesprochene Sprache, Politikersprache, stenographische Aufzeichnung, nachträgliche Korrekturen bzw. Veränderungen, Aufbereitung teilweise mittels optischer Zeichenerkennung (OCR) mit nachträglichen Korrekturen.

Bezüglich der österreichischen und schweizerischen Zeitungskorpora ist darauf hinzuweisen, dass diese relativ wenige verschiedene Zeitungen enthalten, die zudem über die Zeit variieren.

2.2 Interpretation der Ergebnisse

Die Ergebnisse der Kurzstudien sind in erster Linie anhand der im Folgenden dargestellten Frequenzverläufe hinsichtlich möglicherweise vorhandener Trends zu interpretieren. Zur Einschätzung der Zuverlässigkeit der beobachteten relativen Häufigkeiten sind an den Datenpunkten jeweils Konfidenzintervalle (95%-Niveau) dargestellt. Diese geben den Bereich an, in dem mit 95%-iger Wahrscheinlichkeit die tatsächlichen Werte liegen, wenn man das untersuchte Korpus als Stichprobe seiner zugrundeliegenden Sprachdomäne betrachtet. Um die Erkennung von Trends zu erleichtern, sind zu den Frequenzverläufen jeweils auch mithilfe der Standardmethode LOESS („Local regrESSion“) geglättete Kurven in grau dargestellt.

Als weiteren Indikator für die Erkennung von Trends haben wir ergänzend Kendalls Rangkorrelationskoeffizient τ (siehe Hilpert/Griess 2009) verwendet. τ gibt im Wesentlichen an, wie gut ein Frequenzverlauf einer monoton steigenden ($\tau > 0$) bzw. fallenden ($\tau < 0$) Funktion entspricht. Außerdem kann für τ die Wahrscheinlichkeit p eines irrtümlichen Verwerfens der Null-Hypothese ($\tau = 0$) ermittelt werden. Wir werden im Folgenden nur τ -Werte für $p < 0,05$ betrachten. Da τ weniger Information enthält, als typischer-

weise anhand der Zeitverläufe unmittelbar ersichtlich ist, und keine Aussage über die Größenordnung der Änderung macht, sollte τ nicht überbewertet werden (siehe Hilpert/Griess 2009).

2.3 Kurzstudien

2.3.1 Zunehmende Verwendung von Nomina

Hypothese

Der Anteil von Nomina steigt gegenüber dem anderer Wortarten (vgl. Sick 2004, S. 87 ff.).

Operationalisierung

Grundlage der Operationalisierung waren die Wortartannotationen von Connexor MPT. Gemessen haben wir die relative Gesamthäufigkeit des POS-Tags „N“ ohne Eigennamen („Prop“).

Ergebnisse

Wie Abbildung 2 zeigt, weist der Nomenanteil in den Zeitungskorpora insgesamt einen leichten Aufwärtstrend auf ($\tau = 0,48$), wobei in den österreichischen Zeitungen eine deutlichere Zunahme zu verzeichnen ist ($\tau = 0,52$), in den Schweizer Zeitungen jedoch kein Trend erkennbar ist. Wie Abbildung 3 zeigt, ist der Nomenanteil in den Zeitungen stark vom Ressort abhängig.

Der Trend des Nomenanteils in den Plenarprotokollen ist rückläufig ($\tau = -0,87$).

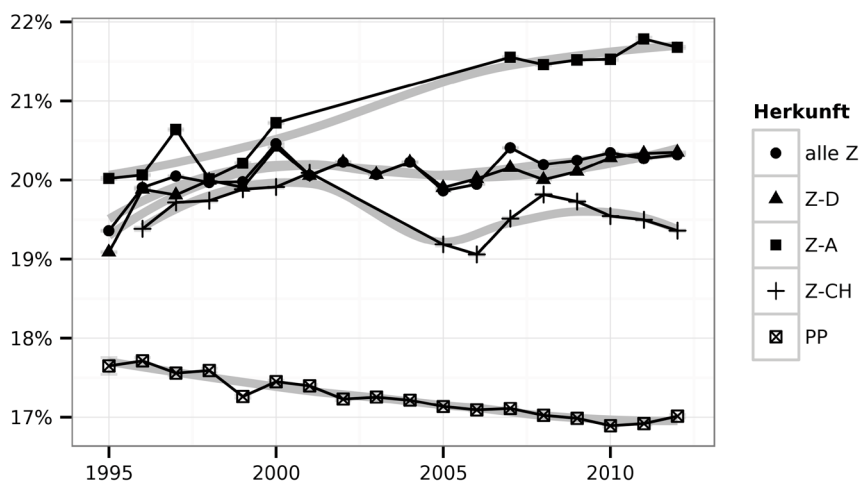


Abb. 2: Häufigkeit von Nomina (ohne Eigennamen)

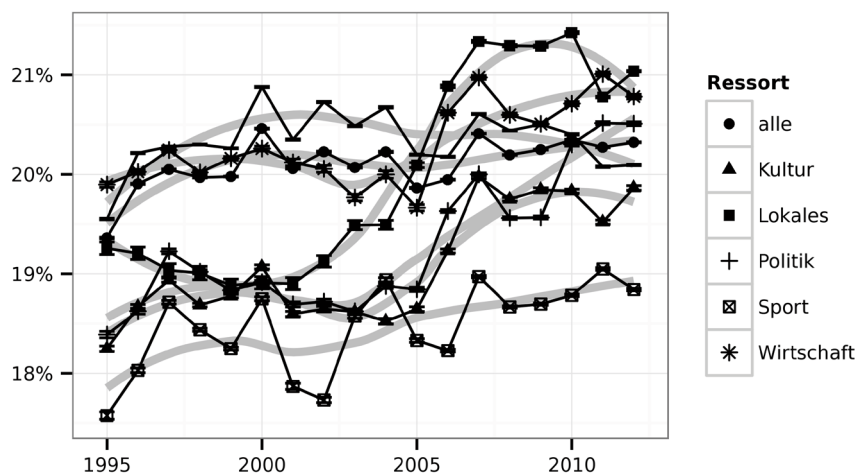


Abb. 3: Häufigkeit von Nomina nach Ressort

2.3.2 Schwache Präteritumsform bei starken Verben

Hypothese

Bei stark flektierbaren Verben wird zunehmend eine schwache Präteritumsform verwendet.

Operationalisierung

Grundlage der Untersuchung war folgende Auswahl stark flektierbarer Verben: backen, befehlen, braten, dringen, erleichen, fechten, flechten, frieren, gießen, gleichen, gleiten, heben, klingen, leiden, meiden, melken, raten, riechen, scheinen, schießen, schleichen, schleifen, schneiden, schreien, singen, sprießen, stinken, streichen, vermeiden, werben.

