

Universität Hamburg
Institut für Germanistik
Modul: GL-M02 Linguistische Empirie
Seminar: 52:145 Statistik für Korpuslinguistik
Dozentin: Prof. Heike Zinsmeister

Gender und soziale Rolle in der Entwicklung der satzinternen
Großschreibung –
eine multifaktorielle Analyse

Lena Schnee
Matrikelnummer 4623268
Lorenzengasse 15
22303 Hamburg
[lena.schnee@studium.uni-hamburg.de](mailto:lana.schnee@studium.uni-hamburg.de)
Vorgelegt am 19. März 2020

Inhaltsverzeichnis

1 Einleitung	1
2 Faktoren in der Entwicklung der satzinternen Großschreibung	2
2.1 Phasen und Steuerungsfaktoren der Majuskelverwendung	3
2.1.1 GENDER und MORALISCHE BEWERTUNG als Faktoren für Großschreibung	6
2.1.2 Einfluss des Lemmas (LEMMAPRESTIGE)	7
2.1.3 SCHREIBERSPEZIFISCHE UNTERSCHIEDE	8
2.1.4 FREQUENZ	9
2.2 Hypothesenaufstellung	9
3 Methodische Grundlegung	10
3.1 Das SiGs-Projekt	10
3.2 Korpusbeschreibung	12
3.3 Untersuchte Teilmenge	13
3.4 Vorgehen	14
4 Datenexploration und Visualisierung	17
5 Ergebnisse	21
5.1 LEMMAPRESTIGE	23
5.2 GENDER und MORALISCHE BEWERTUNG	25
5.3 FREQUENZ	28
6 Fazit	28
7 Literaturverzeichnis	30
7.1 Korpus	30
7.2 Forschungsliteratur	30
7.3 Online Quellen und Programme	31
Anhang	32

Abbildungsverzeichnis

Abbildung 1: Belebtheitsskala mit Konkretheitsgraden.	5
Abbildung 2: Genderdifferenzierte Großschreibung von Personenbezeichnungen (n=549) (aus Barteld et al. 2016:398).....	6
Abbildung 3: Groß- und Kleinschreibung der Personenbezeichnungen nach Protokollen (n=551).	8
Abbildung 4: Bewertung der einzelnen Rollen.	14
Abbildung 5: Rollen in denen die prestigebehafteten Lemmata (n=17) auftreten (67 Belege).	17
Abbildung 6: Groß- und Kleinschreibung von Personenbezeichnungen im Kernkorpus (n= 551).....	18
Abbildung 7: Bewertung der Männer und Frauen im Kernkorpus.....	18
Abbildung 8: Großschreibung und soziale Bewertung bei Männern (n=307) Frauen (n=244).	19
Abbildung 9: Großschreibung und soz. Bewertung bei Männern und Frauen (anteilig).	19
Abbildung 10: Großschreibung Männer bezeichnender Appellativa nach Rolle bei neutralem oder prestigebehaftetem Lemma (n=67).	20
Abbildung 11: Anteil Groß- und Kleinschreibung neutraler und prestigebehafteter männlicher Personenbezeichnungen.	20
Abbildung 12: Groß- und Kleinschreibung der frequentesten Lemmata.	21
Abbildung 13: Vorhergesagte Werte des Modells anhand der Estimates (0 = Großschreibung wahrscheinlich, 1=Kleinschreibung wahrscheinlich).	23
Abbildung 14: Vorhergesagte Werte des Modells anhand der Estimates (0 = Großschreibung wahrscheinlich, 1=Kleinschreibung wahrscheinlich).	26

1 Einleitung

Die satzinterne Großschreibung im Deutschen hat sich maßgeblich im 16. und 17. Jahrhundert, der Zeit des Frühneuhochdeutschen, entwickelt. Mehrere Studien konnten den Einfluss der semantischen Kategorie der Belebtheit auf diesen graphematischen Wandelprozess zeigen: Die satzinterne Majuskel breitete sich vom höchsten Belebtheitsgrad (Personenbezeichnungen) zum niedrigsten (Abstrakta) aus. Bergmann (1999:70) konstatiert für die Zeit um 1590 in den von ihm untersuchten Drucken bereits eine Großschreibungsnorm für Eigennamen und Konkreta, die abstrakten Substantive folgen mit Verzögerung. Auch in den handschriftlichen Hexenverhörprotokollen, die im Rahmen des DFG-Projekts „Entwicklung der satzinternen Großschreibung im Deutschen“ (SiGS)¹ korpuslinguistisch aufbereitet und untersucht wurden, erweist sich der Belebtheitseinfluss als signifikant (vgl. allg. Barteld et al. 2016:396f. oder speziell für den norddeutschen Sprachraum Schutzeichel/Szczepaniak 2015:164). Barteld et al. (2016:398) weisen jedoch auf einen erheblichen Unterschied in der Großschreibung von Personenbezeichnungen hin. Während rund 71% der Männer bezeichnenden Substantive großgeschrieben werden, sind es bei den Frauen nur rund 44%. Sie führen diesen Befund allgemein auf die sehr patriarchale Gesellschaftsstruktur dieser Zeit zurück und speziell auf die Textsorte, die ein zugespitzt misogynen Setting aufweist, da die Hexenverfolgung überwiegend gegen Frauen gerichtet war, während die (juristischen) Autoritäten in den Verfahren Männer waren (vgl. Barteld et al. 2016:398). Die "Nicht-Auszeichnung geschlechterspezifischer Personenbezeichnungen" durch die Majuskel (Szczepaniak/Barteld 2016:46) scheint ein Ausdruck der geringen gesellschaftlichen Relevanz (bedingt durch niedrigen sozialen Status, vgl. Szczepaniak 2011:347) der weiblichen Referentinnen zu sein. Barteld et al. (2016:399) deuten dies zunächst als Zeichen dafür, dass neben Belebtheit pragmatische Faktoren eine Rolle spielen.

In einer unveröffentlichten Hausarbeit konnte ich, ebenfalls am Korpus der Hexenverhörprotokolle, zeigen, dass die konkrete soziale Rolle bzw. die moralische Bewertung dieser sich signifikant auf die Großschreibung der Frauen bezeichnenden Appellativa auswirkt, während die Männer bezeichnenden Appellativa unabhängig von

¹ Das Projekt wurde insgesamt von 2013 bis 2019 von der DGF finanziell unterstützt, s. <https://www.uni-bamberg.de/germ-ling/forschung-und-lehre/forschungsprojekte/sigs/>.

der moralischen Bewertung überwiegend großgeschrieben werden (vgl. Dücker 2020:32). Dieser Befund könnte darauf hindeuten, dass Frauen von den Schreibern² weniger als (prototypische) Menschen und als relevant wahrgenommen wurden und somit weitere Erwägungen Einfluss auf die Schreibung nehmen konnten. Nicht miteinbezogen wurden in dieser Untersuchung die anzunehmende Varianz in der Großschreibung von Frauen bezeichnenden Appellativa je nach Schreiber, welche insgesamt sehr unterschiedliche Großschreibungstendenzen aufweisen (vgl. Schutzeichel/Szczepaniak 2015:165). Ein weiterer möglicher, noch nicht mituntersuchter Faktor ist der des jeweiligen Lemmas, und zwar bei den Männer bezeichnenden Appellativa: Bei diesen handelt es sich teilweise um (Ehren-)Titel, die positiv bewertet und prestigeträchtig sind, was sich über die Ehrerbietungsfunktion der Großschreibung (vgl. Bergmann 1999:73) auf die Großschreibung auswirken könnte. Ein weiterer relevanter Faktor ist LEMMAFREQUENZ (vgl. Barteld et al.2016:404), die Auswirkung soll durch Einbeziehung als fester Effekt kontrolliert werden.

In dieser Arbeit werden die Faktoren LEMMAPRESTIGE und LEMMAFREQUENZ sowie als Störvariable SCHREIBER in eine multifaktorielle Analyse der Auswirkung von GENDER und MORALISCHER BEWERTUNG auf die Großschreibung von Personen bezeichnenden Appellativa miteinbezogen. Dafür wird zunächst ein Überblick über die Faktoren in der Entwicklung der satzinternen Großschreibung gegeben und die für diese Untersuchung relevanten Faktoren beschrieben (Kap. 2). Im Anschluss wird nach Vorstellung des Korpus der Hexenverhörprotokolle und des verwendeten Teilkorpus das Vorgehen in dieser Untersuchung geschildert (Kap. 3). Die vorliegenden Daten werden in dieser Arbeit ausführlich und visuell unterstützt präsentiert (Kap. 4). Schließlich werden die Ergebnisse der Analyse berichtet und interpretiert (Kap. 5). Die Arbeit schließt mit einem Fazit und Ausblick (Kap. 6).

2 Faktoren in der Entwicklung der satzinternen Großschreibung

In diesem Kapitel wird ein Überblick über die verschiedenen Entwicklungsphasen der satzinternen Großschreibung gegeben, die für diese Untersuchung relevanten kognitiv-semantischen und pragmatischen Steuerungsfaktoren vorgestellt und das Phänomen der unterschiedlichen Majuskelverwendung für männliche und weibliche ReferentInnen

² *Schreiber* wird hier bewusst nicht gendert, da diese Tätigkeit nur von Männern verrichtet wurde.

diskutiert (Kap. 2.1). Anschließend werden Untersuchungshypothesen aufgestellt (Kap. 2.2).

2.1 Phasen und Steuerungsfaktoren der Majuskelverwendung

Die Entwicklung der satzinternen Großschreibung von Substantiven im Deutschen erfolgte zunächst ohne kodifizierte Regelungen durch Grammatiken: Sie entstand im Schriftgebrauch und wurde erst später, als sie sich schon weitestgehend durchgesetzt hatte, in die Regelwerke aufgenommen (vgl. Bergmann 1999:93f., Kaempfert 1980:73). Der graphematische Wandel wurde von multiplen, interagierenden Faktoren gesteuert, die in den verschiedenen Phasen der Majuskelausweitung unterschiedlich stark wirkten. Insgesamt lässt sich ein grammatikalisierungsähnlicher Prozess erkennen, in dem die Majuskel erst bestimmte kognitive Kategorien ausdrückt, im Folgenden für die betreffende semantische Gruppe konventionalisiert und schließlich für alle Appellativa und Nominalphrasenköpfe grammatikalisiert wird (vgl. Barteld et al. 2016:394). Die historische Entwicklung der Großschreibung lässt sich in vier Phasen einteilen (vgl. Szczepaniak 2011:357).

Phase	Faktoren	Zunehmende von...	Gr.schr.
1	Relevanz, Individualität, Referentialität	Eigennamen, Sacra	Nomina
2	(Relevanz), Referentialität, Individualität, Belebtheit	Personen	
3	(Referentialität),(Individualität), Belebtheit	Konkrete	
4	(Belebtheit), syntaktische Funktion (Kerne von Nominalphrasen)	Abstrakta	

Tabelle 1: Phasen der Majuskelgrammatikalisierung nach Szczepaniak (2011:357).

In Tabelle 1 bezeichnet die Klammer jeweils eine Abschwächung des betreffenden und vorher stärker wirksamen Steuerungsfaktors. In der ersten Entwicklungsphase wurde die Majuskel expressiv zur Hervorhebung religiös oder gesellschaftlich besonders relevanter Entitäten verwendet.³ Die Ehrerbietungsfunktion der Majuskel führte anfangs zu ihrer Verwendung für Gottesbezeichnungen und Heiligennamen sowie Titel,

³ Andere von der satzinternen Großschreibung betroffene Wortarten als Substantive, wie Pronomina und Adjektive, werden in dieser Arbeit nicht behandelt. Siehe hierzu u.a. Kämpfert 1980 (Pronomina) und Bergmann/Nerius 1998.

