

Die englischen Modalverben im Daumenkino: Zur dynamischen Visualisierung von Phänomenen des Sprachwandels

1. Einleitung

Der amerikanische Strukturalist Leonard Bloomfield gibt in seinem Hauptwerk *Language* (1933) eine Einschätzung ab, die den heutigen Leser verwundern muss:

The process of linguistic change has never been directly observed; we shall see that such observation, with our present facilities, is inconceivable.

Die Untersuchung von Prozessen des Sprachwandels hat sich, befördert nicht zuletzt durch die Erfolgsgeschichte der quantitativen Soziolinguistik (Labov 1994), zu einem wichtigen Zweig aktueller linguistischer Forschung entwickelt, der zudem weiter im Wachsen begriffen ist. So ist beispielsweise in der Disziplin der anglistischen Korpuslinguistik in den vergangenen Jahren ein steter Zuwachs an Datenbanken zu verzeichnen, von denen nicht wenige Texte aus vergleichbaren Gattungen vereinen, die zu verschiedenen Zeiten produziert wurden (Davies 2008; 2010). Das Interesse an diachron ausgerichteter korpuslinguistischer Forschung schlägt sich in zahlreichen Publikationen nieder, in deren Tradition sich auch der vorliegende Beitrag sieht (Lindquist/Mair 2004; Kohnen 2006; Lenker/Meurman-Solin 2007; Mair 2006; Mair u. a. 2009; *inter alia*).

Das Hauptwerkzeug der diachron ausgerichteten Korpuslinguistik ist die Erhebung von Frequenzdaten und deren Vergleich über mehrere Zeitperioden hinweg. Mit solchen Vergleichen lässt sich unter anderem zeigen, ob sich eine sprachliche Form zunehmend etabliert oder ob sie langsam aus dem Sprachgebrauch herausfällt. In dem vorliegenden Beitrag soll ein Werkzeug zur Anwendung kommen, das zwar ebenfalls auf Frequenzdaten beruht, das jedoch über die Darstellung von sich wandelnden Häufigkeiten hinausgeht. Der Ansatz nimmt sich dabei zur Aufgabe, Prozesse des Sprachwandels auf dynamische Art und Weise darzustellen, so dass aus einer Serie statischer Visualisierungen ein Eindruck kontinuierlichen Wandels entsteht, ähnlich wie es bei einem Daumenkino der Fall ist. Um dies zu erreichen, bedarf es einer Datenbank, in der dasselbe sprachliche Phänomen in sequentiell geordneten Zeitfenstern untersucht und visualisiert werden kann. Der hauptsächliche Zweck eines solchen Vorgehens ist, dass komplexe Frequenzdaten aus Korpora in ein Format gebracht werden, das der explorativen Untersuchung von Sprachwan-

delphänomenen und der Bildung von Hypothesen über Sprachwandel im Allgemeinen dienen kann.

Das sprachliche Phänomen, das in diesem Aufsatz im Zentrum der Untersuchung stehen soll, ist der Bereich der englischen Modalverben. Es besteht ein Konsens in der anglistischen Forschungslandschaft, dass sich die Modalverben gegenwärtig in einer bereits länger andauernden Phase des Wandels befinden (Palmer 1990; Krug 2000; Facchinetti u. a. 2000; Mair 2006, *inter alia*), wobei es allerdings unterschiedliche Ansichten gibt, worin dieser Wandel besteht. Nach einer vorherrschenden Meinung geht die Frequenz der ›klassischen‹ Modalverben zurück, während sich eine Riege ›quasimodaler‹ Verben zunehmend etabliert. Wie genau dieser Prozess vonstattengeht ist Gegenstand lebhafter Diskussion (vgl. etwa Millar 2009; Leech 2011). In den beiden Hauptsektionen dieses Beitrags sollen zwei Analysen vorgestellt werden, die Wandelprozesse im Bereich der Modalität auf unterschiedliche Art und Weise darstellen. Die erste der beiden Analysen basiert dabei auf bivariaten Frequenzdaten, die zweite auf multivariaten und statistisch aufbereiteten Daten.

2. Die Modalverben *may*, *might*, und ihre Kollokate

Dieser Abschnitt zeigt eine erste Anwendungsmöglichkeit des sprachlichen Dauerkinos, bei der zwei sprachliche Formen, hier die Modalverben *may* und *might*, anhand einer frequenzbasierten Variable miteinander in Kontrast gesetzt werden. Diese Variable ist die Häufigkeit, mit der *may* und *might* eine Verbindung mit lexikalischen Verben im Infinitiv eingehen. Konkret formuliert: Wie häufig treten in einem Korpus die Verbindungen *may say* oder *might say* auf und wie verhalten sich die Häufigkeiten dieser Verbindungen zueinander oder zu Verbindungen mit anderen lexikalischen Verben? Anhand einer Datenbank wie dem COHA Korpus (Davies 2010) lassen sich die entsprechenden Werte ermitteln. Schaubild 1 liefert eine teilweise graphische Darstellung der Ergebnisse, die sich für die Jahre 1860–69 ergeben. In dem Schaubild sind die Kollokate von *may* und *might* als Kreise dargestellt; die Positionen der Kreise geben an, wie häufig die lexikalischen Verben mit *may* und *might* vorkommen. Beispielsweise lässt sich ablesen, dass auf eine Million Wörter im untersuchten Korpus etwa 15 Tokens von *might say* kommen. Die relative Größe der Kreise zeigt die kombinierte Häufigkeit der Verben in Verbindungen mit *may* und *might* an. Das Schaubild zeigt die 80 häufigsten Verben unter Auslassung von *be* und *have*. Zum Zweck der besseren Lesbarkeit sind nur acht Verben beschriftet.

Erwartbarerweise zeigt sich in dem Schaubild eine positive Korrelation der Frequenzwerte: Verben die häufig mit *may* auftreten, finden sich ebenfalls häufig mit *might*. Allerdings lassen sich auch gewisse Unterschiede ausmachen. Das Verb *seem* beispielsweise ist das zweithäufigste mit *may*, aber nur das zehnthäufigste mit *might*. In einer synchronen Untersuchung der Kollokate von *may* und *might* wären Unterschiede dieser Art genauer zu betrachten. Da in diesem Beitrag allerdings die

