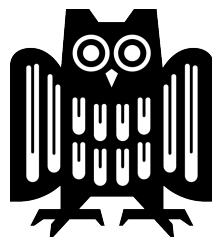


Universität des Saarlandes
Philosophische Fakultät
Germanistik

Sophia Voigtmann

**Wie Informationsdichte Extraposition beeinflusst.
Eine Korpusuntersuchung an wissenschaftlichen
Texten des frühen Neuhochdeutschen**

Dissertation
zur Erlangung des Grades
eines Doktors der Philosophie



Saarbrücken

Dekanin:	Prof. Dr. Stefanie Haberzettel
Erstgutachter:	Prof. Dr. Augustin Speyer
Zweitgutachter:	Prof. Dr. Stefanie Dipper
Drittgutachter:	Prof. Dr. Ingo Reich
Tag der letzten Promotionsleistung:	26. September 2023

Inhaltsverzeichnis

1. Einleitung	1
1.1. Die Hypothesen	5
1.1.1. Hypothese H1: Extrapolation von hochinformationellen Phrasen und Relativsätzen	6
1.1.2. Hypothese H2: Einfluss der Mündlichkeitsnähe auf Ex- trapolation	10
1.1.3. Hypothese H3: Diachrone Perspektive von Extrapolation	12
1.2. Aufbau der Arbeit	14
2. Theoretischer Hintergrund	17
2.1. Informationsdichte	17
2.1.1. Informationstheorie nach Shannon (1948)	17
2.1.2. Information und Syntactic Prediction Locality Theory .	28
2.1.3. Entropy Rate Constancy	35
2.1.4. Smooth Signal Hypothesis und Uniform Information Den- sity Hypothesis	38
2.1.5. Exkurs: Lossy-Context Surprisal und Language-dependent Structural Forgetting	43
2.2. Topologisches Feldermodell	46
2.2.1. Satzklammern und Satzrahmen	50
2.2.2. Vorfeld	56
2.2.3. Mittelfeld	58
2.2.4. Nachfeld	62
2.3. Die Begriffe Auslagerung und Extrapolation	66
2.4. Der Relativsatz	71
2.4.1. Relativsatzarten	71

2.4.2.	Art des Relativums	73
2.4.3.	Relativsatztyp	75
2.5.	Die Nominalphrase	78
2.5.1.	Form von Nominalphrasen	79
2.5.2.	Erweiterbarkeit der Nominalphrasen	80
2.5.3.	Entwicklung der Nominalphrase	84
2.6.	Die Präpositionalphrase	89
2.6.1.	Form der Präpositionalphrase	90
2.6.2.	Funktion der Präpositionalphrase	92
2.6.3.	Bedeutung der Präposition	94
2.6.4.	Entwicklung der Präpositionalphrase	96
2.7.	Mündlichkeit und Schriftlichkeit	98
3.	Extrapolation von Phrasen und Relativsätzen	109
3.1.	Die Extrapolation von PPs und NPs	109
3.1.1.	Grammatiken	109
3.1.2.	Forschung	114
3.2.	Die Extrapolation von Relativsätzen	120
3.2.1.	Grammatiken	120
3.2.2.	Forschung	122
4.	Korpus	135
4.1.	Auswahl der Texte	135
4.2.	Zusammensetzung des Korpus	143
4.3.	Annotationspraxis	149
4.3.1.	Satzklammern	149
4.3.2.	Relativsätze	151
4.3.3.	Präpositionalphrasen	157
4.3.4.	Nominalphrasen	158
4.3.5.	Vergleichspaare für die Phrasen: Minimalpaare	160
4.4.	Bereinigung der Daten	165
5.	Methodik	169
5.1.	Sprachmodelle	169

5.2.	Summiertes und durchschnittliches Surprisal	176
5.3.	Orality Score	179
5.4.	Statistische Analyseverfahren	184
6.	Ergebnisse zum Einfluss auf Extraposition	193
6.1.	Nominalphrasen	194
6.1.1.	Nominalphrasen im gesamten Korpus	194
6.1.2.	Nominalphrasen in den medizinischen Texten	205
6.1.3.	Nominalphrasen in den theologischen Texten	207
6.1.4.	Extraposition von attributiven Nominalphrasen	209
6.2.	Präpositionalphrasen	211
6.2.1.	Präpositionalphrasen im gesamten Korpus	211
6.2.2.	Logistische Regression der Präpositionalphrasen im Gesamtkorpus	222
6.2.3.	Präpositionalphrasen in den medizinischen Texten	224
6.2.4.	Präpositionalphrasen in den theologischen Texten	225
6.2.5.	Extraposition von attributiven Präpositionalphrasen	228
6.3.	Relativsätze	230
6.3.1.	Relativsätze im gesamten Korpus	231
6.3.2.	Logistische Regression der Relativsätze im Gesamtkorpus	237
6.3.3.	Relativsätze in den medizinischen Texten	240
6.3.4.	Relativsätze in den theologischen Texten	243
6.3.5.	Extraposition von Relativsätze unter Beachtung der Distanz	246
6.4.	Nominalphrasen, Präpositionalphrasen und Relativsätze	248
6.5.	Diskussion Hypothese 1: Surprisal und Auslagerung	250
6.5.1.	Nominalphrasen	255
6.5.2.	Präpositionalphrasen	264
6.5.3.	Relativsätze	272
7.	Ergebnisse zum Einfluss von Mündlichkeitsnähe	281
7.1.	Mündlichkeitsnähe	282
7.1.1.	Orality Scores im Korpus	282

7.1.2.	Mündlichkeitsnähe und Auslagerungsfrequenz	286
7.1.3.	Surprisal und Orality	289
7.2.	Diskussion Hypothese 2: Mündlichkeitsnähe und Extraposition	293
7.2.1.	Hypothese 2a: Häufigkeit von Extraposition	293
7.2.2.	Hypothese 2b: Mündlichkeitsnähe und Surprisal	296
8.	Ergebnisse zur diachronen Entwicklung von Extraposition	299
8.1.	Nominalphrasen in den verschiedenen Zeitabschnitten	299
8.2.	Präpositionalphrasen in den verschiedenen Zeitabschnitten	307
8.3.	Relativsätze in den verschiedenen Zeitabschnitten	316
8.4.	Diskussion Hypothese 3: Diachrone Perspektive von Extraposition	325
9.	Zusammenfassung	339
	Literaturverzeichnis	356
	Anhang	387
	A. Informationen zu den Autoren	389
	B. Liste der für die Analysen verwendeten Sätze	395
	C. Texte im Trainingskorpus	423
	D. Zusätzliche Grafiken	427
	E. Zusätzliche Regressionsanalysen	429
E.1.	Regressionsanalysen bei gleicher Länge	429
E.1.1.	Nominalphrasen	429
E.1.2.	Präpositionalphrasen	430
E.2.	Relativsätze ohne Variable Typ	430

Tabellenverzeichnis

2.1. Besetzung des topologischen Feldermodells	47
2.2. Beispiel eines V2-Satzes	48
2.3. Beispiel eines komplexen V2-Satzes und eines W-Fragesatzes	48
2.4. Beispiel V1-Sätze	49
2.5. Beispiel eines VL-Satzes	49
2.6. Beispiel eines Relativsatzes im topologischen Feldermodell	58
2.7. Attributive Erweiterung von NPs	82
4.1. Texte und Tokens pro Zeitabschnitt	143
4.2. Im Korpus verwendete Texte 1	144
4.2. Im Korpus verwendete Texte 2	145
4.2. Im Korpus verwendete Texte 3	146
4.3. Im Korpus verwendete Texte 4	147
4.3. Im Korpus verwendete Texte 5	148
4.4. Anzahl der RelS	156
4.5. Anzahl der extrapolierten PPs	158
4.6. Anzahl der extrapolierten NPs	160
5.1. Verwendetes Language Model	176
6.1. Anzahl der NPs	195
6.2. Satzgliedfunktion der NPs	197
6.3. Länge der NPs	198
6.4. Skipgram-Surprisal der NPs	200
6.5. Regression der NPs, Gesamtkorpus	204
6.6. Deskriptive Werte der NPs, med. Teilkorpus	206
6.7. Regression der NPs, med. Teilkorpus	206

6.8. Deskriptive Werte der NPs, theol. Teilkorpus	207
6.9. Regression der NPs, theol. Teilkorpus	208
6.10. Deskriptive Werte der attributiven NPs	210
6.11. Regression der attributiven NPs	210
6.12. Anzahl der PPs	212
6.13. Länge der PPs	216
6.14. Skipgram-Surprisal der PPs	220
6.15. Regression der PPs, Gesamtkorpus	223
6.16. Deskriptive Werte der PPs, med. Teilkorpus	224
6.17. Regression der PPs, med. Teilkorpus	225
6.18. Deskriptive Werte der PPs, theol. Teilkorpus	226
6.19. Regression der PPs, theol. Teilkorpus	227
6.20. Regression der attributiven PPs	230
6.21. Anzahl der RelS	232
6.22. Länge der RelS	235
6.23. Skipgram-Surprisal der RelS	236
6.24. Regression der RelS, Gesamtkorpus	238
6.25. Deskriptive Werte der RelS, med. Teilkorpus	241
6.26. Regression der RelS, med. Teilkorpus	242
6.27. Deskriptive Werte der RelS, theol. Teilkorpus	243
6.28. Regression der RelS, theol. Teilkorpus	245
6.29. Distanz der RelS zum Antezedens	246
6.30. Regression der RelS mit Distanz, Gesamtkorpus	247
6.31. Regression aller Auslagerungen	249
7.1. Orality Scores pro Text	283
7.2. Orality Scores pro Zeitabschnitt	285
7.3. Anzahl der Auslagerungen in mündlichkeitsnahen Texten . . .	290
7.4. Regression für mündlichkeitsnahe Texte	291
7.5. Anzahl der Auslagerungen in schriftlichkeitsnahen Texten . . .	291
7.6. Regression für schriftlichkeitsnahe Texte	292
8.1. Länge der NPs pro Zeitabschnitt	301

8.2. Skipgram-Surprisalwerte der NPs pro Zeitabschnitt	302
8.3. Regression der NPs von 1650–1700	303
8.4. Regression der NPs von 1700–1750	304
8.5. Regression der NPs von 1750–1800	304
8.6. Regression der NPs von 1800–1850	305
8.7. Regression der NPs von 1850–1900	306
8.8. Länge der PPs pro Zeitabschnitt	309
8.9. Skipgram-Surprisalwerte der PPs pro Zeitabschnitt	310
8.10. Regression der PPs von 1650–1700	311
8.11. Regression der PPs von 1700–1750	312
8.12. Regression der PPs von 1750–1800	313
8.13. Regression der PPs von 1800–1850	314
8.14. Regression der PPs von 1850–1900	315
8.15. Relativsatztypen pro Zeitabschnitt	316
8.16. Länge der RelS pro Zeitabschnitt	318
8.17. Skipgram-Surprisalwerte der RelS pro Zeitabschnitt	319
8.18. Regression der RelS von 1650–1700	320
8.19. Regression der RelS von 1700–1750	321
8.20. Regression der RelS von 1750–1800	322
8.21. Regression der RelS von 1800–1850	323
8.22. Regression der RelS von 1850–1900	324
8.23. Zusammenfassung der Regressionsanalysen pro Zeitabschnitt .	326
C.1. Trainingskorpus	425
E.1. Regression gleich langer NPs	429
E.2. Regression gleich langer PPs	430
E.3. Regression der RelS exklusive Typ	431

Abbildungsverzeichnis

5.1. Beispiel für LM-Output	175
5.2. Beispiel für visuelles Transformationskriterium	189
5.3. Beispiel für visuelles Transformationskriterium, logarithmiert	190
6.1. Anzahl der NPs	196
6.2. Verhältnis zw. Auslagerung und Länge der NPs	199
6.3. Verhältnis zw. Auslagerung und Länge der NPs, log	200
6.4. Skipgram-Surprisal der NPs nach Zeitabschnitt	201
6.5. Verhältnis zw. Auslagerung und Skipgramwerten der NPs	202
6.6. Verhältnis zw. Auslagerung und Skipgramwerten der NPs, log	203
6.7. Anzahl der PPs	213
6.8. Anzahl PPs nach Genre	214
6.9. Valenzstatus der PPs	215
6.10. Länge der PPs pro Zeitabschnitt	217
6.11. Verhältnis zw. Auslagerung und Länge der PPs	218
6.12. Verhältnis zw. Auslagerung und Länge der PPs, log	219
6.13. Skipgram-Surprisal der PPs pro Zeitabschnitt	221
6.14. Verhältnis zw. Auslagerung und Skipgramwerten der PPs	221
6.15. Skipgramwerte der attributiven PPs	229
6.16. Anzahl der RelS pro Zeitabschnitt	233
6.17. Länge der RelS pro Zeitstufe	235
6.18. Skipgramwerte der RelS pro Zeitabschnitt	237
7.1. Orality Scores pro Zeitabschnitt	284
7.2. Auslagerungsanzahl nach Orality Score	286
7.3. Auslagerungsanzahl nach Orality Score, Regressionslinien	288

D.1. Anteil von Extraposition nach RelS-Typ, med. Teilkorpus	427
D.2. Anteil von Extraposition nach RelS-Typ, theo. Teilkorpus	427

1. Einleitung

Das Deutsche zeichnet sich als Sprache mit gewissen Wortstellungsfreiheiten aus. Zu diesen Freiheiten gehört auch die Möglichkeit, Konstituenten ins sogenannte *Nachfeld*, die Position hinter der rechten Satzklammer („RSK“, vgl. Wöllstein 2010; 2014), zu stellen (1). Diese Positionierung von Material im Nachfeld wird auch Extraposition beziehungsweise Auslagerung genannt.

- (1) a. Ich habe das Buch, **das du mir empfohlen hast**, [begonnen]_{RSK}.
→ Ich habe das Buch [begonnen]_{RSK}, **das du mir empfohlen hast**.
- b. Er hat sich schrecklich **über die Klausur** [aufgeregt]_{RSK}.
→ Er hat sich schrecklich [aufgeregt]_{RSK} **über die Klausur**.

Für einige Elemente ist die Positionierung hinter der rechten Satzklammer obligatorisch, andere können im Mittelfeld oder im Nachfeld stehen. An genau dieser Stelle setzt diese Arbeit an: Warum werden manche Nominalphrasen und Präpositionalphrasen sowie Relativsätze ins Nachfeld gestellt, andere aber nicht? Welche Faktoren beeinflussen die Wahl der Position? Und sind in früheren Zeitstufen des Deutschen die gleichen Einflussfaktoren für die Positionierung zu beachten wie im heutigen Standarddeutschen?

Setzt man sich mit diesen Fragen auseinander, findet man große Unterschiede im Ausmaß der Forschung über das Thema. Während viele sowohl synchrone als auch diachrone Studien zur Nachfeldstellung von attributiven Relativsätzen (1a) existieren (vgl. u. a. Lötscher 1972; Shannon 1992; Uszkoreit, Brants, Duchier, Krenn, Konieczny, Oepen und Skut 1998; Sahel 2015; Poschmann und Wagner 2016; Speyer und Lemke 2017), ist das weniger der Fall bei phrasaler Auslagerung (1b). Zwar finden sich hier auch Studien über das Phänomen (vgl. u. a. Schildt 1976; Ebert 1980; Borter 1982; Takada 1998; Demske

2001; Paul 2007; Hinterhölzl 2010; Light 2011; Sapp 2014; Petkova-Kessanlis 2015; Weber 2019), aber insgesamt ist speziell die empirische Datenlage der Extraposition von Phrasen ausbaufähig.

Die bisherige Forschung beschreibt des Weiteren vor allem Gemeinsamkeiten von extraponierten Phrasen beziehungsweise Relativsätzen.¹ Diese Beobachtungen umfassen oft prosodische Voraussetzungen, die für Nachfeldstellung erfüllt sein müssen, und meistens die große Länge des ausgelagerten Materials. Letzteres wird beispielsweise auch von Uszkoreit et al. (1998) und Weber (2019) in ihren Untersuchungen genannt, die daraus den Schluss ziehen, dass die Nachfeldstellung von Material Verarbeitungsvorteile für die Rezipierenden eines Satzes hervorruft. Obwohl diese Erklärung in vielen Untersuchungen und Grammatiken zum Thema nicht explizit genannt wird, scheint diese „Informationsentflechtung“ (Zifonun, Hoffmann und Strecker 1997: 1669f.) doch das verbindende Glied zwischen den Beschreibungen von Extraposition zu bilden. Studien wie die von Hawkins (1992) und Gibson (1998) bemühen sich auch darum, diese Verarbeitungserleichterung quantifizierbar zu machen.

Letzteres bildet auch den Fokus meiner vorliegenden Dissertation. Ihr übergeordnetes Ziel ist zu untersuchen, ob die Informativität des auszulagernden Materials, gemessen in Form von *Informationsdichte* (vgl. Shannon 1948), einen Einfluss darauf hat, ob es tatsächlich extraponiert wird.

Informationsdichte nach Shannon (1948) besagt, dass der Informationsgehalt eines Wortes durch dessen Wahrscheinlichkeit in einem Kontext mathematisch bestimmt werden kann. Jedem Wort beziehungsweise jeder linguistischen Einheit wird so ein bestimmter Wert zugeordnet, das *Surprisal*. Je höher das *Surprisal* ist, desto informativer ist eine sprachliche Einheit. Ihr höherer Informationsgehalt bedingt aber auch, dass die Einheit schwerer zu verarbeiten ist (vgl. z. B. Hale 2001). Für die Verarbeitung mit hohen *Surprisal*werten werden also mehr kognitive Ressourcen verbraucht (vgl. z. B. Hale 2001; Levy 2008). Es ist davon auszugehen, dass sich diese Ressourcen nicht

¹ Im Folgenden werden Nominalphrasen und Präpositionalphrasen zum Teil aus Gründen der verbesserten Lesbarkeit und, um sie sprachlich einfach von den Relativsätzen abzugrenzen, zusammenfassend als *Phrasen* bezeichnet, obwohl alle drei Konstituenten im Sinne der Grammatik natürlich Phrasen bilden (Repp und Struckmeier 2020).

augenblicklich erneuern und wieder zur Verfügung stehen. Wäre dem so, könnte man einen Satz oder einen ganzen Text bezüglich der Wortfolge vollkommen nach Belieben gestalten. Die natürliche Sprache zeigt aber, dass das nicht zutrifft, weil die Anordnung von Worten in erster Linie der Grammatik der jeweiligen Sprache folgt. Allerdings bieten viele Sprachen, wenn nicht die meisten, Freiheiten in der Wortstellung. Die Veränderbarkeit der Wortstellung resultiert natürlicherweise in einer Veränderung des Kontextes, die wiederum Einfluss auf den Informationsgehalt eines Wortes hat (vgl. Shannon 1948). Bei einer nachteiligen Kombination aus Wörtern, also einer ungünstigen Verwendung der Stellungsfreiheit, kann es zu Verarbeitungsproblemen kommen, da zu viel *kognitive Kapazität* verbraucht wird. Bei einer gekonnten Verwendung der Stellungsfreiheit hingegen werden die kognitiven Kapazitäten der Rezipierenden nur so weit beansprucht, dass die Verarbeitung des Satzes noch gelingt. Extraposition bildet, wie bereits gesagt, eine solche Stellungsfreiheit, deren Verwendung zur optimalen Verarbeitung des Satzes beitragen kann.

Diese Arbeit fokussiert sich dabei auf die Extraposition von attributiven Relativsätzen, Nominalphrasen und Präpositionalphrasen. Die Wahl der beiden Phrasenklassen ergibt sich daraus, dass sie häufig sowohl im Mittelfeld als auch im Nachfeld stehen können. Präpositionalphrasen zählen nach Zifonun et al. (1997: 1657) zu den heute am häufigsten im Nachfeld stehenden Konstituenten. Auch Nominalphrasen können dort stehen und wurden früher häufiger hinter die rechte Satzklammer gestellt als heute. Relativsätze wurden stellvertretend für Nebensätze gewählt, die nicht präferenziell im Nachfeld auftreten, sondern auch im Mittelfeld vorkommen können.²

Bereits die Untersuchung von Extraposition im Deutschen mithilfe der Informationstheorie kann als neuartig bezeichnet werden. Die einzige andere Studie, die ebenfalls diese Zielsetzung verfolgt und den gleichen Forschungs-

² Diese Eigenschaft besitzen auch korrelierte Attributsätze (2a) (vgl. z. B. Paranhos Zitterbart 2002) und Komplementsätze (2b), allerdings sind diese in der Auswahl ihres Antezedens begrenzter als Relativsätze. Deshalb wurden sie in dieser Arbeit nicht betrachtet, bieten sich für weiterführende Untersuchungen aber prinzipiell ebenso an.

- (2) a. Er hat sich darüber [gewundert]_{RSK}, **dass der Zug pünktlich war.**
 b. Sie hat die Antwort [gegeben]_{RSK}, **dass ihr das recht sei.**

gegenstand gewählt hat, stammt von Speyer und Lemke (2017) und ist insofern mit dieser Untersuchung assoziiert, da beide Arbeiten im Rahmen des Sonderforschungsbereichs SFB 1102 vorgenommen wurde. Allerdings unterscheidet sich die vorliegende Arbeit von der von Speyer und Lemke (2017) in Bezug auf das verwendete Korpus. Statt wie Speyer und Lemke (2017) Leichenpredigten zu untersuchen, konzentriert sich meine Arbeit auf *wissenschaftliche Texte*, namentlich medizinische und theologische Texte, als Stellvertreter für Natur- und Geisteswissenschaft. Während wissenschaftliche Texte in anderen Sprachen wie dem Englischen auch in historischen Korpora gut untersucht sind (für einen Überblick siehe z. B. Biber 2006: 6ff.), leistet meine Arbeit speziell mit der empirischen Betrachtung des Genres im *frühen Neuhochdeutschen* Pionierarbeit.

Dieser Zeitabschnitt, der für die vorliegende Arbeit von 1650 bis 1900 definiert wird, erfreut sich in der bisherigen Forschung keiner großen Beliebtheit, vor allem nicht hinsichtlich der Untersuchung von Nachfeldbesetzung. Zwar setzen sich Konopka (1996), Takada (1998) und Elspaß (2005) mit Passagen dieses Zeitabschnitts auseinander, doch sind mir keine Studien bekannt, die sich über den gesamten oben genannten Zeitraum mit diachronen Veränderungen befassen. Somit will diese Arbeit auch einen Beitrag dazu leisten, Einblicke in einen untererforschten Abschnitt der deutschen Sprachgeschichte zu liefern.

Sowohl aus der Wahl wissenschaftlicher Texte als auch aus der Wahl des Zeitraums selbst ergeben sich allerdings Herausforderungen. Zum einen muss beachtet werden, dass wissenschaftliches Schreiben in deutscher Sprache erst seit dem siebzehnten Jahrhundert belegt ist (vgl. z. B. Kretzenbacher 1995; Klein 2011; Polenz, Moulin und Harion 2013). Zum anderen lässt sich vermuten, dass der Stil der beiden Wissenschaftsbereiche sehr unterschiedlich ist und vor allem die theologischen Texte vielleicht näher am Pol *konzeptioneller Mündlichkeit* (vgl. Koch und Oesterreicher 2007) zu verorten sind als die medizinischen. Ersteres lässt Veränderungen über den Zeitraum von 250 Jahren erwarten. Letzteres erfordert eine Betrachtung möglicher Unterschiede zwischen konzeptionell schriftlichen und mündlichen Texten, bei denen

sich die Anwendung einer neuartigen Methode anbietet. Diese Methode bildet der *Orality Score* (vgl. Ortmann und Dipper 2024), der bisher nur von Ortmann (2023) angewendet wurde. Auf diese beiden Aspekte und auch auf die Hypothesen, mit denen die Verarbeitung in einen Zusammenhang mit Informationsdichte gestellt wird, geht das folgende Kapitel ein.

1.1. Die Hypothesen

Extraposition wird in dieser Arbeit ausschließlich auf die Oberflächenstruktur des Satzes bezogen. Für das inkrementelle Lesen ist es irrelevant, ob beispielsweise in der Tiefenstruktur der Syntax eine Bewegung nach rechts erfolgt oder andere Elemente beispielsweise vor den Relativsatz gezogen werden (vgl. z. B. Büring und Hartmann 1995; Inaba 2007; Repp und Struckmeier 2020). Für die Rezipierenden sieht diese immer so aus, als sei eine Verschiebung des Materials nach rechts erfolgt. Da die untersuchten Elemente vor allem im Vor- oder Mittelfeld gefunden werden, wird Bewegung in dieser Arbeit als merkmalsgetriebene Operation verstanden. Sie muss folglich durch etwas, beispielsweise ihren Informationsgehalt, ausgelöst sein.

Die zentrale Fragestellung dieser Arbeit befasst sich deshalb damit, ob die Extraposition eines Elements die Verarbeitung des gesamten Satzes vereinfachen kann. In diese Überlegung spielt vor allem die *Uniform Information Density Hypothesis (UID)* (Levy und Jaeger 2007) hinein, die besagt, dass ein gleichförmiges Informationsprofil besonders geeignet ist, um eine optimale Verarbeitung zu gewährleisten. Bisherige Arbeiten über UID haben aber vor allem punktuelle Veränderungen im Informationsprofil gemessen und daraus Schlüsse über die Satzverarbeitung gezogen (vgl. z. B. Jaeger 2005; Frank und Jaeger 2008; Jaeger 2010; 2011). Spielt man das Vorgehen aus den oben genannten Arbeiten aber an einem einzelnen Satz (3) und mit den Mitteln durch, die mir in einer historisch ausgerichteten Arbeit mit dem ausgewählten Korpus (siehe Kapitel 4) zur Verfügung stehen, zeigt sich, warum das für die vorliegende Fragestellung schwierig ist.

- (3) a. Ich habe mich gestern sehr **über die Klausur** aufgeregt.
 b. Ich habe mich gestern sehr *aufgeregt über die Klausur*.

Der Phrase in (3), die extrapониert werden soll, ist fettgedruckt. Verschiebt man sie aus ihrer eingebetteten Position (3a) ins Nachfeld (3b), treten an den kursiven Stellen Veränderungen im Surprisalprofil auf, weil sich für diese kursiven Wörter der Kontext verändert. In (3a) bildet beispielsweise *Klausur* den Kontext zu *aufgeregt*, während in (3b) der Kontext zu *aufgeregt* von *sehr* gebildet wird.³ Es ist zu erwarten, dass diese Veränderungen nicht sehr groß sind und dass sie noch kleiner werden, wenn beispielsweise lemmatisierte Daten verwendet werden (vgl. Kapitel 5.1). Die Einschränkungen machen es schwierig, die UID-Hypothese auf die Art zu testen, wie es in den oben genannten Studien getan wurde, um eine Verbesserung des Informationsprofils durch Extraposition zu postulieren. Deshalb schlage ich, auch unter der Berücksichtigung der gewählten Textgenres und der diachronen Perspektive, drei Teilhypothesen vor, die in den folgenden Teilabschnitten genannt und begründet werden.

1.1.1. Hypothese H1: Extraposition von hochinformativen Phrasen und Relativsätzen

Eine erste Annäherung erfolgt mit H1, die da lautet:

(H1) Relativsätze und Phrasen mit hohen Surprisalwerten werden ausgelagert.

Diese Hypothese H1 unterteilt einen Satz in Anlehnung an Speyer (2015a: 499) in zwei Teile. Der erste Verarbeitungsteil des Satzes reicht vom Vorfeld bis zur rechten Satzklammer,⁴ der zweite umfasst das Nachfeld. Dieses bildet

³ Der Kontext ist natürlich stark abhängig davon, welches Sprachmodell für die Berechnung verwendet wird. Hier gehe ich aus Gründen der einfachen Beschreibbarkeit davon aus, dass der Kontext nur von einem einzigen Wort gebildet wird. Würde ein größerer Kontext einbezogen, würden sich auch das Surprisal innerhalb der extrapониerten Phrase in (3b) verändern.

⁴ Allerdings kann man an dieser Stelle dafür argumentieren, dass das Vorfeld als Position, in der sehr komplexe Satzglieder, einschließlich Nebensätzen, untergebracht werden können, aus dieser Betrachtung auszuschließen sein sollte. Forschung, die zeigt, inwiefern sich ein komplexes Vorfeld auf die weitere Verarbeitung des Satzes auswirkt, ist aktuell ein Desiderat,

laut Zifonun et al. (1997), Hoberg (1981), Vinckel (2006) und Speyer (2015a) den Ort, an dem schwer zu verarbeitende Konstituenten stehen sollen. Das impliziert, dass hinter der rechten Satzklammer mehr kognitive Kapazität zur Verfügung stehen muss oder andere Verarbeitungserleichterungen dort zum Tragen kommen.

Ich gehe davon aus, dass die Verarbeitung des ersten Teils des Satzes kognitiv anspruchsvoller ist als die Verarbeitung des Nachfeldes. Denn vom Vorfeld bis zur rechten Satzklammer muss sowohl der syntaktische Aufbau des Satzes, das heißt, die Satzgliedfunktionen, als auch die Bedeutung der einzelnen Komponenten und mögliche pragmatische Mittel wie Präsuppositionen oder Implikaturen verarbeitet werden. Obwohl Sätze in der Regel in einen größeren Kontext eingebettet sind, enthalten sie gewöhnlich dennoch neue Informationen und schaffen ihrerseits einen neuen Kontext, der seinerseits Erwartungen projiziert.

Bis die rechte Satzklammer erkannt wird, belastet dies das Arbeitsgedächtnis stark. Vorfeld, Satzklammern und Mittelfeld setzen für sich gesehen also ab dem Zeitpunkt ihrer erfolgreichen Verarbeitung einen neuen Kontext und einen Beitrag zum Weltwissen, das zur Verarbeitung des jeweiligen Satzes benötigt wird. Die syntaktischen Besetzungsmöglichkeiten des Nachfeldes sind also bereits eingeschränkt und auch lexikalisch wurde ein Kontext geschaffen, der die Verarbeitung der Bedeutung der extraponierten Wörter erleichtern kann. Diese verringerte *Entropie* (Kapitel 2.1) auf den extraponierten Elementen lässt mehr kognitive Kapazitäten nach der rechten Satzklammer frei werden. Wie der erhöhte Verarbeitungsaufwand vom Vorfeld bis zur rechten Satzklammer zustande kommt, wird später detailliert dargelegt. Zuvor soll aber die Frage geklärt werden, warum diese Voraussetzungen bei den hier untersuchten Elementen eine Rolle spielt.

Weder attributive Relativsätze noch Nominal- noch Präpositionalphrasen können außer in elliptischen Konstruktionen allein stehen und eigenständige Satzeinheiten bilden.⁵ Das bedeutet, dass sie immer auch eine Funktion in-

daher können hierzu keine weiteren Aussagen getroffen werden, ohne eigene Experimente vorzunehmen. Es muss zukünftigen Arbeiten überlassen werden, diese durchzuführen.

⁵ Für appositive Relativsätze wurde diese Lesart zwar diskutiert, doch auch sie benötigen ein Kopfnomen, wie Kapitel 2.4.1 zeigt.

nerhalb ihres Matrixsatzes erfüllen, die mit hoher Wahrscheinlichkeit bis zum Erreichen des Nachfelds bekannt, aber hinsichtlich der lexikalischen Besetzung frei ist. Die kognitive Belastung besteht also nur noch durch die lexikalische Besetzung der Konstituente – auf ihrem lexikalischen Surprisal. Weil dann folglich keine Kapazität mehr für die Unsicherheit auf der Funktion verbraucht wird, kann die freie Kapazität darauf verwendet werden, Phrasen und Teilsätze mit hohen Surprisalwerten zu verarbeiten. Es stehen – in dieser Betrachtungsweise – also mehr kognitive Kapazitäten für die Verarbeitung im Nachfeld zur Verfügung als in dem Teil des Satzes, der vom Vorfeld bis zur rechten Satzklammer reicht. Aus dem Grund sollten auch Phrasen und Relativsätze mit hohen Surprisalwerten im Nachfeld besser verarbeitet werden.

Theoretische Fundierung von Hypothese H1 Ein wichtiges Kriterium für diesen erhöhten Verarbeitungsaufwand ist die Klammerstruktur. Sie ist als Charakteristikum der deutschen Sprache seit spätestens dem Frühneuhochdeutschen benannt (vgl. z. B. Takada 1998; Admoni 1990; Konopka 1996). Wie in Kapitel 2.2 ausführlich beschrieben wird, verteilt sich das Prädikat bei analytischen Verbformen über beide Klammern in Verbzweitsätzen oder steht bei Verbletzstellung vollständig am Ende des Satzes. Für die inkrementelle Verarbeitung des Satzes bietet das den Rezipierenden *eigentlich* schlechte Voraussetzungen, da der Lexikoneintrag des Verbs, der nach Gibson (1998) für die Verarbeitung sehr wichtig ist, erst am Satzende vollständig erfasst wird. Neben dem auf die Vorhersage bezogenen Verbrauch kognitiver Ressourcen werden diese auch dafür verbraucht, um die Zweiteilung im Gedächtnis behalten zu können („Memory Cost“, Gibson 1998: 8).

Außer beim Präteritum und dem Präsens Aktiv, den einzigen synthetischen Verbformen des Deutschen (vgl. Gallmann 2016), ist in der linken Satzklammer die lexikalische Bedeutung des Verbs noch nicht bekannt. Den Rezipierenden ist diese Lücke bewusst. Um sie zu schließen, müssen auch später noch Kapazitäten zur Verarbeitung und Integration des lexikalischen Verbs in das Satzmodell zur Verfügung stehen.

Ein erster Mitspieler an der Verbalhandlung wird zwar bei Verbzweitsätzen bereits im Vorfeld gesehen und in die Erwartungen integriert, die lexika-

lische Bedeutung der hier stehenden Wörter als auch ihre Satzgliedfunktion müssen aber trotzdem gespeichert werden, um sie am Ende in die Gesamtheit des Satzes zu integrieren. Weitere Lücken, die das lexikalische Verb schafft, werden im Mittelfeld geschlossen, wenn singuläre Satzgliedfunktionen besetzt werden und daher nicht mehr auftreten können, wie es beispielsweise bei Objekten – mit Ausnahmen wie den Konstruktionen des doppelten Akkusativs – der Fall ist. Dadurch nimmt die Auftretenswahrscheinlichkeit der anderen, ungesehenen Satzglieder zu und die Entropie darüber, was als nächstes im Satz zu verarbeiten ist, verringert sich. Dies trifft auch auf die Wahrscheinlichkeit zu, dass die rechte Satzklammer erscheint. Je mehr Satzgliedfunktionen, unabhängig davon, ob es sich um Ergänzungen oder Angaben im Sinne der Valenztheorie (vgl. Welke 2011) handelt, aufgetreten sind, desto mehr verringert sich die Zahl der Ergänzungen und Angaben, die noch auftreten können.

Die Idee dieser Art von Verarbeitung geht auf Levy (2008) und vor ihm Gibson (1998) zurück. Sie beschreiben, dass die Rezipierenden Vorstellungen über den weiteren Verlauf des Satzes generieren. Mit jedem Wort, das sie neu lesen, könne diese Vorstellung entweder beibehalten werden oder sie müsse ersetzt und erneuert werden. Zu Beginn eines Satzes gibt es mehr Möglichkeiten, wie er fortgeführt werden kann, daher auch für die Rezipierenden mehr Erwartungen an diese Fortsetzung. Der Verarbeitungsaufwand ist dadurch höher. Je mehr Informationen aber bekannt sind, desto weniger Auswahl bleibt den Rezipierenden. Daher verringere sich der kognitive Aufwand zur Verarbeitung gegen Ende des Satzes (vgl. Levy 2008).

Das menschliche Gehirn verarbeitet allerdings nicht nur die Satzgliedfunktionen der Wörter, sondern auch ihre lexikalische Bedeutung. Auch wenn davon auszugehen ist, dass Skript- und Weltwissen die Möglichkeiten der Wortauswahl einschränken (vgl. z. B. Ostermann 2021), bleibt der Verarbeitungsaufwand aufgrund der Auswahlmöglichkeiten im Bereich der Lexik hoch, was der Verringerung des Arbeitsaufwandes beim Verstehen des Satzes insgesamt bis zu einem gewissen Grad entgegenwirkt.

Es zeigt sich also ein Zusammenspiel von zwei Faktoren, die bei der

Verarbeitung eines Satzes zusammenkommen: Die Verarbeitung der jeweiligen Satzgliedfunktionen, die Wörter und Konstituenten übernehmen, und die Verarbeitung ihrer Bedeutung. Auch wenn dargelegt wurde, dass die benötigte kognitive Kapazität für die Satzgliedverarbeitung gegen Ende des Satzes abnimmt, je weniger Möglichkeiten verbleiben, und der Verarbeitungsaufwand auf der lexikalischen Ebene einem weniger starken Abfall unterliegt, bleiben dennoch zwei Faktoren unberücksichtigt: Gibson (1998), Futrell, Gibson und Levy (2021) sowie Hahn, Degen und Futrell (2021) haben in ihren Untersuchungen gezeigt, dass auch Distanzen bei der Verarbeitung bedacht werden müssen. Das ist besonders im Deutschen aufgrund der oben erwähnten Verteilung des Prädikats auf beide Klammern der Fall.

Je mehr Material im Mittelfeld steht, das sowohl hinsichtlich seiner Funktion als auch hinsichtlich seiner Bedeutung verarbeitet werden muss, desto schwieriger wird es, alle Informationen bis zum Ende des Satzes im Gedächtnis zu behalten. Es entsteht die Gefahr des Informationsverlusts. Um dem entgegenzuwirken, können also lange oder komplexe Satzglieder ins Nachfeld verschoben werden, um die Distanz zwischen den Satzklammern zu verringern und um im Mittelfeld weniger Wörter, deren Bedeutung integriert werden muss, zu verorten. Dieser Gedanke ist beispielsweise in der Literatur zur Relativsatzextraposition von Uszkoreit et al. (1998) beschrieben, aber weniger auf den Informationsverlust und damit auf die informationstheoretische Komponente der Länge bezogen worden.

Ein weiterer Faktor, der bei der Betrachtung von Extraposition eine Rolle spielt, ist die Mündlichkeitsnähe des Textes (Kapitel 2.7), auf die die zweite Hypothese Bezug nimmt.

1.1.2. Hypothese H2: Einfluss der Mündlichkeitsnähe auf Extraposition

Auslagerung gilt als typisches Merkmal mündlichkeitsnaher Texte beziehungsweise der gesprochenen Sprache (vgl. z. B. Auer 1991; Uhlmann 1993). Auch das lässt sich innerhalb des präsentierten theoretischen Rahmens erklären. Bei der Verarbeitung von gesprochenen Informationen werden mehr kognitive

Kapazitäten genötigt als bei der Verarbeitung von geschriebenen Informationen. Das liegt primär an der Natur der Äußerung. Gesprochenes ist flüchtig. Einmal geäußert verschwindet die Information wieder und es erfordert gegebenenfalls die Bitte an die Gesprächsteilnehmenden, sie zu wiederholen. Sprechende bemühen sich also darum, ihre Sätze so zu gestalten, dass die Rezipierenden sie verstehen (zur Hörerorientierung siehe z. B. Stutterheim 2011) und sie von den Sprechenden schnell produziert werden können.

Auf eine erfolgreiche Informationsverteilung achten auch Menschen, die einen Text schreiben. Allerdings sind Sätze hier statisch, das heißt, sie können prinzipiell beliebig häufig gelesen werden – auch wenn das in der Realität ebenso unwahrscheinlich ist, weil es den Lesefluss unterbricht und das nicht gewollt ist. Dennoch bleibt der Unterschied zwischen den ‚Aggregatzuständen‘ der Information, der berücksichtigt werden muss. Daraus ergibt sich die zweite Hypothese H2:

(H2a) Auslagerung wird verstärkt in mündlichkeitsnahen Texten verwendet.

(H2b) In Texten, die mündlichkeitsnäher sind, ist der Einfluss von hohen Surprisalwerten größer als in schriftlichkeitsnahen Texten.

Da es sich hier um eine diachrone Arbeit vom 17. bis zum 20. Jahrhundert handelt, liegen die zu untersuchenden Texte nur in geschriebener Form vor. Weil alle Texte außerdem gedruckt wurden (Kapitel 4), kann davon ausgegangen werden, dass sie Revisionsprozesse durchliefen. Trotzdem ist ein Unterschied bezüglich der Mündlichkeitsnähe zu vermuten, wie er zum Beispiel für das Englische von Degaetano-Ortlieb und Teich (2018) gezeigt wurde.

Der Unterschied zwischen konzeptioneller Mündlichkeit und konzeptioneller Schriftlichkeit geht auf vor allem auf Koch und Oesterreicher (2007) zurück. Unabhängig davon, ob ein Text für die auditive oder visuelle Rezeption verfasst ist, kann er sprachliche Mittel aufweisen, die ihn mehr im mündlichen oder schriftlichen Register verorten. Das bedeutet, dass ein Text, der gedruckt vorliegt, trotzdem eher an einen Text erinnern kann, der für den mündlichen Vortrag geplant ist. Sprachliche Mittel, die das suggerieren, sind beispielsweise eine Häufung von Antwortpartikeln, Vokativen oder ei-

ne bestimmte Art der Satzverknüpfung (vgl. z. B. Selting 2012; Ortman und Dipper 2019; 2020). Es besteht eine begründete Annahme, die auch in Kapitel 7.1.1 verifiziert wird, dass die Mündlichkeitsnähe über die Jahrhunderte hinweg abnimmt, weil sich in diesem Zeitraum der wissenschaftliche Stil zu entwickeln beginnt und die Autoren der Texte diesem auch verstärkt folgen. Eng verknüpft mit der Veränderung der Konzeption ist auch der Einfluss der Zeit. Daher soll dieser Punkt in Hypothese H3 aufgegriffen werden.

1.1.3. Hypothese H3: Diachrone Perspektive von Extraposition

Frühere Arbeiten von Speyer (2015a) und Schildt (1976) haben gezeigt, dass die *Informationsstruktur* über die Jahrhunderte hinweg ihren Einfluss auf Auslagerung verliert. Aus der dargelegten Konzeption über den Verarbeitungsaufwand innerhalb des Satzes lässt sich aber nicht unmittelbar schließen, warum es diese Veränderung geben sollte.

Informationsdichte und -struktur können natürlich nicht gleichgesetzt werden. Dennoch beziehen sich beide auch auf kognitive Kapazitäten, die für die Verarbeitung von Informationen aufgewendet werden. Neue Informationen dürften tendenziell schwerer zu verarbeiten sein als bekannte Informationen und daher – theoretisch und bislang nicht überprüft – mit hohen Surprisalwerten korrelieren. Wenn man also eine Korrelation zwischen Informationsstruktur und Informationsdichte annimmt, sollte sich auch die Einflussnahme von informationstheoretischen Maßen, gemessen als Surprisalwerte, auf Auslagerung über die Jahrhunderte verändert haben. Darauf geht die dritte Hypothese H3 ein:

(H3) Über die Zeit wird der Einfluss der Informationsdichte auf Auslagerung geringer.

Diese Hypothese benötigt gleichzeitig eine vorsichtige Betrachtung, denn sie korreliert mit Hypothese H2 und mit anderen Veränderungen innerhalb der Wissenschaftssprache. Die ersten originalen wissenschaftlichen Texte in deutscher Sprache sind im 17. Jahrhundert belegt (vgl. z. B. Klein 2011; Polenz et al. 2013). Zuvor wurden wissenschaftliche Texte in lateinischer Sprache verfasst

und rezipiert und aus dem Lateinischen oder anderen Sprachen übersetzt. Die Entwicklung dieses Stils steht bei den frühesten in die Untersuchung einbezogenen Texte vermutlich erst am Anfang. Deshalb kann das Lateinische die Gestaltung der Texte (wissentlich und unbeabsichtigt) beeinflusst haben (vgl. z. B. Klein 2011). Von der lateinischen Wortstellung ist bekannt, dass sie stark von informationsstrukturellen Kriterien beeinflusst und sonst praktisch frei ist (vgl. z. B. Panhuis 2011). Auch größere Distanzen zwischen zusammengehörenden Elementen sind in lateinischen Texten häufig zu beobachten und Gleiches gilt für die Länge einer Satzperiode (vgl. z. B. Gulordava und Merlo 2015). Wissenschaftler des 17. Jahrhunderts hatten kaum Vorlagen, nach denen sie sich bei der Gestaltung ihrer Texte orientieren konnten. Im Einzelnen ist es aber nicht möglich, nachzuvollziehen, wie gut die Lateinkenntnisse der einzelnen Autoren waren, deren Texte Bestandteil des Korpus sind.

Zudem hat sich die Klammerstruktur des deutschen Satzes bereits im Frühneuhochdeutschen fest etabliert (vgl. z. B. Admoni 1990).⁶ In den ersten untersuchten Zeitabschnitten treten also zwei konkurrierende Prinzipien zusammen. Zum einen dürften die Wissenschaftler noch an die Stellungsfreiheiten und langen Abhängigkeiten aus dem Lateinischen gewöhnt sein, weil ihnen Vorbilder auf Deutsch fehlen, zum anderen etabliert sich die Distanzstellung der Satzklammern und es beginnt sich bei den Grammatikern des 18. Jahrhunderts eine Tendenz zu festigen, die Satzrahmen möglichst kurz zu halten (vgl. Konopka 1996; Takada 1998). Wenn sich folglich die Gewöhnung an kurze Mittelfelder durchzusetzen beginnt und deren Verarbeitung erleichtert, wie Futrell et al. (2021) argumentieren, und das Lateinische an Bedeutung für das wissenschaftliche Schreiben verliert, können informationelle und strukturelle Gestaltungsprinzipien von Sätzen in Konkurrenz zueinander treten. Im Sinne von Hypothese H3 bedeutet das, dass der Einfluss von Informationsdichte durch das Streben nach einer kurzen Distanz zwischen den

⁶ Auch das korreliert mit der Mündlichkeitsnähe. Gerade für das gesprochene Deutsche gibt es Ansätze, die die Definierbarkeit eines Satzes diskutieren (vgl. Hoffmann 2019). Wenn man davon ausgeht, dass das frühe Neuhochdeutsche insgesamt noch eine stärkere Orientierung am Gesprochenen aufweist, passt das zu dieser Behauptung. Dennoch bleibt in dem vorliegenden Rahmen der Fokus stärker darauf, dass im Nachfeld mehr kognitive Kapazitäten vorhanden sind, die bei medial und konzeptionell mündlicher Kommunikation (vgl. Koch und Oesterreicher 2007) noch optimaler genutzt werden müssen als bei schriftlicher.

Satzklammern abgelöst wird, die beispielsweise durch die Extraposition von langen Konstituenten erreicht wird.

Der genaue Einfluss möglicher lateinischer Vorbilder kann in dieser Arbeit nicht überprüft werden, weshalb auch der Umweg über den Faktor Zeit genommen werden muss. Je jünger die Texte sind, desto mehr ist ein abnehmender Einfluss des Lateinischen anzunehmen, da die Sprachvorbilder in der eigenen Muttersprache zunehmen und andere Sprachkontakte wie mit dem Französischen auftreten (vgl. Polenz et al. 2013). Je bewusster sich die Autoren oben erläuterten Prinzipien werden, desto eher werden sie sie auch umgesetzt haben. Dazu zählt beispielsweise auch die Festigung des Satzrahmens, wie die Untersuchungen von Schildt (1976) und Speyer (2015a) und die deskriptiven Erläuterungen von Admoni (1990) zeigen. Doch auch andere Einflussfaktoren können einen Einfluss auf die Extrapositionsgründe ausgeübt haben. Hierzu zählen beispielsweise das explizite Herausbilden des wissenschaftlichen Stils, der im Laufe der Jahrhunderte auch von den Theologen übernommen wird, sowie Anforderungen daran, auch komplizierte, neue Themen vor allem im medizinischen Bereich erfolgreich darstellen zu können. Es zeigt sich also, dass in der dritten Hypothese verschiedene Faktoren ineinanderfließen.

1.2. Aufbau der Arbeit

Um die eben dargestellten Hypothesen untersuchen zu können, ist die Arbeit wie folgt aufgeteilt: Nach der Einleitung (Kapitel 1) folgt eine Darstellung des theoretischen Hintergrunds (Kapitel 2). Kapitel 2.1 behandelt das Konzept der Informationsdichte nach Shannon (1948), bevor die *Information Theory* nach Levy (2008) und die *Syntactic Prediction Locality Theory* nach Gibson (1998) präsentiert werden und auf ausgewählte Studien und Theorien, die sich anhand von Korpusstudien nachweisen lassen, Bezug genommen wird.

In Kapitel 2.2 folgt aufgrund der Bedeutung des Nachfeldes (Kapitel 2.2.4) für die hier verwendete Definition von Extraposition die Beschreibung des *topologischen Feldermodells* (vgl. v.a. Höhle 1982; Wöllstein 2010; 2014). Zudem greifen die verschiedenen Unterkapitel auf frühere Satzmodelle zurück

und diskutieren diese kritisch (Kapitel 2.2.1 zu den Satzklammern, Kapitel 2.2.2 zum Vorfeld und Kapitel 2.2.3 zum Mittelfeld). Eine genauere Auseinandersetzung mit dem Begriff der Extraposition und eine Abgrenzung zu seiner Verwendung bei anderen Autoren erfolgt unter Kapitel 2.3.

Den Rest von Kapitel 2 bildet die Darstellung der untersuchten Phänomene. Auf Relativsätze geht Kapitel 2.4 ein. Hier werden nacheinander die verschiedenen Relativsatzarten voneinander abgegrenzt (Kapitel 2.4.1), das Relativum besprochen (Kapitel 2.4.2) und schließlich Informationen über die Restriktivität des Relativsatzes, die in dieser Arbeit als Relativsatztyp bezeichnet wird, gegeben (Kapitel 2.4.3). Im Anschluss daran (Kapitel 2.5) werden die Nominalphrasen in folgenden Punkten besprochen: Ihre Form (Kapitel 2.5.1), ihre Erweiterbarkeit (Kapitel 2.5.2) und ihrer Entwicklung mit Fokus auf dem Frühneuhochdeutschen (Kapitel 2.5.3). Der Aufbau des Kapitels zu den Präpositionalphrasen folgt dem gleichen Prinzip (Kapitel 2.6). Nacheinander werden hier die Form (Kapitel 2.6.1), Funktion (Kapitel 2.6.2) und Entwicklung der Präpositionalphrase (Kapitel 2.6.4) sowie der Bedeutung der Präposition (Kapitel 2.6.3) dargestellt. Das Kapitel schließt schließlich mit der Betrachtung der Definitionen von Mündlichkeit und Schriftlichkeit (Kapitel 2.7), wo vor allem auf die Konzeption nach Koch und Oesterreicher (2007) und der Kritik daran eingegangen wird.

Auf die Gründe für Extraposition von Phrasen und Relativsätzen geht Kapitel 3 ein. Aufgrund des Fehlens einer breiten Forschungsbasis zu möglichen Auslösern für die Extraposition von Nominal- und Präpositionalphrasen werden diese zusammengenommen (Kapitel 3.1). Im Anschluss daran folgt die Betrachtung zur Extraposition von Relativsätzen (Kapitel 3.2). Für beide Kapitel werden jeweils ausgewählte Grammatiken zu dem Thema und empirischer Forschung synchroner und diachroner Art herangezogen.

Ab Kapitel 4 wird meine eigene Untersuchung vorgestellt. Das beginnt mit der Korpuserstellung. Zunächst wird hier die Auswahl der wissenschaftlichen Texte begründet (Kapitel 4.1), bevor die Zusammensetzung des Korpus präsentiert wird (Kapitel 4.2). In Kapitel 4.3 wird beschrieben, wie die Phänomene annotiert wurden. Dieser Teil der Arbeit schließt mit Begründungen

dafür, welche Daten aus welchen Gründen aus der Analyse ausgeschlossen werden mussten (Kapitel 4.4).

Kapitel 5 beschreibt *Sprachmodelle* allgemein und stellt dann das in dieser Arbeit verwendete Language Model vor (Kapitel 5.1), bevor die Surprisalmaßße präsentiert werden. Danach erfolgt die Beschreibung, wie der Orality Score als Kennwert für Mündlichkeitsnähe berechnet wird (Kapitel 5.2), und die Darstellung der genutzten statistischen Analyseverfahren (Kapitel 5.4).

Kapitel 6 befasst sich mit den Ergebnissen, die zur Beantwortung von Hypothese H1 notwendig sind, und diskutiert diese in Kapitel 6.5. Zudem ist das Kapitel in die Betrachtung der drei Phänomene, Nominalphrasen (Kapitel 6.1), Präpositionalphrasen (Kapitel 6.2) und Relativsätze (Kapitel 6.3) unterteilt. Der Zusammenhang zwischen der Mündlichkeitsnähe und der Auslagefrequenz beziehungsweise der Informationsdichte der Phänomene wird schließlich in Kapitel 7 beschrieben und in Kapitel 7.2 diskutiert. Die separate Darstellung der drei untersuchten Phänomene wird auch in Kapitel 8 verwendet, in dem die Ergebnisse zur Überprüfung der dritten Hypothese wiedergegeben und in Kapitel 8.4 diskutiert werden.

Die Zusammenfassung der Ergebnisse sowie der Diskussionen (Kapitel 9) schließt die Arbeit ab.

2. Theoretischer Hintergrund

Dieses Kapitel soll einen Überblick über die in dieser Arbeit relevanten linguistischen Termini und Theorien geben. Es beginnt mit der Darstellung der Informationsdichte (Kapitel 2.1), wobei es sich auf die Modelle von Shannon (1948), Gibson (1998), Levy und Jaeger (2007), Levy (2008) und Futrell et al. (2021) fokussiert und Arbeiten in den Mittelpunkt stellt, die einen korpuslinguistischen Hintergrund haben.

Des Weiteren werden in Kapitel 2.2 das topologische Feldermodell nach Wöllstein (2014) und Höhle (1982), die Begriffsverwendung von *Extraposition* (Kapitel 2.3) beschrieben und im Anschluss die in dieser Arbeit betrachteten Phänomene, nämlich Relativsätze (Kapitel 2.4), Nominalphrasen (Kapitel 2.5) und Präpositionalphrasen (Kapitel 2.6), allgemein dargestellt.

2.1. Informationsdichte

2.1.1. Informationstheorie nach Shannon (1948)

Informationsdichte geht auf Shannon (1948) zurück und ist ein Konzept zur mathematischen Beschreibung von Kommunikation. Einer ihrer Vorteile liegt darin, dass sie die oft eher vagen Konzepte von Information (vgl. z. B. Prince 1981; Gundel, Hedberg und Zacharski 1993; Krifka 2008; O'Grady 2016) objektiv analysierbar und auch berechenbar macht.¹

Dabei bedient sie sich eines auch in der Linguistik bekannten Systems, dem Bühlerschen Organon-Modell (vgl. Bühler 1934; Gallmann 2016), bestehend aus den Komponenten Sender, Empfänger, Übertragungsweg und der

¹ Teile des in diesem Kapitel vorgestellten Materials flossen auch in die Veröffentlichungen Voigtmann und Speyer (2021), Speyer und Voigtmann (2021a;b), Voigtmann und Speyer (2022) und Voigtmann und Speyer (2023) ein.