Standes- und Amtsbezeichnungen sozial angesehener Personen (vgl. Bergmann 1999:73). Zusammen mit Relevanz wirkten hier die Faktoren Individuiertheit, also wie distinkt wahrnehmbar eine Entität ist, und Referentialität. Prototypische Individuen sind Menschen, ihr Individuiertheitsgrad hängt noch von verschiedenen grammatisch verstellbaren Parametern wie Definitheit und Singularität ab (vgl. Szczepaniak 2011:345). Eine bestimmte einzelne Person wird distinkter wahrgenommen als eine unbestimmte. Im Plural nimmt die Distinktivität und damit Individuiertheit ebenfalls ab, wie in den Beispielsätzen 1-3 deutlich wird, in denen der Individuiertheitsgrad mit der Reduktion von Definitheit und Singularität abnimmt:

- 1) Ich habe die Bundestagsabgeordnete gesehen.
- 2) Ich habe eine Bundestagsabgeordnete gesehen.
- 3) Ich habe eine Gruppe Bundestagsabgeordnete gesehen.

Den höchsten Individuiertheitsgrad haben Personennamen, die monoreferentiell auf eine einzige, bestimmte Person verweisen.

- 4) Ich habe Detlev Detlevsen gesehen.

Personennamen folgten in der Entwicklung der satzinternen Großschreibung auf die Heiligennamen und Ehrentitel. In den von Bergmann/Nerius untersuchten frühneuhochdeutschen Drucken werden Personen- und auch Ortsnamen bereits um 1530 zu über 90% großgeschrieben (vgl. Bergmann 1999:69). In den Hexenverhörprotokollen werden Eigennamen "almost systematically" (zu 89%) großgeschrieben (Barteld et al. 2016:394f.). Für diese erste Phase der Entwicklung spricht Szczepaniak (2011:351) von "pragmatic strengthening" der Majuskel. Kaempfert (1980), der vier umfangreiche Drucke aus dem 16. Jahrhundert auf ihre Groß- und Kleinschreibung untersucht, nennt zwei weitere pragmatische Großschreibungsprinzipien; zum einen die "thematische Auszeichnung" von Wörtern, die das Thema des Textes bezeichnen (vgl. hierzu auch Szczepaniak 2011:352), zum anderen "die affektive Auszeichnung" von im Kontext beliebten Entitäten (Kaempfert 1980:81).

Die nächste Phase der Majuskelausweitung erfasste nicht-onymische Personenbezeichnungen, deren Individuiertheitsgrad geringer ist als der von Eigennamen, deren Belebtheitsgrad jedoch gleich ist. Die weitere Ausbreitung der

Großschreibung auf andere Appellativa verlief entlang der Belebtheitsskala. Diese anthropozentrische semantische Einteilung von Substantiven beruht auf kognitiven Grundkategorien von Entitäten in der Welt. Grob wird zwischen belebt und unbelebt unterschieden, feiner kann in Menschen, andere Lebewesen, konkrete Gegenstände und Abstrakta eingeteilt werden (s. Abb. 1). Es kursieren hierzu unterschiedlich fein granulierten Schemata.⁴

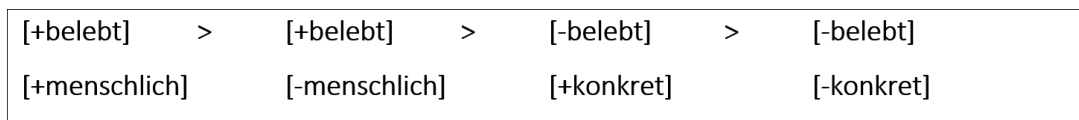


Abbildung 1: Belebtheitsskala mit Konkretheitsgraden.

Nomina Sacra lassen sich in die hier beschriebene Skala nicht einordnen, können aber in vielen Fällen, wie Szczepaniak (2011:352) vorschlägt, an der Spitze der Belebtheitsskala angesetzt werden, da sie als unsterbliche Wesen noch belebter sind als Menschen.

Zeitschnitt	Eigennamen	Nomina sacra	Personenbezeichnungen	Konkrete	Abstrakta
1500	59 %	0 %	11 %	4 %	2 %
1530	74 %	68 %	34 %	8 %	5 %
1560	97 %	90 %	72 %	40 %	18 %
1590	96 %	98 %	91 %	84 %	50 %
1620	98 %	99 %	96 %	91 %	66 %
1650	99 %	100 %	93 %	93 %	72 %
1680	99 %	100 %	96 %	99 %	87 %
1710	99 %	100 %	98 %	94 %	88 %

Tabelle 2: Entwicklung der satzinternen Großschreibung in gedruckten Texten nach Bergmann/Nerius (1998:851).

Tab. 2 zeigt die von Bergmann/Nerius (1998) festgestellte Ausweitung der Großschreibung entlang der Individualitäts- und Belebtheitsskala, angegeben ist jeweils der prozentuale Anteil großgeschriebener Token in den an der Belebtheits- und Individualitätsskala angeordneten Substantivgruppen. Eigennamen und Nomina Sacra, die obere Spitze beider Skalen, sind bereits 1560 zu über 90% großgeschrieben, es folgen Personenbezeichnungen und im nächsten Zeitabschnitt die Konkrete. Abstrakta, das kognitiv unauffällige untere Ende der Skalen, erreichen bis zum Ende des Untersuchungszeitraumes nur 89% Großschreibung. Auch in den handschriftlichen Hexenverhörprotokollen erweist sich der Belebtheitseinfluss als signifikant (vgl. allgemein Barteld et al. 2016:396f. oder speziell für den norddeutschen Sprachraum Schutzeichel/Szczepaniak 2015:164).

⁴ Für eine simplere Skala s. beispielsweise Silverstein (1976); das SiGs-Projekt verwendet eine feinere Abstufung (s. Kap. 3.1).

2.1.1 GENDER UND MORALISCHE BEWERTUNG als Faktoren für Großschreibung

Innerhalb der Personenbezeichnungen scheinen allerdings weitere Faktoren eine Rolle zu spielen, wie Barteld et al. (2016:398) feststellen. Wie eingangs erwähnt, zeigt sich in den Hexenverhörprotokollen zwischen auf Frauen bezogenen Personenbezeichnungen und solchen, die sich auf Männer beziehen, ein signifikanter Unterschied in der Großschreibung.⁵ Wie Abb. 3 zeigt, wurden Frauen bezeichnende Appellativa mit 104 von 233 Belegen zu 45 % großgeschrieben, die Männer bezeichnenden hingegen, mit 224 von 316 Belegen, zu 71 %.

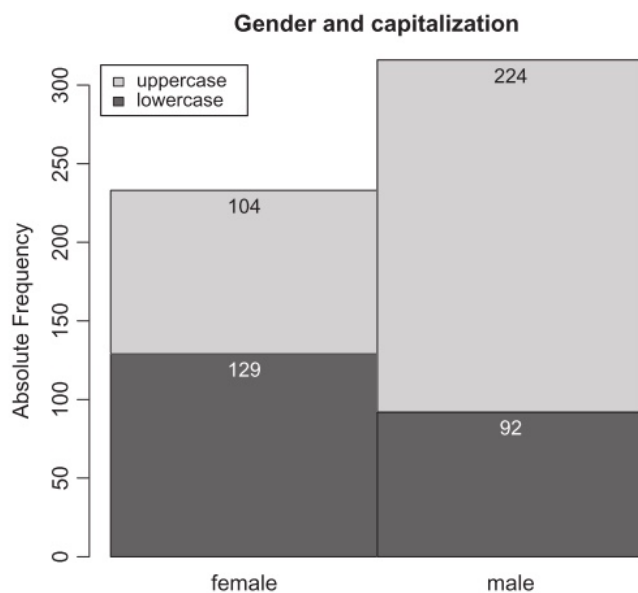


Abbildung 2: Genderdifferenzierte Großschreibung von Personenbezeichnungen (n=549) (aus Barteld et al. 2016:398).

Dieser Unterschied sei, wie eingangs erwähnt, angesichts der patriarchalen Gesellschaft sowie der speziellen Textsorte nicht überraschend:

Across all texts, most of the incriminated persons are women, while the representatives of the legal authorities are men. This obviously amplifies the differences in social status, showing that, in the texts under investigation, pragmatic factors still play a major role in the choice between uppercase and lowercase initials (Barteld et al. 2016:398f.).

Es lässt sich festhalten, dass das in den Hexenverhörprotokollen festgestellte verzögerte Fortschreiten der Großschreibung von Frauen bezeichnenden Appellativa dem erwiesenen Einfluss der Belebtheit, Individuiertheit und Referentialität auf die

⁵ Ergebnis des Chi-Quadrat-Test mit Yates' Continuity Correction für die in Abb. 3 gezeigten Werte: $\chi^2(1) = 37.343$, $p < 0.0005$.

Großschreibung widerspricht, beziehungsweise es aus heutiger Perspektive zu tun scheint. Frauen werden schließlich, ebenso wie Männer, als belebt und menschlich wahrgenommen, sind ebenso prototypische Individuen und relevant. Zu vermuten wäre, dass Frauen durch ihre sozial schlechtere Stellung in der Gesellschaft des 16.–17. Jahrhunderts und ihre im Kontext der Hexenverfolgung noch verstärkte moralische Abwertung als weniger (prototypisch) menschlich, individuell und relevant wahrgenommen wurden⁶ oder, dass die Ehrerbietungsfunktion oder die Funktion der affektiven Auszeichnung durch die Majuskel die Großschreibung bei moralisch schlecht bewerteten Referentinnen blockierte. In dieser Arbeit wird angenommen, dass sich MORALISCHE BEWERTUNG im Zusammenspiel mit GENDER auf die Großschreibung des Appellativums auswirkt und der Faktor daher als fester Prädiktor im gemischten Modell verwendet.

2.1.2 Einfluss des Lemmas (LEMMAPRESTIGE)

Kaempfert 1980 beobachtet in seiner Untersuchung einiger früher Drucke, dass die Großschreibung teilweise nicht

„an dem besonderen Rang der jeweils bezeichneten Person [hängt], sondern am sozialen, allgemeinen Rang der Institution oder Rolle; so werden: z. B. vattr, mutttr, kind in Lukas 2 kleingeschrieben, obwohl sie die Personen der Heilsgeschichte bezeichnen“(Kaempfert 1980:87).

Unter den hier untersuchten Personenbezeichnungen sind einige, bei denen ein besonderer Rang jeweils schon im Lemma verankert ist, wie bei *bürgermeister*⁷. Im Protokoll aus Bamberg (1628) (Macha et al. 2005)⁸ beispielsweise finden sich dazu drei großgeschriebene Belege. Zweimal ist der Referent in der Rolle „Denunzierter“, einmal „Angeklagter“, alle Belege wurden trotz negativer moralischer Bewertung großgeschrieben. Ob neben dem Faktor GENDER die Prestigeträchtigkeit des Lemmas (im Folgenden kurz LEMMAPRESTIGE) für diese Fälle eine Rolle spielt, soll durch die Einbeziehung in die multifaktorielle Analyse überprüft werden.

⁶ Dies müsste kein rein historisches Phänomen bezeichnen, einiges spricht dafür, dass auch heute noch ähnliche Wahrnehmungsunterschiede bestehen, beispielsweise, dass generisches Femininum in Sätzen wie "Alle Menschen werden Schwestern" markiert ist (vgl. Pusch 1984:15-19).

⁷ Die Lemmata werden, unter Beibehaltung der Kleinschreibung, nach dem Deutschen Wörterbuch von Jacob und Wilhelm Grimm zitiert.

⁸ Auf die einzelnen Hexenverhörprotokolle wird im Folgenden durch Entstehungsort und -jahr verwiesen, auf die Nennung der Edition wird der Kürze halber verzichtet.

2.1.3 SCHREIBERSPEZIFISCHE UNTERSCHIEDE

Die bisherige Auswertung des Kernkorpus ergab unter anderem eine starke Variation in der Großschreibungstendenz zwischen den einzelnen Protokollen. Diese variiert zwischen nur 3 % im Göttinger Protokoll und fast 81 % in dem aus Ostrau (vgl. Dücker et al. (i. Vorb.:10)). Die Variation lässt sich nicht mit den angenommen zeitlichen oder räumlichen Entwicklungen in Verbindung bringen (vgl. Dücker et al. (i. Vorb.:10)). Dücker et al. gehen daher davon aus, dass es sich um ideolektale Variation handelt (ebd.:10).⁹ Abb. 4 zeigt die Groß- und Kleinschreibung der Personenbezeichnungen im hier verwendeten Kernkorpus.

