

Lisa Dücker

Die Getrennt- und Zusammenschreibung von Komposita in frnhd. Hexenverhörprotokollen

Abstract: This paper investigates the spelling of compound nouns in a corpus comprised of Early New High German protocols of witch trials from the 16th and 17th century. Previous studies on the spelling of compound nouns in printed texts have found that scribes increasingly write compound nouns as one word during the 16th century. However, this paper will show that there is still much variation in handwritten texts from that time. The study focusses on identifying factors that lead scribes to write compound nouns either as one word or two, such as linking elements and the use of upper case letters. I will argue that while there is more variation in the spelling of compound nouns in the handwritten corpus than in printed texts, there still is a strong tendency to line up the boundaries of the graphemic and syntactic words.

1. Einleitung

Die Zusammen- und Getrenntschreibung von Komposita in der Geschichte des Deutschen wurde in der Vergangenheit hauptsächlich an Drucken untersucht (vgl. u.a. Solling 2012, 2016; Kopf 2017; Schindler 1999; Pavlov 1972, 1983, 2004). Zu handschriftlichen Texten liegen hingegen vereinzelte Studien vor: So analysiert V. Moser (1982) zwei literarische Handschriften Fischarts. Haubold (1914) vergleicht wiederum handschriftliche Manuskripte und Bieldrucke von Martin Luther und stellt dabei unter anderem Folgendes fest: „Über die Schreibung der zusammengesetzten Worte ist zu sagen, daß sie im Druck genau so schwankend ist wie in der Handschrift“ (Haubold 1914: 24). Wie genau sich aber die Regularitäten bei der Worttrennung von Komposita in Handschriften und Drucken unterscheiden, darüber gibt es bisher keine gesicherten Ergebnisse.

Im Fokus des vorliegenden Artikels steht deshalb, ob die Kriterien, die für die Spatiensetzung in gedruckten, meist lektorierten Texten erarbeitet wurden, auch auf Handschriften zutreffen, und welche Unterschiede es gibt. Dies wird anhand eines Korpus von Hexenverhörprotokollen aus dem SiGS-Projekt¹

1 Das SiGS-Korpus ist im Rahmen des DFG-finanzierten Projekts zur „Entwicklung der satzinternen Großschreibung“ (kurz: SiGS), das in der ersten Phase an zwei Universitäten angesiedelt war (Uni Hamburg und Uni Münster), entstanden (SZ 280/2-1 und KO 909/12-1: Münsteraner Gruppe (Klaus-Michael Köpcke, Mark Schutzzeichel), Hamburger Gruppe (Fabian Barteld, Renata Szczepaniak). AnnotatorInnen: Annemarie Bischoff, Lisa Dücker, Julia Hübner, Johanna Legrum, Katja Politt, Nicolai Pudimat, Eleonore Schmitt, Annika Vieregge und Nicholas

untersucht. Diese Texte können als (semi-) spontan gelten und sind als Teil von Gerichtsakten für einen mittleren Grad von Öffentlichkeit bestimmt (vgl. Topalović 2003; Szczepaniak & Barteld 2016: 45).

In Kap. 2 wird zunächst der Reanalyseprozess von der Genitivphrase zum uneigentlichen Kompositum nachvollzogen und in Kap. 3 der Forschungsstand zur Schreibung von Komposita im Frnhd. zusammengetragen. Auf Grundlage des Konzepts des graphematischen Worts nach Fuhrhop (2008) werden in Kap. 4 alle Substantivkomposita aus 56 Hexenverhörprotokollen aus der Zeit von 1570 bis 1665 ausgewertet. Dabei wird der Zusammenhang zwischen Worttrennung und Majuskelsetzung (Kap. 4.6) ebenso beleuchtet wie der Einfluss von Fugenelementen (Kap. 4.5) und der Frequenz (Kap. 4.7).

2. Substantivkomposita im Frühneuhochdeutschen

Das Wortbildungsmuster der Komposition gewinnt im Frnhd. immer mehr an Bedeutung (vgl. Solling 2012: 37). Zu dem aus dem Germanischen ererbten Muster der sog. eigentlichen Komposition tritt nun verstärkt die sog. uneigentliche hinzu. Eigentliche Komposita werden gebildet, indem zwei Substantive mit Hilfe eines (fakultativen) Stammbildungselements verbunden werden (ahd. *sper-a-scaft*, *tag-a-lôn*). Dieses Wortbildungsmuster ist im Frnhd. nicht mehr produktiv; die Stammbildungselemente sind größtenteils ausgefallen oder zu Schwa abgeschwächt (frnhd. *speer-Ø-schaft*, *tag-e-lohn*, vgl. Szczepaniak 2016: 329).

Uneigentliche Komposita sind hingegen diachron betrachtet aus Genitivphrasen entstanden. Demske (1999, 2001) sieht die Entstehung der Determiniererphrase als Ursache für deren Wandel zum Substantivkompositum: Da die vorangestellten Genitivattribute den neu entstandenen Determiniererslot in der Phrase nicht ausfüllen können, setzt ein Stellungswandel ein. Genitivattribute werden zunehmend nachgestellt und die verbliebenen vorangestellten Attribute werden als Kompositionserstglieder reanalysiert:

[[*der Brücken*] *Zoll*] > [*der [Brücken Zoll]*] > [*des [Brücken Zolls]*] > [*des [Brückenzolls]*]

Durch diesen Entstehungsprozess der Komposita aus Genitivphrasen werden an der Grenze zwischen den beteiligten Substantiven ehemalige Flexionsendungen als Fugenelemente konserviert (*Brücke-n-zoll*, *Teufel-s-tanz*; vgl. Nübling & Szczepaniak 2011: 53–55).

Wieling). Die zweite Phase ist an der Universität Hamburg (Lisa Dücker, Stefan Hartmann, Renata Szczepaniak) angesiedelt (SZ 280/2-3). Ich möchte allen an dem Projekt Beteiligten für ihre Arbeit danken. Mein Dank gilt außerdem den anonymen Gutachtern, die mit hilfreichen Hinweisen zur Verbesserung dieses Artikels beigetragen haben.

Im Zuge dieses Prozesses ändern sich die Kongruenzverhältnisse innerhalb der Nominalphrase: Der Determinierer kongruiert nun nicht mehr mit dem ehemaligen Genitivattribut (*der Brücken*), sondern mit dem zweiten Substantiv (*des Zolls*), das als Kopf des Kompositums analysiert wird. Auch auf graphematischer Ebene schlägt sich die Reanalyse nieder, indem die ursprünglich voneinander getrennt geschriebenen Substantive vermehrt auch ohne Spatium vorkommen. Pavlov (2004) konstatiert für frnhd. Texte gar, dass dieses graphematische Kriterium für die Abgrenzung von Genitivkonstruktionen und Komposita hinreichend ist (ähnlich auch Nitta 1987; Solms 1999). Er nimmt an, dass die Schreibung als ein oder zwei graphematische Wörter ein direkter Ausdruck dessen ist, was ein frnhd. Schreiber als ein oder zwei Wörter *empfindet* (Pavlov 2004: 106):

Unserer Ansicht nach handelt es sich bei den Schreibungen wie *lant hërre* vs. *lanthërre* oder *shâf stîge* vs. *shâfstîge* um Wortgruppe auf der einen und Kompositum auf der anderen Seite, trotzdem die beiden Konstruktionsweisen im gegebenen Fall semantisch und funktional, als lexematische Einheiten, identisch sind. (Pavlov 2004: 106)

Wie Szczepaniak & Barteld (2016: 56) und Barteld, Szczepaniak & Zinsmeister (2014: 254–255) anhand der handschriftlichen Hexenverhörprotokolle zeigen, die auch der vorliegenden Untersuchung zugrunde liegen (s.u. Kap. 4.1), kann keineswegs eine 1:1-Zuordnung zwischen graphematischen und syntaktischen Wörtern für das Frnhd. gezogen werden. Sie verweisen darauf, dass diese Uneinheitlichkeit der Grenzen von syntaktischem und graphematischem Wort nicht nur bei Komposita, sondern auch bei Präposition-Artikel-Klisen und weiteren Phänomenen zu beobachten ist (vgl. <auffm Teufelß dantz>, Alme 1630²).

Auch wenn die Zusammen- oder Getrenntschreibung für die vorliegende Studie nicht als hinreichendes Kriterium für die Analyse als Kompositum oder Genitivphrase angesehen werden kann, ist die Schreibung ohne Spatium zweifelsohne ein wichtiger Schritt auf dem Weg von der Phrase hin zu der Analyse als ein nicht nur syntaktisches, sondern auch graphematisches Wort (vgl. Fuhrhop 2008). Im nächsten Kapitel soll deshalb der aktuelle Forschungsstand zur Entwicklung der Worttrennung bei Komposita und mögliche Einflussfaktoren zusammengetragen werden.

Die Unterteilung in eigentliche und uneigentliche Komposita ist nicht unproblematisch (vgl. Solling 2012: 34–36). Bei starken Feminina, die im Genitiv keine Endung zeigen, kann bspw. keine morphologisch basierte Unterscheidung zwischen Kompositionserstglied und Genitivobjekt getroffen

2 Bei Beispielen wird mit einer Chiffre aus Entstehungsort und -zeit nach Macha et al. (2005) auf den entsprechenden Korpus text referiert.

werden. In der vorliegenden Untersuchung wird deshalb in der Auswertung mit der Unterteilung in verfügte und unverfügte Komposita gearbeitet. Die Termini eigentliche und uneigentliche Komposita werden jedoch im Theorie-teil zum Vergleich mit anderen Studien herangezogen.