Nachricht selbst.² Der Sender intendiert, eine Information zu übermitteln. Diese Information kodiert er in einer Form, die durch den Übertragungsweg übertragen werden kann. Im Falle von Sprache geschieht das dadurch, dass die Information in Phoneme bei mündlicher und Grapheme bei schriftlicher Übertragung transformiert wird, wobei sich sowohl Phoneme als auch Grapheme entsprechend der Grammatik der Sprache kombinieren lassen. Beim Empfänger der Nachricht setzen Dekodierungsmechanismen ein, mithilfe derer er aus den überführten Signalen die Information der Nachricht entnimmt. Bei fast allen Sprachen ist dabei zu bemerken, dass es eine große Bandbreite an Variationen im Bereich der Kodierung gibt, die keine Auswirkung auf den übermittelten Inhalt hat (vgl. Shannon 1948: 380). In gesprochener Sprache lässt sich beispielsweise die Länge von Phonen variieren und im Bereich der Morphologie können Sprecher und Schreiber Verkürzungen anwenden (1a). Die Lexik bietet Variationsmöglichkeiten zwischen semantisch sehr ähnlichen Ausdrücken, um den gleichen Sachverhalt auszudrücken (1b). Syntaktisch offerieren das Deutsche und auch zahlreiche andere Sprachen Stellungsfreiheiten (1c) und für die Pragmatik lassen sich beispielsweise unterschiedliche Referenzausdrücke anführen (1d):

- (1) a. Ich **habs** dir doch gesagt. Ich **habe es** dir doch gesagt.
 b. Heute gibt es **Teigwaren** in der Mensa. Heute gibt es **Nudeln** in der Mensa.
 c. **Gestern** habe ich ihm das Buch zurückgegeben. Ich habe ihm das Buch **gestern** zurückgegeben.
 d. Ich habe **deinen Mann** getroffen. Ich habe **ihn** getroffen.

Vergleicht man jeweils beide fettgedruckten Varianten lässt sich erkennen, dass die Proposition der Sätze unverändert bleibt. Wenn man über optimale Kommunikation im Sinne der Informationsdichte spricht, versteht man darunter nicht effizientere Propositionen, sondern deren Präsentations- beziehungsweise

² In diesem Fall handelt es sich um abstrakte Konzepte, die technisch zu verstehen sind und daher nicht auf Personen referieren, da Shannon (1948) das System anhand von Telegrafie erklärt. Daher wird hier nicht gegendert, obwohl es das Modell auch bezogen auf menschliche Sprecher:innen gibt.

ungsweise Kodierungsmöglichkeiten. Im weiteren Verlauf des Kapitels soll darauf eingegangen werden, wie die verschiedenen Repräsentationsformen die effiziente Übertragung der Information beeinflussen. Dies zeigt Shannon (1948) an seinem mathematischen Modell, das die Grundlage aller weiteren Beschäftigungen mit Informationsdichte bildet.

Nach Shannon (1948) sind sowohl der Sender als auch der Empfänger dazu in der Lage, schon während der Übertragung der Nachricht verschiedene Kodierungen aus einem Set von Codes auszuwählen und zu entschlüsseln. Ist das Set der möglichen Nachrichtenkodierungen endlich, dann ist diese endliche Zahl als monotone Funktion beschreibbar (vgl. Shannon 1948: 379). Sie gilt als Maß für die Information, die produziert wird, wenn eine spezielle Kodierung genutzt wird und alle Wahlmöglichkeiten gleich wahrscheinlich sind. Als mathematisches Modell für den Auswahlprozess der verschiedenen, potenziell gleich wahrscheinlichen Kodierungen wird die Logarithmusfunktion zur Basis 2 genutzt (vgl. Shannon 1948: 379). Die Art der Kodierung beeinflusst, wie erfolgreich die Kommunikation verläuft. Die Berechnung dieses Erfolgs lässt sich mithilfe von drei Kommunikationsklassensystemen darstellen: Dem diskreten, kontinuierlichen und gemischten System (vgl. Shannon 1948: 380f.).

Das *diskrete System* lässt sich am besten am Beispiel der Telegrafie verdeutlichen. Jedem Buchstaben des Alphabets kann ein Repräsentant des Morsealphabets zugeordnet werden. Aus einem bestimmten Kontingent elementarer Symbole werden also Symbolsequenzen übertragen, deren mögliche Zeichenkombinationen eingeschränkt sind (vgl. Shannon 1948: 383).

Anhand dieser Voraussetzungen lässt sich die (endliche) Kapazität des Kanals ermitteln. Sie ist abhängig vom System selbst, dem Verhältnis zwischen einem Zeichen und seiner Repräsentation³ und der Kombinierbarkeit der Symbole. Unter Beachtung dieser Einschränkungen kann die maximale Kanalkapazität mathematisch bestimmt werden (vgl. Shannon 1948: 382).

Weitere Einschränkungen betreffen die verschiedenen Zustände in einem Set. Für jeden Zustand können nur bestimmte Signale übermittelt werden.

³ Bei der Telegrafie wird zum Beispiel jedem Symbol eine bestimmte Dauer zugeordnet (vgl. Shannon 1948: 381).

Mit jedem übermittelten Signal entsteht ein neuer Zustand, der abhängig vom vorherigen Zustand und den übermittelten Zeichen ist. Da die Signalübermittlung Regeln unterliegt, ist die Auswahl der transmittierbaren Signale eingeschränkt. Wurde beispielsweise in einem Satz bereits ein Subjekt eingeführt, kann kein weiteres Subjekt auftreten, dafür aber – je nach Satzart – das Prädikat oder eine andere Konstituente.

Neben der Kodierung der Nachricht und ihrer Übertragung ist auch die Art der Informationsquelle bedeutsam (vgl. Shannon 1948: 383f.). Um effizient kommunizieren zu können, wird eine Nachricht in Abhängigkeit des statistischen Wissens über die Quelle beziehungsweise über den Sender kodiert. Wenn hier eine zu den strukturellen Gegebenheiten der Sprache passende Kodierung genutzt wird, wird die Kanalkapazität in einem geringeren Maß ausgelastet (vgl. Shannon 1948: 383). Da diese Strukturen sowohl dem Sender als auch dem Empfänger klar sind, führt dies entweder zu einer Zeitersparnis bei der Übertragung der Nachricht oder zu einer weniger starken Belastung des Kanals, wenn die Nachrichtensequenz angemessen in die Signalsequenz kodiert wurde (vgl. Shannon 1948: 384). Eine effiziente Übertragung geschieht also, indem vorgegebene Strukturen einer Sprache eingehalten werden, beispielsweise die Verbzweitstellung von Verben in den (kanonischen) Aussagesätzen des Deutschen.

Die Übertragung der Symbole geschieht sukzessiv und in Abhängigkeit des vorherigen Symbols sowie des Symbols selbst. Daher lässt sich das System als stochastischer Prozess beschreiben und unterliegt den Gegebenheiten der Wahrscheinlichkeitsrechnung (vgl. Shannon 1948: 384). Der Auswahlprozess ist komplex. Unterschieden werden muss zum einen, ob nur das unmittelbar vorausgehende Element Einfluss auf die Wahl des betreffenden Elements nimmt oder ob zum anderen mehrere vorausgehende Elemente von Bedeutung sind. Mathematisch lässt sich das wie folgt darstellen: $p_i(j)$, womit die Wahrscheinlichkeit beschrieben wird, dass das Element j auf das Element i folgt (vgl. Shannon 1948: 384). Jede sprachliche Einheit kann dabei als Element verstanden werden (vgl. Gibson, Futrell, Piantadosi, Dautriche, Mahowald, Bergen und Levy 2019).

Wird nur das Element j in seiner Häufigkeit betrachtet, spricht man von *Unigram*-Frequenzen. Sie bilden die einfachste Möglichkeit, sich dem stochastischen Prozess zur Auswahl von Elementen zu nähern, weil der Kontext des Elements hier außer Acht gelassen wird. Diese Betrachtungsart spiegelt die natürliche Sprache jedoch nicht ausreichend wider. Dazu muss mehr Kontext betrachtet werden, was dann als *Bigram*, *Trigram* etc. bezeichnet wird. Je größer dieser in Betracht gezogene Kontext wird, desto mehr nähern sich die Ergebnisse der tatsächlichen Sprache an: „[A] sufficiently complex stochastic process will give a satisfactory representation of a discrete source“ (Shannon 1948: 386).

Einen solchen Prozess bildet der *Markoff-Prozess* ab. Dieser besagt, dass die Wahrscheinlichkeit einer zukünftigen Einheit vorausgesagt werden kann, ohne zu weit in die Historie zu schauen (vgl. z. B. Mürmann 2014: 183ff.). Dafür benötigt man eine bestehende, finite Anzahl möglicher Zustände eines Systems (S_1, S_2, \dots, S_n) sowie ein Set aus Übergangsmöglichkeiten und abschließend die Wahrscheinlichkeit $p_i(j)$, dass der Status i zum Status j übergeht. Die verschiedenen Zustände sind also abhängig von dem, was vorher aufgetreten ist („residue of influence“, Shannon 1948: 387).

Neben den diskreten Systemen gibt es auch *gemischte Quellen*. Dabei liegen verschiedene Quellen (L_1, L_2, \dots, L_n) vor. Sie sind ergodisch, haben also die gleichen Auftretenswahrscheinlichkeiten (vgl. Shannon 1948: 388). Zu Beginn der Übertragung herrscht kein Wissen darüber, welche Quelle genutzt wird. Sobald jedoch eine Quelle ausgewählt wird, verhält sie sich gemäß ihrer internen statistischen Komponente. Um das zu erklären, lassen sich wieder die Wahlmöglichkeiten in folgendem Beispiel (2) heranzuführen:

- (2) a. Klaus sagt Ursel, dass er nächste Woche den Rasen mähen möchte.
- b. Klaus sagt Ursel, er möchte nächste Woche den Rasen mähen.

Sobald die Wahl zwischen den verschiedenen Fortführungen der Bitte getroffen wurde, ist diese spezielle Fortführung an die spezifischen Klassifikationen der Konstruktion gebunden. In (2a) determiniert die Wahl des durch eine Sub-

junktion eingeleiteten Nebensatzes, dass „das finite Verb in der rechten Satzklammer“ (Meibauer, Steinbach und Altmann 2013: 232) zu finden sein muss, beziehungsweise das Fehlen der Subjunktion die Verbzweitstellung (2b).

Wenn diese Quellen also auch einem Markoff-Prozess unterliegen, kann die Informationsrate ermittelt werden. Dabei existiert ein Set möglicher Auftretenswahrscheinlichkeiten der einzelnen Quellen p_1, p_2, \dots, p_n . Ihre Wahrscheinlichkeit ist bekannt, aber man weiß nicht, welches Element ausgewählt werden wird. Liegt in der Mathematik ein solches Problem vor, wird die *Entropie* H genutzt, um es zu lösen. Entropie ist „ein Maß dafür [...], wie hoch der Ordnungsgrad des betrachteten Systems ist“ (Tipler, Mosca und Wagner 2015: 621). Mit ihr kann also berechnet werden, wie wahrscheinlich ein Ereignis korrekt erraten werden kann (vgl. Genzel und Charniak 2002: 199). Sie erfüllt dabei verschiedene Voraussetzungen.

Die Entropie eines Elements ist kontinuierlich. Falls alle Auftretenswahrscheinlichkeiten der verschiedenen Quellen gleich sind, kann H als monoton wachsende Funktion beschrieben werden. Das kennzeichnet die größte Unsicherheit über die Wahl des Ereignisses. Die Entropie kann nicht kleiner als 0 werden. Sie ist dann 0, wenn nur ein bestimmtes Ereignis auftreten kann. Wenn aber alle möglichen Ereignisse gleich wahrscheinlich sind, ist die Entropie am größten. Auch ist die Entropie dann größer, wenn alle Ereignisse *ähnlich* wahrscheinlich sind (vgl. Shannon 1948: 390f.).

Dieses Maß für Entropie liegt den meisten weiterführenden Arbeiten über Informationsdichte zugrunde. Mit welcher Güte man Ereignisse vorherzusagen kann, wenn ein bestimmtes anderes Ereignis bereits bekannt ist, lässt sich also berechnen. Beispielsweise ermöglicht die Entropie zweier abhängiger Ereignisse vorherzusagen, wie wahrscheinlich das Auftreten eines Prädikats nach einer subordinierenden Konjunktion im Deutschen ist. Außerdem kann die Entropie auch in Verbindung mit den Informationsquellen gebracht werden. Bei einer diskreten Quelle mit den schon erwähnten Bedingungen gilt Folgendes: H_i bildet die Entropie für jeden möglichen Zustand der Quelle.

Die *relative Entropie* bildet damit das Verhältnis zwischen dem maximalen Wert ab, den eine Quelle haben kann, wenn sie auf dieselben Symbole

beschränkt ist, was auf sprachliche Übertragungen in aller Regel zutrifft, und bildet auch den maximalen Wert der Entropie der Quelle ab (vgl. Shannon 1948: 393). Mit ihrer Hilfe lässt sich berechnen, wie redundant eine Sprache ist. Die Redundanz der Grammatik des Englischen liegt bei etwa 50% (vgl. Shannon 1948: 393). Außerdem spielt bei der Informationsdichte auch die Übermittlung der Informationen durch „transducer[s]“ (Shannon 1948: 394) eine Rolle. Sie übermitteln sowohl den Input als auch den Output als Sequenz diskreter Symbole und stehen in Verbindung zur Übertragung einer Nachricht durch den Kanal. Die Rate, mit der diese Symbole durch den Kanal zu dessen optimaler Auslastung, aber nicht Überlastung fließen, kann ebenfalls mit der Entropie berechnet werden.

Mathematisch lässt sich dieses Theorem auch gut beweisen (vgl. Shannon 1948). Linguistisch und auf menschliche Kommunikation bezogen ist das deutlich schwieriger. Was für menschliche Kommunikation demonstriert werden kann, ist die Verbesserung der Übertragungsrates und damit eine effizientere Kommunikation, wenn die Kodierung der Nachricht dazu dient, die Auftretenswahrscheinlichkeiten von Signalen zu erhöhen, sie also wahrscheinlicher zu machen und die Auswahl von dem, was noch kommen kann, einzuschränken.

Je mehr Elemente übermittelt wurden, desto höher wird die Wahrscheinlichkeit des n -ten Elements im Vergleich zum ersten. Inkludiert man diese Erkenntnis in ein binäres System, lässt sich der Schluss ziehen, dass Nachrichten, die mit hoher Wahrscheinlichkeit auftreten, kürzer kodiert werden und solche mit niedriger Auftretenswahrscheinlichkeit länger sind. Ähnliche Ansätze finden sich auch beispielsweise bei Aylett und Turk (2004), Levy und Jaeger (2007) und Jaeger (2010). Relevant für die weitere Überlegungen ist die Unterschiedlichkeit aller Codes, die ihre Rekonstruierbarkeit zulassen.

Die Kernfragestellung, der Shannon (1948) nachgeht, besteht eigentlich darin, die Voraussetzungen zur optimalen Nachrichtenübermittlung durch einen *störungsbehafteten Kanal* zu beschreiben und mathematisch zu erklären. Die hier bisher vorgestellten Überlegungen beziehen sich auf einen *störungsfreien Kanal* („noiseless channel“, Shannon 1948: 395ff.). Auf menschliche

Kommunikation bezogen ist eine störungsfreie Kommunikation allerdings selten. Geräusche von außen können die Kommunikation ebenso beeinträchtigen wie Beschränkungen des Gedächtnisses, Ablenkungen oder auch das Alter von Personen (vgl. z. B. van Os, Kray und Demberg 2021). Dennoch laufen die meisten Gespräche auch dann erfolgreich, wenn der Sprecher etwas anderes sagt, als der Empfänger versteht, der Input also nicht mehr identisch mit dem Output ist (vgl. Shannon 1948: 398).

Laut Shannon (1948: 398) gibt es zwei Arten von Störungen. Entweder führt die Störung dazu, dass immer der gleiche Fehler auftritt („distortion“, Shannon 1948: 398) oder dass es zu verschiedenen Fehlern kommt. Shannon (1948) nimmt den zweiten Fall an und fragt, ob sich das Empfangene in Abhängigkeit des Lärms durch einen stochastischen Prozess vorhersagen lässt. Für die Betrachtung eines verrauschten Kanals ist anzunehmen, dass eine endliche Anzahl von Zuständen und ein Set aus verschiedenen Wahrscheinlichkeiten vorliegt: $p_{\alpha,i}(\beta, j)$ (vgl. Shannon 1948: 398). Dadurch wird die Wahrscheinlichkeit ausgedrückt, dass ein Kanal im Zustand α ist und Element i übermittelt wurde, beim Empfänger aber Element j ankommt und der Kanal in den Zustand β übergeht. Wenn diese übermittelten Elemente voneinander unabhängig gestört werden, besteht nur ein Zustand zur Beschreibung des Kanals durch das Set der Übergangswahrscheinlichkeiten: $p_i(j)$. Hiermit wird die Wahrscheinlichkeit beschrieben, dass das Inputelement i als Outputelement j empfangen wird.

Während bei einem störungsfreien Kanal eine gemeinsame Entropie von Kanal und Quelle vorliegt, müssen in einem verrauschten Kanal sowohl die Entropie der Quelle als auch die der Störung berechnet werden. Wenn die gemeinsame Entropie ermittelt werden soll, muss die Entropie einmal in Abhängigkeit des Inputs und einmal abhängig vom Output berechnet werden. Diese Unterschiede bieten erste mathematische Hinweise auf die beobachtete Differenzierung zwischen Hörer- und Sprecherorientierung bei der Kodierung von Nachrichten (vgl. z. B. Stutterheim 2011). Einmal ist dabei nur der Input bekannt, was der Sprecherorientierung nahekommt, und einmal nur der Output, was der Hörerorientierung entspricht.

Nach Shannon (1948: 399) besteht die Möglichkeit, die Nachricht so zu kodieren, dass sie auch bei Störungen passend durch den Kanal transportiert werden kann und zumindest ihr Inhalt, in den meisten Fällen aber auch ihre Form vom Empfänger dekodiert und verstanden werden kann.

Zudem widmet sich Shannon (1948) einem hypothetischen Konstrukt, wie eine unvollständig oder falsch übermittelte Nachricht verbessert werden kann, wenn der Moment der Fehlerquelle und die Art des Fehlers bekannt ist. Für die meisten realen Situationen ist dieses Gedankenexperiment nicht anwendbar: Wenn die Fehlerquelle bekannt wäre, könnte sie vermieden werden und man würde ein Ergebnis erhalten, das dem des störungsfreien Kanals entspricht. Dennoch wird im Folgenden auf den Ansatz von Shannon (1948) eingegangen, der Grundlagen für die korrekte Kodierung einer Nachricht für die Übertragung durch einen störungsbehafteten Kanal bildet. Shannon (1948: 399f.) gibt an, dass die richtige Menge an Korrektur der fehlenden Information in der empfangenen Nachricht und der Unsicherheit darüber entspricht, wenn das korrekte Signal übermittelt wurde. Daher ist es sinnvoll, die konditionelle Entropie für den Schlussfolgerungsprozess zu nutzen, wenn man das empfangene Signal kennt und als Maß für die fehlende Entropie nutzen kann. Um das zu verdeutlichen, befasst sich Shannon (1948: 399f.) mit einem kommunikativen System und einem externen Beobachter, der sowohl In- als auch Output mit allen Fehlern sieht, die aufgrund der Störung geschehen. Dieser Beobachter korrigiert die Daten, wobei die Korrektur der Menge an zusätzlicher Information entspricht, die zur Korrektur der Nachricht notwendig ist. Ist der Output bekannt, kann eine Menge möglicher Inputs berechnet werden (vgl. Shannon 1948: 399). Wenn also in einem störungsbehafteten Kanal Informationen redundant transportiert werden, können die Fehler reduziert werden (vgl. Shannon 1948: 401). Theoretisch kann das durch eine unendliche Zahl von Wiederholungen geschehen, was jedoch insgesamt nicht besonders wahrscheinlich ist. Auch die quasi unendliche Erhöhung der Redundanz, die zur vollständigen Beseitigung der Fehler nötig wäre, ist unrealistisch und in der Tat auch nicht notwendig. Stattdessen sorgt die geeignete Kodierung für das Verständnis der Nachricht trotz Übertragungsstörungen.

Ein inkorrekt empfangenes Signal kann grundsätzlich wieder hergestellt werden. Denn die Veränderung ist nie so groß, dass ein völlig anderes Ursprungssignal angenommen werden muss. Erreicht wird dies durch die Redundanz der Sprache, die bestimmte Strukturen bereits vorgibt. Zugleich werden einzelne Elemente beziehungsweise Signale nicht kontextfrei durch den Kanal transportiert, sondern stehen immer in Abhängigkeit von und in Verknüpfung mit anderen Elementen. Sie sind daher durch den Kontext rekonstruierbar. Diese Rekonstruktion basiert auf den Auftretenswahrscheinlichkeiten, dem *Surprisal* eines Wortes.

Shannons (1948) Informationsdichtetheorie lässt sich folgendermaßen zusammenfassen. Durch das zugrundeliegende System aus Sender, Kanal und Empfänger werden Nachrichten übermittelt. Diese Nachrichten werden in einer Form kodiert, die eine möglichst optimale Übertragung durch den Kanal möglich macht. Der Kanal besitzt dabei eine bestimmte Bandbreite, die zur Vermeidung von Informationsverlust nicht überschritten werden sollte. Die Kodierung der Nachricht erfolgt hinsichtlich ihres Informationsgehalts möglichst gleichförmig und in Abhängigkeit vorheriger Signale und der (allgemeinen) Sprachstruktur. Das ist besonders für verrauschte Übertragungsverhältnisse relevant, denn bestimmte Wörter sind in ihrem Kontext erwartbarer als andere Wörter (3).

(3) Sie dürfen die Braut jetzt [küssen].

Die Abfolgen in Satz (3) sind so bekannt und daher so erwartbar, dass *küssen* wenig überraschend ist. Daher ist der Surprisalwert dieses Wortes gering. Das Surprisal wird durch den negativen Logarithmus der Wahrscheinlichkeit eines Elements gegeben eines Kontexts mit folgender Formel berechnet:

$$P(\text{Wort}) = -\log(P(\text{Wort}|\text{Kontext})).$$

Auch hier ist zu unterscheiden, wie weit der Kontext gefasst wird. Im Fall von Unigram-Surprisalwerten ist nur die Häufigkeit des Elements relevant. Bei Bigram-Surprisalwerten wird das Element vor dem betrachteten Element (hier: *jetzt*) noch miteinbezogen, bei Trigram-Surprisalwerten die beiden vorgehenden Elemente (hier: *Braut jetzt*) usw. Je geringer der Surprisalwert,

desto weniger Information beinhaltet dieses Element und desto leichter ist es zu verarbeiten. In diesem Beispiel (3) wurden die Surprisalwerte auf Wörtern beschrieben. Der Bezug zwischen der Vorhersagbarkeit von linguistischem Material und effizienter Kommunikation besteht auf allen sprachlichen Ebenen (vgl. z. B. Gibson et al. 2019) und auch ein Zusammenhang zwischen Verarbeitungsaufwand, also der psycholinguistischen Realität, und der Informationsdichte konnte gezeigt werden (vgl. z. B. Levy 2008).

Gibson et al. (2019) haben in einer Metastudie zusammengefasst, welche sprachlichen Bereiche durch das Streben nach effizienter Kommunikation beeinflusst sind. Sie finden, dass die Kodierung der Nachricht entscheidend ist: Kürzere Signale gelten als effizienter. Beispiele dafür finden sich in der Lexik. Hochfrequente Wörter, unabhängig ihres Kontexts, sind im Schnitt deutlich kürzer (vgl. z. B. Zipf 1965). Dabei müsste für eine optimale Kodierung eigentlich die Auftretenswahrscheinlichkeit in einem Kontext ausschlaggebend für die Wortlänge sein (vgl. Gibson et al. 2019: 394). Tatsächlich trifft das beispielsweise auf die Verwendung von Abkürzungen zu (vgl. Gibson et al. 2019: 394). In einem unerwarteten Kontext würde eher die Vollform genutzt werden, bei einem erwartbaren Auftreten hingegen die Abkürzung.

Wörter werden im sogenannten *Lexikon* gespeichert. Das Lexikon ist zwischen der Informativität eines Wortes und der Komplexität ausbalanciert. Es ist hier hervorzuheben, dass dieser Informationsbegriff von dem von Shannon (1948) abweicht. Die Informativität beschreibt, wie erfolgreich das entsprechende Wort zur Beschreibung eines Zustands in der Welt genutzt werden kann; die Komplexität die Geschwindigkeit, mit der es erlernt und wiedergegeben werden kann (vgl. Gibson et al. 2019: 395). Ein gutes Beispiel dafür bieten Farbkodierungen. Vielleicht beschreibt eine Farbe wie *mauve* den tatsächlichen Zustand eines Gegenstands akkurat. Das Wort könnte aber nicht erfolgreich genutzt werden, wenn die Kommunikationsteilnehmenden es nicht kennen. Eine umständlichere Umschreibung wie *eine Mischung aus altrosa und lila* könnte erfolgreicher sein, obwohl es nicht der optimalen Kodierung entspricht. Trotzdem greift hier das Prinzip der Komplexität, da die Farben altrosa und lila vermutlich früher und schneller erlernt wurden.

Neben der lexikalischen Ebene findet sich der Einfluss von Vorhersagbarkeit auch in der Syntax. So wurde festgestellt, dass die SVO-Reihenfolge⁴, gefolgt von den SOV- und VSO-Reihenfolgen, die effizienteste ist, da sie für das gleichmäßigste Informationsprofil sorgt (vgl. Gibson et al. 2019: 397). Um in einem störbehafteten Kanal auch bei SOV-Sprachen eine optimale Kommunikation zu ermöglichen, sind morphologische Anpassungen wie eine Kasusmarkierung notwendig. Morphologische Markierungen signalisieren Abhängigkeiten, die nicht durch eine feste Wortreihenfolge gekennzeichnet sind und korrelieren nach Gibson et al. (2019: 399) negativ mit fester Wortreihenfolge.⁵

2.1.2. Information und Syntactic Prediction Locality Theory

Die zuvor beschriebenen Annahmen und Beispiele sollen nun in den Kontext syntaktischen Verständnisses gesetzt werden. Dazu soll auf die *Syntactic Prediction Locality Theory* (vgl. Gibson 1998) eingegangen werden, die eine Kombination aus vorhersagebasierten und gedächtnisbasierten Verarbeitungsschwierigkeiten beschreibt und für den weiteren Verlauf der Arbeit besonders relevant ist.

Gibson (1998) beschreibt einen Ansatz, um Satzverarbeitung und Verarbeitungsressourcen in Relation zueinander zu setzen. Bei der Verarbeitung des Satzes werden zwei verschiedene Prozesse angestoßen, die Ressourcen des Arbeitsgedächtnisses verbrauchen:⁶ die *Memory Cost*, mit der der Aufwand beschrieben wird, Informationen aus einer unvollständigen Satzstruktur aufzubewahren, und die *Integration Cost*, unter der der Aufwand verstanden wird, der zur Integration neuer Wörter in eine Struktur verwendet werden muss (Gibson 1998: 8). Beide Prozesse schöpfen aus dem gleichen Pool von kognitiven Ressourcen. Das bedeutet, dass beim Verbrauch von viel *Memory Cost* weniger kognitive Ressourcen für die *Integration Cost* zur Verfü-

⁴ Die Abkürzungen sind wie folgt zu entschlüsseln: S entspricht dem Subjekt, V dem Verb und O dem Objekt.

⁵ Sprachen wie das Isländische (vgl. z. B. Höskuldur 2007: 21f.) oder Afrikaans (vgl. z. B. Donaldson 1993: 362ff.) zeigen allerdings, dass es sich hierbei nur um eine Tendenz handelt, da beide auf unterschiedliche Art von diesem Prinzip abweichen können.

⁶ Die Kapazität des Arbeitsgedächtnisses ist generell begrenzt, wie beispielsweise Untersuchungen von Miller (1956) und Cowan (2010). für allgemeine Gedächtnisleistungen zeigen konnten.

gung stehen und umgekehrt (vgl. Gibson 1998: 9). Das wird für das Satzverständnis relevant, da hier beide Prozesse angesprochen werden.

Das Lesen eines Satzes erfolgt inkrementell. Jedes Wort, das gelesen wird, erzeugt im Sinne der Informationsdichte Erwartungen an die nachfolgenden Wörter. Das geschieht sowohl auf der Ebene der Lexik als auch auf der der Syntax. Das Wort *lesen* kreiert beispielsweise die Erwartung, dass *Buch* oder *Zeitschrift* im weiteren Verlauf des Satzes auftreten. Dadurch wird auf der lexikalischen Ebene eine vorwärtsgerichtete Erwartung an den Inhalt des Satzes hervorgerufen. Auf syntaktischer Ebene wird nach *lesen* ein Akkusativobjekt erwartet. Jedes Wort schafft also aus sich heraus Erwartungen an das, was im Satz noch kommen muss. Zugleich muss das Wort selbst verstanden und in die Erwartungen integriert werden. Bei der Integration der Wörter und der Entwicklung der Erwartungen werden Ressourcen verbraucht, was unter der Integration Cost verstanden wird.

Levy (2008) setzt diese Idee in Verbindung mit Surprisal. Das Verständnis von Informationen basiert auf strukturellen, lexikalischen, pragmatischen und diskursbasierten Quellen (vgl. Levy 2008: 1128). Da diese Quellen zum Verständnis kombiniert werden (vgl. Jurafsky 2003), entsteht ein Wettstreit von ähnlichen Analysen für einen Satz. Dieser wird dann relevant, wenn ein neues Wort in den bestehenden Kontext integriert werden soll, da verschiedene Möglichkeiten zur Fortführung des Satzes bestehen. In (2)⁷ ließe sich so zum Beispiel beschreiben, wie viele kognitive Ressourcen verbraucht werden, um den Anfang des untergeordneten Satzes zu verarbeiten. Den möglichen Fortführungen wird dabei eine Rangfolge zugeordnet, die sich nach Wahrscheinlichkeiten richtet. Der Verarbeitungsaufwand entspricht dann dem Surprisal eines Wortes. Das Surprisal ist dabei die Schnittstelle zwischen der linguistischen Repräsentation während des Verstehens des Satzes und den Verarbeitungsschwierigkeiten, die bei einem bestimmten Wort innerhalb eines Satzes zu finden sind (vgl. Levy 2008: 1128).

⁷ Das Beispiel lautete:

- 2a) *Klaus sagt Ursel, dass er nächste Woche den Rasen mähen möchte.*
- 2b) *Klaus sagt Ursel, er möchte nächste Woche den Rasen mähen.*

Sprachen bestehen normalerweise aus einem endlichen Set komplexer Satzstrukturen. Dabei wird ein Satz dann verstanden, wenn der Rezipient eine Rangliste von Präferenzen über dem Satz aus allen möglichen Strukturen der Sprache konstruiert. Die geschieht ebenfalls inkrementell: Von früh geäußerten Wortfolgen hängt die Bildung späterer Präferenzen ab, sodass ein Update des Sets der Strukturen und Erwartungen nach jedem Wort erfolgt (vgl. Levy 2008: 1129). Schwierigkeiten werden dadurch hervorgerufen, dass die Wahrscheinlichkeitsverteilung immer wieder erneuert werden muss. Diese Schwierigkeit wird dadurch quantifiziert, in welchem Grad die Verteilung aktualisiert werden muss, was wiederum durch die Entropie gemessen wird. Wenn also in einem Satz bereits eine spezielle lokale adverbiale Bestimmung aufgetreten ist, ist deren erneute Verwendung unwahrscheinlich bis unmöglich. Sie wird aus dem Set der Erwartungen gestrichen, sodass den anderen möglichen Ergänzungen des Satzes höhere Wahrscheinlichkeiten zugeschrieben werden. Umgekehrt findet bei einer unerwarteten Konstruktion eine Verringerung der Wahrscheinlichkeit an der Stelle statt. Denn der Rezipient bewahrt dabei das vollständige Set der verschiedenen, wahrscheinlichen, teilverarbeiteten Konstituenten aus dem bereits verarbeiteten Input im Gedächtnis und weist diesem eine mögliche Wahrscheinlichkeitsverteilung über die komplette Struktur zu (vgl. Levy 2008).

Levy (2008) entwickelt den Surprisalbegriff von Shannon (1948) insofern weiter, dass Surprisal als die Schwierigkeit betrachtet wird, eine alte Verteilung mit einer neuen zu ersetzen (vgl. Levy 2008: 1132). Vorhersagen über das nächste Wort werden implizit während der Verarbeitung des restlichen Inputs getroffen. Surprisal vereint die Präsentation von Information und Verhaltensäußerungen: Verschiedene Repräsentationen beeinflussen die Wahrscheinlichkeit des folgenden Wortes im unmittelbaren Kontext.

Dieser Verarbeitungsaufwand ist eher vorwärtsgerichtet und lässt sich sowohl auf die lexikalische Erwartbarkeit eines Wortes als auch auf die syntaktischen Strukturen beziehen. Laut Gibson (1998: 9f.) geschieht die schnellste Verarbeitung dann, wenn alle Bedingungen nur auf eine einzige mögliche Struktur hinweisen. Im Sinne von Shannon (1948) kann damit auch eine ver-

besserte Verarbeitung gemeint sein, wenn die Unsicherheit auf den folgenden Strukturen gering ist. Es wurde bereits erwähnt, dass Gibson (1998) nicht auf die Vorhersagen referiert, die von Shannon (1948) gemacht wurden und in den vorherigen Abschnitten durch den Bezug auf Levy (2008) dargestellt wurden. Gibson (1998: 10) bezieht sich aber auch darauf, dass die Integrationskosten aufgrund der parallelen Verarbeitung des Materials höher sind. Das bedeutet, dass nicht nur die wahrscheinlichste Fortführung des Satzes bei der Verarbeitung im Gedächtnis behalten wird, sondern auch unwahrscheinlichere Strukturen und Fortführungen. Dieses Merken ist besonders für das begrenzte menschliche Verständnis relevant, da der Mensch sowieso nur eine eingeschränkte Anzahl an Quellen im Gedächtnis behalten kann (vgl. u. a. Roark 2001; Henderson 2004).

Auf die Belastung des Gedächtnisses bezieht sich auch die Memory Cost (vgl. Gibson 1998: 8, 13). Sie nimmt einen schwächeren Bezug zu den lexikalischen Verarbeitungsschwierigkeiten als zu den grammatischen. Gibson (1998: 13) definiert sie so: „[T]here is a memory cost associated with remembering each category that is required to complete the current input string as a grammatical sentence.“ Verarbeitungsaufwand entsteht also dadurch, dass sich die Rezipierenden daran erinnern müssen, dass schon etwas im Satz aufgetreten ist, das eine andere Kategorie erforderlich macht. Die Präferenz kurzer Distanzen zwischen verknüpften Wörtern wird als *Dependency Locality* (Gibson et al. 2019: 8) bezeichnet und gilt sprachübergreifend. Durch kurze Distanzen wird die Belastung des Kurzzeitgedächtnisses gering gehalten. Größere Distanzen zwischen zusammengehörenden Materialien verursachen hingegen Verarbeitungsschwierigkeiten, die auch mit der Integration des Wortes in die bestehende Struktur zusammenhängt. Gibson (1998: 11) führt das darauf zurück, dass ein Wort selbst Erwartungen aktiviert, diese Aktivierung aber schwächer wird, je mehr Wörter seit der Verarbeitung des betreffenden Wortes verarbeitet wurden. Auch die syntaktischen Erwartungen müssen zu den bereits bestehenden Wörtern passen und aktivieren ihrerseits weitere Erwartungen (vgl. Gibson 1998: 11). Da die Verarbeitungsressourcen aber begrenzt sind, verringert das wiederum die Aktivierung auf den Wörtern, die weiter

vorne im Satz stehen (vgl. Gibson 1998: 11). Je größer die Distanz wird, desto schwerer wird es, das Wort und dessen Erwartungen wieder zu aktivieren oder es überhaupt noch im Gedächtnis zu behalten.

Als Beispiel im Deutschen lässt sich dafür die Satzklammer anführen. Wenn eine linke Satzklammer aufgetreten ist, muss auch eine rechte folgen.⁸ Auch im Hinblick auf die Relativsätze entsteht eine derartige Verarbeitungsschwierigkeit, wenn beispielsweise ein Signalwort vorliegt, das einen Relativsatz erwartbar werden lässt (Kapitel 2.4.3). Je größer aber die Distanz zwischen dem zusammengehörigen sprachlichen Material ist, hier zwischen dem Signalwort und dem Relativsatzbeginn, desto wahrscheinlicher wird es, dass Informationsverlust durch Vergessen der anfänglichen offenen Stelle entsteht.

Die Dependency Locality erklärt zudem sowohl harmonische Wortreihenfolgen als auch das Gesetz der wachsenden Glieder (vgl. Behaghel 1932: 139f.), also die Beobachtung, dass kürzere Satzglieder vor längeren zu finden sind. Auch wird bei einem ähnlichen Informationsgehalt bestimmter Wörter Nachbarschaft bevorzugt, sodass die Dependency Locality zum Spezialfall der *Information Locality* wird.

Die Surprisalthorie macht letztlich unterschiedliche Vorhersagen zur Lokalität. Je mehr Abhängigkeiten bereits bekannt sind, desto einfacher wird die Verarbeitung des folgenden Materials, da es die Zahl der möglichen Zustände Stück für Stück reduziert (vgl. Levy 2008: 1139ff.). Sobald zum Beispiel klar ist, dass es sich in einem Satz um eine Verbletztkonstruktion handelt, werden die Erwartungen entsprechend angepasst (vgl. Levy 2008: 1146).⁹ Dieser

⁸ Die Satzklammern werden ausführlich in Kapitel 2.2.1 beschrieben.

⁹ Ein empirisches Problem der Surprisalthorie bieten englische Subjektiv- beziehungsweise Objektrelativsätze. Objektrelativsätze sind schwerer zu verarbeiten als Subjektivrelativsätze (vgl. Levy 2008: 1163f.). Wo und warum diese Probleme auftreten, wird von den verschiedenen Theorien unterschiedlich erklärt. Laut der Lokalitätstheorie ist die Distanz zwischen der Auslagerung, der Lücke und dem regierenden Verb zu groß, wodurch die Schwierigkeit auf dem Verb liegt, weil dort die Integration vollbracht werden muss. Die Surprisalthorie sagt aber voraus, dass die Schwierigkeit nicht dort auftreten kann, weil das Relativpronomen bereits den Relativsatz kennzeichnet und dadurch die Struktur vorgegeben wird. Das sollte zu einem langsameren Lesen des Prädikats in Subjektivrelativsätzen im Vergleich zu den Objektrelativsätzen führen. Zugleich sollte der Verarbeitungsaufwand bei Objektrelativsätzen wegen ihrer geringen Frequenz beim Subjekt am größten sein. In diesem Fall zeigt sich die Lokalitätstheorie jedoch als erfolgreicher als die Surprisalthorie. Mögliche Gründe dafür sind *Spillover-Effekte*. Die Schwierigkeit kann eigentlich auf dem Subjekt liegen, wird dort aber noch nicht registriert, weshalb sich der Effekt erst verspätet zeigt. Allerdings wird

Effekt wird von den Surprisalwerten aber nicht gezeigt, was an der Theorie selbst liegt: Sie wird auf Wörtern berechnet, weshalb seltenere Wörter per se ein höheres Surprisal besitzen als hochfrequente. Es besteht also ein Einfluss der lexikalischen Vorhersagbarkeit. Dem Problem kann – auf Kosten des lexikalischen Surprisals – entgegengewirkt werden, wenn beispielsweise nur die POS-Tags und damit grammatische Merkmale untersucht werden. Die Lokalitätstheorie und der Memory Cost-Ansatz von Gibson (1998) sind deshalb besser zur Vorhersage bestimmter grammatischer Konstruktionen geeignet, weil sie mehr Faktoren betrachten können als das klassische Surprisal (vgl. Levy 2008: 1166).

Folgendes Fazit ist zu ziehen: Gibson (1998) und Levy (2008) beschreiben die zur Verarbeitung eines Satzes nötigen Gedächtnisressourcen, die sich in der Terminologie nach Gibson (1998) in die Integration und Memory Cost trennen. Der Integrationsaufwand lässt sich durch Surprisal beschreiben und kann als vorwärtsgerichtete Verarbeitungsstrategie angesehen werden, da ein starker Bezug zu den Erwartungen besteht, die von einzelnen Wörtern oder grammatischen Kategorien hervorgerufen werden. Die Surprisalwerte bieten dabei den Nachteil, dass sie stark von ihrer Berechnungsart abhängen und dadurch in vielen Fällen nicht alle Quellen einbeziehen können, die den menschlichen Parsern aber zur Verfügung stehen. Die Memory Cost hingegen kann zwar auch durch einen gesteigerten Integrationsaufwand beschrieben werden, allerdings entsteht dieser dadurch, dass die Distanz zwischen zusammengehörigem Material zu groß ist und daher das erste Element des zusammengehörenden Materials, vereinfacht ausgedrückt, vergessen wurde. Das ist beim Verständnis des Satzes nicht hilfreich und ist generell eher zu vermeiden. Es kann also gesagt werden, dass die Memory Cost rückwärtsgerichtet ist, weil sie den Blick und die Verarbeitung auf den Satzanfang lenkt.

Neben der Dependency Locality Theorie gibt es folgende andere Ansätze: *Wettbewerbs-* und *dynamische Modelle* skizzieren den Einfluss von nicht-kategorialen, gewichteten Einflussfaktoren linguistischer und extralinguistischer Natur (vgl. Levy 2008: 1141). Dabei stehen die Integration von multi-

das durch experimentelle Untersuchungen widerlegt (vgl. Levy 2008: 1165).

plen, nicht-kategorialen Beschränkungen als Art der syntaktischen Disambiguierung und die Attribution langer Verarbeitungszeiten in einem Wettstreit zueinander. Mit diesen Modellen ist Surprisal kompatibel, da Schwierigkeiten entstehen, wenn Ressourcen auf eine mit der Fortführung des Satzes inkompatible Art verteilt werden (vgl. Levy 2008: 1141).

Auch die *Tuning Hypothese* ist der Surprisaltheorie sehr nahe: Sie beschreibt ein serielles Auswahlmodell von syntaktischer Disambiguierung, wobei die häufigste strukturelle Variante ausgewählt wird (Levy 2008: 1142). Die *Pruning* und *Attention Shift Theorien* ähneln ebenso dem Surprisal: Verarbeitungsschwierigkeiten treten als Resultat einer Reihenfolge möglicher Strukturen vor und nach einem zu integrierenden Wort zutage. Beim Pruning werden unwahrscheinliche Varianten durch die Begrenzungen des Gedächtnisses ausgeschlossen, beim Attention Shift verursacht die Änderung der am besten bewerteten Alternativen Probleme. Es ist nämlich in diesem Ansatz und im Unterschied zur Surprisaltheorie unmöglich, die am besten bewertete Alternative zu bestimmen oder zu verwerfen, ohne dass zuvor schon ein Set aus möglichen Strukturen vorhanden war (vgl. Levy 2008: 1143).

Grundsätzlich gilt die Surprisaltheorie unabhängig von Verbposition und -identität. Auch das nachfolgende Material kann rückblickend die Auswahlmöglichkeiten einschränken und so zu einer Veränderung der Erwartungen führen (vgl. Levy 2008: 1149).

(4) Ich habe einem Helfer gedankt.

In Beispiel (4) kann *habe* so lange als lexikalisches Verb interpretiert werden, bis der Kasus des nachfolgenden Artikels verarbeitet wurde. Sobald das geschehen ist, kann *habe* nur noch als Auxiliar verstanden werden. Die Interpretation des Verbs wird rückblickend geändert. Daher gilt, je mehr Material bereits verarbeitet wurde, desto leichter fällt die Analyse des Folgenden, was auch Lesezeitexperimente aus dem Japanischen und Hindi gezeigt werden konnte (vgl. Levy 2008: 1151).

Es zeigt sich insgesamt, dass die Surprisaltheorie mit den meisten älteren Verarbeitungsmodellen kompatibel ist. Einzig die Lokalitätstheorie verursacht

Schwierigkeiten bei der Vorhersage darüber, wo und warum ein Problem im Satzverständnis auftritt. Alle anderen, skizzierten Modelle treffen Voraussagen über die Fortsetzung eines Satzes. Die Surprisaltheorie macht jedoch als einzige die verschiedenen Vorhersagen zuverlässig berechenbar. Beispiele für weitere Anwendungen der Informationstheorie nach Shannon (1948) werden im Folgenden gegeben.

2.1.3. Entropy Rate Constancy

Die *Entropy Rate Constancy Hypothesis (ERC)* von Genzel und Charniak (2002; 2003) geht davon aus, dass Sätze als natürliche, eigenständige Einheiten zu betrachten sind, für deren Interpretation der Kontext bedeutend ist. Denn ohne Kontext erhöht sich die Entropie des einzelnen Satzes mit der Satznummer. In ihrem jeweiligen Kontext betrachtet haben Sätze aber etwa die gleiche Entropie (vgl. Genzel und Charniak 2003: 65). Je später im Text ein Satz auftritt, desto mehr Information enthält er, wobei auch Paragraphengrenzen überschritten werden können und Topikwechsel eine Informationserhöhung bedingen.

Um diese Hypothese zu überprüfen, wird der Entropiewert jedes Wortes mithilfe eines Trigram-Language-Modells in einem Korpus aus Zeitungstexten und fiktiver Literatur berechnet (vgl. Genzel und Charniak 2003: 66f.). Die Entropie erhöht sich mit der Satzsumme. Trotz Topikwechseln innerhalb eines Kapitels und des generellen Vorhandenseins langer Kapitel ähneln sich die Ergebnisse (vgl. Genzel und Charniak 2003: 69f.). Die große Differenz zwischen dem ersten und letzten Satz eines Kapitels könne zudem nicht durch Zufall erklärt werden. Die Ergebnisse sind auch auf Russisch und Spanisch übertragbar, wobei zu berücksichtigen ist, dass es sich bei den untersuchten Texten um Übersetzungen handelt (vgl. Genzel und Charniak 2003: 69).

Der Einfluss des Genres auf den beobachteten Effekt ist insgesamt allerdings geringer als bei vorherigen Untersuchungen, was darauf zurückgeführt wird, dass Business-Schreiben und Nachrichten generell strukturierter und informativer sein dürften und daher den Leser inhaltlich langsamer an die Thematik heranführen. Ihre höhere Strukturiertheit und das Ziel, Infor-

mationen schnell verständlich zu übermitteln, kann sich auf eine stärkere Korrelation zwischen der Entropie und der Satzzahl auswirken (vgl. Genzel und Charniak 2003: 69). Zugleich wird ein gegenteiliger Effekt bei populären Magazinen, Werbung und Gedichten ermittelt, während akademische Texte einen höheren Korrelationskoeffizienten zwischen Satzzahl und Entropie aufweisen als nicht akademische. Innerhalb der akademischen Texte sind geisteswissenschaftliche und soziologische Texte zudem strukturierter als natur- und ingenieurwissenschaftliche Texte (vgl. Genzel und Charniak 2003: 4).¹⁰

Die Entropie Rate Constancy wurde auch psycholinguistisch von Keller (2004) untersucht. Auch er nimmt keinen Unterschied zwischen der Satzposition und dem Verarbeitungsaufwand an, wenn die Sätze innerhalb ihres Kontexts auftreten, und überprüft das mithilfe einer Lesezeituntersuchung in der Annahme, dass höhere Entropie mit längerer Lesezeit korreliert (vgl. Keller 2004: 2). Die Ergebnisse von Genzel und Charniak (2002; 2003) können bestätigt werden, allerdings findet Keller (2004: 3), dass der Entropy Rate Constancy Effekt eigentlich nur ein Effekt der Satzlänge ist. Zudem zeigt er mithilfe einer Eye-Tracking-Untersuchung, dass der Effekt kleiner ist, als von Genzel und Charniak (2002; 2003) behauptet wird (vgl. Keller 2004: 5). Dazu wurde zunächst die Entropie pro Wort und dann die gesamte Lesezeit für den Satz berechnet, um die Gesamtlesezeit im Hinblick auf die Satzlänge normalisieren zu können. Korrelationen zeigen sich zwischen der Satzposition, -länge und Entropie. Bei der individuellen Satzbetrachtung ist die Korrelation zwischen der Entropie und der Position jedoch eher schwach. Trotzdem bestätigt sich: Je höher die Entropie, desto länger brauchen Probanden, um den Satz zu lesen. Die Satzposition selbst beeinflusst die Lesezeit hingegen nicht (vgl. Keller 2004: 7). Der Verarbeitungsaufwand bleibt also über alle Sätze hinweg durchschnittlich gleich.

¹⁰ Untersucht wurden dabei auch nicht-lexikalische Fälle mithilfe der Penn-Tree-Bank. Die Satzlänge, die mit der mit Schwierigkeiten verbundenen Einbettungstiefe korreliert, und die Zahl der Tochterknoten innerhalb eines Satzes wurden normalisiert, um ihren Effekt ausschließen zu können. Erwartet wurde eine niedrigere Entropie bei Bäumen mit geringer Einbettungstiefe. Die Haupteffekte lagen dabei auf den NP-Knoten, die für die Effekte jedoch nicht allein verantwortlich gemacht werden können (vgl. Genzel und Charniak 2003: 72).

Die bisherigen Ergebnisse beziehen sich auf geschriebene Texte. Doch auch für gesprochene Dialoge wurde die Entropy Rate Constancy von Ward und Vega (2009) untersucht. Grundsätzlich sollte eine konstante Entropie im Sinne von Shannon (1948) der effizienteste Weg zur Kommunikation sein, da so der Verarbeitungsaufwand konstant gehalten wird. In einem medial mündlichen Diskurs sollte der Effekt größer sein als im medial schriftlichen, da die Aufnahmegeschwindigkeit in letzterem einfacher zu variieren ist. Zugleich könnte die Spontanität gesprochener Rede dem jedoch entgegenwirken, da sie gegen die implizite Planung der gleichmäßigen Informationsrate spricht (vgl. Ward und Vega 2009: 1).

Um dies zu überprüfen, werden Dialoge mit zwei Sprechern als Ganzes betrachtet und die Entropiewerte beider Sprecher summiert. So gelangt man zur gesamten Information, die von beiden Kommunikationspartnern verarbeitet werden muss. Es gilt die Annahme: „[I]n spoken dialog, there is a tendency for the rate of information conveyed, summed over the contributions of both participants, to be constant over time“ (Ward und Vega 2009: 2). Bei der Messung wird die Entropie relativ zur Zeit gesetzt, da die Möglichkeit besteht, die Aussprechdauer der Wörter zu variieren. Die Korpusuntersuchung¹¹ lieferte ähnliche Effekte wie die von Genzel und Charniak (2002; 2003) gezeigten, wenn der Kontext herausgerechnet wird: Je später ein Satz in einem Text vorkommt, desto höher ist die Entropie. Diese Zunahme erfolgt graduell. Daraus resultiert die Idee, dass die Verarbeitungskapazität durch gleichzeitiges Zuhören eingeschränkt ist (vgl. Ward und Vega 2009: 2). Der Effekt ist jedoch nur gering. Die Entropie über beide Sprecher ist nur 58 % höher als bei einzelnen Dialogen, obwohl man intuitiv eine 100 % höhere Verteilung erwarten würde. Es findet vielmehr ein Ausgleich zwischen den Kommunikationsteilnehmern statt. Überträgt einer eine hohe Informationsrate, nimmt sich der andere zu diesem Zeitpunkt eher zurück und vermittelt weniger Information (vgl. Ward und Vega 2009: 4). Innerhalb einer Aussage kann beispielsweise durch bestimmte lautliche Signale die Informationsdichte ausgeglichen werden.

¹¹ Das Switchboard Korpus wurde verwendet und daraus Trigramme mithilfe von SRILM (vgl. Stolcke 2002) berechnet. Out-of-Vocabulary-Wörter wurden die gleichen Wahrscheinlichkeiten zugewiesen wie Wörtern, die nur einmal in den Trainingsdaten vorkommen. Geteilt wurde das Korpus anhand von längeren Schweigeperioden.

2.1.4. Smooth Signal Hypothesis und Uniform Information Density Hypothesis

Dass Information gleichmäßig verteilt werden soll, wurde auch von Aylett und Turk (2004: *Smooth Signal Hypothesis*) und von Levy und Jaeger (2007: *Uniform Information Density Hypothesis*) gezeigt. Aylett und Turk (2004) beziehen sich auf spontane, mündliche Sprache. Deren Lautproduktion ist durch eine hohe inter- und intralokutionäre Variabilität gekennzeichnet (vgl. Aylett und Turk 2004: 32). Denn Sprecher zielen auf eine möglichst robuste, von Umgebungsgeräuschen unabhängige Kommunikation ab. Die gleichmäßige Informationsübermittlung in dieser Theorie wird durch die Variation der Silbendauer erzeugt. Je erwartbarer ein Laut ist, desto kürzer wird er ausgesprochen. Auch die Betonung von Silben hängt mit ihrer Erwartbarkeit im Kontext zusammen (vgl. van Son und Pols 2003). Es besteht also eine starke Beziehung zwischen der Silbendauer und der Vorhersagbarkeit durch lexikalische, syntaktische, semantische und pragmatische Faktoren, durch die sichergestellt wird, dass die entsprechende Silbe wiedererkannt wird. Der Zusammenhang ist auch daran zu erkennen, dass die erste Silbe in einem Wort häufig klarer und länger ausgesprochen wird, weil an dieser Stelle noch die meiste Unsicherheit über das Wort selbst herrscht (vgl. Aylett und Turk 2004: 33).

Um die Hypothese zu überprüfen, wurde ein Dialogkorpus mit annotierter Silbendauer verwendet. In einem ersten Schritt wurde die Dauer der Silben unter Normalisierung des phonemischen Inhalts und der Zahl der Segmente analysiert (vgl. Aylett und Turk 2004: 37f.). Die Vorhersagbarkeit wurde mithilfe von Wortfrequenz und eines Trigram-Wahrscheinlichkeitsmodells der Silben berechnet. Zusätzlich wurde der Einfluss von Betonung und Grenzstrukturen miteinbezogen, da auch Unterschiede zwischen reduzierten und Vollvokalen sowie zwischen unbetonten und betonten Silben für die Dauer relevant sein können. Aylett und Turk (2004: 44) bilanzieren, dass 20 % bis 65 % der veränderten Silbendauern in nicht phrasenfinalen monosilbischen Wörtern auf Redundanz zurückzuführen sind und bestätigen, dass eine inverse Beziehung zwischen der Dauer einer Silbe und ihrer Vorhersagbarkeit besteht.