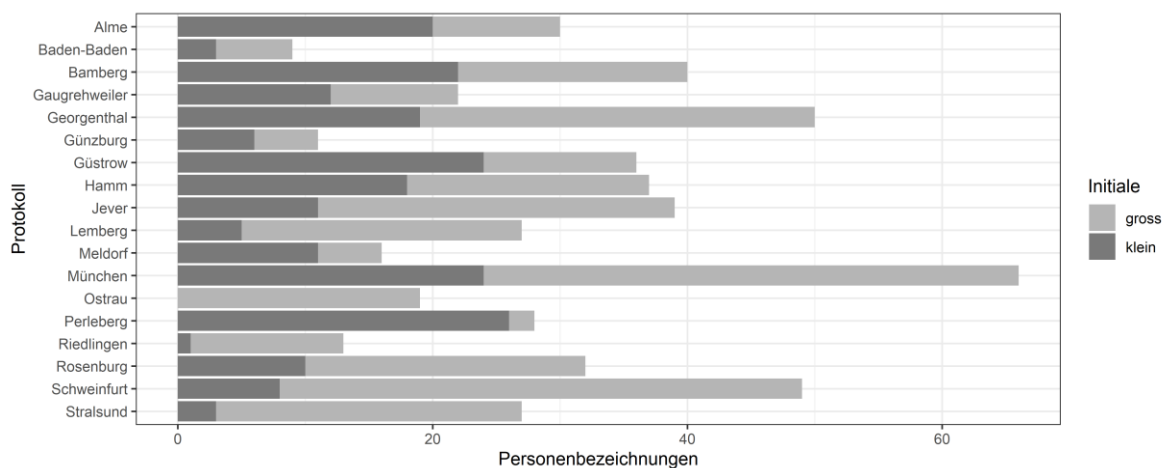


Abbildung 3: Groß- und Kleinschreibung der Personenbezeichnungen nach Protokollen (n=551).¹⁰

In einigen Protokollen, zum Beispiel Stralsund, Riedlingen und Lemberg, werden generell nur ein Bruchteil der Personenbezeichnungen klein geschrieben, in einigen ist das Verhältnis von Groß- und Kleinschreibung beinahe ausgewogen, zum Beispiel in den Texten aus Hamm und Bamberg. In Perleberg, Meldorf und Alme hingegen liegt überwiegend Kleinschreibung vor. Die verschiedenen Steuerungsfaktoren werden von den einzelnen Schreibern also unterschiedlich gewichtet, sie scheinen angesichts der noch nicht kodifizierten variierenden Großschreibungspraktiken individuell unterschiedliche Regeln abgeleitet zu haben (vgl. Schutzzeichel/Szczepaniak 2015:165).

⁹ Bei der Annahme eines Schreivers pro Protokoll handelt es sich um eine grobe Annäherung an die Wirklichkeit. Zwar ist bei den meisten Protokollen von einem einzigen Schreiber auszugehen, jedoch war es nicht immer nachweislich nur einer (vgl. Dücker et al. (i. Vorb.:12)).

¹⁰ Sämtliche nicht als den Arbeiten Anderer entnommen markierte Abbildungen wurden von mir mithilfe von R Studio erstellt. Der verwendete Code für alle Abbildungen, Berechnungen sowie das in Kapitel 5 vorgestellte generalisierte lineare Model sowie weitere Tests ist als digitaler Anhang verfügbar.

2.1.4 FREQUENZ

Barteld et al. (2016) stellen, ebenfalls am Korpus der Hexenverhörprotokolle, Frequenzeffekte¹¹ sowie eine starke Interaktion von Frequenz mit weiteren Faktoren, nämlich BELEBTHEIT und GENDER fest:

The interaction patterns between frequency and animacy suggest that, roughly speaking, cognitive and pragmatic factors enhance or inhibit the degree of capitalization especially in the domain of high-frequency items: both for inanimates and for terms denoting women, capitalization becomes less likely with increasing frequency (Barteld et al. 2016:407).

Frequenz führt also in Zusammenspiel mit hohem Belebtheitsgrad zu verstärkter Großschreibung, allerdings nicht, wenn weibliches Gender vorliegt. Der Faktor FREQUENZ muss daher in dieser Untersuchung kontrolliert werden und wird als fester Effekt in die Untersuchung mit einbezogen. Wie in der Untersuchung von Barteld et al. 2016 wird auch hier als Basis die Lemmafrequenz innerhalb des selben Korpus verwendet, eine Problematisierung dieses Vorgehens erfolgt in Kap. 5.

2.2 Hypothesenaufstellung

In dieser Untersuchung wird angenommen, dass die untersuchte Variable GROßSCHREIBUNG VON PERSONENBEZEICHNUNGEN maßgeblich von einem Zusammenspiel der Faktoren GENDER und SOZIALE BEWERTUNG abhängt. Es werden folgende gerichtete Hypothesen aufgestellt, überprüft wird jeweils die Nullhypothese:

Hypothese 1) Der Faktor SOZIALE BEWERTUNG wirkt sich nur dann auf die Großschreibung der Personenbezeichnung aus, wenn die bezeichnete Person weiblichen Genders ist:

H1₁: Großschreibung von Frauenbezeichnungen tritt bei positiver Bewertung häufiger auf als bei negativer oder neutraler Bewertung.

H1₀: Großschreibung von Frauenbezeichnungen tritt bei positiver Bewertung nicht häufiger auf als bei negativer oder neutraler Bewertung.

Hypothese 2) Der Faktor soziale Bewertung wirkt sich nicht auf die Großschreibung der Personenbezeichnung aus, wenn die bezeichnete Person männlichen Genders ist.

H2₁: Großschreibung von Männerbezeichnungen tritt bei positiver Bewertung nicht häufiger auf als bei negativer oder neutraler Bewertung.

¹¹ Sie arbeiten mit der Lemmafrequenz innerhalb desselben Korpus (vgl. Barteld et al. 2016:399f.).

H2₀: Großschreibung von Männerbezeichnungen tritt bei positiver Bewertung häufiger auf als bei negativer oder neutraler Bewertung.

Desweiteren wird die Vermutung überprüft, dass sich im Lemma mitschwingendes Prestige auf die Großschreibung auswirkt.

Hypothese 3) Der Faktor LEMMAPRESTIGE wirkt sich positiv auf die Großschreibung von Personenbezeichnungen aus:

H3₁: Großschreibung tritt häufiger auf, wenn LEMMAPRESTIGE vorliegt.

H3₀: Großschreibung tritt nicht häufiger auf, wenn LEMMAPRESTIGE vorliegt.

Als Schwellwert wird hier $p < 0.05$ angesetzt, d.h. eine Nullhypothese kann dann abgelehnt werden, wenn die Wahrscheinlichkeit, dass die jeweiligen Ergebnisse oder extremere Ergebnisse erzielt würden, wenn die Nullhypothese wahr ist, unter 0.05 liegt (vgl. Levshina 2015:13).

3 Methodische Grundlegung

In Kapitel 3.1 wird zunächst das SiGs-Projekt vorgestellt. In Kapitel 3.2 wird das in dieser Untersuchung verwendete Korpus beschrieben und in Kapitel 3.3 die methodische Vorgehensweise erläutert.

3.1 Das SiGs-Projekt

Als Korpus für diese Untersuchung dient das Kernkorpus des SiGs-Projekts. Im Projekt wurde die Groß- und Kleinschreibung in den frühneuhochdeutschen handschriftlichen Hexenverhörprotokollen erforscht. In handschriftlichen Texten ist eine langsamere Durchsetzung der satzinternen Großschreibung als in Drucken zu beobachten.¹² Dies ist unter anderem durch das merkantile Interesse der Drucker zu erklären: Die Majuskelsetzung scheint als Verständniserleichterung verstanden worden zu sein, weshalb die Drucker sie verstärkt einsetzten, um somit leichter zugängliche Texte einem breiteren Publikum verkaufen zu können; die Entwicklung der Großschreibung fand in einer Phase der Literarisierung der Gesellschaft und Markterschließung des neuen Druckwesens statt (vgl. Barteld et al. 2016:386). Bei Handschriften besteht durch die geringere Planungszeit und Anzahl involvierter Personen sogenannter "'online'-Charakter", in besonderem Maße ist dies bei Mitschriften der Fall (Szczepaniak/Barteld

¹² Moulin (1990) zeigt dies im Vergleich der Bibeldrucke mit den handschriftlichen Briefen Martin Luthers.

2016:45). Eine Annahme im Projekt ist, dass sich die geringere Planungszeit auf die Majuskelsetzung auswirkt (vgl. Szczepaniak/Barteld 2016:45f.). Durch die relative Spontaneität der Produktion und die noch nicht standardisierte Schriftlichkeit werden im Projekt genauere Erkenntnisse über die folgenden Steuerungsfaktoren der Majuskelausweitung erwartet:

1) (Morpho-)Syntaktische Faktoren

- a. Wortartenzugehörigkeit
- b. Struktur der Nominalphrase
- c. syntaktische Funktion

2) Kognitiv-semantische Faktoren

- a. Belebtheit
- b. Individualität (auch: Individuiertheit)
- c. Referentialität
- d. semantische Rolle

3) Sprachgebrauchsfaktoren

- a. Gebrauchsfrequenz
- b. Idiolektalität
- c. Buchstabenform

Übersicht entnommen aus Szczepaniak/Barteld (2016: 49-50).

Für diese Untersuchung sind die kognitiv-semantischen Faktoren Belebtheit, Individualität und Referentialität relevant. Anstelle von semantischen Rollen wird die konkrete soziale Rolle im Verhörkontext untersucht.

Die Hexenverhörprotokolle wurden im Rahmen des Projekts tokenisiert, lemmatisiert und auf Wortart, Belebtheit und Numerus, syntaktische Funktion und semantische Rolle annotiert (vgl. Barteld et al. 2016:393f.). Die Belebtheitsannotation der Personenbezeichnungen in SiGs differenziert wie folgt:

Abstrakta

abstrakt
abstrakt (Maß)

Konkrete

konkret
konkret (Körperteil)
konkret (Ort)
konkret (Kollektivum)



Tierische Entitäten

tierisch

tierisch (Kollektivum)

Menschliche Entitäten

menschlich

menschlich (Kollektivum)

Übermenschliche Entitäten

übermenschlich (positiv)

übermenschlich (Teufel)

+ Belebtheit

Auflistung entnommen aus den Annotationsrichtlinien des SiGs Projekts (Barteld et al. i. Vorb.:37).

Diese Arbeit beschäftigt sich ausschließlich mit den als „menschlich“ annotierten Personenbezeichnungen, da sich solchen der Kategorien „menschlich (Kollektivum)“ oder „übermenschliche Entitäten“ keine soziale Rolle im Prozess zuweisen lässt. Bezüglich letzterer Kategorie ist erwähnenswert, dass die Tokenisierung der Protokolle, um abweichende graphematische und syntaktische Wortgrenzen abbilden zu können, auf zwei Ebenen angelegt ist (vgl. Barteld et al. 2016:394). Dadurch kann beispielsweise ein auseinandergeschriebenes Kompositum graphisch als zwei Token, syntaktisch jedoch als ein einziges abgebildet werden. Die zwei Ebenen werden auch für semantische Unterscheidungen genutzt, nämlich bei metonymischen Verwendungen. Das Wort *buhle* bedeutet zum Beispiel eigentlich 'liebhaber', 'geliebte', 'unehelich man' oder 'unehelich weip' (<http://www.woerterbuchnetz.de/DWB?lemma=buhle>), womit es also Menschen bezeichnet, es steht aber auch für den Teufel. In solchen Fällen wird auf der graphematischen Ebene „menschlich“, auf der syntaktischen Ebene „übermenschlich“ annotiert. Solche metonymischen Verwendungen, die für übermenschliche Entitäten stehen, können für diese Untersuchung ebenfalls nicht berücksichtigt werden.