3. Frnhd. Komposita als graphematische Wörter

Ein graphematisches Wort ist laut Fuhrhop (2008: 193) eine ununterbrochene Graphemkette, die zwischen zwei Spatien steht. Prototypische graphematische Wörter erlauben zudem nur Großschreibung am Wortanfang (vgl. Fuhrhop 2008: 194), aber prinzipiell können sie Wortzeichen wie Bindestriche oder Apostrophe enthalten (z. B. *M'gladbach*, *Software-Entwickler*). Auch die Binnenmajuskelschreibung bei Komposita (mit oder ohne Bindestrich) führt laut Fuhrhop (2008: 225) nicht dazu, dass diese als zwei graphematische Wörter interpretiert werden: „Eine Schreibung wie *BahnCard* erscheint zwar einerseits als künstliche Schreibung, kann aber quasi als Zwischenstufe zwischen einfachen Komposita (*Bahnkarte*) und Bindestrichkomposita gesehen werden“ (Fuhrhop 2008: 225; Hervorhebungen im Original).

In Dücker & Szczepaniak (2017) wurde ein Modell mit verschiedenen Graden der graphematischen Wortwertigkeit von Komposita erarbeitet (Tabelle 1).

Tab. 1: *Graphematische Wortwertigkeit von Komposita (nach Dücker & Szczepaniak 2017: 49).*

zwei graphematische Wörter			ein graphematisches Wort		
mit Binnenmajuskel (und fakultativer Initialmajuskel)	ohne Majuskel	mit Initialmajuskel	mit Binnenmajuskel (und fakultativer Initialmajuskel)	ohne Majuskel	mit Initialmajuskel
<i>Teufels Tanz</i> <i>teufels Tanz</i>	<i>teufels</i> <i>tanz</i>	<i>Teufels tanz</i>	<i>TeufelsTanz</i> <i>teufelsTanz</i>	<i>teufelstanz</i>	<i>Teufelstanz</i>

Alle Komposita auf der rechten Seite der Darstellung können jeweils als ein graphematisches Wort gelten. Obwohl alle drei Unterkategorien mit dem syntaktischen Wort übereinstimmen, gibt es dennoch Abstufungen: Das graphematische Wort ohne Majuskel stellt im Deutschen das unmarkierte graphematische Wort dar. Eine Majuskelsetzung betont den Anfang eines graphematischen Worts. Eine Initialmajuskel hebt somit ein Kompositum besonders hervor, eine Binnenmajuskel wie in *TeufelsTanz* betont hingegen die einzelnen Bestandteile der Zusammensetzung und verschlechtert die Wahrnehmung als *ein* zusammengehöriges graphematisches und syntaktisches Wort.

Die Schreibung als zwei graphematische Wörter, die auf der linken Seite der Darstellung zu sehen ist, erschwert die Analyse als ein syntaktisches Wort.

Das gilt insbesondere dann, wenn die zweite Konstituente mit einer Majuskel beginnt. Großschreibung beim Bestimmungsglied hebt hingegen auch hier den Beginn des syntaktischen Worts besonders hervor.

Auch wenn Schreibungen auf der linken Seite der Tabelle mit frühen Entwicklungsstadien der Substantivkomposition assoziiert werden und die unmarkierte Schreibung im Nhd. diejenige ist, die ganz rechts steht, verläuft die graphematische Entwicklung nicht so gradlinig für alle Substantivkomposita, wie eine solche Darstellung vermuten lassen könnte. Stattdessen zeigen sich große Unterschiede zwischen den eigentlichen und uneigentlichen Komposita.

Noch für das 14. und die erste Hälfte des 15. Jhs. wird in bisherigen Studien ein hoher Anteil an Zusammenschreibung von eigentlichen Komposita dokumentiert.³ Ab dem Ende des 15. Jahrhunderts gilt dies jedoch nicht mehr für alle Substantivkomposita gleichermaßen: Pavlov (1972, 1983) zeigt für den Zeitraum von 1470–1530, dass eigentliche Komposita 81,5% Zusammenschreibung (Token) aufweisen. Die nun häufiger auftretenden uneigentlichen Komposita, entstanden aus Phrasen, werden noch häufiger getrenntgeschrieben; in Pavlovs (1983: 33–34) Daten werden sie nur in 33,4% der Belege zusammengeschrieben. Um 1700 finden sich wiederum insgesamt knapp 3% Getrenntschreibung, die mehrheitlich auf uneigentliche Komposita entfallen (Pavlov 1983: 105–106, 2009: 64–65).

Die Arbeiten von Solling (2012, 2016), der gedruckte Predigten aus den Jahren 1550–1710 untersucht, zeigen ein ähnliches Bild. Wie in Tab. 2 dargestellt, steigt bei den eigentlichen Komposita der Anteil der getrenntgeschriebenen Belege nie über 6%, während bis zu 45% der uneigentlichen Komposita mit *Spatium* geschrieben werden. Ab Zeitschnitt V (1660) kommt Getrenntschreibung bei allen Komposita nur noch selten vor (<7% pro Zeitschnitt), was auf den starken Anstieg der Bindestrichschreibung zurückzuführen ist.

3 Für einen Überblick über die Worttrennung vom Ahd. an vgl. Solling 2012: 73–98.

Tab. 2: Worttrennung bei *eigentlichen* und *uneigentlichen* Komposita 1550–1710 nach Solling (2016: 296).

	Eigentliche Komposita			Uneigentliche Komposita		
	zusammen	mit Bindestrich	getrennt	zusammen	mit Bindestrich	getrennt
Zeitschnitt I (1550)	94% (280)	0	6% (19)	55% (74)	0	45% (60)
Zeitschnitt II (1570)	97% (1046)	0	3% (35)	69% (346)	0	31% (152)
Zeitschnitt III (1600)	96% (603)	<0,5% (1)	4% (26)	79% (209)	0	21% (55)
Zeitschnitt IV (1620)	96,5% (631)	<0,5% (1)	3,5% (22)	88% (354)	0	12% (47)
Zeitschnitt V (1660)	85% (544)	13% (82)	2% (12)	53,5% (219)	40,5% (165)	6% (24)
Zeitschnitt VI (1710)	46% (382)	53,5% (444)	0,5% (4)	17,5% (101)	78,5% (448)	4% (22)

Sowohl bei *eigentlichen* als auch bei *uneigentlichen* Komposita sinkt somit der Anteil der Getrennschreibung ab der zweiten Hälfte des 16. Jhs. stark, *uneigentliche* Komposita zeigen über den gesamten Zeitraum hinweg jedoch wesentlich höhere Anteile von Spatiensetzung als die *eigentlichen*.

Kopf (2017: 197) sieht in ihrer Studie, die sich auf ein Korpus von kirchlichen Texten und Sachtexten aus dem Zeitraum von 1500 bis 1710 stützt, eine noch deutlichere Abnahme der Worttrennung bei verfügbaren Komposita: 1500 werden 81% aller Komposita mit Fugenelement getrenntgeschrieben, 1710 sind es nur noch 8%. Bei den unverfügbaren Komposita sinkt der Anteil an Getrennschreibung im gleichen Zeitraum von 22% auf 2% (vgl. Kopf 2017: 197).

Während nach relativ konstanter Zusammenschreibung bis zur Mitte des 15. Jhs. die Getrennschreibung von *eigentlichen* Komposita um 1500 Werte von knapp 20% erreicht, sinkt der Anteil im 16. und 17. Jh. bei beiden Klassen von Substantivkomposita so sehr, dass um 1700 kaum noch getrenntgeschriebene Komposita vorkommen. *Eigentliche* und *uneigentliche* Komposita nähern sich also in ihrer graphischen Gestalt einander immer mehr an. Dabei weisen *uneigentliche* Komposita in den bisherigen Studien durchweg einen höheren Anteil an Getrennschreibung auf.

Sowohl Pavlov (2004: 108) als auch Schindler (1999: 318) und Solling (2016: 300) machen das Aufkommen der *uneigentlichen* Komposita für die vorübergehende Zunahme der Getrennschreibung bei den *eigentlichen* Komposita verantwortlich. Pavlov (2004: 108) sieht die Gründe dafür in einer „gegenseitige[n] funktionale[n] Annäherung“ der beiden Kompositionstypen. Diese

führte einerseits zur breiteren Übernahme des getrennten Schriftbildes der Fügungen mit dem vorangestellten flexionslosen Attribut, andererseits aber kamen die genitivischen Komplexe funktional auf gleiche Linie mit den ‚Stammkomposita‘, was ihre Zusammenschreibung stimulierte. (Pavlov 2004: 108)

Auch Solling (2016) erklärt die Schwankungen bei der Worttrennung von Substantivkomposita in gedruckten Predigttexten mit wechselseitigen Analogiebeziehungen zwischen eigentlichen und uneigentlichen Komposita. Dass die uneigentlichen Komposita am Ende seines Untersuchungszeitraums nur noch 4% Getrenntschreibung aufweisen, wertet Solling (2016: 299) als ein Zeichen dafür, dass diese nun als Wortbildungsmuster etabliert sind und (wie eigentliche Komposita) zunehmend als ein syntaktisches Wort interpretiert werden:

In the first part of the period covered by the study [um 1550, L.D.], the artificial compound nouns [uneigentliche Komposita, L.D.] adapted to match the genuine ones [eigentliche Komposita, L.D.], but the opposite tendency can be observed over the last two thirds of the 17th century. (Solling 2016: 300)

Neben dem Auftreten von Fugenelementen ist die Majuskelsetzung beim Kopf entscheidend für die Getrennt- oder Zusammenschreibung von Komposita. Zwar kommen in den Daten von Solling alle Kombinationen von Getrennt-, Zusammen- und Bindestrichschreibung mit Groß- oder Kleinschreibung des Kopfs vor, doch werden Komposita mit kleingeschriebenem Kopf beinahe durchweg zusammengeschrieben, „während Substantivkomposita, in denen eine Binnenmajuskel vorhanden ist, dagegen in den früheren Zeitschnitten überwiegend getrennt geschrieben und in den späteren Zeitschnitten vorwiegend mit einem Bindestrich in der Fuge auftreten“ (Solling 2012: 179).