Dass dieses Prinzip nicht nur auf phonetischer Ebene besteht, zeigen die Untersuchungen von Levy und Jaeger (2007). Sie stellen die These auf, dass Informationsdichte auf jeder kommunikativen Ebene relevant sei und Sprecher ihre Äußerungen so gestalteten, dass starke Schwankungen im Informationsprofil ausbleiben. Das wird einerseits durch die phonetische Kodierung (vgl. z. B. Aylett und Turk 2004) erreicht, andererseits aber auch dadurch, dass Stellungsfreiheiten, die von Sprachen geboten werden, ausgenutzt werden oder optionales Material weggelassen werden kann. Um dies für die syntaktische Gestaltung von Äußerungen zu beweisen, untersuchen Levy und Jaeger (2007) syntaktische Reduktionen, „a phenomenon in which speakers have the choice of either marking a phrase with an optional word, or leaving it unmarked“ (Levy und Jaeger 2007: 2). Ihr Forschungsgegenstand ist das optionale *that* in englischen Relativsätzen. Sie finden mit ihrer Korpusstudie, dass *that* eingefügt wird, wenn das Surprisal auf dem ersten Wort des Relativsatzes sonst zu hoch wäre und die angenommene Kanalkapazität dadurch überschritten und ein Informationsverlust eingeleitet würde (vgl. Levy und Jaeger 2007: 7f.). Somit fanden sie erste Evidenzen für das, was als Uniform Information Density Hypothesis (UID) bekannt ist.

Anwendung der Uniform Information Density Hypothesis Jaeger (2005; 2010; 2011) und Frank und Jaeger (2008) haben sich ausführlich mit dem Thema des optionalen *that* befasst und führen es als ein Beispiel an, um die optimale Strategie zur Sprachproduktion darzulegen. Dazu wird zum einen der Bezug zur *availability-based production* (vgl. Frank und Jaeger 2008) gezogen. Diese beschreibt eine Beschränkung der Äußerungsplanung durch Schwierigkeiten dabei, Informationen aus dem Gedächtnis zu holen und in das bereits bestehende oder vorgeplante Material der Äußerung zu integrieren (vgl. Frank und Jaeger 2008: 1). Im Gegensatz zu den vorher vorgestellten Arbeiten beziehen sich Frank und Jaeger (2008) also auf Sprecher- statt Hörerorientierung. Daraus resultieren zwei Herausforderungen bei der Sprachproduktion, nämlich eine effiziente Kommunikation zu bedienen und dabei trotz der eigenen Aufmerksamkeits- und Gedächtniseinschränkungen die eigene Produktion so optimal zu gestalten wie möglich. Dass Sprecher ein optionales

Relativpronomen fallen lassen, kann sowohl für gesprochenes als auch geschriebenes Englisch gezeigt werden (vgl. Jaeger 2011: 163). Das Phänomen lässt sich zudem in bestehende Verarbeitungsansätze und -präferenzen integrieren. Es ist nach Jaeger (2011: 165) sowohl mit *dependency processing accounts* kombinierbar, die davon ausgehen, dass Varianten bevorzugt werden, die kürzere Abhängigkeitsverhältnisse innehaben, als auch mit *alignmet accounts*, die den Zugang zu Referenten als Hauptfaktor für sprachliche Präferenzen ansehen und sich auf die konzeptionelle Erreichbarkeit und die Vorerwähtheit von Referenten stützen. Auch mit *availability accounts*, die mehr den Sprecher in den Blick nehmen und dafür plädieren, dass Material, das dem Sprecher schnell kognitiv zur Verfügung steht, früher im Satz erscheint, lässt sie sich vereinen (vgl. Jaeger 2011: 165). In diesem Bezug steht auch die inkrementelle Sprachproduktion: Früher Verfügbares kann früher geäußert werden. So gelten bekannte oder vorerwähnte Referenten, die oftmals auch als Pronomen dargeboten werden, als erreichbarer als neue Referenten, die noch eine längere Vollform verlangen (vgl. z. B. Tily und Piantadosi 2009).

Das optionale *that* in (Subjekt-)Relativsätzen sorgt dafür, dass die Information über den folgenden Relativsatz bei Vorhandensein des Relativums über zwei Wörter verteilt werden kann, wodurch sich der Informationsgehalt insgesamt reduziert (vgl. Jaeger 2011). Das Pronomen kündigt den kommenden Relativsatz an, wodurch das erste Wort des Relativsatzes nur noch seine lexikalische Information trägt. Fehlt das Pronomen hingegen, so muss das erste Wort des Relativsatzes auch noch die grammatische Information der vorliegenden Konstruktion tragen. Daher wird *that* dann platziert, wenn der Relativsatz unerwartet ist. Ist er erwartbar, wird es weggelassen.

Überprüft wurde diese Annahme mithilfe des British National Corpus und den darin vorkommenden finiten, restriktiven, adjazenten Relativsätzen (vgl. Jaeger 2011: 171ff.). Die Länge, Häufigkeit der Wortformen vor dem Relativsatz und des ersten Inhaltswortes im Relativsatz, sowie Ambiguitäten wurden als zusätzliche Faktoren in die Analyse einbezogen (vgl. Jaeger 2011: 169). Zur Schätzung des Informationsgehalts des Anfangswortes des Relativsatzes wurden Bigram-Sprachmodell verwendet (vgl. Jaeger 2011: 179f.).

Die Ergebnisse der logistischen Regression bestätigen die von der UID aufgestellten Erwartungen. Spillover-Effekte vom Kopfnomen konnten ebenso ausgeschlossen werden wie die Vermeidung von Ambiguitäten durch das Setzen des Relativpronomens. Zugleich beeinflusst die Länge des Relativsatzes die Eliminierung des Pronomens, jedoch sind die Ergebnisse in dieser Hinsicht nicht eindeutig. Auch wenn bei dieser Untersuchung ein schriftliches Korpus verwendet wurde, so zeigt Jaeger (2005: 4), dass Störungen des Redeflusses als Hinweis für das Vorkommen der Subjunktion dienen können. Zugleich deuten diese Störungen darauf hin, dass ein hoher Informationsgehalt vorliegt, der durch sie ein wenig abgeschwächt wird.

Frank und Jaeger (2008) zeigen auch auf morphologischer Ebene die Bedeutung einer gleichmäßigen Informationsverteilung, indem sie englische Vollformen mit ihren bedeutungsgleichen Klitika vergleichen. Ihre Hypothese ist, dass die Vollform dann verwendet ist, wenn das Auftreten von negiertem *to be* und *to have* im Kontext unerwartet ist (Frank und Jaeger 2008: 2). Bei ihrer Analyse des gesprochen sprachlichen Switchboard Korpus zeigt sich für beide Auxiliare der beschriebene Effekt: Liegt auf diesem Verb mit seiner Negation mehr Information, wird die klitische Form nicht verwendet (vgl. Frank und Jaeger 2008: 4). Das nachfolgende Wort hat dabei einen geringeren Einfluss als das vorangehende. Die Wahl zwischen Klitika und Vollformen beeinflusst ebenfalls die Gleichförmigkeit des Informationsprofils.

Während Jaeger (2005; 2010; 2011) und Frank und Jaeger (2008) also für die Sprecherorientierung bei syntaktischer Variation argumentieren, gibt es nur wenige Studien, die einen Beleg dafür finden, dass auch Rezipienten sensibel für Surprisal und gleichmäßige Informationsprofile sind. Sikos, Greenberg, Drenhaus und Crocker (2017) untersuchen vor diesem Hintergrund die Verarbeitung von prä- und postnominalerweiterten Nominalphrasen im Deutschen. Ihre Hypothese ist, dass Relativsätze eher geeignet seien, Informationen gleichmäßig zu verteilen als pränominale Erweiterungen und daher in weniger vorhersagbaren Kontexten präferiert werden (vgl. Sikos et al. 2017: 1). In ihrem self-paced-reading-Experiment fanden Sikos et al. (2017), dass pränominale Modifizierungen in offenen Kontexten das Verständnis eher als

postnominale Erweiterungen erleichtern und das auch mit den Vorhersagen der Uniform Information Density übereinstimmt. Bei vorhersagbaren Kontexten zeigt sich ein geringerer Effekt durch eine schnellere Lesezeit. Pränominal modifizierte, erwartbare Wörter enthalten weniger Information (vgl. Sikos et al. 2017: 4), sodass diese Form der Nominalphrasenattribuierung genutzt werden kann, um eine gleichmäßigere Informationsdichteverteilung zu erzeugen (vgl. Sikos et al. 2017: 6).

Neben diesen Untersuchungen gibt es auch Studien, die den Einfluss des *Common Grounds* (vgl. z. B. Krifka 2008) auf Informationsdichte untersuchen, weil auch extralinguistischer Kontext Einfluss auf die Vorhersagbarkeit ausübt. Der Kontext kann dabei unterschiedlich schnell adaptiert werden, um die Informationsdichte konstant zu halten (vgl. Doyle und Frank 2015: 1588).

Hochinformativ extralinguistische Events im Common Ground werden durch linguistisch weniger informative Nachrichten kompensiert. Je komplexer das Extralinguistische ist, desto einfacher sollte das Linguistische sein. Untersucht wurden dazu Twitterkommentare zu einem Baseballspiel (vgl. Doyle und Frank 2015: 1588). Die Hypothese ist, dass sich die linguistische Entropie in späteren Tweets erhöht, da der Common Ground durch das gemeinsam erlebte, nicht-linguistische Material vergrößert wird. Beispielsweise werden sich im Common Ground im späteren Verlauf mehr gemeinsame Referenten befinden und bestimmte Erwartungen an den Spielverlauf gerichtet werden. Da das Extralinguistische nicht mehr so viel Verarbeitungskapazität durch seine Bekanntheit verbraucht, können im Umkehrschluss die Tweets mehr Informationen enthalten (vgl. Doyle und Frank 2015: 1590). Der Informationsgehalt eines Events wird mithilfe der Zahl der Tweets zu dem Zeitpunkt ermittelt (vgl. Doyle und Frank 2015: 1591). Der Zusammenhang zwischen dem Einfluss des Common Grounds und der Informationsdichte kann gezeigt werden (vgl. Doyle und Frank 2015: 1594).

Die in diesem Unterkapitel dargestellten Studien und Theorien stammen aus den vergangenen zwanzig Jahren, im Fall der Dependency Locality Theorie von Gibson (1998) sogar aus dem letzten Jahrhundert, bilden aber trotzdem relevante Faktoren ab, die auch für die vorliegende Untersuchung wich-

tig sind, wie die Prämisse der gleichförmigen Verteilung von Information oder die verschiedenen kognitiven Ressourcen. Letztere wurden von einer neueren Studie aufgegriffen, die im Folgenden dargestellt wird.

2.1.5. Exkurs: Lossy-Context Surprisal und Language-dependent Structural Forgetting

Das *Lossy-Context Surprisal* (vgl. Futrell et al. 2021) verbindet die Ansätze von gedächtnis- und erwartungsbasierten Modellen, die auch für diese Arbeit relevant sind. Unter den gedächtnisbasierten Modellen wird vor allem der Aspekt der Memory Cost nach Gibson (1998) angeführt: Verarbeitungsschwierigkeiten entstehen durch Probleme mit dem Kurzzeitgedächtnis. Der Rückgriff auf bereits Verarbeitetes misslingt, weshalb auch neu im Text Vorkommendes nicht mehr gut verarbeitet werden kann. Die Verarbeitungsschwierigkeit wird größer, je mehr Material zwischen Zusammengehörendem steht.¹²

Der zweite Bestandteil des Modells nach Futrell et al. (2021), die erwartungsbasierten Modelle und Beschreibungen von Verarbeitung kombinieren, betrachtet, wie vorhersagbar ein Wort in seinem Kontext ist. Problematisch ist allerdings, dass erwartungsbasierte Modelle Verarbeitungsschwierigkeiten, die mit langen Abhängigkeiten zusammenhängen, nicht korrekt vorhersagen können (vgl. Futrell et al. 2021: 10). Hier liefern die gedächtnisbasierten Ansätze bessere Erklärungen.

Beim Lossy-Context Surprisal werden beide Ansätze zusammengebracht, indem die Vorhersagbarkeit eines Wortes in einem unvollständigen Kontext berechnet wird (vgl. Futrell et al. 2021: 3).¹³ Die Präzision, mit der eine Vorhersage bei einem unvollständigen Kontext getroffen wird, hängt unter anderem davon ab, ob der Kontext selbst vorhersagbar ist oder nicht:

[C]omprehenders will make less accurate predictions from less probable contexts, as compared with more probable contexts. Equivalently, comprehenders

¹² Damit liegt diese Theorie, wenn auch selten so bezeichnet, den Erklärungsansätzen zugrunde, die Extrapolation vor allem von Länge oder zu überbrückenden Distanzen abhängig machen (Kapitel 3).

¹³ Unter einem unvollständigen Kontext verstehen Futrell et al. (2021) den unmittelbaren Kontext vor dem relevanten Ausdruck beziehungsweise Wort, aus dem beliebige Wörter entfernt werden, als seien sie quasi vergessen worden.

in lossy-context surprisal theory are less able to use information from low-frequency contexts as compared to high frequency contexts when predicting following words. (vgl. Futrell et al. 2021: 14)

Je weniger vom Kontext gegeben ist, desto eher greifen Rezipierende auf die Erwartungen zurück, die auf bloße Häufigkeiten und daher auf Unigram-Surprisalwerte zurückzuführen sind (vgl. Futrell et al. 2021: 15). Wenn man zudem davon ausgeht, dass ein weiter entfernter Kontext eher vergessen wird und daher nicht mehr zur Verfügung steht, gewinnt der unmittelbare Kontext vor dem zu verstehenden Wort an Bedeutung, wodurch sich diese Art der Surprisalberechnung wiederum dem Bigram-Surprisal annähert, aber nicht mit ihm gleichzusetzen ist (vgl. Futrell et al. 2021: 15).

Futrell et al. (2021: 29) skizzieren das Language-dependent Structural Forgetting und die Information Locality. Hinter der Information Locality steht die Annahme, dass eine größere Distanz zwischen linguistischen Elementen die Vergessenswahrscheinlichkeit erhöht (vgl. Futrell et al. 2021: 29f.). Die von Gibson (1998) präsentierte Dependency Locality sehen Futrell et al. (2021: 35) als Unterart der Information Locality an, weil durch syntaktische Abhängigkeiten verbundene Elemente einander natürlicherweise vorhersagen und daher näher zusammenstehen sollten, da sie vermutlich zueinander ähnlich hohe Surprisalwerte besitzen. Wird diese Bedingung erfüllt, sollte letztlich auch die UID-Hypothese, die von Futrell et al. (2021) allerdings an der Stelle nicht angeführt wird, erfüllt sein, da so Höhen und Tiefen im Surprisalprofil vermieden werden, die aufgrund der grammatischen Vorkommensnotwendigkeit des Materials nicht ausgeglichen werden können. Je näher die Wörter zusammenstünden, umso vergleichbarer sei der Informationsgehalt der Wörter (vgl. Futrell et al. 2021: 36). Stehen diese Wörter weiter entfernt, treten die Dependency Locality Effekte auf, weil das Wort, das zur Vorhersage des nun weiter entfernt stehenden Wortes hilfreich gewesen wäre, vergessen wurde (vgl. Futrell et al. 2021: 37). Fraglich ist an dieser Stelle allerdings, ob die Dependency Locality Effekte auch durch Gewöhnungseffekte und die verwendeten Language Models zustande gekommen sein können, zumal die Studienlage für dieses Phänomen keine klaren und vor allem keine sprachübergreifenden Effekte aufweist (vgl. Futrell et al. 2021: 37ff.).

Auch wenn das Lossy-Context Surprisal für diese Untersuchung nicht als Berechnungsmethode angewendet wird, ist die Untersuchung von Futrell et al. (2021) vor allem wegen der Hintergründe zum Language-dependent Structural Forgetting erwähnenswert. Darunter werden Gewöhnungseffekte an die Strukturen verschiedener Sprachen verstanden, die bei Futrell et al. (2021) dazu verwendet werden, Unterschiede zwischen den Sprachen zu erläutern. Beispielsweise erklärt die Theorie, warum Partikelverben im Englischen, wo man an kurze Abhängigkeiten gewöhnt ist, zu Verarbeitungsproblemen führen, wenn zu viele Konstituenten oder eine zu lange Konstituente die Verbbestandteile trennen, was im Deutschen aber beispielsweise nicht der Fall ist, weil die Sprechenden dieser Sprache an lange Mittelfelder und die Verteilung des Verbs auf zwei Positionen gewöhnt sind. Frank, Trompenaars und Vasishth (2016) konnten in einer experimentellen Studie zeigen, dass ihre Testpersonen die strukturellen Gegebenheiten von Englisch und Niederländisch kannten und die Items so bewertet haben, dass es die Präferenzen beider Sprachen widerspiegelte. Allerdings machen sie abgesehen von Alter und Studienort der Teilnehmenden keine Angaben zu ihnen. Es ist also denkbar, dass es sich bei der eher kleinen Gruppe um Personen handelt, die einen sehr reflektierten Zugang zur untersuchten Fremdsprache haben oder ein sehr hohes Sprachniveau aufweisen. Es bliebe zu überprüfen, ob sich die Ergebnisse für naive Sprechende ebenfalls in der Ausprägung zeigen. Zudem können Unterschiede in der Verarbeitung auch auf Orthografie und Zeichensetzung zurückzuführen sein (vgl. Frank et al. 2016: 565f.).

Es ist allerdings auch denkbar, dass dieser Ansatz genutzt werden kann, um Unterschiede zwischen Sprachstufen oder sprachlichen Vorbildern erklärbar zu machen, wenn die Gewöhnung in den Vordergrund gestellt wird. Menschen, die überwiegend mündlich kommunizieren und kaum schriftsprachliche Vorbilder haben, wenden diesen Stil beispielsweise auch dann an, wenn sie einen Text schreiben (vgl. z. B. Elspaß 2005). Die Ansätze bieten also eine interessante Basis, um die gegenläufigen Erklärungen, die von gedächtnis- und erwartungsbasierten Ansätzen getroffen werden, zu vereinen, und werden daher in den Diskussionskapiteln 6.5, 7.2 und 8.4 wieder aufgegriffen.

Kapitel 2.1 hat die theoretischen Konzepte von Information behandelt. Im Folgenden ändert sich die Perspektive und nimmt die linguistischen Konzepte, beginnend mit dem topologischen Feldermodell, und die in dieser Arbeit untersuchten Phänomene in den Blick.

2.2. Topologisches Feldermodell

Die hier verwendete und in Kapitel 2.3 ausführlicher beschriebene Definition von Auslagerung basiert auf der Einteilbarkeit des deutschen Satzes in verschiedene Felder, beschrieben durch das topologische Feldermodell (vgl. Höhle 1982; Wöllstein 2010; 2014). Dieses Modell soll im Folgenden dargestellt werden. Ziel des Kapitels ist folglich, einerseits die Besetzungsmöglichkeiten der einzelnen Felder zu beschreiben, da die Konstituenten, die ins Nachfeld gestellt werden, von der Grammatik auch in anderen Positionen lizenziert sind. Dabei ist der Anspruch dieses Kapitels nicht, das topologische Feldermodell theoretisch vollständig abzuhandeln. Stattdessen dient es dem Verständnis des Phänomens der Auslagerung, das gerade für die Nominal- und Präpositionalphrasen davon ausgeht, dass diese aus dem Vor- oder Mittelfeld herausbewegt wurden. Daher wird der Blick auf die verschiedenen Besetzungsmöglichkeiten der Felder und Serialisierungen umso relevanter, weil es Aufschlüsse darüber gibt, von wo die betreffende Phrase ins Nachfeld gelangt sein könnte. Da die Wortstellung innerhalb der Felder und in begrenztem Maße auch innerhalb des Modells jedoch frei ist, wird diese Frage hier nicht abschließend zu klären sein. Auch beim Nachfeld als Ort des Interesses wird in diesem Kapitel lediglich auf mögliche Unterteilungen und Besetzungsmöglichkeiten eingegangen (vgl. z. B. Altmann 1981; Zifonun et al. 1997). Gründe für die Platzierung von Elementen im Nachfeld werden in Kapitel 3 besprochen.

Des Weiteren wird der Tatsache Rechnung getragen, dass Höhle (1982) und Wöllstein (2010; 2014) das topologische Feldermodell nicht erfunden, aber besser nutzbar gemacht haben. Daher wird auch ein Überblick über frühere Arbeiten gegeben, der aber keinen Anspruch auf Vollständigkeit erhebt.

Den Satz in verschiedene Positionen zu unterteilen, versuchen auch die früheren Arbeiten, jedoch variieren vor allem deren Beschreibungen und Besetzungsmöglichkeiten des Rahmens so stark, dass sie zum Teil schwer zu verallgemeinern sind, was in Kapitel 2.2.1 wiederholt exemplarisch aufgegriffen wird. Diesen Nachteil gleicht das topologische Feldermodell nach Höhle (1982) und Wöllstein (2010; 2014) vollständig aus. Es trägt vor allem der Verbalklammerstruktur als Hauptmerkmal des Deutschen Rechnung. Diese Struktur ist immer gegeben, die Positionen um sie herum sind fest und existieren auch dann, wenn sie nicht lexikalisch besetzt sind (vgl. Reis 1997). Das setzt natürlich einen definierbaren, einheitlichen Satzbegriff voraus. Ich schließe mich dem Verständnis von Wöllstein (2010: 3) an, die verbhaltige Konstruktionen als Sätze definiert. Ebenfalls von Wöllstein (2010; 2014) übernehme ich die Nomenklatur der Felder.

Wöllstein (2014: 145) teilt den Satz in folgende Felder auf: *Vorvorfeld* (VVF), *Vorfeld* (VF), *linke Satzklammer* (LSK), *Mittelfeld* (MF), *rechte Satzklammer* (RSK) und *Nachfeld* (NF). Im topologischen Feldermodell ist das Vorfeld das Feld vor der linken Satzklammer, das Mittelfeld steht mittig zwischen den Klammern und das Nachfeld ist als das Feld nach der rechten Satzklammer klassifiziert. Prototypisch werden diese Felder wie in Tabelle 2.1 beschrieben besetzt.

VVF	VF	LSK	MF	RSK	NF
Ko- und parordinierende Elemente, Linksversetzung	eine Konstituente, Relativpronomen und -partikeln	finites Verb, Subjunktionen	beliebig viele Konstituenten	Verbzusätze, finites, infinite Verben	Nebensätze, Phrasen

Tabelle 2.1.: Elemente im topologischen Feldermodell.

Durch die Besetzungsmöglichkeiten des Feldermodells lassen sich auch die verschiedenen Satzarten bestimmen (vgl. z. B. Brandt, Reis, Rosengren und Zimmermann 1992). Dabei wird das Vorfeld als erste Position des Satzes behandelt, weshalb wir im Deutschen auch von einer Verbzweitstellung spre-

chen, wenn „das finite Verb in der linken Klammer“ (Schierholz und Uzonyi 2022: 948) auftritt und das Vorfeld besetzt ist (V2-Sätze, Beispiel 2.3a). Ist letzteres nicht der Fall, handelt es sich um einen Verberstsatz (V1-Satz). Mit Verberstsätzen werden in der Regel Entscheidungsfragesätze (2.4a) und Imperativsätze (2.4b) realisiert. Sätze, in denen das finite Verb an letzter Stelle steht, werden Verbletztsätze (VL-Sätze) genannt (vgl. Pittner und Berman 2021: 88f.). W-Fragen (2.3b) und Aussagesätze sind klassische Beispiele für Verbzweitsätze, wobei Eigenständigkeit kein dezidiertes Merkmal für Verbzweitsätze ist (2.2). Die Abhängigkeit von einem übergeordneten Satz zeichnet aber die meisten Verbletztsätze aus. Eine Ausnahme bildet Beispiel 2.5, das zugleich für einen infiniten Satz steht. Gerade der zuletzt erwähnte Beispielsatz zeigt jedoch, dass Verberst-, Verbzweit- und Verbletztsätze nicht vollständig mit ihren Funktionen korrelieren. Die Funktion des Satzes in Beispiel 2.5 ist die eines Befehls, die Form aber nicht die eines Verberstsatzes, sondern eines Verbletztsatzes. Die Hauptleistung des topologischen Feldermodells ist daher eine Form- und keine Funktionsbeschreibung.

VF	LSK	MF	RSK	NF
Klaus ...er	sagt habe	den Film schon	gesehen.	...

Tabelle 2.2.: Beispiel für einen abhängigen V2-Satz (fett).

Das Beispiel in Tabelle 2.3 illustriert neben den Besetzungsmöglichkeiten des topologischen Feldermodells in einem komplexen Aussagesatz auch die eines Nebensatzes in der zweiten Zeile. Der Nebensatz ist fettgedruckt.

VVF	VF	LSK	MF	RSK	NF
a) Aber	gestern	hat weil	Ursel einen Kuchen Klaus Geburtstag	gebacken hatte.	Nebensatz
b)	Wo	hast	du das Rührgerät	hingestellt?	

Tabelle 2.3.: Beispiel eines komplexen V2-Satzes mit abhängigem adverbialen Nebensatz (a) und Beispiel eines W-Fragesatzes (b).

In diesem Beispiel (2.3a) steht ein Konnektor im Vorvorfeld. Die Position kann aber auch von Appellativa oder Linksversetzungen gefüllt werden (vgl. z. B. Altmann 1981: 47ff.). Eine einzige Konstituente, in diesem Fall eine Temporaladverbiale, steht im Vorfeld. Die linke Satzklammer wird in der ersten Zeile von einem finiten Verb und im Nebensatz in der zweiten Zeile von einer Subjunktion besetzt. Das Mittelfeld beherbergt jeweils zwei Konstituenten. In der rechten Satzklammer im Hauptsatz (erste Zeile) ist der infinite Teil des Prädikats zu finden und im Nebensatz stehen sowohl die finiten als auch die infiniten Teile des Prädikats in der rechten Klammer. Im Nachfeld des Hauptsatzes befindet sich der kausale Nebensatz, dessen Nachfeld unbesetzt bleibt.

VF	LSK	MF	RSK	NF
a)	Hast	du meine Schlüssel	gesehen?	
b)	Gib	mir mal die Fernbedienung.		

Tabelle 2.4.: Beispiele für einen V1-Entscheidungsfragesatz (a) und für einen V1-Imperativsatz (b).

Das Satzverständnis nach Wöllstein (2010: 3) bedingt auch, dass Konstruktionen ohne finites Verb zu Sätzen zu zählen. In Beispiel 2.5 liegen zwei derartige Sätze vor. In 2.5a ist sogar nur die rechte Satzklammer besetzt. Sie tragen die Form von Verbletztsätzen, fungieren aber als Imperativsatz.

VF	LSK	MF	RSK	NF
a)			Aufgepasst!	
b)		Die Butter zügig zum Teig	hinzufügen!	

Tabelle 2.5.: Beispiele für eigenständige VL-Imperativsätze.

Was die Beispiele 2.2, 2.3, 2.4 und 2.5 gemein haben, sind unbesetzte Felder. Die Charakteristik der einzelnen Satztypen verlangt zwar, dass zu deren Bestimmung immer mindestens eine Klammer besetzt sein muss, alle anderen Felder können aber lexikalisch unrealisiert bleiben.

Die nächsten Abschnitte befassen sich mit der detaillierten Betrachtung der verschiedenen Felder und beginnen mit den Satzklammern.

2.2.1. Satzklammern und Satzrahmen

Die Positionen, auf die sich im deutschen Satz das Verb verteilen lässt, werden als Satzklammern (vgl. Wöllstein 2014) oder als (Satz-)Rahmen bezeichnet (vgl. z. B. Presch 1977). In der Literatur werden Klammer und Rahmen oft synonym verwendet.¹⁴ Trotz der verschiedenen Bezeichnungen ist die Satzklammer die Position im Satz, über die die größte Einigkeit in der Literatur herrscht.

Bei Drach (1937: 17) wurde die linke Satzklammer als „Mitte“ bezeichnet. Sie bildete also das Zentrum des Satzes, um das sich Vor-, Mittel- und Nachfeld drehen. Drach (1937) vernachlässigt mit der Begrifflichkeit jedoch, dass rechts von der linken Satzklammer viel häufiger mehr Material steht als links von ihr, wodurch die Mitte entweder als sehr breit angesehen werden muss oder eben nicht mehr die Mitte bildet. Trotzdem verweist Drach (1937) dabei richtig auf die große Bedeutung der Klammer. In ihr werde nämlich vor allem die „Finitheitsinformationen“ des syntaktischen Prädikats realisiert (Zifonun et al. 1997: 1500). Unter dem syntaktischen Prädikat wird hier nur der verbale Kern des Satzes verstanden, sprich „die gruppe aus finitem verb und infinitiv, partizip, präfix usw.“ (Presch 1977: 80).

Presch (1977) greift zwar die verschiedenen Bezeichnungen der Satzklammer auf und kontrastiert auch Klammer und Felder elaborierter als Drach (1937), jedoch definiert er die Klammer vor allem durch die Unterscheidung zwischen der Distanz- und Kontaktstellung des Verbs. Er spricht davon, dass dann Distanzstellung vorliege, wenn beide Klammern mit verbalem Material gefüllt werden, und Kontaktstellung, wenn nur eine Klammer gefüllt sei (vgl. Presch 1977: 84). Die Klammern sind damit keine festen Positionen im Satz, wie anhand der Beispiele 2.2, 2.3 und 2.4 verdeutlicht wurde. Stattdessen sieht Presch (1977: 85) in seinen Beispielsätzen auch bei Voranstellung des infiniten Prädikatsteils nur eine Klammer besetzt, wie durch die runden Klammern in (5) angedeutet wird:

¹⁴ In den Fällen, in denen es zwischen den Bezeichnungen Abweichungen gibt, wird das hier explizit hervorgehoben.

- (5) (gesehen habe) ich ihn gestern schon, (gesprochen habe) ich ihn leider nicht. (Presch 1977: 85, Beispiel (d))

Presch (1977: 85) bemerkt allerdings selbst, dass dieses Fronting nur in bestimmten, kontrastiven Kontexten möglich ist. Hier offenbart sich eine der Schwachstellen dieses Ansatzes. Zwar legt Presch (1977) dar, dass vor dem finiten Prädikatsteil in Aussagesätzen nur ein Satzglied stehen kann, aber er zieht daraus nicht die Konsequenz, seine Satzrahmendefinition zu ändern. Das bedeutet, dass er theoretisch einen Satz wie in (6) so strukturieren würde, dass das Partizip in der linken Satzklammer stehen müsste, was aber dem Fintheitskriterium und dessen festgelegter Position bei Höhle (1982) und Wöllstein (2010; 2014) widersprechen würde. Zudem gilt Distanzstellung in seinem Modell nur für Verbzweit- und Verberstsätze, während Kontaktstellung mit Ausnahme des eben erwähnten Frontings bei Verbletztsätzen zu finden ist (vgl. Presch 1977: 81, 85).

- (6) Ihn [[gesehen]_{LSK?} hab]_{LSK?} ich gestern schon, ihn [[gesprochen]_{LSK?} hab]_{LSK?} ich aber leider nicht.

Die Position des finiten Prädikatsteils grenzt zwar die Satzart ab und gilt als festgelegt, aber das Vorfeld ist hier noch nicht klar als die Position vor dem finiten Verb, die nur von einer einzigen Konstituente gefüllt werden kann (vgl. Wöllstein 2014; Höhle 1982), definiert.¹⁵ Deshalb kann für Preschs (1977) das Modell zwar verallgemeinernd von einer Satzklammer gesprochen werden, die jedoch für jeden Satz anders aussehen kann, was eine allgemeine Beschreibung der Struktur daher erschwert. Dennoch ist dieser Ansatz generalisierbarer für syntaktische Analysen als diverse andere Ansätze.

Brinkmann (1971) klassifiziert die Satzklammer beispielsweise nach ihrer Funktion und in Abhängigkeit der Verbsemantik. Für eine allgemeine syntaktische Analyse müssten also auch hier zu viele Möglichkeiten für die Klammern vermutet werden, was die Anwendung solcher Modelle bei einer Fragestellung wie der vorliegenden kaum umsetzbar macht.

¹⁵ Fronting bereitet allerdings auch den Modellen von Wöllstein (2014) und Höhle (1982) Probleme. Lösungs- und Beschreibungsansätze finden sich zum Beispiel in Müller (2011).

Ein anderes, syntaktisches Problem bilden *Funktionsverbgefüge* (7). Auch bei ihnen sind die Bestandteile des Verbs getrennt, weshalb Presch (1977: 100) sie zu den sich ähnlich verhaltenden Partikelverben zählt. Das impliziert, dass der nominale Bestandteil der Funktionsverbgefüge, das Funktionsnomen, in der rechten Klammer steht. Durch dessen nicht-verbale Charakter wäre aber auch die Position am rechten Rand des Mittelfeldes denkbar.

(7) Ich [sage]_{LSK} dir [Bescheid]_{RSK?}.

Dass das Funktionsnomen in den Verbalkomplex wie in (8) eingeordnet werden kann, kontradiert seine Stellung am rechten Rand des Mittelfeldes. Bech (1955) und Askedal (1986a) beschreiben die gleiche Stellungsmöglichkeit aber auch für andere nicht-verbale Elemente (9).

(8) ...weil ich dir [hätte **Bescheid** sagen können]_{RSK}.

Das einzige Argument gegen die Zuordnung von Funktionsnomen zur rechten Satzklammer ist der nominale Charakter. Es ist kein Bestandteil eines analytischen Verbs, da keine grammatische Information darauf realisiert wird (vgl. Modrián-Horváth 2019) und da es auf jeden Fall ein Verb braucht, um in dieser Bedeutung verwendet werden zu können. Sein nicht-verbaler Charakter schließt es also eigentlich aus den Bestandteilen der rechten Satzklammer nach dem Verständnis von Wöllstein (2014: 146) aus. Ähnlich problematisch ist das Prädikativum (vgl. Presch 1977: 107f.). Wie das Funktionsnomen kann auch das nicht-verbale Prädikativum als eng zum verbalen Prädikatsteil betrachtet werden. Letzteres gilt als Argument für seine Einordnung an den rechten Rand des Mittelfeldes.

Die Diskussion um die Besetzungsmöglichkeiten der rechten Satzklammer mit nicht-verbalem Material (8) weist auf einen Schwachpunkt des Modells nach Wöllstein (2014) hin. Bevor auf die Besetzung der linken Satzklammer mit Subjunktionen eingegangen wird, ist deshalb eine kurze Beschäftigung mit der rechten Satzklammer nötig.

Trotz der vielfältigen Besetzungsmöglichkeiten des Nachfeldes (Kapitel 2.2.4) markiert die rechte Satzklammer sowohl das rechte Ende von Neben- als

auch von Hauptsätzen und damit, wie in der Einleitung angelehnt an Speyer (2015a) beschrieben, das Ende einer Verarbeitungseinheit. Für Ramers (2006: 104) gilt die rechte Satzklammer als „Träger der Kernbedeutung, Regens [...] für die Elemente in X und Y“.¹⁶ Laut Pafel (2009: 40) stehen in der Position „(fakultativ) beliebig viele Verbformen und dazu gehörende Verbzusätze, falls vorhanden“. Die Definition der rechten Satzklammer nach Wöllstein (2014: 146, 149) umfasst nur diese Elemente.

Das oben geschilderte Problem, dass die Sätze in den Beispielen (8) und (9) nicht ungrammatisch sind, weist auf die notwendige Erweiterung der Beschreibung der rechten Satzklammer hin. Gemein ist den Beispielsätzen, die ein nicht-verbales Element eindeutig in der rechten Satzklammer verorten, jedoch die Besetzung des rechten Rands der rechten Satzklammer mit verbalem Material (vgl. z. B. Askedal 1986b; Pafel 2009; Schallert 2014). Nicht-verbales Material kann also nur in der rechten Satzklammer vorkommen, wenn diese auch mit verbalem Material gefüllt ist. Das wiederum spricht für die Grenzziehung des Satzes durch die rechte Satzklammer, eröffnet aber die Frage, warum in (9) nicht drei Verbpositionen im topologischen Feldermodell existieren sollten, deren mittlere Position, hier fettgedruckt, in den meisten Fällen unbesetzt bliebe.

(9) Klaus sagt, dass Ursel ihn [hätte **danach** fragen können]_{RSK}.

Diese Frage lässt sich mit einem Verweis auf Ockhams Rasiermesser klären. Konstruktionen wie (9) sind nur möglich, wenn mindestens zwei Verbbestandteile in der rechten Klammer stehen und werden erst mit höherer Anzahl der Verbbestandteile wahrscheinlicher. Es handelt sich also keinesfalls um ein frequentes Phänomen und auch nie um eine obligatorische Wortfolge. Alle anderen Felder des Modells sind hingegen mindestens in manchen Konstruktionen obligatorisch. Ein Beispiel dafür bildet die verbindliche Stellung von *sodass*-Sätzen im Nachfeld. Die Einführung einer dritten ‚Klammer‘ wäre daher unökonomisch. Stattdessen wird für die Erklärung des Phänomens eine Unterteilung der rechten Satzklammer in ein *Ober-* und *Unterfeld* angenom-

¹⁶ Ramers (2006: 102) bezeichnet das Mittelfeld als X- und das Nachfeld als Y-Feld.

men (vgl. Bech 1955). Pafel (2009: 56) gibt das System wie folgt wider:

Bech gliedert das Schlussfeld in Unterfeld und fakultatives Oberfeld, wobei das Unter- auf das Oberfeld folgt. Im Unterfeld folgt auf ein Verb (wenn das Unterfeld aus mehr als einem Verb besteht) das Verb, von dem es abhängig ist und das seinen Status regiert. Im Oberfeld ist es gerade umgekehrt, da folgt das abhängige, statusregierte Verb auf das Verb, von dem es abhängig ist. [...] Doch kann nach Bech im Unterfeld auch nichtverbales Material (und dabei nicht nur ein Verbzusatz bzw. eine Verbpartikel) vorkommen.

Die rechte Satzklammer kann folglich in Felder unterteilt werden, die auch von Material besetzt werden können, das nach der Definition von Wöllstein (2014) eigentlich nicht dort stehen kann. Diese Unterteilung besteht auch, wenn die linke Satzklammer wie in Verbletztsätzen mit einer Subjunktion besetzt ist (2.2). Admoni (1970: 296) schreibt dazu: „[I]m Nebensatz wird der Rahmen anders gebildet: Hier werden die Konjunktionen und das Subjekt, das oft am Anfang oder an zweiter Stelle des Nebensatzes steht, von dem Prädikatsverb distanziert, das am Ende des Satzes erscheint“.¹⁷

Nach Admoni (1970) müssten im Nebensatz von (10) entweder drei Klammern vorliegen oder die klammerbildende Funktion eines Elements ignoriert werden, da gemäß dieser Definition sowohl die Subjunktion als auch das Subjekt das einleitende Element für die Klammer bilden können. Grundsätzlich ist die Position des Subjekts allerdings zu variabel, um von ihr ausgehend von einer verallgemeinerbaren Einteilung des Satzes zu sprechen. Die Subjunktion hingegen steht immer an einem festen Platz. Wird davon abgewichen, wird ein Satz ungrammatisch (11):

(11) *Ich weiß, mein Vater dass den Film gekauft hat.

Im topologischen Feldermodell bedingt die Subjunktion in der linken Satzklammer, dass sowohl der finite als auch der infinite Teil des Prädikats in

¹⁷ Von der klammerbildenden Funktion des Subjekts muss sich an dieser Stelle jedoch klar distanziert werden, denn die Position des Subjekts kann auch in mit Subjunktionen eingeleiteten Nebensätzen variiert werden, wie folgendes Beispiel (10) zeigt.

(10) Ich weiß, dass ihm_{Dativobjekt} das Buch_{Subjekt} gut gefallen hat.

der rechten Satzklammer stehen müssen (vgl. Wöllstein 2014: 151). Somit erfüllt die Subjunktion die für ? : 114 [sic!Persch.1977 notwendigen Kriterien der Klammer: „distanzstellung, die abgrenzung von gruppen von klammeröffnenden und klammerschließenden gliedern und stellungsregelmäßigkeiten dieser gruppen im satz.“ Sowohl die Abgrenzung als klammeröffnende Gruppe als auch die Stellungsregelmäßigkeit wird dadurch erfüllt, dass nur Subjunktionen für diese Art der Klammeröffnung genutzt werden können und diese in der Regel am ganz linken Rand des Nebensatzes zu finden sind. Das problematische Kriterium der Distanzstellung ist zumindest bei finiten Gliedsätzen erfüllt, da diese mindestens ein lexikalisch realisiertes Subjekt zwischen den Klammern erfordern. Eine Ausnahme davon bilden allerdings subjektlose Sätze, beispielsweise im unpersönlichen Passiv (12a). Im Fall von infiniten Nebensätzen kann jedoch auch hier Kontaktstellung vorliegen (12b).

- (12) a. Ursel fragte, [**ob**]_{LSK} dort [**getanzt wurde**]_{RSK}.
 b. Klaus geht in den Garten, [**um**]_{LSK} [**zu mähen**]_{RSK}.

Im Rahmen des topologischen Feldermodells ist aber auch die Beschreibung solcher Sätze kein Problem, da sie die Besetzung der Klammern durch bestimmte Elemente definiert, aber gleichzeitig leere Felder erlaubt.

Die Besetzungsmöglichkeiten der linken Klammer durch finite Verformen oder Subjunktionen werden von Höhle (1982) und Ramers (2006) aufgegriffen. Sie verwenden zwei Bezeichnungen für diese Klammer, die FINIT-Position, wenn die linke Satzklammer mit Finitheitsmerkmalen tragenden Elementen gefüllt wird, und die C-Position, die für Subjunktionen reserviert ist. Da beide Positionen nie gleichzeitig gefüllt werden können, liegen sie an derselben Stelle im topologischen Feldermodell.

Zusammenfassend ist also zu sagen, dass die Definition des Satzrahmens und der rahmenbildenden Elemente im Laufe der Zeit Veränderungen unterworfen war. Es ist einer der größten Verdienste des topologischen Feldermodells, eine einheitliche Definition des Satzrahmens zu schaffen, die praktisch ohne Ausnahmen auskommt. Solche Ausnahmen werden davon gebildet, dass die linke Satzklammer sowohl für finite Verbformen als auch für

Subjunktionen genutzt werden kann. Selbst diese Ausnahme kann aber durch die Übertragung der topologischen Analyse in die generative Grammatik behoben werden. Sabel (2000) beschreibt beispielsweise diese Position als Kopf-Position, deren Füllung mit der Integration des Satzes und der damit einhergehenden Verbbewegung zusammenhängt.

Allerdings muss auch ein Umweg zur Erklärung von Fällen genommen werden, die in der rechten Satzklammer nicht-verbales Material aufweisen. Das gelingt durch die Einteilung der rechten Satzklammer in Ober- und Unterfeld, wobei das Unterfeld permissiv für nicht-verbales Material ist. Relevant für die hier vorgenommene Definition von Satzklammern ist dabei jedoch Folgendes: Den rechten Rand einer lexikalisch besetzten rechten Satzklammer bildet stets verbales Material, wodurch der Beginn des Nachfeldes klar als Position rechts des letzten verbalen Bestandteils der rechten Satzklammer definiert wird. Der Beginn des Satzrahmens mit der linken Satzklammer ist unstrittig. Mit dieser Klammerstruktur lassen sich die restlichen Felder des Modells definieren. Über ihre Besetzungsmöglichkeiten und Einschränkungen geben die folgenden Kapitel Auskunft.

2.2.2. Vorfeld

Wöllstein (2014: 154) und Höhle (1982) definieren das Vorfeld als die Stelle vor der linken Satzklammer, die nur von einer einzigen, dafür aber beliebig komplexen Konstituente besetzt werden darf (13).¹⁸

(13) [Weil Ursel Klaus versprochen hat, an diesem Wochenende sein Lieblingsgericht zu kochen, das er sonst nur sehr selten bekommt, da es sehr lange köcheln muss]_{VF}, [steht]_{LSK} sie heute früh [auf]_{RSK}.

(13) zeigt ein Vorfeld, in dem ein mit einem Infinitivsatz kombinierter kausaler Nebensatz steht, der wiederum durch einen Relativsatz erweitert wird, in dem ein Kausalsatz steckt. Der Satz erstreckt sich über vier Ebenen im Feldermodell und belegt die Besetzungsmöglichkeit des Vorfeldes mit Nebensät-

¹⁸ Diese Definition gilt zumindest für das moderne Standarddeutsche. Über die Möglichkeiten, das Vorfeld mit zwei Konstituenten zu besetzen, schreibt beispielsweise Speyer (2008a).

zen. Während Nebensätze also im Vorfeld stehen können, dürfen dort weder „unbetonte Modalpartikel (*halt, eben, wohl* usw.), *sodass*-Sätze, *weil-/obschon-/obgleich*-Sätze im Verb-Zweit-Stellung“ (Ramers 2006: 108) auftreten.

Neben den Verboten gibt es auch Elemente, die präferiert im Vorfeld stehen. Die Artikel von Speyer (2008*b*; 2009) geben einen Überblick über dieses Thema, wobei Speyer (2009) vor allem darauf eingeht, unter welchen Bedingungen des Vorfeld-*es* an seine namensgebende Position gestellt wird. Die Besetzungsmöglichkeiten des Vorfelds werden nicht von syntaktischen Regeln vorgegeben. Stattdessen wird es diskursbedingt meistens von referentiellen Ausdrücken besetzt (vgl. Speyer 2008*b*: 269). Die Vorfeldausdrücke werden überdurchschnittlich häufig von Topiks, kontrastiven Elementen und Rahmenbildnern gebildet (vgl. Speyer 2008*b*: 269). Das Topik wird dabei als eine Entität definiert, die diskursalt ist und über die im Verlauf des Satzes weitere Informationen gegeben werden (vgl. Speyer 2008*b*: 276). Das erste Kriterium wird in Speyer (2009: 339) dahingehend abgeschwächt, dass auch eine diskursneue Entität im Vorfeld stehen kann, wenn sie makrostrukturell relevant ist, also wieder in anderen Sätzen aufgegriffen wird. Im Vorfeld können ebenfalls kontrastive Entitäten auftreten, die vorerwähnt sein müssen (vgl. Speyer 2008*b*: 276*f.*) und auf die im nachfolgenden Satz referiert wird (vgl. Speyer 2008*b*: 279). Das gilt auch für die Rahmenbildner. Darunter werden meistens temporale oder lokale Adverbiale verstanden, die die Situation, in der die Proposition wahr ist, essenziell einschränken (vgl. Speyer 2008*b*: 281).

Speyer (2008*b*) kann unter Einbeziehung der Optimalitätstheorie zeigen, dass die Verteilung folgendermaßen lautet: Rahmenbildner stehen am häufigsten im Vorfeld, gefolgt von Kontrasten, gefolgt von Topiks. Dieses Ranking wird auch dann nicht verletzt, wenn die verschiedenen Ausprägungen des Vorfeld-*es* in die Betrachtung miteinbezogen werden (vgl. Speyer 2009). Da das Vorfeld im deutschen Deklarativsatz gefüllt werden muss, um dessen Grammatikalität zu gewährleisten, gibt es neben den oben genannten Möglichkeiten, auch die Option, das Vorfeld mit dem Pronomen *es* zu füllen. Dieses kann als Vorfeld-*es* nur die Funktion der Verbzweitsatzbildung erfüllen, als expletives *es* bei leeren Subjektstellen fungieren oder als Korrelat-*es* einen

im Nachfeld stehenden Nebensatz ankündigen (vgl. z. B. Speyer 2009). Die Bewegung dieses Pronomens ins Vorfeld findet allerdings nur als „Last Resort“ (vgl. Speyer 2009: 351) statt, wenn keine der anderen in Speyer (2008b) benannten Kandidaten für die Vorfeldstellung in Frage kommen. Dies gilt prinzipiell auch im gesprochenen Register, wobei hier die Etablierung des *origos* größere Bedeutung gewinnt und mehr Vorfelder mit *dann* besetzt sind (vgl. Speyer 2010: 287ff). Auch das Relativum wird im Vorfeld statt in der linken Satzklammer angenommen (2.6).

VF	LSK	MF	RSK	NF
Ursel den	hat	Klaus einen Film sie beide	vorgeschlagen mögen.	Relativsatz

Tabelle 2.6.: Beispiel eines Relativsatzes (fettgedruckt) im topologischen Feldermodell.

Denn in der linken Satzklammer können „Köpfe aber keine Phrasen“ (Wöllstein 2014: 148) stehen. Mit seinen in Kapitel 2.4.2 beschriebenen Eigenschaften fällt das Relativum unter die Kategorie der Phrasen, obwohl es gleichzeitig die Funktion eines Relativsatzeinleiters erfüllt und den Relativsatz seinem Matrixsatz unterordnet, was es funktionell, aber nicht formal für die linke Satzklammer prädestiniert. Neben der aus der generativen Syntax stammenden Argumentation, zeigen auch Beispiele aus den Dialekten (vgl. z. B. Bayer 1984), dass bei Relativsätzen die linke Satzklammer durch ein weiteres, nebensatzeinleitendes und Subordination anzeigendes Element wie *dass* oder *wo* besetzt sein kann (vgl. auch Fleischer 2005: 175f.). Dieses Phänomen lässt sich in der Standardsprache allerdings nicht mehr nachweisen (vgl. Fleischer 2005: 184f.), kann aber dennoch als weiteres Argument für die Platzierung des Relativums im Vorfeld betrachtet werden.

2.2.3. Mittelfeld

Noch bei Drach (1937: 18) wurde das Feld nach der linken Satzklammer als Nachfeld bezeichnet und für den Ort gehalten, wo „Ergänzungen und Erläuterungen“ stehen. Es galt als „Eindrucksstelle (Sinnwort als Denkergebnis

oder Belehrungsmittel)“ (Drach 1937: 18).¹⁹ Neben Drach (1937) verwenden auch andere Autoren andere Namen für das Mittelfeld. Griesbach (1986: 47) bezeichnet es beispielsweise als „Satzfeld“. Höhle (1982) verwendet den Buchstaben *X* als Bezeichnung für die Position zwischen den Satzklammern. Unabhängig von seiner Bezeichnung stimmen die Charakteristika dieses Feldes aber miteinander überein.

So stellt Ramers (2006: 110) fest, dass im Mittelfeld „mehrere einfache oder komplexe Konstituenten von beliebiger Kategorie“ stehen können. Einschränkungen gibt es lediglich für Subjekt- und Objektsätze, die „nur marginal möglich sind“ (Ramers 2006: 110), sowie Attributsätze, die für gewöhnlich eher im Nachfeld zu finden sind. Wie schon der Name vermuten lässt, gilt auch das Vorfeld-*es* als nicht mittelfeldfähig. Auch Fragewörter und -phrasen sind nur in Echofragen in diesem Bereich möglich. Die Position von *sodass*-Sätzen ist umstritten.²⁰

Alle anderen Satzglieder können im Mittelfeld stehen. Die Variabilität der Wortfolge im Mittelfeld definiert das Deutsche als Sprache mit freier Wortstellung (vgl. Sabel und Mamoru 2005: 1). Arbeiten, welche die komplette Abfolge von Elementen im Mittelfeld betrachten, sind rar (vgl. z. B. Engel 1970; Lenerz 1977). Ein Großteil der Forschung konzentriert sich nur auf das Verhältnis einzelner Objekte zueinander. Hier wurde wohl am häufigsten die Abfolge von indirektem und direktem Objekt im modernen Standarddeutschen untersucht (vgl. Musan 2002; Frey und Pittner 1998; Frey 2004). Arbeiten über die Abfolgevarietäten im historischen Deutsch wurden beispielsweise von Speyer

¹⁹ Damit verweist er bereits auf die Bedeutung der Thema-Rhema-Gliederung (vgl. z. B. Lutz 1981; Krifka 2008), die für die Stellung von Elementen innerhalb des Mittelfeldes von großer Bedeutung ist.

²⁰ Sowohl die Frage nach diesen Sätzen als auch die nach der Position von Relativsätzen, löst unweigerlich die Frage nach dem Problem der Abgrenzung hervor. Wie im Abschnitt 2.2.1 zu sehen, ist die Satzklammer auch bei nicht-lexikalischer Realisierung trotzdem vorhanden. Ohne dieses Abgrenzungssignal ist es allerdings schwer zu bestimmen, wo das Mittelfeld endet und das Nachfeld anfängt. Dies gilt insbesondere für Elemente am rechten Rand des Mittelfeldes.

(14) *Sie haben ihren Koffer gestern schon, sodass sie früher losfahren können, gepackt.

An Beispiel (14) lässt sich erkennen, dass sich der *sodass*-Satz tatsächlich nur im Nachfeld befinden kann, da die Position am äußeren Rand des Mittelfeldes bei einer lexikalisch gefüllten rechten Satzklammer von ihm nicht eingenommen werden kann (vgl. Zifonun et al. 1997).

(2011), der feststellt, dass der Einfluss der Pragmatik auf die Dativ-Akkusativ-Abfolge erst im 16. Jahrhundert nachweisbar ist, und von Rauth (2020) vorgestellt, der die Objektfolge in Dialekten seit dem Althochdeutschen untersucht. Wenige Arbeiten gibt es hingegen zu den Stellungsmöglichkeiten von Adverbialen im Verhältnis zu Objekten (mit der Ausnahme von Frey und Pittner 1998). Im Folgenden wird oberflächlich auf die verschiedenen Wortfolgen von Nominal- und Präpositionalphrasen im Mittelfeld eingegangen.

Zwischen den mittelfeldfähigen Elementen herrscht ein Wettbewerb um die Abfolge. Die Mitstreiter dieses Wettbewerbs sind pragmatischer, semantischer und syntaktischer Natur. So findet Jacobs (1988: 19f.), dass Agentien vor anderen thematischen Rollen stehen, der Dativ, wenn er als Rezipient, Adressat, Benefaktiv oder belebtes Ziel verwendet wird, vor dem Patiens, Pronomen vor nominalen Nominalphrasen, definite vor indefiniten Phrasen und Elemente, die dem Hintergrund zugeordnet werden können vor denen, die im Fokus stehen. Jacobs' (1988) Modell zeigt eine Vermischung von pragmatischen, semantischen und syntaktischen Kategorien bei der Beschreibung der Objektreihenfolge. Die thematischen Rollen sind semantischen Kategorien zuzuordnen, Definitheit und die Fokus-Hintergrund-Gliederung hingegen pragmatischen. Der syntaktische Bezug ist durch die Wortart gegeben. Dass Pronomen vor Nomen realisiert werden, steht im Spannungsfeld von Syntax und Pragmatik. Pronomen können in der Regel nur genutzt werden, wenn sie auf bereits in den Diskurs eingeführte Referenten verweisen, wodurch sie Bekanntheit markieren (vgl. z. B. Gundel et al. 1993).

Während sich Jacobs (1988) auf den binären Vergleich verschiedener Elemente konzentriert, versucht sich Engel (1970) an einer umfassenden Beschreibung des Mittelfeldes. Er nimmt Nominativ-, Akkusativ-, Dativ- und Genitivergänzungen, sowie Präpositiv-, Situativ-, Direktiv-, Nominal- und Adjektivergänzungen an. Aus den letzten fünf Klassen kann nur ein Element gewählt werden, da sie als Ergänzungen zu den Prädikativen gehören. Ihre Position ist immer am rechten Rand des Mittelfeldes und läuft daher in geringerer Konkurrenz zu den Objekten. Engel (1970) integriert die Präpositionalobjekte aber nicht in seine Analyse, obwohl diese auch von den Verben

selegiert werden. Eine Begründung für ihren Ausschluss fehlt.