Gemessen haben wir den Anteil schwach flektierter Varianten an allen Präteritumsformen dieser Verben.

Ergebnisse

Wie Abbildung 4 zeigt, sind bzgl. des Anteils schwach flektierter Varianten stark flektierbarer Verben über den untersuchten Zeitraum keine Trends zu beobachten. Lediglich in den schweizerischen und österreichischen Zeitungen könnte in den letzten Jahren möglicherweise ein Aufwärtstrend begonnen haben. Mit einem jeweiligen durchschnittlichen Anteil von unter 0,5% bei den Zeitungen und nahezu 0% bei den Plenarprotokollen ist das Phänomen insgesamt sehr niedrigfrequent.

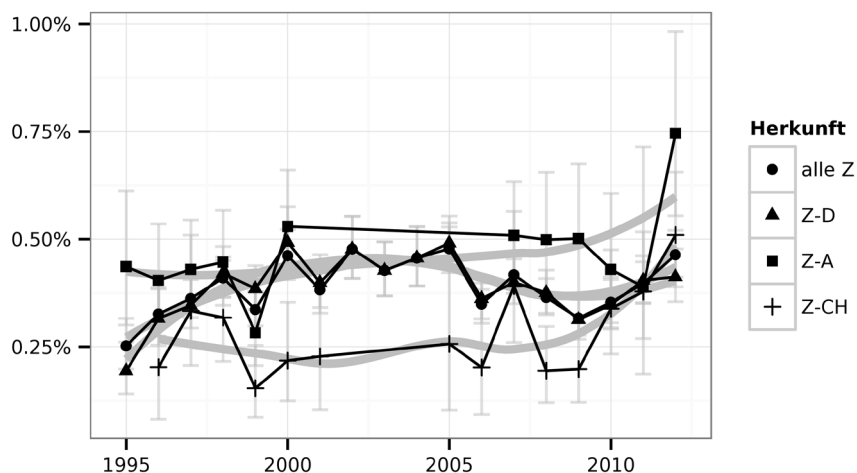


Abb. 4: Anteil der schwach flektierten Varianten

2.3.3 Adverbien auf „-weise“ als Adjektiv

Hypothese

Adverbien auf „-weise“ werden zunehmend auch attributiv wie Adjektive verwendet (vgl. Sick 2004, S. 110 ff.).

Operationalisierung

Die naheliegende Operationalisierung anhand von Wortart-Annotationen war in diesem Fall nicht möglich, da sowohl Connexor MPT als auch der RFTagger (mit der von uns verwendeten Parameter-Datei) Formen auch bei adverbialer Verwendung häufig als Adjektive klassifizierten. Daher haben wir stattdessen eine halbautomatisch gewonnene Liste von Adverbien auf „-weise“ als Grundlage genommen (z.B. andeutungsweise, ansatzweise, beispielsweise, bekannterweise, dummerweise, ersatzweise, haufenweise, schrittweise, stellenweise, stufenweise, teilweise, vergleichsweise, zwangsweise). Gemessen haben wir den Anteil der flektierten Formen (Wörter aus der Liste mit den Endungen m, n, r, s) am Vorkommen aller Formen der ausgewählten Wörter.

Ergebnisse und Diskussion

Abbildung 5 zeigt, dass der Trend sowohl in den Zeitungs- als auch bei den Plenarprotokoll-Korpora leicht rückläufig ist ($\tau = -0,48$ bzw. $\tau = -0,57$), wobei er bei schweizerischen Zeitungen etwas deutlicher ausfällt ($\tau = -0,74$).

Insgesamt ist also eher ein Rückgang des Phänomens zu beobachten. Zur Aufklärung der unflektierten adjektivischen Verwendungen wäre allerdings eine elaboriertere Operationalisierung, z.B. durch eine manuelle Annotation einer Stichprobe, notwendig.

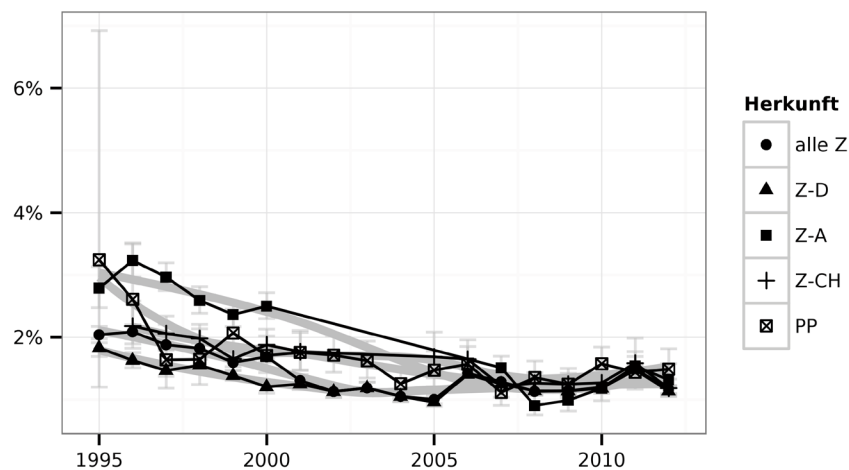


Abb. 5: Anteil flektierter Varianten einer Liste von Adverbien auf „-weise“

2.3.4 Analytische Konjunktivbildung

Hypothese

Der Konjunktiv wird immer häufiger analytisch gebildet (vgl. Sick 2005, S. 77 ff.).

Operationalisierung

Als Grundlage konnten wir hier die morphosyntaktischen Annotationen von Connexor MPT verwenden. Gemessen haben wir den Anteil der Formen von „würde“ an allen Verben im Konjunktiv.

Ergebnisse und Diskussion

Wie Abbildung 6 zeigt, ist der Anteil analytischer Konjunktivmarkierungen recht konstant. Lediglich bei den Plenarprotokollen ist ab 1999 ein leichter Aufwärtstrend zu beobachten ($\tau = 0,5$ für den gesamten beobachteten Zeitraum). Der Anteil in den österreichischen Zeitungen und insbesondere in den Plenarprotokollen ist insgesamt höher als in den deutschen Zeitungen.