3.2 Korpusbeschreibung

Das Kernkorpus beinhaltet achtzehn der von Macha et al. (2005) edierten Hexenverhörprotokolle, die zwischen 1588 und 1630 entstanden und gleichmäßig auf sechs Dialektgebiete verteilt sind.¹³ Es handelt sich dabei um Jever (1592), Meldorf (1618) und Alme (1630) aus dem nordwestlichen, Perleberg (1588), Güstrow (1615) und Stralsund (1630) aus dem nordöstlichen, Hamm (1592), Gaugrehweiler (1610) und

¹³ Eine Karte zur regionalen Verteilung der Protokolle findet sich im Anhang.

Lemberg (1630) aus dem mittelwestlichen, Georgenthal (1597), Rosenberg (1618) und Ostrau (1628) aus dem mittelöstlichen, Riedlingen (1596), Günzburg (1613) und Baden-Baden (1628) aus dem südwestlichen und München (1600), Schweinfurt (1616) und Bamberg (1628) aus dem südöstlichen Dialektraum. Der Umfang der Protokolle variiert zwischen 948 (Gaugrehweiler) und 1.934 (München) Token.¹⁴ Die Handschriften wurden von unterschiedlichen, größtenteils unbekannten Schreibern geschrieben, die namentlich bekannten unter ihnen waren professionelle Stadtschreiber oder Notare (vgl. Szczepaniak/Barteld 2016:48). Teilweise handelt es sich um Mitschriften, teilweise um Abschriften, für den Großteil fehlt diese Information (vgl. Szczepaniak/Barteld 2016:46–48). Da in der zumeist verwendeten gotischen Kurrentschrift bei den Graphemen <h/H>, <v/V> und <z/Z> nicht eindeutig zwischen Majuskel und Minuskel differenziert werden kann, werden die Wörter mit diesen Initialen (im Kernkorpus betrifft dies insgesamt 961 Token) nicht berücksichtigt (s. hierzu Macha et al. 2005:XXIII und Szczepaniak/Barteld 2016: 69).

3.3 Untersuchte Teilmenge

Diese Arbeit untersucht sämtliche mit dem Belebtheitsgrad „menschlich“ annotierten Nomina (NN) und Berufsbezeichnungen (NB) des Kernkorpus, denen eine der Genderkategorien weiblich oder männlich zugeordnet werden kann.¹⁵ Damit scheiden aus den Personenbezeichnungen Wörter wie *kind* oder *Persohn* aus, außerdem Pluralformen wie *Gefangene*, *anwesendte* oder *auererschen* ('Außerirdische'). Diese Vorauswahl ergibt letztendlich eine Liste von 244 weiblichen und 307 männlichen Personenbezeichnungen. Die Differenz zu den Zahlen der Untersuchung von Barteld et al. (2016) (233 mal weiblich und 316 mal männlich) ergibt sich aus einigen Korrekturen meinerseits. Der Beleg *Sandtman* (Güstrow 1615) wurde eliminiert, da dieser anhand des Kontextes eindeutig als Eigenname (NE) identifiziert werden konnte.^{16,17} Die zu den Lemmata *buhle* und *buhlin* gehörenden Belege *Puhlin* (Bamberg 1628), *Buol* (Riedlingen

¹⁴ Eine Tabellarische Übersicht über die Texte im Kernkorpus findet sich bei Szczepaniak/Barteld (2016:46f).

¹⁵ Diese Daten samt zusätzlicher Genderannotation wurden mir freundlicherweise von Dr. Stefan Hartmann als csv-Datei zur Verfügung gestellt.

¹⁶ Eigennamen verhalten sich auch in Bezug auf die Großschreibung anders als Appellativa und werden daher aus dieser Untersuchung ausgeklammert (vgl. Kap. 2).

¹⁷ Die ebenfalls der Zauberei beschuldigte Tochter der Angeklagten sowie deren Mann tragen diesen Namen, die weiteren 8 Vorkommen des Wortes wurden nicht als Nomina (NN) annotiert, weshalb von einem Fehler in der PoS-Annotation ausgegangen wurde

1596) und *Bule* (Georgenthal 1597) wurden aussortiert, da sie in diesem Kontext für Übermenschliches (den Teufel) stehen. Dadurch sind sie für die Untersuchung der sozialen Rolle und moralischen Bewertung nicht verwertbar. Vermutlich handelt es sich um Annotationsfehler, bei denen die graphische und syntaktische Ebene durcheinandergebracht wurden (vgl. Kap. 3.1). In zehn Fällen wurde die vorliegende Genderannotation von männlich zu weiblich korrigiert.¹⁸ Eine Übersicht über die vorgenommenen Korrekturen samt begründenden Informationen zu Kontext oder Editions Hinweisen findet sich in Tabelle 1 im Anhang 2.

3.4 Vorgehen

Neben dem (kategorialen) Merkmal GENDER (zwei Ausprägungen: weiblich vs. männlich) wurde die (kategoriale) ROLLE DER REFERENZPERSON IM PROZESS (sechs Ausprägungen: Angeklagt, Denunziert, Geschädigt, Kläger/in, Zeug/in, Justiz) annotiert. Die Rollen-Annotation wurde unter SOZIALE BEWERTUNG zu drei Kategorien (positiv, neutral, negativ) zusammengefasst (s. Abb. 5).

Rolle	Wertung
Angeklagte/r	negativ
Denunzierte/r	negativ
Geschädigte/r	positiv
Kläger/in	positiv
Justiz	positiv
Zeuge/Zeugin	neutral

Abbildung 4: Bewertung der einzelnen Rollen.

In den Hexenverhören führt die instrumentalisierte, dämonisierende Anschuldigung der angeblichen Hexerei zur negativen Wertung der Rollen „Angeklagte/r“ und „Denunzierte/r“. „Kläger/in“, „Geschädigte/r“ und „Justiz“ hingegen stehen auf der anklagenden, moralisch nicht angezweifelte und damit im Kontext positiv gewerteten Seite (s. Abb. 5). Letztere ist zudem nicht nur selbst die moralische Bewertungsinstanz,

¹⁸ Dies betrifft acht Belege mit dem weiblichen Suffix -in, wobei zusätzlich jeweils auch der direkte Kontext, für *Beürin* (Bamberg 1628) der in der männlichen Form nicht vorhandene Umlaut sowie für *schaffmeisterin* (Rosenburg 1618) die Erklärung der Edition (Macha et al. 2005:320, FN 96) für weibliches Genus sprechen. Die zwei anderen Belege weisen die niederdeutsche feminine Endung -sche auf (*Swegersche* (Jever 1592) und *Geuatters* (Hamm 1592)). Die Edition weist *Geuatters* als Variante von nd. *Gevattersche* aus (MACHA ET AL. 2005:229, FN 155). Zudem spricht der Kontext für weibliches Genus.

die Schreiber stehen als rangniedrigere Angestellte der gleichen Institution auch noch in einem besonderen Respekt- und Abhängigkeitsverhältnis zu den Richtern. Die Rolle „Zeuge/Zeugin“ ist mit Vorsicht zu behandeln. Man könnte zwar meinen, sie stünden ebenfalls auf der moralisch nicht angezweifelte Seite im Prozess, allerdings wurden die belastenden Aussagen nicht selten im Zuge anderer Prozesse gemacht, weshalb die moralische Bewertung der Zeugen und Zeuginnen schwer einzuschätzen ist. Um die Belege nicht ausschließen zu müssen, wurden sie zunächst bezüglich der Bewertung als neutral kategorisiert.

Die Rollen sind teilweise nicht scharf abgrenzbar, es wurde jedoch versucht, die Zuordnung möglichst klar zu halten. Generell wurde in dieser Untersuchung nur dann eine Rolle annotiert, wenn aus dem Kontext eindeutig hervorgeht, dass sich auf eine bestimmte Person bezogen wird, die eine der oben aufgelisteten Rollen innehat. Grammatisch liegt ein eindeutiger Bezug bei Prädikativa vor, da bei Kopula eindeutig ist, dass die Referenz gleichbleibt. Auch konjunktivisch angebrachte Prädikativa wurden gewertet, so zum Beispiel der Beleg *Schelmb* aus dem Bamberger Protokoll (1628), in dem die Aussage des angeklagten Bürgermeisters Johannes Junius wiedergegeben wird.

5) [...] wann Er ein solcher Schelmb wehre, wolte Er sich nicht also Martter[n]
lassen [...]

Die Referenz des Prädikativums *Schelmb* ist mit dem des Subjekts *Er* identisch. Nicht gewertet wurden Vergleiche, da Identität der verglichenen Entitäten grundsätzlich ausgeschlossen ist, sowie generalisierende und damit nicht referentielle Verwendungen. Im Folgenden wird das spezifische Vorgehen bei der Zuordnung der einzelnen Rollen beschrieben. Wer im jeweiligen Prozess Angeklagte/r war, geht zumeist aus der dem Protokoll vorangestellten Inhaltsangabe der Edition hervor und wird außerdem häufig am Anfang des Protokolls klargestellt. Teilweise sind mehrere Personen, in München (1600) beispielsweise eine ganze Familie, angeklagt. Als Denunzierte/r zählt, wer als Teilnehmer/in nächtlicher Teufelstänze aufgezählt wird und wem zauberische Fähigkeiten (wozu auch jegliche Kenntnisse über Kräuter oder Hausmittel zählen) und auch das Weitergeben solcher an andere nachgesagt werden. Des Weiteren wird über verdächtige Tätigkeiten berichtet, wie beispielsweise das Vergraben von Gegenständen oder auch eindeutiger schwerwiegende Fälle wie Mord oder Handel mit Kinderhänden. Es ist im Prozesskontext relativ eindeutig, was als

Denunziation zu gelten hat, da beispielsweise das Schenken von Molchen zum Schutz der Hausbewohner (Georgenthal 1597), wenn es nicht als falsch oder verdächtig erachtet würde, vor Gericht nicht angesprochen würde. Geschädigte/r ist, wer Opfer angeblicher Hexerei wurde, sei es, dass eine Person selbst zu Schaden kam, sei es, dass ihre Angehörigen oder Angestellten oder ihr Besitz oder Vieh geschädigt wurden. In der Logik der Prozesse zählen hierzu auch Opfer von Heilungszaubern. Es wurde nicht differenziert, ob die jeweilige Person nur zu Schaden kommen sollte oder ob es tatsächlich dazu kam. Die Tatsächlichkeit der großenteils abstrusen Anschuldigungen ist ohnehin fragwürdig, oft wird zudem nicht erwähnt, ob die Schädigung tatsächlich stattfand. Die Nennung als 'angedachtes Opfer' reicht m. E. aus, um in der konkreten Situation als Opfer wahrgenommen zu werden. So fallen auch Belege wie *Sohn* im folgenden Beispiel aus Bamberg (1628) in diese Kategorie.

6) Seine Puhlin hette alsobaldt[en] nach seiner verführung begehrt, Er solte seinen Jüngsten **Sohn** hannß Geörgen vmbbring[en] (Hervorhebung L.S.).

Wer der oder die Kläger/in war, geht aus den Inhaltsangaben der Edition sowie dem Protokoll selbst hervor. Bei der Hälfte der Belege zu dieser Rolle handelt es sich um die Wörter *kläger* und *klägerin*. In die Kategorie „Justiz“ wurden neben Richtern und Schöffen auch sämtliche andere dieser Institution angehörige Personen gewertet, so zum Beispiel ein *Gerichtsknecht* (Georgenthal 1597) und ein *Ambtman* (Lemberg 1630). Als Zeuge oder Zeugin zählt, wessen Aussage im Protokoll wiedergegeben wurde, wie es sehr deutlich im folgenden Beispiel aus Alme (1630) der Fall ist:

7) Scheper Gerdt ein zeit hero in starkem Zau- ber gerücht gewesen **testantib[us]** **des Grauen** [INT] !Jorgen schmes Jacobo Timmermans (Hervorhebung L.S.)