Unter Beachtung der Tendenz der verschiedenen Ergänzungen, an ähnlichen Stellen zu stehen, lässt sich folgende Reihenfolge festlegen: Pronominale Nominalphrasen (einschließlich *man*) stehen vor akkusativisches Reflexivpronomen. Diese stehen vor pronominalen Akkusativen einschließlich expletivem *es*, die wiederum pronominalen Dativergänzungen einschließlich reflexiven Dativen vorausgehen. Ihnen folgen definite Nominativergänzungen einschließlich Namen, Possessiv- oder Demonstrativpronomen, worauf schließlich indefinite Nominalphrasen und Indefinitpronomen folgen. Nach ihnen können definite Dativergänzungen stehen, die definiten Akkusativergänzungen vorausgehen. Nach diesen kommen indefinite Dativergänzungen, danach indefinite Akkusativergänzungen. An letzter Stelle stehen Genitivergänzungen, unabhängig davon, ob sie pronominal oder nominal, definit oder indefinit sind. Zusammenfassend ist zu sagen, dass Pronomen vor Nomen und definite vor indefiniten Objekten und Ergänzungen stehen.

Hoberg (1981) weist allerdings zurecht auf Abweichungsmöglichkeiten von dieser Reihenfolge aufgrund semantischer Charakteristika wie der Belebtheit von Objekten hin. In (15) steht ein indefiniter Akkusativ vor einem definiten Dativ; eine Abfolge, die allein deshalb zweifach gegen Abfolgevarianten verstößt. Zugleich besteht hier aber auch die von Hoberg (1981) geforderte Abfolge von belebten und unbelebten Objekten.²¹

- (15) Jemand hat heute ganz skrupellos [einen Welpen]_{Akkusativ, indefinit, belebt} [dem Tierheim]_{Dativ, definit, unbelebt} gegeben.

Das Mittelfeld kann beliebig viele Konstituenten enthalten. Zugleich haben nur sehr wenige Studien versucht, für jedes mögliche Element eine eigene Position im Mittelfeld zu erfassen, wobei diese Versuche oft an der Einfluss-

²¹ Nach Fortmann und Frey (1997: 145) und vor ihnen Haider (1992) besitzt das in (15) verwendete Verb *geben* allerdings zwei unmarkierte Abfolgen, weshalb der Satz unmarkiert wäre, was wiederum von Rauth (2020: 33f.) aufgrund des verwendeten Tests von Fortmann und Frey (1997) kritisiert wird: Statt zwei semantische Lesarten darzustellen, würden Fortmann und Frey (1997: 158ff.) lediglich die Bekanntheitsbedingung der Objekte verändern (vgl. Rauth 2020: 34f.). Die Definitheitsabfolge und die damit assoziierte Bekanntheitsabfolge sind somit immer noch in kontextlosen Beispielen ausschlaggebender als semantische Lesarten, wie Fortmann und Frey (1997: 160) vorschlagen.

nahme von semantischen, pragmatischen und syntaktischen Faktoren auf die Position von Elementen scheitern. Für diese Untersuchung ist das insofern relevant, als diese Untersuchungslage es beinahe unmöglich macht, die genaue ursprüngliche Mittelfeldposition einer Phrase, die nun im Nachfeld zu finden ist, zu determinieren. Selbst Aussagen über eine wahrscheinliche Position sind nur schwer möglich, da diese auf Prinzipien basieren, deren Allgemeingültigkeit ebenfalls in Frage zu stellen ist. Deshalb wird an dieser Stelle auf den Versuch, die Ursprungsposition der (eigenständigen) Phrasen zu definieren, verzichtet, wodurch Aussagen beispielsweise über die Distanz zwischen der Ursprungsposition einer ausgelagerten Phrase und dem Nachfeld unmöglich sind.

2.2.4. Nachfeld

Als Nachfeld oder Y-Feld bezeichnet man den Bereich hinter der rechten Satzklammer (vgl. z. B. Höhle 1982; Zifonun et al. 1997; Wöllstein 2014). Auch bei einer unbesetzten rechten Satzklammer sind Mittelfeld und Nachfeld immer voneinander getrennt (vgl. Reis 1997). Sichtbar wird die Grenze zum Nachfeld nach Zifonun et al. (1997) zum Beispiel durch die Stellung der Prädikative am rechten Rand des Mittelfeldes. Wenn aber weder Prädikative noch verbales Material vorhanden sind, können Mittel- und Nachfeld nicht sicher voneinander abgegrenzt werden.

Die Position hinter der rechten Satzklammer ist jedoch nicht ausschließlich dem Nachfeld vorbehalten. Zifonun et al. (1997: 1646) beschreiben zusätzlich das *rechte Außenfeld*, das nicht den rechten Rand des Nachfeldes begrenzt, sondern variabel im Bereich hinter der rechten Satzklammer auftreten kann, als Position für beispielsweise Interjektionen oder Anredeformen. Im rechten Außenfeld stehen auch Thematisierungsausdrücke. Diese werden durch einen kataphorischen oder katadeiktischen Ausdruck in Vorfeld oder Mittelfeld eingeführt, die dann im Nachfeld wieder aufgegriffen werden (16).

- (16) Dass **der**_{Kataphorikum} sich bei dem Fest so ins Zeug gelegt hat,
der Mann_{Thematisierungsausdruck}!

Besteht zudem ein Satz- oder Verbgruppenbezug, werden sie auch als „Nachträge“ der echten ‚Ausklammerung‘ (Zifonun et al. 1997: 1647) bezeichnet. Eines ihrer Kennzeichen ist die Einleitung mit *zwar*. Sie sind meistens Adverbiale mit konnektivem, modifizierendem oder graduierendem Charakter und werden als eigene Intonationseinheit realisiert. Für gewöhnlich sind sie auch grafisch abgegrenzt oder anderweitig hervorgehoben. Die Unterscheidung zwischen rechtem Außenfeld und Nachfeld wird also sowohl durch die syntaktische Integriertheit als auch durch die prosodische Realisierung der jeweils in diese Position verschobenen Einheit getroffen. Im Gegensatz zum rechten Außenfeld ist das Nachfeld syntaktisch und prosodisch integriert, aber auch hinsichtlich seiner Besetzungsmöglichkeiten eingeschränkt.

Allgemein kann über die Besetzungsmöglichkeiten des Nachfeldes gesagt werden, dass sowohl Phrasen als auch Sätze in dieser Position stehen können (vgl. z. B. Ramers 2006; Zifonun et al. 1997). Ramers (2006: 119) beschreibt das Nachfeld sogar als die klassische Position von Nebensätzen. Dabei sei es unabhängig, ob die Nebensatzart von der Valenz gefordert wird, eine freie Angabe bildet oder attributiven Charakter hat (vgl. Ramers 2006: 119).

Nachfeldfähigkeit wird unter anderem von der Größe des Elements im Sinne des Potenzials an struktureller Komplexität bestimmt (vgl. Zifonun et al. 1997: 1651). Dadurch wird beispielsweise auch die hohe Anzahl von Sätzen im Nachfeld erklärt. Komplementsätze sind dort häufiger als Adverbialsätze zu finden und auch Attributsätze stehen etwas seltener an dieser Position, wie in Kapitel 3.2 gezeigt wird. Denn die Positionierung hinter der rechten Satzklammer steht im Spannungsfeld zur Adjazenzbedingung, die sie in unmittelbarer Nähe zu ihrem Kopfelement verortet (vgl. z. B. Hawkins 1992).

Neben Sätzen sind auch Phrasen in beschränkterem Maße nachfeldfähig (Kapitel 3.1.1). Zu den Phrasen, die häufig im Nachfeld stehen, zählen Adjunktorphrasen, wenn sie einen Teil eines Vergleichsgefüges bilden (vgl. Zifonun et al. 1997: 1654f.). Abhängig von ihrem Verwendungszweck stehen auch Präpositionalphrasen eher im Nachfeld als andere Phrasen. So beschreiben

Zifonun et al. (1997: 1657f.), dass sie häufiger in Funktion einer Angabe als in der eines Komplements ausgelagert werden. Die Funktion der Phrase ist also wichtiger als beispielsweise die Art der Präposition. Diese nimmt selbst keinen Einfluss auf die Auslagerung.

Grundsätzlich können auch Nominalphrasen ins Nachfeld gestellt werden, vor allem, wenn sie kombiniert oder erweitert sind (vgl. Zifonun et al. 1997: 1659f.). Die Auslagerung von Nominalphrasen gilt dabei aber als typisch für den mündlichen Diskurs (vgl. z. B. Beneš 1979; Zahn 1991; Brooks 2006), wobei sie dort durch „mangelnde Planung der Satzstruktur, Reparatur oder den Versuch, das Rederecht länger zu behaupten“ (Zifonun et al. 1997: 1660) begünstigt wird. Bestimmte Kasus fördern Extraponierbarkeit hingegen nicht. Nach Ansicht von Zifonun et al. (1997: 1661) sind koordinative und appositive Phrasen durchaus nachfeldfähig, obwohl Adjektiv- und Adverbialphrasen generell selten im Nachfeld auftreten. Im Gegensatz zu Nebensätzen stehen phrasale Elemente wie Adverbien, Präpositionalphrasen sowie Vergleichselemente, die in der Regel durch *als* oder *wie* eingeleitet werden, und Nominalphrasen niederfrequent im Nachfeld. Die nicht-vorfeld- oder -mittelfeldfähigen Attribute und erweiternde Komplementeile können hinter der rechten Satzklammer auftreten, obwohl auch sie eher selten hier stehen (vgl. Zifonun et al. 1997). Elemente, die zumindest in der neuhochdeutschen Sprachperiode nicht ausgelagert werden können, sind Pronomen und „nominale oder adjektivische Teile des Prädikatsausdrucks im engeren Sinne“ (Zifonun et al. 1997: 1651) sowie Partikeln oder der Verbalkomplex. In der Regel sind diese angeschlossenen Phrasen von der gleichen Art wie ihre Bezugsphrasen und werden durch die Nachfeldstellung als Zusätze charakterisiert. Ihre Extraposition ist unabhängig von der Konzeption des Textes (vgl. Koch und Oesterreicher 2007).

Neben der strukturellen Komplexität von Elementen ist auch die Informationsstruktur nicht unbedeutend für die Besetzung der Position hinter der rechten Satzklammer. Das Nachfeld kann es zur Einführung des Diskurstopiks genutzt werden und so einen Wechsel anzeigen (vgl. Speyer 2015a: 498) oder einen Kontrastfokus beinhalten (vgl. Speyer 2015a: 499). Zudem ist es be-

deutend für die Verteilung der Information innerhalb einer kommunikativen Einheit und dient der Übermittlung von Hintergrund- und Vordergrundinformationen und ist somit ein sprachökonomisches Mittel (Zifonun et al. 1997: 1669). Wie in Kapitel 1 erwähnt, ist eine wichtige Aufgabe des Nachfelds die Informationsentflechtung. Sie tritt vor allem bei besonders komplexen Informationen in Kraft und soll die Verarbeitung der Gesamtinformation erleichtern.

Dies kann die bevorzugte Stellung von Nebensätzen hinter der rechten Satzklammer erklären, da sie die höchsten Anforderungen zur Verarbeitung stellen. Ähnliches gilt für erweiterte phrasale Einheiten, die einen Attributsatz zu sich nehmen (17).

- (17) Morgen wird im Fernsehen eine Doku laufen **über die Ruinen, die wir im Sommer besichtigt haben.**

Wird die Bezugseinheit zusammen mit dem Attributsatz ins Nachfeld gestellt, tritt das Prinzip der besseren Verarbeitbarkeit in Kraft, wobei zugleich das Adjazenzprinzip befolgt wird. Eine solche Wortstellung sei oft „kommunikativ motiviert“ (vgl. Zifonun et al. 1997: 1670). Ein weiterer Punkt, der Informationsentflechtung und strukturelle Komplexität kombiniert, ist der Befund, dass längere Sätze mit größerer Wahrscheinlichkeit im Nachfeld stehen. Dies ist jedoch auch abhängig von der Art der Satzklammer: Verbletztsätze mit einem vollständigen Verbalkomplex verhindern Auslagerung eher, wohingegen in der rechten Satzklammer stehende trennbare Verbpartikeln Auslagerung eher begünstigen.

Bisher wurde vor allem die Besetzung des Nachfeldes durch einzelne Elemente beschrieben. Zifonun et al. (1997) heben aber auch hervor, dass es prinzipiell möglich ist, das Nachfeld mehrfach zu besetzen. Dabei ist die Anzahl der ausgelagerten Konstituenten in der Regel auf nicht mehr als zwei Elemente beschränkt, die keinen zusammenhängenden Komplex bilden (vgl. Zifonun et al. 1997: 1662ff.). Die Kombination zweier Komplementsätze ist im Gegensatz zur gemeinsamen Nachfeldstellung zweier nicht von der Valenz geforderter Sätze ausgeschlossen. Dabei muss aber die eindeutige Zuordnung

der beiden nicht von der Valenz geforderten Sätze zum gleichen Matrixsatz gewährleistet sein. Bei der Kombination von Phrasen mit Sätzen stellt sich diese Schwierigkeit hingegen nicht. Zudem ist ihre Abfolge festgelegt; der Nebensatz folgt der Phrase (vgl. Zifonun et al. 1997: 1664).²² Auch hier gilt aber die Beschränkung, dass man zuordnen können muss, worauf sich der nachgestellte Satz bezieht. Da das bei Attributsätzen oft nicht automatisch gegeben ist, werden diese seltener kombiniert ins Nachfeld gestellt, wenn sie nicht auch von der im Nachfeld stehenden Phrase abhängen. Die letzte mögliche Kombination ist die von zwei Phrasen, die vor allem in oralen Diskursen als Nachtrag vorkommt (vgl. Zifonun et al. 1997: 1665).

2.3. Die Begriffe Auslagerung und Extraposition

Inhalt dieser Arbeit ist die Untersuchung der Besetzung des Nachfeldes durch Nominal- und Präpositionalphrasen sowie durch Relativsätze. Diese Nachfeldbesetzung wurde und wird von mir als Auslagerung beziehungsweise Extraposition bezeichnet, wobei ich die beiden Begriffe vollständig synonym benutze. Gemeint ist damit ausschließlich die Positionierung der erwähnten syntaktischen Einheiten im Nachfeld, wobei das Nachfeld in Anlehnung an Höhle (1982), Zifonun et al. (1997) und Wöllstein (2010; 2014) als das Feld hinter der rechten Satzklammer definiert wird.

Um die begriffliche Verwendung von Auslagerung und Extraposition nachvollziehen zu können, würde diese Definition vermutlich ausreichen. Beschäftigt man sich aber eingehender mit der Thematik, zeigt sich, dass meine Begriffsverwendung keinesfalls die einzige ist, die in der Literatur genutzt wird, um die Stellung von Elementen im Nachfeld zu beschreiben. Mit dem Phänomen der Nachfeldbesetzung befasst sich beispielsweise Altmann (1981).²³

²² Diese Abfolge ließe sich auch schon als weiterer Hinweis für das Zutreffen des Gesetzes der wachsenden Glieder nach Behaghel (1932: 139) betrachten. Überhaupt ist Komplexität neben der syntaktisch-semantischen Integriertheit einer der Hauptfaktoren, die über die Reihenfolge von zwei ausgelagerten Elementen im Nachfeld bestimmen. Es gilt, dass Komplementsätze vor Angaben, aber nicht vor Attributsätzen stehen und einfache Phrasen eher vor Sätzen (vgl. Zifonun et al. 1997: 1667).

²³ Genauer gesagt beschreibt er *Herausstellungen* nach links und nach rechts, wobei letzteres

Anachronistisch wird mit der Darstellung der von Altmann (1981) getroffenen Unterscheidung von Extraposition und Ausklammerung begonnen, da ich diese Begriffe synonym verwende. Für Altmann (1981: 65) werden ausschließlich Sätze extraponiert. Dabei ist es unabhängig, ob diese eigenständig oder attributiv sind (Altmann 1981: 66), wobei bei Letzteren ein Bezugselement im Vor- oder Mittelfeld des Matrixsatzes in Form einer Nominalphrase, Präpositionalphrase oder Proform vorliegen muss. Wenn dieses Bezugselement bei eigenständigen Gliedsätzen von einem resumptiven *es* gebildet wird, markiert dieses die Position, wo die nominale Realisierung des entsprechenden Gliedsatzes stehen würde (vgl. Altmann 1981: 66). Des Weiteren charakterisiert die Extraposition das Verbot von einleitenden Floskeln wie *und zwar* und allen anderen Zusätzen und das Fehlen einer Intonationspause. Vor allem aber unterscheidet sie sich von der Ausklammerung dadurch, dass die Extraposition nicht nachfeldfüllend ist, weil „daneben ganz normale Ausklammerung auftreten kann“ (Altmann 1981: 65).

Bei diesen Ausklammerungen handelt es sich um die felderfüllende Positionierung von Elementen, die ausschließlich von Phrasen gebildet werden, in der Position nach der rechten Satzklammer (Altmann 1981: 67), aber vor Rechtsversetzungen²⁴ und vor extraponierten Ausdrücken (vgl. Altmann 1981: 68). Darauf, was diese Differenzierung in felderfüllende und nichtfelderfüllende Bereiche hinter der rechten Satzklammer bedeutet, werde ich im Anschluss an die Beschreibung der Ausklammerung zurückkommen. Während es dialektal mehr Möglichkeiten gebe, dieses Phänomen zu nutzen, sei

hier von größerer Bedeutung ist. Dabei versucht er, verschiedene Arten von Herausstellungen anhand von linguistischen Kriterien zu definieren und eindeutig voneinander abzugrenzen. Wichtig ist aber, dass die Herausstellung nicht per se als Platzierung von Elementen ins Nachfeld anzusehen ist, sondern auch auf eine Bewegung ins Mittelfeld verweisen kann, solange die Richtung der Bewegung nach rechts oder links geht.

Es ist an der Stelle auch zu bemerken, dass er nicht der erste Autor ist, der das Thema des Satzendes aufgreift, und dass diese Thematik bereits in den fünfziger und sechziger Jahren Interesse hervorgerufen hat. Ich sehe an dieser Stelle trotzdem davon ab, deren - wenn man Altmann (1981) beim Wort nimmt - noch diffusere Klassifizierung von Auslagerung zu beschreiben, da sich der Begriff des Nachfelds in der von mir verwendeten Weise erst in den achtziger Jahren durch Höhle (1982) herausgebildet hat, wodurch die Vergleichbarkeit der verschiedenen Definitionen erschwert wird.

²⁴ Unter Rechtsversetzungen versteht Altmann (1981: 54) die Platzierung von Nominal- oder Präpositionalphrasen am rechten Satzrand von Sätzen, die eine „koreferente Proform enthalten, gewöhnlich ein Pronomen, aber auch andere Referenzmittel mit geringem deskriptivem Gehalt“. Diese Definition wird auch von Repp und Struckmeier (2020: 149) geteilt.

es im Standarddeutschen praktisch auf fakultative Satzglieder und höchstens noch Präpositionalobjekte beschränkt (vgl. Altmann 1981: 67). Allerdings dürfen diese in Form von Aufzählungen, Koordinationen, Kontrastierungen und „Koordinationsreduktion[en] ohne Konjunktionen“ auftreten (Altmann 1981: 69). Auch gibt es bei ihnen kein Resumptiv im Matrixsatz und sie dürfen ebenfalls nicht mit Floskeln und Zusätzen verknüpft werden (vgl. Altmann 1981: 68). Es werden auch noch intonatorische Differenzen zwischen den beiden Herausstellungsarten ausführlicher beschrieben, auf die hier jedoch nicht weiter eingegangen werden soll.

Altmann (1981: 68ff.) unterteilt die Position am Satzende also in ein Feld und eine unbestimmte andere ‚Aufbewahrungsform‘ für sprachliche Elemente, wodurch er glaubt, diese Satzposition klarer zu machen. Wenn man versucht, das in die Struktur des topologischen Feldermodells, wie es von Wöllstein (2014) beschrieben wird, zu übertragen, ergibt sich daraus entweder die Notwendigkeit, zwei verschiedene Felder oder Positionen hinter der rechten Satzklammer zu vermuten oder das Nachfeld in verschiedene Unterpositionen zu unterteilen, wie das auch mit der rechten Satzklammer getan wird (vgl. z. B. Bech 1955; Askedal 1986a;b; Pafel 2009).

Einen solchen unterteilenden Ansatz beschreiben Zifonun et al. (1997: 1649). Sie unterscheiden in Analogie zum Vorvorfeld zwischen dem Nachfeld und dem rechten Außenfeld, wobei im rechten Außenfeld „keine syntaktisch integrierten Bestandteile des betreffenden Satzes“ (Zifonun et al. 1997: 1646) stehen dürfen. Die Position ist also Gesprächspartikeln oder Vokativen vorbehalten. Im eigentlichen Nachfeld können hingegen sowohl phrasale als auch sententiale Konstituenten untergebracht werden. Daraus könnte man schließen, dass die Abfolge hinter der rechten Satzklammer in Analogie zum Vorfeld so aussieht, dass das rechte Außenfeld die rechte Grenze des Nachfeldes bildet. Das rechte Außenfeld kann aber auch mitten im Nachfeld stehen, wodurch das Nachfeld in ein enges und weites Nachfeld unterteilt werden kann (18). Unterscheiden lassen sich die Positionen vor allem durch die Intonation.

- (18) Hast du [**erzählt**]_{RSK} ([von der Sache,]_{enges Nachfeld} [Ursei,]_{rechtes Außenfeld} [wenn du schon dabei bist]_{weites Nachfeld})_{Nachfeld} ?

An diesem Beispiel (18) sollten neben der gefühlten Mündlichkeitsnähe, die sich allein durch die Besetzung des rechten Außenfeldes, bei dem es sich auch um eine einfache Parenthese handeln kann, mit einem Vokativ ergibt, auch auffallen, dass der gesamte Komplex hinter der rechten Satzklammer auch von Zifonun et al. (1997: 1649) als Nachfeld bezeichnet wird.

Das Nachfeld ist also als Position anzusehen, in der vor allem in der gesprochenen Sprache mehrere Elemente stehen können, die aber keine so komplexe und explizit differenzierte Unterteilung wie die von Altmann (1981) benötigt. Hilfreich ist die Unterteilung des Nachfeldes in den seltenen Fällen, in denen wirklich mehrere Elemente extraponiert werden, weil so die Reihenfolge der Konstituenten besser dargestellt werden kann.

Für die vorliegende Arbeit nicht von Bedeutung ist hingegen die von Altmann (1981) getroffene Unterscheidung zwischen nachfeldfüllenden Herausstellungen und solchen, die das nicht erfüllen sollen. Was bedeutet diese Unterscheidung bei der Platzierung von Elementen hinter der rechten Satzklammer? Es scheint so, als plädiere Altmann (1981: 65) für die Schaffung einer weiteren, neuartigen Position hinter der rechten Satzklammer, die aber weder einer mit Köpfen zu füllenden Klammer noch den mit Phrasen zu füllenden Feldern (vgl. Wöllstein 2014) entsprechen würde und sehr variabel ist. Denn Altmann (1981: 67) argumentiert selbst, dass auch zur felderfüllenden Ausklammerung noch zusätzliches Material treten kann. In der Einleitung (Kapitel 1) wurde beschrieben, warum die rechte Satzklammer auch als Ende eines Verarbeitungsabschnitts angesehen wird und dass rechts von ihr ein neuer Verarbeitungsabschnitt auftritt.²⁵ Auch wenn es wegen der Reihenfolge

²⁵ Die rechte Satzklammer ist auch insofern von Bedeutung für die Abgrenzung von Altmann (1981), weil er sie nicht für alle Herausstellungsarten, die er beschreibt, als Grenzlinie ansieht. Zuvor wurde schon die Rechtsversetzung als Platzierung einer Nominal- oder Präpositionalphrase, die auf eine links von der Phrase stehende Proform referiert, beschrieben (Altmann 1981: 54f.). Die Rechtsversetzung landet aber, wie aus den Beispielen hervorgeht, nicht zwangsläufig hinter der rechten Satzklammer. Das gilt auch für das freie Thema, ein „selbstständiger satzwertiger Ausdruck [...] z w i s c h e n den Sätzen“ (Altmann 1981: 56) mit einer engeren Anbindung an den vorangehenden Satz, Vokative und Appositionen, bei denen die Herausstellung auch darin besteht, dass sie rechts von einem zur Ambiguitätsvermeidung notwendigen Bezugselement stehen (Altmann 1981: 58). Auch sind diese Elemente von koreferenten Formen abhängig, was sie als eher mündlichkeitsnah definiert (vgl. z. B. Koch und Oesterreicher 2007; Ortman und Dipper 2019) und was auch in den in dieser Arbeit untersuchten Texten, wenn überhaupt, nur selten auftritt. Eine genauere Analyse ist daher an dieser Stelle nicht zielführend.

mehrerer Elemente im Nachfeld verlockend ist, dieses wieder in Unterfelder zu unterteilen, macht das in der Verarbeitung einen geringen Unterschied. Das gilt vor allem für die hier behandelten medial schriftlichen Texte, deren Rezeption lesend erfolgt, und in denen doppelte Nachfeldbesetzung nur niederfrequent vorkommt. Für die Rezeption und auch für die Berechnung von Surprisalwerten ist nur die sukzessive Abfolge der Wörter relevant. Ob diese im rechten Außenfeld oder im Nachfeld stehen, ist daher irrelevant.

Bedingt durch diese Verarbeitungsart ergibt sich auch, dass es für diese Arbeit nicht notwendig ist, zwischen Ausklammerung und Extraposition im Sinne von Altmann (1981) zu unterscheiden. Bedeutend ist, dass Elemente, die auch zwischen den Satzklammern auftreten können, hinter der rechten Satzklammer stehen, ohne dass der Satz ungrammatisch wird. Das ist sowohl bei den Nominal- und Präpositionalphrasen der Fall als auch bei den Relativsätzen so. Ihre Nachfeldstellung muss merkmalsgetrieben sein (Kapitel 3). Obwohl unter den generativen Grammatikern keine Einigkeit darüber herrscht, ob wirklich eine Bewegung nach rechts oder Basisgenerierung bei manchen Formen von Extraposition vorliegt (vgl. z. B. Inaba 2007; Repp und Struckmeier 2020), besteht aus der Sicht der Rezipierenden dennoch eine Bewegung über die rechte Satzklammer hinweg. Deshalb wird hier weiterhin synonym von der Extraposition beziehungsweise Ausklammerung beziehungsweise Nachfeldplatzierung von linguistischem Material gesprochen.

Die Bewegung über die rechte Satzklammer und die mit ihr verbundenen Herausforderungen, aber auch Möglichkeiten für das Verständnis des Satzes sind Gründe, warum das hiesige Verständnis von Extraposition schwer mit dem in Einklang zu bringen ist, was für das Englische über das Phänomen geschrieben wurde. Auch diese Sprache kennt Extraposition, versteht sie aber natürlich unabhängig von der rechten Satzklammer, da das topologische Feldermodell nur auf Sprachen zu übertragen ist, die die gleiche Klammerstruktur wie das Deutsche aufweisen. Das Englische besitzt zwar auch analytische Verbformen, doch diese stehen meistens adjazent zueinander, werden also auch gemeinsam verarbeitet und geben daher auch sofort die Valenz des Verbs preis, bevor dessen Aktanten verarbeitet werden. Daher wird in der

weiteren Untersuchung nur vereinzelt auf das Phänomen der Extraposition im Englischen zurückgegriffen werden, obwohl es gerade im Bereich der Relativsätze, mit denen sich das nächste Kapitel beschäftigt, eine recht breite Forschungsbasis gibt (vgl. z. B. Francis und Michaelis 2014; 2017; Levy, Fedorenko, Breen und Gibson 2012).

2.4. Der Relativsatz

Ein Fokus dieser Arbeit liegt auf der Position von Relativsätzen. Das folgende Kapitel widmet sich daher dieser Klasse von Nebensätzen, beschreibt die verschiedenen Arten und befasst sich kurz mit der Art des Relativums und ihrer Restriktivität. Auf die Position wird in Kapitel 3.2 eingegangen.²⁶

2.4.1. Relativsatzarten

Relativsätze gehören zur Gruppe der Nebensätze. Sie besitzen ein einleitendes Element, ein Relativum (vgl. Gallmann 2016: 1030), und weisen Verbletzstellung auf. Am häufigsten werden sie als Attributsätze verwendet (vgl. Eisenberg und Thieroff 2020: 293). Hentschel und Weydt (2013: 384) klassifizieren Relativsätze sogar als „Prototyp“ von Attributsätzen. *Attributive Relativsätze* sagen „etwas über ein Element im übergeordneten Satz aus“ (Gallmann 2016: 1029). Es muss also ein Bezugswort, ein Antezedens, vorhanden sein, das normalerweise von Nomen oder Pronomen gebildet wird (vgl. Gallmann 2016: 1030) und zusammen mit dem Relativsatz eine Konstituente darstellt (vgl. Fritsch 1990: 107), die jedoch durch die Positionsvariabilität von Relativsätzen diskontinuierlich sein kann. Erkennbar ist der Konstituentenstatus durch die Vorfeldfähigkeit von Antezedens und Relativsatz (19).

(19) [Das Restaurant, **das sie am liebsten mögen,**]_{Vorfeld} ist ein Grieche.

Der fettgedruckte Relativsatz trifft eine Aussage über sein Bezugswort *Re-*

²⁶ Teile des in diesem Kapitel vorgestellten Materials flossen auch in die Veröffentlichungen Voigtmann und Speyer (2021), Voigtmann (2022) und Voigtmann und Speyer (2022) ein. Da der Fokus des Kapitels auf verarbeitungsbezogenen Beschreibungen des Relativsatzes liegt, werden generative Analyseansätze (vgl. z. B. Inaba 2007) hier ignoriert.

staurant, weist Verbletzstellung des finiten Verbs auf und wird durch ein Relativpronomen eingeleitet, das eine „Identitätsbeziehung“ (Birkner 2008: 23) mit dem Antezedens eingeht. Dieser Aufbau und auch die identitätsstiftende Funktion der Relativsätze ist prototypisch für attributive Relativsätze. Der Aufbau lässt sich auch bei den anderen Relativsatzarten, mit denen sich diese Arbeit nicht befasst, die aber trotzdem kurz erwähnt werden sollen, nachweisen. Diese unterscheiden sich jedoch vor allem durch die Wahl und Art des Bezugswortes von den attributiven Relativsätzen.

Gallmann (2016) unterscheidet neben den attributiven *weiterführende* und *freie Relativsätze*. Weiterführende Relativsätze beziehen sich auf den gesamten Satz (vgl. Gallmann 2016: 1037). Deshalb definieren Eisenberg und Thieroff (2020: 299) sie als konjunktionale Nebensätze anstatt als Attributsätze. Dem schließen sich auch Helbig und Buscha (2005: 566) an. Sie negieren den Attributstatus des Relativsatzes und sehen daher auch den übergeordneten Satz nicht als dessen Bezugselement an. Hier ist zu bemerken, dass ihre Definition des Relativsatzes stark an seine Funktion geknüpft ist und weniger an seine Form. Denn formal sind weiterführende Relativsätze nicht von freien Relativsätzen zu unterscheiden (vgl. Wöllstein, Heilmann, Stepan und Vikner 1997: 48). Gleichzeitig weisen sie jedoch auch Ähnlichkeiten mit eingebetteten Fragen auf und lassen sich in Verbzweitsätze umformen, ohne den Sinn der Aussage zu verfälschen oder ihren Bezugssatz ungrammatisch zu machen (20).²⁷ Zu dem können weiterführende Relativsätze ausschließlich im Nachfeld stehen. Dadurch und wegen des fehlenden (pro-)nominalen Antezedens sind sie für die vorliegende Arbeit irrelevant.

- (20) a. Sie zog nicht zu ihm, was alle ein wenig überraschte.
 b. Sie zog nicht zu ihm. Das überraschte alle ein wenig.

Auch die freien Relativsätze weisen kein lexikalisch realisiertes Antezedens auf, haben sonst aber die Eigenschaften eines attributiven Relativsatzes. In

²⁷ Die Umformung eines attributiven Relativsatzes in einen Verbzweitsatz ist hingegen nur bei nicht-restriktiven Relativsätzen in Form einer Parenthese möglich. Handelt es sich bei dem attributiven Relativsatz aber um einen restriktiven Relativsatz, wird der gesamte Satz durch den Versuch der Umformung ungrammatisch.

der Schulgrammatik werden sie oft über ihre Funktion als Subjekts- oder Objektsätze klassifiziert (vgl. Pittner und Berman 2021: 118). Formal unterscheiden sie sich kaum von attributiven Relativsätzen mit Bezugsnomen. Auch sie sind Verbletztsätze, die durch ein Relativum eingeleitet werden. In der Wahl des Relativums liegt aber der Unterschied zu den attributiven Relativsätzen. Bei freien Relativsätzen sind d-Pronomen ausgeschlossen und die Fragepronomen *wer* oder *was* werden häufiger verwendet. In ihrer Position sind sie zudem freier als attributive und weiterführende Relativsätze und können beispielsweise allein im Vorfeld stehen, wenn sie als Subjekt- oder Objektsätze verwendet werden (vgl. z. B. Hentschel und Weydt 2013). Wenn man sie, wie von Wöllstein et al. (1997) vorgeschlagen, in einen Relativsatz mit Bezugselement umformuliert (21), unterliegen sie wieder den Bedingungen, die auch für attributive Relativsätze gelten.

- (21) a. Wer viel lernt, schreibt gute Noten.
b. Gute Noten schreibt, wer viel lernt.
c. Der, der viel lernt, schreibt gute Noten.
d. Der schreibt gute Noten, der viel lernt.

Das hier gezeigte Beispiel (21) ist ein simples, wenn auch typisches für freie Relativsätze. Prinzipiell wäre ein Vergleich zwischen Satz (21a) und Satz (21b) denkbar und im theoretischen Rahmen dieser Arbeit zu untersuchen. Dennoch habe ich mich dazu entschieden, sie ebenfalls nicht zu analysieren. Sie bilden eine neue Klasse von Sätzen, was für die vorliegende Arbeit zu weit führen würde. Eine Untersuchung ihrer Position und deren Variabilität wäre dennoch wünschenswert für zukünftige Arbeiten.

2.4.2. Art des Relativums

Das Antezedenz des Relativsatzes wird durch ein *Relativum* beziehungsweise ein Relativpronomen aufgegriffen. Zu den Relativpronomen gehören die d-Relativpronomen *der*, *die*, *das* und die w-Relativpronomen *welcher*, *welche*, *welches* sowie *wer*, *was* (vgl. Gallmann 2016: 1029). Letztere zählen zur Gruppe der relativ gebrauchten Interrogativa (vgl. Hentschel und Weydt 2013: 383). Es

ist zu beobachten, dass *welcher* hauptsächlich zur Vermeidung von Wiederholungen verwendet ist, und *der* nur dann ersetzen kann, wenn das Antezedens nicht im Genitiv steht (vgl. Fritsch 1990). Die Verwendung von *wer, was* ist zudem auf freie Relativsätze beschränkt. Helbig und Buscha (2005: 236) zählen die Relativpronomen dabei nicht zu einer eigenen Gruppe, sondern bezeichnen sie als morphologisch veränderte Demonstrativpronomen (*der*), Interrogativa (*welcher*) und Pronominaladverbien (*wo-*), die oft statt der Kombination aus Präpositionen und Relativpronomina verwendet werden. Zusätzlich können Pronominaladverbien, die mit *wo-* beginnen (vgl. z. B. Hentschel und Weydt 2013: 383), als Relativsatzeinleiter fungieren. Zudem kann der Relativsatz durch Pronomen, Adverbien und Subjunktionen wie *was, wo, als* und *wenn* eingeleitet werden (vgl. Gallmann 2016: 1030ff.). Während die ersteren jedoch bei allen Nomina und Pronomina verwendet werden können, sind gerade die von Helbig und Buscha (2005) genannten Subjunktionen stark von der inhärenten temporalen beziehungsweise lokalen Bedeutung ihres Bezugswortes abhängig und auch sonst semantisch stark eingeschränkt (22).

- (22) a. Der Moment, als ihr Kind geboren wurde, war der glücklichste in ihrem Leben.
 b. Der Moment, in dem ihr Kind geboren wurde, war der glücklichste in ihrem Leben.

Moment ist ein Nomen mit temporaler Bedeutung, weshalb die Relativsatzeinleitung durch *als* in Satz (22a) möglich ist. Der Satz lässt sich aber auch mit nur einem geringfügigen stilistischen Unterschied (22b) in eine Präpositionalphrase mit Relativpronomen umwandeln. Diese Operation ist bei allen nicht durch die klassischen Relativa eingeleiteten Relativsätzen möglich.

Die Relativpronomen erhalten durch ihre Funktion innerhalb des Relativsatzes den Kasus (Fritsch 1990: 55), Genus und Numerus hingegen von ihrem Bezugswort (vgl. Fritsch 1990; Gallmann 2016; Eisenberg und Thieroff 2020). Wenn sich das Relativpronomen auf ein Bezugsnomen in der ersten oder zweiten Person oder auf eine Höflichkeitsform bezieht, wird diese wiederholt, wenn das Relativpronomen das Subjekts bildet (23).

- (23) Ich, die **ich** diesen Kuchen gebacken habe, will mich ja eigentlich nicht selbst loben, aber mir schmeckt er wirklich gut.

Brooks (2006) beschreibt für das *Frühneuhochdeutsche* („1350-1650“, Nübling, Dammel, Duke und Szczepaniak 2013: 6) eine relative, zeitliche Stabilität der oben genannten Relativa. Im Frühneuhochdeutschen gab es die Möglichkeit, Relativsätze mit der Partikel *so* einzuleiten. *So* wurde ähnlich verwendet wie das englische *that* und weist keine Kasusmarkierungen auf. Nach Brooks (2006: 122) erwies sich das als Nachteil,

als dadurch die Identifizierung der Bezugsgröße im Obersatz erschwert wird. Die mangelnde Kasusmarkierung der Partikel bewirkt nicht nur den Zusammenfall von Nominativ und Akkusativ, sondern erklärt darüber hinaus auch, weshalb die Rolle eines Genitiv- oder Dativobjekts beinahe gar nicht übernommen werden kann.

Die fehlende Flektierbarkeit nimmt er auch als Grund für die abnehmende Verwendung von *so* im 18. Jahrhundert (vgl. Brooks 2006: 123). Die Partikel ist zum Kanzleistil zu zählen und wurde niemals so häufig verwendet wie andere Relativsatzeinleiter, obwohl sich eine Verwendungshäufung „im ersten Drittel des 17. Jahrhunderts“ (Brooks 2006: 131) verzeichnen lässt. Im Verlauf des Frühneuhochdeutschen zeigt sich gleichzeitig eine enorme Zunahme des aus dem Niederdeutschen stammenden *welch-* in Kombination mit Präpositionen (vgl. Brooks 2006: 138), die im modernen Deutschen nicht mehr in dem Ausmaß belegt ist. Auch die Verwendung des semantisch eingeschränkten *da* nimmt im Verlauf des 17. Jahrhunderts ab (vgl. Brooks 2006: 131).

2.4.3. Relativsatztyp

Relativsätze lassen sich in zwei verschiedene Typen unterscheiden: *restriktive* und *appositive* (auch nicht-restriktive beziehungsweise expletive) (vgl. Fritsch 1990: 99ff.). Weiterführende Relativsätze sind ausschließlich appositiv und freie Relativsätze kommen nur restriktiv vor. Aufgrund ihres Bezugswortes können attributive Relativsätze beide Typen bilden. Der Typ des Relativsatzes gibt an, ob das Antezedens durch den Relativsatz eindeutig identifiziert

wird oder nur zusätzliche Informationen über das Bezugswort gegeben werden (vgl. u.a. Zifonun et al. 1997; Helbig und Buscha 2005; Hentschel und Weydt 2013; Gallmann 2016; Eisenberg und Thieroff 2020). Erstere werden als restriktive Relativsätze bezeichnet; letztere gelten als appositive Relativsätze.

Hentschel und Weydt (2013) und Eisenberg und Thieroff (2020) sind sich darüber einig, dass die Hauptfunktion von restriktiven Relativsätzen (24) die Einschränkung der Bezugsmenge des Antezedens ist. Somit leistet ein restriktiver Relativsatz einen „Beitrag zur Bestimmung des Referenten oder allgemeiner des Denotats der NP, in der er vorkommt, er schränkt die Klasse der möglichen Referenten ein“ (Fritsch 1990: 99). Bei prädikativen Konstruktionen steht ausschließlich die restriktive Lesart zur Verfügung (vgl. Fritsch 1990: 100). Genauso sind bestimmte Pronomina damit verbunden: Artikelwörter wie *derjenige*, *der*, *ein*, negative Indefinita wie *nichts*, *kein*, *niemand* und verallgemeinernde Indefinita wie *einige*, *manche*, *ein paar*, *mehrere*, *viele*, *solche* liefern Hinweise darauf, dass ein nachfolgender Relativsatz restriktiv ist (vgl. z. B. Fritsch 1990; Gallmann 2016). In vielen Fällen ist zudem bereits im Lexikoneintrag dieser Pronomen und Indefinita vermerkt, dass sie ein w-Relativpronomen zu sich nehmen (vgl. Fritsch 1990: 103).

(24) Ursula erinnert sich an alle Schüler, die bei ihr in der Klasse waren.

(24) ist eindeutig als restriktiver Relativsatz zu lesen. Zum einen enthält er den Allquantor vor dem Kopf des Antezedens. Zum anderen müsste der Satz bei Fehlen des Relativsatzes so gelesen werden, dass sich Ursula an alle Schüler erinnert. Eine Einschränkung der Menge an Schülern dürfte nicht vorgenommen werden, obwohl das Weltwissen es unwahrscheinlich macht, dass keine Einschränkung vorgenommen werden muss. Erst der Relativsatz schränkt die Menge der Schüler eindeutig ein. Ursula erinnert sich nur an die spezifische Menge der von ihr unterrichteten Schüler.

Die Betonung in (24) liegt dabei auf dem Quantor. Der Relativsatz ist intonatorisch eng an den Matrixsatz angeschlossen, was ebenfalls typisch für restriktive Relativsätze ist und als Unterscheidungskriterium zu den appositiven Relativsätzen zu verstehen ist (vgl. Wöllstein et al. 1997: 1997).

Im Schriftlichen ist die Unterscheidung zwischen appositiven und restriktiven Relativsätzen schwer möglich, wenn nicht gerade die oben genannten Signalwörter zur Verfügung stehen. Wöllstein et al. (1997: 47) schlagen eine Ersetzungsprobe vor, um herauszufinden, ob ein Relativsatz restriktiv ist: Der Artikel des Antezedens soll durch eine Form von *derjenige* ersetzt werden. Im modernen Deutschen und bei gleichem Weltwissen der Gesprächsteilnehmenden mag das auch funktionieren. Doch gibt es speziell in historischen Texten immer wieder Kontexte, in denen trotzdem nicht entschieden werden kann, ob die Ersetzungsprobe zu einem semantisch äquivalenten Satz führt. Daher sind oft beide Lesarten möglich und hängen stark vom Weltwissen des Rezipienten ab (vgl. Eisenberg und Thieroff 2020).

Restriktive Relativsätze identifizieren ihr Bezugswort also im Idealfall eindeutig und lassen sich durch bestimmte Determinierer, Indefinita und Allquantoren erkennen. Außerdem weisen die Betonung auf dem Determinierer und ein intonatorisch engerer Anschluss an den Matrixsatz ebenfalls auf Restriktivität hin.

Appositive Relativsätze leisten quasi das genaue Gegenteil. Wird ein appositiver Relativsatz verwendet, ist das Bezugswort bereits eindeutig identifiziert. Daher kann der Relativsatz nur zusätzliche Informationen über sein Antezedens beitragen (vgl. z. B. Engel 2009; Hentschel und Weydt 2013; Eisenberg und Thieroff 2020). Deiktische Personalpronomen, Demonstrativpronomina, Possessivpronomina, pränominale Genitive und Eigennamen sind typische Formenmerkmale des Antezedens', die auf die appositive Lesart des Relativsatzes hinweisen (vgl. Fritsch 1990: 99). All diese Signalwörter disambiguieren die Referenz. Zusätzlich zu diesen Hinweiswörtern im Matrixsatz können auch Bestandteile des Relativsatzes selbst auf dessen appositiven Typ deuten. Abtönungspartikeln wie *ja* und Kommentationsaverbien wie *übrigens* und *selbstverständlich* können ausschließlich in appositiven Relativsätzen stehen. Daher ist deren Insertion auch eine Methode, um den Typ herauszufinden (vgl. z. B. Fritsch 1990; Wöllstein et al. 1997). Gallmann (2016: 1036) beschreibt appositive Relativsätze zudem als durch Parenthesen ersetzbar. Neben diesen lexikalischen Signalen spielt wiederum die Intonation eine Rol-

le bei der Unterscheidung. Poschmann und Wagner (2016) sehen appositive Relativsätze als eigenständige Sätze und Sprechakte an, die prosodisch unabhängig sind. Gerade im Gesprochenen findet ein „Intonationsbruch“ (Zifonun et al. 1997: 2007) zwischen Matrixsatz und Relativsatz statt. Die Betonung liegt außerdem im Gegensatz zu den restriktiven Relativsätzen auf dem Nomen statt auf dem Determinierer.

- (25) a. Er hat sein Auto, das **ja** sowieso selten geputzt wird, auch jetzt nicht in die Waschstraße gefahren.
 b. Er hat sein Auto - das wird ja sowieso selten geputzt - auch jetzt nicht in die Waschstraße gefahren.

In Beispiel (25a) weisen sowohl das Possessivpronomen in der Determiniererposition, das das Fahrzeug als Besitz von Klaus identifiziert, und die Abtönungspartikel *ja* auf einen appositiven Relativsatz hin. Prinzipiell wäre es anhand des Matrixsatzes noch möglich, auch hier von einem restriktiven Relativsatz zu sprechen. Wenn er mehrere Autos besitzt, könnte es sich um das häufig schmutzige oder sehr schmutzige Auto handeln. Erst die Partikel kennzeichnet den Satz eindeutig als nicht-restriktiv. Zusätzlich kann der Relativsatz wie in Beispiel (25b) in eine Parenthese umgewandelt werden, ohne dass eine Bedeutungsveränderung auftritt.

Insgesamt ist die Bestimmung des Relativsatztyps stark vom Kontext des Satzes und des Weltwissens der Rezipienten abhängig ist. Oft geben weder die Determinierer einen Hinweis auf die Lesart noch lexikalische Signalwörter im Relativsatz. Ebenso ist im Schriftlichen die prosodische Unabhängigkeit appositiver Relativsätze nur schwer zu erkennen.

2.5. Die Nominalphrase

Ein weiteres Element, das in der vorliegenden Arbeit hinsichtlich seiner Stellungsmöglichkeit im Nachfeld untersucht wird, ist die Nominalphrase. Die folgenden Unterkapitel sollen den Verwendungszweck der Nominalphrase knapp umreißen und mögliche Komplemente und Attribute benennen.

2.5.1. Form von Nominalphrasen

Die hier verwendeten Definitionen von *Nominalphrasen* entstammen einerseits einer generativen Sichtweise, die unter anderem durch Haider (1988), Glück (2000), Bhatt (2005) und Schlipphak (2008) vertreten ist, und andererseits einer deskriptiven Sichtweise, die beispielsweise von Zifonun et al. (1997) und Gallmann (2016) repräsentiert wird. Beide Definitionen stimmen darin überein, dass der Kopf einer Nominalphrase von einem Nomen oder Substantiv gebildet wird.²⁸ Allerdings divergieren die beiden Ansätze in ihrem Umgang mit Determinierern und Pronomen.

Gallmann (2016: 797) beschreibt die Nominalphrase als Phrase, deren Kern „ein Substantiv, eine Substantivierung oder ein Pronomen“ ist. Somit würden auch reine Pronomen als Nominalphrase analysiert werden. Dahingegen hebt Bhatt (2005: 1) hervor, dass der Kopf der Nominalphrase immer von einem Nomen gebildet wird, auch wenn dieses nicht zwangsläufig lexikalisch realisiert werden muss. In der generativen Grammatik werden Determinierer als Köpfe von *Determiniererphrasen* angesehen (vgl. z. B. Glück 2000), bilden somit also eine eigene Klasse, an deren Komplementstelle Nominalphrasen hängen. Die Determinierer können „unbetont“ und „unrealisiert bleiben“ (Bhatt 2005: 20).²⁹ Pronomen zählen nach diesem Ansatz also zu den Determinierern, die man klar von den Nominalphrasen unterscheiden kann, wodurch auch Unterschiede in der Erweiterbarkeit der beiden Phrasen beispielsweise hinter der rechten Satzklammer erklärt werden können.

Auf der Ebene bloßer Oberflächenstruktur macht es keinen Unterschied, ob man die Nominalphrase als Determiniererphrase oder als Nominalphrase, in deren Spezifikator der Determinierer steht, interpretiert. Die Analyse als Determiniererphrase bietet allerdings einige Vorteile, was die gramma-

²⁸ Auch wenn Zifonun et al. (1997: 981ff.) den Eindruck erwecken, Substantiv und Nomen bezeichneten verschiedene Dinge, ohne jedoch klar erkenntlich darzulegen, worin diese Unterschiede bestehen sollen, werden die beiden Bezeichnungen in dieser Arbeit synonym verwendet.

²⁹ Hinsichtlich der lexikalischen Realisierung unterscheiden sich der deskriptive und der generative Ansatz. Bhatt (2005: 1) beschreibt zum Beispiel, dass der Kopf der Nominalphrase nicht realisiert werden muss. Dass eine Phrase auch bei nicht lexikalisch realisiertem Kopf als Nominalphrase identifiziert werden kann, liegt auch daran, dass sie zur Determiniererphrase erweitert werden kann.

tische Analyse betrifft. Die Determinierer werden als funktionales Element ohne Thetarollenzuweisung klassifiziert und können keine anderen Komplemente als Nominalphrasen zu sich nehmen (vgl. Bhatt 2005: 28). Sie bilden den Sitz der grammatischen Merkmale Person, Numerus, Genus und Kasus und können auch durch Quantoren mit Definitheitsmerkmalen gebildet werden (vgl. z. B. Haider 1988; Bhatt 2005). Sie müssen allerdings auch nicht lexikalisch realisiert werden, um die Merkmale an die Nominalphrase weiterzugeben. Insgesamt zählen Determinierer beziehungsweise Artikelwörter (vgl. Gallmann 2016: 800) zu den häufigsten Erweiterungen der Nominalphrase.

2.5.2. Erweiterbarkeit der Nominalphrasen

Determinierer Die bekannteste und häufigste Erweiterung der Nominalphrase wird, wie erwähnt, von der Klasse der Determinierer gebildet. Darunter sind „Ausdrücke, die zusammen mit einem Nomen Ausdrücke der Kategorie **Nominalphrase** bilden“, zu verstehen (Zifonun et al. 1997: 1929).

Zu den deiktischen Determinierern zählen neben den definiten auch *der-/die-/dasjenige, derselbe* und *solcher*. Einer der Unterschiede zwischen den Artikeln besteht in ihrer Verwendung. Während der definite Artikel fast uneingeschränkt verwendet werden kann (vgl. Zifonun et al. 1997: 1930f.), bestehen Restriktionen hinsichtlich des indefiniten Artikels. Er kann beispielsweise nicht bei Nomen im Plural stehen.³⁰ Laut Gallmann (2016: 800) sind die Artikel meistens nicht weglassbar und mindestens zur Wohlgeformtheit der Nominalphrase, wenn nicht sogar zu ihrer Vollständigkeit, essentiell notwendig. Bei den possessiven Determinierern unterscheidet man zwischen „sprecherbezogene[n]“ und „genusdifferenzierte[n] gegenstandsbezogene[n] possessive[n] Determinativ[en]“ (Zifonun et al. 1997: 1933). Die verschiedenen Determinierer divergieren bezüglich ihrer Deixis (vgl. Zifonun et al. 1997: 1933) und kongruieren mit dem Nomen in Kasus, Numerus und Genus (vgl. Repp und

³⁰ Die Verwendung des definiten Artikels kann allerdings auch eingeschränkt sein. Das gilt für die Kombination des Artikels mit Eigennamen (vgl. Zifonun et al. 1997: 1931f.). Außerdem können Artikel bei Firmennamen stehen und bei „staatenbezeichnende Abkürzungen“ für Ländernamen (vgl. Zifonun et al. 1997: 1931). Auch Orts- und Eigennamen stehen in der Regel ohne Artikel. Werden Familiennamen mit einem Artikel versehen, wird dadurch Berühmtheit angezeigt (vgl. Zifonun et al. 1997: 1933).

Struckmeier 2020: 93). Zu den quantifizierenden Determinierern zählen „*einiger, etlicher, irgendein, irgendwelcher, aller, jeder, jedweder, jeglicher, mancher, mehrere und kein*“ (vgl. Zifonun et al. 1997: 1938) und zu den w-Determinierern werden „*welch, welcher wessen, wie vieler, wieviel* sowie *was für*“ gezählt (vgl. Zifonun et al. 1997: 1947). Deskriptiv stehen alle Determiniererklassen vor der Nominalphrase.³¹

Komplemente und Adjunkte Auch Attribute können zu den Nominalphrasen treten. Sie lassen sich aufgrund ihrer Form in acht Untergruppen aufteilen (Tabelle 2.7). In ihrer Funktion werden sie in die Klassen *Komplement* und *Adjunkt* unterschieden (vgl. z. B. Gallmann 2016: 800).

Komplemente sind vor allem bei „relationalen Nomina“, die von verbalen Nomen gebildet werden, zu finden (Bhatt 2005: 89). Die Erweiterung durch meistens im Genitiv stehende Nominalphrasen ist möglich (vgl. Gallmann 2016: 801). Werden Nominalphrasen durch Adjunkte erweitert, geschieht das meistens durch Adjektivphrasen, Präpositionalphrasen und Relativsätze (vgl. Bhatt 2005: 67). Oft ist es bereits durch das Nomen oder die Erweiterung selbst klar, ob es sich bei der Erweiterung um ein Adjunkt oder ein Komplement handelt, allerdings trifft das nicht auf alle Fälle zu.

Zur Bestimmung, ob die Erweiterung einer Nominalphrase ein Komplement oder Adjunkt bildet, können verschiedene Kriterien wie die *Fixiertheit* (vgl. Zifonun et al. 1997: 1969) herangezogen werden. Sie beschreibt, ob die Erweiterung obligatorisch ist. Ist sie obligatorisch, handelt es sich um ein Komplement, andernfalls um ein Adjunkt. Dieses Kriterium ist aber bei Verben, bei denen es ursprünglich Anwendung fand, zuverlässiger als bei Nomen. Mit ihm verknüpft ist das *Rektionskriterium*, das besagt, dass der Komplementstatus auch durch Formveränderungen des Kasus oder der Präposition offenbart wird (vgl. Zifonun et al. 1997: 1969).