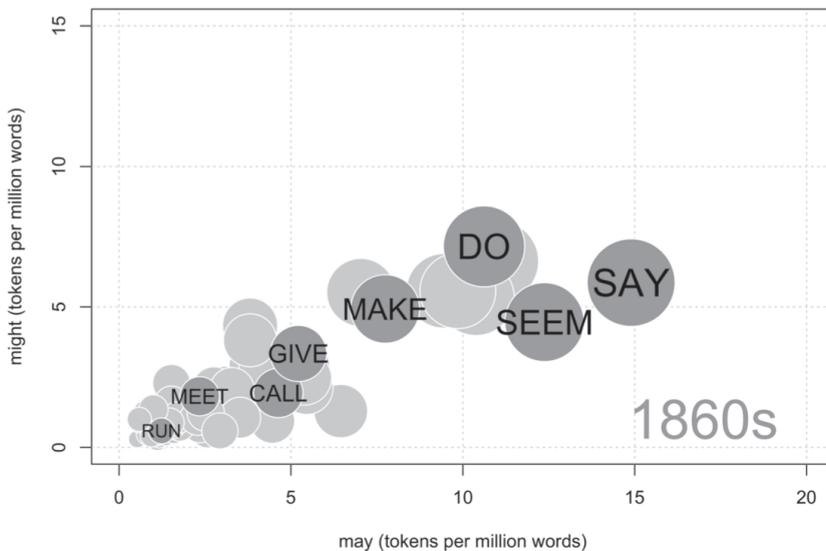


Schaubild 1: Die Kollokate von *may* und *might* in den 1860er Jahren

Diachronie im Vordergrund stehen soll, kommen wir direkt zu einem Vergleich der Daten über mehrere Zeitperioden hinweg. Schaubild 2 zeigt die Frequenzen der Kollokate von *may* und *might* über den Verlauf von insgesamt 150 Jahren. Die Darstellung der Ergebnisse folgt dem Modell von Schaubild 1, so dass die Werte von *may* auf der x-Achse abgetragen sind, die Werte von *might* auf der y-Achse. Anders als beim ersten Schaubild gilt es aber hier, wie bei einem Daumenkino, Unterschiede von einem Bild zum nächsten zu entdecken. Einige davon sollen im Folgenden kurz diskutiert werden.

Global zeigt sich in dem Schaubild ein Frequenzrückgang des Modalverbs *may*. Während in den ersten Dekaden die häufigsten Verbindungen mit *may* durchaus zwischen 10 und 15 mal in einer Million Wörtern zu finden sind, fallen diese im Laufe der Zeit deutlich unter 10 zurück. Für *might* lässt sich ein allgemeiner Frequenzrückgang nicht feststellen, wobei einzelne Verbindungen aber durchaus Häufigkeitsschwankungen unterworfen sind. Besonders gekennzeichnet sind in Schaubild 2 die Verben *give*, *say* und *seem*. Die Entwicklungen dieser Verben zeigen exemplarisch, was für Veränderungen in einem sprachlichen Daumenkino erfasst und visualisiert werden können. Das Verb *say* ist in den 1860er Jahren das häufigste Element, das eine Verbindung mit *may* eingeht. Im Verlauf der nächsten 50 Jahre entwickelt es sich zunächst auch zum häufigsten Kollokat von *might*, nach 1920 setzt eine sukzessive Auflösung der Verbindung *may say* ein. Im aktuellen Gebrauch des amerikanischen Englisch dominieren andere lexikalische Verben als Hauptkollokate von *may*. Ganz anders verläuft die Entwicklung von *seem*, das sich über die Jahrzehnte als häufiges Kollokat von *may* hält. Als Kollokation bleibt *may*

seem trotz des generellen Frequenzrückgangs von *may* noch eine häufige Wendung. Nicht alle Verben zeigen Besonderheiten, die entweder *may* oder *might* betreffen. Das Verb *give* weist über die Jahre hinweg Frequenzwerte auf, die sich im Wesentlichen als Funktion der generellen Häufigkeit von *may* und *might* lesen lassen.

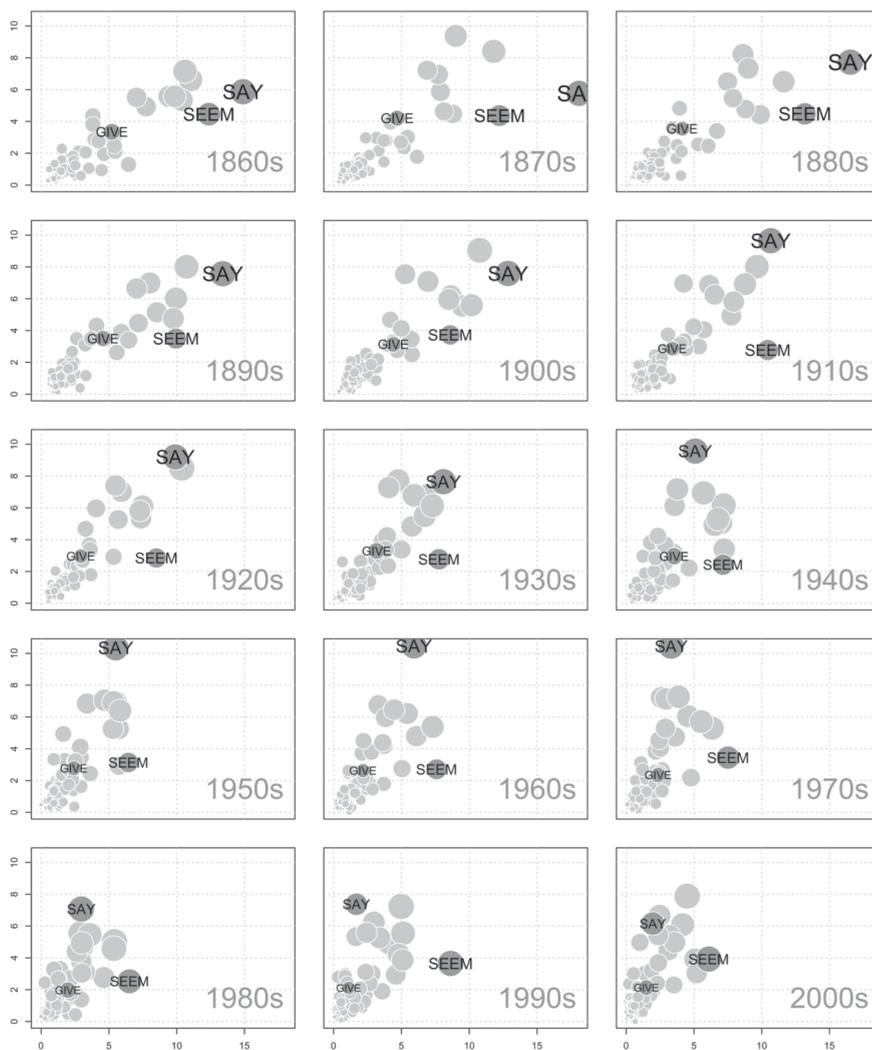


Schaubild 2: Die Kollokate von *may* und *might* über 15 Dekaden hinweg

Weitere Aspekte des Wandels von *may*, *might*, und ihren verbalen Kollokaten könnten hier diskutiert werden; stattdessen sei auf einige generelle Möglichkeiten des verwendeten Ansatzes verwiesen. Die Betrachtung sprachlicher Wandelpro-

zesse im Daumenkino bereitet eine große Menge quantitativer Information so auf, dass der Betrachter Entwicklungen intuitiv erfassen und in einem zweiten Schritt Hypothesen über mögliche Erklärungen anstellen kann. Die technische Hilfe kann dabei beliebig ausgeweitet werden, beispielsweise können diejenigen Elemente hervorgehoben werden, die über die Zeit hinweg am meisten ›in Bewegung‹ sind. Ein wichtiger Aspekt ist, dass die Darstellung eine gleichzeitige Betrachtung globaler und partikulärer Entwicklungen zulässt. Interessant ist die Entwicklung eines einzelnen Verbs ja erst im Kontext einer größeren Gruppe von Verben, die diese Entwicklung nicht teilen. Elemente, die ›gegen den Strom schwimmen‹, können durch ein sprachliches Daumenkino leicht identifiziert werden.