Belege, für die keine konkrete Rolle im Prozess festgestellt werden können, weil sie entweder nicht referentiell sind, die Referenz aus dem Kontext nicht eindeutig ersichtlich ist oder die bezeichnete Person am Prozess unbeteiligt ist und zum Beispiel nur ihr Haus als örtliche Orientierung genannt wird, wurden nicht in die Untersuchung einbezogen.¹⁹

¹⁹ Das betrifft insgesamt 135 Belege. 27 sind Frauen- und 108 Männerbezeichnungen. In den Protokollen werden Frauen kaum einfach erwähnt, sie kommen in der Regel in einer Rolle im Prozess vor und erfahren dadurch Bewertung, während Männer häufig auch als neutrale Außenstehende, ohne direkten Zusammenhang zur Prozesssituation Erwähnung finden.

Eine weitere annotierte Kategorie ist PRESTIGE DES LEMMAS (kategorial, zwei Ausprägungen: ja vs. nein). „ja“ wurde bei Lemmata annotiert, in deren Bedeutung eine hierarchische Höherstellung enthalten ist und deren Bezeichneten dadurch besonderer Respekt oder Verehrung zukommt. Dies trifft lediglich bei 18 Lemmata zu, womit insgesamt 67 Belege betroffen sind. Es handelt sich dabei ausschließlich um Männer bezeichnende Appellativa. Abb. 6 zeigt, in welchen Rollen diese auftreten. Die Belege für *amman* wurde im Folgenden als Varianten mit zu *amtman* gezählt.

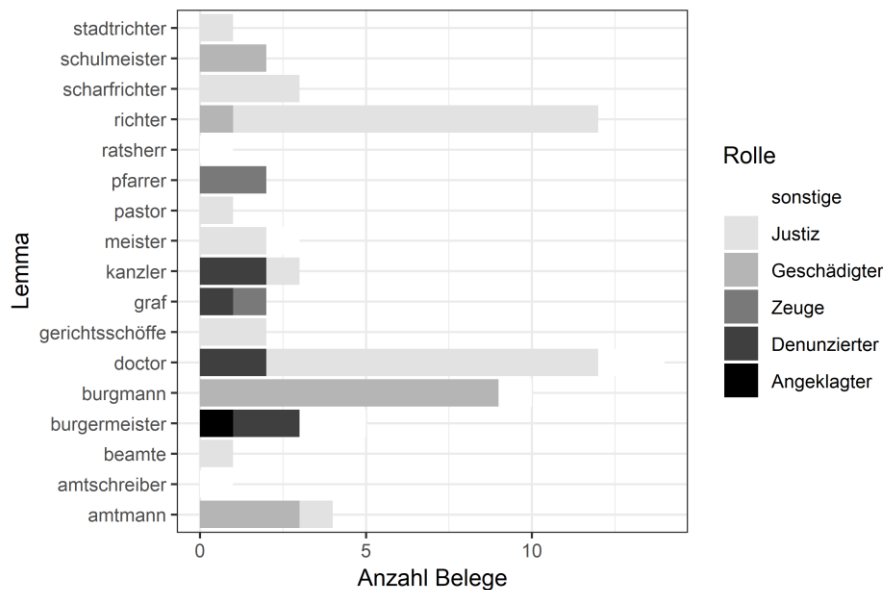


Abbildung 5: Rollen in denen die prestigebehafteten Lemmata (n=17) auftreten (67 Belege).

Die mit prestigebehafteten Lemmata bezeichneten Männer tauchen überwiegend in den positiv gewerteten Rollen „Justiz“ und „Geschädigter“ auf, marginal sind auch „Zeuge“, „Denunzierter“ und „Angeklagter“ vertreten.

4 Datenexploration und Visualisierung

Die Datenexploration und Visualisierung wurde mit Hilfe von R Studio (R Studio Team 2018) und *ggplot2* (Wickham 2016) durchgeführt. Untersucht werden insgesamt 551 Belege, 244 weibliche und 307 männliche Personenbezeichnungen. Abb. 7 zeigt die Anzahl der groß- bzw. kleingeschriebenen Belege für männliche und weibliche Appellativa.

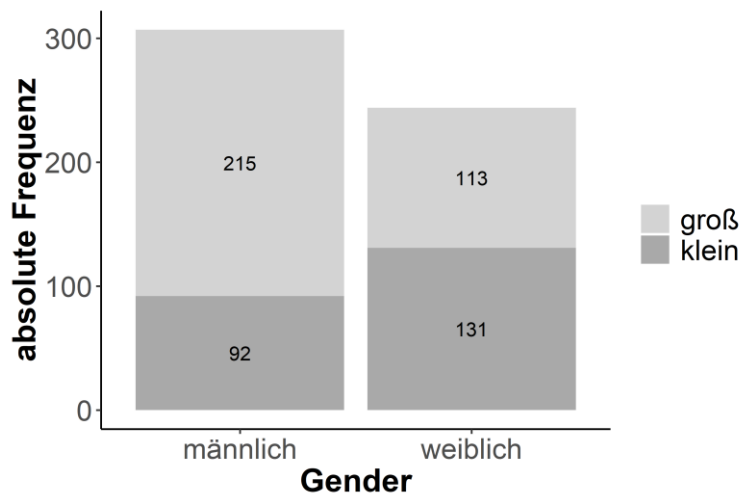


Abbildung 6: Groß- und Kleinschreibung von Personenbezeichnungen im Kernkorpus (n= 551).

Bei den männlichen liegt zu 70 % Großschreibung vor, bei den weiblichen lediglich zu 46,3%. Die beiden Gruppen unterscheiden sich hoch signifikant bezüglich der Großschreibung.²⁰

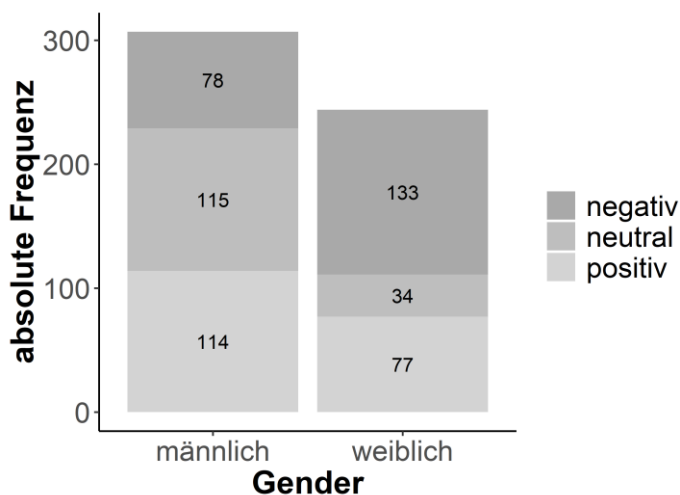


Abbildung 7: Bewertung der Männer und Frauen im Kernkorpus.

Auch die moralische Bewertung ist bei den männlichen und den weiblichen Belegen unterschiedlich verteilt: Abb. 8 stellt den Anteil positiver, neutraler und negativer Bewertung bei den Frauen dem bei den Männern gegenüber.

Bei über der Hälfte der weiblichen Belege ist die Referenz negativ bewertet, bei den männlichen ist es ein Viertel. Bei den weiblichen Belegen gibt es im Gegensatz zu den männlichen kaum neutrale bzw. nicht bewertete Referenzen. Bei den männlichen kommen ebensoviele neutrale wie positiv bewertete Belege vor. Abb. 9 zeigt die

²⁰ $\chi^2 = 30.775$, $df = 1$, $p\text{-value} = 2.897e-08$.

Verteilung der groß- und kleingeschriebenen Belege auf die drei Bewertungsgruppen negativ, neutral, positiv für die männlichen (links) und weiblichen (rechts) Personen.

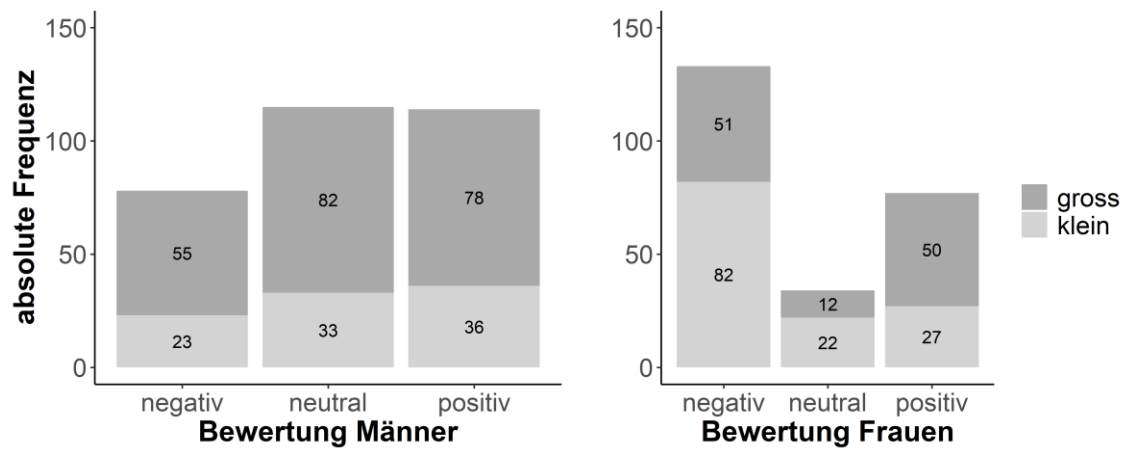


Abbildung 8: Großschreibung und soziale Bewertung bei Männern (n=307) Frauen (n=244).

Der Gender-Unterschied zeigt sich noch deutlicher in den anteiligen Darstellungen in Abb. 10: Bei den Männern liegt unabhängig von der Bewertung zu etwa zwei Dritteln Großschreibung vor, bei den Frauen ist der Anteil der großgeschriebenen Appellativa bei den positiv bewerteten Referentinnen erheblich höher als bei den negativ und neutral bewerteten.

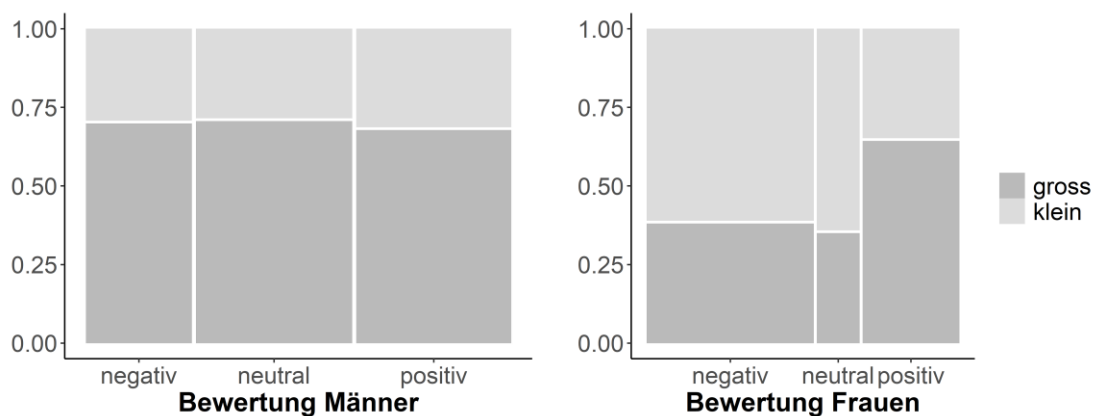


Abbildung 9: Großschreibung und soz. Bewertung bei Männern und Frauen (anteilig).

Bei den Frauen wirken neben der Belebtheit pragmatische/affektive Faktoren auf die Großschreibung (vgl. Kap. 2). Für die Männer wird überprüft, ob LEMMAPRESTIGE einen Einfluss auf die Großschreibung der Appellativa hat. Abb. 11 zeigt die Großschreibung der Männer nach Rollen bei neutralen (links) und prestigebehafteten Lemmata.