Ein weiteres Kriterium ist *Konstanz*. Sie bezieht sich darauf, dass die Präposition von der Nominalphrase festgelegt und nicht beliebig austauschbar ist

³¹ Im Ansatz der NP-Analyse stehen sie in deren Spezifikatorposition, wohingegen sie im Ansatz der DP-Analyse linksköpfig sind und die NP nachgestellt als Komplement zu sich nehmen (vgl. z. B. Haider 1988; Bhatt 2005).

Art des Attributs	Beispiel
Adjektivphrasen (Gallmann 2016: 800)	Die blaue Tasse ist in der Spülmaschine.
Adverbphrasen (Gallmann 2016: 801)	Die Klausur morgen wird bestimmt gut.
Genitiv-Nominalphrasen (Gallmann 2016: 801)	Das Haus der Frau wird bald abgerissen.
Apposition (Gallmann 2016: 801)	Bundeskanzler, Olaf Scholz , spricht heute vorm Bundestag.
Präpositionalphrasen (Gallmann 2016: 802)	Sie besuchte das Haus am See .
Konjunkionalphrasen	„Annas Bekanntheit als gute Schachspielerin...“ (Gallmann 2016: 802)
Fokuspartikel	„sogar Anna“ (Gallmann 2016: 802)
Nebensätze	Ein Buch, das jeder gelesen haben sollte , ist Harry Potter. Der Bericht, dass Cäsar Rom erobern will , erreichte den Senat.

Tabelle 2.7.: Beispiele für Nominalphrasen, die durch Attribute erweitert sind.

(vgl. Zifonun et al. 1997: 1971). In dieser Hinsicht ist das Kriterium nicht klar vom Rektionskriterium zu trennen. Im Fall von Präpositionalphrasenkomplementen regiert das Nomen die Präposition und den Kasus des Komplements der Präpositionalphrase (vgl. Zifonun et al. 1997: 1971). (26) verdeutlicht das.

- (26) a. Vertrauen in [den Staat]_{Akkusativ}
 b. das Buch in [der Vitrine]_{Dativ}

In (26a) bestimmt die Nominalphrase *Vertrauen* sowohl die Präposition als auch, dass diese Präposition den Akkusativ regiert. In (26b) besteht dieses Rektionskriterium nicht. Es handelt sich daher um ein Adjunkt statt um ein Komplement. Wird eine Präpositionalphrase als Komplement interpretiert, selegiert die Nominalphrase die Präposition (vgl. Zifonun et al. 1997: 1977).

Wird die Nominalphrase durch eine Präpositionalphrase erweitert, ergeben sich zwei Lesarten, wenn sie als Adjunkte an die Nominalphrase ange-

geschlossen werden: eine *restriktive* und eine *appositive* Lesart (vgl. Zifonun et al. 1997: 2033). Die intonatorische Integriertheit in die Nominalphrase klassifiziert ein Präpositionalattribut als restriktiv (vgl. Zifonun et al. 1997: 2033). Zudem können die restriktiven Präpositionalattribute sowohl post- als auch pränominal verwendet werden, wohingegen appositive Präpositionalphrasen nur postnominal stehen können. Die pränominale Position dieser Attribute ist jedoch selten und markiert (vgl. Zifonun et al. 1997: 2033).

Unterschiedliche Lesarten finden sich auch bei der Erweiterung der Nominalphrase durch Adjektive und Relativsätze, nicht jedoch für „Nominalphrasen im Genitiv, die ausschließlich restriktiv sind“ (Zifonun et al. 1997: 1989). Auf Komplementebene können Nominalphrasen durch „Numerativkonstruktionen“ (Zifonun et al. 1997: 1979) erweitert werden. Diese Phrasen bestehen „aus einem Maßausdruck (z. B. *Gramm, Liter, Gros, Dutzend, Bogen*), einem Zahlattribut und einem substanzdenotierenden Substantiv oder einem Substantiv im Plural“ (Zifonun et al. 1997: 1979).

Neben den sowohl restriktiv als auch appositiv zu lesenden Relativsätzen (Kapitel 2.4.3) können Nominalphrasen auch durch weitere Nebensatzarten, die als Komplemente und Adjunkte fungieren, erweitert werden. Als Komplemente können *dass*-Sätze und Infinitivkonstruktionen (vgl. Zifonun et al. 1997: 1978) zur Nominalphrase hinzutreten. Weitaus häufiger fungieren Nebensätze allerdings als Adjunkte. Sie denotieren gewisse Charakteristika dieser Phrasen (vgl. Zifonun et al. 1997: 1987), gehören fest zur Nominalphrase und sind ihrerseits erweiterbar.

Bereits in Tabelle 2.7 wurden Adjektivattribute aufgelistet. Sie zählen neben Nominalphrasen im Genitiv zu den gängigsten Erweiterungen der Nominalphrase. Bis auf die als archaisch zu bewertenden Fälle, in denen das Adjektiv unflektiert hinter dem Nomen steht (z. B. *Freude pur*), werden Adjektivattribute pränominal verwendet. Sie stehen nach Zifonun et al. (1997: 1991) in Kasus-Numerus-Genus-Kongruenz (KNG-Kongruenz) zu ihrem Kopfnomen. Ob sie stark oder schwach flektiert werden, hängt vom Determinierer ab, der bei der Nominalphrase steht (vgl. Gallmann 2016: 256f.). Adjektivattribute weisen Skopus über das rechts von ihnen stehende Nomen zu (vgl.

Zifonun et al. 1997: 1992). Wird eine Nominalphrase mit einer Adjektivphrase zusätzlich durch einen pränominalen Genitiv erweitert, steht dieser vor dem Adjektiv und niemals zwischen Adjektiv und Nomen.

Auch die Kombination einer Nominalphrase mit einer nachgestellten Genitivnominalphrase wird restriktiv gelesen. Diese Genitivphrase ist „intonatorisch in die NP integriert“ (Zifonun et al. 1997: 2017) und kann nicht mit Partikeln kombiniert werden (vgl. Zifonun et al. 1997: 2018). Sie ist bedeutungsgleich zum nachgestellten Genitiv, aber nicht endozentrisch (vgl. Zifonun et al. 1997: 2020). Zifonun et al. (1997: 2023) beschreiben sie zudem als rekursiv, da auch die Genitivphrase durch Attribute erweitert werden kann.

Die letzte hier zu erwähnende Erweiterungsklasse, die in Nominalphrasen enthalten sein kann, sind *appositive Erweiterungen* oder Appositive. Diese Zusätze können selbst keine Illokution tragen und sind semantisch eigenständig (vgl. Zifonun et al. 1997: 2036). Diese meistens vollständigen Nominalphrasen stehen zu ihrer Bezugsphrase in KNG-Kongruenz und können, wenn auch eingeschränkt, ihrerseits erweitert werden (vgl. Zifonun et al. 1997: 2042).

Die bisherigen Abschnitte haben sich mit Nominalphrasen im modernen Standarddeutschen befasst. Im Folgenden soll der Fokus auf früheren Sprachstufen, mit einem Hauptaugenmerk auf dem Frühneuhochdeutschen, liegen.

2.5.3. Entwicklung der Nominalphrase

Seit dem *Althochdeutschen* („ca. 750 bis 1050“, Nübling et al. 2013: 6) besteht die erwähnte Kongruenz zwischen Kasus, Numerus und Genus (vgl. Demske 2001). Auch im Frühneuhochdeutschen kongruieren Adjektive und Artikel mit dem Kopf der Nominalphrase, wobei zwischen verschiedenen Flexionen beim Adjektiv unterschieden werden muss. Schon diese Periode kennt die Distinktion zwischen stark flektierten, schwach flektierten und unflektierten Adjektiven, die im Frühneuhochdeutschen in ähnlicher Häufigkeit vorkommen (vgl. Demske 2001: 76). Anders als im Gegenwartsdeutschen wird die Wahl der Flexion aber nicht vom Artikelwort, sondern vom Definitivitätsmerkmal auf dem jeweiligen Determinierer bestimmt (vgl. Demske 2001: 76). Star-

ke Flexion findet sich daher auch in Zusammenhang mit Possessivpronomen, dem unbestimmten Artikel, „jeder, all-, solch-“ (Demske 2001: 77) sowie dem bestimmten Artikel, dem Demonstrativpronomen und bei unflektierten Artikelwörtern (vgl. Demske 2001: 78). Sie kann außerdem auch bei Feminina und Vokativen auftreten. Ein weiterer Unterschied zur Flexion, die man im modernen Deutschen findet, besteht in der Nutzung unflektierter Adjektive. Sie kommen heutzutage nur noch in prädikativer oder archaischer, nachgestellter Verwendung vor, wohingegen sie im Frühneuhochdeutschen auch nach unflektierten Artikelwörtern und Possessivpronomen und im Femininum auftreten können (vgl. Demske 2001: 78). Wird eine Nominalphrase von zwei Attributen näher bestimmt, kann allerdings nur das erste unflektiert bleiben (vgl. Demske 2001: 79). Im 16. Jahrhundert ist diese Entwicklung abgeschlossen. Bis dahin war es auch möglich, prädikative Adjektive stark zu flektieren (vgl. z. B. Ebert, Reichmann, Solms und Wegera 1993; Demske 2001).

Erklärt wird diese Entwicklung durch das Adjektivparadigma. Genusunterschiede sind im Frühneuhochdeutschen nur noch im Singular sichtbar (vgl. Demske 2001: 84). Für den Erstspracherwerb in dieser Periode bedeutet das, dass der Zusammenhang zwischen Flexion und Definitheit, den es noch im Althochdeutschen gab, nicht mehr nachvollziehbar ist (vgl. Demske 2001: 84). Demske (2001) vernachlässigt bei dieser Erklärung jedoch sowohl Entwicklungsschritte im *Mittelhochdeutschen* („ca. 1050 bis 1350“ Nübling et al. 2013: 6) als auch die Frage, ob diese Unterscheidbarkeit durch andere sprachliche Mittel ausgeglichen wurde. Denkbar ist, dass die Ambiguität hinsichtlich der Definitheit gewollt war oder durch die Stellung der Nominalphrase im Satzgefüge oder durch prosodische Mittel beseitigt wurde.³² Die Unterscheidung zwischen dem bestimmten Artikel und dem Possessivpronomen sowie „die erweiterten Formen *denen, deren, derer, dessen* im Genitiv Singular sowie Genitiv und Dativ Plural des pronominalen Paradigma“ entwickeln sich erst im 14. Jahrhundert (Demske 2001: 87f.).

Auch im Frühneuhochdeutschen lassen sich die Nominalphrasen mit Genitivattributen erweitern. Im Gegensatz zur heutigen Verwendung und zum

³² Eine genauere Ursache-Wirkungs-Forschung ist an dieser Stelle nicht zu leisten.

heutigen Aufbau wurde der pränominaler Genitiv häufiger und mit weniger Beschränkungen verwendet. Beispielsweise steht er nicht nur bei Eigennamen, sondern modifiziert auch Individualnomen (vgl. Demske 2001: 215). Selbst partitive Genitive können pränominal auftreten (vgl. Demske 2001: 216). Für die postnominalen Genitive gibt es keine semantischen Restriktionen (vgl. Demske 2001: 216). Insgesamt scheint die Stellung des Genitivs zu Beginn des Frühneuhochdeutschen bis ins 15. Jahrhundert variabel zu sein. Erst im 16. Jahrhundert findet die schon im Althochdeutschen beginnende Entwicklung ihren Abschluss; Individualnomen, die sich auf unbelebte Referenten beziehen, mit einem pränominalen Genitiv gelten als markiert (vgl. Demske 2001: 216f.). Eine Ausnahme davon bilden Verbindungen, die sich später zu Komposita entwickeln. Zu nennen sind hier „des reichs sachen“ oder „des halßgerichts ordnung“ (Ebert 1986: 217).

Auch die Nachstellung belebter Nominalnamen beginnt im Frühneuhochdeutschen. Bereits um 1700 gilt ihre pränominaler Stellung als stark markiert. Sie sind daher fast nur noch postnominal zu finden (vgl. Demske 2001). Postnominal stehen auch Eigennamen, allerdings nur, wenn sie lateinisch flektieren (vgl. Demske 2001: 219). Durch die zunehmende Beschränkung der Elemente, die pränominal stehen können, erhöht sich insgesamt der Anteil der postnominalen Genitive (vgl. Demske 2001: 220), wie Demske anhand ihres Zeitungskorpus nachweisen kann. Nach Ebert (1986) beruht die Verschiebung auf dem Gesetz der wachsenden Glieder (vgl. Behaghel 1932: 139), da unbelebte Individualnomen häufiger durch Attribute erweitert werden. Das macht sie länger, sodass sie nach diesem Gesetz weiter hinten stehen müssen. Braunmüller (1982) hingegen sieht eine typologische Veränderung, die mit der veränderten Rektionsrichtung von Nominalphrasen korreliert ist. Statt nach links regieren Nomina nach rechts. Demske (2001: 221, nach Braunmüller (1982)) führt hier allerdings an, dass diese Analyse problematisch ist, da pränominaler Adjektive für die Rektionsrichtung nach links sprechen und auch die OV-Stellung im abhängigen Nebensatz nicht auf Rektion nach rechts deutet. Wieder einen anderen, nämlich pragmatischen Ansatz für Sprachwandel verfolgt Oubouzar (1990). Während pränominaler Genitive im Althochdeutschen

noch die Funktion des Artikels übernehmen, fällt das durch die Etablierung des bestimmten Artikels und des Demonstrativpronomens im Frühneuhochdeutschen weg. Demske (2001) bleibt der Erklärung, warum die Eigennamen dann immer noch diese Funktion übernehmen können, schuldig.

Syntaktisch haben prä- und postnominale Genitive ähnliche Eigenschaften, sind sogar durch Relativsätze erweiterbar (vgl. Demske 2001: 223) und erlauben die Distanzstellung der Erweiterungen und der Attribute (vgl. Demske 2001: 224). Zudem müssen weder Possessivpronomen noch pränominalen Genitive phraseninitial stehen. Eine zusätzliche Erweiterung durch Adjektive ist möglich (vgl. Demske 2001: 227). Beide gelten dann semantisch als Argumente des Nomens, weil sie bei sortalen Konzepten mit dem bestimmten Artikel vorkommen (vgl. Demske 2001: 227) und das Nomen nur dann indefinit ist, wenn der Genitiv vor dem Nomen stehen kann (vgl. Demske 2001: 228).

Außerdem können pränominalen Genitive mit Adjektivattributen links des Kopfnomens auftreten, was in späteren Sprachstufen nicht mehr der Fall ist (vgl. Demske 2001: 225).³³ Eine Nominalphrase kann zudem indefinit sein und gleichzeitig einen pränominalen Genitiv besitzen, wenn sie mit einem indefiniten Artikel versehen ist (vgl. Demske 2001: 225). Neben dem pränominalen Genitiv können auch Demonstrativpronomen und *solch-* links des Kopfnomens der Nominalphrase stehen (vgl. Demske 2001: 226). Letzteres hat dabei ähnliche Eigenschaften und Erweiterungsmöglichkeiten wie Possessivpronomen. Anders als im modernen Deutschen werden Nominalphrasen mit einem pränominalen Genitiv als indefinit interpretiert:

Damit lassen sich auch diese Konstruktionen in die Geschichte der Definitheitsanzeiger einreihen, anders ausgedrückt gilt auch für die Fälle indirekter Referenz wie für die anderen Instanzen semantischer Definitheit, dass erst im Verlauf des Frnhd. der bestimmte Artikel obligatorisch wird. (Demske 2001: 229f.)

Bisher wurde die Erweiterung der Nominalphrase durch Determinierer und pränominalen Genitive thematisiert. Im Frühneuhochdeutschen besteht außerdem eine Möglichkeit, Nominalphrasen durch possessive Dative (z. B. [*dem*

³³ Allerdings ist hier zu bemerken, dass die von Demske (2001) angeführten Beispiele sich auf Individualnomina, die in der Periode sowieso schon seltener werden und nicht auf Eigennamen, die im Gegenwartsdeutschen pränominal stehen können, beziehen. Wie vergleichbar diese Entwicklung daher mit dem Gegenwartsdeutschen ist, ist fraglich.

Vater]_{Dat sein Geschäft}) zu erweitern (vgl. Demske 2001: 258f.). Dieser kann dabei in den Pertinenzdativ und in „interne possessive Dative“ (Demske 2001: 259) unterteilt werden. Die Verwendung ist unabhängig vom Kopfnomen und erst seit dem Frühneuhochdeutschen belegt (vgl. Demske 2001: 259).

Ebert (1986) führt den generellen Rückgang adnominaler Genitive, der durch die veränderte Nominalmorphologie zustande kam, welche die Entwicklung komplexer Nominalphrasen im Frühneuhochdeutschen ermöglicht, als Grund für die Nutzung des possessiven Dativs an. Dieser breitete sich im 14. und 15. Jahrhundert vor allem in mündlichkeitsnahen Texten aus (vgl. Demske 2001: 260). Allerdings gehört der possessive Dativ nur dann zur Nominalphrase, wenn die Relation zwischen der pränominalen Nominalphrase und dem Kopf der Hauptnominalphrase eindeutig interpretiert werden kann (vgl. Demske 2001: 260). Diese eindeutige Relation drückt die Zugehörigkeit der adjazent stehenden, adnominalen Dativnominalphrase zur eigentlichen Nominalphrase aus. Die Dativnominalphrase bildet dann ein Adjunkt (vgl. Demske 2001: 263, nach Johnson (1988)). Sonst muss statt des possessiven Dativs ein postnominaler Genitiv verwendet werden (vgl. Demske 2001: 261).

Einen weiteren Gegensatz zum modernen Deutschen bildet die liberalere Möglichkeit, diskontinuierliche Nominalphrasen im Frühneuhochdeutschen zu bilden. Während sich das heutige Standarddeutsche durch eine eher strikte Serialisierung in der Nominalphrase auszeichnet (vgl. Demske 2001: 268), die Bewegungen aus der Nominalphrase nur marginal erlaubt, können im Frühneuhochdeutschen Teile der Nominalphrase nach rechts verschoben werden oder durch „*was für*-Split und [...] Quantorenfloating“ (Demske 2001: 269) getrennt werden. Im Gegensatz zu der vorliegenden Arbeit fokussiert sich Demske (2001: 270) in ihrem Text allerdings auf die Verschiebung nach links, die nur für Nominalphrasen „mit Kongruenzkasus sowie Präpositionalphrasen“, nicht aber für Phrasen im Genitiv oder mit satzwertigen Attributen möglich ist. Außerdem müssen diese Nominalphrasen bestimmte syntaktische Funktionen erfüllen, um verschoben werden zu können. So kommt es nur bei Akkusativobjekten und Subjekten vor, solange diese nicht von Eigenschaftsprädikaten abhängen, die keine partitiven Splitkonstruktionen bilden

(vgl. Demske 2001: 270). Generell gibt es für nicht-partitive Attribute mehr Beschränkungen bei der Extraktion³⁴ als für Partitive (vgl. Demske 2001: 271). Die nicht-partitive Distanzstellung ist beispielsweise nur bei indefiniten Nominalphrasen möglich, bei partitiven besteht diese Einschränkung nicht (vgl. Demske 2001: 271).

Unabhängig vom partitiven Charakter der Nominalphrase, beschreibt Demske (2001: 274) diese Konstruktion als sehr häufig. Unterbrechungen zwischen den Teilen einer Nominalphrase sind möglich und nicht nur auf die Verschiebung ins Vorfeld beschränkt (vgl. Demske 2001: 275). Oft wird der abgetrennte Teil unmittelbar vor die zugehörige Nominalphrase gestellt (vgl. Demske 2001: 277). Bei partitiven Präpositionalphrasen, die aus der Nominalphrase extrahiert werden, ist sogar nur die unmittelbare Voranstellung möglich (vgl. Demske 2001: 279). Die Komplexität des vorgelagerten Teils ist nicht entscheidend, da auch sehr komplexe partitive Genitive beispielsweise belegt sind (vgl. Demske 2001: 277). Vereinzelt finden sich auch Belege für diskontinuierliche Dativ-Nominalphrasen und auch für partitive Genitive (vgl. Demske 2001: 275f). Bei ihnen hängt die Diskontinuität davon ab, ob sie eine Maßangabe mit Kasusflexion beinhalten oder von Qualitätsadjektiven in substantivierter Verwendung gebildet werden oder nach Indefinitpronomen stehen (vgl. Demske 2001: 276). Das Kopfnomen selbst ist nicht ausschlaggebend für die Diskontinuität der Nominalphrase (vgl. Demske 2001: 277).

2.6. Die Präpositionalphrase

In den vorherigen Kapiteln wurden Relativsätze und Nominalphrasen beschrieben. Dieses Kapitel befasst sich nun mit den Präpositionalphrasen, einer der Phrasenklassen, die am häufigsten im Nachfeld stehen können (vgl. z. B. Zifonun et al. 1997: 1657). Das Kapitel beginnt mit einer Betrachtung der Form der Präpositionalphrase (Kapitel 2.6.1), wird vom Teilkapitel über ihre Funktion im Satz (Kapitel 2.6.2) fortgeführt und schließlich mit einer jeweils

³⁴ Darunter wird das Herausbewegen von linguistischem Material von seiner ursprünglichen Position, ohne notwendigerweise die Verschiebung ins Nachfeld zu bezeichnen, verstanden (vgl. z. B. Repp und Struckmeier 2020: 149).

knappen Betrachtung der Bedeutung der Präposition (Kapitel 2.6.3) und der Entwicklung der Präpositionalphrase (Kapitel 2.6.4) abgeschlossen.

2.6.1. Form der Präpositionalphrase

Namens- und kopfgebend für *Präpositionalphrasen* ist die in ihnen enthaltene Präposition. Präpositionen werden der funktionalen Klasse zugeordnet (vgl. Fries 1991; Breindl 2006), können nicht flektiert werden (vgl. Fries 1991: 75) und werden in primäre und sekundäre Präpositionen unterschieden (vgl. Breindl 2006: 938). Die sekundären werden wiederum nach monolexematischen und polylexematischen Präpositionen differenziert (vgl. Breindl 2006: 938). Letztere bestehen aus mehreren Wörtern (zum Beispiel die Zirkumposition *um... willen* oder *auf Grund*), die aber als eine Präposition angesehen werden können.

Präpositionen stehen zwischen den lexikalischen und funktionalen Kategorien (vgl. Breindl 2006: 937). Sie können zwar eine eigenständige lexikalische Bedeutung haben, fungieren aber auch als Kopf der Phrase und nehmen ein einziges Komplement zu sich (vgl. Fries 1991: 74), dem sie Kasus zuweisen (vgl. Breindl 2006: 938). In der Regel handelt es sich bei diesem Argument um eine Nominalphrase (vgl. Nübling 2016: 602), aber auch Adjektive, Adverbien und andere Präpositionen sind als Komplemente möglich (vgl. z. B. Breindl 2006; Nübling 2016).

Die Präposition regiert den Kasus der Nominalphrase, wobei es prinzipiell keine Einschränkungen hinsichtlich der Kasusauswahl gibt (vgl. Nübling 2016: 606). Lediglich der Genitiv ist im modernen Standarddeutschen als Dependens im Rückgang und wird häufig durch den Dativ ersetzt, wie bei der Präposition *wegen* zu sehen ist.³⁵ Die Kasusreaktion ist nach Zifonun et al. (1997: 2077) ein ausschlaggebendes Kriterium zur Bestimmung von Präpositionen. Sie zählen aufgrund der fehlenden Kasusreaktion beispielsweise *als*, *wie* und *lang* nicht zu den Präpositionen und bezeichnen sie stattdessen als Adjunktoren beziehungsweise als Satzadverbien.

³⁵ Der Dativ wird genutzt, „wenn dem Bezugssubstantiv seinerseits ein Genitivattribut folgt“ (Nübling 2016: 612). Auch die Art, wie der Genitiv markiert wird, beeinflusst seine Verwendung.

Die Präpositionen *bis*, *seit* und *von* können auch Adverbialphrasen und Präpositionalphrasen als Argumente zu sich nehmen (vgl. Zifonun et al. 1997: 2078). Dass dies möglich ist, steht im Lexikoneintrag dieser Präpositionen (vgl. Breindl 2006: 942). Nur mit anderen Adverbphrasen, aber nicht mit Präpositionalphrasen kombinierbar sind die Präpositionen *ab*, *nach*, *für* und *seit* (vgl. Zifonun et al. 1997: 2079). Hängt von der Präposition eine weitere Präpositionalphrase ab, wird deren Präposition nicht von der ersten selektiert. Die zweite Präposition „muss nur ein semantisch kompatibles temporales oder lokales Adverbial bilden können“ (Breindl 2006: 942). Die Kasusreaktion des in der zweiten Präpositionalphrase enthaltenen Arguments wird ausschließlich von der zweiten Präposition bestimmt (vgl. z. B. Breindl 2006; Nübling 2016). Die Füllung der Argumentstelle einer Präpositionalphrase durch eine Adjektivphrase ist nach Breindl (2006: 942) nur bei Phraseologismen, bei der Präposition *von* und *bis* und als Komplement zu bestimmten Verben wie *erachten* möglich.

Präpositionen können auch in Form von Post- und Zirkumpositionen vorkommen (vgl. z. B. Zifonun et al. 1997; Breindl 2006; Nübling 2016). Sie stehen dann entweder nach ihrem Komplement oder umrahmen es. Im Fall der Präpositionen *durch* und *entlang* hat das Einfluss auf die Rektionsbedingung (vgl. Zifonun et al. 1997: 2079): Werden diese Präpositionen als Postposition bei Adverbialen verwendet, regieren sie ihr Komplement nicht mehr. Die Präposition kann nicht weggelassen werden, ohne dass die Phrase selbst ungrammatisch wird oder ihren Status als Präpositionalphrase verliert (vgl. Breindl 2006: 939).

Keinen Einfluss auf die Rektionsbedingungen haben Modifikationen und Modifikatoren der Präpositionalphrasen (vgl. Breindl 2006; Nübling 2016). Die Modifikatoren stehen immer links von und adjazent zur Präpositionalphrase (vgl. Breindl 2006: 943). Pro Präpositionalphrase kann es nur eine Modifikation geben (vgl. Breindl 2006: 943). Trotz gleicher Bezeichnung divergiert das Verständnis von Modifikationen erheblich zwischen Breindl (2006) und Zifonun et al. (1997). Nach Breindl (2006: 943) können sie durch „Maß und Dimensionsbezeichnungen“ erweitert werden, die von Adjektiven, Ad-

verbien oder Nominalphrasen im Akkusativ gebildet werden. Eine Modifikation durch weitere Präpositionalphrasen sei ebenso ausgeschlossen wie eine mehrfache Modifikation (vgl. Breindl 2006: 943).

Zifonun et al. (1997: 2090) hingegen führen auch ein Beispiel mit einer im Nominativ stehenden Nominalphrasen an und sagen, dass es zum einen Präpositionalphrasenerweiterungen geben kann und dass zum anderen Erweiterungen sowohl links als auch rechts von der Präpositionalphrase stehen können. Das gelte auch für Adverbphrasen (vgl. Zifonun et al. 1997: 2090).

Zur Form der Präpositionalphrasen ist abschließend zu sagen, dass Präpositionen auch klitisch vorkommen können (vgl. Fries 1991: 74), vor allem wenn es sich um eine hochfrequente Präposition handelt, die mit einem Artikel kombiniert wird (vgl. Nübling 2016: 615). Dies ist allerdings nur möglich, wenn mit dem nachfolgenden Nomen kein „konkretes Einzelobjekt (Gegenstand, Person, Institution, Zeitangabe) gemeint ist, das kataphorisch näher definiert (identifiziert wird), z. B. durch einen Relativsatz“ (Nübling 2016: 617). Handelt es sich bei dem von der Präposition abhängenden Komplement um ein Pronomen, ist es üblich, diese Kombination durch ein Präpositionaladverb zu ersetzen. Aus der Präpositionalphrase wird dann gemäß Gallmann (2016: 837) eine Adverbphrase.³⁶

2.6.2. Funktion der Präpositionalphrase

Die Funktion der Präpositionalphrase im Satz ist variabel. Als Objekte bezeichnen sie handlungsbetroffene Personen und Sachen und können nur durch Präpositionaladverbien, in denen die Präposition selbst vorhanden ist (vgl. Eroms 1991: 45), oder durch Nebensätze ersetzt werden (vgl. u. a. Eroms 1991; Breindl 2006; Gallmann 2016). Ihre Präposition ist desemantisiert, fixiert (vgl. Breindl 2006: 936) und vom Verb regiert (vgl. Gallmann 2016). Sie kennzeichne nach Breindl (2006: 936) nur „den Argumentstatus“, wobei Eroms (1991) ihr dahingehend widerspricht, wie unten gezeigt wird. Die Präposition kann zudem nicht ausgetauscht oder weggelassen werden, ohne den Satz ungram-

³⁶ Breindl (2006: 942) argumentiert in der Hinsicht zwar dafür, dass Präpositionalphrasen und Adverbialphrasen als Komplemente die gleiche Struktur in der x-Bar-Theorie aufweisen. Da aber der Kopf ein anderer ist, kann die Phrase nicht der gleichen Klasse zugeordnet werden.

matisch zu machen oder seine Bedeutung zu verändern (vgl. Eroms 1991: 44). Beispiel (27) nutzt das Verb *warten*, das ein Präpositionalobjekt mit der Präposition *auf* fordert.

(27) Er wartet **auf** den Briefträger.

Die fehlende Eigensemantik und die Fixiertheit der Präposition bei Präpositionalobjekten werden als eines der Hauptkriterien zur Unterscheidung von Präpositionalobjekten und Adverbialen, die in Form von Präpositionalphrasen auftreten, genannt. Eroms (1991: 47) hebt die Leistung der Präpositionalobjekte hervor, „eine tiefer eingebettete Prädikation“ zu komprimieren. Ihre Satzeinbettung postuliere die lexikalische Bedeutung, sodass auch bei schwer zu abstrahierenden Relationen die Grundsemantik der Präposition selbst noch zu erkennen sei (vgl. Eroms 1991: 48).

Werden Präpositionalphrasen als Adverbiale gebraucht, können sie sowohl von der Valenz gebunden als auch freie Angaben sein (vgl. Eroms 1991: 39). Als Präpositionalergänzungen kommen sie vor allem als Situativ- und Direktionalergänzungen (vgl. Eroms 1991: 39) sowie in temporalen Rollen (vgl. Breindl 2006: 936) vor. In dem Fall behalten die Präpositionen ihre lexikalische Bedeutung bei. Zudem ist die Präposition auch austauschbar, solange der Charakter der Ergänzung nicht verändert wird (28).

(28) Ich werfe den Ball **über die Hecke / in den Teich**.

Die direktionale, präpositionale Adverbiale ist in (28) hervorgehoben. Ihre Funktion ist gleich, doch trotz unterschiedlicher Präpositionen wird der Satz nicht ungrammatisch, weil die Wahl dieser Präpositionen „nach semantischen Gesichtspunkten“ erfolgt (Gallmann 2016: 840).

Unabhängig davon, ob die Präpositionalphrase als Adverbial von der Verbvalenz gefordert wird oder nicht, können sie in „Adverbialkombinationen“ vorkommen (Breindl 2006: 943). Es handelt sich dann immer noch um Adverbiale, die aber mit anderen Adverbialphrasen, anderen Präpositionalphrasen und Sätzen kombiniert werden können (vgl. Breindl 2006: 943). Ihre Position im Satz ist frei und die einzelnen Bestandteile, deren Anzahl nicht

durch die Grammatik begrenzt ist, sind voneinander trennbar (vgl. Breindl 2006: 943). Strukturell können sie ambig sein, was sich jedoch nicht auf ihre Semantik ausweiten muss, und weder durch Position noch Intonation behoben werden kann (vgl. Breindl 2006: 944).

Des Weiteren können Präpositionalphrasen prädikativ und attributiv vorkommen (vgl. Gallmann (2016: 839), Breindl (2006: 937) und andere). In ihrer prädikativen Verwendung erfüllen sie „die gleiche Aufgabe wie der prädikative Nominativ oder Akkusativ, d.h., sie bezieh[en] sich je nachdem auf das Subjekt oder auf das Objekt“ (Gallmann 2016: 839). Ihre attributive Verwendung ist mit der aller anderen Attribute vergleichbar (vgl. Gallmann 2016: 839). Das heißt, sie können vor allem bei Nomen und Adjektive auftauchen, wobei letztere auch fixe Präpositionen verlangen können (29):

(29) stolz **auf** ihn

Wenn die Präposition bei Attributen nicht valenzgebunden ist, ist sie frei austauschbar und von der Semantik bestimmt (30):

(30) das Buch **unter/auf/neben** dem Tisch

Wie bereits eingangs erwähnt, können Präpositionalphrasen also in den Funktionen von Objekten, Adverbialen und Attributen vorkommen. Sie können valenzgebunden und frei sein, wobei sich dieser Status auf die Wahlmöglichkeiten der Präposition, über die das folgende Teilkapitel weitere Informationen gibt, sowohl bei Attributen als auch Objekten auswirkt. Ihre Ersetzbarkeit kann im Fall ihrer Verwendung als Objekt eingeschränkt sein und bei adverbialer Nutzung kann jede Argumentrolle nur einmal vergeben werden (vgl. Zifonun et al. 1997: 2092).

2.6.3. Bedeutung der Präposition

Präpositionen bilden den Kopf von Präpositionalphrasen und stellen zweistellige Beziehungen zwischen Entitäten her (vgl. Eroms 1991: 40), die Nominalphrasen formal repräsentieren. Das unterscheidet sie von den satzver-

knüpfenden Konjunktionen und von den Infinitivphraseneinleitern, die Beziehungen zwischen Sätzen herstellen (vgl. Breindl 2006: 937f.).

Die Bindungskraft der Präposition wird von Eroms (1991: 41) als asymmetrisch beschrieben. Logisch-semantisch ist sie zweiteilig, syntaktisch jedoch nur einwertig (auch Breindl 2006: 937), da immer nur ein Komplement gebunden werden kann, dem die Präposition den Kasus zuweist. Die Präposition selbst kann von der Verbvalenz oder ihrer Eigensemantik bestimmt werden. Die zweiteilige logisch-semantische Verbindung besteht vor allem bei Präpositionen, die nicht von Verben regiert werden.³⁷ (31) verdeutlicht das.

(31) die Spinne [auf dem Tisch]_{PP, Dativ}

Syntaktisch kann die Präposition nur ein Komplement an sich binden und regieren. Wenn man versucht, das Beispiel mit einem zweiten Komplement zu erweitern, zeigt sich, dass das in einer ungrammatischen Phrase resultiert (32). Daher ist die syntaktische Relation einwertig.

(32) * die Spinne [auf dem Tisch dem Stuhl]_{PP}

Logisch-semantisch besteht eine zweiwertige Beziehung zwischen der *Spinne* und dem *Tisch*. Durch die Präposition wird verdeutlicht, wo sich die Spinne befindet. Die Relation besteht sowohl zwischen dem *Tisch* als auch der *Spinne*. Die Bedeutung, die die Präposition denotiert, ist eine räumliche, die in (31) auch durch den Kasus angegeben ist.

Die verschiedenen Bedeutungen der Präpositionen können durch einen Merkmalskatalog bestimmt werden (vgl. Fries 1991). Sie können unter anderem räumlich, zeitlich und kausal sein (vgl. Eroms 1991). Ausschlaggebend für die Bedeutung ist das Aufzeigen von Relationen (vgl. Eroms 1991: 40). Das gilt allerdings nur für Präpositionen, bei denen die semantische Bedeutung noch voll erhalten ist. Nübling (2016: 603) beschreibt auch „neutrale (leere)

³⁷ Während die Funktion eines Kasus, der von einem Verb selektiert wird, einfach bestimmt werden kann (vgl. Brinkmann 1971: 403ff.), ist das für von Präpositionen regierte Kasus schwieriger. Der Akkusativ gilt beispielsweise als „Zielkasus bzw. Kasus der Alterität“ und der Dativ als „Personenkasus“ (Eroms 1991: 40). Dem Genitiv ist keine solche Beziehung zugeordnet (Eroms 1991: 46).

Präpositionen“, wenn die Präposition von einem Verb oder Adjektiv gefordert wird, aber bereits desemantisiert wurde.

Eine Präposition kann verschiedene Bedeutungen haben, also homonym sein. Deutlich wird das beispielsweise daran, dass der Wechsel des mit der Präposition verbundenen Kasus auch den Wechsel zwischen einer statischen Position und einer direktionalen anzeigen kann (vgl. Nübling 2016: 604). In (31) befindet sich die Spinne statisch auf dem Tisch. Kombiniert man den Kasus der Nominalphrase in der Präpositionalphrase mit einem Akkusativ, verändert sich die Bedeutung der Präposition dahingehend, dass Bewegung impliziert wird (33).

(33) die Wanderung [auf den Berg]_{PPAkkusativ}

Die vorherigen Abschnitte implizieren das allgemeine Wissen, dass Präpositionen eigentlich nicht allein stehen können. Eroms (1991: 41) beschreibt zwar, dass Präpositionen auch ohne ein Komplement vollständig und eigenständig sein können, dieses Vorkommen aber stark kontextabhängig ist und nur in kommunikativen Situationen vorkommen kann. Aufgrund der eingeschränkten Verwendbarkeit werden sie in dieser Arbeit ausgeklammert.

2.6.4. Entwicklung der Präpositionalphrase

Präpositionalphrasen sind seit dem Althochdeutschen nachweisbar (vgl. z. B. Fries 1991). Trotzdem ist die Forschungslage über die Form und Veränderungen der Präpositionalphrasen eher spärlich. Exemplarisch soll hier Waldenberger (2009) aufgeführt werden, die sich mit Präpositionalphrasen im Mittelhochdeutschen befasst. Unter anderem gibt sie einen Überblick über die möglichen Erweiterungen der Präpositionalphrase, die sich auf deren Komplexität auswirken (vgl. Waldenberger 2009: 68ff.).³⁸ Schon im Mittelhochdeutschen bestehen Präpositionalphrasen aus einer Präposition und einem Komplement, das von ihr regiert wird. Mehr als 90 % der Präpositionalphrasen im Korpus von Waldenberger (2009) setzen sich aus weniger als vier Elementen zusam-

³⁸ Daneben fokussiert Waldenberger (2009: 95ff.) sich auf die semantische Bedeutung der Präpositionalphrase, was hier weitestgehend ausgeklammert wird.

men, wobei in Verstexten kürzere und in Prosa- oder Urkundentexten längere Präpositionalphrasen vorkommen, es also einen Unterschied zwischen den Textsorten gibt (vgl. Waldenberger 2009: 72).

In einem Drittel der Fälle werden Präpositionalphrasen mit einem Artikel und einem Substantiv kombiniert (vgl. Waldenberger 2009: 75), in einem Viertel regiert die Präposition nur eine Nominalphrase ohne Artikel, in sieben Prozent der gefundenen Präpositionalphrasen nimmt die Nominalphrase ein Adjektivattribut und einen Artikel zu sich und in sechs Prozent der Fälle tritt die gleiche Kombination ohne Determinierer auf (vgl. Waldenberger 2009: 76). Die restlichen Belege teilen sich in Präpositionalphrasen, deren nominale Komplemente durch mehrere Adjektive erweitert werden, und solche mit erweiterten Adjektivattributen auf (vgl. Waldenberger 2009: 79f.). Weniger als ein Prozent der Präpositionalphrasen haben als Komplement eine Adverbialphrase (vgl. Waldenberger 2009: 82). Zudem findet Waldenberger (2009: 91) vor allem die auch im Neuhochdeutschen etablierte Rektionsrichtung nach rechts, wodurch „die in der Forschung häufig in den Vordergrund gerückten [...] strukturellen Besonderheiten wie Linksversetzungen als randständige Phänomene erscheinen“.

Im Frühneuhochdeutschen ist schließlich die Entstehung neuer Präpositionen zu beobachten, was nach Erben (2000a: 1584f.) der zunehmenden grafischen Rezeption des Deutschen geschuldet sei. Neben diesen neuen Präpositionen lässt sich auch ein veränderter Gebrauch von Präpositionalphrasen beobachten. Vor allem beginnen sie sowohl Genitivobjekte als auch adnominale Genitive zu ersetzen (vgl. Erben 2000a: 1587, 1589). Ersteres wird von Fischer (1987) in Zusammenhang mit der Textsorte beziehungsweise deren Funktion gebracht. Wenn die Texte für ein breiteres Publikum verfasst werden, werden eher Präpositionalobjekte als Genitivobjekte verwendet (vgl. Fischer 1987: 320f.), was dafür spricht, dass durch Präpositionen ausgedrückte Relationen einfacher zu verstehen sind.

Diese Entwicklungstendenzen setzen sich im Zeitabschnitt von 1650 bis 1900 fort. Unter anderem werden auch hier präpositionale Attribute für die Erweiterung von Nominalphrasen verwendet (vgl. Erben 2000b: 1860f.). Auf

Satzebene treten ab dem 18. Jahrhundert die Agentien der Passivbildung in Form von Präpositionalphrasen mit der Präposition *durch* beziehungsweise *von* auf (vgl. Erben 2000b: 1862), wobei die Wahl der Präposition nicht mehr von den Belebtheitsmerkmalen des regierten Nomens abhängt (vgl. Erben 2000b: 1863). Auch der bereits im Frühneuhochdeutschen beginnende Schwund des Genitivobjekts und dessen Ersetzung durch Präpositional- und Akkusativobjekte setzt sich fort (vgl. Erben 2000b: 1870f.).

Insgesamt sei die Forschungslage zur Verwendung von Präpositionalphrasen über die Zeit hinweg laut Erben (2000b: 1870ff.) eher ein Desiderat und meistens nur im Kontext zu anderen Phänomenen wie dem Genitivschwund zu finden. Ob sich die Komplexität der Präpositionalphrasen ändert, wurde bisher nicht überprüft, was insofern bedauerlich ist, da Extraposition oft mit Länge in Verbindung gebracht wird (Kapitel 3).³⁹

2.7. Mündlichkeit und Schriftlichkeit

Das Thema der Nachfeldbesetzung wird aber nicht nur im Zusammenhang mit der Länge des ausgelagerten Materials betrachtet, sondern auch im Zusammenhang mit gesprochener Sprache (vgl. z. B. Averintseva-Klisch 2007; Imo 2015). Die Notwendigkeit von Auslagerung wird mit kognitiven Verarbeitungsbeschränkungen begründet (vgl. Averintseva-Klisch 2007; Imo 2015; Ortman und Dipper 2019). Durch die flüchtige und unmittelbare Verarbeitung beziehungsweise Produktion des Gesagten werden hohe Anforderungen an das Arbeitsgedächtnis gestellt, die oft nicht erfüllt werden können. Dadurch werden beispielsweise Reparaturen oder Nachträge notwendig, die dann im Nachfeld zu finden sind (Kapitel 2.3). Auch die interaktionale Dimension des gesprochenen Deutsch wird immer wieder zur Begründung für Nachfeldbesetzungen angeführt (vgl. Selting 2012). So finden sich im gesprochenen Deutschen beispielsweise häufiger Rechtsversetzungen (vgl. Altmann 1981), die der Reparatur, der Spezifizierung einer pronominalen Referenz

³⁹ Ich verzichte an dieser Stelle darauf, auf mögliche Veränderungen der von den Präpositionen regierten Kasus hinzuweisen, weil die vorgenommenen Analysen auf Lemmata basiert, sodass Unterschiede im Kasus übergangen werden (Kapitel 5).

oder zur Vermeidung einer kognitiven Überlastung des Hörers dienen (vgl. Averintseva-Klisch 2007: 166). Generell scheint das Nachfeld in der gesprochenen Sprache permissiver zu sein als in der geschriebenen Sprache. Imo (2015) findet beispielsweise in seinem Korpus, dass Phrasen, die Rahmenbedingungen beschreiben, häufig im Nachfeld auftreten.

Die angeführten Studien erheben keinen Anspruch darauf, das Thema vollständig abzudecken. Doch sie erlauben zwei Beobachtungen: zum einen die Relevanz, sich mit *Mündlichkeit* und *Schriftlichkeit* auseinanderzusetzen, wenn man sich mit Nachfeldbesetzung beschäftigt; zum anderen die Vermischung von zwei Konzepten. Bei den erwähnten Studien von Averintseva-Klisch (2007) und Imo (2015) wurden *Gespräche* als Untersuchungsgegenstand genommen. Das bezieht sich auf die Darbietungsform oder das *Medium*. Der flüchtige, das Arbeitsgedächtnis stark fordernde Charakter eines Gesprächs ist offensichtlich und aus den Untersuchungen wird deutlich, dass es für diese ‚Textform‘ Abweichungen von der Standardsprache gibt, die untersucht werden können. Diese Abweichungen beziehen sich auf die *Konzeption* des Gesprächs. Es bestehen also zwei Pole: Medium und Konzeption.

Dass diese beiden Pole nicht immer deckungsgleich sein müssen, legen Koch und Oesterreicher (2007) dar. Diese Unterscheidung ermöglicht eine Klassifizierung von Merkmalen des gesprochenen Deutsch oder der Mündlichkeit in Texten, die nicht nur in der flüchtigen Darbietungsform eines Gesprächs vorliegen müssen. Man findet sie zum Beispiel auch bei Texten, die in einem geschriebenen Medium vorliegen, zum Beispiel bei einem Tagebucheintrag oder einem Brief. Das macht Unterscheidung zwischen den Polen auch anwendbar auf historische Daten.

Die Idee hinter der Unterscheidung der Konzeption von Mündlichkeit und Schriftlichkeit von Koch und Oesterreicher (2007) ist an folgendem Beispiel zu verdeutlichen: Ein Chat ist in schriftlicher Form präsentiert, doch die Produktion der Nachrichten erfolgt unter einem ähnlichen Zeitdruck und unter ähnlichen Ansprüchen an das Arbeitsgedächtnis wie ein orales Gespräch und kann daher ähnliche sprachliche Charakteristika aufweisen wie zum Beispiel Reparaturen oder Rechtsversetzungen. Es kann also sein, dass Texte,

obwohl sie in geschriebener Form vorliegen, mündlich gestaltet sind. Koch und Oesterreicher (2007) bezeichnen das als *konzeptionelle Mündlichkeit* beziehungsweise *konzeptionelle Schriftlichkeit*.

Ein Text kann zwischen vier Polen verortet sein. Zum einen gibt es die Pole des Mediums (vgl. Koch und Oesterreicher 2007: 348). Sie beschreiben, wie der Name schon sagt, in welchem Medium ein Text vorkommen kann. Unterschieden wird zwischen *graphischen* und *phonischen* Texten. Diese Differenzierung ist binär. Es gibt keine Zwischenschritte.

Das ist bei den beiden anderen Polen anders. Diese bezeichnen die Konzeption des Textes und werden von Koch und Oesterreicher (2007: 348) als *gesprochen* versus *geschrieben* bezeichnet: „[H]ier geht es also – unabhängig von der medialen Realisierung – um varietätenbezogene und diskurspragmatische Optionen im sprachlichen Ausdruck (einschließlich der entsprechenden rezipientenseitigen Erwartungen).“

Konkret bedeutet das, dass ein Text Charakteristika enthalten kann, die entweder typisch für gesprochene Sprache oder für geschriebene Sprache sein können. Er kann sich auf einer Art Skala zwischen den beiden Punkten ‚geschrieben/gesprochen‘ bewegen. Wo genau er auf dieser Skala liegt, ist davon abhängig, welche Charakteristika überwiegen.

Dabei besteht ein gewisser Zusammenhang zwischen der Realisierung als Medium und der Konzeption. Graphisch Realisiertes tendiert eher dazu, auch schriftlich konzipiert zu sein, und phonisch Realisiertes eher mündlich (vgl. Koch und Oesterreicher 2007: 348). Schon durch Alltagserfahrungen ist aber festzustellen, dass ein Wechsel des Mediums verhältnismäßig einfach möglich ist. Gesprächsprotokolle, in denen erwartungsmäßig mündlichkeitsnahe Texte aus dem Phonischen ins Graphische übertragen werden, sind ein Beispiel dafür. Die Übertragung beziehungsweise Übersetzung von einer Konzeption in die andere ist hingegen schwerer möglich und auch nicht Gegenstand der Theorie von Koch und Oesterreicher (2007). Stattdessen bemühen sie sich darum, die Konzeption und damit die Klassifizierbarkeit von Texten als mündlichkeits- beziehungsweise schriftlichkeitsnah beschreibbar zu machen. Hier liegt allerdings auch die größte Schwachstelle des sonst zugängli-

chen Konzepts der Autoren, wie Ágel und Henning (2012) darlegen.

Obwohl es auch zum Zeitpunkt, zu dem Koch und Oesterreicher (2007) ihre Theorie entwickelt haben, schon ausführliche Literatur zur Grammatik der gesprochenen Sprache gab, verzichteten sie in ihrem Konzept auf linguistische Parameter zur Klassifikation der Konzeption. Stattdessen teilen Koch und Oesterreicher (2007: 351) Mündlichkeitsnähe ein durch „Privatheit[,] Vertrautheit der Kommunikationspartner[,] starke emotionale Beteiligung[,] Situations- und Handlungseinbindung[,] referenzielle Nähe[,] raumzeitliche Nähe (*face-to-face*)[,] kommunikative Kooperation[,] Dialogizität, Spontanität[,] freie Themenentwicklung usw.“, wohingegen Schriftlichkeitsnähe durch „Öffentlichkeit[,] Fremdheit der Kommunikationspartner[,] geringe emotionale Beteiligung[,] Situations- und Handlungsentbindung[,] referenzielle Distanz[,] raumzeitliche Distanz[,] keine kommunikative Kooperation[,] Monologizität[,] Reflektiertheit[,] Themenfixierung“ gekennzeichnet ist (Koch und Oesterreicher 2007: 351). Daraus ergibt sich, dass ein Gespräch (unter Freunden) beispielsweise höchst mündlichkeitsnah ist, eine wissenschaftliche Publikation allerdings sehr schriftlichkeitsnah. Dieses intuitive Beispiel zeigt, dass die Pole nach Koch und Oesterreicher (2007) auch an bestimmte Gattungen und Textsorten geknüpft sind, die als mehr oder weniger prototypisch gelten können.

Ein scheinbarer Vorteil dieser Klassifizierung liegt darin, dass die Texte durch einmaliges Lesen als mehr oder weniger mündlichkeitsnah bestimmt werden können. Zudem sind sie ohne linguistische Kenntnis anwendbar. Eine Gewichtung einzelner Faktoren geben Koch und Oesterreicher (2007) aber nicht an und gehen auch nicht auf ein weiteres Problem der Klassifizierung anhand ihrer Kriterien ein, das jedoch offenkundig wird, wenn man sich ihre Vorschläge zur Anwendbarkeit in der historischen Linguistik anschaut. Sie beschreiben eine Entwicklung von mündlichen zu schriftlichen Texten. Dabei beziehen sie sich unter anderem auf die „Verschriftlichung von Idiomen“ und die Standardisierung als „Aus- und Abwahl der einzelsprachlichen Mittel im Distanzbereich“ (Koch und Oesterreicher 2007: 364). Sie implizieren damit, dass die Charakterisierung nicht nur auf den situativen Aspekten wie räumli-

cher und zeitlicher Nähe basiert, sondern sich auch in der Syntax niederschlagen kann. Allerdings machen sie diesbezüglich keinerlei weitere Angaben. Aus den genannten Kriterien kann sich auch die „erhebliche sprachliche Diskrepanz zwischen Nähe- und Distanzbereich“ (Koch und Oesterreicher 2007: 365) nicht allein ergeben. Diese Diskrepanz konstituiert sich in beobachtbaren, sprachlichen Mustern wie einer gehäuften Verwendung des Perfekts im Gesprochenen Deutschen (vgl. z. B. Schwitalla 2006: 136) oder der Nachfeldbesetzung durch Herausstellungen im Sinne von Altmann (1981) und Repp und Struckmeier (2020), die in Kapitel 2.3 beschrieben wurden.

Somit ist die Kritik, die Ágel und Henning (2012: 181ff.) formulieren, nicht unberechtigt. Sie zielen in ihrer Arbeit darauf ab, Mündlichkeits- und Schriftlichkeitsnähe kategorisierbar zu machen. Allerdings beziehen auch sie sich stark auf Kriterien, die im vorherigen Abschnitt situationsbezogen genannt wurden. Mündlichkeitsnahe Kommunikation, die von Ágel und Henning (2012: 185) als „Nähesprechen“ bezeichnet wird, definieren sie als Kommunikation, in der die Rollen von Sprechenden und Rezipierenden vertauschbar sind, was bei Distanzsprache, sprich schriftlichkeitsnaher Kommunikation, nicht möglich ist. Diese Unterscheidung bezeichnen sie als „Rollenparameter“ (Ágel und Henning 2012: 186). Daneben definieren sie „Zeitparameter“, die bestimmen, ob der Zeitpunkt der Äußerung dem der Planung entspricht, „Situationsparameter“, „Parameter des Codes“ und „Parameter des Mediums“ (Ágel und Henning 2012: 186). Letzteres ist mit der Distinktion zwischen phonischem und graphischem Medium nach Koch und Oesterreicher (2007) identisch.

Auch die Situationsparameter ähneln denen von Koch und Oesterreicher (2007). Aus linguistischer Perspektive interessanter ist aber der Parameter des Codes, da hier beispielsweise die Variabilität der Deixis Erwähnung findet. Beim Nähesprechen, vor allem bei face-to-face-Kommunikation, können andere Mittel der Deixis verwendet werden als beim Distanzsprechen, wenn sich die Kommunikationsteilnehmenden nicht zur gleichen Zeit an einem Ort aufhalten. Neben der Deixis nutzen Ágel und Henning (2012) weitere linguistische Parameter zur Unterscheidung. So wird schriftlichkeitsnahe Kommu-

nikation über „wohlgeformte Strukturen, komplexe Strukturen, längere Diskurseinheiten, hypotaktischere Diskurse, komplexere Hypotaxen, Nebensätze [und] anhängige Nebensätze“ (Ágel und Henning 2012: 191) klassifiziert. Für mündlichkeitsnahe Kommunikation gelten die jeweiligen Gegenteilte.

Die Nutzung linguistischer Kriterien ist bereits ein erheblicher Vorteil gegenüber Koch und Oesterreicher (2007). Dennoch ist es nach dem Konzept von Ágel und Henning (2012) notwendig, die Texte manuell zu klassifizieren und sich auf die Beschreibung von Experten zu verlassen. Das birgt das Risiko des Objektivitätsverlusts und ist zudem, wie Ortmann und Dipper (2020: 1293) anmerken, nicht einfach auf große Datenmengen anzuwenden.