An der bisherigen Vorstellung der Methode ließe sich kritisieren, dass sich nur eine begrenzte Anzahl sprachlicher Phänomene als bivariate Frequenzwerte darstellen lassen, und die Anwendungsmöglichkeiten daher beschränkt seien. Diesem Einwand begegnet die zweite Beispielanalyse, mit der sich die folgende Sektion befasst.

3. Eine multivariate Analyse der englischen ›core modals‹

In dieser Sektion soll ein sprachliches Daumenkino vorgestellt werden, das nicht, wie das oben gezeigte, auf bivariaten Frequenzdaten beruht, sondern dass ungleich komplexere Daten zur Grundlage hat. Wie im bereits gezeigten Beispiel geht es jedoch auch hier um die Erstellung zweidimensionaler Schaubilder, aus deren Sequenz sich sprachliche Entwicklungen ablesen lassen. In diesem Fall jedoch sind die Schaubilder das Ergebnis eines aufbereitenden statistischen Verfahrens, der multi-dimensionalen Skalierung, kurz MDS (Kruskal/Wish 1978; Wheeler 2005). Schaubilder dieser Art werden zumeist als ›Karten‹ bezeichnet, die ein komplexes Phänomen vereinfachend als Konstellation mehrerer Punkte auf einer zweidimensionalen Fläche darstellen. In solch einer Darstellung finden sich Elemente in großer räumlicher Nähe, die sich auf messbare Art und Weise ähnlich sind. Umgekehrt platziert das Verfahren unähnliche Elemente in größtmöglicher Entfernung zueinander. Um zu den jeweiligen Koordinaten zu gelangen, müssen die relativen Ähnlichkeiten der untersuchten Elemente ermittelt werden. Dies geschieht anhand von Ähnlichkeitsmerkmalen, die im Vorwege der Untersuchung ausgewählt werden müssen. Die Methode der multidimensionalen Skalierung kommt in der Sprachwissenschaft in verschiedenen Bereichen zum Einsatz, nicht zuletzt in der Soziolinguistik (Labov 1994), der Varietätenlinguistik (Szmrecsanyi/Kortmann 2009) oder der Sprachtypologie (Croft/Poole 2008).

Das Ziel einer MDS-Analyse ist für gewöhnlich die Untersuchung eines komplexen sprachlichen Phänomens im aktuellen Sprachgebrauch, also in der Synchronie. Eine diachrone Anwendung ist jedoch durchaus möglich, so unternimmt beispielsweise Szmrecsanyi (2010) einen Vergleich von zwei Zeitstufen. In der vorliegenden Analyse sollen insgesamt 15 Zeitstufen berücksichtigt werden, nämlich wieder die Dekaden von 1860 bis 2000, die in dem COHA Corpus (Davies

2010) repräsentiert sind. Die Datenmenge der Teilkorpora ist dabei groß genug, um auch komplexe Phänomene sinnvoll untersuchen zu können.

Konkret geht es in dieser Sektion um die neun englischen ›core modals‹ *can*, *could*, *may*, *might*, *shall*, *should*, *will*, *would*, und *must*. Die Form und Bedeutung der englischen Modalverben ist Gegenstand intensiver Forschung, auf die der vorliegende Beitrag zurückgreifen kann (Coates 1983; Palmer 1990; Krug 2000, Facchinetti u. a. 2000, auch die einschlägigen Referenzgrammatiken Quirk u. a. 1985, Biber u. a. 1999, Huddleston/Pullum 2002 liefern ausführliche Informationen). Sämtliche der neun Modalverben sind hinsichtlich ihrer Bedeutung sehr variabel (Coates 1983), was eine semantische Untersuchung zu einer anspruchsvollen und nicht unproblematischen Aufgabe macht, besonders wenn nur dekontextualisierte und konstruierte Beispiele in Betracht gezogen werden. Die vorliegende Fallstudie hat wieder die Häufigkeiten verbaler Kollokate als Datengrundlage. In der ersten Beispielanalyse wurden diese für die zwei Modalverben *may* und *might* direkt verglichen; in dieser Fallstudie wird der Vergleich auf sämtliche neun Modalverben ausgedehnt. Das Ziel der Untersuchung ist zu ermitteln, wie sich die Modalverben hinsichtlich ihrer verbalen Kollokate ähneln. Welche Modalverben treten mit denselben Kollokaten in ähnlichen Häufigkeiten auf, welche Modalverben zeigen ein eher eigenständiges Profil? Eine zugrundeliegende Annahme für die Untersuchung dieser Fragen ist die sogenannte ›distributionelle Hypothese‹ (Firth 1957; McDonald/Ramscar 2001), die wohl am treffendsten von J.R. Firth (1957, S. 11) mit dem Slogan »You shall know a word by the company it keeps!« zum Ausdruck gebracht wurde. Elemente mit ähnlichen Kollokaten und Kollokatfrequenzen sind damit als semantisch und pragmatisch ähnlich anzusehen. Die distributionelle Hypothese liegt vielen aktuell gebräuchlichen Verfahren in der Korpuslinguistik und Computerlinguistik zugrunde, beispielsweise der Arbeit zu semantischen Vektor-Modellen (Pado/Lapata 2007).

Im vorliegenden Fall liefert Firths Hypothese die Voraussage, dass semantisch ähnliche Modalverben, wie etwa *must* und *should*, die beide Verpflichtungen zum Ausdruck bringen, auch mit ähnlichen Sets von Kollokaten in ähnlichen Häufigkeiten vorkommen. In einer graphischen Umsetzung einer synchronen MDS-Analyse wäre mithin zu erwarten, dass *must* und *should* in kurzer Distanz zueinander abgebildet werden. Da sich die Modalverben derzeit in einem längeren Prozess des Wandels befinden (Millar 2009), ist es weiterhin zu erwarten, dass sich die Kollokathäufigkeiten für die einzelnen Modalverben verschieben, so dass sich Veränderungen der Ähnlichkeitsbeziehungen im Paradigma der Modalverben ergeben. Ein Daumenkino von sequentiell geordneten MDS-Karten kann einer solchen Entwicklung Rechnung tragen und konkrete Anhaltspunkte zu ihrer Analyse liefern. Die folgenden Absätze beschreiben das Verfahren, eine genauere Schilderung findet sich in Hilpert (2011a).