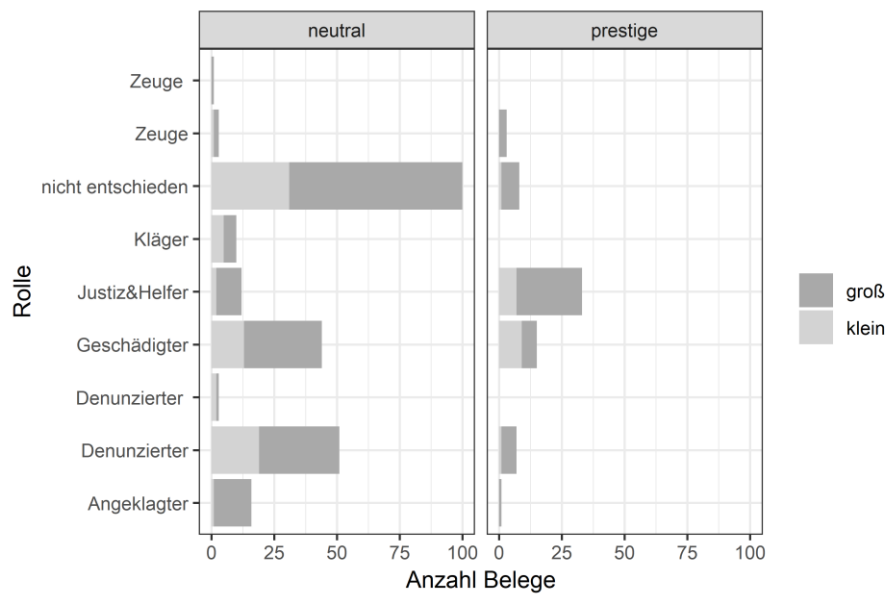


Abbildung 10: Großschreibung Männer bezeichnender Appellativa nach Rolle bei neutralem oder prestigebehaftetem Lemma (n=67).

Die prestigebehafteten Lemmata machen generell nur einen geringen Teil der Belege aus, und zwar 67 (von 304 Männer bezeichnenden Appellativa). Davon sind 46 groß- und 18 klein geschrieben. Das Verhältnis von Groß- und Kleinschreibung ist also sehr ähnlich wie bei den nicht prestigebehafteten, neutralen Lemmata (s. Abb. 12). Abb. 11 zeigt, dass in den negativ bewerteten Rollen auftretende prestigebehaftete Lemmata beinahe ausschließlich großgeschrieben werden, während bei den neutralen Lemmata zumindest bei den Denunzierten auch Kleinschreibung vorkommt.

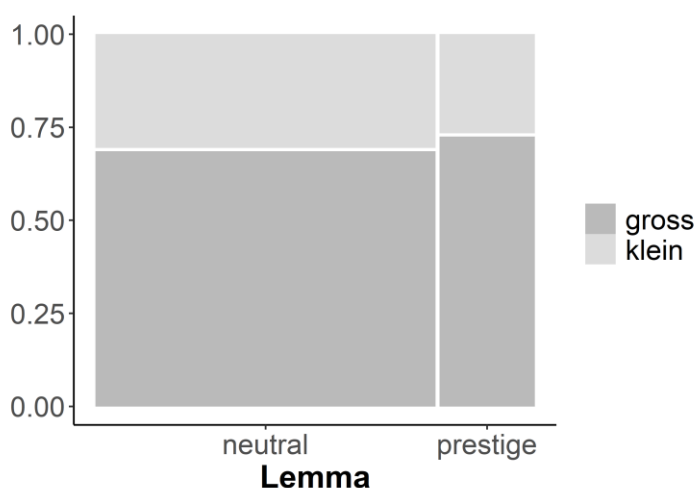


Abbildung 11: Anteil Groß- und Kleinschreibung neutraler und prestigebehafteter männlicher Personenbezeichnungen.

Es sieht nicht danach aus, als ob der Faktor LEMMAPRESTIGE sich besonders auf die Großschreibung der Männer bezeichnenden Appellativa auswirkt, dennoch wird er in die Multifaktorielle Analyse miteinbezogen werden.

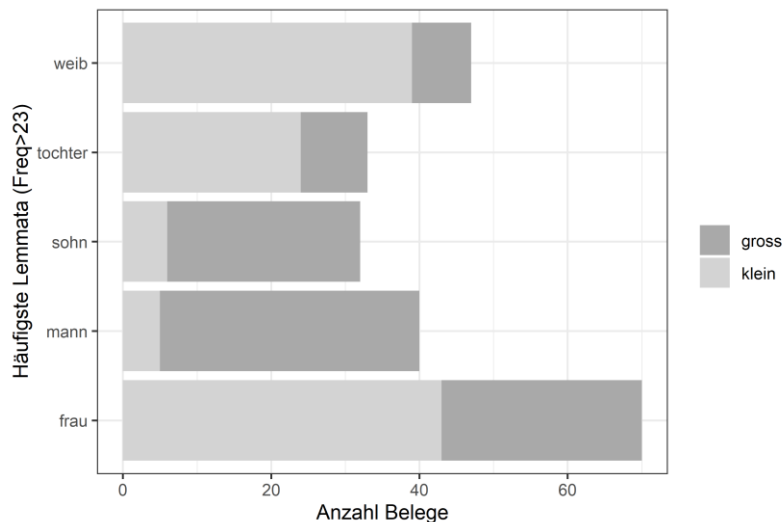


Abbildung 12: Groß- und Kleinschreibung der frequentesten Lemmata.

Abb. 13 zeigt die fünf häufigsten Lemmata im untersuchten Subkorpus: *weib*, *tochter*, *sohn*, *mann* und *frau* sind mit Abstand die frequentesten Items. Dieses Ergebnis hängt mit der speziellen Situation zusammen, die in den Texten wiedergegeben wird, und darf keinesfalls als repräsentativ für den Sprachgebrauch im Allgemeinen gesehen werden. Im Folgenden Kapitel werden die multifaktorielle Analyse vorgestellt und deren Ergebnisse präsentiert.

5 Ergebnisse

Die statistische Analyse der Daten wurde mit Hilfe von R Studio (R Studio Team 2018) und *lme4* (Bates, Maechler & Bolker 2012) durchgeführt. Um die Auswirkung der verschiedenen oben diskutierten Faktoren auf die abhängige Variable MAJUSKELSETZUNG zu errechnen, wurde ein generalisiertes gemischtes lineares Modell verwendet. Da sowohl feste als auch zufällige Effekte (fixed/ random effects) miteinzubeziehen sind, ist ein gemischtes Modell notwendig und da die abhängige Variable kategorial ist (GROßSCHREIBUNG: ja/nein), musste ein generalisiertes (logistisches) Modell verwendet werden. Für das Modell (fullmodel²¹) wurde die folgende Spezifikation verwendet:

²¹ Für alle im Folgenden vorzustellenden Modelle wird jeweils auf den entsprechenden Variablennamen im anhängten R-Skript verwiesen.

$$\begin{aligned} \text{MAJUSKELSETZUNG} \sim & \text{GENDER} * \text{MORALISCHE BEWERTUNG} + \text{LEMMAPRESTIGE} \\ & + \text{FREQUENZ} \\ & + (1 + \text{GENDER} | \text{ORT}) \end{aligned}$$

Als feste Effekte wurden GENDER, MORALISCHE BEWERTUNG und LEMMAPRESTIGE eingespeist, wobei aufgrund der Erkenntnisse aus vorherigen Untersuchungen bei GENDER und MORALISCHE BEWERTUNG von Interaktion ausgegangen wird (*).

Um zu vermeiden, dass durch Frequenzeffekte bedingte Varianz den hier untersuchten Faktoren zugeschrieben wird, wird zusätzlich FREQUENZ als Kontrollfaktor einbezogen. Die Datenpunkte sind, da mehrere Belege aus demselben Text und damit vom selben Schreiber stammen, nicht unabhängig voneinander (vgl. Kap. 2.4). Da dadurch gegen die *Independence Assumption* verstoßen wird (vgl. Winter 2020:233), wurden random Y-Achsenabschnitt (intercepts) für jeden ORT (als Annäherung an Schreiber)²². Dadurch wird für jedes Protokoll eine unterschiedliche baseline-Großschreibungsrate angenommen (vgl. Winter 2020:234). Außerdem wurden random Steigungen (slopes) für den Effekt von GENDER und MORALISCHE BEWERTUNG pro ORT einbezogen, um so die idiosynkratische Varianz in der Wirksamkeit dieses Faktors zu erfassen (vgl. Kap. 2.4).

R-Studio gibt für das so errechnete Modell eine Konvergenz-Warnung aus. Konvergenzprobleme kommen häufig vor, wenn komplexere Modelle auf zu kleine Datensets angewendet werden (Winter 2020:266). Oft ist das Einbeziehen von random slopes der Grund für Konvergenzprobleme (Winter 2020:266), wesshalb sie jedoch nicht einfach aus der Analyse herausgelassen werden sollten, da dies die Typ-1-Fehlerrate erhöht (Winter 2020:266). In diesem Fall ließen sich die Konvergenzprobleme durch eine Änderung des Optimierungsalgorithmus zu „bobyqa“ beheben.²³

Der Anteil der (Großschreibungs-) Varianz in den Daten, der durch das Modell erklärt wird, ist nicht sehr hoch, er liegt zwischen 36 –44 % (R^2 Werte: theoretical 0.44, delta: 0.37)²⁴. Die R^2 Werte nur für die festen Effekte sind deutlich niedriger (theoretical: 0,

²² Vgl. Fußnote 8.

²³ Ermittelt mithilfe der Funktion „allFit“ aus dem R-Paket „afex“. Die vorhergesagten Werte des Optimierers unterscheiden sich nicht maßgeblich von den für andere Optimierer ausgegebenen Werten oder denen des zuvor gerechneten Modells (vgl. im angehängten Skript „summary(allFit)“).

²⁴ Die Berechnung eines R^2 -Werts für generalisierte gemischte lineare Modelle im Gegensatz zu gemischten linearen Modellen ist nicht unproblematisch (Nakagawa et al. 2017, S. 1-2). Für

11, delta: 0,09), was darauf hindeutet, dass durch Einbeziehen der random Effekte die Erklärungskraft des Modells erheblich erhöht wird.

Im Folgenden wird zunächst die Auswirkung von LEMMAPRESTIGE untersucht. Da die visuelle Inspektion (s. Kap. 4) bereits nahelegt, dass kein Effekt vorhanden ist, muss der Faktor im Sinne des Parsimonieprinzips (vgl. Levshina 2015: 149) potenziell aus dem Modell ausgeschlossen werden. Kriterium hierfür sind p Werte, R^2 Werte und Akaikes's Information Criterion (AIC) (vgl. Levshina 2015:149).

5.1 LEMMAPRESTIGE

Es zeigt sich kein Effekt dieses Faktors. Die Odds ratios liegen bei 0.58, die Odds für Großschreibung sind also größer, wenn LEMMAPRESTIGE vorliegt, allerdings in zu vernachlässigendem Ausmaß ($p=0.17$). Abb. 13 zeigt die vom Modell vorhergesagten Werte für neutrale und prestigebelastete Lemmata samt Konfidenzintervallen.

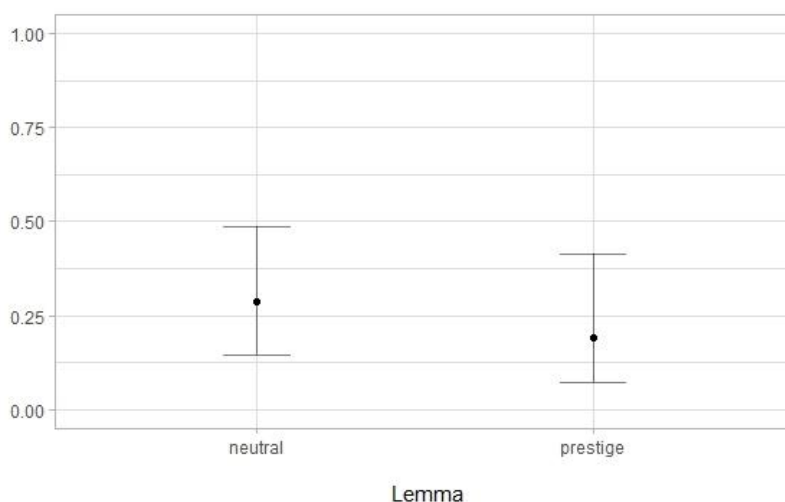


Abbildung 13: Vorhergesagte Werte des Modells anhand der Estimates (0 = Großschreibung wahrscheinlich, 1=Kleinschreibung wahrscheinlich).