Weidman (1941: 96) stellt zudem für den von ihm untersuchten Codex Manesse fest, dass frequente Lemmata häufiger zusammengeschrieben werden als weniger frequente. Die Ergebnisse von Solling (2012: 267) zeigen hingegen, dass in den Predigttexten frequente Komposita nicht durchweg häufiger Zusammenschreibung aufweisen. Stattdessen schwanken Komposita, die in seinem Korpus häufig vorkommen, nicht so sehr in Bezug auf die Worttrennung wie weniger frequente Komposita, sondern zeigen entweder überwiegend Getrennt- oder überwiegend Zusammenschreibung. Diese beiden Beobachtungen können als Hinweise auf eine stärkere Konventionalisierung („Lexikalisierung“) der Schreibung frequenter Lemmata interpretiert werden. Im Gegensatz dazu werden Hapax Legomena häufiger getrennt als zusammengeschrieben (vgl. Solling 2012: 101, Pavlov 1983: 28–29).

Zur Zusammen- und Getrenntschreibung von Substantivkomposita speziell in Handschriften gibt es bisher nur wenige empirische Untersuchungen: V. Moser (1982: 418) stellt für die von ihm untersuchten Handschriftenbruchstücke fest, dass am Ende des 16. Jhs. „zusammenschreibung [sic] das

gewöhnliche“ ist und Getrenntschreibung bei eigentlichen wie uneigentlichen Komposita nur „vereinzelt“ belegt ist. Haubold (1914) findet in seiner Studie keinen signifikanten Unterschied in der Getrenntschreibung von Komposita zwischen den Drucken und Handschriften Martin Luthers. Folglich wäre davon auszugehen, dass in einem Korpus von handschriftlichen Texten aus dem Zeitraum vom letzten Drittel des 16. Jhs. bis zur Mitte des 17. Jhs. überwiegend mit Zusammenschreibung von Substantivkomposita zu rechnen ist.

4. Korpusstudie in frnhd. Hexenverhörprotokollen

Die Analyse der Schreibung von Komposita in Hexenverhörprotokollen zeigt insgesamt ein wesentlich uneinheitlicheres Bild, als auf Grundlage der bisherigen Forschung zu erwarten wäre. In Kap. 4.1. wird das SiGS-Korpus vorgestellt und in Kap. 4.2 und 4.3 dargelegt, wie Substantivkomposita sowie Getrennt- und Zusammenschreibung für die Korpusanalyse definiert bzw. operationalisiert wurden. Kap. 4.4 bietet einen Überblick über die Zusammen- und Getrenntschreibung der Komposita, bevor anschließend Faktoren für die Worttrennung überprüft werden, darunter das Auftreten von Fugenelementen (4.5), die Majuskelsetzung bei den Kompositionsgliedern (4.6) und Lexikalisierung häufiger Lemmata (4.7).

4.1. Das SiGS-Korpus

Abb. 1: SiGS-Gesamtkorpus.



Im Gegensatz zu bisherigen Analysen von Substantivkomposita wird die vorliegende Untersuchung in einem Korpus aus 56 handschriftlichen Hexenverhörprotokollen, die von Macha et al. (2005) ediert wurden (vgl. Abb. 1), durchgeführt. Die Texte stammen aus der Zeit zwischen 1570 und 1665 und sind nach Macha et al. (2005) in sechs Regionen eingeteilt: Nordwestdeutschland (Region I), Nordostdeutschland (Region II), Mittelwestdeutschland (Region III), Mittelostdeutschland (Region IV), Südwestdeutschland (Region V) und Südostdeutschland (Region VI). Für die Untersuchung der zeitlichen Entwicklung wurde eine Einteilung in Zeitschnitte von jeweils zehn Jahren gewählt (1570–1579, 1580–1589 etc.).⁴

Es handelt sich bei den Texten um Verhörmit- und -abschriften, die in der Regel von einem anonymen, aber geübten Schreiber verfasst wurden (vgl. Szczepaniak & Barteld 2016: 47–48).⁵ Sie enthalten zwischen 550 und 2848 graphische Token und sind in der Kanzleikursive abgefasst. Das Korpus wurde mithilfe des Programms GATE (Cunningham et al. 2014) linguistisch aufbereitet und annotiert. Für diese Studie wurden stichprobenartig Faksimiles der handschriftlichen Akten zur Überprüfung der Transkription herangezogen.

4.2. Operationalisierung der Substantivkomposita

In die Untersuchung gehen alle Komposita ein, die in ihrer Bedeutung transparent sind und im Frnhd. eindeutig aus zwei Substantiven bestehen, was anhand des *Deutschen Wörterbuchs* der Gebrüder Grimm überprüft wurde. So werden Belege mit Erstgliedern, die potentiell einen Verbstamm darstellen (bspw. *Bademagd* oder *Spielmann*) und stark lexikalisierte Belege wie *Herzog* oder *Sonntag* ausgeschlossen (anders: Solling 2012). Wortbestandteile wie *-werk* in *Handwerk*, die im Untersuchungszeitraum zwar als eigenständige Substantive vorkommen, aber bereits dabei sind, sich zu Affixoiden zu entwickeln, werden als Kompositionsbestandteile gewertet.⁶ Zudem werden ausschließlich Appellativa betrachtet: Zusammengesetzte Eigennamen (*Rheinwiesen*, *Rosenfeld*) werden von der Untersuchung ausgeschlossen. Zusätzlich werden Belege ausgeschlossen, bei denen aus der Transkription hervorgeht,

4 Dieses Vorgehen führt zu Unterschieden in der Größe der einzelnen Teilkorpora. So sind für die Hochzeit der Hexenverfolgung in den 1620er-Jahren mehr Texte bei Macha et al. (2005) ediert als aus den Zeitschnitten zu Beginn und am Ende des Untersuchungszeitraums.

5 Für Informationen über den Entstehungsprozess einzelner Texte s. Macha et al. (2005), die zu jedem transkribierten Protokoll, wo möglich, Informationen über den Status des Dokuments als Mit- oder Abschrift und die Anzahl und Identität der Schreiber vor jedem Transkript angeben.

6 Das gilt allerdings nicht für *-tum*, *-schaft* und *-heit*, die bereits im Mhd. ihren Status als freie Lexeme eingebüßt haben (vgl. Klein, Solms & Wegera 2009: 158).

dass ein Großteil des Worts bzw. eines Kompositionsglieds aus aufgelösten Abkürzungen besteht, bspw. <gg> für *Goldgulden*. Erstglieder koordinierter Komposita werden ebenfalls nicht aufgenommen; in Beispielen wie *Rueben oder Pflanzgärtl* (München 1600) wird nur der zweite Teil (in diesem Fall *Pflanzgärtl*) in die Untersuchung miteinbezogen.

Die Unterscheidung zwischen Substantivkomposita und Phrasen mit vorangestelltem Genitivattribut erfolgt im Rahmen dieser Untersuchung nach dem morphologischen Kriterium der Artikelkongruenz (vgl. Dücker & Szczepaniak 2017). Die Schreibung als ein oder zwei Wörter hat hingegen keinen Einfluss auf ihre Interpretation als Kompositum. Wie in Kap. 2 beschrieben, unterscheiden sich Genitivphrasen und Substantivkomposita dadurch, dass der Determinierer bei einer Genitivphrase mit dem Attribut (vgl. Bsp. (1)) und bei einem Kompositum mit dessen Kopf kongruiert (vgl. Bsp. (2)).⁷

- (1) Genitivkonstruktion: *solchem Pauren Auch Ain Rosß Inns Theüfels Nammen Zue Todt geschlagenn* (Rosenfeld 1603)
- (2) Kompositum: *Seie sie bei ¼ Jahr auf einem hexen tanz beim Salzstadel gewesen* (Memmingen 1665)

Wie in Kap. 2 erwähnt wurde, wird für die Auswertung keine Unterteilung in eigentliche und uneigentliche Komposita vorgenommen. Stattdessen wird zwischen verfügt und unverfügt Komposita unterschieden. Ein Fugenelement wird hier definiert als das Wortmaterial, das zwischen Bestimmungsglied und Kopf eines Kompositums steht und nicht zum Nominativ Singular des Bestimmungsglieds gerechnet werden kann (vgl. Kürschner & Szczepaniak 2013: 1). Dabei ist es unerheblich, ob es sich um ein aus einem ehemaligen Genitivflexiv entstandenes Fugenelement handelt (*gericht+s+personen*, Gaugrehtweiler 1610) oder um den erstarrten Kompositionsvokal eines eigentlichen Kompositums (*hertz+e+leidt*, Perleberg 1588).⁸

4.3. Operationalisierung der Getrennt- und Zusammenschreibung

Die Operationalisierung von Getrennt- und Zusammenschreibung erfolgt wegen der Besonderheiten von handschriftlichen Texten nicht nach den Kriterien, die Solling (2012) auf Grundlage von Saenger (1997) anwendet. Dass bisherige Studien zur Worttrennung bei Komposita sich auf gedruckte Texte

7 Für das Vorgehen in Fällen mit ambiger Determiniererkongruenz vgl. Dücker & Szczepaniak (2017: 38–39). Belege, die auch mit dem dort vorgestellten semantischen Kriterium nach Solling (2011, 2012) nicht eindeutig erfasst werden können, werden von der Untersuchung ausgeschlossen. Es handelt sich dabei um 24 Belege.

8 Solling (2012: 35–36) geht bei der Einteilung anders vor; er zählt Belege vom Typ *Herzeleid* zu den eigentlichen Komposita.

konzentriert haben, liegt auch daran, dass gängige Methoden zur Spatienerkennung hier einfacher anwendbar sind: „The decisive factor here [bei der Wahl von Drucken als Untersuchungskorpus, L.D.] was that it is easier to determine whether a word is written as one word or two in printed texts than in manuscripts“ (Solling 2016: 291).