Die Urdaten für die vorliegende Fallstudie stammen aus den letzten 15 Dekaden des COHA Corpus. Für jedes der neun Modalverben wurden Konkordanzanzen erstellt, die sämtliche Sequenzen aus Modalverb und einer infinitivischen Verbform erfassen. Diese Prozedur vernachlässigt damit unterbrochene Verbindungen

von Modalverb und Infinitiv, was bei der Interpretation der Ergebnisse berücksichtigt werden muss. Tabelle 1 zeigt einen Ausschnitt aus den erhobenen Daten, der das Format der verwendeten Information deutlich machen soll.

Tab. 1: Kollokathäufigkeiten der englischen Modalverben, 1860–69

	can	could	may	might	must	shall	should	will	would
be	3.702	3.152	8.099	3.467	5.700	3.109	5.998	8.417	9.219
do	810	606	190	126	196	167	191	836	554
see	484	701	178	101	150	360	157	439	152
make	416	389	137	89	189	118	183	570	634
have	383	1.899	1.538	2.825	2.593	1.034	2.101	1.096	7.331
tell	367	196	39	20	110	55	42	345	74
get	356	358	68	80	134	73	71	208	111
give	264	185	94	62	171	106	106	561	475
...

Anhand der gezeigten Werte lassen sich bereits jetzt Ähnlichkeiten und Unterschiede zwischen den Modalverben erkennen: Alle Modalverben treten häufig mit den generellen Verben *be* und *do* auf, aber bereits bei *have* öffnet sich die Schere zwischen den Modalverben *can* und *could*. Die Kollokation *can have* ist deutlich weniger häufig als *could have*, was aller Wahrscheinlichkeit nach auf die Verwendung des Perfekts mit *could* zurückgeht, die *can* verwehrt ist (*could have done*, **can have done*). Vor der Analyse durch MDS werden die Werte in Tabelle 1 normalisiert, so dass nicht rohe Frequenzdaten verglichen werden, sondern Häufigkeiten einer Kollokation per einer Million Wörter. Eine dergestalt modifizierte Tabelle bildet die Grundlage für die weiterführenden Schritte der Analyse. Der nächste Schritt in dem Verfahren ist die Berechnung einer sogenannten Distanzmatrix für sämtliche neun Modalverben in allen 15 Zeitfenstern. Die Distanzmatrix enthält für sämtliche logisch möglichen Paarungen der Modalverben jeweils einen Ähnlichkeitswert, der hier die euklidische Distanz zwischen zwei Reihen von Häufigkeitswerten abbildet. Eine Besonderheit in diesem Zusammenhang ist, dass nicht nur Vergleiche angestellt werden, die beispielsweise *may* und *might* gegenüberstellen. Vielmehr wird *may* (1860–69) mit *might* (1860–69) und sämtlichen anderen Modalverben in sämtlichen anderen Zeitfenstern verglichen. Die Distanzmatrix vergleicht damit auch das Modalverb *may* (1860–69) mit ›sich selbst‹ zu späteren Zeitpunkten. Dabei ist natürlich zu erwarten, dass ein Modalverb sich von einer Dekade zur nächsten nicht radikal verändert, so dass Unterschiede zwischen unterschiedlichen Modalverben größer sind als Unterschiede in der Verwendung eines einzigen Modalverbs über die Zeit hinweg. Der dritte und letzte Schritt des Analyseverfahrens ist die Anwendung eines Algorithmus, der die Werte der Dis-

tanzmatrix in eine zwei-dimensionale Karte umsetzt. Diese Umsetzung kann mehr oder minder erfolgreich sein; der Algorithmus ›bemüht sich‹, die Werte der Distanzmatrix so exakt wie möglich umzusetzen, allerdings erlauben die wechselseitigen Verhältnisse der Zahlen das nicht immer. Mit der Darstellung eines Phänomens auf einer zwei-dimensionalen Karte geht also ein Informationsverlust einher, der im Idealfall gering ist, der aber auch ein Ausmaß erreichen kann, welches die Aussagekraft einer Karte stark schmälert. Die Ergebnisse der vorliegenden Fallstudie rechtfertigen eine zweidimensionale Darstellung; die Anpassungen der Werte der Distanzmatrix liegen in einem verträglichen Rahmen.

Anders als bei synchronen MDS-Analysen ist es bei einer diachronen Analyse nicht sinnvoll, sämtliche Datenpunkte auf einer einzigen Karte anzuzeigen. Die Ergebnisse der Analyse werden stattdessen nach Zeitfenstern geordnet und so dem Betrachter ›nach und nach‹ präsentiert. Schaubild 3 liefert ein Daumenkino, das die Entwicklung der neun Modalverben im Hinblick auf die Ähnlichkeit ihrer Kollokathäufigkeiten nachzeichnet. Alle neun Modalverben sind als Kreise dargestellt, die Größe der Beschriftung gibt die Frequenz des Modalverbs in der jeweiligen Dekade an.

Wie aus der Sequenz von Schaubildern ersichtlich ist, bewahrheitet sich die Vermutung, dass Unterschiede zwischen gänzlich verschiedenen Modalverben stärker sind als Unterschiede, die dasselbe Modalverb von einer Dekade zur nächsten betreffen. Die grobe Konfiguration der neun Modalverben bleibt über die letzten 150 Jahre hinweg bestehen, so dass *would* nach wie vor in der oberen rechten Ecke zu finden ist, *will* einen Außenposten in der unteren Mitte bildet, und *shall* den linken Rand markiert. Darüber hinaus lassen sich natürlich diverse Entwicklungen ausmachen, die im Folgenden näher diskutiert werden sollen.