Ortmann und Dipper (2019) entwickeln daher eine Methode zur automatischen Bestimmung der Mündlichkeitsnähe für moderne Daten und übertragen diese in Ortmann und Dipper (2020) auch auf historische Daten, was für die vorliegende Untersuchung von besonderer Relevanz ist. Als Merkmale ziehen Ortmann und Dipper (2019: 66f.) folgende heran: Es wird angenommen, dass in mündlichkeitsnahen Texten mehr Pronomen und Demonstrativpronomen zu finden sind (z. B. Schwitalla und Tiittula 2009). Begründen lässt sich das beispielsweise mit veränderten Möglichkeiten der Deixis im mündlichkeitsnahen und vor allem phonischen Diskurs. Für letzteren ist die Voraussetzung, dass sich beide Gesprächsparteien räumlich und zeitlich am gleichen Ort aufhalten. Das wiederum bedeutet, dass sie auch die gleichen Entitäten in ihrem Umfeld sehen und es entsprechend leichter möglich ist, darauf mit Demonstrativpronomen zu verweisen, was hingegen in einem schriftlichkeitsnahen, graphischen Diskurs praktisch nur textintern möglich ist. Für eine vermehrte Verwendung von Pronomen spricht eine größere Themenkonstanz im mündlichkeitsnahen Dialog. Außerdem ist die Komplexität der Sätze in der mündlichen Konzeption geringer als bei schriftlichkeitsnahen Texten, weil mündlichkeitsnahe Texte häufiger unmittelbar verarbeitet werden. Es muss angemerkt werden, dass diese Begründung wie eine Mischung aus Konzept und Medium wirkt (vgl. Koch und Oesterreicher 2007). Der flüchtige Charakter eines Textes kommt nur bei der phonischen Realisierung zum Tragen. Bei graphischer Realisierung hingegen können Passagen

beliebig häufig aufgenommen werden. Allerdings zeigen diverse Studien (vgl. z. B. Tomczyk-Popínska 1987; Richter 1985; Rehm 2002; Bader 2002), dass in mündlichkeitsnahen Texten sowohl Sätze als auch Wörter weniger komplex sind. Außerdem ist die Zahl der Subordinationen hier geringer als die der Koordinationen (vgl. Ortmann und Dipper 2019: 67). Für schriftlichkeitsnahe Texte ist der Nominalstil und ein höherer Anteil an Inhaltswörter typisch (vgl. Ortmann und Dipper 2019).

Auch hinsichtlich der Syntax gibt es Unterschiede. Im mündlichkeitsnahen Deutsch gibt es mehr Ellipsen, das bevorzugte Genus Verbi ist das Aktiv und es kommen mehr analytische als syntaktische Verbformen vor (vgl. Ortmann und Dipper 2019: 66). Auch Konstruktionen, die laut der Schulgrammatik im Deutschen unmöglich sind, lassen sich im gesprochenen Deutschen finden, beispielsweise das *am*-Progressiv (vgl. Ortmann und Dipper 2019: 67). Im gesprochenen Deutsch gibt es zudem mehr Partikeln, darunter auch Antwort- und Modalpartikeln und Interjektionen als im geschriebenen Deutschen (Ortmann und Dipper 2019: 67). Schriftlichkeitsnahe Kommunikation zeigt dafür die größere Variation an Wortschatz und Gestaltung (vgl. Ortmann und Dipper 2019: 67). Diese Kriterien erweisen sich auch als nützlich für die automatische Bestimmung der Mündlichkeit eines Textes und lassen sich auch auf historische Texte übertragen (Ortmann und Dipper 2020).

Darauf soll im Detail eingegangen werden, da die von Ortmann und Dipper (2020) vorgestellte Analysemethodik die Grundlage für die Zuordnung des *Orality Scores* bildet, der auch in dieser Arbeit genutzt wird.⁴⁰ Dieser Score ordnet Texten objektiv einen Wert zu, der klassifiziert, wie ausgeprägt die Merkmale sind, die der mündlichen Konzeption zugeordnet sind. Anhand dessen können innerhalb eines Korpus⁴¹ Aussagen getroffen werden, ob ein Text im Vergleich zu einem anderen mündlich- oder schriftlichkeitsnäher ist.

Um die Methodik auch auf historische Daten anwenden zu können, tes-

⁴⁰ Hegendüs (2012) bemüht sich beispielsweise auch um eine objektivere, an eine Maßzahl gebundene Klassifizierung von historischen Texten auf einer Skala der Mündlichkeit, allerdings ist sein Ansatz wie auch der von Ágel und Henning (2012), auf den er sich stark bezieht, nicht auf große Textmengen anwendbar.

⁴¹ Da es keine definierbaren Endpunkte der Skala der Mündlichkeit und Schriftlichkeit gibt, ist es nur möglich, einzelne Texte hinsichtlich ihres Verhältnisses zueinander zu klassifizieren (vgl. Ortmann und Dipper 2024).

ten Ortmann und Dipper (2020: 1293) die Methode an vier Textsorten aus dem Deutschen Textarchiv („DTA“, vgl. BBAW 2019), genauer gesagt an wissenschaftlichen Texten, Zeitungsartikel, Erzählungen und Predigten von 1700 bis 1900. Zunächst wurde basierend auf den an Koch und Oesterreicher (2007) und Ágel und Henning (2012) angelehnten Kriterien eine prototypische Zuordnung zur Konzeption durchgeführt. So wurden wissenschaftliche Texte aufgrund ihres spezifischen Themas und dem eingeschränkten Rezipientenkreis am ehesten als schriftlichkeitsnah angesehen, gefolgt von Zeitungsartikeln, die aufgrund der breiteren Rezipientenbasis einen Schritt Richtung Mündlichkeitsnähe machen, aber immer noch am schriftlichkeitsnahen Pol zu verorten sind (vgl. Ortmann und Dipper 2020: 1294). Als mündlichkeitsnäher werden Erzählung wegen des Wechsels aus dialogischen und narrativen Passagen empfunden. Innerhalb dieses Korpus gelten die Predigten als am mündlichkeitsnächsten, da sie für die auditive Rezeption gestaltet wurden, obwohl angemerkt werden muss, dass sie vor dem Druck noch aufbereitet und verändert wurden (vgl. Ortmann und Dipper 2020: 1294).

Aus diesen Texten wurden pro 50-Jahres-Abschnitt 1000 Tokens analysiert, um auch die aus anderen Sprachen bekannte Veränderung über die Jahrhunderte hinweg betrachten zu können. Diese Teilkorpora wurden nach den folgenden Kriterien untersucht (vgl. Ortmann und Dipper 2020: 1295):

- Durchschnittliche Satzlänge und Median der Satzlänge
- Durchschnittliche Wortlänge und Median der Wortlänge
- Verhältnis von Subjunktionen zu Vollverben
- Anteil der Sätze, die mit einer koordinierenden Konjunktion beginnen
- Verhältnis von Vollverben zu Nomen
- Anteil der Inhaltswörter an allen Wörtern
- Anteil von Pronomen der 1. Pers. Sg./Pl. an allen Wörtern
- Anteil der Demonstrativpronomen an allen Wörtern
- Anteil von Demonstrativpronomen mit Lemmata wie „diese“ oder „die“, inklusive der Kurzformen, deren Lemma mit „die“ angegeben ist.
- Anteil der Antwortpartikel an allen Wörtern
- Verhältnis von ein-Wort-Interjektionen an allen Wörtern

Für Schriftlichkeitsnähe sprechen ein hoher Anteil beziehungsweise ein hohes Verhältnis der Faktoren Satzlänge, Wortlänge und Subjunktionen, die sich auf Komplexität beziehen, sowie der Nominalstil, der über das Verhältnis von Nomen zu Vollverben bestimmt wird, wobei ein hoher Anteil von Nomen den Nominalstil klassifiziert. Ähnliches gilt für den Anteil der Inhaltswörter – auch ein gehäuftes Vorkommen von ihnen deutet auf Schriftlichkeitsnähe hin –, wohingegen sowohl ein großer Anteil an Pronomen als auch an Demonstrativpronomen, unabhängig davon, ob sie gekürzt sind oder nicht, für Mündlichkeitsnähe spricht. Beide Pronomenarten weisen auf eine Deixis hin, die typisch für die face-to-face-Kommunikation von Sprechenden und Rezipierenden ist. Ebenfalls einen Hinweis auf die Nähe zur Mündlichkeit liefern die Anteile von Antwortpartikeln und Ein-Wort-Interjektionen.

Ortmann und Dipper (2020: 1295) finden einen Zusammenhang zwischen Wortlänge und Schriftlichkeit sowie zwischen den Personal- und Demonstrativpronomen und der Mündlichkeit. Auch die restlichen oben genannten Features deuten auf Unterschiede und die Nutzbarkeit für die Bestimmung des Grads der Mündlichkeit hin und finden schließlich auch Niederschlag in der Berechnung des Orality Scores (Kapitel 5.3).

Vom 18. bis 20. Jahrhundert können sie zeigen, dass der Nominalstil für wissenschaftliche Texte typisch wird und sich die Satzlänge im Laufe der Zeit verkürzt (vgl. Ortmann und Dipper 2020: 1297). Auch der Stil der Zeitungsartikel wird nominaler und somit schriftlichkeitsnäher, was im Kontrast zu den narrativen Texten steht. Hier werden Wörter und Sätze kürzer, die Zahl der Subordinationen nimmt ab, während Fragen und Ausrufe mehr werden, was für eine Hinwendung zum mündlichkeitsnahen Stil spricht (vgl. Ortmann und Dipper 2020: 1297). Für die Predigten ist keine klare Tendenz zu erkennen (vgl. Ortmann und Dipper 2020: 1297).

Durch die Konzentration auf die oben genannten Charakteristika aus Ortmann und Dipper (2019; 2020) können auch innerhalb eines Genres und innerhalb der gleichen situativen Parameter Hinwendungen zu der einen oder anderen Konzeption beschrieben werden. Würde man ausschließlich nach der Klassifikation von Koch und Oesterreicher (2007: 351) gehen, so wären bei-

spielsweise alle wissenschaftlichen Texte in gleichem Maße schriftlichkeitsnah konzipiert. Denn sie weisen immer Öffentlichkeit durch den Druck als Buch oder Artikel auf, wodurch die Fremdheit der Kommunikationspartner gewährleistet ist, was auch die Situations- und Handlungsentbindung, referentielle Distanz und mangelnde kommunikative Kooperation und Themenfixierung bedingt. Doch sogar Koch und Oesterreicher (2007: 364) bedenken, dass Standardisierungen, also die „Aus- und Abwahl der einzelsprachlichen Mittel im Distanzbereich“, möglich ist und die Integration linguistischer Charakteristika in die Bestimmung umso notwendiger macht. Die von Ortman und Dipper (2020) beobachtete Veränderung innerhalb eines Genres unterstützt die Integration von Merkmalen der einen Konzeption in die andere (vgl. Koch und Oesterreicher 2007). Diese Übertragung kommt insbesondere im Bereich der historischen Linguistik zum Tragen.

Ein weiterer Vorteil der Arbeit von Ortman und Dipper (2019; 2020) ist folgender: Die herangezogenen Charakteristika sind unabhängig von Phänomenen, die als typisch für das gesprochene Deutsche gelten. Beispielsweise werden hier die Verbzweitstellung in *weil*-Sätzen (vgl. Selting 2012) oder Auslagerungsphänomene angeführt. Durch den Verzicht auf solche syntaktischen Features lässt sich objektiver beschreiben, ob diese wirklich ein Phänomen der mündlichen oder schriftlichen Konzeption sind. Auch Probleme der Zirkularität werden dadurch vermieden, dass sich nicht auf diese Phänomene konzentriert wird. Ein Text, in dem sie (vermeintlich) überdurchschnittlich häufig vorkommen, kann gemäß seiner Mündlichkeitsnähe beschrieben werden. Wenn das überdurchschnittliche Vorkommen von beispielsweise Extraposition mit einer hohen Übereinstimmung sonstiger Mündlichkeitsmerkmale zusammenfällt, kann es tatsächlich als weiteres Merkmal aufgefasst werden.

3. Extraposition von Phrasen und Relativsätzen

Während in Kapitel 2.4, 2.5 und 2.6 Relativsätze, Nominal- und Präpositionalphrasen beschrieben und in Kapitel 2.3 die hier verwendete Definition von Extraposition dargestellt wurde, bringt dieses Kapitel beides zusammen. Anhand von Grammatiken und empirischen Untersuchungen wird geschildert, welche Faktoren nach dem aktuellen Forschungsstand für die Auslagerung von Phrasen und Relativsätzen verantwortlich gemacht werden. Die Darstellung der empirischen Forschung konzentriert sich dabei auf ausgewählte Werke, die sich auf die Informationsstruktur oder Verarbeitungsherausforderungen beziehen. Aufgrund der eher spärlichen Forschungslage werden Nominal- und Präpositionalphrasen dabei zusammengefasst.

3.1. Die Extraposition von PPs und NPs

3.1.1. Grammatiken

Zifonun et al. (1997: 1657) führen Präpositionalphrasen als diejenigen Phrasen an, die im heutigen Deutsch am häufigsten im Nachfeld stehen.¹ Ihre Positionierung an dieser Stelle ist nicht der Normalfall. Sie sind viel häufiger im Vor- oder Mittelfeld zu finden, wo sie im Allgemeinen als natürlicher empfunden werden (vgl. Weber 2019). Ihre Extraposition muss also durch andere Faktoren wie die Gewährleistung erfolgreicher Kommunikation motiviert sein (vgl. z. B. Zifonun et al. 1997; Wöllstein 2014).

¹ Teile des in diesem Kapitel vorgestellten Materials flossen auch in die Veröffentlichung Voigtmann und Speyer (2023) ein.

Extraposition verkürzt die Distanz zwischen den Satzklammern. Das ist insofern relevant, als beispielsweise Gibson (1998) und Futrell et al. (2021) argumentieren, dass eine Vergrößerung des Abstands zwischen zusammengehörendem Sprachmaterial auch die Wahrscheinlichkeit erhöht, den ersten Teil dieses Materials aufgrund der größeren Belastung des Gedächtnisses zu vergessen. Daraus resultiert schließlich Informationsverlust und ein mögliches Scheitern der Kommunikation (Kapitel 2.1). Daher heißt es, dass vor allem lange Phrasen ausgelagert werden.

Der Faktor *Länge* wird oft synonym mit Komplexität verwendet. Je mehr Wörter eine Phrase enthält, desto wahrscheinlicher ist es, dass sie ins Nachfeld verschoben wird. Die Zunahme der Länge kann durch Koordinationen oder (komplexe) Attribute erfolgen (1).

- (1) a. Ich [habe]_{LSK} **auf dem Bahnhof, wo wir uns damals getroffen haben, als du zum ersten Mal in meiner Stadt warst**, gestern lange auf dich [gewartet]_{RSK}.
- b. Ich [habe]_{LSK} gestern lange auf dich [gewartet]_{RSK} **auf dem Bahnhof, wo wir uns damals kennengelernt haben, als du zum ersten Mal in meiner Stadt warst**.

In (1a) ist der Abstand zwischen der linken und der rechten Satzklammer länger als in Beispiel (1b). Durch die Extraposition in (1b) wird das Mittelfeld hingegen so kompakt wie möglich gehalten. Dies gilt insbesondere für die mündliche Kommunikation (vgl. z. B. Beneš 1979; Zahn 1991; Brooks 2006), wo Informationen aufgrund der Flüchtigkeit bloß ein einziges Mal zur Verfügung stehen und bei Nichtverstehen mühsam nachgefragt werden müssen. Damit kann auch das häufigere Vorkommen von Extraposition in mündlichkeitsnahen Texten zusammenhängen. In schriftlicher Kommunikation hingegen stehen die Informationen länger zur Verfügung; sie können erneut gelesen werden, wenn das Verständnis andernfalls behindert würde.

Sowohl in der grafischen als auch in der oralen Kommunikation steht die Extraposition zwischen zwei unterschiedlichen Arten von Distanz. Wie oben gezeigt, besteht ein Ziel der Extraposition darin, den Abstand zwischen

den Satzklammern zu verkürzen, aber auch den Abstand zwischen der möglichen Position im Mittelfeld, an der die Phrase sonst stehen würde, und dem Nachfeld so klein wie möglich zu halten. Das Hauptziel dahinter ist die Vermeidung von Informationsverlusten.

Diese Gründe haben sich in der Geschichte des Deutschen nicht wesentlich geändert. Da der Fokus auf dem Frühneuhochdeutschen und dem frühen Neuhochdeutschen liegt, wird nur ein kurzer Überblick über die Extraposition im Althochdeutschen und im Mittelhochdeutschen gegeben.

Auch wenn Hinterhölzl (2010) und Borter (1982) Extraposition im Althochdeutschen beschreiben und Korpusbelege finden, muss klargestellt werden, dass die Verbzweitstellung im Althochdeutschen noch nicht etabliert war und Objekt-Verb- und Verb-Objekt-Linearisierungen erlaubt und unmarkiert waren (vgl. Hinterhölzl 2010: 128). Dennoch sind Tendenzen einer VO-Abfolge erkennbar, die mit Gewichts- und Informationsaspekten des jeweiligen Objekts zusammenhängen. Hinterhölzl (2010) findet Unterschiede zwischen der Extraposition aus Haupt- und Nebensätzen. Extraposition erfolgt häufiger aus Hauptsätzen heraus.

Für Notkers Psalmen, die vor 1020 entstanden sind, beschreibt Borter (1982) eine Korrelation zwischen der Position einer Phrase im lateinischen Original und der Extraposition in der deutschen Übersetzung. Darüber hinaus stellt er in Übereinstimmung mit Belegen aus dem Gegenwartsdeutschen und dem Mittelhochdeutschen fest, dass Phrasen eher extrapponiert werden, wenn sie durch einen Relativsatz erweitert werden (vgl. Borter 1982: 117ff.).

Bereits die Hälfte aller Sätze im Mittelhochdeutschen weisen die rechte Satzklammer als Begrenzungszeichen auf (vgl. Paul 2007), sodass der Satzrahmen schon zu dieser Zeit als evident, aber als noch nicht vollständig etabliert gilt (vgl. Bergmann, Moulin und Ruge 2019). Schon damals wurden Präpositionalphrasen vermehrt bei langen Matrixsätzen (vgl. Morlicchio 1991) und in Fällen, in denen eine Phrase durch ein sententielles Attribut, beispielsweise einen Relativsatz (vgl. Paul 2007: 460) eingeleitet wurde, extrapositioniert. Paul (2007: 460) beschreibt auch, dass die Linearisierung von infinitem vor finitem Material in der rechten Satzklammer mit der Extraposition verbun-

den ist. Außerdem findet er in Hauptsätzen höhere Extrapositionsraten als in Nebensätze (vgl. Paul 2007: 460).

Im Frühneuhochdeutschen gilt die Verbzweitstellung als voll etabliert. Eichinger (1995) bezieht daher außersprachliche Faktoren wie gesellschaftliche Veränderungen und höhere Anforderungen durch wissenschaftliche Entdeckungen sowie die Notwendigkeit, möglichst effizient zu kommunizieren, in die Betrachtung von Extraposition ein. Die Satzklammern lenkt zudem die Aufmerksamkeit des Rezipienten auf grammatische (linke Satzklammer) und lexikalische Informationen (rechte Satzklammer) (vgl. Dal und Eroms 2014: 202). Dal und Eroms (2014: 202) postulieren, dass dies mit der zunehmenden Verteilung von Informationen auf unabhängige Wörter seit dem Althochdeutschen zusammenhänge. Dennoch sei das Prinzip des ungebrochenen Satzrahmens ein Phänomen des schriftlichen und nicht des mündlichen Diskursmodus (vgl. Dal und Eroms 2014: 203). Die Extraposition von Phrasen war im Frühneuhochdeutschen sehr verbreitet, wie zum Beispiel Schildt (1976: 271ff.) zeigt, nimmt aber im Laufe dieser Periode sukzessive ab.

Ebert (1980) weist den Einfluss der Länge des Matrixsatzes und der extraponierten Phrase auf Auslagerung in seinem Korpus aus frühneuhochdeutschen Texten aus Nürnberg nach.² Auch Sapp (2014) bekräftigt diese Ergebnisse. Er stützt zudem den Befund von Zifonun et al. (1997: 1657), dass Präpositionalphrasen die am häufigsten extraponierten Phrasen sind. Sapp (2014) bestätigt auch die Unterscheidung zwischen konzeptionell mündlichen und schriftlichen Quellen. Extraposition findet sich häufiger in mündlichkeitsnahen Texten. Zusammenfassend lässt sich sagen, dass in Texten von 1350 bis 1650 in erster Linie grammatikalische und pragmatische Faktoren die Extraposition beeinflussen.

Dies wird auch von Grammatikern aus der späten, frühneuhochdeutschen Periode behauptet (vgl. Takada 1998). Ihre Konzeption der Stellungsregularitäten des deutschen Satzes ist stark mit der des topologischen Feldermodells vergleichbar (vgl. Takada 1998). Im Allgemeinen akzeptieren sie die Satzklammer als ein Hauptmerkmal des Deutschen, sind sich aber nicht

² Aufgrund der Beschränkung des Korpus' auf Nürnberg können die von Ebert (1980) getroffenen Aussagen nicht oder nur mit Vorsicht verallgemeinert werden.

darüber einig, aus welchen Gründen sie durchbrochen werden kann. Einerseits könne so die Bedeutung des Satzes beibehalten werden, die sonst aufgrund des zunehmenden Abstands zwischen linker und rechter Satzklammer verloren geht (vgl. Takada 1998: 229).³ Andererseits wird ein vollständiger, undurchbrochener Satzrahmen als prestigeträchtig bezeichnet (vgl. Takada 1998: 229) und aus soziolinguistischen Gründen angewendet (vgl. Takada 1998: 230).

Demske (2001: 275) befasst sich am Rand mit der Extraposition von (attributiven) Nominalphrasen im Frühneuhochdeutschen. Sie interpretiert solche Nominalphrasen postnominal und beschreibt, dass Extraposition meist aus prädikativ genutzten Nominalphrasen heraus geschieht und nicht nur aus Subjekten belegt ist (vgl. Demske 2001: 276). Zwar gebe es wenige Belege, „in denen eine attributive Präpositionalphrase entweder links oder rechts vom Kopfnomen erscheint“ (vgl. Demske 2001: 278), aber in den Fällen sei Extraposition im Gegensatz zu Extraktion ein wenig häufiger und vor allem bei Akkusativobjekten und Subjekten belegt. Vermehrt kommt Extraposition von Nominal- und Präpositionalphrasen dann vor, wenn diese gemeinsam mit ihrem Relativsatz ins Nachfeld geschoben werden (vgl. Demske 2001: 280). Demske (2001) nimmt allerdings keine Auszählungen von Korpusbelegen vor und beschreibt die Extraposition von Nominalphrasen nur exemplarisch.

Das frühe Neuhochdeutsche („1700 bis 1900“, Nübling et al. 2013: 6) erzeugt die Schwierigkeit, dass systematische Forschung sehr selten ist und auch Beschreibungen nicht häufig vorgenommen wurden. Eine Ausnahme ist Admoni (1990), der einen Überblick über die gesamte deutsche Sprache gibt, aber dabei eher unsystematisch vorgeht. In der Zeit von 1500 bis 1700 werde die Extraposition seltener und vor allem für „Nebensätze, Infinitivkonstruktionen, Satzglieder mit Vergleichssemantik, präpositionale Konstruktionen“ (Admoni 1990: 201) verwendet. Für das 19. Jahrhundert wird behauptet, dass die Satzstrukturen kürzer, einfacher und klarer werden (vgl. Admoni 1990: 220). Die Tendenz des Extrapositionsrückgangs scheint sich also fortzusetzen. Auch hier ist es wichtig festzustellen, dass Präpositionalphrasen zu den häu-

³ Es kann davon ausgegangen werden, dass hier auch das Prinzip der Memory Cost nach Gibson (1998) eine Rolle spielt.

figsten Phrasentypen für die Nachfeldstellung gehören. Dies wird auch von Konopka (1996) für das 18. Jahrhundert bestätigt. Dabei spielt es keine Rolle, ob die Präpositionalphrasen Adjunkte oder Argumente sind. Beide Arten können extrapониert werden (vgl. Konopka 1996: 167). Die von den Gelehrten aus dieser Zeit angeführten Gründe ähneln denen, die oben genannt wurden: Die Extraposition wird verwendet, um den Satzbau kompakter zu halten und dabei eine erfolgreiche Kommunikation zu gewährleisten (vgl. Konopka 1996: 178). Phrasen, die einen Relativsatz enthalten, sind häufiger im Nachfeld zu finden als nicht-erweiterte Phrasen. Darüber hinaus fasst er mehrere pragmatische Faktoren für die Extraposition zusammen. Zu ihnen zählen beispielsweise, weniger wichtige Informationen hinzuzufügen oder einen zweiten Schwerpunkt im Satz zu bilden, da wichtigere Informationen tendenziell später im Satz auftreten sollen (vgl. Konopka 1996: 171).⁴

Die Forschung über spätere Perioden ist noch spärlicher oder nicht auf die Extraposition ausgerichtet. Engel (1970: 55) fasst zusammen, dass es keine einheitliche Tendenz zur Extraposition gebe und diese zwischen den Textsorten variere.

3.1.2. Forschung

Synchrone Forschung Obwohl diese Arbeit das Phänomen der Extraposition im Englischen weitestgehend ausklammert (Kapitel 2.3), wird an dieser Stelle mit der Untersuchung von Arnold, Losongco, Wasow und Ginstrom (2000) begonnen, weil sich deren Arbeit zwar mit Abweichungen von der unmarkierten Wortfolge im Englischen befassen und diese aber mit Faktoren erklärt werden, die auch in dieser Arbeit genutzt werden. Arnold et al. (2000) beschreiben anhand von einer Korpusuntersuchung und Experimenten den Einfluss der Länge und des Informationsstatus der Satzglieder auf das Phänomen des *Heavy NP Shifts* und der *Dativalternation*. Als Messung der Länge nehmen sie die Differenz zwischen den beiden infrage kommenden Satzgliedern und finden erneut Bestätigung dafür, dass schwere Konstituen-

⁴ Wenn Extraposition in modernen deutschen, schriftlichen Texten vorkommt, wird sie auch häufig genutzt, um ein neues Thema einzuleiten (vgl. z. B. Petkova-Kessanlis 2015).

ten später im Satz auftreten. Um den Einfluss des Informationsstatus auf die beiden Phänomene zu beschreiben, überprüfen Arnold et al. (2000: 30) auch, ob bekannte vor neue Information gestellt wird. Begründet wird das Prinzip sowohl mit der Sprecher- als auch mit der Hörerorientierung (vgl. Arnold et al. 2000: 31ff.). Sowohl die Rezeption als auch die Produktion können durch die Stellung von bekannter Information vor neuer Information erleichtert werden. Die Untersuchung erfolgt an Transkripten von Debatten des kanadischen Parlaments. Die Texte wurden also medial mündlich produziert, dürften aber in der Mitte zwischen konzeptioneller Mündlichkeit und Schriftlichkeit liegen (vgl. Arnold et al. 2000: 35). Die Hörenden sollten also mehr kognitive Kapazitäten verbrauchen als die Schreibenden, die mehr Zeit in der Planung hatten. Arnold et al. (2000: 37f.) bestätigen, dass neue und lange Konstituenten später im Satz auftreten:

[T]he effect of newness appears to be greater for the middle categories of heaviness; that is, newness has more of an effect when heaviness does not make strong predictions either way. When one constituent is considerably longer than the other, newness does not account for much of the data, but when the two constituents are closer in length, newness effects are more clearly observable. (Arnold et al. 2000: 38)

Ihre Produktionsstudie brachte ähnliche Ergebnisse, wobei sich dort die Interaktion zwischen der Länge des extrapponierten Materials und des Bekanntheitsstatus deutlicher zeigt und die Länge dann einen größeren Effekt aufweist, wenn beide Objekte bekannt sind (vgl. Arnold et al. 2000: 48). Der Verarbeitungsbezug zeigt sich aufgrund von Länge und Bekanntheit, der sowohl den Sprechenden als auch den Adressierten nützt. Extraposition geschieht folglich auch im Englischen aufgrund von Verarbeitung. Dies lässt sich sowohl in Korpora als auch in Experimenten nachweisen.

Experimentell betrachtet die Studie von Weber (2019) die Auslagerung von deutschen Präpositionalphrasen aus Nominalphrasen heraus. Dabei geht sie der Frage nach, ob eine Präpositionalphrase dann wahrscheinlicher ausgelagert wird, wenn sie mit einem Relativsatz zusammen ausgelagert werden kann oder aus einer definiten Nominalphrase heraus bewegt wird (vgl. Weber 2019: 63). Die Attribuierung eines Relativsatzes an die Präpositionalphra-

se legt nahe, diese Phrase als komplex zu charakterisieren, was Weber (2019) auch in Anlehnung an Hawkins (1992) und Gibson (1998) tut. Diese Komplexität kann also zum Beispiel durch die Anzahl der verbalen Knoten gemessen werden oder aber durch die bloße Anzahl an Wörtern. Weber (2019) fokussiert sich unter anderem darauf, ob die Komplexität der Phrase gemessen in verbalen Knoten tatsächlich Einfluss hat.

Weber (2019: 70) untersucht ihre Fragestellung mit einem Akzeptabilitätsexperiment. Darin geht sie der Frage nach, ob eine Präpositionalphrase mit oder ohne Relativsatz natürlicher klingt, wenn sie über verbales Material bewegt wird. Um diesen Faktor konstant zu halten, hält Weber (2019: 70) die Länge der beiden Auslagerungen gleich. Die Extrapositionen werden unabhängig von ihrer grammatischen Komplexität generell schlechter bewertet als Einbettungen (vgl. Weber 2019: 72). In der Interpretation zieht Weber (2019: 78) einen Bezug zu ihrer Studie über Relativsätze (vgl. Weber 2018) und beschreibt beide Auslagerungsarten als fundamental unterschiedlich. Erklärt wird das durch die Existenz beziehungsweise das Fehlen des Relativums, das durch die Kongruenz mit dem Bezugswort der Disambiguierung des Bezugs der Auslagerung dient. Bei der Auslagerung attributiver Präpositionalphrasen verbrauchen die Hörenden mehr kognitive Kapazitäten für die Verarbeitung, was in deren generell schlechterer Bewertung resultiert.

In ihrem zweiten Experiment findet Weber (2019: 74ff.) eine bessere Bewertung von extrapositionierten Präpositionalphrasen mit Relativsatz. Der Unterschied zwischen den beiden Experimenten ist die verschiedene Länge der ausgelagerten Phrasen. Längere Phrasen erzielten bessere Bewertungen, wenn sie in Wörtern gemessen wurden (vgl. Weber 2019: 79). Die Anzahl von verbalen Knoten beeinflusst die Akzeptabilität einer Auslagerung hingegen weniger (vgl. Weber 2019: 79). Es zeigt sich also insgesamt ein Einfluss der Länge auf die Extraposition von Präpositionalphrasen (vgl. Weber 2019: 79).

Diachrone Forschung Sapp (2014) untersucht Nachfeldbesetzung im Mittel- und Frühneuhochdeutschen und weist durch logistische Regression den Einfluss von Mündlichkeitsnähe und Informationsstruktur nach. Dazu wurden Prosatexte aus dem Bochumer Mittelhochdeutschkorpus (vgl. Klein und Dip-

per 2016) und aus dem Bonner Frühneuhochdeutschkorpus (vgl. Fisseni 2017) untersucht (Sapp 2014: 132). Es wurden 2312 Teilsätze mit Auslagerungen und 1629 Sätze ohne Auslagerungen, aber mit Material, das grundsätzlich extrapponiert werden kann, wurden gefunden. Sapp (2014) untersucht die Variablen Art der Auslagerung (Nominal-, Präpositional- und Adjektivphrasen), Konstituentenart der Konstituente vor der Auslagerung, Wort- und Silbenzahl der Extraposition, Fokus, Satzart, morphosyntaktische Erscheinungsform der Elemente im Verbalkomplex des Verbalkomplexes, Reihenfolge im Verbalkomplex, Zeitraum, Dialekt und Genre (vgl. Sapp 2014: 132ff.).

Bei der Analyse der Ergebnisse bestätigt er zunächst Funde aus der synchronen Betrachtung von Extrapositionen (vgl. z. B. Zifonun et al. 1997: 1657): Präpositionalphrasen zählen zu den häufigsten, ausgelagerten Elementen, gefolgt von Nominalphrasen, einer Kombination mehrerer Elemente, Adverbialphrasen und Pronomen (vgl. Sapp 2014: 135), die im heutigen Deutschen als ungrammatisch gelten, wenn sie im Nachfeld auftreten (vgl. Zifonun et al. 1997: 1651). Auch Sapp (2014) bestätigt den großen Einfluss von Länge auf die Position von Phrasen. Alle gefundenen Auslagerungen hatten mindestens drei Wörter. Auch der Fokus wird als einflussreicher Prädiktor für Extraposition angesehen, wenn er entweder auf der Verbalphrase liegt, wenn Objektfokus besteht oder er auf einer andere Konstituente zu finden ist (vgl. Sapp 2014: 137f.). Dahingegen verhindern das Ausbleiben von Fokus, Verbfokus, Subjektfokus und Fokus auf dem ganzen Satz die Nachfeldbesetzung einzelner Konstituenten (vgl. Sapp 2014: 137f.). Fokus wird hier als Charakteristikum für Informationsstatus verwendet (vgl. Sapp 2014).

Außerdem untersucht Sapp (2014) das Genre der Texte, das er mit der Mündlichkeitsnähe gleichsetzt. Anhand dieser Klassifikation bestätigt er, dass Auslagerungen häufiger in Texten zu finden sind, die sehr wahrscheinlich gesprochen wurden (vgl. Sapp 2014). Dahingegen bleibt der Dialekt ohne signifikanten Einfluss, was nicht auf den Zeitraum der Texte zutrifft (Sapp 2014: 141): „The frequency of extraposition remains stable from the 12th through the 15th century at around 37-39 %, before dropping precipitously in the 16th to less than five percent.“ Trotz dieses Rückgangs merkt er an, dass Länge und

Fokus nicht an Bedeutung verlieren (vgl. Sapp 2014: 141f.).

Hervorzuheben ist die Studie von Sapp (2014) aufgrund der Größe seines Datensatzes und der verwendeten Statistik, nämlich logistische Regression. Sapp (2014) findet, dass sich die Einflussfaktoren auf Auslagerung nicht signifikant zwischen dem Mittelhochdeutschen und dem Neuhochdeutschen unterscheiden. Wie im vorherigen Abschnitt dargestellt wurde, sind die genannten Faktoren auch in der synchronen Literatur, die sich mit Auslagerung befasst wiederzufinden.

Auch Ebert (1980) untersucht die Entwicklung des Satzrahmens im Frühneuhochdeutschen und dessen Durchbrechung. Dabei greift er auf die soziolinguistischen Faktoren Alter, Geschlecht, sozialer Rang, Bildung und Beruf zurück. Außerdem versucht er Extraposition mit folgenden, linguistischen Klassifikationen zu erklären: Der Art der rechten Satzklammer, der ausgelagerte Konstituente, ihre Anzahl und Länge, sowie der Bekanntheitsstatus der Konstituenten und die Unterscheidung in Fokus und Hintergrund (vgl. Ebert 1980: 360). Dazu nutzt Ebert (1980: 360) logistische Regression.

Sein Korpus stammt aus Nürnberg und besteht aus informellen und formellen Briefen, Tagebüchern und Chroniken (vgl. Ebert 1980: 362f.). Untersucht werden die Zeitabschnitte 1300 bis 1400, 1400 bis 1500, 1500 bis 1550 und 1550 bis 1600 (vgl. Ebert 1980: 362f.). Dabei zeigt sich, dass die Anzahl der vollständigen Satzrahmen⁵ im 15. Jahrhundert geringer ist als im 14. und 16. Jahrhundert, und es einen Unterschied macht, welche Textsorte untersucht wird, da in informellen Briefen mehr vollständige Satzrahmen gefunden werden (vgl. Ebert 1980: 368). In der Regel werden nur einzelne Konstituenten, meistens eine Nominal- oder Präpositionalphrase, extraponiert (vgl. Ebert 1980: 374). Bei vollständigen Satzrahmen sind diese eher lang und weisen mit höherer Wahrscheinlichkeit mehr als eine Konstituente im Nachfeld auf (vgl. Ebert 1980: 380). Zudem finden sich vollständige Satzrahmen auch bei mündlichkeitsnahen Texten (vgl. Ebert 1980: 382).

Im Gegensatz zu Ebert (1980) fokussiert sich Light (2011: 319) auf einen kleineren Zeitabschnitt, indem sie die Auslagerung von Subjekten im 1522

⁵ Ebert (1980: 357f.) definiert den Satzrahmen in weitestgehender Übereinstimmung mit Wöllstein (2014).

erschienenen Septembertestament von Martin Luther untersucht und sie in Zusammenhang mit engem Fokus und Satzakzent setzt (vgl. Light 2011: 315). Sie findet insgesamt 115 eindeutig ausgelagerte Subjekte (vgl. Light 2011: 319), deren Anzahl sie aufgrund von Längenfaktoren reduziert,⁶ um diesen Einfluss zu kontrollieren. In die Untersuchung wurden sowohl bei den ausgelagerten Subjekten als auch bei ihren eingebetteten Gegenständen nur Phrasen mit maximal 15 Silben einbezogen. Sie kommt zu folgendem Ergebnis: „[Q]uantified/numeric subjects are extraposed more frequently than definites“ (vgl. Light 2011: 320).

Wird ein Subjekt im Nachfeld realisiert, handelt es sich in fast 40 % der Fälle um einen diskursneuen Referenten, während das auf nur rund 12 % der eingebetteten Subjekte zutrifft (vgl. Light 2011: 321). Diese sind dafür in 61 % der Fälle bekannt. Eine inferenzstatistische Analyse wurde nicht durchgeführt. Die Auszählung allein weist aber bereits darauf hin, dass die Bekanntheit eines Subjekts einen Einfluss auf die Position hat. Auch wenn Light (2011) es nicht auf die Verarbeitbarkeit der Phrase in der Position bezieht, lassen sich aus den Ergebnissen diese Rückschlüsse im Rahmen der in Kapitel 1.1 getroffenen Erläuterungen ziehen. Wenn neue Referenten tatsächlich schwerer zu verarbeiten sind als bekannte, könnten sie deshalb ins Nachfeld gestellt werden. Wie bereits erwähnt, zieht Light (2011) diesen Schluss jedoch nicht, sondern versucht, einen einheitlichen Analyseansatz für die Extraposition von Subjekten und Objekten, die aus der Arbeit von Bies (1996) genommen werden, zu finden. Light (2011: 322ff.) erklärt die Auslagerung beider Phrasentypen mit Prosodie. Subjektextraposition geschehe, damit der Satzakzent auf dem Subjekt bleiben kann, ohne zugleich eine kontrastive Lesart zu provozieren (vgl. Light 2011: 323). Sie untermalt ihre Behauptung allerdings nur mit Beispielsätzen, statt quantitativ auszuwerten, ob extraponierte Subjekte diese Funktion (signifikant) häufiger erfüllen als andere ins Nachfeld gestellte Objekte. Generell ist die fehlende statistische Analyse der eigentliche Schwachpunkt der Studie. Besonders, da das Beispiel, das ihren Punkt zur Prosodie klar stützen soll, zugleich eine außergewöhnlich lange Nomi-

⁶ Die ausgelagerten Subjekte sind im Schnitt 13,07 Silben lang, während die eingebetteten nur 3,29 Silben besitzen (vgl. Light 2011: 320).

nalphrase extrapониert und sie nicht darauf eingeht, ob das der eigentliche Grund der Extraposition ist.

Sowohl die synchronen als auch diachronen Forschungsdaten zeigen zusammenfassend den Einfluss der Länge auf die Positionierung von Nominal- und Präpositionalphrasen. Diese Faktor wird auch bei der Betrachtung der Extraponierbarkeit von Relativsätzen von besonderer Relevanz sein, wie im nächsten Kapitel gezeigt wird.

3.2. Die Extraposition von Relativsätzen

Die folgenden Abschnitte befassen sich mit der Extraposition von Relativsätzen. Wie im vorherigen Abschnitt auch wird dabei zwischen dem unterschieden, was geläufige Grammatiken zu dem Thema schreiben (Kapitel 3.2.1), und dem, was die Forschung über die Position von Relativsätzen herausgefunden hat (Kapitel 3.2.2).⁷ Im Gegensatz zum Phänomen der phrasalen Extraposition ist die Abdeckung des Themas sowohl in der Forschung als auch in Grammatiken gut. Letztere beschreiben aber oft entweder nur, dass es verschiedene Stellungsoptionen von Relativsätzen gibt, oder führen Gründe an, die auch mithilfe von Korpusuntersuchungen oder Experimenten überprüft wurden. Um Wiederholungen auf ein Minimum zu begrenzen, wird Kapitel 3.2.1 eher kurz gehalten. Kapitel 3.2.2 fokussiert sich auf Arbeiten, die Extraposition mit ähnlichen Faktoren wie den von mir verwendeten erklären.⁸

3.2.1. Grammatiken

Wie bereits in Kapitel 1 und Kapitel 2.4.1 gesagt wurde, können Relativsätze adjazent zu (2a) und separiert von ihrem Bezugswort (2b) stehen (vgl. auch Lehmann 1984). Im Fall von adjazenten Relativsätzen steht keinerlei sprachliches Material zwischen den Bezugswort und dem Relativsatz, während bei

⁷ Teile des in diesem Kapitel vorgestellten Materials flossen auch in die Veröffentlichungen Speyer und Voigtmann (2021a;b) und Voigtmann (2022) ein.

⁸ Da zur Extraposition von Relativsätzen insgesamt mehr geforscht wurde, fokussiert sich die hiesige Darstellung auch auf Forschung ab dem Frühneuhochdeutschen. Trotzdem sollen hier beispielhaft die Untersuchungen über das Mittelhochdeutsche von Sapp (2014) und über das Althochdeutsche von Coniglio, Linde und Ruetter (2017) genannt werden.

ausgelagerten Relativsätzen beispielsweise die rechte Satzklammer und/oder andere Konstituenten intervenieren können.

- (2) a. Ursel hat den Kuchen, **den Klaus sich gewünscht hat**, gebacken.
b. Ursel hat den Kuchen gebacken, **den Klaus sich gewünscht hat**.

Fritsch (1990: 114) beschreibt die Trennung vom Antezedens als Bewegung des Relativsatzes „an den Rand des kleinsten Satzes, in dem er enthalten ist“. Dabei darf diese Bewegung die Organisation anderer Sätze nicht beeinflussen, weshalb der Relativsatz außerhalb der rechten Satzklammer stehen muss. Zwar verwendet Fritsch (1990: 114) in seinen Beispielen nur Relativsätze, die über eine lexikalisch realisierte Satzklammer bewegt werden. Der Landepunkt ist aber auch dann das Nachfeld, wenn die Extraposition über nicht-verbale Konstituenten im Mittelfeld erfolgt.

Dabei steht die Auslagerung von Relativsätzen dem Gesetz von Behaghel (1932) entgegen, das besagt, das geistig Zusammengehörendes auch zusammenstehen sollte. Durch den Attributcharakter der Relativsätze können sie als geistig eng mit dem Bezugsnomen verbunden betrachtet werden, weshalb sie mit den Kopfnomen zusammen eine gemeinsame Konstituente bilden (vgl. Fritsch 1990: 107). Trotzdem werden sowohl in modernen als auch historischen Korpora fast genauso viele Beispiele für ausgelagerte wie für eingebettete Relativsätze gefunden.⁹

Die Begründung dafür lautet, dass andere, vor allem pragmatische Einflussfaktoren (vgl. Shannon 1992: 253), schwerer wiegen als die Zusammengehörigkeit von Kopfnomen und Relativsatz. Die Literatur ist sich weitestgehend darüber einig, welche Faktoren Einfluss darauf nehmen können, wo ein Relativsatz steht. Zu ihnen zählen die Länge (vgl. u.a. Uszkoreit et al. 1998; Poschmann und Wagner 2016), sowohl vom Relativsatz als auch vom Matrixsatz, die Distanz zwischen dem Antezedens und dem Nachfeld (u.a. Uszkoreit et al. 1998; Poschmann und Wagner 2016), die Restriktivität des

⁹ Uszkoreit et al. (1998: 131) findet zum Beispiel 340 extraponierte und 449 eingebettete Relativsätze für seine Korpusuntersuchung. Und auch die hier vorliegende Untersuchung historischer Texte kann auf einen etwa gleichen Anteil von eindeutig extraponierten und eindeutig eingebetteten Relativsätzen zurückgreifen (Kapitel 6.3).

Relativsatzes (vgl. Shannon 1992; Marillier 1993) sowie Fokussierungsbedingungen und Akzentuierungen. Immer wieder werden auch Verarbeitungsbedingungen herangeführt (vgl. z. B. Uszkoreit et al. 1998: 129). Was bisher selten getan wurde, ist, die zuerst genannten Faktoren mit dem Faktor der Informationsentflechtung (vgl. Zifonun et al. 1997: 1669) zusammenzubringen und sie explizit zu einem einzigen Grund zu verdichten: Durch die verschiedenen Positionierungen von Relativsätzen soll erfolgreiche Kommunikation gesichert werden (vgl. Voigtmann und Speyer 2021).

3.2.2. Forschung

Synchrone Forschung Dass *Fokus* und *Akzentuierung* nicht leicht von der Informationsentflechtung zu trennen sind, wird dadurch klar, dass Poschmann und Wagner (2016) beschreiben, dass Fokus häufig mit neuen Informationen verbunden ist und neue Informationen dazu tendieren, komplexer beziehungsweise länger zu sein als alte Informationen. Daher lässt sich die Verbindung von Fokus und Relativsatzauslagerung nicht eindeutig von dem Punkt neuer beziehungsweise bekannter Information trennen (vgl. z. B. Prince 1981; Gundel et al. 1993; Krifka 2008). Auch gelingt es Poschmann und Wagner (2016) nicht, Länge und Fokus bei der Begründung für die Relativsatzextraposition klar voneinander abzugrenzen. Sie können jedoch zeigen, dass Relativsätze signifikant häufiger ausgelagert werden, wenn das Antezedens zur Hintergrundinformation gehört (vgl. Poschmann und Wagner 2016: 1027). Auch die intervenierenden Elemente spielen für die Auslagerungstendenzen eine Rolle. Nominalphrasen zwischen dem Antezedens und dem Relativsatz verhindern dann die Auslagerung, wenn die Zuordnung des Relativsatzes durch ihr Vorkommen wie in Beispiel (3a) ambig wird (vgl. Poschmann und Wagner 2016: 1025). Das gilt auch bei Determiniererphrasen, da sich Relativsätze auch auf Pronomen beziehen können, obwohl diese meistens unbetont sind (3b). Verglichen damit ist die Bewegung über direktionale oder lokale Adverbien hinweg einfacher, da diese seltener potenzielle Antezedenzen sein können (vgl. Kathol und Pollard 1995). In ihren Experimenten zeigen Kathol und Pollard (1995), dass die Verteilung der Akzente im Satz einen großen

Einfluss auf die Akzeptabilität der Auslagerung hat.

- (3) a. Gestern hat der Mann den Jungen gesehen, der die Straße entlang gegangen ist.
 b. Gestern hat er ihn gesehen, der den Rasen gemäht hat.

Auch Lötscher (1972) befasst sich mit dem Verhältnis von Auslagerung und Akzentuierung und findet, dass die Auslagerungstendenz steigt, wenn der Satzakkzent auf dem Antezedens liegt. Shannon (1992: 273) ist strikter in seiner Aussage: Laut ihm sei Auslagerung nur dann möglich, wenn das Bezugswort rhematisch ist und fokussiert wird. Das treffe selten auf Bezugswörter im Vorfeld zu und ebenso selten auf definite Bezugswörter. Problematisch an dieser absoluten Aussage ist, dass Shannon (1992) dennoch auch Fälle anerkennt, bei denen der Fokus nicht auf dem Antezedens liegt. Shannon (1992: 275f.) löst dieses Problem aber nicht. Auch für ältere Zeitstufen hat sich die Bedeutung von Fokus und Akzentuierung auf Auslagerung gezeigt.

Poschmann und Wagner (2016) verknüpfen Fokus mit neuer Information, aber im gleichen Zug auch mit einer größeren Länge des fokussierten Materials. Das zeigt nicht nur, wie schwer die einzelnen Bestandteile voneinander zu trennen sind, sondern auch den Einfluss von *Länge*.

Hinsichtlich Auslagerungen ist es allgemein so, dass umfangreiche Konstituenten eher im Nachfeld zu finden sind (vgl. Wöllstein et al. 1997).¹⁰ Der Einfluss von Länge auf Auslagerung ist gut belegt in der Literatur. Schon Behaghel (1932) spricht davon, dass längere Satzglieder eher am Ende eines Satzes vorkommen und kürzere weiter vorne stehen. Er formuliert das im „Gesetz der wachsenden Glieder“ (Behaghel 1932: 139). Auch neuere Studien

¹⁰ Die Definition von ‚umfangreich‘ variiert dabei zwischen der Zählweise einfacher Wörter, Silben und der grammatischen Komplexität, der Erweiterung des ausgelagerten Materials durch weitere Sätze. Wasow (1997) zeigt, dass alle Zählweisen zum gleichen Ergebnis führen können, indem er die verschiedenen Interpretationen von Gewicht beziehungsweise Länge beziehungsweise Komplexität anhand von „heavy NP shift (HNPS), particle movement [...] (MP), and the dative alternation (DA)“ (Wasow 1997: 83) im Englischen miteinander vergleicht. Dabei unterscheidet er zwischen der Länge einer Konstituente in Wörtern und der grammatischen Komplexität, die sich nach Anzahl von (phrasalen) Knoten richtet (vgl. Wasow 1997: 85). Beide nicht voneinander zu trennenden Maße eignen sich sehr gut für die Vorhersage der Konstituentenabfolge (vgl. Wasow 1997: 91). Je mehr Wörter eine Phrase hat, desto mehr phrasale Knoten kann sie auch besitzen (vgl. Wasow 1997: 93).

bestätigen dies und führen es auf die leichtere Verarbeitung kürzerer Satzglieder zurück (vgl. Poschmann und Wagner 2016), weil die Länge von Wörtern mit der Häufigkeit ihres Vorkommens zusammenhängt, was kürzere Wörter präsenter und per se einfacher zu verarbeiten macht (vgl. z. B. Shannon 1948; Gibson et al. 2019).

Eine der ausführlichsten Studien zum Zusammenhang von Länge und Relativsatzauslagerung wurde von Uszkoreit et al. (1998) vorgenommen. Sie gehen von der Annahme aus, dass Relativsätze aufgrund von Performanzeffekten ausgelagert werden, um das Kurzzeitgedächtnis zu entlasten (vgl. Uszkoreit et al. 1998: 130). Dabei greifen sie auf Hawkins (1992) zurück, der darlegt, dass Abweichungen von der unmarkierten Wortstellung zur schnellstmöglichen Verarbeitung genutzt werden. Für Relativsätze bedeutet das, dass die Distanz zwischen dem Antezedens und dem Relativpronomen so klein wie möglich gehalten werden soll. Daher müsste Adjazenz theoretisch immer zu bevorzugen sein. Das ist jedoch nicht der Fall, weil man zugleich die Distanz zwischen den beiden Satzklammern beachten muss. Auch für sie gilt der Grundsatz der verringerten Distanz. Daraus entsteht ein Spannungsfeld, mit dem auch begründet wird, warum die Länge eines Relativsatzes Extraposition bedingen kann. Die Distanz zwischen den Satzklammern wird so nämlich wie generell bei Extrapositionen verringert.

Diesen Ansatz greifen Uszkoreit et al. (1998: 131) auf und untersuchen ihn anhand einer Korpusanalyse basierend auf Zeitungsartikeln der Frankfurter Rundschau und anhand eines Rating-Experiments. In der Korpusuntersuchung schauen sie nach der Verteilung der Relativsätze im Text¹¹ und erfassen automatisch die Länge der Relativsätze und die Entfernung zwischen Antezedens und Relativsatz. Die Messung erfolgt in Wörtern (vgl. Uszkoreit et al. 1998: 131). Die Autoren finden heraus, dass extraponierte Relativsätze in der Regel 10,3 Wörter lang sind, während eingebettete nur 9 Wörter bemessen. Die maximale Distanz, die überbrückt wird, beträgt 9 Wörter. Im Schnitt trennen 1 bis 4 Wörter den Relativsatz von seinem Antezedens, wobei

¹¹ Von 12000 Sätzen sind nur 340 eindeutig ausgelagert und 499 eindeutig adjazent. Adjazente Relativsätze, die am Ende des Satzes stehen, ohne dass eine rechte Satzklammer lexikalisch realisiert ist, werden nicht in die Auswertung einbezogen (vgl. Uszkoreit et al. 1998: 131).

die Mehrzahl der Auslagerungen über verbales Material hinweg erfolgt (vgl. Uszkoreit et al. 1998: 131). Die anschließende experimentelle Betrachtung testet die Akzeptabilität von ausgelagerten Relativsätzen in Abhängigkeit von Länge und Distanz. Dabei vertreten Uszkoreit et al. (1998: 132) die Hypothese, dass ein über zwei Wörter ausgelagerter Relativsatz die beste Bewertungen enthält. Tatsächlich zeigen sie, dass eine Vergrößerung der Distanz zu signifikant besseren Bewertungen für Adjazenz führt (vgl. Uszkoreit et al. 1998: 132) und dass die Länge des Relativsatzes dieses Kriterium erst dann überschreiben kann, wenn er zwischen neun und elf Wörtern lang ist. Generell wird Adjazenz besser bewertet, was den Ergebnissen der Korpusuntersuchung in gewisser Weise widerspricht. Die dort gefundene Anzahl an Relativsätzen ist etwa gleich auf beide Positionen verteilt, was eigentlich dafür sprechen sollte, dass auch die Bewertung der unterschiedlichen Positionen etwa gleich sein sollte. Uszkoreit et al. (1998: 132) schließen daraus, dass die Distanz generell eigentlich die besseren Vorhersagen über die Position eines Relativsatzes treffen kann als Länge. *Distanzen* fordern das menschliche Verarbeitungssystem grundsätzlich heraus. Auch für das Deutsche, dessen Sprecher:innen durch die Satzklammern an Distanzstellung gewöhnt sind, konnten Futrell et al. (2021) beispielsweise zeigen, dass zu große Distanzen trotz der Gewöhnung an sie zu Informationsverlust führen können. Das gilt für die Relativsatzextraposition besonders dann, wenn durch das überbrückte Material Ambiguitäten in der Zuordnung des Relativsatzes entstehen können (vgl. Poschmann und Wagner 2016). (4) illustriert diese Schwierigkeit der Distanzstellung, die auch gegeben ist, wenn das Antezedenz nicht ambig ist.

- (4) Klaus sagte zu Ursel, dass sie *den Tannenbaum* dieses Jahr ausnahmsweise wirklich nicht in das große Wohnzimmer neben das Highboard stellen sollten, **den sie in der Stadt gekauft hatten.**

Zwischen dem Bezugswort *Tannenbaum* und dem Relativsatz stehen 14 Wörter. Der Satz ist zwar grammatisch, doch die Verarbeitung des Relativsatzes wird durch die große Distanz erschwert, obwohl keine Wörter intervenieren, die mit dem Relativpronomen kongruent wären, sodass das Antezedenz

nicht ambig ist. Erschwerend kommt aber hinzu, dass dieser Relativsatz nicht eindeutig als restriktiv oder appositiv klassifiziert werden kann. Die Kürze des Relativsatzes (7 Wörter) zusammen mit der weiten Strecke, die bei der Stellung ins Nachfeld überbrückt werden muss, sprechen in diesem Beispiel dafür, dass der Relativsatz eigentlich eingebettet werden sollte.