Für gedruckte Texte kann nach Saenger (1997: 27) ganz allgemein gelten, dass ein Abstand zwischen zwei Buchstaben, der 1,5-mal so groß ist wie „the distance between the minum stroke of the ‚lowercase‘ letters *u*, *m*, or *n*“, eine Grenze zwischen zwei Wörtern konstituiert. Diese recht mechanische Vorgehensweise ist für die handschriftlichen Mitschriften, aus denen das hier untersuchte Korpus besteht, allerdings nicht anwendbar. Busch & Fleischer (2015: 566) nennen das Spatium in handschriftlichen Texten einen „skalare[n] Begriff“ und weisen darauf hin, dass es für Handschriften „keineswegs eine Definition, ab welchem Abstand ein worttrennendes Spatium vorliegt“, gibt (Busch & Fleischer 2015: 566). Zur Bestimmung von Spatien setzen sie den horizontalen Abstand zwischen zwei Graphien mit der vertikalen Höhe der umgebenden Graphien ohne Ober- und Unterlänge ins Verhältnis: Entspricht der horizontale Abstand mindestens der Hälfte der vertikalen Höhe, sprechen sie von einem „deutliche[n] Spatium“ (Busch & Fleischer 2015: 568).

Die vorliegende Untersuchung stützt sich vorrangig auf die Transkription von Macha et al. (2005). Die Edition liefert aber leider keine Hinweise darauf, wie die Unterscheidung zwischen Getrennt- und Zusammenschreibung von Wörtern gemacht wurde. In der vorliegenden Untersuchung wird deshalb eine konservative Operationalisierung der Getrennt- und Zusammenschreibung von Komposita vorgenommen: Prinzipiell wird von Getrenntschreibung der Kompositionsglieder ausgegangen. Zwei Wörter gelten nur dann als zusammengeschieden, wenn sie in der Handschrift durch eine physische Linie miteinander verbunden sind, da nur in solchen Fällen eindeutig davon ausgegangen werden kann, dass der Schreiber Zusammenschreibung intendiert hat. Dazu kommen einige Regeln für besondere Fälle: Enthält das Kompositum bspw. auch innerhalb einer Konstituente Buchstaben, die nicht miteinander verbunden sind, so gilt das Kompositum als getrenntgeschrieben, wenn der Abstand zwischen den Buchstaben an der Grenze zwischen den Konstituenten größer ist als der im Inneren der Konstituente.

17 der Transkriptionen von Macha et al. (2005) wurden auf Grundlage dieser Kriterien mit Faksimiles der Handschriften verglichen. Da die Transkription nur in sehr wenigen Fällen von den hier vorgestellten Maßstäben abweicht, werden die Texte von Macha et al. (2005) als authentische Darstellung der Zusammen- und Getrenntschreibung von Komposita in frnhd. Handschriften angesehen. So wird es möglich, erstmals auch handschriftliche historische Texte in Bezug auf die Getrennt- und Zusammenschreibung von

Komposita zu untersuchen. Gedruckte Texte gehen von der Konzeption bis zur Drucksetzung durch viele verschiedene Hände und unterliegen oftmals Veränderungen durch Lektorat und Drucker. Die Untersuchung handschriftlicher Texte verspricht entsprechend einen direkteren Einblick in spontane Schriftlichkeit und die „online-Graphematik“ der Schreibenden (vgl. Szczepaniak & Barteld 2016: 45).

4.4. Zusammen- und Getrenntschreibung im SiGS-Korpus

Anhand der in Kap. 4.3 genannten Kriterien wurden insgesamt 880 Komposita ermittelt, die sich auf 472 unterschiedliche Types verteilen. Dies entspricht einem Anteil von 5,1% an allen im Korpus vorhandenen Substantiven (n=17152). In den einzelnen Texten sind Substantivkomposita unterschiedlich oft vertreten: In Helmstedt 1580 ist bspw. nur ein einziges Kompositum belegt (*kerkhoue*), was einem Anteil von 0,6% an allen in dem Protokoll enthaltenen Substantiven entspricht. In Dillenburg 1631 sind hingegen 31 Komposita verzeichnet, die 14,4% aller Substantive ausmachen.

Das Verhältnis zwischen getrennt- und zusammengeschriebenen Komposita ist wesentlich ausgeglichener als in den Daten von Solling (2012): Von den 880 im Korpus befindlichen Substantivkomposita werden 529 (60,1%) aller Belege zusammengeschrieben. Darunter befinden sich 46 Komposita, die an der Kompositionsfuge am Zeilenende mit einem Bindestrich getrennt werden und ein Beleg mit Bindestrichschreibung innerhalb der Zeile (*Sondern Ihr daumen und Schien-schrauben Angelegt worden*, Rhens 1629). Von den 351 (39,9%) getrenntgeschriebenen Belegen werden 55 an ihrer Kompositionsfuge am Zeilenende getrennt (vgl. Tab. 3).⁹

9 Die am Zeilenende getrennten Komposita mit und ohne Bindestrich werden hier nicht als eigenständige Gruppen aufgefasst, sondern im weiteren Verlauf der Untersuchung werden sie jeweils als Untergruppe der zusammen- bzw. getrennt geschriebenen Komposita betrachtet. Dieses Vorgehen wird dadurch begründet, dass sich die am Zeilenende mit Bindestrich getrennten Komposita in Bezug auf die hier untersuchten Eigenschaften genauso verhalten wie die zusammengeschriebenen Belege. Auch die Komposita mit Trennung innerhalb der Zeile und am Zeilenende ohne Bindestrich verhalten sich gleich. Die Aufteilung in Trennung am Zeilenende mit und ohne Bindestrich ist deshalb notwendig, da nur bei Bindestrichsetzung ein klarer Wille zur Zusammenschreibung erkennbar ist.

Tab. 3: *Getrennt- und Zusammenschreibung im Gesamtkorpus.*

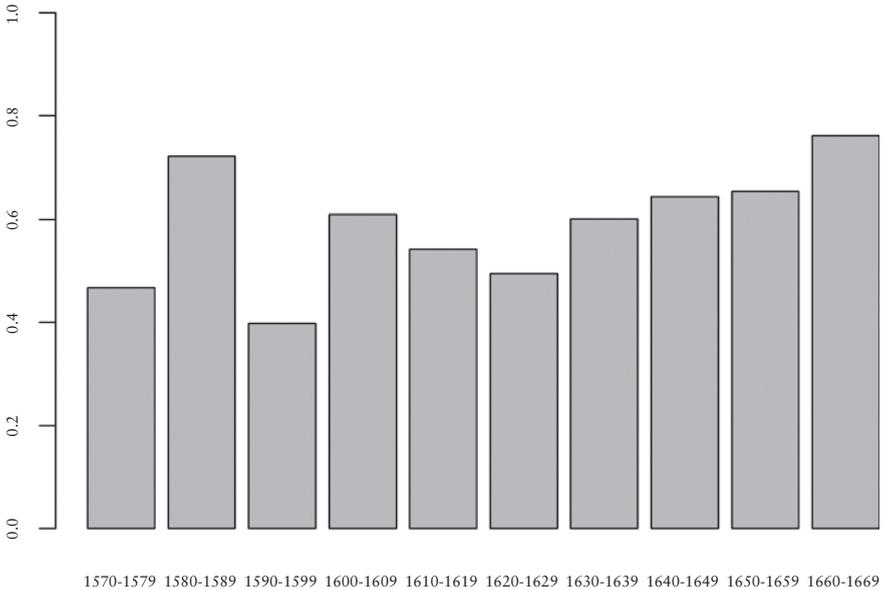
Getrennt- und Zusammenschreibung	Absoluter Wert	Prozentwert	Beispiel
Getrennt	351	39,9%	<i>Ross koth</i> (Baden-Baden 1628)
davon am Zeilenende	55		<i>Donnerstag Abendt</i> (Werl 1630)
zusammen	529	60,1%	<i>stutenbrei</i> (Alme 1630)
davon mit Bindestrich am Zeilenende	46		<i>hosenbandt</i> (Schweinfurt 1616)
davon mit Bindestrich innerhalb der Zeile	1		<i>Schien-schrauben</i> (Rhens 1629)
Gesamt	880		

Diese Zahlen sind weit von Sollings (2012) Ergebnissen entfernt, dessen Daten für die Zeitschnitte 1550 bis 1620 stets mehr als 80% Zusammenschreibung aufweisen (vgl. Solling 2012: 121); in den handschriftlichen Texten herrscht deutlich mehr Getrenntschreibung vor (60,1%). Wie Abbildung 2 zeigt, wird diese Marke für keinen der hier untersuchten Zeitschnitte erreicht: Zwei von drei Zeitschnitten vor 1600 weisen weniger als die Hälfte zusammengeschriebene Komposita auf; der Höchstwert liegt im letzten Zeitschnitt bei 76,2%.

Über den gesamten Zeitraum hinweg ist allerdings ein Anstieg der Zusammenschreibung zu beobachten (Kendall's $\tau = 0,6$).¹⁰

10 Der hohe Anteil an Zusammenschreibung im zweiten Zeitschnitt ist auf den starken Einfluss des Schreibers des Protokolls Perleberg 1588 zurückzuführen: Bis auf einen Beleg stammen alle in diesem Zeitraum erfassten Komposita aus diesem Text. In die Berechnung von Kendall's τ sind alle Belege aus dem Zeitschnitt eingegangen.

Abb. 2: Anteil der Zusammenschreibung pro Zeitschnitt (n=880, Kendall's $\tau = 0,6$, $p=0,01667$).



Bei der räumlichen Verteilung ergeben sich in Bezug auf den Anteil der Zusammenschreibung keine signifikanten Unterschiede zwischen den Regionen (vgl. Tab. 4). Der höchste Anteil an Zusammenschreibung findet sich im westlichen Mitteldeutschland (64,6%) und östlichen Norddeutschland (64,5%), während der niedrigste Anteil an zusammengesetzten Komposita im östlichen Mitteldeutschland gemessen wird (47,8%). Es ist kein eindeutiger Ausbreitungsverlauf der Zusammenschreibung von Substantivkomposita erkennbar, was sich mit den Ergebnissen von Solling (2012: 160) und Pavlov (1983: 12) deckt.