Da die Schaubilder nach der distributionellen Hypothese semantische Unterscheidungen abbilden sollen, ist der Frage nachzugehen, was die Konfiguration der neun Modalverben im Hinblick auf ihre Semantik zu bedeuten hat. Idealerweise sollte die Anordnung der Elemente für den linguistisch geschulten Betrachter einen Sinn ergeben, so dass beispielsweise die x-Achse und y-Achse jeweils einem semantischen oder pragmatischen Kontinuum zugewiesen werden kann. Eine informelle Betrachtung der Schaubilder lässt einige vorläufige Schlüsse zu. In der ersten Karte sind die verpflichtenden Modalverben *must* und *should* in direkter Nähe voneinander dargestellt. Die möglichkeitsorientierten Modalverben *might* und *could* befinden sich ebenfalls recht nah beieinander. Wenn man die erste Karte horizontal oberhalb von *must* unterteilt ergibt sich ein Kontrast von *might*, *could*, und *would* auf der oberen Hälfte und *can*, *shall*, *must*, *should*, *may*, und *will* auf der unteren Hälfte. Eine Arbeitshypothese könnte demnach sein, dass die y-Achse eine Unterscheidung zwischen epistemischer Modalität auf der oberen Hälfte und deontisch gelagerter Modalität auf der unteren Hälfte abbildet. Wie ließe sich eine solche Hypothese überprüfen? Da die relativen Positionen der Modalverben auf den Karte das Ergebnis ihrer jeweiligen Kollokathäufigkeiten sind, treten Elemente von der unteren Hälfte der Karte notwendigerweise mit anderen lexikalischen Kollokaten auf als Elemente von der oberen Hälfte. Ein Kontrast zwischen diesen Kollokaten

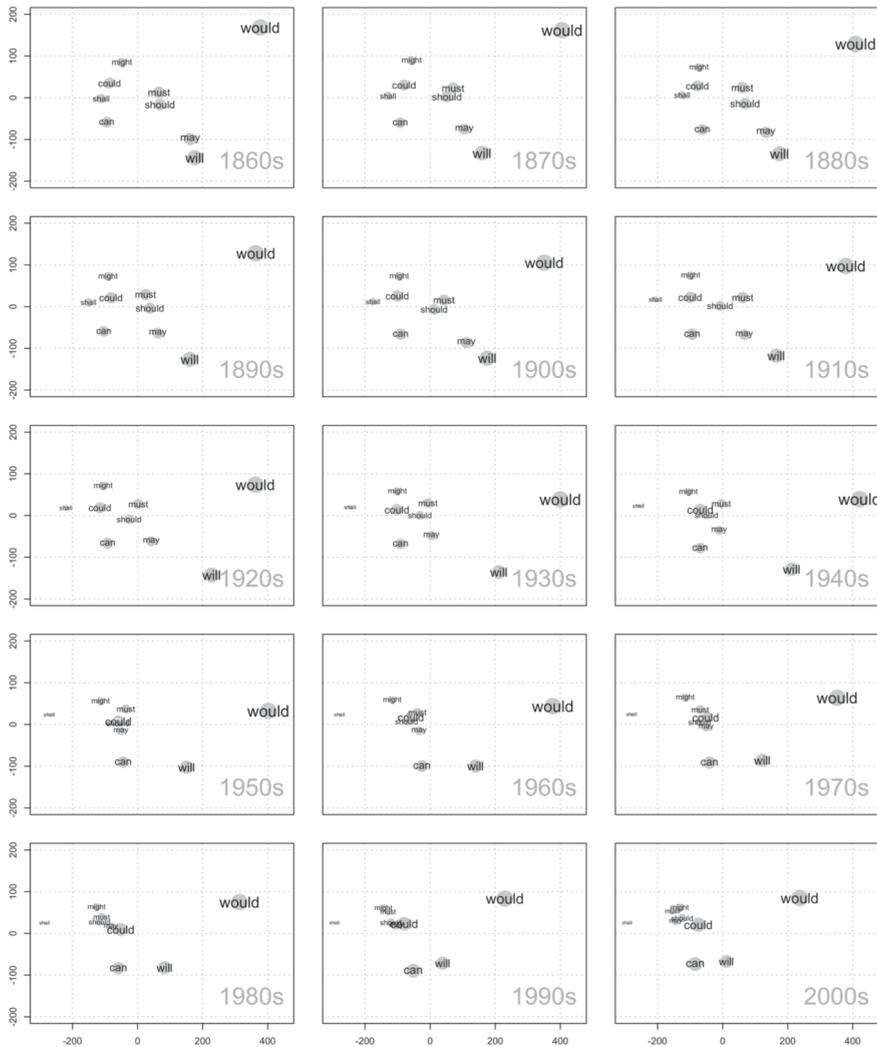


Schaubild 3: MDS-Analyse der Kollokathäufigkeiten der englischen ›core modals‹

kann herangezogen werden, um zu überprüfen, ob die Unterscheidung zwischen deontischer und epistemischer Modalität wirklich der entscheidende Faktor ist.

Eine partikuläre Entwicklung, die sich aus Schaubild 3 herauslesen lässt, ist, dass sich das Modalverb *may* im Verlauf der letzten 150 Jahre relativ stark gewandelt hat. Auf der ersten Karte ist *may* der direkte Nachbar von *will*; über die nächsten 50 Jahre hinweg nähert es sich *can* an; danach führt es seine Entwicklung sukzessive in einen Zusammenfall seiner Position mit denen von *must*, *might*, und *should*. Aus dieser Entwicklung lässt sich folgende Vorhersage ableiten: Wenn die

y-Achse eine Unterscheidung zwischen deontischer und epistemischer Modalität abbildet, so ist zu erwarten, dass *may* in den 1860er Jahren vorwiegend permissive Bedeutung zum Ausdruck gebracht hat, während *may* in den 2000er Jahren vorwiegend für die Verbalisierung logischer Möglichkeit verwendet wurde. Diese Vorhersage kann anhand der verbalen Kollokate von *may*₁₈₆₀ und *may*₂₀₀₀ überprüft werden. Tabelle 2 zeigt die Resultate einer distinktiven Kollexemanalyse (Gries/Stefanowitsch 2004), die die typischen Verbalen Kollokate von *may* in den zwei kritischen Zeitfenstern kontrastiert. Die Tabelle zeigt für jede der beiden Zeitfenster diejenigen Kollokate, deren beobachtete Häufigkeiten am weitesten (und jeweils hochsignifikant) über den erwarteten Häufigkeiten lagen.

Tab. 2: Die distinktiven Kollexeme von *may*, 1860er Jahre und 2000er Jahre

1860s	OBS.	EXP.	COLL.STR	2000s	OBS.	EXP.	COLL.STR
be	8099	7517.13	46.43	have	2089	1412.88	130.24
say	277	206.94	15.95	help	153	70.9	35.45
do	190	136.74	14.31	want	134	61.94	31.36
add	79	50.67	12.30	need	159	77.91	31.12
judge	51	31.13	10.94	sound	72	30.38	22.39
hope	46	28.08	9.87	experience	30	11.69	12.29
form	36	21.98	7.72	include	35	15.58	9.56
trust	34	20.76	7.29	explain	42	20.26	9.09
meet	43	27.47	7.02	provide	41	19.87	8.77
suppose	31	18.92	6.65	play	27	12.08	7.38

Eine Betrachtung der einzelnen Kollokate lässt den Schluss zu, dass in der Tat die erste Periode eher von Verben gekennzeichnet ist, die mit permissiver Bedeutung harmonisieren, während die zweite Periode Verben enthält, deren Verwendung vorwiegend epistemischer Natur ist. Beispielsweise beschreiben die Verben *want* und *need*, die distinktiv für die 2000er Jahre sind, Zustände, die nicht durch die Erlaubnis einer dritten Person zustande kommen. Belegstellen aus dem COHA Corpus unterstützen diese Beobachtung. ›Alte‹ Beispiele mit *be*, *say*, *do*, und *judge* bringen permissive Bedeutungen zum Ausdruck; ›neue‹ Beispiele mit *have*, *help*, *want*, und *need* tragen epistemische Bedeutung.

