

REINHARD FIEHLER

Korpusbasierte Analyse von Univerbierungsprozessen

Abstract

Vorgestellt werden Ziele und erste Ergebnisse des Projektes „Univerbierung“ am Institut für Deutsche Sprache. Das Projekt untersucht in verschiedenen Korpora, ob sich Prozesse der Univerbierung quantitativ belegen lassen. In Form von Univerbierungsprofilen sollen Univerbierungsverläufe dargestellt werden, d.h. die quantitativen Veränderungen, die zeitlich im Verhältnis der Getrennt- und Zusammenschreibungen eintreten (Kap. 1 und 2). Zugleich wird dabei methodologisch reflektiert, ob und inwieweit diese Korpora für solche Untersuchungen geeignet sind (Kap. 3). Exemplarisch werden einige Univerbierungsprofile vorgestellt (Kap. 4). Es handelt sich zum einen um Beispiele, bei denen sich die Normlage im Zuge der Rechtschreibreform nicht geändert hat, und zum anderen um solche, bei denen sie im Untersuchungszeitraum (1985-2008) verändert wurde. Die Untersuchungen zielen in der Perspektive darauf ab, Faktoren herauszuarbeiten, die Univerbierungsprozesse fördern bzw. hemmen, und aufzuklären, was Schreiber(-innen) als *ein* Wort gilt. Dies kann dazu beitragen, empirisch gestützt Komponenten des Wortkonzepts zu ermitteln (Kap. 5).

1. Univerbierung

Unter Univerbierung wird hier mit Bußmann (1983: 563) Folgendes verstanden:

In der Wortbildung Vorgang und Ergebnis des Zusammenwachsens mehrgliedriger syntaktischer Konstruktionen zu einem Wort, z.B. *ob + schon* zu *obschon*.

Univerbierung ist damit ein Prozess, der zur Bildung neuer Lexeme führt (Lexikalisierung). Zugleich ist sie in vielen Fällen auch ein Prozess der Grammatikalisierung, wenn das neue Lexem einer anderen grammatischen Kategorie angehört als die Ausgangsbestandteile (*auf Grund* → *aufgrund*: Präpositionalphrase → Präposition).

Löst man die Metapher des Zusammenwachsens auf, so ist Univerbierung zum einen ein mentales Phänomen: Sprachliche Elemente werden nicht mehr als getrennt empfunden bzw. verstanden, sondern als zusammengehörig. Dies manifestiert sich nach außen und für andere erkennbar in der Zusammenschreibung (grafische Univerbierung) von Elementen, die bisher getrennt geschrieben wurden. In dieser Hinsicht betrachtet ist Univerbierung damit zum

anderen ein schriftsprachliches Phänomen, dessen Realisierung und Erkennbarkeit die Trennung von Wortformen durch Spatien zur Voraussetzung hat. Werden zwei Wörter, die bisher mehrheitlich durch ein Spatium getrennt wurden, nicht mehr getrennt-, sondern zusammengeschrieben, so liegt Univerbierung vor.¹ In der Spatiensetzung kommt zum Ausdruck, was Schreiber als *ein* Wort empfinden. Univerbierung reflektiert, dass bestimmte Wortfolgen im Zuge des Sprachwandels nicht mehr als solche, sondern zunehmend als *ein* Wort empfunden werden. Dabei unterliegen nicht alle Wortfolgen einer Tendenz zur Univerbierung, sondern nur bestimmte, d.h. es muss Faktoren geben, die die Univerbierung zweier Wortformen befördern, und solche, die ihr entgegenstehen.

Univerbierung ist ein diachroner Prozess, er führt von einer mehrheitlichen Getrennt- zu einer mehrheitlichen Zusammenschreibung.² Synchron betrachtet ist sein Resultat eine systematische Varianz: Die Koexistenz von getrennt- wie zusammengeschriebenen Formen (in einem sich zeitlich verändernden quantitativen Verhältnis). Univerbierung kann in Bezug auf einzelne Personen, auf Textsorten oder in Hinblick auf eine Sprach- bzw. Schreibgemeinschaft als Ganzes betrachtet werden. Univerbierung ist zudem ein fortwährender historischer Prozess, der immer neue Wortfolgen erfasst.

[...] dass im Deutschen seit längerer Zeit die Tendenz zu beobachten ist, bei bestimmten Typen von Wortgruppen, die eine relativ abgegrenzte, einheitliche Gegebenheit der objektiven Realität (Gegenstand, Eigenschaft, Vorgang, Beziehung) benennen und damit eine den Einwortlexemen ähnliche Nominationsfunktion übernehmen, auch die geschriebenen Formative denen von Einwortlexemen anzugleichen und sie in eine geschlossene graphische Wortform zu überführen (graphische Univerbierung) (Nerius et al. 2000: 163)

Viele in der Schriftsprachgeschichte entstandene Univerbierungen werden heute nicht oder kaum mehr als solche erkannt (*seinerzeit, zufrieden, immerhin, durchaus* etc.). Analoge Prozesse finden sich nicht nur in der deutschen Schriftsprache, sondern auch in anderen Sprachen.