Tab. 4: Anteil der Zusammenschreibung pro Dialektgebiet ($\chi^2=8,1207$, $p=0,1497$).

	Westdeutschland	Ostdeutschland	gesamt
Norddeutschland	94/164 (57,3%)	49/76 (64,5%)	143/240 (59,6%)
Mitteldeutschland	153/237 (64,6%)	33/69 (47,8%)	186/306 (60,1%)
Süddeutschland	76/121 (62,8%)	124/213 (58,2%)	200/334 (59,9%)
gesamt	323/522 (61,9%)	206/358 (57,5%)	

In fünf von sechs Regionen wird die Mehrzahl der Komposita zusammengeschieden. Im Gegensatz zu dem in Dücker & Szczepaniak (2017) untersuchten SiGS-Kernkorpus, das 18 Protokolle umfasst, ist nach der Auswertung

des SiGS-Gesamtkorpus mit 56 Texten kein eindeutiges Ost-West-Gefälle mehr erkennbar; die westdeutschen Protokolle weisen (anders als im Kernkorpus) mehr Zusammenschreibung auf als die ostdeutschen Texte. Dies ist mit der Heterogenität innerhalb der Dialekträume zu erklären, auf die auch Schutzeichel & Szczepaniak (2015) in Bezug auf den regionalen Einfluss der Substantivgroßschreibung hinweisen. Wichtiger als die Region ist für die Zusammenschreibung möglicherweise der Idiolekt oder auch der Einfluss einzelner Schreibschulen und Kanzleien, was sich allerdings aufgrund der fehlenden Daten zur Identität und Ausbildung der jeweiligen Schreiber derzeit nicht nachweisen lässt (vgl. Fußnote 5).

4.5. Zusammenhang zwischen Worttrennung und Verfugung

Von den 880 Substantivkomposita, die in den Hexenverhörprotokollen vorkommen, weisen 255 ein Fugenelement auf (29,0%). Beinahe drei Viertel aller Belege sind somit unverfugt (71,0%). Im SiGS-Korpus sind dabei sechs unterschiedliche Fugenelemente belegt:

- (1) *-e-*: *Schwineborsten* (Flensburg 1608)
- (2) *-s-*: *weibsbildt* (Bamberg 1628)
- (3) *-es-*: *kindes händtel* (Münüchen 1600)
- (4) *-n-*: *katzenbirn* (Lemberg 1630)
- (5) *-en-*: *Narren werk* (Mergentheim 1629)
- (6) *-er-*: *Kinder tauff* (Ahaus 1608)

Die Hexenverhörprotokolle lassen keine zeitliche Entwicklung bezüglich des Verhältnisses von unverfugten und verfugten Komposita erkennen (vgl. Tab. 5). Der Anteil der Komposita mit Nullfuge¹¹ liegt über den gesamten Zeitraum bei mindestens zwei Dritteln.

Tab. 5: Zeitliche Entwicklung des Anteils von unverfugten und verfugten Komposita (Kendall's $\tau = 0,06667$).

Zeitschnitt	Anzahl der Texte pro Zeitschnitt	Anzahl aller Wörter pro Zeitschnitt	unverfugte Komposita	verfugte Komposita	gesamt
1570–1579	1	1127	11	4	15
1580–1589	2	2206	17	1	18
1590–1599	8	11366	69	39	108
1600–1609	4	7571	48	21	69

11 Darunter werden hier neben Komposita, in denen die beiden Wortstämme zusammengefügt werden (*Gaißbokh*, Bamberg 1628), auch Komposita mit Reduktionsfuge (*Schulmeister*, Georgenthal 1597) verstanden.

Zeitschnitt	Anzahl der Texte pro Zeitschnitt	Anzahl aller Wörter pro Zeitschnitt	unverfugte Komposita	verfugte Komposita	gesamt
1610–1619	10	12229	108	25	133
1620–1629	13	21303	166	71	327
1630–1639	11	17297	121	62	183
1640–1649	5	10104	52	18	70
1650–1659	1	2287	17	9	26
1660–1669	1	1463	16	5	21
gesamt	56	86953	625	255	

Solling (2012: 295) und Kopf (2017: 197) weisen darauf hin, dass der Status als uneigentliches oder eigentliches Kompositum bzw. das Auftreten eines Fugenelements einen großen Einfluss auf die Zusammen- oder Getrenntschreibung eines Kompositums hat. Die Daten aus den Hexenverhörprotokollen stützen diese Befunde (vgl. Tab. 6).

Tab. 6: Worttrennung und Verfugung ($\chi^2=102,73$, $p<0,0001$).

	Zusammenschreibung	Getrenntschreibung	gesamt
Fugenelement	86 (33,6%)	169 (66,3%)	255 (100%)
kein Fugenelement	443 (70,9%)	182 (20,7%)	625 (100%)
Gesamt	529	351	

Verfugte und unverfugte Belege zeigen in den untersuchten Daten ein komplementäres Bild: Während Komposita mit Fugenelement in nur einem Drittel aller Fälle zusammengeschrieben werden, zeigen Belege mit Nullfuge über 70% Zusammenschreibung.

Die Aufschlüsselung in Komposita mit und ohne Fugenelemente in Abb. 3 und 4 zeigt deutlich, wie sehr die Verfugung die Worttrennung der Komposita beeinflusst: Unverfugte Komposita wie *beinschraube* (Göttingen 1649) weisen bereits zu Beginn des Untersuchungszeitraums einen Anteil von über 60% Zusammenschreibung auf, der in der Folge nur einmal unterschritten wird. Bis auf den Zeitschnitt 1600–1609 (49,3%) weisen die unverfugten Komposita stets überwiegend Zusammenschreibung auf. Im letzten Zeitschnitt werden 87,5% aller Komposita mit Nullfuge zusammengeschrieben. Auch wenn der Anteil der Zusammenschreibung nicht kontinuierlich ansteigt, so ist über den gesamten Untersuchungszeitraum hinweg doch ein Trend hin zu mehr Zusammenschreibung zu beobachten (Kendall's $\tau = 0,64$).

Abb. 3: Zeitliche Entwicklung der Zusammenschreibung bei unverfügten Komposita (Kendall's $\tau=0,64$, $p=0,009148$).

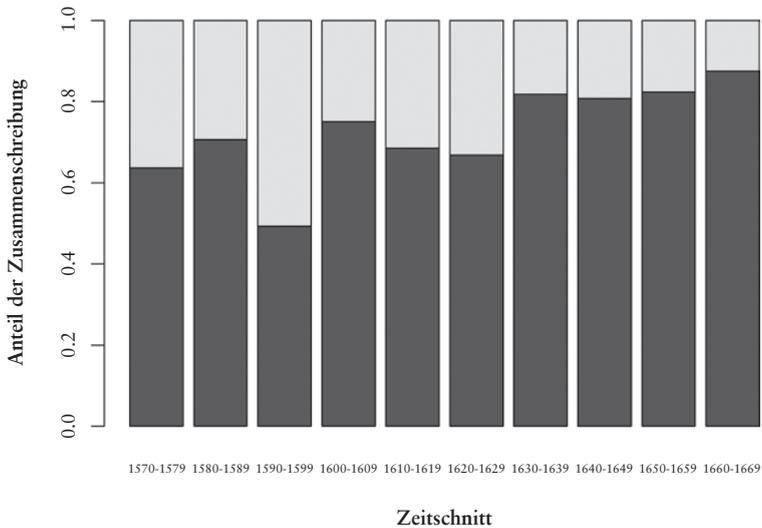
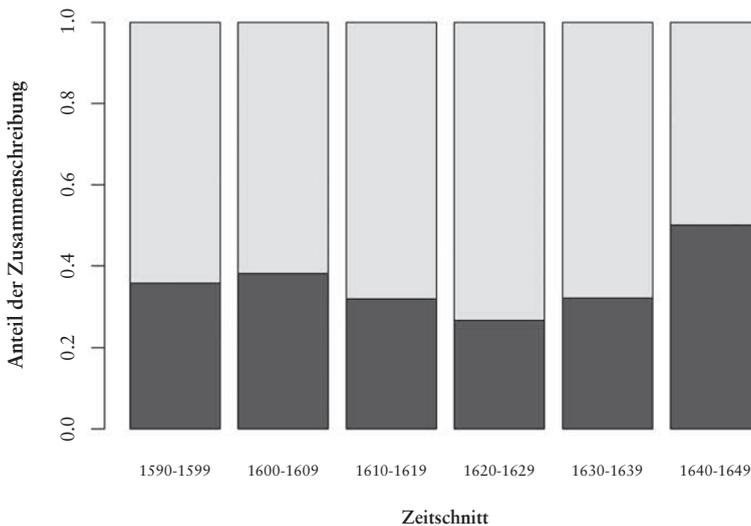


Abb. 4: Zeitliche Entwicklung der Zusammenschreibung bei verfügten Komposita (Kendall's $\tau=0,0667$, $p=1$)¹².



12 Für die graphische Darstellung der Entwicklung der Worttrennung bei verfügten Komposita wurden nur die Zeitschnitte ausgewählt, für die mindestens zehn Belege vorliegen. Die zwei frühesten und die zwei spätesten Zeitschnitte sind entsprechend nicht in der Graphik abgebildet.

Auch bei verfügten Komposita wie *teufels hure* (Leipzig 1640) ist am Ende des Untersuchungszeitraums mehr Zusammenschreibung zu beobachten als zu Beginn, jedoch ist hier keine eindeutige Aufwärtskurve auszumachen, sondern vielmehr ein Auf und Ab, bei dem die zusammengeschrriebenen Komposita jedoch nie in der Mehrzahl sind (Kendall's $\tau = 0,0667$). Im Zeitschnitt 1640–1649 stehen neun zusammengeschrriebene neun getrennten Komposita gegenüber; in allen anderen Zeitschnitten überwiegt die Getrenntschreibung. Der Anteil der Zusammenschreibung ist bei den Komposita mit Fugenelement also deutlich geringer als in den gleichen Zeitschnitten bei den unverfügten Komposita.