Die Distanz zwischen der potenziellen Landeposition und dem Relativpronomen ist jedoch nicht die einzige, die berücksichtigt werden muss. In (4) ist das Mittelfeld relativ lang, aber da es sich dabei um einen Nebensatz mit Verbletzstellung handelt, ist dem Rezipienten aufgrund dieser Konvention bewusst, dass alle Bestandteile des Prädikats erst folgen müssen. Während des inkrementellen Lesens des Satzes, verringert sich zudem die Auswahlmöglichkeit des Prädikats (vgl. z. B. Levy 2008).

Die Erwähnung eines Tannenbaums und die Existenz eines Lokaladverbials lassen *stellen* (oder ein semantisch ähnliches Verb) wahrscheinlicher werden. Trotzdem besteht bei allen sehr großen Distanzen die Gefahr, dass Informationen verloren gehen. Das kann sich sowohl auf die Satzklammern, aber auch auf alle anderen Bestandteile im Mittelfeld beziehen. Schon im 17. Jahrhundert galt es als erstrebenswert, wenn auch nicht unbedingt als prestigeträchtig, die Mittelfelder möglichst kurz zu halten und dazu auch Auslagerungen zu nutzen (vgl. Konopka 1996; Takada 1998). Das hat sich im modernen Deutschen nicht geändert. Hier besteht jedoch auch die Verknüpfung zur Länge der Relativsätze. Je länger ein adjazenter Relativsatz ist, desto länger wird auch das entsprechende Feld, was die gesamte Satzverarbeitung erschwert (5).

- (5) Ursel [hat]_{LSK} das Geschirr, **das sie vor über einem Jahr bei einem Besuch in dem kleinen Geschäft im Westen des Saarlandes gesehen hatte**, nach langen Debatten mit Klaus über dessen Preis nun doch [gekauft]_{RSK}.

Länge und Distanz sind aufgrund ihrer Interaktion schwer voneinander zu trennen. Zu bevorzugen sind kurze Distanzen, vor allem über die rechte Satzklammer hinweg, und vor allem dann, wenn die Relativsätze gleichzeitig lang

sind (vgl. Uszkoreit et al. 1998). Kurze Relativsätze, die zu einem Bezugswort zu Beginn des Mittelfeldes oder sogar zu einem im Vorfeld stehenden Antezedens gehören, sollten dieser Argumentation folgend daher eher eingebettet werden. Gerade das seltene Vorkommen von im Vorfeld stehenden Bezugselementen von extraponierten Relativsätzen wird durch diese Faktoren erfolgreich erklärt.

Nach Shannon (1992) findet Extraposition immer dann statt, wenn Negationen, Adjektive im Superlative oder Quantoren vor dem Antezedens stehen. Wie in Kapitel 2.4.3 gezeigt wurde, deckt sich diese Beobachtung mit der *Restriktivität* von Relativsätzen. Auch Zifonun et al. (1997: 1652f.) schreiben, dass restriktive Relativsätze häufiger dazu tendieren, ausgelagert zu werden. Erklärt wird das damit, dass der Relativsatz dafür notwendig ist, den Satz zu vervollständigen. Es bleibt quasi eine Lücke in der Erwartung der Rezipierenden oder Schreibenden übrig, für die auch kognitive Kapazitäten weiter freigelassen werden, die dann darauf verwendet werden können, den restriktiven Relativsatz zu verarbeiten und das Bezugswort damit eindeutig zu identifizieren. Intuitiv klingt die Auslagerung eines restriktiven Relativsatzes über eine größere Distanz auch bei einer geringen Länge des Relativsatzes natürlicher (6a) als es bei einem appositiven Relativsatz der Fall ist (6b).

- (6) a. Klaus hat denjenigen Mitarbeiter wegen der mangelnden Qualität seiner Arbeit gerügt, **bei dem das schon vergangene Woche nötig gewesen war.**
- b. Klaus hat seiner Nichte gestern per E-Mail zum Geburtstag gratuliert, **die ja zurzeit in den USA wohnt.**

Bei Korpusuntersuchungen, die die Restriktivität des Relativsatzes als Faktor in ihre Untersuchung aufnehmen (vgl. Marillier 1993), hat sich jedoch gezeigt, dass der Einfluss des Relativsatztyps gering bis unbedeutend ist. Faktoren wie Länge und Distanz überschreiben seinen Einfluss konsequent (vgl. Marillier 1993).

An dieser Stelle lässt sich also zusammenfassen, dass für die Extraposition moderner Relativsätze verschiedene Faktoren verantwortlich sind, die alle

auf erfolgreiche Kommunikation und die Vermeidung von Informationsverlusten abzielen. Ein wichtiges Kriterium, das hier noch keinen Niederschlag gefunden hat, ist die Definition von Information. Bislang kann Information vor allem mit Komplexität und Länge gleichgesetzt werden oder wurde mithilfe der Fokus-Hintergrund-Gliederung beschrieben. Für frühere Zeitstufen sind hingegen auch Untersuchungen, die Information im Hinblick auf bekannte und neue Information (im Sinne von „Givenness“, vgl. z. B. Krifka 2008; Prince 1981; Gundel et al. 1993) durchgeführt worden, die unter anderem im folgenden Abschnitt dargelegt werden.

Diachrone Forschung Sahel (2015: 165) befasst sich mit der Nachfeldstellung von Relativsätzen und Vergleichsphrasen im frühen Neuhochdeutschen mithilfe des GerManC Korpus (vgl. Bennett, Durrell, Ensslin, Scheible und Whitt 2007). Interessant an seinem Ansatz ist die Behauptung, dass die Auslagerungspraxis im Frühneuhochdeutschen nicht mit der im (frühen) Neuhochdeutschen zu vergleichen sei, da von 1350 bis 1650 mehr Elemente auslagerungsfähig waren als im Neuhochdeutschen (vgl. Sahel 2015: 168). Damit begründet er die Beschränkung auf Vergleichsphrasen und Relativsätze. Gerade erstere stünden heute unmarkiert im Nachfeld (vgl. Sahel 2015: 169). Dabei werden jedoch Untersuchungen vernachlässigt, die zeigen, dass im modernen Deutschen auch Nominal-, Präpositional-, Adjektiv- und Adverbphrasen prinzipiell im Nachfeld stehen können, ohne dass die Sätze ungrammatisch werden (vgl. z. B. Sapp 2014). Auch die vorliegende Untersuchung zeigt, dass bis ins 20. Jahrhundert Nominal- und Präpositionalphrasen im Nachfeld gefunden werden (Kapitel 6.1 und 6.2). Im vorliegenden Kapitel sollten zudem ausreichend viele Studien aufgeführt worden sein, die sogar die gleichen, zeitlich stabilen Faktoren als relevant für Extraposition bezeichnen.

Die Behauptung über unterschiedliche Verhaltensweisen von Extraposition ist nicht haltbar, was dennoch wenig an der Relevanz der Studie von Sahel (2015) ändert. Sie ist eine der einzigen zur Relativsatzextraposition im frühen Neuhochdeutschen. Sahel (2015: 171) unterteilt seinen Korpus in 50-Jahres-Abschnitte (1650 bis 1700, 1700 bis 1750, 1750 bis 1800) und untersucht Prosatexte, Zeitungen und wissenschaftliche Texte. In die Untersuchung wer-

den nur Sätze einbezogen, die eine linke und eine rechte Satzklammer haben (vgl. Sahel 2015: 171). Es werden 1108 Relativsätze und 213 Vergleichsphrasen untersucht (vgl. Sahel 2015: 171), wobei eine steigende Tendenz hinsichtlich der Auslagerung der Relativsätze beobachtet wird (vgl. Sahel 2015: 172). Am geringsten ist der Anteil der ausgelagerten Relativsätze in den wissenschaftlichen Texten, am höchsten in der Prosa. Allerdings ist die Zunahme in allen Textsorten erkennbar (vgl. Sahel 2015: 173).¹² Eine Erklärung für dieses Phänomen bleibt er schuldig und auch die Motivation für die Positionierung der beiden Elemente wird nicht untersucht. In der Darlegung der theoretischen Grundlage erwähnt er zwar, dass die erweiterten Auslagerungsmöglichkeiten des Frühneuhochdeutschen im Vergleich zum Neuhochdeutschen Hinweise darauf liefern, dass die Klammerstruktur noch nicht ausreichend motiviert sei (vgl. Sahel 2015: 168f.). Vielleicht haben sich aber auch die Gründe für Extraposition dieser Konstituenten geändert. Auch wäre es zur Klärung der Festigung des Satzrahmens sinnvoller gewesen, Elemente zu untersuchen, die im heutigen Deutschen in geringerer Zahl im Nachfeld vorkommen wie Nominal- oder Präpositionalphrasen. Trotzdem zeigt die Arbeit die Entwicklung von Relativsatzauslagerung.

Ein weiterer Forschungsbeitrag, der sich mit Relativsatzextraposition beschäftigt, stammt von Voigtmann (2022). Mithilfe eines Korpus aus Briefen aus dem 16. bis 19. Jahrhundert (vgl. Voigtmann 2022: 238) wird einerseits untersucht, ob adjazente Relativsätze einen höheren Anteil bekannter Referenten haben, und umgekehrt, ob extraponierte Relativsätze einen höheren Anteil neuer Referenten besitzen. Außerdem liegt ein weiterer Fokus auf der Frage, ob die Extraposition von Relativsätzen sich im Lauf der untersuchten Zeit verändert hat (vgl. Voigtmann 2022: 237). Die erste Fragestellung bedarf einiger Erläuterungen. Sie bezieht sich auf die Annahme, dass neue Referenten schwerer zu verarbeiten sind als bekannte Referenten und deshalb in die Position des Nachfelds verschoben werden, wo mehr kognitive Kapazitäten frei sind, um auch verarbeitungsaufwändige Sätze und Phrasen zu verstehen.

¹² Zu ähnlichen Ergebnissen gelangt er bei den Vergleichsphrasen. Auch hier ist eine, wenn auch entgegen der Erwartung fürs Neuhochdeutsche geringere, lineare Zunahme erkennbar (vgl. Sahel 2015: 175). Die meisten Auslagerungen wurden in wissenschaftlichen Texten gefunden und die größte Zunahme bei den Prosatexten (vgl. Sahel 2015: 176).

Es wird also die gleiche Fragestellung wie in der hier untersuchten Hypothese H1 untersucht, jedoch mit unterschiedlichen Ansätzen für Verarbeitungsaufwand und der Operationalisierung davon. Voigtmann (2022: 240f.) operationalisiert den Bekanntheitsstatus in Anlehnung an Prince (1981) und Gundel et al. (1993) und unterscheidet zwischen diskursneuen, inferierbaren, nicht-salienten und salienten Referenten. Die letzten beiden Kategorien gehören eigentlich beide zu den „bekannten“ Referenten. Die Unterscheidung wird jedoch getroffen, um dem Rechnung zu tragen, dass das Arbeitsgedächtnis nicht unendlich groß ist und ein Referent vergessen werden kann, auch wenn er bereits im Diskurs aufgetreten ist (vgl. Voigtmann 2022: 241f.). Zudem werden sowohl der Matrixsatz als auch der Relativsatz selbst analysiert.

Die zu untersuchenden Sätze wurden aus drei Zeitabschnitten genommen: 1530 bis 1600, 1630 bis 1700 und 1800 bis 1870 (vgl. Voigtmann 2022: 238). Pro Zeitabschnitt wurden je 100 Sätze händisch auf die verschiedenen Referentenstatus untersucht. Dabei zeigte sich ein ausgeglichener Anteil von ausgelagerten und eingebetteten Relativsätzen im 16. und 17. Jahrhundert (vgl. Voigtmann 2022: 243f.), aber nicht im 19. Jahrhundert. Hier wurden nur 33 % ausgelagerte Relativsätze gefunden (vgl. Voigtmann 2022: 244). Auch die Auszählung der bekannten und unbekannt Referenten zeigt im 16. und 17. Jahrhundert eine klare Tendenz. Hier wurden mehr bekannte Referenten in den adjazenten Relativsätzen gefunden (vgl. Voigtmann 2022: 244). Umgekehrt wurden in den ausgelagerten Relativsätzen mehr neue Referenten annotiert (vgl. Voigtmann 2022: 245). Voigtmann (2022: 244) interpretiert das als Bestätigung für die Hypothese, dass Relativsätze ausgelagert werden, weil sie aufgrund vieler neuer Referenten schwerer zu verarbeiten sind.¹³ Anhand der deskriptiven Statistik ist zu erkennen, dass im 19. Jahrhundert mehr bekannte Referenten in den ausgelagerten Relativsätzen gefunden wurden (vgl. Voigtmann 2022: 244), während etwa gleich viele neue Referenten in den beiden Positionen im 19. Jahrhundert entdeckt wurden (vgl. Voigtmann 2022: 245). Daher werden Hinweise darauf gesehen, dass Relativsätze im 19. Jahrhun-

¹³ Die Ergebnisse werden durch Chi-Quadrat-Tests unterstützt, allerdings wird ein p-Wert von $p < 0,15$ als Signifikanzniveau festgelegt (vgl. Voigtmann 2022: 243), was die Aussagekraft der statistischen Analyse ebenso mindert wie fehlende Werte zur Effektstärke.

dert dann eingebettet werden, wenn sie viele neue Referenten enthalten (vgl. Voigtmann 2022: 245). In einem nächsten Schritt wurden die Bekanntheitsstatus der Referenten in den Matrixsätzen ermittelt und verglichen. Dabei zeigen die Matrixsätze im 16. und 17. Jahrhundert keinen Einfluss auf die Position der Relativsätze (vgl. Voigtmann 2022: 246). Dafür sind sie im 19. Jahrhundert besonders einflussreich: Wenn der Matrixsatz in diesem Jahrhundert besonders viele neue Referenten enthält, steht ein Relativsatz eher im Nachfeld (vgl. Voigtmann 2022: 246). Voigtmann (2022: 247ff.) sieht darin für die ersten beiden Jahrhunderte eine Bestätigung der Hypothese zum Zusammenhang zwischen der Relativsatzextraposition und Verarbeitungsbedingungen aufgrund von Bekanntheit, was auch in Bezug zur untersuchten Textsorte der Briefe gesetzt wird, die mehr dem mündlichkeitsnahen Spektrum (vgl. Koch und Oesterreicher 2007) zugeordnet werden können. Für das 19. Jahrhundert kann der Zusammenhang nicht bestätigt werden. Im Gegensatz zu den in den anderen Zeitstufen verwendeten Briefen wurden diese Briefe allerdings zur Veröffentlichung geschrieben, weshalb von einer geringeren Mündlichkeitsnähe ausgegangen werden muss (vgl. Voigtmann 2022: 249). Doch auch für das 19. Jahrhundert wird ein Einfluss von Verarbeitung gesehen, allerdings geht er hier vom Matrixsatz aus (vgl. Voigtmann 2022: 249): Enthält ein Matrixsatz mehr neue Referenten, wird Extraposition wahrscheinlicher. Durch diesen Wechsel des Einfluss ausübenden Satzteils wird auch die Hypothese als belegt angesehen, dass es eine Veränderung über die Jahrhunderte gibt (vgl. Voigtmann 2022: 246f.). Trotz der erwähnten statistischen Mängel und auch einer spärlichen Datenmenge, hat diese Studie den Vorteil, dass sie nah an der auch hier untersuchten Fragestellung ist und zum frühen Neuhochdeutschen durchgeführt wurde.

Das trifft auch auf die Studie von Speyer und Lemke (2017) zu. Auch sie untersuchen den Zusammenhang von Relativsatzauslagerung, Informationsstruktur und Informationsmanagement. Neben der Studie von Voigtmann (2022) bildet sie auch eine der Grundlagen für die hier untersuchten Zusammenhänge, allerdings auf einer kleineren Datenbasis. Speyer und Lemke (2017) versuchen den Bekanntheitsstatus von Referenten mit der Informati-

onstheorie nach Shannon (1948) zu kombinieren und so einen Beleg für die schwerere Verarbeitung von neuen Referenten zu finden und das zusätzlich mit Relativsatzextraposition zu kombinieren. Auch ihre Fragestellung ist also, ob Relativsatzauslagerung mit informationellen Beschränkungen und Herausforderungen zusammenhängt. Das untersuchen sie anhand eines Korpus aus Texten aus dem Deutschen Textarchiv (vgl. BBAW 2019), die unter die Rubrik Leichenpredigten fallen (Speyer und Lemke 2017: 4) und zwischen 1600 und 1630 veröffentlicht wurden. Das Korpus enthält mehr als 85.000 Sätze, aber nur vier Texte wurden für die eigentliche Auswertung untersucht.¹⁴ Da auch Bezug auf die Informationsdichte genommen wird, wurde anhand dieser Daten ein bigram-Modell auf den Lemmata der Texte mit dem SRILM-Tool (vgl. Stolcke 2002) berechnet (vgl. Speyer und Lemke 2017: 4). Die Relativsatzpositionen wurden manuell bestimmt (vgl. Speyer und Lemke 2017: 4) und die Annotation der Bekanntheitsstatus erfolgt, wie zuvor in Voigtmann (2022) beschrieben. Neben diesen beiden Maßen für Verarbeitungsaufwand nutzen Speyer und Lemke (2017: 18f.) auch noch ein selbstentwickeltes Maß, das Informationsstatus und Surprisal kombiniert. Die Idee dahinter ist folgende:

[A]ny lexical item should come with a high ‚surprisal‘ in the non-technical sense of the word when it is unused in a text for the first time [...] This ‚novelty bonus‘ should wear out after a while, and after it has been mentioned a certain number of times, it should make no difference any more. (Speyer und Lemke 2017: 5f.).

Referenten wird also ein Neuheitsbonus zu dem eigentlichen Surprisalwert addiert, der bei der ersten Erwähnung höher ist als bei später auftretenden Referenten und wieder erhöht werden kann, wenn der Referent längere Zeit nicht erwähnt wurde (vgl. Speyer und Lemke 2017: 6f.).

Die erste Auszählung der Referenten von Speyer und Lemke (2017: 6) bestätigt die Ergebnisse von Voigtmann (2022). In adjazenten Relativsätzen stehen mehr bekannte Referenten. Die zweite Auswertung betrifft die bloßen Surprisalwerte ohne den Neuheitsbonus, die auch als durchschnittliche und kumulierte Werte (Kapitel 5.2) berechnet wurden, um Aussagen über den ganzen Relativsatz und den Matrixsatz treffen zu können (vgl. Speyer und Lemke

¹⁴ Speyer und Lemke (2017) erwähnen dabei nicht, wie groß dieses Trainingskorpus ist.

2017: 7). Dabei wurden adjazente Relativsätze so behandelt, als seien sie ausgelagert (vgl. Speyer und Lemke 2017: 7). Dabei zeigt sich, dass der Einfluss des Matrixsatzes zu vernachlässigen ist und der durchschnittliche Surprisalwert der eingebetteten Relativsätze kleiner ist als der der ausgelagerten Relativsätze (vgl. Speyer und Lemke 2017: 7). Ein ähnlicher Effekt zeigt sich bei der Kombination aus Informationsstatus und Informationsdichte (vgl. Speyer und Lemke 2017: 7). Auch hier weisen die ausgelagerten Relativsätze deutlich höhere Werte auf als die eingebetteten (vgl. Speyer und Lemke 2017: 7). Allerdings unterscheiden sich die beiden Berechnungsarten darin, dass bei dieser auch der Matrixsatz an Bedeutung gewinnt: Bei eingebetteten Relativsätzen besitzen Matrix- und Relativsatz fast die gleichen Surprisalwerte (vgl. Speyer und Lemke 2017: 7).

Die Ergebnisse zwischen den einzelnen Berechnungsarten sind so ähnlich, dass die Hypothese als bestätigt angesehen und Relativsatzauslagerung in Verbindung mit kognitiver Verarbeitung gebracht wird (vgl. Speyer und Lemke 2017: 26). Einschränkend zu den Ergebnissen ist allerdings zu sagen, dass es sich um einen sehr kleinen Datensatz handelt und nur der Faktor der Informationsdichte untersucht wird, obwohl Studien zeigen und Grammatiken beschreiben, dass Relativsatzauslagerung auch mit anderen Faktoren wie der Länge, der Restriktivität oder der Distanz zwischen Nachfeld und Antezedens in Verbindung gebracht werden muss. Zudem erfasst ein Bigram-Sprachmodell den Kontext nur unzureichend, auch wenn die Anwendung anderer Modell bei einem kleinen Datensatz entsprechend schwierig ist (vgl. Kapitel 5.1). Insgesamt kann diese Studie als Inspiration für die in dieser Arbeit durchgeführte Untersuchung dienen, wobei die Probleme wie die Auswahl des Sprachmodells und die fehlende Betrachtung der anderen Faktoren aber aufgegriffen und optimiert wurden (Kapitel 5.1).

4. Korpus

Um die im Rahmen dieses Dissertationsprojekts durchgeführte Studie geht es ab diesem Kapitel. Es beginnt mit der Korpusbeschreibung und der Legitimation des frühen Neuhochdeutschen als Untersuchungszeitraum sowie der Wahl wissenschaftlicher Texte (Kapitel 4.1). Zudem werde ich das Deutsche Textarchiv („DTA“, BBAW 2019) als Bezugsquelle des Korpus‘ vorstellen. Im Anschluss daran erfolgt die Auflistung der verwendeten Texte (Kapitel 4.2). Danach werden das Annotationsvorgehen der Satzklammern, der Relativsätze sowie der Präpositional- und Nominalphrasen (Kapitel 4.3) sowie abschließend die Bereinigung der Daten (Kapitel 4.4) dargestellt.

4.1. Auswahl der Texte

Einer der größten Vorteile der *Korpuslinguistik* ist, dass sie die Auseinandersetzung mit natürlichen Daten ermöglicht (vgl. Lemnitzer und Zinsmeister 2015), und es zusätzlich erlaubt, Sprachstufen zu untersuchen, bei denen Experimentdaten natürlicherweise nicht mehr zur Verfügung stehen. Somit sind Korpora für historische Fragestellungen das Mittel der Wahl.

Unabhängig davon, ob mithilfe einer Korpusstudie synchrone oder diachrone Fragestellungen bearbeitet werden, muss das Korpus diverse Anforderungen erfüllen. Dazu zählt zum einen eine Größe, die auch eine quantitative Analyse ermöglicht, gefolgt von der Vergleichbarkeit der Texte, die Vermeidung von Einflüssen wie Übersetzungen, wenn diese nicht explizit Thema der Untersuchung sind, und vom Versmaß, was besonders in historischen Daten problematisch werden kann (vgl. z. B. Lemnitzer und Zinsmeister 2015; Andresen und Zinsmeister 2019). Zudem sollte eine breite Streuung von Text-

sorten und/oder Autoren gewährleistet, das Korpus also ausgewogen sein (vgl. Andresen und Zinsmeister 2019: 17). Im Fall diachroner Untersuchungen müssen diese Kriterien auf jeden der untersuchten Zeitabschnitte zutreffen. Für die im Rahmen dieser Arbeit verfolgte Fragestellung müssen einige zusätzliche Kriterien erfüllt sein, darunter die Etablierung des Satzrahmens (vgl. Wöllstein 2014).

Aufgrund dieses Kriteriums der fest etablierten Satzklammer ergibt sich, dass das Frühneuhochdeutsche die frühestmögliche Zeitstufe für die Untersuchung bildet (vgl. z. B. Erben 2000b: 1588). Diese Periode ist allerdings noch stark von dialektalen Einflüssen geprägt, da sich eine gemeinsame Hochsprache erst im 16. Jahrhundert zu entwickelt beginnt (vgl. z. B. Polenz und Moulin 2021: 100ff.). Im frühen Neuhochdeutschen, seit etwa der Mitte des 17. Jahrhunderts also, hat sich der Einfluss der Dialekte aufgrund des Buchdrucks und der Verfügbarkeit von einheitlich geschriebenen Texten im gesamten deutschen Sprachraum verringert (vgl. z. B. Sonderegger 1979; Polenz und Moulin 2021). Daraus resultiert, dass sich auch die Anzahl syntaktisch vergleichbarer Texte im 17. Jahrhundert erhöht. Da außerdem eine große Textgrundlage für die computerlinguistisch gestützte Analyse notwendig ist (Kapitel 5.1), ergibt sich der Vorteil dieses Untersuchungszeitraums. Innerhalb dieses Zeitraums stehen zudem verschiedene Textsorten zur Verfügung, die sich zur Untersuchung von Auslagerung anbieten. Dazu zählen beispielsweise *Kanzleitexte*, *Briefe*, *Zeitungsartikel* und die von mir verwendeten *wissenschaftlichen Texte*.

Die Untersuchung von Extraposition im Kontext bestmöglicher Informationsverteilung (vgl. z. B. Shannon 1948; Levy und Jaeger 2007) setzt voraus, dass die Texte einen kommunikativen Zweck erfüllen und sich nicht als Kunstwerk verstehen. Literarische Texte wie Romane sind daher für die Untersuchung eher ungeeignet. Dagegen erfüllen Kanzleitexte das Kriterium der Informationsvermittlung. Sie waren dazu gedacht, rechtliche Angelegenheiten der Bürger zu verschriftlichen (vgl. z. B. Polenz und Moulin 2021). Aufgrund ihres informationellen Charakters sollten sie ein ebenmäßiges Informationsprofil aufweisen. Da ihnen aber zugleich eine Orientierung an älteren

Sprachstufen und ein Streben nach Prestige zugeschrieben wird, das dem natürlichen Sprachgebrauch entgegengesetzt ist (vgl. z. B. Konopka 1996; Takada 1998), sind sie für diese Untersuchung ungeeignet.

Elspaß (2005) zeigt ausführlich die Vorteile von Briefen für die Untersuchung alltagsnaher Sprache auf. Zu seinen Argumenten für ihre Verwendung zählen ihre Natürlichsprachlichkeit, ein kaum erkennbarer Einfluss von Prestigegegedanken und auch die Mündlichkeitsnähe, wo sprachliche Veränderungen nach Koch und Oesterreicher (2007) eher ihren Ursprung nehmen. Um den Einfluss regionaler Varianten einzugrenzen, bietet es sich an, nur Texte einer bestimmten Region zu betrachten, um die Vergleichbarkeit zu gewährleisten, wie es von Ebert (1980) getan wurde.¹ Elspaß (2005) erläutert auch soziale Einflussfaktoren, welche die Schreibtätigkeit beeinflussen können wie der Zugang zu Textvorlagen aller Art und die Häufigkeit, mit der das eigene Schreiben trainiert wird. Obwohl diese Argumente valide für die Wahl der Textsorte sind, können Briefe für meine Studie nicht herangezogen werden, weil es kein ausreichend großes, digitalisiertes und korpuslinguistisch aufbereitetes Briefkorpus des frühen Neuhochdeutschen gibt. Der Aufwand, ausreichend viele Briefe digital aufzubereiten, übersteigt die Praktikabilität für diese Arbeit.

Abgesehen von Briefen würden sich für die Untersuchung auch Zeitungstexte anbieten. Demske-Neumann (1990: 240) nutzt diese für ihre Untersuchungen, führt aber selbst die Schwachstellen des Genres aus. Weil Zeitungen bereits im Frühneuhochdeutschen unter Zeitdruck produziert wurden, ergeben sich kommunikativ ungünstige Konstruktionen (vgl. Demske-Neumann 1990). Dennoch sind Zeitungstexte näher an den Kriterien, die für diese Untersuchung gelten, als zum Beispiel Kanzleitexte. Zeitungen existieren zwar auch in digital aufbereiteter Form bis ins 21. Jahrhundert hinein, doch besteht das Problem, dass die Texte im *Referenzkorpus Frühneuhochdeutsch* (vgl. Demske 2019) nur das Frühneuhochdeutsche abdecken und in computerlinguistisch bearbeitbarer Form nicht bis ins 19. Jahrhundert reichen.

¹ Er bezieht sich in seiner Untersuchung beispielsweise auf ein Korpus aus Nürnberger Briefen, die jedoch nicht digitalisiert vorliegen und einen zu geringen Umfang besitzen, um den Anforderungen einer modernen Korpusanalyse gerecht zu werden.

Neben den oben genannten grundsätzlichen Kriterien der Authentizität, Größe und Vergleichbarkeit, muss also allgemein auch ein Einfluss von diatopischen Unterschieden, Prestige und auch soziolinguistischen Faktoren wenigstens kontrolliert werden. Für die vorliegende Untersuchung ist neben den genannten Faktoren auch die Digitalisierung und die korpuslinguistische Bearbeitung der Texte ausschlaggebend.

Daher beschäftige ich mich mit wissenschaftlichen Texten.² Zum einen besitzen diese einen überregionalen Anspruch: Die von mir betrachteten Autoren hatten größtenteils verschiedene Wirkungsstätten (Anhang A), was zusätzlich zum allgemeinen Abbau dialektaler Einflüsse in gedruckten Texten dieser Zeit (vgl. z. B. Polenz et al. 2013) gegen relevante diatopische Unterschiede im Korpus spricht. Zum anderen sind auch immer wieder Zitate und Verweise auf andere wissenschaftliche Arbeiten oder Wissenschaftler in den Texten anzutreffen, sodass ebenfalls davon ausgegangen werden kann, dass sich die Schreiber mit weiteren Werken aus anderen Sprachregionen auseinandergesetzt haben. In den frühesten Texten finden sich zwar noch vermehrt lateinische oder griechische Passagen, doch nimmt deren Anteil über die Jahrzehnte hinweg ab. Darin spiegelt sich ebenfalls ein Wandel der Wissenschaftssprache wider (vgl. z. B. Klein 2011), die zunehmend auf die Veröffentlichung in der Muttersprache der Schreibenden ausgerichtet und in dieser Zeitspanne nicht so stark auf die Nutzung einer *Lingua Franca* fokussiert ist, wie es heute zum Beispiel mit dem Englischen der Fall ist.

Des Weiteren steht bei wissenschaftlichen Texten auch die Wissensvermittlung im Vordergrund. Daher und weil sich wissenschaftliches Schreiben in der Muttersprache erst im 17. Jahrhundert entwickelt hat (vgl. z. B. Polenz und Moulin 2021: 165), ist es unwahrscheinlich, dass sie einen archaischen oder prestigeträchtigen Stil vorweisen oder sich an älteren Texten dieses Genres orientieren. Sie wollen informieren und grenzen sich insofern von den Kanzleitexten ab, als wissenschaftliche Texte nicht zusätzlich auf eine lang bestehende Herrschaftsstruktur verweisen müssen. Von den Zeitungstexten unterscheidet sie, dass sie mit weniger Zeitdruck verfasst und publiziert wur-

² Wie bereits in der Einleitung (Kapitel 1) gesagt wurde, spielt auch die Förderung dieser Untersuchung durch den SFB 1102 eine (allerdings eher untergeordnete) Rolle.

den. Überarbeitungen vor dem Druck werden so wahrscheinlicher. Dadurch wird die Wahrscheinlichkeit verringert, dass die hier untersuchten Nachfeldbesetzungen reine Korrektur- oder Ergänzungsphänomene sind (vgl. z. B. Altmann 1981). Daher kann davon ausgegangen werden, dass Extraposition in dieser Textsorte öfter zu einem kommunikativen und verarbeitungsrelevanten Zweck genutzt wird und dass die Fälle, in denen es sich nur um das Hinzufügen eines im Mittelfeld vergessenen Elements, sprich um einen Nachtrag, handelt, statistisch nicht ins Gewicht fallen.

Konträr zu Briefen ist sowohl der Verfasser- als auch der Adressatenkreis der wissenschaftlichen Texte homogen.³ Wissenschaftler schreiben in aller Regel für andere Wissenschaftler, die einen ähnlichen Bildungsgrad besitzen und ein mindestens ähnliches Fachgebiet haben. Auch das kann sich in effizienter Kommunikation widerspiegeln. Neben der Homogenität der Adressaten kann auch davon ausgegangen werden, dass die Autoren selbst miteinander vergleichbar sind. Sie alle besitzen einen hohen Bildungsgrad und kommen regelmäßig in Kontakt mit geschriebenen Texten. Abweichungen durch unterschiedliche Bildung, Literarizität oder auch geschlechtsspezifische Unterschiede – bei den Autoren handelt es sich ausschließlich um Männer – können daher weitestgehend ausgeschlossen werden, weshalb die gefundene sprachliche Variation mit einer höheren Wahrscheinlichkeit auf tatsächliche kommunikative Prozesse und weniger auf individuelle Differenzen zurückgeführt werden kann. Wissenschaftliche Texte lassen sich zudem fest in der grafischen Rezeption verankern (vgl. Koch und Oesterreicher 2007). Ein eher praktisches Argument, das für die Verwendung wissenschaftlicher Texte spricht, ist, dass Texte dieses Genres in großer Zahl über das Deutsche Textarchiv (vgl. BBAW 2019) verfügbar sind, wodurch eine breit aufgestellte quantitative Analyse möglich wird.

Allerdings birgt die Nutzung wissenschaftlicher Texte auch einige Nachteile. Besonders im frühesten der untersuchten Zeitabschnitte (1650 bis 1700) kann ein starker Einfluss des Lateinischen angenommen werden, da es noch nicht viele deutsche Quellen dieser Textsorte gab, muttersprachlich konzipier-

³ Diese Behauptung wird in den Diskussionskapiteln 6.5, 7.2 und 8.4 relativiert.

te Texte überhaupt erst seit dem 15. Jahrhundert angenommen werden und es sich dabei nicht unmittelbar um wissenschaftliche Texte im heutigen Sinn handelt (vgl. Drozd und Seibicke 1973: 16). Zudem bestritten die Schreiber zwangsläufig einen großen Teil ihrer Ausbildung mit lateinischen oder griechischen Quellen (vgl. Polenz und Moulin 2021: 165f.). Selbst im 16. und 17. Jahrhundert handelte es sich bei wissenschaftsnahen Texten häufig noch um Übersetzungen aus dem Lateinischen (vgl. Drozd und Seibicke 1973: 17).

Das kann problematisch sein. Das Lateinische zeichnet sich vor allem durch seine freie Wortstellung und teils große Distanzen zwischen sprachlich Zusammengehörendem aus. Diese Vorbilder könnten Ausklammerung in den frühen Zeitabschnitten begünstigen. Gleichzeitig besteht hier jedoch auch ein Spannungsfeld zwischen den literarischen, lateinischen Vorlagen und den zeitgenössischen, deutschsprachigen Vorlagen. Auch muss man sich die Frage stellen, inwiefern Mündlichkeit – denn die Autoren sprachen ja Deutsch – und Schriftlichkeit voneinander unbeeinflusst ablaufen können. Untersuchungen von Elspaß (2005) ergeben, dass das Mündliche durchaus auch Einfluss auf die schriftliche Sprache ausübt, wenn beispielsweise dialektale Besonderheiten in Briefen nachweisbar sind. Es kann daher auch für das wissenschaftliche Schreiben nicht ausgeschlossen werden, dass das Sprechsprachliche Niederschlag im Schriftlichen findet, auch wenn die schriftlichen Vorlagen anderen Mustern und Besonderheiten folgen. Mögliche Abweichungen in den Texten des frühen 17. Jahrhunderts lassen sich also noch nicht eindeutig auf die lateinischen Vorlagen attribuieren. Die generellen Veränderungen über den untersuchten Zeitraum von 250 Jahren werden bei der Betrachtung von Hypothese H3⁴ berücksichtigt, auch wenn es nicht Zweck dieser Arbeit ist, den Einfluss des Lateinischen zu analysieren.

Zusätzlich muss bedacht werden, dass sich das Genre der wissenschaftlichen Texte wiederum in geistes- und naturwissenschaftliche Texte unterteilen lässt.⁵ Dieser Tatsache trage ich mit der Wahl zweier Subkorpora Rechnung.

⁴ H3: „Über die Zeit wird der Einfluss der Informationsdichte auf Auslagerung geringer.“

⁵ Auch wenn sich die Geisteswissenschaften erst im 19. Jahrhundert als eigenes Feld etabliert haben (vgl. z. B. Dilthey 1910), finden sich auch im 17. und 18. Jahrhundert Texte, die nach heutiger Definition in den Fachbereich der Geisteswissenschaften einzuordnen sind, obwohl man sie nach heutigen Maßstäben oft eher populärwissenschaftlich verorten würde.

Als Vertreter für naturwissenschaftliche Texte werden medizinische Artikel und Monografien untersucht. Die geisteswissenschaftlichen Texte werden von theologischen Publikationen repräsentiert.

Medizinische Texte bieten den Vorteil, dass sie zwar die Fachterminologie der naturwissenschaftlichen Texte teilen, aber weniger mathematische oder chemische Formeln enthalten als beispielsweise Texte aus der Chemie, Biologie oder Physik.⁶ Die theologischen Texte wurden gewählt, da sie im untersuchten Zeitraum in größerer Zahl zur Verfügung stehen und sich die Theologie nicht erst als eigenständige Wissenschaft im 19. Jahrhundert herausgebildet hat. Zudem ist das Thema der theologischen Texte häufig biblisch. Dieses hat zudem einen hohen Verbreitungsgrad und eine große Relevanz für den Alltag der Menschen.

Zwischen den beiden Genres sind allerdings auch unabhängig von ihren thematischen Schwerpunkten Unterschiede zu vermuten. Zwar wurden beide unter „Wissenschaft“ im DTA (vgl. BBAW 2019) eingeordnet, aber ein Blick auf die genaue Auswahl der Texte (Tabellen 4.2 und 4.3) zeigt, dass einige der theologischen Texte nach heutigen Gesichtspunkten keine klassischen wissenschaftlichen Texte darstellen und eher für ein breiteres Publikum gedacht waren. Letzteres ist bei den medizinischen Texten nicht zu erwarten. Auch dürfte die Schreibpraxis der beiden Gruppen unterschiedlich sein. Während die Mediziner wohl tatsächlich für andere Mediziner schreiben und dabei zahlreiche Neuerungen in ihrem Fachgebiet aufgreifen mussten (vgl. z. B. Eckart 2021), waren die Theologen an ein heterogeneres Publikum und das Verfassen anderer Textsorten, beispielsweise Predigten, gewöhnt. Beides kann zu Unterschieden hinsichtlich der Sprecher- und Hörerorientierung zwischen den beiden Gruppen führen, die beachtet werden müssen.

Zusammenfassend ist zu sagen, dass wissenschaftliche Texte den Untersuchungsgegenstand dieser Arbeit bilden, weil sie der Informationsvermittlung dienen, der Einfluss von Prestige und archaischen Strukturen weitestgehend ausgeschlossen werden kann, sie mit weniger Zeitdruck publiziert wur-

⁶ Für naturwissenschaftliche, englische Texte zeigen Degaetano-Ortlieb und Teich (2018) und Degaetano-Ortlieb, Kermes, Khamis und Teich (2019) eine Veränderung von Fachtermini, die sich vor allem in der Auswahl des Trainingskorpus widerspiegeln muss (Kapitel 5.1).

den als Zeitungsartikel und große soziolinguistische Unterschiede zwischen den Autoren aufgrund der eher heterogenen Gruppe ausgeschlossen werden können. Ein weiteres, praktisches Argument für ihre Wahl ist das Vorhandensein einer großen, digitalisierten Datenbasis im *Deutschen Textarchiv* (vgl. BBAW 2019), über das der folgende Abschnitt einige Informationen liefert.

Das Deutsche Textarchiv Das Deutsche Textarchiv („DTA“, BBAW 2019) besteht aus einer Sammlung verschiedener Gattungen und Epochen mit dem Ziel, einen Überblick über die Entwicklung der deutschen Sprache vom 16. bis zum 20. Jahrhundert mit Stichproben aus Zeitungen, Romanen, Gebrauchsliteratur und wissenschaftlichen Texten zu geben. Die Texte sind digitalisiert und für die korpuslinguistische Analyse vorverarbeitet, sprich normalisiert, tokenisiert, lemmatisiert und mit POS-Tags versehen (vgl. BBAW 2019).

Die beiden Vorteile des DTA sind die große Vielfalt an Genres und der Vorverarbeitungsschritt der Lemmatisierung, da für diese Untersuchung besonders die lexikalische Verarbeitung statt der grammatischen relevant ist. Daher wird die Analyse der Informationsdichte an Lemmata durchgeführt, weil Verzerrungen durch unterschiedliche Flexionen oder orthographische Inkonsistenzen, die typisch für das Frühneuhochdeutsche sind (vgl. z. B. Ebert et al. 1993), so nicht ins Gewicht fallen. Allerdings gestaltet sich die Textauswahl im DTA sehr willkürlich, was wiederum den eigenen Auswahlprozess für das Testkorpus einschränkt, wenn vergleichbare Texte gewählt werden sollen, die nicht vom gleichen Autor stammen.

Die Probleme, die das DTA außerdem mit sich bringen, liegen in der Qualität der Annotation. Diese wurde rein automatisch getätigt, das bedeutet, dass die Fehlerquote in den Annotationen sehr hoch und die Genauigkeit der Annotationen eher schlecht ist (vgl. Ortmann 2020; 2023). Ein manueller Goldstandard scheint nicht erstellt worden zu sein und auch das Training der Annotationen wurde nicht auf historischen Daten getätigt, sondern mit modernen, was wiederum vor allem die Qualität der POS-Tags stark beeinträchtigt. Aus diesem Grund ist eine automatische Suche nach dem Phänomen nicht möglich. Dennoch überwiegen die eben skizzierten Vorteile den Nachteilen und rechtfertigen daher die Verwendung des DTA.

4.2. Zusammensetzung des Korpus

Das Korpus besteht aus zufällig ausgewählten medizinischen und theologischen Texten aus dem DTA (vgl. BBAW 2019), unter der Voraussetzung, dass es sich nicht um Übersetzungen handelt und sie zum Zeitpunkt des Downloads (2018) in die DTA-Kategorien ‚Medizin‘ oder ‚Theologie‘ eingeordnet waren.⁷ Insgesamt wurden so zehn medizinische und 12 theologische Texte ausgewählt. Diese Texte verteilen sich wie folgt auf diese Abschnitte und die Genre (Tabelle 4.1).

	1650-1700	1700-1750	1750-1800	1800-1850	1850-1900
Medizin	2	2	1	2	2
Theologie	2	3	2	2	3

Tabelle 4.1.: Anzahl der medizinischen und theologischen Texte pro Zeitabschnitt und die Tokens im Testkorpus.

Die Tabellen 4.2 und 4.3 listen die verwendeten Texte auf und stellen dar, wann und wo sie veröffentlicht wurden. In Anhang A werden kurze Informationen über die Autoren gegeben. Dies dient einem Überblick über die Berufe der Autoren und ihren Werdegang. Auch wenn im Untersuchungszeitraum (1650 bis 1900) die dialektalen Einflüsse auf die Schriftsprache weitestgehend vernachlässigbar werden können (vgl. z. B. Polenz 2020: 109ff.), soll mit dem Wissen über die Reisetätigkeiten und die Wirkungsplätze der Autoren ein weiteres Argument gegen den Einbezug von Dialekträumen in die Analyse geschaffen werden.⁸ Aufgrund der unterschiedlichen Verfügbarkeit von medizinischen und theologischen Texten im DTA werden die im Korpus abgedeckten 250 Jahre in 50 Jahresabschnitte unterteilt. Im Anschluss wird ermittelt, wie viele ausgelagerte Relativsätze, Nominalphrasen und Präpositionalphrasen pro 50-Jahres-Stufe gefunden werden.

⁷ Auf das Korpus inklusive der Annotationen kann zugegriffen werden über: <https://github.com/SFB1102/C6Samples>.

⁸ Bei der Annotation wurde ich von den studentischen Hilfskräften Nils Dörr, der die Texte Ludwig (1856), Egger (1895) und Löhe (1847) bearbeitet hat, und Katharina Wilhelm, die die Texte Roth (1692), Benner (1739), Bengel (1751), Hanssen (1731) und Strauß (1835) bearbeitet hat, unterstützt. Ihre Arbeit wurde von mir überprüft und, wo nötig, korrigiert.

Autor	Genre	Ort	Jahr	Titel	Tokens ⁹
Abel, Heinrich Kaspar	Medizin	Leipzig	1699	Wohlerfahrner Leib-Medicus Der Studenten, welcher So wohl allen auf Schulen Gymnasiis und Universitäten Lebenden oder auf Reisen begriffenen gelehrten Personen/ als auch allen Menschen insgemein die nöthigsten Reguln und herrlichsten Artzeneyen mittheilet/ Krafft deren sie nicht allein die Gesundheit nechst Gott erhalten/ sondern auch die zugestossenen Kranckheiten abwenden und vertreiben können	49 972
Bengel, Johann Albert	Theologie	Stuttgart	1751	Abriß der so genannten Brüdergemeine, in welchem die Lehre und die ganze Sache geprüft, das Gute und Böse dabey unterschieden, und insonderheit die Spangenbergische Declaration erläutert wird	58 344
Benner, Johannes Hermann	Theologie	Frankfurt am Main, Leipzig	1739	Christliches Bedencken von dem vorsetzlichen Meineid, wodurch das Recht des Nechsten wissentlich gekräncket wird Ob, und wie ein solcher Meineid von Gott vergeben werde?	10 150

Tabelle 4.2.: Im Korpus verwendete Texte 1.

⁹ Die Anzahl der Tokens wurde der Dokumentation des DTA entnommen (vgl. BBAW 2019).

Autor	Genre	Ort	Jahr	Titel	Tokens
Bräuner, Johann Jacob	Medizin	Frankfurt am Main	1714	Pest-Büchlein: Oder Kurtzer/ doch gründlicher Unterricht Von der [...] Seuche der Pestilenz, In welchem wohlmeynend gezeiget wird/ Woher solche kommet/ wie sie ansteckend und fortgeplantzet/ auch wie sich [...] welche mit solchen Pest-inficirten Patienten umgehen müssen/ auch sonst jederman darbey zu verhalten habe: Nicht weniger auch/ wie man solche Seuche von sich durch Gottes Gnade abwenden/ und darwider glücklich präserviren und curiren kann Alles nach der sicherest und bewährtesten Methode [...] treulich entworffen/ und zu jedermans Nutzen und Gebrauch zum Druck befördert	73 371
Carus, Carl Gustav	Medizin	Leipzig	1820	Lehrbuch der Gynäkologie. Bd. 1.	119 871
Daumer, Georg Friedrich	Theologie	Münster	1895	Die dreifache Krone Rom's. Versuch einer neuen Beleuchtung und Charakterisirung des römisch-katholischen Priester- und Kirchenthums, namentlich was dessen elementare und principielle Inhaltsbestimmungen und deren vorläufige Begründung und Erscheinung in vorchristlicher Zeit und Welt betrifft	32 656

Tabelle 4.2.: Im Korpus verwendete Texte 2.

Autor	Genre	Ort	Jahr	Titel	Tokens
Egger, Augustinus	Theologie	unbekannt	1895	Der christliche Vater in der modernen Welt, Erbauungs- und Gebetbuch	86 146
Gall, Franz Joseph	Medizin	Wien	1791	Philosophisch-Medicnische Untersuchungen über Natur und Kunst im kranken und gesunden Zustande des Menschen	154 263
Hanssen, Petrus	Theologie	Lübeck, Hamburg	1731	Achtzig erläuterte Grund-Fragen. in welchen Die Lehre der Evangelisch-Lutherischen Kirche von dem Mittler-Ampt Jesu Christi und dessen Einfluß zu einem heiligen Leben wider ihre Feinde überhaupt, insonderheit aber wider den beruffenen Johann Conrad Dippel [...] solchergestalt befestiget, daß die dahin gehörige Warheiten nach Vernunft und Schrift auseinander hergeleitet, folglich in einer nohtwendigen Verbindung und unüberwindlichen Gewißheit dargestellt werden: [...]	47 869
Hasak, Max	Theologie	Berlin	1893	Die Predigtkirche im Mittelalter	10 193
Koch, Robert	Medizin	Leipzig	1878	Untersuchung über die Aetiologie der Wundinfectionskrankheiten	24 246
Kraepelin, Emil	Medizin	Jena	1892	Ueber die Beeinflussung einfacher psychischer Vorgänge durch einige Arzneimittel. Experimentelle Untersuchungen	81 738

Tabelle 4.2.: Im Korpus verwendete Texte 3.

Autor	Genre	Ort	Jahr	Titel	Tokens
Löhe, Johann Konrad Wilhelm	Theologie	Nürnberg	1847	Prediget das Evangelium aller Creatur! Predigt am Missionsfeste zu Nürnberg den 17. Juni 1847	81 738
Ludwig, Karl Friedrich Wilhelm	Medizin	Heidelberg, Leipzig	1856	Lehrbuch der Physiologie des Menschen	190 702
Niviandts, Friedrich	Theologie	Köln	1708	Güldenes Schwerd. Womit Wider jetziger Zeit ein einreissende Ketzeren bevorab wider den Lutheranismum und Calvinismum	95 403
Purmann, Gottfried	Medizin	Halberstadt	1680	Der rechte und wahrhafftige Feldscher Oder/ Die rechte und warhafftige Feldschers-Kunst	46 971
Pahl, Johann Gottfried von	Theologie	Heilbronn	1799	Leben und Thaten des ehrwürdigen Paters Simpertus	43 288

Tabelle 4.3.: Im Korpus verwendete Texte 4.

Autor	Genre	Ort	Jahr	Titel	Tokens
Reil, Johannes Christian	Medizin	Halle (Saale)	1803	Rhapsodien über die Anwendung der psychischen Curmethode auf Geisteszerrüttungen	96 532
Rotth, Albrecht Christian	Theologie	Halle	1692	Eylfertiges Bedencken über M. August Hermann Franckens/ Pastoris zu Glauche vor Halle/Seine Schutz-Predigt/ Ob Er Durch dieselbe seinen Zweck/ den er auf dem Titul gedachter Predigt/ berühret hat/ erlanget oder nicht? [...]	7 934
Spener, Philipp Jacob	Theologie	Frankfurt am Main	1676	Pia Desideria. Oder Hertzliches Verlangen/ Nach Gottgefälliger Besserung der wahren Evangelischen Kirchen/ sampt einigen dahin einfältig abzweckenden Christlichen Vorschlägen Philipp Jacob Speners. [...]	58 680
Strauß, David	Theologie	Tübingen	1835	Das Leben Jesu, kritisch bearbeitet, Band 1	204 870
Unzer, Johann August	Medizin	Halle	1746	Gedanken vom Einfluß der Seele in ihren Körper	34 027

Tabelle 4.3.: Im Korpus verwendete Texte 5.

4.3. Annotationspraxis

In den in Kapitel 4.2 aufgelisteten Texten wurden manuell die Sätze gesucht, die mindestens eines der gesuchten Phänomene enthalten.¹⁰ In diesen Sätzen habe ich neben den Satzklammern exhaustiv die extraponierten Nominalphrasen, Präpositionalphrasen und Relativsätze hinsichtlich ihrer Position annotiert. Bei allen Phänomenen wurden die Länge und die Surprisalwerte der Nominal- und Präpositionalphrase beziehungsweise des Relativsatz ermittelt und bei Letzteren zusätzlich das Antezedens und die Restriktivität annotiert. Des Weiteren erfolgte für die Relativsätze die Bestimmung der Distanz zwischen dem Antezedens und dem Relativsatzbeginn. Die Annotationen erfolgten im Programm *WebAnno* (vgl. Yimam, Gurevych, de Castilho und Biemann 2013).¹¹ *WebAnno* wurde ausgewählt, weil es webbasiert ist, verschiedene Annotatoren zulässt und die Einrichtung eigener Annotationsmasken erlaubt. Die Kriterien, nach denen die einzelnen Annotationen erfolgten, werden im Folgenden beschrieben.¹²

4.3.1. Satzklammern

Ein grundlegender Bestandteil der Annotation ist die Markierung der Satzklammern. Sie bildet die Grundlage dafür, Auslagerung überhaupt bestimmen zu können. Ich habe mich bei der Annotation der Satzklammern stark an den in Kapitel 2.2.1 dargestellten Kriterien nach Wöllstein (2010; 2014) orientiert, weiche aufgrund der bessere Operationalisierbarkeit für die Arbeit mit den Texten aus dem DTA (vgl. BBAW 2019) in manchen Aspekten davon ab.

¹⁰ Teile des in diesem Kapitel vorgestellten Materials flossen auch in die Veröffentlichungen Voigtmann und Speyer (2021; 2022) und Voigtmann und Speyer (2023) ein.

¹¹ Das Programm wurde von Stefanie Dipper und Katrin Ortman für die vorliegende Arbeit bereitgestellt und von letzterer während der Arbeit damit technisch betreut.

¹² Zusätzlich zu den hier beschriebenen Elementen wurden auch Adjektiv- und Adverbialphrasen und komparative Elemente annotiert. Deren Anzahl war allerdings so gering, dass eine statistisch fundierte Auswertung der Daten nicht möglich war. Deshalb wurden sie aus der Analyse dieser Arbeit ausgeschlossen und ihre Annotation wird in diesem Kapitel auch nicht besprochen.

Linke Satzklammer Die linke Satzklammer, die im Annotationstool als LK bezeichnet wird, grenzt zu ihrer Linken das Vorfeld und zu ihrer Rechten das Mittelfeld ab und enthält neben subordinierenden Konjunktionen finite Verbformen, Auxiliare und Modalverben (vgl. Wöllstein 2014).

Neben den subordinierenden Konjunktionen zähle ich auch Relativpronomen und -partikeln, inklusive möglicher dazugehöriger Präpositionen zu den Elementen, die in der linken Satzklammer stehen können. Die Entscheidung, diese Elemente als linke Satzklammer zu annotieren, fällt aufgrund der Operationalisierbarkeit und zu Ungunsten der kanonischen Position der Relativsatzeinleiter. Um der tatsächlichen Position der Relativsatzeinleiter im topologischen Feldermodell Rechnung zu tragen, müsste auch mindestens das Vorfeld annotiert werden. Zum einen wäre das in WebAnno (vgl. Yimam et al. 2013) allerdings zu zeitaufwendig gewesen und würde für alle anderen Sätze keinen nennenswerten Vorteil bringen, da die relevanten Felder durch ihr Verhältnis zu den Satzklammern differenziert werden können. Zum anderen kommt in keinem der gefundenen Relativsätze eines der beispielsweise von Fleischer (2005) beschriebenen, Nebensatzeinleitenden Elemente zusätzlich zum Relativum vor. Daher wird die Funktion des Relativums als Nebensatzeinleiter über seine Form und weiteren Eigenschaften, die es im Vorfeld verorten, gestellt.

Rechte Satzklammer Die Annotation der rechten Satzklammer, die RK in WebAnno genannt wird, erfolgt in vollständiger Übereinstimmung mit Wöllstein (2014). Zur rechten Satzklammer zählen verbale Partikel, Affixe und verbale Bestandteile wie der infinite Teil des Prädikats in Verbzweitsätzen. In Verbletztsätzen zählen sowohl alle infiniten als auch finiten Bestandteile des Prädikats zur rechten Satzklammer. Funktionsnomen werden in Übereinstimmung mit Presch (1977) in der rechten Klammer gesehen. Auch andere nicht-verbale Bestandteile, die im Oberfeld der rechten Satzklammer (vgl. z. B. Bech 1955) stehen, gehören zu dieser Klammer. Negationen und Prädikative werden aber als Elemente des Mittelfelds betrachtet.