Getrennt- und Zusammenschreibung unterliegen in vielen Bereichen einer Normierung durch die kodifizierte Rechtschreibung. Die faktische Schreibung ist jedoch nicht allein durch die Norm bestimmt, sondern eine Resultante aus

¹ Ehemals nominale Bestandteile werden bei der Univerbierung – falls vorhanden – unter Vermeidung von Binnenmajuskeln kleingeschrieben.

² Zu dem der Univerbierung komplementären Prozess der Desintegration, der Getrenntschreibung von Bestandteilen eines bisher überwiegend zusammengeschriebenen Wortes (z.B. *Besucher Parkplatz*), vgl. Scherer (2010).

drei Faktoren: dem intuitiven Konzept (Sprachgefühl) des Schreibers, was *ein* Wort ist, der Normkenntnis des Schreibers und dem Willen des Schreibers, die Norm zu beachten.

Das primäre Interesse des Projekts gilt den faktischen Schreibungen, so wie sie in verschiedenen Korpora vorfindbar sind.³ Die Normierung der Getrennt- und Zusammenschreibung ebenso wie Veränderungen dieser Normierung (z.B. im Rahmen der Rechtschreibreform) sind nur insoweit von Interesse, wie sie als – sicherlich zentrale – Faktoren diese Schreibungen mitbeeinflussen.

2. Ziele des Projekts

Motivation für die Beschäftigung mit Univerbierungsprozessen im Rahmen des Projekts „Univerbierung (KG-U)“⁴ am Institut für Deutsche Sprache war zum einen die Erfahrung, dass sich die Veränderungen der Normen im Bereich der Getrennt- und Zusammenschreibung im Zuge der Rechtschreibreform, die ohne hinreichende Berücksichtigung faktischer Sprachwandelprozesse erfolgten, als problematisch erwiesen haben (vgl. Günther 1997, Bredel / Günther 2000), zum anderen aber auch eine punktuelle Beobachtung wie die, dass die Univerbierung *ausversehen* für mich völlig überraschend und von niemandem (außer den Schreiber(-innen)) beachtet offensichtlich dabei ist, Platz zu greifen: So sind bei Google Web mehr als ein Viertel der Schreibungen von *aus_Versehen*⁵ univerbiert. Eine Vergleichsanalyse von Google Web, dem Deutschen Referenzkorpus (DeReKo) und dem Korpus des Digitalen Wörterbuchs der Deutschen Sprache (DWDS) ergab folgendes Bild:

- Google Web (2.9.09)

<i>ausversehen</i> :	196 000 Vorkommen	(26,7%)
<i>aus Versehen</i> :	734 000 Vorkommen	
- DeReKo (2.9.09)

<i>ausversehen</i> :	15 Vorkommen	(0,25%)
<i>aus Versehen</i> :	6 085 Vorkommen	

³ Dieser Aspekt ist in der Literatur bisher wenig beachtet worden. Univerbierung wird dort vor allem unter den Gesichtspunkten Normierung (z.B. Herberg/Baudusch1989), Theorie des Wortes (z.B. Gallmann 1999, Weinberger 2005), Wortbildung (z.B. Eisenberg 2004, Fuhrhop 2007), Systemhaftigkeit (z.B. Jacobs 2005, 2007) und Sprachgeschichte (z.B. Nerius et al. 2000) behandelt.

⁴ Mitarbeiter(-innen): Reinhard Fiehler, Kerstin Güthert. Das Projekt ist eine Vorstudie für das geplante Projekt „Grammatische Variation im standardnahen Deutsch (Korpusgrammatik)“.

⁵ Zur Kennzeichnung, dass sowohl auf die univerbierten wie auch auf die nicht univerbierten Schreibungen referiert wird, wird das Spatium mit einem Unterstrich versehen.

- DWDS (2.9.09)
ausversehen: 0 Vorkommen (0%)
aus Versehen: 235 Vorkommen

Diese gravierenden Unterschiede verweisen deutlich auf die Bedeutsamkeit, die Korpora und ihrer je spezifischen Zusammensetzung für solche Untersuchungen zukommt.

Aufgrund dieser Vorüberlegungen ergaben sich für das Projekt folgende Zielsetzungen:

- Sichtung und Prüfung deutschsprachiger Korpora auf ihre Eignung für die quantitative Untersuchung von Sprachwandelprozessen;
- Grammatische Kategorisierung von Univerbierungen und Bildung von Fallklassen;
- Korpusbasierte quantitative Beschreibung von Univerbierungsverläufen (Erarbeitung von Univerbierungsprofilen);
- Systematisierung univerbierungsfördernder und -hemmender Faktoren;
- Analyse des Einflusses von Normänderungen (im Zuge der Rechtschreibreform) auf Univerbierungsverläufe;
- Erarbeitung von Empfehlungen für die Normierung der Getrennt- und Zusammenschreibung auf empirischer Basis;
- Entwicklung eines empirisch basierten Wortkonzepts auf der Grundlage der univerbierungsfördernden und -hemmenden Faktoren.