In Bezug auf die Verfügung lässt sich somit festhalten, dass unverfügte Komposita in späten Zeitschnitten besonders zur Zusammenschreibung tendieren, während Zusammensetzungen mit Fugenelementen häufiger getrenntgeschrieben werden, unabhängig vom Entstehungsjahr des entsprechenden Textes.

Angesichts der Tatsache, dass sich das Verhältnis zwischen verfügten und unverfügten Komposita im Zeitraum von 1570 bis 1669 nicht wesentlich verändert und die unverfügten Komposita so eindeutig in der Mehrzahl sind, stellt sich außerdem die Frage, ob die verfügten Komposita im 15. Jh. überhaupt schon frequent genug waren, um einen analogischen Sog auf die unverfügten auszuüben und zu mehr Getrenntschreibung bei den unverfügten Komposita zu führen, wie in der bisherigen Forschungsliteratur angenommen wird.

4.6. Zusammenhang zwischen Worttrennung und Majuskelsetzung

Für den Status als graphematisches Wort ist neben der Zusammen- oder Getrenntschreibung auch die Majuskelsetzung von großer Bedeutung (vgl. Kap. 3, Tab. 1). Im SiGS-Korpus sind alle möglichen Kombinationen aus Majuskelsetzung bei Bestimmungsglied oder Kopf und Getrennt- oder Zusammenschreibung des Kompositums belegt (vgl. Tab. 7).

Tab. 7: Majuskelsetzung und Worttrennung (n=608).¹³

Majuskelsetzung bei Bestimmungsglied und Kopf	Zusammenschreibung	Getrennschreibung	gesamt
Groß + klein	219	96	315
klein + klein	110	87	197
Groß + Groß	23	58	81
klein + Groß	3	12	15
Gesamt	355	253	

Am häufigsten ist mit 219 Belegen jedoch die zusammengesriebene Kombination aus majuskelhaltigem Erstglied und kleingeschriebenem Kopf (*Melthau*, Ellingen 1590), doch auch durchgängige Kleinschreibung (*gerichtscheffen*, Gaugrehweiler 1610) ist mit 197 Belegen häufig vertreten. Großschreibung des Kopfs kommt überwiegend in Kombination mit einem majuskelhaltigen Erstglied vor (*Garten Thuer*, Ostrau 1628). Nur dreimal sind hingegen Komposita belegt, die ein kleingeschriebenes Erstglied und einen großgeschriebenen Kopf nicht mit einem Spatium trennen (z. B. *kunstGeygerin*, Leipzig 1640).

Es gibt keinen signifikanten Zusammenhang zwischen der Worttrennung und der Großschreibung des Bestimmungsglieds ($\chi^2=3,1518$, $p=0,07584$); vielmehr wird der Anfangsrand des Erstglieds unabhängig von der Worttrennung bevorzugt mit einer Majuskel markiert, was die Wortwertigkeit der gesamten Konstruktion hervorhebt. Der Zusammenhang zwischen der Worttrennung und der Großschreibung des Kopfs ist hingegen hochsignifikant ($p<0,0001$): Nur 26 von 96 Komposita mit majuskelhaltigem Kopf werden zusammengesrieben (*AmbtMan*, Dillenburg 1631). Wird der Kopf kleingeschrieben, überwiegt stattdessen die Zusammenschreibung (329 von 512 Belegen), wodurch die Wortwertigkeit auch auf graphematischer Ebene betont wird. Diese Werte bestätigen die Ergebnisse, die Dücker & Szczepaniak (2017: 47) bereits für das SiGS-Kernkorpus gezeigt haben. Lediglich der Anteil an Binnenmajuskelschreibungen ist im Gesamtkorpus wesentlich höher (15,8%) als bei der Auswertung von 18 Texten (7,9%).

13 Prinzipiell ist bei der in den Hexenverhörprotokollen verwendeten Kanzleikursive eine Unterscheidung zwischen Majuskeln und Minuskeln möglich. Macha et al. (2005: XXIII) stellten bei der Transkription jedoch Probleme bei der Unterscheidung von <v>/<V>, <h>/<H> und <z>/<Z> fest. Aus diesem Grund sind alle Komposita, deren Bestimmungsglied oder Kopf mit einem dieser drei Grapheme beginnt, von der Untersuchung in Bezug auf die Majuskelsetzung ausgeschlossen, wodurch sich die Gesamtzahl hier auf 608 reduziert.

Majuskelsetzungen, die die graphematische Wortwertigkeit des Kompositums beeinträchtigen können, werden also im Allgemeinen vermieden.¹⁴ Weitere Gemeinsamkeiten zwischen Komposita mit majuskelhaltigem Kopf lassen sich nicht erkennen: Sie stammen aus allen Dialektgebieten und sind in den meisten Zeitschnitten vertreten. Auch in Bezug auf Lexik, Semantik oder die Stellung am Zeilenanfang lässt sich kein Einfluss auf die Großschreibung des Zweitglieds erkennen.

4.7. Das Lemma als Faktor

Als weiterer Faktor für die Getrennt- und Zusammenschreibung von Komposita wird das Lemma in den Blick genommen. Dabei wird untersucht, ob bestimmte Lemmata zu der Schreibung mit oder ohne Spatium tendieren und ob es sich dabei um besonders frequente Komposita handelt.

Mithilfe des R-Pakets *Collostructions* (Flach 2017, Version 0.1.0) wird eine distinktive Kollexemanalyse durchgeführt (vgl. Gries & Stefanowitsch 2004). Dabei werden zwei Konstruktionen miteinander verglichen und berechnet, in welcher der Konstruktionen (hier Getrenntschreibung und Zusammenschreibung) ein bestimmtes Lemma häufiger auftritt, als bei einer zufälligen Verteilung zu erwarten wäre. Als Maß der Assoziationsstärke wird in dieser Untersuchung die Log-Likelihood Ratio verwendet. Tab. 8 enthält die zehn Types mit der höchsten Log-Likelihood Ratio für Zusammenschreibung, während Tab. 9 die zehn Types mit der größten Affinität zur Getrenntschreibung auflistet. Die Daten sind jeweils absteigend nach der Stärke der Assoziation angeordnet.

Tab. 8: Distinktive Kollexemanalyse, Assoziation: Zusammenschreibung.

Type	Tokenanzahl Getrennt- schreibung	Tokenanzahl Zusammen- schreibung	Stärke der Assoziation	Signifikanzniveau
<i>hausfrau</i>	2	45	18.37306	*****
<i>amtman</i>	0	18	9.27824	*****
<i>rathaus</i>	1	16	6.21995	*****
<i>steinmetzer</i>	0	10	5.12359	*****

14 Von den 26 Belegen mit großgeschriebenem Kopf bei Zusammenschreibung der Kompositionsglieder entfallen allein neun auf die Berufsbezeichnung *SteinMetzer* (Dillenburg 1631). Dieser Begriff wird stets gebraucht, um einen bestimmten Mann zu bezeichnen. Es handelt sich dabei somit möglicherweise um eine konventionalisierte Schreibung (vgl. Kap. 4.2) oder einen Eigennamen, worauf auch Macha et al. (2005: 192) hinweisen, die den Titel der Gerichtsakte als „Inquisitio contra Hansen Holttschenhäwer sonsten Steinmetzger genant zu Dillenburgk“ wiedergeben.

Type	Tokenanzahl Getrennt- schreibung	Tokenanzahl Zusammen- schreibung	Stärke der Assoziation	Signifikanzniveau
<i>ehebruch</i>	0	9	4.60778	****
<i>bürgermeister</i>	1	12	4.39068	****
<i>burgmann</i>	1	9	3.0759	***
<i>schafmeister</i>	0	6	3.06497	***
<i>kirchtag</i>	0	6	3.06497	***
<i>ehemann</i>	1	8	2.65286	**

Tab. 9: Distinktive Kollexemanalyse, Assoziation: Getrenntschreibung.

Type	Tokenanzahl Getrennt- schreibung	Tokenanzahl Zusammen- schreibung	Stärke der Assoziation	Signifikanzniveau
<i>teufelssalbe</i>	7	0	6.47023	*****
<i>nachtzeit</i>	6	0	5.54069	*****
<i>mannsperson</i>	4	0	3.68687	***
<i>teufelspulver</i>	4	0	3.68687	***
<i>dienstverwalter</i>	3	0	2.76256	**
<i>gerichtsperson</i>	3	0	2.76256	**
<i>leinentuch</i>	3	0	2.76256	**
<i>reichstaler</i>	3	0	2.76256	**
<i>teufelskind</i>	3	0	2.76256	**
<i>verdachtspunkt</i>	3	0	2.76256	**

Belege für *hausfrau* und *amtman*n kommen demnach überzufällig häufig mit Zusammenschreibung vor, *teufelssalbe* und *nachtzeit* bevorzugen hingegen die Schreibung mit Spatium. Es zeigt sich, dass die stark zur Zusammenschreibung tendierenden Komposita insgesamt häufiger im Korpus vertreten sind als solche, die zur Getrenntschreibung tendieren, deren Frequenz relativ niedrig ist. Nur die zwei Types, die am stärksten zur Getrenntschreibung tendieren, sind häufiger als fünfmal im Korpus belegt. Entsprechend werden für die Assoziation mit Zusammenschreibung häufiger höhere Signifikanzniveaus erreicht. Bei einem Type wie *dienstverwalter*, der mit nur drei Token belegt ist, fällt ein Beleg mit Zusammenschreibung wesentlich mehr ins Gewicht als ein Beleg mit Spatium bei *hausfrau*, das auch bei Getrenntschreibung bei zwei von 47 Token die stärkste Assoziation zur Zusammenschreibung zeigt. So lässt sich auch der auffällige Umstand erklären, dass diverse der mit Zusammenschreibung assoziierten Types auch in getrennter Schreibweise vorliegen, die am deutlichsten zur Getrenntschreibung tendierenden Types aber ausschließlich mit Spatium geschrieben werden.