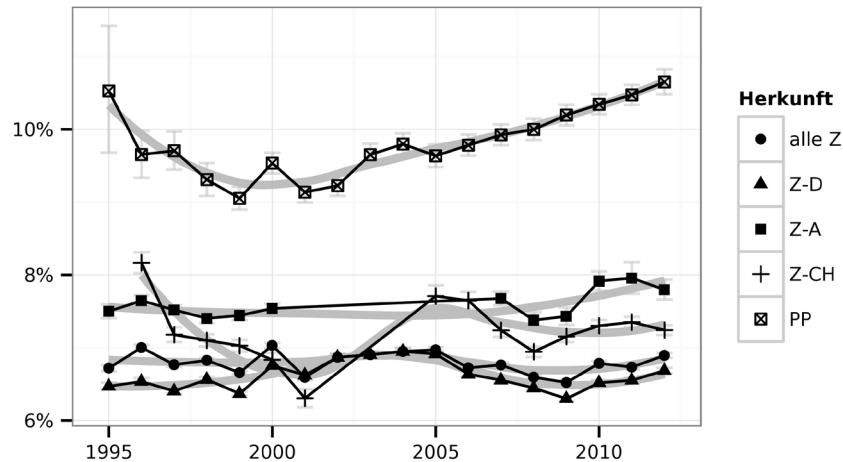


Abb. 6: Anteil analytischer Konjunktivmarkierungen

2.3.5 „Wegen“ mit Dativ

Hypothese

Die Präposition „wegen“ wird immer häufiger mit Dativ verwendet (vgl. Sick 2004, S. 15 ff.).

Operationalisierung

Auch hier konnten weder Connexor MPT noch RFTagger verwendet werden, da beide bei ambigem Kasus, also z.B. bei Fem.-Sg.-Ergänzungen, nur einen möglichen Kasus-Wert (Genitiv oder Dativ) und keine Information bzgl. der Ambiguität lieferten. Daher haben wir als Basis der Operationalisierung Bigramm-Frequenzlisten mit jeweils „wegen“ als erstem Element verwendet, die wir halbautomatisch in die drei Kategorien i) eindeutig Dativ, ii) eindeutig Genitiv und iii) ambig klassifiziert haben. Gemessen haben wir den Anteil von „wegen“+Dativ an „wegen“-Gesamt.

Ergebnisse und Diskussion

Der Anteil ambiger Fälle mit unklarem Kasus lag zwischen etwa 70% und 80%, so dass je nach Bedingung nur etwa 20-30% der Treffer weiter untersucht werden konnten. Die in Abbildung 7 dargestellten Ergebnisse zeigen starke Veränderungen des Dativ-Anteils bei den österreichischen und schweizerischen Zeitungen mit allerdings schon in 2005 bzw. 2009 überschrittenen Höhepunkten. In den deutschen Zeitungen ist der Anteil recht konstant bei

etwa 0,2% evtl. mit leicht fallender Tendenz ($\tau = -0,41$, $p < 0,05$). Die Spitze in 2005 ist nach Sichtung der einzelnen Belege auf Sätze wie „Heißt es nun wegen dem oder wegen des schlechten Wetters?“¹ in Buchbesprechungen von Bastian Sicks „Der Dativ ist dem Genitiv sein Tod“ zurückzuführen. Dies deutet darauf hin, dass es bei dieser Frage besonders wichtig wäre, genauer zwischen echtem Sprachgebrauch und Erwähnung in Zitaten, Titeln, Metasprache zu unterscheiden (vgl. Perkuhn/Keibel/Kupietz 2012, S. 15).

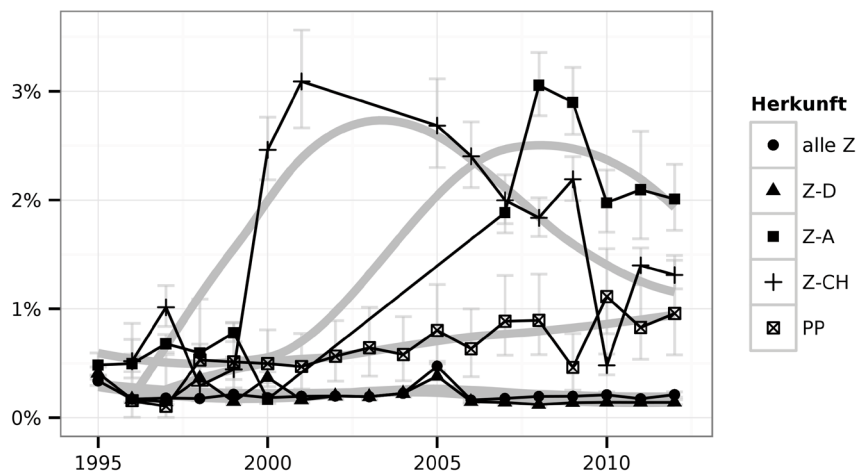


Abb. 7: Anteil der Verwendung von „wegen“ mit Dativ an der Gesamt-Verwendung von „wegen“

2.3.6 Anglizismen

Hypothese

Es werden zunehmend Anglizismen verwendet (vgl. Sick 2004, S. 80 ff.).

Operationalisierung

Da es, falls überhaupt, nur mit sehr hohem Aufwand möglich gewesen wäre, alle Anglizismen in unserem Korpus zu erfassen, haben wir uns in diesem Fall für eine besonders einfache Operationalisierung entschieden und eine recht kurze, aber dafür recht bekannte Liste von Anglizismen gewählt, nämlich die so genannte Ramsauer-Liste verbotener Wörter.² Aus dieser Liste haben wir alle Anglizismen, die aus einem Wort (bzw. mehreren mit Bindestrich verbundenen) bestehen, ausgewählt (z.B. Backup, Beamer, Benchmark, Berechnungs-Tools, VIP-Lunch, Voting, Whiteboard, Workshop) und ihre relative Häufigkeit (einschließlich flektierter Formen) gemessen.

¹ BZ, 10.1.2005, S. 27 nach DEREKO-2013-I:B05/JAN.02372.

² Online-Ausgabe der Rhein-Zeitung vom 29.12.2010.

Ergebnisse und Diskussion

Die in Abbildung 8 dargestellten Ergebnisse zeigen durchweg einen Anstieg der Verwendung der untersuchten Anglizismen – in den Zeitungskorpora auf etwa das Doppelte des Anteils von 1995 ($\tau = 0,91$). In den österreichischen Zeitungen wurde der Höhepunkt anscheinend bereits 2007 erreicht. In den Plenarprotokollen ist eine deutlich niedrigere Häufigkeit und ein langsamerer Anstieg zu verzeichnen.

Die Frage, ob der Anteil der Anglizismen insgesamt oder nur der der Ramsauer-Liste steigt, bleibt allerdings aufgrund der vereinfachten Fragestellung und Operationalisierung offen.