3. Korpora

Im Rahmen des Ziels „Sichtung und Prüfung deutschsprachiger Korpora auf ihre Eignung für die quantitative Untersuchung von Sprachwandelprozessen“ wurde am Beispiel von *statt_dessen* (und weiteren hier nicht dokumentierten Beispielen) eine Reihe frei zugänglicher Korpora getestet.⁶ Bei den Korpora handelt es sich um das Deutsche Referenzkorpus (DeReKo) des IDS⁷, das

⁶ Für einen ähnlichen Korpusvergleich (allerdings zu anderen Zwecken) vgl. Stuyckens/Brône (2009).

⁷ Da im DeReKo ab 1985 die Anzahl der Wortformen pro Jahr sehr deutlich größer ist als im Zeitraum zuvor, liegt es nahe, den Zeitraum 1985-2008 als gesondertes Teilkorpus zu betrachten. Im Zeitraum vor 1985 gibt es lediglich in den Jahren 1949, 1954, 1959, 1964, 1969 und 1974 überdurchschnittliche Anzahlen von Wortformen pro Jahr. Diese sechs Messpunkte wurden als zweites DeReKo-Teilkorpus untersucht.

Digitale Wörterbuch der Deutschen Sprache (DWDS), das Korpus C4⁸, das Schweizer Textkorpus (CHTK), die Datenbank gesprochenes Deutsch (DGD) des IDS, Google Web und Google Bücher⁹. In der folgenden Tabelle sind jeweils die Gesamtzahl der im betreffenden Korpus enthaltenen Wortformen, die Anzahl der Vorkommen von *statt_dessen*, der Anteil der Univerbierungen an der Gesamtzahl der Vorkommen und die Frequenz von *statt_dessen* (Vorkommen / Million Textwörter) angegeben (4.9.2009):

	Wortformen	Vorkommen <i>statt_dessen</i>	Anteil Univerbierungen	Frequenz Vork. / Mio
DeReKo 1949-1974	3,4 Mio	93	12 %	27,4
DeReKo 1985-2008	3219 Mio	153 050	63 %	47,6
DWDS	100 Mio	1338	19 %	13,4
C4	45,8 Mio	406	14 %	8,9
CHTK	20 Mio	116	22 %	5,8
DGD	5,8 Mio	4	0 %	0,7
Google Web	?	5 900 000	77 %	?

Die Untersuchungsergebnisse zu *statt_dessen* in den einzelnen Korpora möchte ich an dieser Stelle nicht im Detail diskutieren. Betrachtet man die Ergebnisse im Überblick, so wird deutlich, dass die untersuchten Korpora in jeder Hinsicht gewaltige Unterschiede aufweisen. Wie auch immer die unterschiedlichen Resultate im Einzelnen zu erklären sind, machen sie doch überaus deutlich, dass man sich bei Untersuchungen der angestrebten Art nicht auf ein Korpus wird beschränken können.

Geht man von einer statistisch relevanten Untergröße von 1 000 Vorkommen aus, so reduziert sich für *quantitative Untersuchungen* das Feld der Korpora auf Google Web, DeReKo 1985-2008 und DWDS. Für *quantitative Untersuchungen von zeitlichen Veränderungen* kommen nur DeReKo 1985-2008 und DWDS infrage. Google Web lässt (mit Ausnahme der Abfrage der Vorkommen im

⁸ Gemeinsames Korpus des DWDS, des Austrian Academy Corpus (AAC), des Korpus Südtirol und des Schweizer Textkorpus (CHTK).

⁹ Google Bücher wird im Weiteren wegen der noch unausgereiften Recherchemöglichkeiten nicht berücksichtigt. Vgl. Brückner (2009).

letzten Jahr) keine zeitliche Differenzierung zu.¹⁰ Auch wenn DeReKo in quantitativer Hinsicht keine Wünsche offenlässt, so besteht sein Nachteil doch darin, dass, wenn man hinreichend große Anzahlen von Wortformen und Vorkommen pro Jahr zur Bedingung macht, nur ein für Sprachwandelprozesse sehr kleiner Zeitraum von 23 Jahren (1985-2008) betrachtet werden kann. Ein weiterer Nachteil ist, dass DeReKo ganz überwiegend aus Zeitungstexten besteht und nur eine geringe Textsortendiversifikation aufweist. DWDS überspannt den wesentlich größeren Zeitraum von 100 Jahren (1900-2000), jedoch liegt seine Größe und entsprechend die Anzahl der Vorkommen für gegenwärtige Ansprüche an der unteren Grenze des quantitativ Erforderlichen.