Der vorausgesagte Wert für prestigeträchtige Lemmata ist minimal niedriger als für neutrale Lemmata, das Modell würde also für prestigeträchtige Lemmata geringfügig eher Großschreibung annehmen als für neutrale. Die starke Überlappung der Konfidenzintervalle weist auf Effektivlosigkeit hin, das Modell kann kaum differenzierende Vorhersagen über die beiden Gruppen treffen. Ein Vergleich des Modells mit einem in allen anderen Punkten identischen Modell ohne den Faktor LEMMAPRESTIGE (model2)

mittels eines Likelihood Ratio Tests bestätigt ebenfalls, das LEMMAPRESTIGE keinen statistisch verlässlichen Effekt hat ($\chi^2(1,11)=1.88$, $p>0.5$).

Die Nullhypothese (H_{30}) kann also für diese Daten nicht abgelehnt werden, und es kann keine positive Auswirkung des Lemmaprestiges auf die Großschreibung angenommen werden. Es ist zu beachten, dass für diesen Teil der Untersuchung mit nur 67 Belegen kein zufriedenstellend großes Datenset vorlag. Davon abgesehen passt dieses Ergebnis in das sich allgemein ergebende Bild, dass bei Männern keine weiteren Faktoren Einfluss auf die bereits relativ stabile Großschreibungstendenz nehmen konnten.

Das Modell ohne Einbeziehung von LEMMAPRESTIGE ist besser als das erste, maximale Modell; bei ähnlichen R^2 Werten (theoretical 0.43, delta: 0.36) ist der AIC Wert niedriger ($627.53 > 627.41$). Es wird daher gegenüber dem maximalen Modell vorgezogen.

Random effects:					
Groups	Name	Variance	Std.Dev.	Corr	
Ort	(Intercept)	1.5494	1.2448		
	Genderweiblich	0.8273	0.9096	0.00	
Number of obs: 551, groups: Ort, 18					
Fixed effects:					
	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.159621	0.435163	-2.665	0.0077	**
Genderweiblich	0.896991	0.452805	1.981	0.0476	*
Moralische.Bewertungneutral	-0.127020	0.381374	-0.333	0.7391	
Moralische.Bewertungpositiv	-0.175027	0.387744	-0.451	0.6517	
Freq	0.018252	0.005594	3.263	0.0011	**
Genderweiblich:Moralische.Bewertungneutral	0.470016	0.645587	0.728	0.4666	
Genderweiblich:Moralische.Bewertungpositiv	-0.712268	0.561663	-1.268	0.2047	

Tabelle 3: Ausschnitt aus der Zusammenfassung der vom Modell (modell2) ausgegebenen Werte.

Der vom besseren Model (modell2) geschätzte durchschnittliche Y-Achsenabschnitt (Intercept) liegt bei -1.16. Die vom Modell geschätzten Y-Achsenabschnitte für die einzelnen Protokolle (Ort) liegen im Bereich von 1.24 Stellen um den durchschnittlichen Y-Abschnitt. Die Steigungen (slopes) für den Effekt von Gender in den einzelnen Protokollen schwanken um 0.9 Stellen.

Beim Intercept wird von jeder der kategorialen Variablen das Referenzlevel angenommen. Durch die Voreinstellungen von R bedingt ist das jeweils das im Alphabet zuerst stehende Level: bei den unabhängigen Variablen (Predictors) sind die Referenzlevel GENDER[männlich], MORALISCHE.BEWERTUNG[negativ] und PRESTIGE_LEMMA[neutral]. Das Referenzlevel der abhängigen Variable ist GROßSCHREIBUNG

(upperCase[groß]). Die Estimates sind bei kategorialen abhängigen Variablen schwierig direkt interpretierbar, sie werden daher im Folgenden mithilfe weiterer Visualisierungen interpretiert. Anhand von Tab. 3 lässt sich jedoch bereits erkennen, dass GENDER[weiblich] (bei negativer moralischer Bewertung, also dem Referenzlevel, bleibend), GENDER[weiblich] bei neutraler Bewertung sowie steigende FREQUENZ entgegengesetzte Effekte zu den anderen festen Effekten aufweisen, da hier positive Estimates vorliegen. Der Standardfehler ist überall recht hoch, ausser bei FREQUENZ. Den Alphaswert von 0,05 unterschreiten FREQUENZ(**) und GENDER(*). Der Abgleich mit den Hypothesen 1 und 2 erfolgt nach der weiteren Interpretation.

Predictors	Odds Ratios	CI	p
(Intercept)	0.31	0.13 – 0.74	0.008
Gender [weiblich]	2.45	1.01 – 5.96	0.048
Moralische.Bewertung [neutral]	0.88	0.42 – 1.86	0.739
Moralische.Bewertung [positiv]	0.84	0.39 – 1.79	0.652
Freq	1.02	1.01 – 1.03	0.001
Gender [weiblich] * Moralische.Bewertung [neutral]	1.60	0.45 – 5.67	0.467
Gender [weiblich] * Moralische.Bewertung [positiv]	0.49	0.16 – 1.47	0.205

Tabelle 4: Odds Ratios mit Konfidenzintervallen und p-Werten (exportiert mit dem R-Paket sjPlot).

Tab. 4 zeigt den Einfluss der festen Effekte auf die Odds Ratios, dass eine Personenbezeichnung großgeschrieben wird. Das Alphalevel unterschreitende p-Werte sind hervorgehoben. Im Folgenden wird auf die unabhängigen Variablen GENDER und MORALISCHE BEWERTUNG genauer eingegangen.

5.2 GENDER und MORALISCHE BEWERTUNG

Die Odds für Großschreibung sind in signifikantem Maße geringer, wenn das Gender weiblich ist (die anderen festen Effekte auf Referenzlevel bleibend, also bei negativer Bewertung).²⁵

Der Unterschied in der moralischen Bewertung wirkt sich bei männlichem Gender kaum aus. Die Odds von Großschreibung bei negativer vs. neutraler oder positiver Bewertung sind in etwa gleich.²⁶ Die unteren beiden Zeilen zeigen die Odds Ratios im Zusammen-

²⁵ Odds Ratios > 1 zeigen an, dass die Odds unter Einwirkung des Faktors geringer sind als ohne den Faktor (vgl. Levshina 2015:208).

²⁶ Odds ratios nahe 1 bedeuten, dass die Odds unter Einwirkung des Faktors etwa gleich sind wie ohne Einwirkung des Faktors (vgl. Levshina 2015:208).

spiel von GENDER und MORALISCHE BEWERTUNG. Die neutrale moralische Bewertung wirkt sich auch in Kombination mit weiblichem Gender nicht maßgeblich aus, sie sind zwar niedriger als bei weiblichem Gender und negativer Bewertung, liegen aber immer noch deutlich über 1. Positive Bewertung zeigt in Zusammenspiel mit weiblichem Gender erhebliche Auswirkung: Die Odds von Großschreibung sind bei weiblichem Gender größer, wenn positive moralische Bewertung vorliegt, als wenn negative Bewertung vorliegt.²⁷ Abb. 13 zeigt die Wahrscheinlichkeiten für Großschreibung, die das Modell für hypothetische neue Daten voraussagen würde.

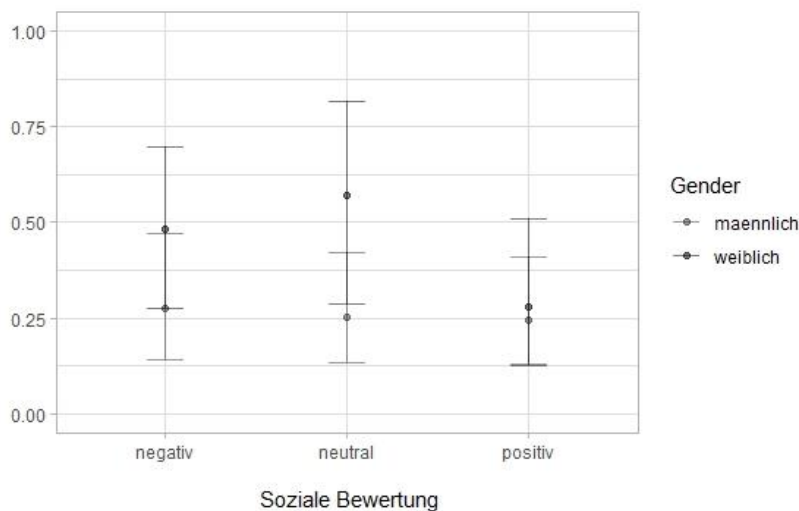


Abbildung 14: Vorhergesagte Werte des Modells anhand der Estimates (0 = Großschreibung wahrscheinlich, 1=Kleinschreibung wahrscheinlich).

Dargestellt sind für die drei Gruppen moralischer Bewertung jeweils die (vorhergesagte) durchschnittliche Großschreibungsrate (als Punkt) bei weiblichem (schwarz) und männlichem (grau) Gender, außerdem die jeweiligen Konfidenzintervalle. Das Modell sagt für Bezeichnungen von Personen mit männlichem Gender für alle drei Bewertungsgruppen stabil überwiegende Großschreibung voraus (0 steht für wahrscheinliche Großschreibung und 1 für wahrscheinliche Kleinschreibung). Die Konfidenzintervalle unterscheiden sich leicht in ihrer Weite (bei negativer Bewertung besteht am wenigsten Sicherheit über die Voraussage), sie liegen jedoch durchgehend zwischen 0.25 und 0.50, also im Bereich tendenzieller Großschreibung. Für die Bezeichnungen von Personen mit weiblichem Gender sagt das Modell bei negativer wie neutraler Bewertung etwa 50% Großschreibung voraus, die Konfidenzintervalle sind relativ breit und reichen etwa von 0.25,

²⁷ Odds ratios <1 zeigen also, dass die Odds unter Eintwirkung des Faktors höher sind also ohne den Faktor (vgl. Levshina 2015:208).

tendenziell großgeschrieben, bis über 0.75 (tendenziell kleingeschrieben).²⁸ Das Modell kann also relativ schlecht Voraussagen für diese Gruppen treffen. Allerdings differenziert es bei neutraler und negativer Bewertung klar zwischen beiden unterschiedlichen Gendergruppen: Die Konfidenzintervalle überschneiden sich zwar, der vorhergesagte Durchschnitt liegt aber jeweils außerhalb des Konfidenzintervalls der anderen Gruppe.

Anders sieht es bei den Werten für positive Bewertung aus: Hier überlappen die Konfidenzintervalle (beide im Bereich unter 50%) der vorhergesagten Großschreibungsrate für weibliches und männliches Gender, und der Durchschnittswert für die weiblichen Personenbezeichnungen rückt fast an den für die männlichen heran. Das Modell sagt also für Datenpunkte mit weiblichem Gender eine sehr ähnliche Großschreibungsrate voraus wie für die mit männlichem Gender, aber nur, wenn positive Bewertung vorliegt, was bei männlichem Gender hingegen keine Rolle spielt.

Die Nullhypothese (H_{10}) („Großschreibung von Frauenbezeichnungen tritt bei positiver Bewertung nicht häufiger auf als bei negativer oder neutraler Bewertung“) kann nicht abgelehnt werden ($p > 0.05$). Es kann also nicht ausgeschlossen werden, dass die ermittelte Großschreibungsrate bei Frauenbezeichnungen auch ohne Wirksamkeit der (positiven) moralischen Bewertung genauso aussähe. Die vorhergesagten Werte deuten jedoch auf eine Interaktion von Gender und Großschreibung hin. Eine Untersuchung mit größerer Datengrundlage (zum Beispiel das gesamte SiGs-Korpus anstelle nur des Kernkorpus) könnte aufschlussreich sein und außerdem helfen, die Erklärkraft des Modells zu erhöhen. Die zweite Nullhypothese (H_{20}) (Großschreibung von Männerbezeichnungen tritt bei positiver Bewertung häufiger auf als bei negativer oder neutraler Bewertung) kann abgelehnt werden ($p > 0.05$). Es kann also davon ausgegangen werden, dass soziale Bewertung sich nicht auf die Großschreibung von Männerbezeichnungen auswirkt.