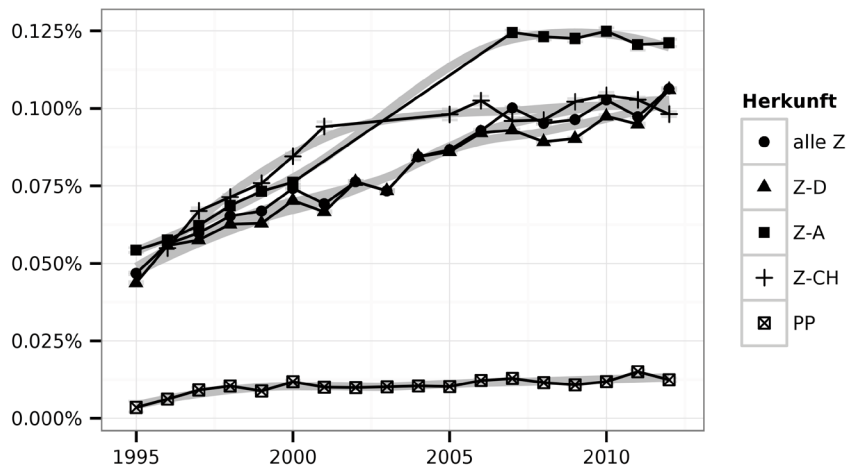


Abb. 8: Häufigkeit der untersuchten Anglizismen aus der Ramsauer-Liste

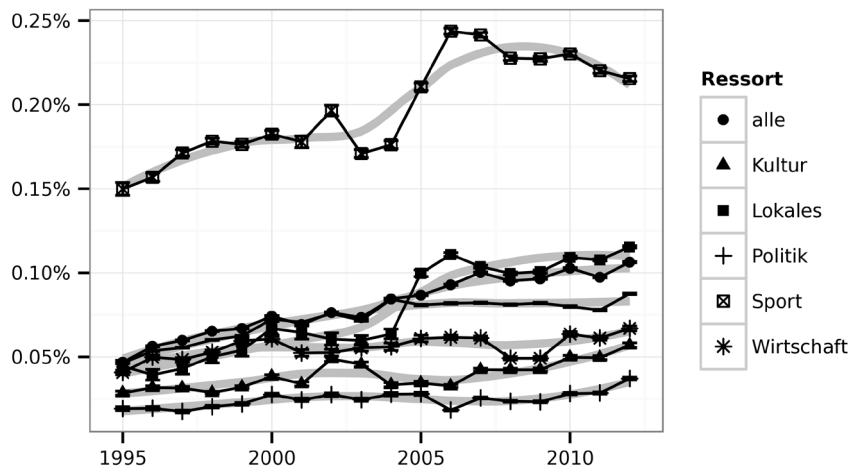


Abb. 9: Häufigkeit der untersuchten Anglizismen aus der Ramsauer-Liste nach Ressort

2.4 Zusammenfassung und Diskussion

Wie Tabelle 1 zeigt, ergibt sich bezüglich der untersuchten Hypothesen mit fünf ansteigenden Tendenzen, drei rückläufigen Tendenzen und vier annähernd konstanten Verläufen ein gemischtes Bild. Das heißt, selbst wenn man annimmt, dass die untersuchten Phänomene für die Frage nach einem Verfall der deutschen Sprache relevant sind, lassen unsere Untersuchungen keine sehr eindeutigen Schlüsse zu.

Hypothese	Zeitungen	Plenarprotokolle
1. Nomen	+	–
2. Schwache Flexion starker Verben		
3. Adverbien als Adjektive	–	–
4. „wegen“ mit Dativ		+
5. Analytischer Konjunktiv		+
6. Anglizismen	+	+

Tab. 1: Zusammenfassung der Ergebnisse der Kurzstudien.

+ = steigende Tendenz, – = fallende Tendenz

Die meisten der vorgestellten Untersuchungen müssten noch elaborierter operationalisiert und weiter verfolgt werden, um gesichertere Schlussfolgerungen zu erlauben. Dies gilt insbesondere auch für die Untersuchung bzgl. der Anglizismen, die als einzige eine stark ansteigende Tendenz ergeben hat.

Die Untersuchungen haben außerdem gezeigt, dass Herkunftsfaktoren (Zeitung vs. Plenarprotokoll, Land, Ressort) bzgl. der beobachteten Phänomene meist eine größere Rolle spielen als der Faktor Zeit. Dies ist zwar an sich nicht überraschend, zieht jedoch die Frage nach sich, inwieweit der Anstieg der Häufigkeit eines Phänomens um wenige Prozent überhaupt wahrnehmbar ist.

3. Weitere Ansätze zur Untersuchung spezieller Aspekte von „Sprachverfall“

3.1 Sprachverfallswahrnehmung als Phänomen der Textrezeption

Im vorangegangenen Abschnitt sind wir der Frage nachgegangen, ob sich anhand geläufiger Hypothesen zu seinen möglichen Erscheinungsformen ein Verfall des Deutschen in Korpora beobachten, d.h. messen lässt. In die-

sem Abschnitt wollen wir der Frage nachgehen, ob es Gründe gibt, warum Sprachverfall im beobachteten Zeitraum besonders stark wahrgenommen worden sein könnte – Gründe, die zunächst weniger mit Sprachwandel zu tun haben als mit veränderten Umständen der Textproduktion und insbesondere Textrezeption.

Diese Möglichkeit scheint plausibel, da sich die Umstände der Textrezeption seit 1995 sicherlich stark verändert haben. Während 1995 noch Tageszeitungen, Bücher, redaktionell bearbeitete Lexika etc. vermutlich den Großteil der alltäglich rezipierten Texte geprägt haben, sind diese heute teilweise verdrängt durch Texte aus dem World Wide Web, deren Entstehungs- und Auswahlbedingungen sich stark von den damaligen unterscheiden können. So ist z.B. davon auszugehen, dass heute deutlich mehr unredigierte Texte, die von sehr viel mehr unterschiedlichen und sehr viel häufiger nicht-professionellen Autoren verfasst wurden, rezipiert werden. Darüber hinaus ist vermutlich davon auszugehen, dass die Streuung bzgl. der regionalen Herkunft der Autoren größer ist und dieser und ähnliche Parameter beim Lesen seltener im Voraus bekannt sind, da Texte etwa über ein einfaches Verfolgen von Suchmaschinen-Links dekontextualisierter rezipiert werden können als etwa im Falle einer Tageszeitung oder eines Buchs. So ist es nicht ganz unplausibel, davon auszugehen, dass wir heute erstens häufiger mit geschriebenen Texten konfrontiert sind, die nicht unseren Sprachgewohnheiten entsprechen, und dass es zweitens häufiger Situationen gibt, in denen wir keine spezifischen Erwartungen bzgl. der Sprache aufbauen können, die wir rezipieren. Dass solche Kontexte in Bezug auf Merkmale, die für Sprachqualität als relevant betrachtet werden, größere Auswirkungen haben können als der Faktor Zeit, hatten wir eingangs gezeigt (siehe z.B. Abb. 3).