4. Beispiele für Univerbierungsprofile

Im Folgenden werden nun exemplarisch einige Univerbierungsverläufe in DeReKo 1985-2008 im Detail betrachtet.

4.1 *statt_dessen*

Die Recherche für *statt_dessen* im DeReKo 1985-2008 ergibt folgende Ergebnisse (4.9.2009):

Bei 3 219 Mio. Textwörtern finden sich 153 050 Vorkommen. Davon zeigen 57 021 Getrennschreibung, 96 029 (= 63%) Zusammenschreibung. Die Frequenz beträgt 47,6 Vorkommen / Million Textwörtern.

Fasst man die Jahresergebnisse zu 3-Jahres-Zeiträumen zusammen, so ergeben sich folgende Anteile von Zusammenschreibungen:

1985-87: 36%	1997-1999: 31% ¹¹
1988-90: 26%	2000-2002: 79%
1991-93: 20%	2003-2005: 80%
1994-96: 19%	2006-2008: 95%

Norm (bis 1998): getrennt

Norm (ab 1998): zusammen

¹⁰ Anders ist dies bei Google Bücher, wo eine zeitliche Differenzierung über das Erscheinungsjahr erfolgen kann. Aber – wie gesagt – sind die Recherchemöglichkeiten dort im Moment noch nicht ausgereift.

¹¹ In den Zeitraum 1997-1999 fällt das Inkrafttreten der Rechtschreibreform 1998. Er ist damit gewissermaßen ein Scharnierzeitraum, wobei der Vergleich der Zeiträume davor und danach von besonderem Interesse ist.

Abbildung 1 zeigt diese Ergebnisse in grafischer Darstellung:

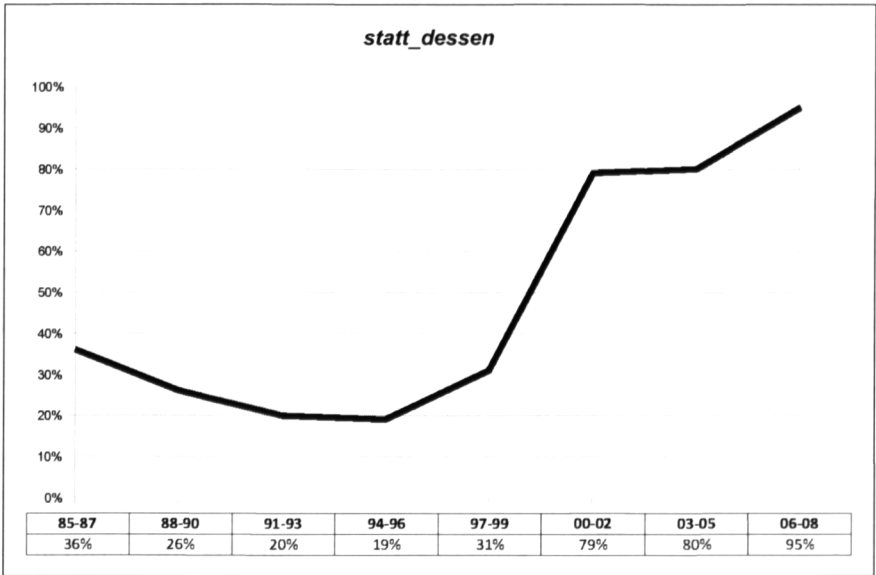


Abb. 1: Anteile der Univerbierungen von *statt_dessen* im DeReKo 1985-2008

In Zeitraum 1985-1996 ist eine kontinuierliche Abnahme der Anzahl der Zusammenschreibungen, also eine Entwicklung in Richtung auf die normentsprechende Getrennschreibung, festzustellen. Der Anteil der Zusammenschreibungen bewegt sich dabei aber auch noch am Ende des Zeitraums auf vergleichsweise hohem Niveau.

Die Normänderung katalysiert dann im Zeitraum 2000-2008 die Univerbierung bis hin zu einer fast ausschließlichen Zusammenschreibung.¹²

4.2 *genauso_gut*

Die Recherche für *genauso_gut* im DeReKo 1985-2008 ergibt folgende Ergebnisse (4.9.2009):

¹² Inwieweit diese Entwicklung durch Rechtschreibkorrekturprogramme mitverursacht ist, bleibt zu prüfen. DeReKo ist ganz überwiegend ein Zeitungskorpus. Man kann davon ausgehen, dass in den Zeitungsredaktionen solche Programme verwendet werden. Im konkreten Fall ist festzuhalten, dass beispielsweise der Duden-Korrektor die Getrennschreibung nicht als falsch ausweist.

Bei 3 219 Mio. Textwörtern finden sich 9 076 Vorkommen. Davon zeigen 6 672 Getrennschreibung, 2 404 (= 26%) Zusammenschreibung. Die Frequenz beträgt 2,8 Vorkommen / Million Textwörtern.