- (3) a. However, there are a few old roads that may be trodden with profit.
 b. Without vanity, I may say that I succeeded.
 c. »You may do so if you like,« said Mrs Sortridge composedly.
 d. Mrs. Chapman, I may add here, had a great weakness for distinctions.
 e. There has been much festivity in this saloon, if I may judge by the character of its frescos.

- (4) a. I may have told you that Miller bailed out.
 b. If the hives are itchy, antihistamines may help.
 c. The police may want to speak with you.
 d. If fillets are large, you may need to cook them in two batches.
 e. This is not so radical a step as it may sound.

Das Resultat, dass *may* im gegenwärtigen Sprachgebrauch zumeist epistemisch verwendet wird, deckt sich mit mehreren Aussagen in der Literatur zu diesem Thema. Biber u. a. (1999, S. 491) beispielsweise geben an, dass die epistemische Funktion von *may* in mehreren Textsorten des Britischen Englisch substantiell häufiger ist als die permissive. Millar (2009, S. 204) stellt im amerikanischen TIME Corpus zwischen den 1920er und 2000er Jahren eine stabile Entwicklung hin zum vermehrten Ausdruck von Möglichkeiten fest. Die Karten in Schaubild 3 und die Ergebnisse in Tabelle 2 ergänzen sich mit diesen Beobachtungen.

Eine weiterhin offene Frage ist, wie sich die x-Achse in den Karten von Schaubild 3 sinnvoll interpretieren lässt. Die Verteilung der Modalverben ist auf der x-Achse noch deutlich weiter gestreut als auf der y-Achse, so dass eine Beantwortung dieser Frage einigermaßen dringlich erscheint. Als erster Annäherungsversuch mag ein Vergleich der Modalverben *might* und *would* dienen. Beide Elemente finden sich in der oberen Hälfte der Karten, was auf vorwiegend epistemische Verwendung hinweist. Weiterhin sind die Verben in geeigneten Kontexten quasi-synonym, wie Minimalpaare belegen: Die Sätze *That might be a good idea* und *That would be a good idea* bringen ähnliche Bedeutungen zum Ausdruck. Die Tatsache, dass die beiden Elemente jedoch in relativ weiter Entfernung voneinander auf der x-Achse positioniert sind, deutet auf semantisch-pragmatische Unterschiede hin, die der näheren Untersuchung bedürfen. Dazu kommt mit der distinktiven Kollexemanalyse (Gries/Stefanowitsch 2004) dieselbe analytische Methode noch einmal zum Einsatz.

Tabelle 3 zeigt die Resultate eines Kontrasts der Kollokate von *might* und *would* auf der Basis von Daten aus den 1860er Jahren. Für beide Modalverben zeigt die Tabelle jeweils die zehn Kollokate mit den größten Diskrepanzen zwischen beobachteter und erwarteter Häufigkeit, sämtliche Ergebnisse sind hochsignifikant.

Es fällt auf, dass mehrere distinktive Kollexeme von *would* die große Häufigkeit von konversationellen Stereotypen wie *I would like*, *it would seem*, oder *if you would permit* reflektieren. In diesen Wendungen hat *would* die Funktion, Distanz zwischen dem Sprecher und einem Sprechakt zu schaffen, und es somit dem Hörer zu erleichtern, unter Wahrung des Gesichts auf den Sprechakt zu reagieren. Die Tatsache, dass *would* häufig in dieser pragmatischen Funktion verwendet wird, ist hinlänglich dokumentiert (Coates 1983, S. 216).

Im Gegensatz zu *would* bewegt sich *might* eher im semantischen Bereich der logischen Möglichkeiten. Die distinktiven Kollexeme beinhalten Verben des Passierens (*happen*, *befall*) sowie mentale Verben (*judge*, *expect*, *see*, *prove*). Die Beispiele in (5) und (6) veranschaulichen die Unterschiede zwischen *might* und *would* noch etwas mehr. Während *might* in Kontexten verwendet wird, in denen es um die

Tab. 3: Die distinktiven Kollexeme von *might* und *would*, 1860er Jahre

MIGHT	OBS.	EXP.	COLL.STR	WOULD	OBS.	EXP.	COLL.STR
judge	13	3.36	7.65	like	457	352.34	37.89
happen	52	26.34	7.32	make	634	536.3	18.84
have	2825	2622.63	7.10	give	475	398.33	15.87
befall	12	3.1	7.06	seem	437	382.01	8.54
expect	27	10.85	6.75	permit	76	57.86	7.53
see	101	65.33	6.21	allow	77	59.34	6.80
get	80	49.32	6.02	do	554	504.4	5.43
be	3467	3275.96	5.73	pay	74	58.6	5.11
live	39	20.14	5.39	leave	114	94.2	5.00
prove	66	40.03	5.38	care	34	25.22	4.41

Möglichkeit eines Ereignisses geht, steht die Verwendung von *would* häufig im Zusammenhang mit dem emotionalen oder evaluativen Engagement des Sprechers.

- (5) a. Thus both parties were at liberty to act as they might judge best for the public interest.
 b. I was prepared for any thing which might happen.
 c. Then came the sudden picture of what might have been, and now.
 d. Who could tell what might befall her in the eventful voyage she was bound on?
 e. Practically the class was not so large as we might expect.
- (6) a. »I have a whim,« he said, dreamily, »that I would like to satisfy.«
 b. »Don't you think I would make a successful farmer, Mr. Morton?« he asked.
 c. »I would give a good deal to be out of this scrape,« continued Paul.
 d. The young gentleman, it would seem, hardly knew his own heart.
 e. If you would permit me to advise, I would suggest that Count Tristan should remain undisturbed.