3.1.1 Kurzstudie: Internetbasierte Texte

Die Frage, ob es aufgrund der Auflösung von soziokulturellen und regionalen Plurizentrik-Barrieren und -Markern durch das WWW eine verstärkte Wahrnehmung von Sprachverfall gibt, kann an dieser Stelle natürlich nicht beantwortet werden, zumal man bezweifeln kann, ob Sprachverfall im untersuchten Zeitraum überhaupt stärker wahrgenommen wurde als zu einem anderen Zeitpunkt bzgl. irgendeiner anderen Zeitperiode (siehe Durrell in diesem Band). Auch zur Veränderung der Textrezeption liegen keine ausreichenden Daten vor, die eine genauere Untersuchung zulassen würden, ebenso wenig wie Korpora, die als Stichprobe für eine veränderte Textrezeption geeignet wären. Aus diesen Gründen haben wir auch an dieser Stelle die Fragestellung vereinfachen müssen und versucht, anhand bestehender webbasierter Korpora zu untersuchen, ob es bezüglich der eingangs vor-

gestellten Hypothesen zur Manifestation von Sprachverfall Indizien gibt, die für oder gegen die Vermutung sprechen, dass die verstärkte Rezeption von Web-Texten zu einer verstärkten Wahrnehmung von Sprachverfall führen kann.

Als Grundlage der Untersuchung haben wir in diesem Fall drei DEREKO-Subkorpora, nämlich die Wikipedia-Konvertierungen von 2005 (Subkorpusname in DEREKO/COSMAS II: wpd) und 2011 (wpd11) und die Konvertierung der Wikipedia-Diskussionen von 2011 (wdd11) sowie eine 10%-Zufallsstichprobe aus dem Deutschen Web-Korpus deWaC in einer Version von 2005 (Baroni/Kilgarriff 2006) verwendet. Die so getroffene Korpusauswahl war allein durch das Fehlen anderer Daten bedingt und hat weder den Anspruch einer repräsentativen Stichprobe bzgl. des Word-Wide-Web noch bzgl. der Rezeption webbasierter Texte. Untersucht haben wir nur den Anteil schwacher Präteritumsbildung bei stark flektierbaren Verben sowie den Anteil des adjektivischen Gebrauchs von Adverbien und die analytische Konjunktivverwendung.

Ergebnisse und Diskussion

Was den Anteil schwacher Präteritumsbildung bei stark flektierbaren Verben betrifft (Abb. 10), ist das Phänomen insbesondere beim Wikipedia-Korpus eher weniger stark ausgeprägt als in den Zeitungskorpora. Bei den beiden anderen Web-Korpora wird – erkennbar an den großen Konfidenzintervallen – das Präteritum insgesamt so selten verwendet, dass die Ergebnisse kaum aussagekräftig sind.

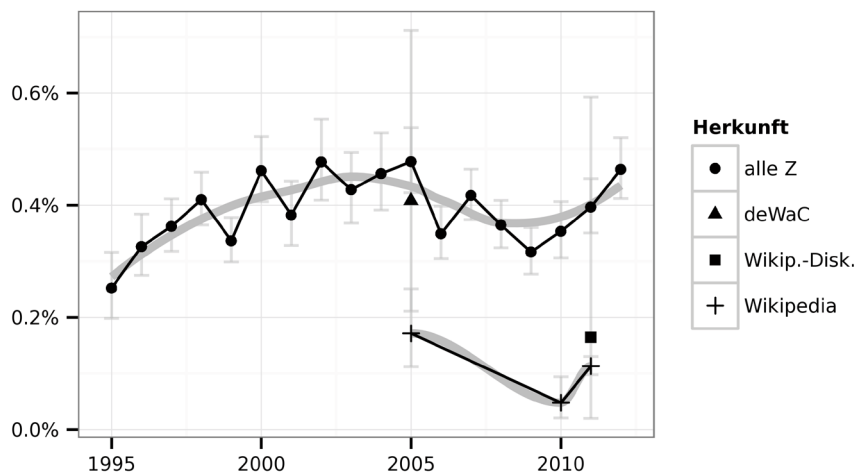


Abb. 10: Anteil schwacher Präteritumsbildung bei stark flektierbaren Verben

Anders sehen die Ergebnisse beim Anteil des flektierten Gebrauchs von Adverbien (Abb. 11) aus. Hier ist der Anteil bei deWaC und den Wikipedia-Diskussionen deutlich größer als in den Zeitungskorpora.

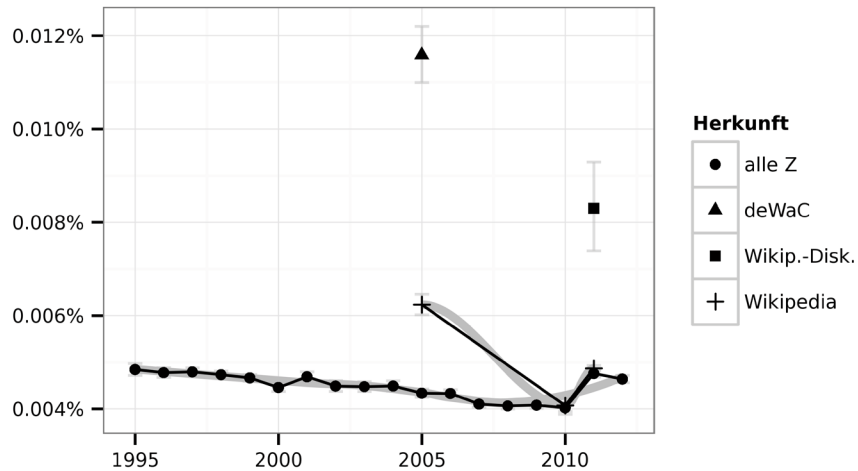


Abb. 11: Anteil des flektierten Gebrauchs von Adverbien

Deutlich größer als in den Zeitungskorpora ist in deWaC auch der Anteil analytischer Konjunktivverwendungen (Abb. 12). In diesem Fall verhalten sich jedoch auch die Wikipedia-Diskussionen ähnlich unauffällig wie die Wikipedia-Artikel.