Die Anteile der Zusammenschreibungen in den 3-Jahres-Zeiträumen betragen:

1985-1987: 57%	1997-1999: 49%
1988-1990: 60%	2000-2002: 10%
1991-1993: 63%	2003-2005: 9%
1994-1996: 60%	2006-2008: 2%

Norm (bis 1998): zusammen

Norm (ab 1998): getrennt

Abbildung 2 zeigt diese Ergebnisse in grafischer Darstellung:

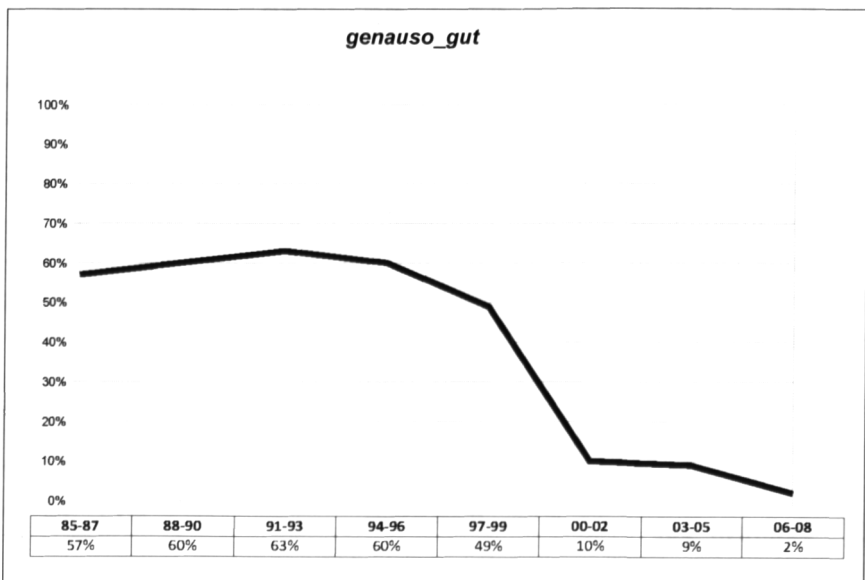


Abb. 2: Anteile der Univerbierungen von *genauso_gut* im DeReKo 1985-2008

Im Zeitraum 1985-1996 zeigt sich ein gleichbleibender Anteil (um 60%) der normentsprechenden Zusammenschreibung bei relativ hohem Anteil der Getrennschreibungen. D.h. *genauso_gut* zeigt keine hohe Tendenz zur Univerbierung.

Von 1997-2008 katalysiert dann die veränderte Norm die Getrenntschreibung.¹³

4.3 zu_Hause

Die Recherche für *zu_Hause* im DeReKo 1985-2008 ergibt folgende Ergebnisse (4.9.2009):

Bei 3 219 Mio. Textwörtern finden sich 282 640 Vorkommen. Davon zeigen 223 690 Getrenntschreibung, 58 959 (= 21%) Zusammenschreibung. Die Frequenz beträgt 87,8 Vorkommen / Million Textwörtern.

Die Anteile der Zusammenschreibungen in den 3-Jahres-Zeiträumen betragen:

1985-1987: 16,05%	1997-1999: 19,24%
1988-1990: 16,44%	2000-2002: 20,63%
1991-1993: 14,45%	2003-2005: 20,38%
1994-1996: 19,72%	2006-2008: 23,69%

Norm (bis 1998): getrennt

Norm (ab 1998): getrennt + zusammen möglich

Abbildung 3 zeigt diese Ergebnisse in grafischer Darstellung:

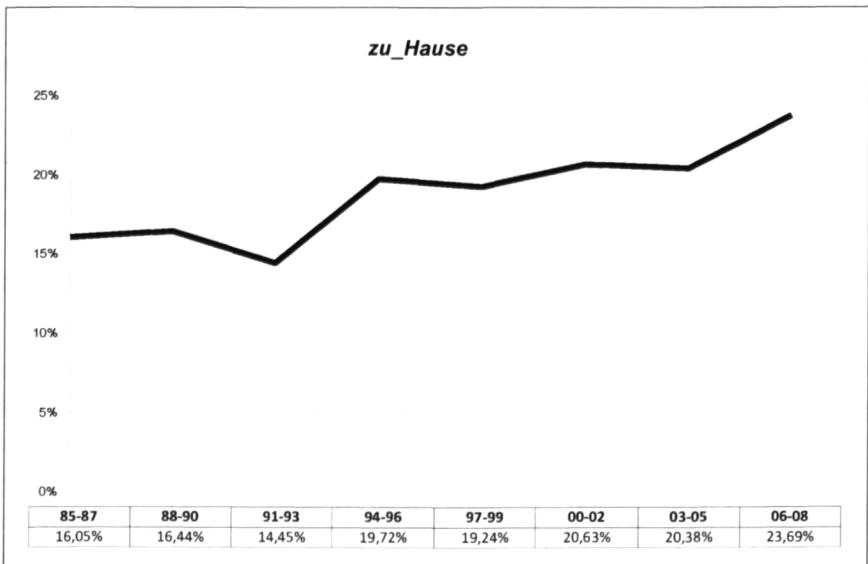


Abb. 3: Anteile der Univerbierungen von *zu_Hause* im DeReKo 1985-2008

¹³ Der Duden-Korrektor markiert die Zusammenschreibung.