Beispiele wie diese legen nahe, dass der Unterschied zwischen *might* und *would* auf der x-Achse ein Kontinuum abbildet, das Biber (1988, S. 107) als Involviertheit bezeichnet. Textsorten unterscheiden sich hinsichtlich ihrer Involviertheit, wobei stark involvierte Texte mit informierenden Texten kontrastieren. Informierende Texte sind vom Sprecher distanziert und beinhalten präzise Angaben, die in hoher Dichte präsentiert werden. Involvierte Texte lassen sich als interaktiv und affektiv beschreiben. Könnte es also sein, dass *would* ein Kennzeichen involvierter Texte ist, während *might* eher in informierenden Texten auftritt? Frequenzdaten aus dem COCA Corpus (Davies 2009), einer großen Datenbank, die verschiedene Textsorten des gegenwärtigen Amerikanischen Englisch repräsentiert. Daten aus dem

COCA Corpus zeigen, dass manche Modalverben eher in informierenden Textsorten zu finden sind, während andere Modalverben eher in involvierten Texten häufig sind. Eine Gegenüberstellung von Daten aus den Textsorten ›spoken‹ und ›academic writing‹, die auf dem Kontinuum der Involviertheit entgegengesetzte Pole einnehmen, zeigt, dass die relativen Häufigkeiten der neun Modalverben in den beiden Textsorten signifikant mit den Werten der x-Achse aus Schaubild 3 korrelieren. Eine lineare Regression, hier veranschaulicht durch die gestrichelte Linie, zeigt, dass Elemente die weiter rechts auf der x-Achse von Schaubild 3 liegen, typischerweise eher in involvierten Texten zu finden sind ($p = 0.02$, $R^2 = 0.56$). Die Modalverben *will* und *could*, die in Schaubild 3 eher in der rechten Hälfte der Karten zu finden sind, sind ebenfalls eher in gesprochenen Texten häufig. Die Modalverben *must*, *may*, und *shall* dahingegen haben eine höhere relative Frequenz in akademischen Texten. Damit zeigt sich, dass die Dimension der Involviertheit von Texten eine sinnvolle erste Annäherung an die x-Achse der Karten in Schaubild 3 darstellt. Die starke Streuung der Modalverben *must*, *may*, *might*, *should*, *can*, und *could* im mittleren Bereich von Schaubild 4 deutet freilich an, dass es weitere Unterscheidungen gibt, die es zu untersuchen gilt.

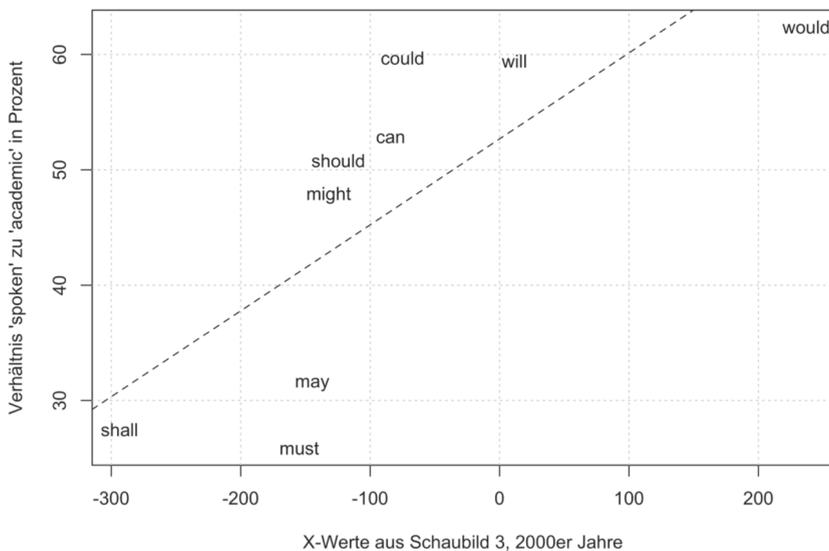


Schaubild 4: Das Verhältnis der Häufigkeiten der Modalverben in gesprochenen und akademischen Texten korreliert mit x-Werten aus Schaubild 3

Zusammenfassend lässt sich festhalten, dass das Daumenkino der sich wandelnden englischen Modalverben sich in einer Weise analysieren lässt, die Anschluss an bestehende theoretische Konzepte erlaubt. Die Karten in Schaubild 3 repräsentieren das Spannungsfeld zwischen deontischer und epistemischer Modalität auf der einen

Seite und die Affinitäten der Modalverben mit verschiedenen Textsorten auf der anderen. Entlang dieser beiden Dimensionen zeigen die Karten gleichermaßen Stabilität und Wandel. Semantische und pragmatische Stabilität kann insbesondere für die Elemente *would*, *could*, und *can* konstatiert werden. Diese Modalverben halten ihre relativen Positionen in dem sich wandelnden Paradigma weitestgehend konstant. Insgesamt lassen sich keine Wandel feststellen, nach denen ein Modalverb sich von links nach rechts entwickeln würde. Eine solche Entwicklung ließe sich als Kolloqualisierung erklären. Umgekehrt lässt sich die Bewegung von *shall* in eine zunehmend isolierte Position am linken Rand der Karten als eine Nischenbildung sehen: *shall* wird zunehmend nur noch in stark formalen Kontexten verwendet. Die wichtigste Entwicklung in Schaubild 3 jedoch betrifft die y-Achse. Wie bereits diskutiert entwickelt *may* eine stärkere Ausprägung epistemischer Bedeutungen und bewegt sich in diesem Zuge aufwärts. Entwicklungen in entgegengesetzter Richtung sind nicht erwartet, da epistemische Bedeutungen zwar aus deontischen hervorgehen, der umgekehrte Fall aber als typologisch unüblich eingeschätzt wird (Narrog 2005). Auch hier decken sich die beobachteten Entwicklungen also mit theoretischen Annahmen, die unabhängig von der Studie selbst formuliert wurden.

4. Abschluss und Ausblick

Die Untersuchungen in diesem Beitrag haben hoffentlich gezeigt, dass der Pessimismus von Leonard Bloomfield durch eine vorsichtig optimistische Grundhaltung ersetzt werden kann. Die aktuell verfügbaren korpuslinguistischen Ressourcen, die sowohl textueller als auch technischer Natur sind, erlauben die Beobachtung von vielen Phänomenen des Sprachwandels in einem Ausmaß, das vor 80 Jahren nicht vorhersehbar war. Der vorliegende Beitrag hat sich die englischen Modalverben zum Untersuchungsgegenstand gemacht und als Frequenzmaß die Häufigkeiten von Kollokaten in diachronen Korpusdaten ermittelt. Anhand zweier Daumenkino-Analysen konnten Beobachtungen gemacht werden, die sich teilweise mit bestehenden Befunden decken, zumeist aber in ihrer Detailschärfe über sie hinausgehen. Ein Nachweis für den praktischen Nutzen eines sprachlichen Daumenkinos sollte damit erbracht sein. In den abschließenden Zeilen dieses Beitrags sei nun noch auf zwei Desiderate für weiterführende Forschung hingewiesen.

Zum einen steht außer Frage, dass in Bezug auf Wandel in Textfrequenzen schon vieles von der diachron ausgerichteten Korpuslinguistik geleistet worden ist. Was noch aussteht, und in Zukunft stärker vorangetrieben werden könnte, ist ein Ausbau von korpuslinguistischen Studien, die mit dem vollen Spektrum erhebbarer Frequenzdaten arbeiten. Die Beiträge dieses Themenheftes veranschaulichen, dass Frequenz ein Begriff ist, der sehr verschiedene Aspekte der Häufigkeit sprachlicher Formen unter sich vereint. Alle Arten von Frequenz eignen sich für diachrone Analysen, so dass sich hier ein weites Feld für zukünftige Forschung öffnet.