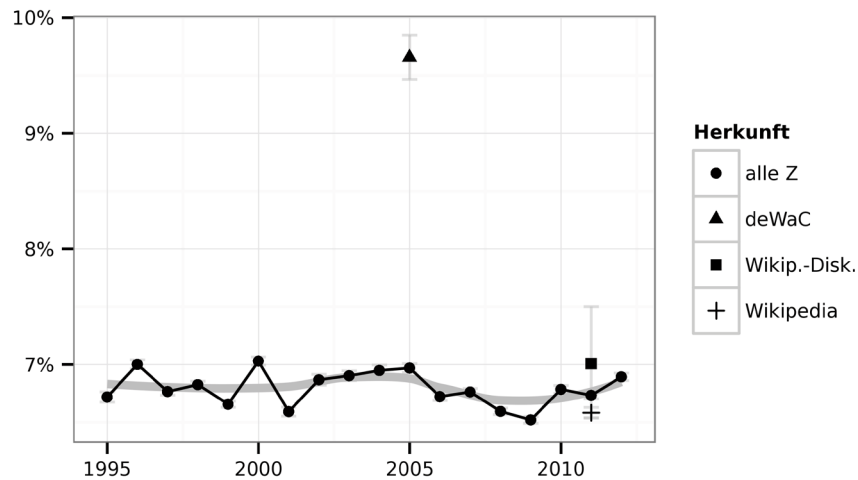


Abb. 12: Anteil analytischer Konjunktivverwendungen

3.1.2 Fazit

Die recht heterogenen Ergebnisse lassen sich vielleicht so zusammenfassen, dass es bzgl. des allgemeinen deWaC keine Veranlassung gibt, die Hypothese, dass ein eventuell wahrgenommener Sprachverfall auch auf WWW-bedingt veränderte Umstände der Textrezeption zurückzuführen ist, zu verwerfen. Darüber hinausgehende Schlussfolgerungen lassen sich anhand der wenigen punktuellen Messungen, die auch gezeigt haben, dass sich Texte aus dem Web sehr unterschiedlich verhalten und dass daher Ergebnisse entscheidend von der Stichprobensammensetzung abhängen, nicht ziehen.

3.2 Lexikalische Vielfalt

Neben den eingangs untersuchten möglichen Kriterien für die Bewertung von Sprachwandel, die eine Bewertungsrichtung nahelegen und prinzipiell eine Verallgemeinerung der auf unserer Korpusgrundlage gewonnenen Beobachtungen zulassen, gibt es im Bereich der Stilometrie zahlreiche weitere quantifizierbare Kriterien, die grundsätzlich für eine Bewertung von Sprachwandel in Frage kommen, wie z.B. lexikalische Vielfalt, Satzkomplexität (syntaktische Einbettungstiefe, Aufzählungslängen, Satzlänge), Funktionswortdichte, Gebrauch von Mehrworteinheiten, Idiomatizität (Abdeckung durch gute/sprachtypische n-Gramme, Kollokationen).

Aus diesen potenziellen Kriterien haben wir uns für eine weitere Untersuchung die lexikalische Vielfalt herausgegriffen, da die Bewertungsrichtung vergleichsweise eindeutig und eine leichte Operationalisierbarkeit gegeben ist. Als Maß haben wir das Measure of Textual Lexical Diversity (MTLD) (McCarthy 2005) gewählt, da es im Gegensatz zur standardisierten Type-Token-Ratio auch auf Texte sehr unterschiedlicher Länge anwendbar ist. Berechnet haben wir das durchschnittliche MTLD auf der Basis von Texten und Wortformen.

3.2.1 Ergebnisse und Diskussion

Wie Abbildung 13 zeigt, hat das MTLD aller untersuchten Korpora eine negative Tendenz. Allerdings scheint der Rückgang in den Zeitungskorpora zum größten Teil auf die Ressorts Lokales und Kultur zurückzugehen (siehe Abb. 14). Dies legt die Vermutung nahe, dass die wechselnde Zusammensetzung der Zeitungskorpora und speziell der Anteil an Artikelsorten mit schematischer Sprache (z.B. Veranstaltungsankündigungen und Todesanzeigen) für den Rückgang mit verantwortlich sein könnte. Um diese Vermutung zu überprüfen, haben wir versucht, den Anteil solcher Artikel mit

schematischer Sprache anhand des Anteils an partiellen Dubletten (mindestens 50% Abdeckung durch gemeinsame normalisierte 5-Gramme, siehe Kupietz 2005) zu approximieren. Das Ergebnis legt den Schluss nahe, dass unsere Vermutung richtig war: Der Anteil an solchen partiellen Dubletten zeigt von 2004 bis 2008 einen markanten Anstieg von 0,2% auf 12,1% und weist insgesamt eine starke negative Korrelation mit den MTLD-Werten auf ($r = -0,73, p < 0,001$).

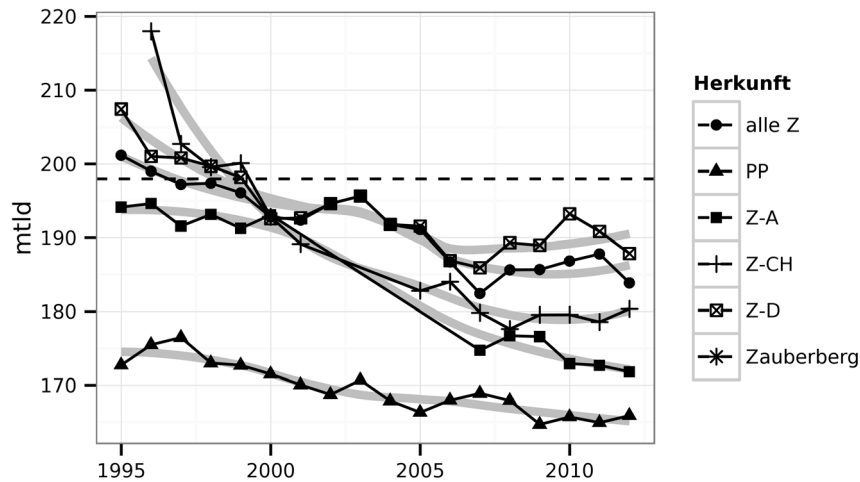


Abb. 13: Durchschnittliches MTLD über Texte aus den jeweiligen Jahren. Zum Vergleich das MTLD des Zauberbergs von Thomas Mann als gestrichelte Linie