Über den ganzen Zeitraum 1985-2008 findet sich ein durchgehender leichter Anstieg der Zusammenschreibungen. Man kann dies als leichte Tendenz zur Univerbierung deuten. Die Normliberalisierung hat keinen erkennbaren Einfluss auf diese Tendenz.

Ein Korpusvergleich mit Google Web erbringt folgende Ergebnisse (7.9.09):

Google Web:

27,2 Mio. Vorkommen: Davon 15,3 Mio. Getrennschreibungen und 11,9 Mio. (=44%) Zusammenschreibungen.

DeReKo 1985-2008:

282 640 Vorkommen: Davon 223 690 Getrennschreibungen und 58 950 (= 21%) Zusammenschreibungen.

Sowohl die Beobachtung, dass die Anzahl der Vorkommen bei Google Web ca. um den Faktor 100 größer ist, wie auch die Feststellung, dass der Anteil der Univerbierungen deutlich über dem Wert von DeReKo liegt (hier mehr als doppelt so groß ist), sind Befunde, die sich der Tendenz nach auch bei weiteren Recherchen immer wieder belegen lassen.

4.4 zu_Ende

Die Recherche für *zu_Ende* im DeReKo 1985-2008 ergibt folgende Ergebnisse (4.9.2009):

Bei 3 219 Mio. Textwörtern finden sich 134 556 Vorkommen. Davon zeigen 132 887 Getrennschreibung, 1 669 (= 1,24%) Zusammenschreibung. Die Frequenz beträgt 41,8 Vorkommen / Million Textwörtern.

Die Anteile der Zusammenschreibungen in den 3-Jahres-Zeiträumen betragen:

1985-1987: 4,74%	1997-1999: 1,58%
1988-1990: 4,78%	2000-2002: 0,59%
1991-1993: 2,04%	2003-2005: 0,98%
1994-1996: 1,42%	2006-2008: 0,73%

Norm (durchgehend): getrennt

Diagramm 4 zeigt diese Ergebnisse in grafischer Darstellung:

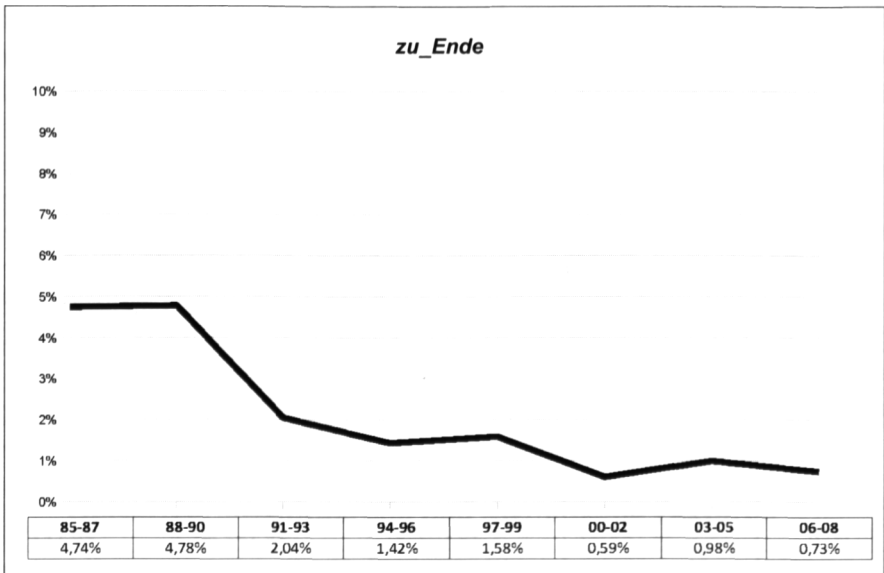


Abb. 4: Anteile der Univerbierungen von *zu_Ende* im DeReKo 1985-2008

Ein Korpusvergleich mit Google Web erbringt folgende Ergebnisse (7.9.2009):

Google Web:

5,287 Mio. Vorkommen: Davon 4,92 Mio. Getrennschreibungen und 0,367 Mio. (= 6,94%) Zusammenschreibungen.

DeReKo 1985-2008:

134556 Vorkommen: Davon 132887 Getrennschreibungen und 1669 (= 1,24%) Zusammenschreibungen.

Der Anteil der Univerbierungen von *zu_Ende* ist bei Google Web zwar ca. 5-fach größer, absolut ist er aber in beiden Korpora gering.