Unter den Komposita, die zur Zusammenschreibung tendieren, befinden sich nur solche, deren Nennform laut *Deutschem Wörterbuch* kein Fugenelement aufweisen und die entsprechend nicht als uneigentliche Komposita gelten. Mit Getrenntschreibung sind hingegen sowohl verfugte (*teufelskind*) als auch unverfugte Komposita (*dienstverwalter*) assoziiert. Die verfügten Komposita sind mit acht Lemmata allerdings dominant. Hier zeigt sich erneut der in Kap. 3.3 beobachtete Zusammenhang zwischen Verfugung und Worttrennung. Mit Blick auf einen möglichen analogischen Sog in Bezug auf die Getrenntschreibung, wie er in der Forschungsliteratur vorgeschlagen wurde, muss aber auch hier auf die niedrigen Tokenzahlen bei den verfügten Komposita verwiesen werden.

Ein Urteil über den Zusammenhang zwischen Tokenzahl und Tendenz zur Zusammenschreibung, wie ihn Pavlov (1983) formuliert, kann anhand dieser Daten nicht gefällt werden, da Types, die ein ausgeglichenes Verhältnis zwischen Zusammen- und Getrenntschreibung zeigen, mit der distinktiven Kollexemanalyse nicht erfasst werden.

Ein Blick auf den Anteil der Zusammenschreibung bei den zehn häufigsten Komposita im SiGS-Korpus (vgl. Tab. 10) zeigt aber deutlich, dass bei neun von zehn Lemmata die Zusammenschreibung bei weitem überwiegt.

Tab. 10: Anteil der Zusammenschreibung bei den frequentesten Komposita.

Lemma	Anteil der Zusammenschreibung
<i>hausfrau</i>	45/47
<i>amtman</i>	18/18
<i>rathaus</i>	16/17
<i>bürgermeister</i>	12/13
<i>burgmann</i>	9/10
<i>kubhirte</i>	5/10
<i>steinmetzer</i>	10/10
<i>ehebruch</i>	9/9
<i>ehemann</i>	8/9
<i>teufelstanz</i>	4/9

Von den häufigsten Komposita sind nur *kubhirte* und *teufelstanz* nicht in Tab. 8 enthalten. Sie sind die einzigen frequenten Komposita, die nicht stark zu Zusammenschreibung tendieren, sondern ein ausgewogenes Verhältnis zwischen Getrennt- und Zusammenschreibung aufweisen. *Teufelstanz* ist zudem das einzige verfugte Kompositum, das mehr als acht Token im SiGS-Korpus aufweist. Anders als bei Solling (2012: 267), in dessen gedruckten Predigttexten frequente Komposita je nach Zeitschnitt entweder mit Bindestrich, mit Spatium oder zusammengeschrieben werden, zeigen die Hexenverhörprotokolle eine

klare Tendenz dafür, häufig vorkommende Komposita zusammenzuschreiben. Genau wie bei Solling (2012) ist diese Beobachtung nicht an einzelne Texte gebunden, sondern auch bei Lemmata wie *hausfrau* zu sehen, die in Texten aus allen Regionen und Zeitschnitten belegt sind.

Um einer möglichen Lexikalisierung der Komposita nachzuspüren, wurden die zehn häufigsten Komposita im *Deutschen Wörterbuch*, dem mhd. Wörterbuch von Lexer sowie im ahd. Wörterbuch von Schützeichel nachgeschlagen. Wo möglich, wurde das ahd. Wörterbuch von Karg-Gasterstädt & Frings herangezogen.¹⁵ Alle in Tabelle 10 enthaltenen Zusammensetzungen sind im DWB belegt und nur *teufelstanz* hat keinen Eintrag bei Lexer. Im ahd. Wörterbuch sind wiederum nur *amtmann* (ahd. *ambabtman*) und *steinmetzer* (ahd. *steinmezzo*) aufgelistet. Damit steht einerseits fest, dass die im SiGS-Korpus frequenten Komposita überwiegend schon in mhd. Zeit vorkommen. Ob dies andererseits aber ausreicht, um daraus eine Lexikalisierung als Grund für die Zusammenschreibung abzuleiten, ist fraglich, zumal fünf der am stärksten mit Zusammenschreibung assoziierten Komposita auch mit Spatium vorkommen. Für das SiGS-Korpus lässt sich somit festhalten, dass die frequentesten Komposita überwiegend unverfugte sind und normalerweise zusammengeschieden werden und dass Komposita, die durchgehend getrenntgeschrieben werden, im Vergleich weniger frequent sind. Inwiefern eine mögliche Lexikalisierung von frequenten Lemmata einen Einfluss auf die Worttrennung v.a. bei eigentlichen Komposita hat, muss nach wie vor als Desiderat angesehen werden.

5. Fazit: Die zunehmende graphematische Wortwertigkeit der frnhd. Komposita

Die Untersuchung von frnhd. handschriftlichen Hexenverhörprotokollen hat gezeigt, dass unterschiedliche Faktoren dafür verantwortlich sind, ob ein Kompositum mit oder ohne Spatium geschrieben wird. Neben einer zeitlichen Entwicklung hin zu mehr Zusammenschreibung hat das Auftreten von Fugenelementen eine blockierende Wirkung und führt eher zu Getrenntschreibung. Auch ist die Majuskelsatzung beim Kopf des Kompositums ein guter Indikator für Getrenntschreibung. Ob das Bestimmungsglied groß oder klein geschrieben wird, ist hingegen nicht von Bedeutung. Auch eine hohe Tokenfrequenz wirkt sich begünstigend auf die Zusammenschreibung von Substantivkomposita aus.

¹⁵ Das *Deutsche Wörterbuch*, das mhd. Wörterbuch von Lexer und das ahd. Wörterbuch von Karg-Gasterstädt & Frings wurden über www.woerterbuchnetz.de eingesehen (gesehen am 8. Juni 2017).

Die höchste Wahrscheinlichkeit für eine Zusammenschreibung der Konstituenten ist gegeben, wenn das Kompositum in einem Text aus der Mitte des 17. Jhs. häufig auftaucht, es nicht verfugt ist und der Kopf nicht mit Majuskel geschrieben wird. Ein Spatium ist hingegen in solchen Fällen wahrscheinlich, wenn ein verfugtes Kompositum am Ende des 16. Jhs. nur selten vorkommt und einen majuskelhaltigen Kopf aufweist.

Somit sind in den handschriftlichen Hexenverhörprotokollen dieselben Faktoren einflussreich, die bereits für die Spatiensetzung in Drucken herausgearbeitet wurden. Der Anteil der Getrenntschreibung ist aber wesentlich höher als in gedruckten Texten aus der gleichen Zeit.

In Kap. 3 wurde eine graphische Darstellung von verschiedenen Stufen der graphematischen Wortwertigkeit frnhd. Komposita vorgestellt. Tab. 11 zeigt, wie sich die Belege aus dem SiGS-Gesamtkorpus in dieses Schema eingliedern.

Tab. 11: Die graphematische Wortwertigkeit der im Korpus belegten Komposita.

zwei graphematische Wörter			ein graphematisches Wort		
mit Binnenmajuskel (und fakultativer Initialmajuskel)	ohne Majuskel	mit Initialmajuskel	mit Binnenmajuskel (und fakultativer Initialmajuskel)	ohne Majuskel	mit Initialmajuskel
<i>Teufels Tanz</i> <i>teufels Tanz</i>	<i>teufels tanz</i>	<i>Teufels tanz</i>	<i>TeufelsTanz</i> <i>teufelsTanz</i>	<i>teufelstanz</i>	<i>Teufelstanz</i>
70 Belege	87 Belege	96 Belege	26 Belege	110 Belege	219 Belege

Deutlich ist zu sehen, dass die frnhd. Schreiber alle ihnen zur Verfügung stehenden Möglichkeiten in Bezug auf Worttrennung und Majuskelsetzung ausnutzen. Sowohl bei den Komposita mit als auch ohne Spatium nimmt die Belegzahl von links nach rechts ständig zu, d.h. dass immer mehr Substantivkomposita eine Gestalt mit möglichst hoher graphematischer Wortwertigkeit annehmen. Auch insgesamt gesehen nimmt die Anzahl der Belege zu, je höher die Wortwertigkeit eines Kompositums einzuschätzen ist. Die Ausnahme bilden dabei diejenigen Komposita, die trotz Zusammenschreibung einen majuskelhaltigen Kopf aufweisen, was die Interpretation als ein syntaktisches Wort erschwert. Somit zeigt sich bei aller Varianz eine klare Tendenz hin zu einer graphematischen Gestalt, die im Einklang mit dem Status als syntaktisches Wort ist.

Literatur

Althochdeutsches Wörterbuch. Auf Grund der von Elias v. Steinmeyer hinterlassenen Sammlungen im Auftrag der Sächsischen Akademie der Wissenschaften zu Leipzig.