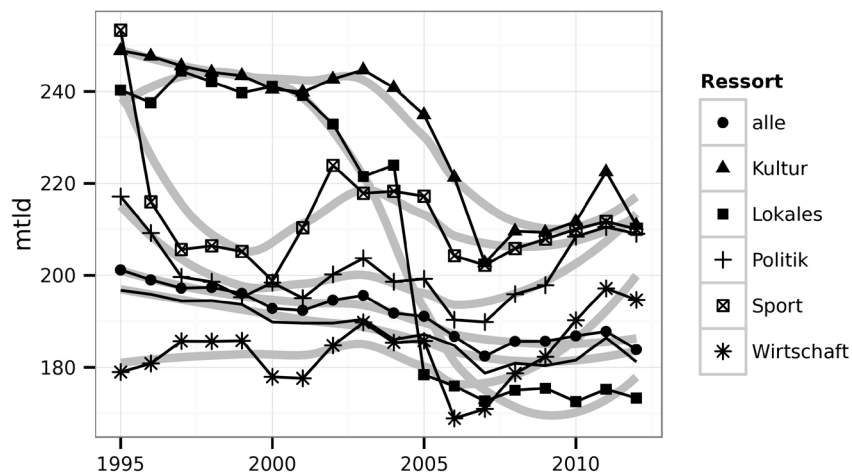


Abb. 14: Nach Ressort aufgeschlüsselte MTLD-Entwicklung der Zeitungskorpora

Dies kann allerdings nicht den Rückgang der MTLD-Werte in den Plenarprotokoll-Korpora erklären, da diese sehr homogen zusammengesetzt sind. Unser Haupt-Fehlerverdacht in diesem Fall bezog sich auf die technische Aufbereitung des Korpus. So könnte die abnehmende Vielfalt hier etwa mit dem Rückgang der Verwendung von OCR bei der Aufbereitung der Rohdaten zusammenhängen. D.h., dass größere Vielfalt nicht unbedingt auf mehr seltene Wörter sondern auf mehr fehlerhaft erkannte Zeichenketten zurückzuführen sein könnte. Um diese Fehlerhypothese zu untersuchen, haben wir die Typelisten der verschiedenen Jahrgänge mit dem Lemmatisierer und morphologischen Analysewerkzeug *glemm* (Belica 1994) untersucht, in der Erwartung, dass mit einem höheren Anteil an Datenmüll bzw. Nicht-Wörtern der Anteil der durch *glemm* nicht analysierbaren Zeichenketten wächst. Die Ergebnisse haben unseren Verdacht in gewissem Maße bestätigt: Der Anteil nicht analysierbarer Types fällt im Plenarprotokoll-Korpus von 38% in 1999 auf 28% in 2007 während er in den Zeitungskorpora konstant bei etwa 42% liegt. Die Korrelation ist in diesem Fall ($r = 0,65$, $p < 0,005$) allerdings nicht so stark wie im Falle der Dublettenanalyse oben. Obwohl wir mit diesen beiden Korrelationsstudien noch keine ausschließliche kausale Erklärung für das rückläufige MTLD in den Zeitungs- und Plenarprotokoll-Korpora nachgewiesen, sondern nur plausible alternative Erklärungsmöglichkeiten aufgezeigt haben, legen diese doch den Schluss nahe, dass die Beobachtungen nicht auf den Sprachgebrauch schlechthin verallgemeinert werden können – zumal es ohnehin fraglich ist, ob die Bedingungen für die Verallgemeinerbarkeit eines stil- und textsortenabhängigen Merkmals wie der lexikalischen Vielfalt durch die verwendete Korpusgrundlage grundsätzlich erfüllt sind.

Um eine weitere mögliche methodische Fehlerquelle auszuschließen, haben wir zuletzt noch die Validität und Adäquatheit des MTLD ansatzweise überprüft, indem wir als ein alternatives Maß den durchschnittlichen Frequenzrang aller Types für alle Jahrgänge der untersuchten Korpora gemessen haben (vgl. Steinig et al. 2009, S. 121 f.). Die Frequenzränge wurden dabei anhand der Gesamtkorpora (z.B. Lokal-Ressort aus allen Zeitungen über alle Jahrgänge) ermittelt. Das alternative Maß sollte innerhalb eines Gesamtkorpus Werte liefern, die von der Größe der Teilkorpora ausreichend unabhängig sind und die die durchschnittliche Häufigkeit der in den Teilkorpora verwendeten Types innerhalb der durch das Gesamtkorpus repräsentierten Sprachdomäne ausdrücken: je höher der durchschnittliche Häufigkeitsrang, desto seltener die durchschnittlich verwendeten Types und desto reichhaltiger der verwendete Wortschatz. Der Vergleich zwischen mittlerem Rang und MTLD ergab eine starke Korrelation (beim Lokal-Ressort-Korpus $r = 0,89$, $p < 0,001$) und hat damit die Eignung des durchschnittlichen MTLD für unseren Zweck bekräftigt.

3.3 Gewichtung von Bewertungskriterien

Wir haben einige mögliche Bewertungskriterien von Sprach- bzw. Sprachgebrauchswandel vorgestellt. Offen geblieben ist bisher die Frage, wie man die Relevanz der einzelnen Merkmale ermitteln oder festlegen kann.

Ein zwar nicht korpusbasierter aber grundsätzlich datengestützter, alternativer Ansatz, die Merkmalsgewichtung anzugehen, ist eine Vorgehensweise, die häufig in der angewandten Informatik verwendet wird: Man definiert zunächst einen Goldstandard durch manuelle Bewertung einer Stichprobe – in diesem Fall von Texten bzgl. einer Skala [gutes Deutsch ... schlechtes Deutsch] – durch Experten (oder andere Gruppen von *Ratern*) und trainiert auf diesen Daten (Merkmalsausprägungen und Ratings) einen automatischen Klassifikator. Wenn der Klassifikator nicht erfolgreich trainiert werden kann, hätte man einen Hinweis darauf, dass möglicherweise die gewählten Merkmale für die Entscheidungen der Rater nicht ausreichend relevant waren. Andernfalls hätte man ein Modell der Sprachbewertung zu einem bestimmten Zeitpunkt – jedoch kein objektiviertes Modell, da die Gewichtung der Faktoren durch Rater implizit enthalten ist, ebenso wie die Auswahl der Rater. Der Klassifikator kann anschließend auf neue Daten angewendet werden, um Vorhersagen zu erhalten und (im Prinzip) analysiert werden, um Erkenntnisse darüber zu erhalten, welche Faktoren wie relevant sind und wie sie miteinander verknüpft sind. Eine weitergehende Frage wäre dann, ob das gewonnene Modell auch dazu taugt, einen Zeitungsartikel von 1920 (oder 2020) aus der Perspektive von 1920 (oder 2020) zu bewerten. Um sich dem Phänomen Sprachverfall weiter anzunähern und weitere Erkenntnisgewinne zu erhalten, müsste man die Invarianzen mehrerer solcher Modelle über die Zeit untersuchen.