Zweitens ist es wünschenswert, dass sich die Analyse von Frequenzdaten nicht als eigener Zweig der Sprachwissenschaft begreift, sondern vielmehr als integraler

Teil einer Forschungspraxis, die auf Theoriebildung ausgelegt ist. Messen lässt sich vieles, aber interessant sind diese Messungen in erster Linie, wenn sie eine bestehende theoretische Annahme bekräftigen oder aber in Frage stellen. Dass eine solche Ausrichtung Raum für explorative Untersuchungen lässt versteht sich dabei von selbst; die Fallstudien in diesem Beitrag sind ein Beispiel dafür. Als theoretischer Rahmen für diese Untersuchungen bietet sich vor allem die gebrauchsbasierte Konstruktionsgrammatik an (Goldberg 1995; 2006), da der Frequenz in dieser Theorie über das Konzept des ›entrenchment‹ eine zentrale Rolle zukommt. Wie genau die Analysemethoden der diachronen Korpuslinguistik für die Theorie der Konstruktionsgrammatik fruchtbar gemacht werden können ist Gegenstand der Diskussion in Hilpert (2011b i.Dr.). Aspekte der Frequenz haben für dieses Unterfangen eine nicht zu unterschätzende Bedeutung.

Summary

A ›Flipbook‹ of the English Modals. Visualizing Language Change in a Dynamic Way

How can frequency data from diachronic corpora facilitate research into processes of language change? This paper uses an approach that dynamically visualizes complex patterns of language change (Hilpert 2011a). This technique can be likened to a ›flipbook‹ of language change: a sequence of graphs visualizes a linguistic phenomenon during successive periods of real time. The purpose of this approach is to transform frequency data into a format that facilitates the explorative analysis of change and the formation of research hypotheses. The studies in this paper are based on the ›Corpus of Historical American English‹ (Davies 2010), which contains authentic language use from the past two centuries. The case studies presented here focus on current changes in the domain of the English modal auxiliaries.

Literatur

- Biber, Douglas: *Variation across speech and writing*. Cambridge 1988.
- Biber, Douglas/Johansson, Stig/Leech, Geoffrey/Conrad, Susan/Finegan, Edward: *The Longman grammar of spoken and written English*. Essex 1999.
- Bloomfield, Leonard: *Language*. London 1933.
- Coates, Jennifer: *The semantics of the modal auxiliaries*. London 1983.
- Croft, W./Poole, K. T.: »Inferring universals from grammatical variation: multidimensional scaling for typological analysis«. In: *Theoretical Linguistics* 34 (2008), S. 1–37.
- Davies, M.: The Corpus of Contemporary American English (COCA): 400+ million words, 1990-present. In: <http://www.americancorpus.org> (2008).
- Davies, M.: The Corpus of Historical American English (COHA): 400+ million words, 1810–2009. In: <http://corpus.byu.edu/coha> (2010).
- Facchinetti, Roberta/Krug, Manfred/Palmer, Frank R. (Hg.): *Modality in Contemporary English*. Berlin 2000.
- Firth, John R.: »A synopsis of linguistic theory, 1930–1955«. In: *Studies in Linguistic Analysis*. Special Volume (Philological Society). Oxford 1957, S. 1–32.
- Goldberg, Adele E.: *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago 1995.

- Goldberg, Adele E.: *Constructions at Work: The Nature of Generalization in Language*. Oxford 2006.
- Gries, Stefan Th./Stefanowitsch, Anatol: »Extending collostructional analysis: A corpus-based perspective on ›alternations««. In: *International Journal of Corpus Linguistics* 9/1 (2004), S. 97–129.
- Hilpert, Martin: »Dynamic visualizations of language change: Motion charts on the basis of bivariate and multivariate data from diachronic corpora«. In: *International Journal of Corpus Linguistics* 16/4 (2011a), S. 435–461.
- Hilpert, Martin: »Was ist Konstruktionswandel?«. In: Alexander Ziem und Alexander Lasch (Hg.): *Konstruktionsgrammatik III: Aktuelle Fragen und Lösungsansätze*. Tübingen 2011b, S. 59–76.
- Hilpert, Martin: *Constructional Change in English: Developments in Allomorphy, Word Formation, and Syntax*. Cambridge (i. Dr.).
- Kohnen, Thomas: »Historical corpus linguistics: perspectives on English diachronic corpora«. In: *Anglistik*, 17/2 (2006), S. 73–91.
- Krug, Manfred: *Emerging English modals: A corpus-based study of grammaticalization*. Berlin 2000.
- Kruskal, J./Wish, M.: *Multidimensional Scaling*. London 1978.
- Labov, William: *Principles of Linguistic Change. Volume 1: Internal Factors*. Oxford 1994.
- Leech, Geoffrey: »The modals ARE declining: Reply to Neil Millar's ›Modal verbs in TIME: frequency changes 1923–2006««. In: *International Journal of Corpus Linguistics* 16/4 (2011), S. 547–564.
- Majid, A./Boster, J. S./Bowerman, M.: »The cross-linguistic categorization of everyday events: A study of cutting and breaking«. In: *Cognition* 109/2 (2008), S. 235–250.
- Mair, Christian: *Twentieth-Century English: History, Variation and Standardization*. Cambridge 2006.
- Mair, Christian/Leech, Geoffrey/Hundt, Marianne/Smith, Nicholas: *Change in Contemporary English: A Grammatical Study*. Cambridge 2008.
- McDonald, Scott/Ramscar, Michael: »Testing the distributional hypothesis: The influence of context on judgements of semantic similarity«. In: *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (2001), S. 611–616.
- Millar, Neil: »Modal verbs in TIME: frequency changes 1923–2006«. In: *International Journal of Corpus Linguistics* 14/2 (2009), S. 191–220.
- Narrog, Heiko: »Modality, mood, and change of modal meanings: A new perspective«. In: *Cognitive Linguistics* 16/4 (2005), S. 677–731.
- Pado, Sebastian/Lapata, Mirella: »Dependency-based Construction of Semantic Space Models«. In: *Computational Linguistics* 33/2 (2007), S. 161–199.
- Szmrecsanyi, Benedikt/Kortmann, Bernd: »The morphosyntax of varieties of English worldwide: a quantitative perspective«. In: *Lingua* 119/11 (2009), S. 1643–63.
- Wheeler, E. S.: »Multidimensional Scaling for Linguistics«. In: R. Koehler, G. Altmann, R. G. Piotrowski (Hg.): *Quantitative Linguistics. An International Handbook*. Berlin 2005, S. 548–553.