- Bearbeitet und herausgegeben von Elisabeth Karg-Gasterstädt und Theodor Frings. Leipzig: 1952–2015ff.
<http://awb.saw-leipzig.de> (accessed 08 June 2017).
- Barteld, Fabian, Renata Szczepaniak & Heike Zinsmeister. 2014. The definition of tokens in relation to words and annotation tasks. In V. Henrich, E. Hinrichs, D. de Kok, P. Osenova & A. Przepiórkowski (eds.), *Thirteenth International Workshop on Treebanks and Linguistic Theories (TLT13)*. December 12–13, 2014, Tübingen, Germany. 250–257.
<http://tlt13.sfs.uni-tuebingen.de/tlt13-proceedings.pdf> (letzter Aufruf 13.06.2017).
- Busch, Nathanael & Jürg Fleischer. 2015. Zusammenschreibung im Althochdeutschen und Altniederdeutschen bis ca. 975. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 137(4). 563–598.
- Cunningham, Hamish, Diana Maynard & Kalina Bontcheva. 2014. *GATE Developer*. General Architecture for Text Engineering. Version 8.3. Sheffield.
- Demske, Ulrike. 1999. Case Compounds in the History of German. In Matthias Butt & Nanna Fuhrhop (eds.), *Variation und Stabilität in der Wortstruktur: Untersuchungen zu Entwicklung, Erwerb und Varietäten des Deutschen und anderer Sprachen*, 150–176. Hildesheim [u.a.]: Olms.
- Demske, Ulrike. 2001. *Merkmale und Relationen. Diachrone Studien zur Nominalphrase des Deutschen*. Berlin & New York: de Gruyter.
- Deutsches Wörterbuch* von Jacob und Wilhelm Grimm. 16 Bde. in 32 Teilbänden. Leipzig 1854–1961. Quellenverzeichnis Leipzig 1971.
<http://woerterbuchnetz.de/dwb> (accessed 08 June 2017).
- Dücker, Lisa & Renata Szczepaniak. 2017. „Auffm Teuffelßdantz haben sie auff knotten korffen linen gedantzet“. Die graphematische Markierung von Komposition in den Hexenverhörprotokollen aus dem 16./17. Jh. In Florentine Oehme, Hans Ulrich Schmid & Franziska Spranger (eds.), *Wörter. Wortbildung, Lexikologie und Lexikographie, Etymologie* (Jahrbuch für Sprachgeschichte Band 8), 30–51. Berlin & Boston: de Gruyter.
- Flach, Susanne. 2017. *Collostructions: An R implementation for the family of collostructional methods*. R package version 0.1.0, www.bit.ly/sflach (accessed 14 December 2017).
- Fuhrhop, Nanna. 2008. Das graphematische Wort (im Deutschen). Eine erste Annäherung. *Zeitschrift für Sprachwissenschaft* 27. 189–228.
- Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collostructional analysis: A corpus-based perspective on “alternations”. *International Journal of Corpus Linguistics* 9(1). 97–129.
- Haubold, Fritz. 1914. *Untersuchung über das Verhältnis der Originaldrucke der Wittenberger Hauptdrucker Lutherscher Schriften: Grunenberg, Lother, Döring-Cranach und Lufft zu Luthers Druckmanuskripten*. Zugl. Diss., Univ. Jena, 1914. Borna-Leipzig: Noske.
- Klein, Thomas, Hans-Joachim Solms & Klaus-Peter Wegera. 2009. *Mittelhochdeutsche Grammatik. Band 3: Wortbildung*. Tübingen: Niemeyer.

- Kopf, Kristin. 2017. Fugenelement und Bindestrich in der Compositions-Fuge. In Nanna Fuhrhop, Renata Szczepaniak & Karsten Schmidt (eds.), *Sichtbare und hörbare Morphologie*, 177–204. Berlin & New York: de Gruyter.
- Kopf, Kristin. in press. From genitive suffix to linking element. A corpus study on the genesis and productivity of a new compounding pattern in (Early) New High German. In Horst Simon, Christian Zimmer & Tanja Ackermann (eds.), *Germanic Genitives*, 2–25. Amsterdam & Philadelphia: Benjamins.
- Kürschner, Sebastian & Renata Szczepaniak. 2013. Linking elements – origin, change, and functionalization. *Special issue of Morphology* 23(1). 1–6.
- Lexner, Matthias. 1872–1878/2011. *Mittelhochdeutsches Handwörterbuch*. 3 Bände. Leipzig: Hirzel. <http://woerterbuchnetz.de/Lexer/> (accessed 08 June 2017).
- Macha, Jürgen, Elvira Topalović, Iris Hille, Uta Nolting & Anja Wilke (eds.). 2005. *Deutsche Kanzleisprache in Hexenverhörprotokollen der Frühen Neuzeit*. 2 Bände. Berlin & New York: de Gruyter.
- Moser, Virgil. 1982. *Schriften zum Frühneuhochdeutschen*. Herausgegeben von Hugo Stopp, mit einem Register erstellt von Angelika Schmitt unter Mitw. von Maria Walch. 2 Bände. Heidelberg: Winter.
- Nitta, Haruo. 1987. Zur Erforschung der ‚uneigentlichen‘ Zusammensetzungen im Frühneuhochdeutschen. *Zeitschrift für deutsche Philologie* 106. 400–416.
- Nübling, Damaris & Renata Szczepaniak. 2011. Merkmal(s?)analyse, Seminar(s?)arbeit und Essen(s?)ausgabe: Zweifelsfälle der Verfungung als Indikatoren für Sprachwandel. *Zeitschrift für Sprachwissenschaft* 30(1). 45–73.
- Pavlov, Vladimir M. 1972. *Die substantivische Zusammensetzung im Deutschen als syntaktisches Problem*. München: Hueber.
- Pavlov, Vladimir M. 1983. *Zur Ausbildung der Norm der deutschen Literatursprache im Bereich der Wortbildung (1470–1730). Von der Wortgruppe zur substantivischen Zusammensetzung*. Berlin: Akademie-Verlag.
- Pavlov, Vladimir M. 2004. Zur Entwicklung der substantivischen Zusammensetzung im Frühneuhochdeutschen. In Klaus J. Mattheier & Haruo Nitta (eds.), *Sprachwandel und Gesellschaftswandel – Wurzeln des heutigen Deutsch: Studien des deutsch-japanischen Arbeitskreises für Frühneuhochdeutschforschung*, 99–119. München: Iudicium.
- Pavlov, Vladimir M. 2009. *Deutsche Wortbildung im Spannungsfeld zwischen Lexikon und Syntax. Synchronie und Diachronie*. Frankfurt a.M. [u. a.]: Peter Lang.
- Saenger, Paul. 1997. *Space between words: the origins of silent reading*. Stanford: Stanford University Press.
- Scherer, Carmen. 2012. Vom „Reisezentrum“ zum „Reise Zentrum“: Variation in der Schreibung von N+N-Komposita. In Livio Gaeta (ed.), *Das Deutsche als kompositionsfreudige Sprache: Strukturelle Eigenschaften und systembezogene Aspekte*, 57–81. Berlin & Boston: de Gruyter.
- Schindler, Wolfgang. 1999. Bindestrich-Komposita im Frühneuhochdeutschen. In Wolfgang Schindler & Jürgen Untermann (eds.), *Grippe, Kamm und Eulenspiegel. Festschrift für Elmar Seebold zum 65. Geburtstag*, 313–329. Berlin & Boston: de Gruyter.

- Schutzzeichel, Marc & Renata Szczepaniak. 2015. Die Durchsetzung der satzinternen Großschreibung in Norddeutschland am Beispiel der Hexenverhörprotokolle. In Markus Hundt & Alexander Lasch (eds.), *Deutsch im Norden. Varietäten des norddeutschen Raums*, Bd. 1, 151–167. Berlin & Boston: de Gruyter.
- Schützeichel, Rudolf. 2012. *Althochdeutsches Wörterbuch*. 7., durchgesehene und verbesserte Auflage. Berlin & Boston: de Gruyter.
- Solling, Daniel. 2011. Zur Problematik der Unterscheidung zwischen pränominalen Genitivattribut und getrennt geschriebenem Substantivkompositum im Frühneuhochdeutschen. In Jörg Riecke (ed.), *Historische Semantik*, 294–311. Berlin & Boston: de Gruyter.
- Solling, Daniel. 2012. *Zur Getrennt-, Zusammen- und Bindestrichschreibung von Substantivkomposita im Deutschen (1550–1710)*. Zugl. Diss. Univ. Uppsala, Schweden, 2012. Uppsala: Univ.
- Solling, Daniel. 2016. Compound nouns in German (1550–1710): open, closed and hyphenated forms. In Paola Cotticelli-Kurras und Alfredo Rizza (eds.), *Variation within and among writing systems. Concepts and methods in the analysis of ancient written documents*, 291–302. Wiesbaden: Reichert.
- Solms, Hans-Joachim. 1999. Der Gebrauch uneigentlicher Substantivkomposita im Mittel- und Frühneuhochdeutschen als Indikator kultureller Veränderung. In Andreas Gardt, Ulrike Haß-Zumkehr & Thorsten Roelcke Thorsten (eds.), *Sprachgeschichte als Kulturgeschichte*, 225–246. Berlin & Boston: de Gruyter.
- Szczepaniak, Renata. 2016. Is the development of linking elements in German a case of exaptation? In Muriel Norde & Freek van de Velde (eds.), *Exaptation and Language Change*, 317–340. Amsterdam & Philadelphia: Benjamins.
- Szczepaniak, Renata & Fabian Barteld. 2016. Hexenverhörprotokolle als sprachhistorisches Korpus. In Sarah Kwekkeboom & Sandra Waldenberger (eds.), *Perspektivwechsel oder: Die Wiederentdeckung der Philologie. Bd. 1: Sprachdaten und Grundlagenforschung in der Historischen Linguistik*, 43–70. Berlin: Erich Schmidt Verlag.
- Topalović, Elvira. 2003. *Sprachwahl – Textsorte – Dialogstruktur: Zu Verhörprotokollen aus Hexenprozessen des 17. Jahrhunderts*. Trier: WTV.
- Weidman, Robert H. 1941. The orthographic conflation of nominal compounds in mhg based on a study of the Manesse manuscript. In Arno Schirokauer & Wolfgang Paulsen (eds.), *Corona: Studies in celebration of the eightieth birthday of Samuel Singer, professor emeritus, University of Berne, Switzerland*, 89–99. Durham, North Carolina: Duke Univ. Press.

Lisa Dücker
 Otto-Friedrich-Universität Bamberg
 Lehrstuhl für Deutsche Sprachwissenschaft
 Hornthalstraße 2
 96045 Bamberg
lisa.duecker@uni-bamberg.de