Alternativ zur extensionalen Festlegung eines statischen Goldstandards könnte man auch versuchen, einen dynamischen Kriterienstandard zu definieren, z.B. „Gutes Deutsch ist, was im Zeitraum Δt im Feuilleton der FAZ stand.“. Damit würde eine gewisse Unabhängigkeit von der Änderung der außersprachlichen Realität und der Bewertungsmaßstäbe über die Zeit erreicht, wissenschaftlich wäre damit aber wohl nichts gewonnen.

4. Zusammenfassung und Schlussfolgerungen

Wir haben eingangs versucht, die Untersuchung von Sprachverfall zu strukturieren und zwar im Wesentlichen in zwei Teile: (1.) Bereiche von Einflussfaktoren und (2.) die Gewichtung und Einordnung der Faktoren (siehe Abb. 1). Von den zahlreichen potenziellen Einflussbereichen haben wir einige derer, die sowohl im gesellschaftlichen Diskurs präsent als auch empirisch zugänglich sind, punktuell untersucht. Dabei hat sich mit fünf ansteigenden

Tendenzen, drei rückläufigen und vier konstanten Verläufen ein gemischtes Bild ergeben, wobei bei den Tendenzen größtenteils unklar geblieben ist, ob sie eine wahrnehmbare Größenordnung erreicht hatten (Ausnahme: Anglizismen). Es hat sich auch gezeigt, dass die meisten Untersuchungen aufwändiger hätten angelegt werden müssen, um eindeutige Ergebnisse zu erzielen und die Beobachtungen im Hinblick auf die Frage nach Sprachverfall interpretieren zu können und alternative Erklärungen auszuschließen. Sowohl hinsichtlich der Ergebnisse als auch hinsichtlich der möglichen alternativen Erklärungen und der Frage nach der Verallgemeinerbarkeit war dabei besonders die Untersuchung der lexikalischen Vielfalt interessant: Zwar wies das MTLD (als Maß der lexikalischen Vielfalt) sowohl in den Zeitungs- als auch in den Plenarprotokoll-Korpora einen Rückgang über die Zeit auf, jedoch konnte dieser größtenteils auf außersprachliche, v.a. auf korpusmethodische bzw. technische Ursachen zurückgeführt werden.

Was die Gewichtung von Einflussfaktoren betrifft, haben wir argumentiert, dass ohne die Festlegung einer solchen aus der empirischen Untersuchung einzelner möglicher Faktoren selbst keine Antwort auf die Frage nach Sprachverfall abgeleitet werden kann. Je nachdem, wie man die in Tabelle 1 zusammengefassten Beobachtungen gewichtet, lautet die Antwort ja oder auch nein. Wir haben außerdem skizziert, wie man auch induktiv auf Grundlage manuell bewerteter Texte zu einer möglichen Gewichtung gelangen kann. Den verschiedenen Gewichtungsmöglichkeiten entsprechend gibt es unseres Erachtens auch nicht den einen Sprachverfall, den es zu entdecken gilt, sondern sicherlich mehrere denkbare Konzepte. Wir selbst haben Sprachverfall als Thema eines gesellschaftlichen Diskurses aufgegriffen und mit „Sprachverfallswahrnehmung als Phänomen der Textrezeption“ im Zusammenhang mit der Nutzung des WWW ein weiteres alternatives Konzept ansatzweise untersucht.

Literatur

- Baroni, Marco/Kilgarriff, Adam (2006): Large linguistically-processed Web corpora for multiple languages. In: Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Trient, S. 87–90.
- Belica, Cyril (1994): WP-2 Lemmatizer. MECOLB final report: MLAP93-21. Internet: www1.ids-mannheim.de/fileadmin/kl/dokumente/glemmrep.pdf (Stand: 19.6.2013).
- Belica, Cyril et al. (2011): The Morphosyntactic Annotation of DEREKO: Interpretation, opportunities, and pitfalls. In: Konopka, Marek et al. (Hg.): Grammatik und Korpora 2009. Dritte Internationale Konferenz. Mannheim, 22.–24.9.2009. (= Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache 1). Tübingen, S. 451–469.

- Hilpert, Martin/Griess, Stefan T. (2009): Assessing frequency changes in multistage diachronic corpora: Applications for historical corpus linguistics and the study of language acquisition. In: *Literary and Linguistic Computing* 24, S. 385–401.
- IDS (2013): Deutsches Referenzkorpus. Archiv der Korpora geschriebener Gegenwartssprache 2013-I, Institut für Deutsche Sprache, Release vom 15.03.2013. Internet: www.ids-mannheim.de/kl/projekte/korpora/archiv.html (Stand: 15.8.2013).
- Kupietz, Marc (2005): Near-Duplicate Detection in the IDS Corpora of Written German (Tech. Rep. KT-2006-01). Institut für Deutsche Sprache. Internet: <http://ftp.ids-mannheim.de/kt/ids-kt-2006-01.pdf> (Stand: 19.6.2013).
- Kupietz, Marc et al. (2010): The German Reference Corpus DEREKO: A primordial sample for linguistic research. In: Calzolari, Nicoletta et al. (Hg.): *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC 2010)*. Valletta, S. 1848–1854.
- McCarthy, Philip M. (2005): *An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD)*. Diss., Univ. Memphis.
- Perkuhn, Rainer/Keibel, Holger/Kupietz, Marc (2012): *Korpuslinguistik*. Paderborn.
- Schmid, Helmut/Laws, Florian (2008): Estimation of conditional probabilities with decision trees and an application to fine-grained POS tagging. In: *Proceedings of the 22nd International Conference on Computational Linguistics (COLING) 2008*. Bd. 1. Manchester, S. 777–784.
- Sick, Bastian (2004): *Der Dativ ist dem Genitiv sein Tod. Ein Wegweiser durch den Irrgarten der deutschen Sprache*. Köln.
- Sick, Bastian (2005): *Der Dativ ist dem Genitiv sein Tod. Folge 2: Neues aus dem Irrgarten der deutschen Sprache*. Köln.
- Steinig, Wolfgang et al. (2009): *Schreiben von Kindern im diachronen Vergleich. Texte von Viertklässlern aus den Jahren 1972 und 2002*. Münster u.a.