Zu_Ende gehört ebenso wie das eben behandelte *zu_Hause* in das Feld der *zu_** Wortfolgen, es zeigt aber deutlich andere Ergebnisse als *zu_Hause*. Über den ganzen Zeitraum 1985-2008 ist eine durchgehende Anpassung auf hohem Niveau an die normentsprechende Getrennschreibung festzustellen. *Zu_Ende* lässt keine Tendenz zur Univerbierung erkennen.¹⁴ Der Unterschied in den Ergebnissen lässt u.U. darauf schließen, dass *zu_Hause* zunehmend nicht mehr als Präpositionalphrase verstanden wird, sondern als adverbiale Ortsangabe

¹⁴ Der Duden-Korrektor markiert die Zusammenschreibung.

(analog zu bereits abgeschlossenen Univerbierungen wie *daheim* oder *hierzulande*), während *zu_Ende* nach wie vor als Präpositionalphrase mit einem Nomen als Kern gesehen wird.

4.5 *beieinander/ beisammen_**

Im Folgenden werden nicht einzelne Wortfolgen betrachtet, sondern die Felder, die *beieinander_** und *beisammen_** jeweils mit den zweiten Bestandteilen *bleiben*, *haben*, *liegen*, *sitzen* und *stehen* bilden. Die folgende Tabelle zeigt die Ergebnisse, die die Recherche im DeReKo in allen Korpora des Archivs W ergibt, im Überblick (8.9.2009):

	<i>beieinander_</i>		<i>beisammen_</i>	
	zusammen	getrennt	zusammen	getrennt
<i>bleiben</i>	11 (11%)	90 (89%)	40 (29%)	96 (71%)
<i>haben</i>	7 (8%)	79 (92%)	31 (9%)	317 (91%)
<i>liegen</i>	301 (21%)	1164 (79%)	42 (24%)	133 (76%)
<i>sitzen</i>	49 (28%)	127 (72%)	207 (58%)	151 (42%)
<i>stehen</i>	70 (26%)	195 (74%)	37 (45%)	46 (55%)

Norm vor 1998: zusammen
 Norm zwischen 1998 und 2006: *beieinander_*: getrennt
 beisammen_: zusammen
 Norm nach 2006: zusammen

Frappierendes Ergebnis ist, dass alle Wortfolgen dieser Felder (bis auf *beisammen_sitzen*) mehrheitlich getrennt geschrieben werden. Und dies, obwohl die Norm (vor 1998 und nach 2006 wieder) die Zusammenschreibung vorsieht. D.h. es ist zu vermuten, dass in diesen Wortfolgen Faktoren wirksam sind, die einer Univerbierung entgegenstehen und die die mehrheitliche Getrenntschreibung – auch gegen die Norm – befördern. Vermutlich ist es die hohe Silbenzahl, die die Zusammenschreibung erschwert.

beisammen_sitzen ist der einzige Fall in diesem Feld, bei dem die Zusammenschreibung überwiegt. Vermutlich ist dies auf die relativ frequente Existenz der Substantivierung *Beisammensitzen* (das nach den Belegen im DeReKo ganz überwiegend ein *gemütliches* ist) zurückzuführen, dessen Zusammenschreibung „abfärbt“.

Umgekehrt weisen *beieinander_haben* und *beisammen_haben* besonders hohe Anteile der Getrennschreibung auf (92% bzw. 91%). Dies mag daran liegen, dass *haben* in dem Feld *bleiben, liegen, sitzen* und *stehen* semantisch eine Sonderstellung einnimmt. *haben* ist weniger auf Personen bezogen, als es die anderen Zweitbestandteile sind. Diese Sonderstellung kann unter Umständen die Tendenz zur Getrennschreibung verstärken.

Die Analyse solcher Fälle, in denen in großen Anteilen bzw. sogar mehrheitlich gegen die Norm geschrieben wird, auf der einen Seite und die Untersuchung von quantitativen Unterschieden in Reihen (**_dessen*) bzw. Feldern (*zu_**; *beieinander/beisammen_**) auf der anderen Seite können Aufschluss geben über Faktoren, die eine Univerbierung entweder befördern oder behindern.

5. Univerbierungsfördernde und -hemmende Faktoren

Im Folgenden sollen einige mögliche Kandidaten für univerbierungsfördernde und -hemmende Faktoren benannt werden. Sie sind exemplarisch gemeint. Beim jetzigen Stand der Arbeit können sie nicht anders als hypothetisch und unvollständig sein. Wenn möglich wird auf Beispiele aus den vorstehenden Analysen zurückgegriffen.