²⁸ Dass der Durchschnitt bei den neutral bewerteten Frauenbezeichnungen sogar noch höher, also weiter in Richtung wahrscheinliche Kleinschreibung, liegt, verwundert – allerdings ist die Wahrscheinlichkeit groß, dass dieses Ergebnis durch einen Annotationsschritt bedingt ist. Es könnte sein, dass sich hier die Einstufung der Rolle „Zeugin“, die möglicherweise im Prozesskontext eindeutig negativ war, wenn allen Beteiligten klar war, dass es sich ihrerseits um eine der Hexerei angeklagte und andere Frauen denunzierende Person handelte. Dem Ohnehin nicht signifikanten Unterschied wird hier keine weitere Bedeutung zugesprochen, allerdings zeigt sich, dass für zukünftige Untersuchungen der Umgang mit der Rolle der „Zeugin“ bzw. der neutralen Bewertung überdacht werden muss.

5.3 FREQUENZ

Der als Kontrollfaktor einbezogene feste Effekt FREQUENZ zeigt in der logistischen Regression hochsignifikanten Einfluss auf die Großschreibung der Personenbezeichnungen (Estimate 0.018252, $p < 0.005$). Es ist also unwahrscheinlich, dass der Estimate oder die Odds Ratios ohne die Wirksamkeit dieses Faktors zustande gekommen wären. Die Wahrscheinlichkeit, dass eine Personenbezeichnung großgeschrieben wird, sinkt mit steigender Häufigkeit des Lemmas im Korpus. Das FREQUENZ kontrollierende Modell ist adäquater als ein Modell, das FREQUENZ nicht einbezieht: dieses würde ein verzerrtes Bild abgeben, da die durch FREQUENZ verursachte Varianz fälschlich anderen Faktoren zugeschrieben wird.

Dieses Ergebnis deckt sich mit denen von Barteld et al. (2016) (s. Kap. 2.1.4). Wie in Kap. 4 gezeigt, ist *frau* die mit Abstand am häufigsten auftretende Personenbezeichnung im Korpus, gefolgt von *weib*, *tochter*, *sohn* und *mann*. Es ist zu vermuten, dass die frequentesten Items prototypisierungseffekte aufweisen (vgl. Barteld et al. 2016:392). Da es sich zu großem Anteil um überwiegend kleingeschriebene Frauenbezeichnungen handelt, passt das Ergebnis, dass bei steigender Lemmafrequenz die Wahrscheinlichkeit sinkt, dass ein Item großgeschrieben wird. Insgesamt, für alle Nomina, nicht nur die hier untersuchten Personenbezeichnungen, stellen Barteld et al (2016:404) eine mit steigender Frequenz steigende Großschreibungstendenz sowie signifikante Interaktion zwischen FREQUENZ und BELEBTHEIT fest. Es wäre ein wichtiger nächster Schritt, generalisierbarere Frequenzwerte aus einem möglichst ausgewogenen Referenzkorpus heranzuziehen, um zu überprüfen, inwiefern abweichende Effekte festzustellen sind, sobald die korpuspezifische Häufung von Frauenbezeichnungen (in negativen Kontexten) nicht mehr doppelt in die Analyse hineinspielt.

6 Fazit

Diese Arbeit konnte das Zusammenspiel der verschiedenen Einflussfaktoren auf die Großschreibung der Personenbezeichnungen im Kernkorpus der Hexenverhörprotokolle ein Stück weit aufdecken. Während der von Barteld et al. bereits festgestellte Effekt von Gender und von Frequenz bestätigt wurden, konnte für den ebenfalls erwogenen Einfluss von prestigehaltigen Lemmata kein Effekt festgestellt werden. Auch die Interaktion von Gender und Moralischer Bewertung der benannten Person stellte sich

durch das Modell als nicht signifikant heraus, allerdings zeigen die vorhergesagten Werte deutlich den Unterschied, den positive moralische Bewertung bei weiblichem Gender macht. Es wäre lohnenswert, das Modell auf ein vergrößertes Datenset anzuwenden, was die möglicherweise die Erklärfähigkeit des Modells vergrößern und zu verlässlicheren Ergebnissen führen könnte. Nicht nur ein quantitativ größeres Datenset, auch ein qualitativ anderes wäre hilfreich für weitere Untersuchungen: Die Effekte von Gender und von Frequenz sowie deren Interaktion sollten unbedingt an einem weniger sexistisch geprägten Korpus überprüft werden, da nicht auszuschließen ist, dass es sich hier um Korpusartefakte handelt. Zumindest sollten Werte für die Lemmafrequenz aus einem repräsentativeren Korpus getestet werden.

Es hat sich angedeutet, dass ein wichtiger Schritt für eine weitere Untersuchung die Überarbeitung der Zuweisung der Bewertungskategorien wäre, die theoretisch begründete Problematik der Einordnung der Zeuginnen als „neutral“ zeigt sich möglicherweise auch in den Vorhersagen des Modells.

7 Literaturverzeichnis

7.1 Korpus

Macha, Jürgen/ Topalovi, Elvira/ Hille, Iris/ Nolting, Uta/ Wilke, Anja (Hrsg.) (2005): Deutsche Kanzleisprache in Hexenverhörprotokollen der Frühen Neuzeit. 2 Bde., Berlin/New York.

7.2 Forschungsliteratur

Barteld, Fabian / Hartmann, Stefan / Szczepaniak, Renata (2016): The usage and spread of sentence-internal capital letters in Early New High German – A multifactorial approach. In: *Folia Linguistica* 50/2, 385–412.

Bergmann, Rolf (1999): Zur Herausbildung der deutschen Substantivgroßschreibung. Ergebnisse des Bamberg-Rostocker Projekts. In: Walter Hoffmann u.a. (Hrsg.): *Das Frühneuhochdeutsche als sprachgeschichtliche Epoche*. Werner Besch zum 70. Geburtstag. Frankfurt am Main, 59-79.

Dücker, Lisa/ Hartmann, Stefan & Szczepaniak, Renata (i. Druck): Satzinterne Großschreibung in Hexenverhörprotokollen. Multifaktorielle Analyse des Majuskelgebrauchs. Pragmatische, semantische und syntaktische Einflussfaktoren. In: ebd. (Hg.): *Hexenverhörprotokolle als Korpus*. Berlin/Boston.

Dücker, Lisa (2020): Großschreibung von Personenbezeichnungen in Hexenverhörprotokollen – eine Analyse von Gender, Lemma und Rolle im Gerichtsprozess. In: *Kolloquium Forschende Frauen 2019. Beiträge Bamberger Nachwuchswissenschaftlerinnen*. Hg. von den Universitätsfrauenbeauftragten der Otto-Friedrich-Universität Bamberg (=Forschende Frauen in Bamberg 11). Bamberg.

Levshina, Natalia (2015): *How to do Linguistics with R. Data exploration and statistical analysis*. Amsterdam/Philadelphia.

Kaempfert, Manfred (1980): Motive der Substantivgroßschreibung. Beobachtungen an Drucken des 16. Jahrhunderts. In: *Zeitschrift für deutsche Philologie* 99, 72–98.

Moulin, Claudine (1990): *Der Majuskelgebrauch in Luthers deutschen Briefen (1517-1546)*. Heidelberg.

Pusch, Luise (1984): *Das Deutsche als Männersprache. Aufsätze und Glossen zur feministischen Linguistik*. Frankfurt am Main.

Schutzeichel, Marc/ Szczepaniak, Renata (2015): Die Durchsetzung der satzinternen Großschreibung in Norddeutschland am Beispiel der Hexenverhörprotokolle. In: Hundt, Markus/ Lasch, Alexander (Hrsg.): *Deutsch im Norden. Varietäten des*

norddeutschen Raumes (Jahrbuch für Germanistische Sprachgeschichte 6). Berlin/Boston, 151-167.

Silverstein, Michael (1976): Hierarchy of Features and Ergativity. In: Dixon, Robert M. W. (Hrsg.): Grammatical Categories in Australian Languages. New Jersey, 112-171.

Szczepaniak, Renata / Barteld, Fabian (2016): Hexenverhörprotokolle als sprachhistorisches Korpus. In: Kwekkeboom, Sarah / Waldenberger, Sandra (Hrsg.): PerspektivWechsel oder: Die Wiederentdeckung der Philologie. Bd. 1: Sprachdaten und Grundlagenforschung in Historischer Linguistik. Berlin, 43-70.

Szczepaniak, Renata (2011): Gemeinsame Entwicklungspfade in Spracherwerb und Sprachwandel? Kognitive Grundlagen der onto-und historiogenetischen Entwicklung der satzinternen Großschreibung. In: Köpcke, Klaus-Michael/Ziegler, Arne (Hrsg.): Grammatik verstehen lernen. Berlin/New York, 341–359.

Wasserstein, Ronald L./ Lazar Nicole A. (2016): The ASA Statement on p-Values: Context, Process, and Purpose, The American Statistician, 70:2, 129-133.

Winter, Bodo (2019): Statistics for linguists : an introduction using R. New York/London.

7.3 Online Quellen und Programme.

DWB = Deutsches Wörterbuch von Jacob und Wilhelm Grimm. 16 Bde. in 32 Teilbänden. Leipzig 1854-1961. Quellenverzeichnis Leipzig 1971 (<http://www.woerterbuchnetz.de/DWB>).

R Core Team (2018): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Bates, Douglas / Maechler, Martin / Bolker, Ben and Walker, Steve (2015): Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Wickham, H. et al. (2019): Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, <https://doi.org/10.21105/joss.01686>.

Wickham, H. (2016): ggplot2: Elegant Graphics for Data Analysis. New York: Springer Verlag.

Anhang

Anhang 1: Karte: Räumliche Verteilung der Texte des Kernkorpus



Anhang 2 : Tabelle 1: Korrekturen in der Gender Annotation

Die Tabelle enthält die für diese Untersuchung vorgenommen Änderungen an den bei Barteld et al (2016) verwendeten Gender-Annotationen. Neben einigen weiteren in Kapitel 3.3 erläuterten Tilgungen erklären die folgenden Fälle die Abweichungen zwischen den hier und den bei Barteld et al. (2016) untersuchten Gesamttokenzahlen.

Beleg	Ort, Jahr	Änderungsbegründung
<i>wehrwulffin</i>	Alme 1630	Kontext: "Herde Hanß <i>weib</i> sei <i>die</i> wehrwulffin"
<i>Beürin</i>	Bamberg 1828	Umlaut, -in
<i>Beürin</i>	Bamberg 1828	Umlaut, -in
<i>schaffhirtin</i>	Gaugrehweiler 1610	Kontext: "Will nicht gestehen das sie vnd <i>die</i> schaffhirtin die knie gegen einand[e..r]"
<i>Geuatters</i>	Hamm 1592	Kontext: "hett er Schweißdhall angefang[en] Geuatters warlich meine haußf[rau]" & Fußnote 155: Geuatters: 'Gevattersche'
<i>Swegersche</i>	Jever 1592	Kontext: "Ir Swegersche Indemet Loutet Teilen <i>weib</i> , " plus -sche
<i>Schaffmeisterin</i>	Rosenburg 1618	Kontext & FN. 96: Schaffmeisterin: Frau des Oberschäfers
<i>Schafmeisterin</i>	Rosenburg 1618	Kontext & FN. 96: Schaffmeisterin: Frau des Oberschäfers
<i>Schafmeisterin</i>	Rosenburg 1618	Kontext & FN. 96: Schaffmeisterin: Frau des Oberschäfers
<i>Schafmeisterin</i>	Rosenburg 1618	Kontext & FN. 96: Schaffmeisterin: Frau des Oberschäfers

Digitaler Anhang

In digitaler Form angehängt liegen vor

- Das für die Visualisierung sowie Berechnung der multifaktoriellen verwendete R-Skript (Ha_Statistik_Schnee.R)
- Das vollständige annotierte Datenset (LogRegData0211.csv)