Beispiele für mögliche univerbierungsfördernde Faktoren:

- Existenz von Substantivierungen (*beisammen_sitzen*)
- Vorläuferuniverbierungen (*indessen, unterdessen* etc. für die Reihe **_dessen*)
- Hauptakzent auf 2, 1 als Auftakt¹⁵ (*zu_Hause*)
- Hauptakzent auf 1 (*allzu_viel, vielmehr*)
- Hohe Kookkurrenz von 1 und 2 (*naja, um_Gottes_willen*)
- Ikonizität (*zusammenschreiben*)
- Auslaut 1 = Anlaut 2 (*warmmachen*)
- Intransparente Morphologie (*zuhauf, zulasten*)
- etc. etc.

Beispiele für mögliche univerbierungsbehindernde Faktoren:

- Hohe Silbenzahl von 1 (*beieinander_**)
- Semantische Andersartigkeit von 2 in einem 2-Feld (*beieinander_haben*)
- Semantische Eigenständigkeit von 1 und 2

¹⁵ 1 meint den ersten Bestandteil der Wortfolge, 2 den zweiten.

- Unterschiedliche Bedeutung von Zusammenschreibung und Getrenntschreibung (*sitzen bleiben* vs. *sitzenbleiben*)
- Ikonizität (*getrennt schreiben*)
- etc. etc.

Die univervierungsfördernden wie -behindernden Faktoren betreffen alle sprachsystematischen Ebenen. Die Entscheidungen, ob Schreiber(-innen) eine Wortfolge zusammen oder getrennt schreiben, sind – neben dem Normbezug – Resultante des Zusammenspiels dieser Faktoren. Sie sind Erscheinungsformen bzw. Ausdruck eines individuellen wie kollektiven Wortkonzepts. Die Explikation dieser Faktoren trägt dazu bei, zu operationalisieren, was Personen als *ein* Wort verstehen, und so ein empirisch fundiertes Wortkonzept zu erarbeiten.

Literatur

- Bredel, Ursula / Günther, Hartmut (2000): Quer über das Feld das Kopfadjunkts. Bemerkungen zu Peter Gallmanns Aufsatz *Wortbegriff und Nomen-Verb-Verbindungen*. In: Zeitschrift für Sprachwissenschaft 19: 103-110.
- Brückner, Dominik (2009): Die Google Buchsuche als Hilfsmittel für die Lexikographie. In: Sprachreport 25, 3: 26-31.
- Bußmann, Hadumod (1983): Lexikon der Sprachwissenschaft. Stuttgart: Kröner.
- Eisenberg, Peter (2004): Das Wort. Grundriss der deutschen Grammatik Bd. 1. 2. Aufl. Stuttgart / Weimar: Metzler.
- Fuhrhop, Nanna (2007): Zwischen Wort und Syntagma. Zur grammatischen Fundierung der Getrennt- und Zusammenschreibung. Tübingen: Niemeyer.
- Gallmann, Peter (1999): Wortbegriff und Nomen-Verb-Verbindungen. In: Zeitschrift für Sprachwissenschaft 18: 269-304.
- Günther, Hartmut (1997): Alles Getrennte findet sich wieder – Zur Beurteilung der Neuregelung der deutschen Rechtschreibung. In: Eroms, Hans W. / Munske, Horst H. (Hg.): Die Rechtschreibreform: Pro und Kontra. Berlin: Erich Schmidt, 81–93.
- Herberg, Dieter / Baudusch, Renate (1989): Getrennt oder zusammen? Ratgeber zu einem schwierigen Rechtschreibkapitel. Leipzig: Bibliographisches Institut.
- Jacobs, Joachim (2005): Spatien. Zum System der Getrennt- und Zusammenschreibung im heutigen Deutsch. Berlin/New York: de Gruyter.
- Jacobs, Joachim (2007): Vom (Un-)Sinn der Schreibvarianten. In: Zeitschrift für Sprachwissenschaft 26: 43-80.

- Nerius, Dieter et al. (2000): Deutsche Orthographie. Von einem Autorenkollektiv unter der Leitung von Dieter Nerius. 3. Aufl. Mannheim/Leipzig/Wien/Zürich: Dudenverlag.
- Scherer, Carmen (2010): Vom *Friseursalon* zum *Friseur Salon* – Desintegrationsprozesse in der Schreibung von Komposita. Ms.
- Stuyckens, Geert / Brône, Geert (2009): Brauchbarkeit von Korpora des geschriebenen Deutsch für DaF-Lehrende. Eine Fallstudie. In: Deutsch als Fremdsprache 46, 1: 3-9.
- Weinberger, Elke (2005): *Kopf stehen* oder *kopfstehen*? Versuch einer grammatischen Sichtung des Grundproblems der Getrennt- und Zusammenschreibung im Bereich der Nomen-Verb-Verbindungen. Seminararb. Univ. Zürich. Internet: www.ds.uzh.ch/lehrstuhlduerscheid/docs/seminararb/weinberger-gzs-05.pdf (Stand: 03/2010).