4.3.2. Relativsätze

In diese Arbeit fließen ausschließlich attributive Relativsätze, in WebAnno als RELC bezeichnet, ein, weil freie Relativsätze kein im Satz vorhandenes Antezedens besitzen und weiterführende Relativsätze ausschließlich im Nachfeld stehen und ihren ganzen Matrixsatz als Antezedens wählen. Die Definition von attributiven Relativsätzen folgt Hentschel und Weydt (2013: 383f.): Relativsätze sind Sätze, die sich attributiv auf ihr Antezedens beziehen und die durch die Relativpronomen *der, die, das* oder *welcher, welche, welches* eingeleitet werden (vgl. z. B. Eisenberg und Thieroff 2020; Helbig und Buscha 2005).¹³ Neben diesen Relativpronomen wird auch die Relativpartikel *so* zu den hier verwendeten Relativsatzeinleitern gezählt. Zwar fehlen ihr Flexionseigenschaften, die Relativpronomen auszeichnen, aber *so* ist unabhängig von der Art des Nomens, welches das Antezedens zum Relativsatz bildet, und es besitzt keine Eigensemantik und dient nur der Anzeige des Relativsatzes (vgl. z. B. Brooks 2006).

Position Die bedeutendste Annotation ist die Relativsatzposition. Ich differenziere zwischen *eingebetteten/in situ*, *ausgelagerten/extraponierten* und *ambigen Relativsätzen*. Entscheidend für diese Unterscheidung ist die Stellung von Antezedens und Relativsatz zueinander.

Wenn das Antezedens zusammen mit dem Relativsatz in Adjazenz im Vorfeld oder in einem Mittelfeld steht, das von beiden, lexikalisch gefüllten Satzklammern begrenzt wird, wird der Relativsatz eindeutig als eingebettet bestimmt (2a). Wird der Relativsatz jedoch durch sprachliches Material vom Antezedens getrennt, liegt ein extraponierter Relativsatz vor. Dabei ist es un-

¹³ Relativsätze, die mit subordinierenden Konjunktionen eingeleitet werden und die funktionalen Merkmale eines Relativsatzes besitzen (1a), wurden nicht in das Korpus aufgenommen.

- (1) a. Der Moment, **als ihr Kind geboren wurde**, war ganz besonders.
b. Der Moment, **in dem ihr Kind geboren wurde**, war ganz besonders.

Diese subordinierenden Konjunktionen besitzen eine zusätzliche Bedeutung wie einen Hinweis auf Zeit, Ort oder Art und Weise und können nur mit Nomen kombiniert werden, die ebenfalls eine solche Bedeutung haben. Relativpronomen sind hingegen neutral, bis sie mit Präpositionen kombiniert werden, und können, wie (1b) zeigt, die Subjunktionen in Kombination mit einer Präposition ersetzen.

abhängig davon, ob die beiden durch die rechte Satzklammer wie in (2b) separiert werden oder ob eine oder mehrere andere Konstituenten zwischen dem Kopfnomen und dem Relativsatz stehen. Auch eine Kombination aus Konstituenten und der – meistens rechten – Satzklammer ist denkbar. Wie in Kapitel 2.2.1 beschrieben wurde, trennt die rechte Satzklammer das Mittelfeld auch dann vom Nachfeld, wenn sie nicht lexikalisch realisiert ist. Zudem ist das Nachfeld die definierte Landeposition der meisten Nebensätze. Daher spreche ich auch unabhängig von der lexikalischen Füllung der rechten Satzklammer von Auslagerung, sofern die erwähnte Separierung zwischen Antezedenz und Relativsatz vorliegt.

Ein Sonderfall liegt bei den ambigen Relativsätzen vor. Ich bezeichne einen Relativsatz als *ambig*, wenn er zwar adjazent zu seinem Bezugswort ist, aber gemeinsam mit ihm am rechten Rand des Mittelfelds (2c) oder im Nachfeld steht, weil die rechte Satzklammer in diesen Beispielen lexikalisch leer ist. Diese Fälle werden aus der Analyse ausgeschlossen, weil der Relativsatz anders als bei eingebetteten Relativsätzen nicht mehr von seinem Bezugswort getrennt werden kann. Die Entscheidung, ob er aufgrund von Komplexität oder Informativität an einer anderen Stelle stehen könnte, entfällt für diese Sätze also, was sie für die Analyse ungeeignet macht.

- (2) a. *Die alteration aber / die aus [...] Wasser entstehet / [geschichte]_{LSK} auf solche Art: (Abel 1699: Satz 113)*
- b. *Streng genommen müsste man dazu alle diejenigen Krankheiten [rechnen]_{RSK}, welche eine Folge von Verwundungen [...] sind. (Koch 1878: Satz 4)*
- c. *Also [sind]_{LSK} die Brüche eine [...] Zerschmetterung der harten Knochen so aneinander hangen. (Purmann 1680: Satz 169)*

Komplexe Relativsätze Die zuvor gezeigten Beispiele (2) sind einfache Relativsätze, von denen keine weiteren Nebensätze abhängen. Sehr häufig werden in den Daten jedoch adverbiale Bestimmungen und auch Objekte im Relativsatz sentential gebildet (3).

- (3) so bildete sich *die andere Hauptform der Auslegung*, **welche eher geneigt ist, den geschichtlichen Hergang zuzugeben, nur aber denselben nicht als einen göttlichen, sondern als einen menschlichen fasst**, zunächst bei den Gegnern des Christenthums ... (Strauß 1835: Satz 33)

In diesen Fällen zähle ich den ganzen fettgedruckten Teil in (3) als einen Relativsatz, da sich die weiteren Nebensätze inhaltlich nicht davon ablösen lassen und notwendig sind, um den gesamten Inhalt der Aussage zu verstehen. Dies gilt sowohl bei freien Angaben als auch bei valenzgebundenen Ergänzungen in Satzform.

Wenn die vom Relativsatz abhängenden Nebensätze von diesem getrennt würden, könnten Referenzen innerhalb des Relativsatzes nicht mehr korrekt verarbeitet werden, da sie sich dann theoretisch auch auf den Matrixsatz beziehen könnten. Neben dem Verständnis ist auch die Bewegung des Relativsatzes relevant. Ungeachtet ihrer Länge und ihrer syntaktischen Komplexität können Relativsätze adjazent stehen. Daher muss davon ausgegangen werden, dass Relativsätze mitsamt ihren satzförmigen Ergänzungen und Angaben verschoben werden.

Ähnlich werden koordinierte Relativsätze behandelt. Hängen sie vom gleichen Antezedens ab, werden sie als ein einziger Relativsatz annotiert und gewertet. Dies gilt unabhängig davon, ob das Relativpronomen beziehungsweise die -partikel wieder aufgenommen wird oder zu Beginn des zweiten Relativsatzes nicht mehr realisiert wird (4).

- (4) daß der Mensch umsonst nach Dingen strebe, **fuer welche er keine Sinne hat, und folglich schlechterdings nicht in das Gebiet seiner Erkenntniß gehoeren**. (Gall 1791: Satz 12)

Antezedens Untrennbar mit der Annotation der Relativsätze ist die Annotation des Antezedens verknüpft. Hier wird die Nominal-, Determinierer- oder Präpositionalphrase markiert, von der der Relativsatz abhängt. WebAnno (vgl. Yimam et al. 2013) erlaubt die Unterscheidung zwischen dem Antezedens, womit die gesamte Phrase inklusive aller attributiven Erweiterungen

annotiert wird, wie in (4) durch die Kursivierung angezeigt wird, und dem Kopf. Letzteres bezieht sich auf das Wort in der Phrase, das als tatsächliches Kopfnomen zu verstehen ist. In der Regel handelt es sich dabei entweder um ein Nomen oder ein Pronomen, wie an dem unterstrichenen Wort in (4) zu erkennen ist.

Restriktivität Des Weiteren werden Relativsätze dahingehend unterschieden, ob sie restriktiv oder appositiv sind. Die Klassifizierung orientiert sich dabei an der Darstellung aus Kapitel 2.4.3. Ein Relativsatz gilt dann als restriktiv, wenn die eindeutige Identifizierbarkeit des Antezedens ermöglicht. Einfacher ist die Bestimmbarkeit, wenn die klassischen Indikatoren für Restriktivität vorliegen. Dazu zählen Pronomen wie *der-/die-/dasjenige* oder *etwas*. Ebenso gelten die meisten definiten Nomen, die diskursneu sind, als Indikatoren für Restriktivität. Sonst ist der Relativsatz appositiv. Ausschlaggebend für die Bezeichnung ist zum einen der Kontext, der verdeutlicht, ob das Antezedens beispielsweise bereits aufgetreten und daher identifizierbar ist. Unmittelbare kontextuelle Vorerwähntheit sowie Eigennamen sind zwei der Indikatoren für Appositivität.

In einigen Fällen ist es nicht leicht festzustellen, ob ein Relativsatz restriktiv ist oder nicht, wenn weder der Kontext noch das Weltwissen den Relativsatztyp eindeutig bestimmen. Dies ist vor allem bei den vorliegenden historischen Daten von Bedeutung, da moderne Leser möglicherweise nicht in der Lage sind, das Wissen des Autors und seines Publikums zu rekonstruieren. Das fiktive Beispiel (5) soll dies verdeutlichen:

- (5) Er behandelt den Patienten *mit einem Bezoar*, **der einer Katze entnommen wurde**.

Ein moderner Leser kann hinsichtlich der Herkunft eines Bezoars, eine Zutat für ein Heilmittel, die tatsächlich in zwei Texten im Korpus erwähnt wird (vgl. Purmann 1680; Abel 1699), keine genaue Aussage treffen und ist daher nicht in der Lage, den Relativsatztyp zu bestimmen. In derartigen Fällen wurde der Typ nicht bestimmt und mit dem Platzhalter NA versehen.

Länge Die Annotation der Länge ist vor allem durch die bestehende Forschungslage begründet (Kapitel 3.2). Ich messe als Länge die Anzahl der Tokens pro Einheit, die als Relativsatz markiert wurde. Die Bestimmung erfolgt automatisch über die Bearbeitung in R (R Core Team 2018; 2022). Das birgt allerdings in zweifacher Hinsicht die Gefahr unsauberer Ergebnisse. Zu dem erfolgte die Tokenerkennung bereits automatisch im DTA, weshalb es wegen einiger orthografischer Besonderheiten des frühen Neuhochdeutschen zu falschen Worttrennungen kommen kann, die dann zu einer größeren Anzahl an Tokens führen würde, als tatsächlich im Satz stehen. Das kann beispielsweise auch bei Trennungs- und Bindestrichen im Originaltext passieren.

Problematisch ist zudem auch, wie mit Zahlen und Referenzen im Text umgegangen wird. Ihr Einfluss auf Verarbeitung wurde noch nicht untersucht. Da sie jedoch vom Leser auf jeden Fall gesehen werden, werden sie in dieser Untersuchung nicht ausgeschlossen. Insgesamt ist davon auszugehen, dass die Fehlerrate durch die automatische Annotation vernachlässigt werden kann. Zu erwähnen ist jedoch, dass nur die Länge des Relativsatzes, nicht auch seines Matrixsatzes gemessen werden konnte, da die Satzgrenzbestimmung des DTA in zu vielen Fällen zu fehlerhaft ist (vgl. Ortmann 2020; 2023), um die Daten nicht signifikant zu verzerren.

Distanz Die Bestimmung der Distanz zwischen dem Antezedens und dem Relativsatz begründet sich dadurch, dass sie sowohl Hawkins (1992), Gibson (1998) als auch Uszkoreit et al. (1998) als einen der relevantesten Faktoren für die Extraposition von Relativsätzen definieren. In dieser Arbeit erfolgt sie zunächst automatisch¹⁴ und wird über die intervenierenden Tokens zwischen Antezedenz und Relativsatzeinleiter gemessen. Um eine Vergleichbarkeit gewährleisten zu können, wird für adjazente Relativsätze die Distanz zwischen dem Ende des Relativsatzes und dem Ende der rechten Satzklammer gemessen, weil das die Strecke wäre, die diese Sätze überbrücken müssten, um ausgelagert zu werden.

In vielen Fällen zeigte die automatische Distanzbestimmung große und unrealistische Schwankungen, weshalb eine manuelle Überprüfung vorge-

¹⁴ Ich bedanke mich an der Stelle bei Katrin Ortmann für die Bereitstellung dieser Werte.

nommen wurde. Fehlerhafte Zählungen wurden manuell korrigiert. Im Fall der *in situ* Relativsätze wurden sämtliche Werte der Distanzen manuell gezählt, da die Satzgrenzenbestimmung, wie bereits erwähnt, zu fehlerhaft ist, um sich auf ein automatisches Maß verlassen zu können.

Belege Auf diese Art konnten insgesamt 5332, davon 2706 extraponierte, Relativsätze gefunden werden, deren Verteilung auf die einzelnen Zeitstufen und Textgenres in Tabelle 4.4 dargestellt ist. Die ambigen Relativsätze werden hier nicht gezählt. Zur Darstellung des relativen Anteils der Relativsätze dient die dritte Spalte der Tabelle, die die Anzahl von Sätzen mit extraponierten Relativsätzen proportional zu allen Sätzen im Korpus sieht. Es zeigt sich, dass Relativsatsauslagerung insgesamt nur in zehn Prozent aller Sätze vorkommt. Eine nähere Betrachtung dieser Ergebnisse erfolgt in Kapitel 6.3.

Zeitabschnitt	Gesamtzahl RelS (ex- trap.)	Anteil Extrapo- sition an Sätzen in Prozent, gerundet	Medizinisch (extrap.)	Theologisch (extrap.)
1650 - 1700	769 (408)	9,9	369 (202)	400 (206)
1700 - 1750	1435 (665)	11,3	659 (358)	776 (307)
1750 - 1800	1038 (425)	17,5	354 (125)	684 (300)
1800 - 1850	1175 (677)	16,1	988 (575)	187 (102)
1850 - 1900	915 (531)	8,2	501 (287)	414 (244)
gesamt	5332 (2706)	10	2871 (1547)	2461 (1159)

Tabelle 4.4.: Anzahl der Relativsätze nach Zeitabschnitt und Genre.

4.3.3. Präpositionalphrasen

Bei der Annotation dieses Phänomens wurden zunächst alle Präpositionalphrasen gesucht, die im Nachfeld zu finden sind.¹⁵ Anders als bei den Relativsätzen muss zwingend eine rechte Satzklammer vorhanden sein, um eine Präpositionalphrase als ausgelagert bezeichnen zu können.

Die Präpositionalphrasen wurden mit allen von ihnen abhängigen Attributen und anderen Erweiterungen annotiert. Dazu zählen Adjektivattribute und Relativsätze, die in diesen Fällen als ambig gekennzeichnet und aus der weiteren, eigenen Analyse ausgeschlossen werden (6a). Wenn zwei Nominalphrasen von der gleichen Präposition regiert werden und koordiniert sind, werden sie als eine Präpositionalphrase gewertet (6b). Hängen sie von unterschiedlichen Präpositionen ab, werden sie als zwei (oder mehr) Phrasen annotiert. Auch weitere satzförmige Ergänzungen, die vom Komplement der Präpositionalphrase abhängen, werden zu der Präpositionalphrase gezählt.

- (6) a. Da nun diese letztern nur abhängig waren **von dem senkrechten Abstand, in dem sie unter dem Wasserspiegel lagen**, so folgt daraus, ... (Ludwig 1856: Satz 182)
- b. Der [...] Cyclus des weiblichen Lebens, welchen wir, entsprechend den drei großen Stadien des allgemeinen Lebens, eingetheilt haben **in Schwangerschaft (Entwicklung), Geburt (Wendepunkt), und Wochenperiode (Rückbildung)**, beginnt mit dem geheimnißvollen Akte der Empfängniß (Conceptio) (Carus 1820: Satz 5)

Zudem wurde die Länge über die Anzahl der Tokens in den als Präpositionalphrase markierten Wortspannen bestimmt (vgl. Kapitel 6.2). Die Berechnung der Länge erfolgte wie bei den Relativsätzen mit R (vgl. R Core Team 2022). Zusätzlich wurde der jeweilige Valenzstatus für die gefundenen extrapolierten Präpositionalphrasen manuell bestimmt. Unterschieden wurde zwischen obligatorischen und fakultativen Ergänzungen sowie freien Angaben (vgl. z. B. Welke 2011: 48).

¹⁵ Inhalte dieses Kapitels flossen in die Veröffentlichung Voigtmann und Speyer (2022) und Voigtmann und Speyer (2023) ein.

Belege Auf diese Art konnten insgesamt 1043 extrapониerte Präpositionalphrasen gefunden werden, deren Verteilung auf die einzelnen Zeitstufen und Textgenres in Tabelle 4.5 dargestellt ist. In 98 % der Fälle wurden eigenständige Phrasen ins Nachfeld gestellt (Kapitel 6.2). Zur Darstellung des relativen Anteils der Präpositionalphrasenextraposition dient die dritte Spalte der Tabelle, die die Anzahl von Sätzen mit extrapониerten Präpositionalphrasen proportional zu allen Sätzen im Korpus sieht. Es zeigt sich, dass Auslagerung insgesamt nur in weniger als fünf Prozent aller Sätze vorkommt. Eine nähere Beschreibung dieser Ergebnisse erfolgt in Kapitel 6.2.

Zeitabschnitt	Gesamtzahl extrapониerte PP	Anteil Extraposition an Sätzen in Prozent, gerundet	Medizinisch	Theologisch
1650 - 1700	239	3,08	156	83
1700 - 1750	103	0,812	19	84
1750 - 1800	132	2,22	10	122
1800 - 1850	70	0,96	35	35
1850 - 1900	499	2,63	421	78
gesamt	1043	4,05	641	402

Tabelle 4.5.: Anzahl der Präpositionalphrasen nach Zeitabschnitt und Genre.

4.3.4. Nominalphrasen

Das Vorgehen zum Erkennen der Nominalphrasen ähnelt dem der Präpositionalphrasen. Eine Nominalphrase ist dann extrapониert, wenn sie hinter der lexikalisch realisierten rechten Satzklammer steht. Zu ihr zählen auch die von ihr abhängigen Attribute, also Adjektive, Präpositionalphrasen, weitere Nominalphrasen und Relativsätze (siehe Tabelle 2.7 in Kapitel 2.5). In (7a) findet sich ein Beispiel für eine durch einen komplexen Relativsatz erweiterte Nominalphrase. Diese Relativsätze werden als ambig gekennzeichnet und von der weiteren Analyse der Relativsätze ausgeschlossen. Auch Reihungen und Aufzählungen werden als eine Nominalphrase gekennzeichnet, wenn die Rei-

hung vollständig ins Mittelfeld geschoben werden könnte (7b).

- (7) a. Daraus würde sich ergeben, dass der Alkohol hier die Neigung hat, den gewöhnlichen Parallelismus zwischen Arbeitsgrösse und Schwankung zu stören, indem er gleichzeitig mit einer Herabminderung der rechnerischen Leistung dieselbe auffallend häufig auch ungleichmässiger gestaltet, **eine Störung, die sich erst dann wieder ausgleicht, wenn auch die Beeinträchtigung der Arbeitsfähigkeit durch den Alkohol nachlässt.** (Kraepelin 1892: Satz 15)
- b. Quantitativ sind v. Bibra bestimmt worden **die in Aether löslichen und unlöslichen Bestandtheile, das Wasser und die Aschen am nerv. opticus, brachialis, cruralis, ein oberer und unterer Abschnitt des ischiadicus bei Menschen von 3 bis 93 Jahren, männlichen und weiblichen Geschlechts.** (Ludwig 1856: Satz 1281)

Für alle annotierten, extraponierten Nominalphrasen wurde nachträglich ihre Satzgliedfunktion im Sinne der klassischen Syntax bestimmt. Die Auszählung ist Tabelle 6.2 zu entnehmen.¹⁶ Auch ihre Länge wird annotiert (vgl. Kapitel 6.1, Tabelle 6.3).

Belege Nach diesem Verfahren wurden insgesamt 306 extraponierte Nominalphrasen gefunden. Ihre Verteilung auf die einzelnen Zeitabschnitte und Genre ist in Tabelle 4.6 dargestellt. In der dritten Spalte der Tabelle wird der relative Anteil an Sätzen mit Extraposition an allen Sätzen des Zeitabschnitts dargestellt. Es ist zu erkennen, dass die Extraposition von Nominalphrasen nur in insgesamt weniger als 2% aller Sätze vorkommt und daher ein eher unproduktives Phänomen darstellt, das zudem Schwankungen unterliegt, auf die in Kapitel 6.1 näher eingegangen wird.

¹⁶ Eine Nominalphrase im Nachfeld wurde auch dann als Subjekt oder Objekt annotiert, wenn es sich um eine koordinierte Nominalphrase handelt, deren erstes Glied noch im Mittelfeld steht, was allerdings eher niederfrequent vorgekommen ist. Attribute wurden nach Tabelle 2.7 bestimmt. Vokative oder Phrasen, die mit *sondern* eingeleitet werden, wurden in die Kategorie ‚Sonstiges‘ eingeordnet.

Zeitabschnitt	extra- ponierte NP	Anteil an Sätzen in Prozent, gerundet	Medizinisch	Theologisch
1650 - 1700	98	1,26	60	40
1700 - 1750	32	0,25	8	24
1750 - 1800	71	1,19	8	63
1800 - 1850	20	0,25	13	7
1850 - 1900	85	0,45	68	17
gesamt	306	1,19	157	151

Tabelle 4.6.: Anzahl der extrapponierten Nominalphrasen nach Zeitabschnitt und Genre.

4.3.5. Vergleichspaare für die Phrasen: Minimalpaare

In Kapitel 4.3.2 wurde beschrieben, dass sowohl alle adjazenten als auch alle extrapponierten Relativsätze annotiert wurden und die ambigen Relativsätze aus der Analyse ausgeschlossen wurden. Durch diese Praxis ist es im Fall der Relativsätze eindeutig möglich, einen Vergleich zwischen den im Nachfeld stehenden und den im Vor- oder Mittelfeld befindlichen Relativsätzen zu ziehen. Bei den Nominalphrasen und den Präpositionalphrasen ist das hingegen aus operationalen Gründen nicht so einfach machbar.

Wie ich in Kapitel 4.1 beschrieben habe, war die Qualität der Vorverarbeitungsschritte des DTAs zum Startzeitpunkt dieser Dissertation mit starken Mängeln behaftet, die eine automatische Suche nach Konstituenten erschwert. Um wie bei den Relativsätzen einen Vergleich zwischen allen eingebetteten und allen ausgelagerten Phrasen im Korpus durchführen zu können, hätte ich folglich *alle* eingebetteten Nominal- und Präpositionalphrasen im Korpus händisch annotieren müssen. Die Annotation aller *in situ*-Phrasen ist vor allem wegen des Aufwands der händischen Annotation, die den (zeitlichen) Umfang dieses Projekts überschritten hätte, unpraktikabel.

Ich habe mich folglich dazu entschieden, ein Sample aus Nominalphrasen und Präpositionalphrasen zu erstellen, die allerdings nicht vollständig dem Zufallsprinzip folgend gewählt wurden. Im Idealfall sollte so jeder ausgelagerten Phrase ein eingebettetes Gegenstück zugeordnet werden, das eben-

falls hätte ausgelagert werden können. Ich bezeichne dieses so entstehende Paar als Minimalpaar, weil sich diese Phrasen von den extraponierten vor allem in dem einen Merkmal der Position unterscheiden sollen, und entlehne den Begriff aus der Phonetik (vgl. z. B. Glück und Rödel 2016: 432).

Bei der Wahl der Phrasen kamen folgende Prinzipien zum Einsatz: Um als eingebettet annotiert zu werden, muss die Präpositional- beziehungsweise Nominalphrase entweder im Mittel- oder im Vorfeld stehen. Um im Mittelfeld eingebettet zu sein, muss entweder die rechte Satzklammer lexikalisch gefüllt sein (8a) oder es muss mindestens eine weitere Konstituente zwischen dem Ende des (Teil-)Satzes und der betreffenden Präpositionalphrase stehen (8b). Ich habe mich zur besseren Operationalisierung entschieden, auch das Vorfeld als eingebettete Position zu betrachten (8c). Das Vorfeld ist zwar eine Position mit Einschränkungen und besonderen Besetzungsbedingungen (vgl. u. a. Speyer 2008b; 2009; 2010). Trotzdem erscheint es vertretbar, dieses Feld einzuschließen, weil die zur Verarbeitung benötigte Kapazität vergleichbar mit den am linken Rand des Mittelfelds liegenden Phrasen sein sollte.

- (8) a. Diese Erfahrung ist zu typisch, [um]_{LKS} uns nicht **an die ganz analoge Erscheinung beim Zahlenlernen, namentlich hinsichtlich der Wiederholungsschnelligkeit** [zu erinnern]_{RKS}. (Kraepelin 1892: 67)
- b. Er versteht **unter Septicämie** eine Krankheit hervorgerufen durch Blutveränderungen, die eine Folge der Aufnahme von Fäulnisproducten sind. (Koch 1878: 11)
- c. **Den Nutzen der Diaet** erklärt auch gar fein Ovvenus sagend: (Abel 1699: 254)

Das Vorfeld ebenfalls als eingebettet zu betrachten, wird auch deshalb notwendig, weil die eingebetteten Phrasen eine *vergleichbare* Länge und Komplexität wie ihre ausgelagerten Gegenstücke aufweisen sollen. Wenn zum Beispiel nur sehr kurze eingebettete Phrasen berücksichtigt würden, ist die Wahrscheinlichkeit natürlich größer, dass ihr Surprisalwert kleiner ist als der von komplexeren, ausgelagerten Präpositionalphrasen, sodass die Daten ver-

zerren sein können (Kapitel 5.2 und Voigtmann und Speyer (2023: 16)). In der in Kapitel 3.1 dargestellten Literatur wurde bereits ausführlich gezeigt, dass Länge einen starken Einfluss auf die Position einer Phrase ausübt. Ein Ziel dieser Arbeit ist es aber auch zu untersuchen, ob Informationsdichte einen noch bedeutenderen Faktor darstellt. Da dieser Wert aber eben verzerrt sein könnte, wenn nur kurze Phrasen annotiert werden, wird die Länge, aber vor allem die Komplexität, zu berücksichtigen versucht.¹⁷

Oben wurde erwähnt, dass die bisherige Forschung Länge und Komplexität in der Regel in Anlehnung an Wasow (1997) synonym verwendet. In dieser Arbeit wird davon abgewichen, weil die Komplexität als *bedeutender* angesehen wird als die Länge in Wörtern. Unter Komplexität wird in dieser Arbeit verstanden, dass die Erweiterungen der eingebetteten Phrase denen der extraponierten entsprechen. In der Praxis ist damit vor allem gemeint, dass das Dependens von einem satzförmigen Attribut erweitert ist. Dabei handelt es sich meistens um einen Relativsatz, der bei extraponierten Phrasen als ambig annotiert wird, bei eingebetteten Phrasen aber auch als *in situ*-Relativsätze in die Analyse eingeflossen sein kann. Bei Nominalphrasenreihungen im Nachfeld wurde außerdem versucht, eine eingebettete Nominalphrase zu finden, die fast genauso viele Wörter umfasst.

Das und die Satzattribute machen es aber sowieso außerordentlich unwahrscheinlich, dass eine eingebettete Phrase mit satzförmiger Erweiterung und/oder der exakt gleichen Wortzahl wie ihr ausgelagertes Gegenstück ge-

¹⁷ (9) und (10) bieten Beispiele für Minimalpaare gleicher Länge.

- (9) a. Laß alles mit einander zergehen **über gelindem Feuer** / exprimir es durch ein Tuch... (Purmann 1680: Satz 78)
 b. ... lege ein Oppodeltoch **über die Wunden** und ein gutes Defensiv, gemacht von Synnau und Bonenmehl darauff / damit keine Schmertzen und Entzündung darzu schlagen könne. (Purmann 1680: Satz 114)
- (10) a. Und weil an dem Magen das meiste gelegen / so sol man denselben mit allen Fleiß in acht nehmen / daß er nicht zornig werde / und allen Kranckheten Thür und Angel aufmache / weil er ohne diß von den meisten Medicis genennet wird **ein Schmid aller Kranckheiten**. (Abel 1699: Satz 225)
 b. denn vor 7. Jahren ist einem Manne ein Beil aus der Hand entfallen / welches biß auf ietzige Stund **den Grund des Meeres** noch nicht berühret hat (Abel 1699: Satz 117)

funden werden kann, wie (11) zeigt. Somit kommt es dazu, dass sich die Länge der eingebetteten und ausgelagerten Phrasen gemessen in Tokens stark unterscheidet, was auch in Kapitel 6 zu sehen ist.¹⁸

- (11) a. ... daß sie durch aus keine Verwandtschaft [haben]_{RSK} **mit dem alten Buch, das man die Bibel heisset.** (Bengel 1751: 62)
- b. ... und die Reflexionen **mit ihren Beylagen, unter denen sich auch die Gewissens-Rüge befindet,** [zielen]_{LSK} auf Verantwortung... (Bengel 1751: 60)

Bei den Präpositionalphrasen soll außerdem folgendes Kriterium beachtet werden: Die eingebettete Phrase muss durch dieselbe Präposition eingeleitet werden wie die extraponierte Präpositionalphrase, um mögliche semantische Effekte der Präposition auszuschließen. Nach Höllein (2019) führt dieses Kriterium dazu, dass sich die Funktionen der Präpositionalphrasen ähneln.¹⁹

Die Suche nach den Minimalpaaren geschah in der Regel im unmittelbaren Kontext der ausgelagerten Phrase, was auch (11) zeigt: Die ausgelagerte Phrase stammt aus Satz 62 in Bengel (1751), die eingebettete Phrase aus Satz

¹⁸ Eine Längenausählung zufällig ausgewählter Phrasen im Korpus (Stichprobe von je 25 eingebetteten, nicht im Zuge der Minimalpaaranalyse annotierten Nominalphrasen und 50 eingebetteten, nicht im Zuge der Minimalpaaranalyse annotierten Präpositionalphrasen pro Text) hat ergeben, dass diese zufällig ausgewählten Phrasen insgesamt signifikant kürzer sind als die von mir im Zuge der Minimalpaaranotation gefundenen. Dieses Zufallssample der Nominalphrasen besitzt eine durchschnittliche Länge von 4,26 Tokens, die Minimalpaar nominalphrasen besitzen eine Länge von 13,08 Tokens. Ein Wilcoxon-Test, durchgeführt mit William Revelle (2024), ergibt einen signifikanten Unterschied zwischen den Phrasen ($p < 0,001$): Die Länge des randomisierten Sample ist signifikant kürzer als die der Minimalpaarphrasen. Bei den Präpositionalphrasen wird das gleiche Ergebnis gefunden. Das Zufallssample der Präpositionalphrasen hat eine durchschnittliche Länge von 4,43 Tokens, während die Minimalpaarpräpositionalphrasen eine Länge von 19,77 Tokens aufweisen. Auch hier hat der Wilcoxon-Test einen signifikanten Unterschied belegen können ($p < 0,001$).

Somit erlaubt diese Auszählung auch Rückschlüsse auf die Gesamtheit der Phrasen und daher auch die Aussage, dass nicht im Nachfeld stehende Phrasen insgesamt und grundsätzlich kürzer sind, was sich schließlich auch, wie in den Kapiteln 6.1, 6.2 und 6.3 gezeigt wird, auch in den Minimalpaaren widerspiegelt. Mit den Phrasen des Zufallssamples wurden *keine* Analysen in dieser Arbeit durchgeführt. Die Berechnungen in den Kapiteln 6.1 und 6.2 wurden ausschließlich auf den Minimalpaaren getätigt.

¹⁹ Höllein (2019) beschreibt für verschiedene Präpositionen semantische Nischen, die sich an den thematischen Rollen orientieren. Gleiche Präpositionen stünden für gleiche thematische Funktionen. So beschreibt er zum Beispiel „hoffen mit PO_{auf+Akk} als Prototyp der Nische PROSPECTUM“ (Höllein 2019: 318) und widerspricht insgesamt der These der desemantisierten Präpositionen (vgl. u. a. Nübling 2016).

60. Aber es wurde auch der gesamte Text in Betracht gezogen, sodass die ausgelagerte Phrase am Anfang des Textes und ihr Gegenstück in der Mitte oder am Anfang stehen kann oder umgekehrt.

Aufgrund der oben genannten Einschränkungen war es in vielen Fällen allerdings nicht möglich, eine eingebettete Phrase zu finden, die allen Kriterien entspricht. Daher wurden die oben beschriebenen Kriterien für die Minimalpaarsuche bei Auswahlprozess der passenden Phrase gewichtet. 'Eins' stellt dabei das bedeutendste Kriterium dar, 'fünf' das unbedeutendste.

1. Die Phrase befindet sich eindeutig *nicht* im Nachfeld.
2. Das Minimalpaar besitzt die gleiche Komplexität, das heißt, es wird durch ein ähnliches Attribut erweitert.
3. Handelt es sich um eine Präpositionalphrase, stimmen die Präpositionen überein. Dieses Kriterium entfällt bei Nominalphrasen.
4. Die eingebettete Phrase befindet sich im Mittelfeld. Wenn das nicht möglich ist, soll die eingebettete Phrase im Vorfeld stehen.
5. Die eingebettete Phase besitzt die gleiche Länge, gemessen in Tokens, wie ihr Gegenstück.

Wenn also eine Präpositionalphrase im Nachfeld mit beispielsweise *durch* gefunden wurde, sich im Text jedoch keine den Einbettungskriterien entsprechende weitere Präpositionalphrase mit dieser Präposition finden ließ, wurde eher eine mit einer anderen Präposition gesucht, die dafür der ausgelagerten in der Komplexität ähnlich war. Trotzdem war es nicht immer möglich, für jede ausgelagerte Phrase ein passendes Gegenstück zu finden, was in den Kapiteln 6.2 und 6.1 verdeutlicht wird. Für die Analyse der Daten mithilfe von logistischen Regressionen ist dieser Unterschied allerdings unerheblich (vgl. Winter 2020; Gries 2021).

Während bei den ausgelagerten Phrasen zusätzlich der Valenzstatus für Präpositionalphrasen und die Satzgliedfunktion der Nominalphrasen ermittelt wurde, entfällt dieser Schritt bei den eingebetteten Phrasen, da davon auszugehen ist, dass eingebettete Phrasen sowieso jede mögliche Satzgliedfunktion und jeden möglichen Valenzstatus aufweisen können, weil sie auf jeden Fall häufiger im Vor- und Mittelfeld auftreten können.

4.4. Bereinigung der Daten

Die vorherigen Teilkapitel haben beschrieben, welche Daten annotiert wurden. Hier wird nun kurz darauf eingegangen, welche Daten nicht in das Korpus einbezogen wurden, und Gründe für ihren Ausschluss aufgezeigt. Dabei versteht es sich von selbst, dass generell nur Sätze in die Analyse einfließen, die eine als NP, PP oder RELC annotierte Wortspanne aufweisen,²⁰ unabhängig davon, ob diese eingebettet oder ausgelagert ist. Alle anderen Sätze wurden aussortiert.

Satzzeichen Aus den verbleibenden Sätzen wurden alle Satzzeichen entfernt. Das schließt folgende Elemente ein: Kommas, Punkte, Fragezeichen, Ausrufezeichen, Semikolons, Doppelpunkte, Binde- und Gedankenstriche sowie eckige und runde Klammern. Der Filterungsprozess nach diesen Zeichen richtet sich nach ihren POS-Tags nach dem STTS-Schema (vgl. Schiller, Teufel, Stöckert und Thielen 1999). Die Daten liegen nach dem Download aus WebAnno im .tsv-Format vor. Das heißt, jedes Token steht in seiner eigenen Zeile und erhält eine Lemma- und POS-Tag-Zuordnung in einer anderen Spalte. Da auch Satzzeichen als Token gelten, kann mithilfe von R (R Core Team 2018) nach den Zeilen gefiltert werden, die „\$,“, „\$.“ oder „\$(“ enthalten. Diese Zeilen wurden verworfen, die restlichen behalten.

Natürlich beeinflusst dieser Schritt das Endergebnis. Allerdings ist er nötig, um die unterschiedlichen Texte aus den verschiedenen Zeitabschnitten vergleichbar zu machen. Eine einheitliche, festgelegte Zeichensetzung für das Deutsche gibt es erst seit dem 20. Jahrhundert, also erst nach dem hier betrachteten Zeitraum (vgl. z. B. Polenz et al. 2013). Grundlegende Regeln, wie einen Satz mit einem Punkt, Komma oder Ausrufezeichen zu beenden, lassen sich bereits in den frühen Texten erkennen, doch ist gerade für die Kommasetzung über die Jahrhunderte kein einheitlicher Umgang zu erkennen.

Zwar zeigt sich in (12a), dass Kommas durchaus zur Abgrenzung von Nebensätzen genutzt wurden, aber das Korpus enthält auch zahlreiche Bele-

²⁰ Die Abkürzungen beziehen sich hier auf die in WebAnno (vgl. Yimam et al. 2013) verwendeten Label der Annotationen.

ge, in denen das nicht der Fall ist. Häufig findet sich in den Texten des 17. und 18. Jahrhunderts eine Abgrenzung von Satzgliedern durch Virgeln (12b). Beides erschwert die Generalisierung.

- (12) a. Ob nicht in der Natur selbs eine Ursache liege, warum wir dasjenige, was GOtt thut und gebeut, was Christus gethan und geboten hat, für gut erkennen und preisen sollen. (Bengel 1751: Satz 388)
- b. DAs Reiten bekömmst auch nicht einem iedweden / denn entweder der eine bricht den Hals / wie dem Philippo, Königs in Franckreich Ludovici Sohn wiederfahren / oder leidet sonsten Schaden an seinem Leibe / als da sind Brüche / Zerquetschungen der Beine / und andere Ubel / welche von vehementer Bewegung oder Concussion im Reiten entstehen / die einem iedweden behutsam machen solten. (Abel 1699: Satz 91)

Die verwendeten Sprachmodelle (Kapitel 5) sind im Vergleich zu menschlichen Lesenden unflexibel. Die Berechnung der Surprisalwerte und der Länge erfolgen auf der Basis von Tokens. Beides ist also stark davon beeinflusst, ob in einem Satz ein Satzzeichen steht oder nicht, da diese auch als Tokens gelten und daher in das Modell einfließen würden. Dadurch würde in einigen Stellen nicht die kontextuelle Wahrscheinlichkeit, die auf den zu verarbeitenden Wörtern beruht, berechnet werden, sondern die Wahrscheinlichkeit, dass ein Wort nach einem Satzzeichen verwendet wird. Dieser Wert gibt also weder einen für meine Untersuchung relevanten Wert an, noch wäre dieser über die Jahrhunderte hinweg vergleichbar.

Die weiteren Schritte zur Bereinigung der Daten haben im Gegensatz zum Löschen der Satzzeichen keinen Einfluss auf die Ergebnisse, da sie vor allem technischer Natur sind und dazu dienen, das aus WebAnno (vgl. Yimam et al. 2013) heruntergeladene Format für die weitere Betrachtung nutzbar zu machen.²¹

²¹ Beispielsweise wurden Spalten gelöscht, die POS-Tags enthielten.

Sätze mit mehreren ausgelagerten Phänomenen Ein größeres Problem als der Umgang mit Satzzeichen oder das Löschen von ungenutzten Informationen bildet der Umgang mit Sätzen, die mehr als ein untersuchtes Phänomen enthalten. In vielen der Sätze wurde sowohl eine eingebettete oder ausgelagerte Nominal- oder Präpositionalphrase zusammen mit einem Relativsatz gefunden und annotiert. Wenn diese Elemente, wie gerade beschrieben, unterschiedlichen Kategorien angehören, kann man sie problemlos voneinander unterscheiden und analysieren. Ähnliches gilt für Elemente der gleichen Kategorie, für die unterschiedliche Positionen vermerkt wurden. (13) stellt einen solchen Satz dar. Er enthält zwei Relativsätze, die durch Fettdruck markiert sind. Der erste ist eingebettet, der zweite ausgelagert. Beide Sätze können durch die unterschiedliche Annotation separat analysiert werden.

- (13) Ein ernstliche Christliche Zucht / **die Gott in seinem Wort so ernstlich gebeut / und von seinen Christen haben will** / muß da (in dem Lutherthumb) ein new Pabstthumb / ein new Müncherey seyn / wir haben / sprechen sie / gelehr. net / daß wir allein durch den Glauben an Christum seelig werden / **der mit seinem Todt all unser Sünd bezahlt hat** . (Niviandts 1708: Satz 237)

Das funktioniert mit dem von mir geschriebenen R-Skript (R Core Team 2018; 2022) auch, wenn zwei Relativsätze in unterschiedlichen Einbettungstiefen vorkommen (14). Beide Relativsätze in (14) sind extraponiert, allerdings ist der kursiv gedruckte Relativsatz ein Teil des fett gedruckten Relativsatzes. Er liegt also auf einer anderen Ebene. Dies wurde in den Annotationen mit Zahlen vermerkt und ließ sich so voneinander trennen.

- (14) Auch den angeblichen unmittelbaren Beobachtungen von Bakterien im normalen Blute stehen die Angaben von zuverlässigen Mikroskopikern, wie Rindfleisch und Riess gegenüber **die bestimmt erklären, dass das normale Blut frei von Bakterien ist, dagegen, wie Riess nachgewiesen hat, mehr oder weniger reichlich kleine rundliche Körperchen enthält, die höchst wahrscheinlich Zerfallsproducte der**

weissen Blutkörperchen sind und wegen ihrer Aehnlichkeit mit Mikrokokken zu Verwechslungen mit diesen Veranlassung gegeben haben. (Koch 1878: Satz 56)

Die klare Trennung und einfache Verarbeitung ist allerdings nicht der Fall, wenn (mindestens) zwei Relativsätze mit gleicher Position auf der gleichen Einbettungsebene in einem Satz auftreten. Am häufigsten kommt das bei *in situ*-Relativsätze vor, aber (15) zeigt auch, dass extraponierte Relativsätze von diesem Phänomen betroffen sein können. Die Relativsätze hängen zwar von verschiedenen Kopfnomen ab, werden vom Skript aber so betrachtet, als befänden sie sich auf der gleichen Ebene im Satz, weil der zweite Relativsatz nicht im ersten eingebettet ist (oder umgekehrt).

- (15) Wenn man [die Worte des N. Testamentes] zusammenstellen und vorlesen sollte, **welche sich daraus beziehen**; so würden sich alle, die Ohren haben zu hören, am Schlusse durch die Erkenntnis der Sünde tief gedemüthigt fühlen, daß sie [eine solche Sache] gering, jeden Falls zu gering geachtet haben, **von welcher Gottes Wort so Vieles und Großes spricht.** (Loehe 1847, Satz 3)

Durch das verwendete Skript werden solche Relativsätze als ein einziger Relativsatz gewertet und zusammengezogen. Daraus resultieren verzerrte Werte für die Berechnung der Länge und der Surprisalwerte. Für die aktuelle Untersuchung wurde daher beschlossen, Sätze wie (15) aus der Analyse auszuschließen. Dazu wurde zunächst allen Sätzen ein eindeutiger Name zugeteilt, der sich aus dem Autor, dem Veröffentlichungsjahr, dem annotierten Element, dessen Position und der Satznummer zusammensetzt.²² Anschließend wurden die problematischen Sätze (n=392, 6,84%) manuell gesucht und mithilfe von R (vgl. R Core Team 2022) über die Namensgebung aus der Gesamtzahl an Sätzen ausgeschlossen. Bei der Suche nach passenden Minimalpaaren für diese Phrasentypen wurde zudem darauf geachtet, dass nicht zwei eingebettete Phrasen des gleichen Typs in einem Satz annotiert wurden.

²² Für Beispiel (15) würde dieser lauten: Loehe_1847_RELC_extrap_3.

5. Methodik

Das vorliegende Kapitel befasst sich mit der in dieser Arbeit genutzten Methodik zur Analyse und Auswertung der Daten, deren Annotation im vorherigen Kapitel beschrieben wurde. Zunächst werden das Sprachmodell (Kapitel 5.1) und die verwendeten Berechnungsmethoden für die Surprisalwerte vorgestellt (Kapitel 5.2). Anschließend (Kapitel 5.3) erfolgt die Beschreibung des Orality Scores nach Ortmann und Dipper (2024) und abschließend die ausführliche Beschreibung der zur Hypothesentestung verwendeten statistischen Analyseverfahren (Kapitel 5.4).

5.1. Sprachmodelle

Sprachmodelle, im Folgenden *Language Models* genannt, werden dazu genutzt, um mithilfe von statistischen und probabilistischen Methoden Vorhersagen über das Auftreten eines Wort oder ganzer Wortsequenzen zu treffen. Sie bedienen sich der Annahme, dass mit einem Hidden-Markoff-Modell durch vorher Gegebenes Aktuelles bestimmt werden kann (vgl. Manning und Schütze 2001: 192).

Bei dem vorher Gegebenen handelt es sich entweder um den durch das Markoff-Modell eingegrenzten unmittelbaren Kontext oder um den von *Trainingsdaten* gebildeten Kontext. Trainingsdaten werden verwendet, um mithilfe eines Hidden-Markoff-Modells Schätzwerte über die Auftretenswahrscheinlichkeit des Wortes in diesem Kontext zu gewinnen. Die Trainingsdaten werden dazu genutzt, allgemeine Regeln über Sprache zu lernen. Diese sollen ausreichen, um präzise Vorhersagen über neue, unbekannte Daten treffen zu können. Die Markoff-Annahme besagt schließlich, dass die Wahrscheinlich-

keit einer zukünftigen Einheit vorhergesagt werden kann, ohne zu weit in die Vergangenheit zu schauen (vgl. Mürmann 2014: 183ff.). Für Sprachen bedeutet dies, dass nicht jede jemals produzierte sprachliche Äußerung in die Berechnung einfließen muss, sondern dass ein Teil der sprachlichen Äußerungen ausreicht, um akzeptable Aussagen zu treffen.

Im hiesigen Kontext werden Sprachmodell genutzt, um die Surprisalwerte im Sinn von Shannon (1948) zu generieren, indem man Wahrscheinlichkeiten in den Trainingsdaten berechnet und diese auf die Testdaten abbildet. Dies geschieht mit Hilfe der *Maximum Likelihood*-Schätzung:

The maximum likelihood estimate is so called because it is the choice of parameter values which gives the highest probability to the training corpus. [...] It does not waste any probability mass on events that are not in the training corpus, but rather it makes the probability of observed events as high as it can subject to the normal stochastic constraints. (Manning und Schütze 2001: 198)

Weitere Smoothing-Methoden ermöglichen es, Schätzwerte für im Trainingsmodell ungesehene Wortkombinationen zu geben. Über Smoothing wird im Verlauf des Kapitels noch berichtet.

Markoff-Annahmen werden abhängig von der Größe des betrachteten Kontextes in „(n-1)th order Markov model(s)“ (Manning und Schütze 2001: 193) unterteilt. Diese Ränge der Markoff-Annahme werden als N-Gramme bezeichnet. Grundsätzlich ist es wünschenswert, ein N-Gramm-Modell mit möglichst hohem Wert für n zu generieren, da vor allem ein großer Kontext präzise Vorhersagen über die Auftretenswahrscheinlichkeit eines Wortes machen kann. Ein Unigram-Modell ignoriert beispielsweise den umliegenden Kontext vollständig, weil hier nur die Häufigkeiten der einzelnen Wörter miteinbezogen werden (vgl. Manning und Schütze 2001: 201). Zur Bestimmung von Auftretenswahrscheinlichkeiten, speziell zur Vorhersage der nachfolgenden Wörter, eignen sich daher Bigramme, Trigramme oder höhere N-Gramme, die jeweils noch den Kontext miteinbeziehen und abhängige Wahrscheinlichkeiten berechnen, besser.

(1) The girl swallowed a big, green... .

Anhand dieses englischen Beispiels (1) lässt sich der Wechsel der Erwartungen an die Weiterführung des Satzes demonstrieren. Das Prädikat lässt erwarten, dass es sich vielleicht um eine Tablette handelt, die von dem Mädchen geschluckt wird. Diese Vermutung wird aber vielleicht verworfen, wenn die Worte *groß* und *grün* auftreten, nach denen man eher *Frosch* erwarten würde als *Tablette*. Je größer man den Kontext in diesem Beispiel also wählt, desto eher geraten verschiedene Erwartungshaltungen miteinander in Konflikt.¹

Deshalb wird empfohlen, die Größe des Trainingsmodells proportional mit der Höhe n des N-Gramme zu korrelieren. Da das wegen begrenzter Ressourcen nicht immer gewährleistet werden kann, werden Bi- und Trigramme am häufigsten für Language Models verwendet. Neben der Verkleinerung des Kontexts der N-Gramme können auch weitere Unterschiede wie verschiedene Flexionsendungen zwischen dem Trainings- und Testset abgebaut werden (vgl. Manning und Schütze 2001). Dadurch werden weniger Trainingsdaten benötigt, da die Zahl ungesehener N-Gramme durch die geringere Variation in den Daten verkleinert wird. Die Zeichensetzung wird hingegen eher selten eliminiert, da zum Beispiel die Kenntnis des Satzanfangs wichtig ist (vgl. Manning und Schütze 2001: 195).²

Language Models werden klassischerweise über die „Maximum Likelihood estimate“ (Manning und Schütze 2001: 197) berechnet. Dabei werden die Auftretenswahrscheinlichkeiten nach der tatsächlichen, in den Trainingsdaten vorkommenden Häufigkeit zugeteilt. Problematisch ist dabei der Umgang mit Wortkombinationen, die in den Trainingsdaten nicht aufgetreten sind. Diesen Fällen wird eine Wahrscheinlichkeit von 0 zugeordnet (vgl. Manning und Schütze 2001: 198). Gerade bei natürlichsprachlichen Daten ist das wegen der großen Vielfalt und der nicht unendlichen Größe des Trainingskorpus ein Problem, da es zahlreiche Kombinationen gibt, die so unerfasst bleiben. Der Effekt kann auch nicht durch eine beliebige Erhöhung der Trai-

¹ In dieser Form funktioniert das Beispiel aber nur im Englischen, weil das Englische kaum Kasus- und Genusmarkierungen kennt. Im Deutschen wäre ein Hinweis darauf, welche Erwartung korrekt ist, bereits im Genus des Artikels gegeben.

² Manning und Schütze (2001) beziehen sich allerdings nicht auf historische Daten, sodass die Aussagekraft und Vergleichbarkeit von Satzzeichen in ihren Beispielen eher gegeben sind als in historischen Texten, in denen noch keine diesbezügliche Systematisierung stattfand.

ningskorpusgröße eliminiert werden (vgl. Manning und Schütze 2001: 199). Behoben wird die Schwierigkeit stattdessen durch *Smoothing*.

The term smoothing describes techniques for adjusting the maximum likelihood estimate of probabilities to produce more accurate probabilities. The name comes from the fact that these techniques tend to make distributions more uniform, by adjusting low probabilities such as zero probabilities upward, and high probabilities downward. Not only do smoothing methods generally prevent zero probabilities, but they also attempt to improve the accuracy of the model as a whole. (Chen und Goodman 1999: 360)

Im Zuge dieser Definition weisen Chen und Goodman (1999) auch darauf hin, dass es verschiedene Smoothing Methoden gibt, die von ihnen empirisch miteinander verglichen werden. Dabei stellen sie fest, dass das auch von mir verwendete *Kneser-Ney-Smoothing* (vgl. Kneser und Ney 1995) in leicht veränderter Form eine der Methoden ist, die für kleine Trainingsdatensätze am besten funktionieren. Grob zusammengefasst gleicht dieses Smoothing hohe Auftretenswahrscheinlichkeiten von Wörtern, die aber nur in einem sehr speziellen Kontext auftreten können, dort aber praktisch immer vorkommen, dadurch aus, dass es die Unigram-Wahrscheinlichkeiten nicht proportional zur Auftretenshäufigkeit eines (bestimmten) Worts setzt, sondern auch in Betracht zieht, ob das betreffende Wort auch mit anderen Wörtern kombiniert werden kann (vgl. Chen und Goodman 1999: 366).³ Chen und Goodman (1999: 366) verdeutlichen das am Beispiel des Wortpaares *San Francisco*. Durch das häufige Vorkommen des Stadtnamens ist auch die Unigram-Auftretenswahrscheinlichkeit von *Francisco* hoch. Durch das Smoothing wird der Wert jedoch verringert und mehr der Wirklichkeit angepasst, da das Wort praktisch ausschließlich in Kombination mit *San* vorkommen kann.

Verwendetes Language Model Das hier verwendete Korpus ist nicht groß genug, um Trigram-Surprisalwerte sinnvoll berechnen zu können, die neben Bigram-Modellen am häufigsten verwendet werden.⁴ Die Problematik bei der

³ Für eine mathematische Berechnung, die das von mir verwendete Tool (vgl. Kusmirek, Greenberg, Oualil und Klakow 2023) automatisch im Hintergrund durchführt, siehe Kneser und Ney (1995) und Chen und Goodman (1999: 367ff.).

⁴ Teile des in diesem Kapitel vorgestellten Materials flossen auch in die Veröffentlichung Voigtmann und Speyer (2021) und Voigtmann und Speyer (2023) ein.

Verwendung von Bigram-Modellen wurde bereits in Kapitel 1 skizziert.⁵

Eine Lösung dieses Problems schlagen Guthrie, Allison, Liu, Guthrie und Wilks (2006) mit dem *Skipgram-Surprisal*-Modell vor. Dieses Modell berücksichtigt nicht nur die unmittelbar benachbarten Wörter, sondern erlaubt auch das Überspringen von Tokens. Dadurch erhöht sich die Anzahl der möglichen N-Gramme pro Satz, was zu einer größeren Abdeckung von möglicherweise wichtigen N-Grammen führt. Guthrie et al. (2006: 1222) erläutern dies am Beispiel „Insurgents killed in ongoing fighting“. Bei einem klassischen Bigram-Ansatz würde das Modell die folgenden vier Bigramme sehen: „Insurgents killed, killed in, in ongoing, ongoing fight“. Wenn ein 2-Skip-Bigramm-Modell verwendet wird, erhöht sich die Anzahl der Kombinationen auf neun, weil immer wieder ein Element übersprungen wird: „Insurgents killed, insurgents in, insurgents ongoing, killed in, killed ongoing, killed fight, in ongoing, in fight, ongoing fight“ (Guthrie et al. 2006: 1222).

Dieses Modell erfasse den Kontext besser und erreiche eine bessere Abdeckung der Daten, wenn das Trainingskorpus klein ist (vgl. Guthrie et al. 2006: 1223ff.). Obwohl einige Kombinationen nicht sinnvoll sind, verschlechtern sie die Abdeckung des Modells nicht (vgl. Guthrie et al. 2006: 1225). Außerdem kann der positive Effekt ansonsten nur durch eine Vergrößerung des Trainingskorpus, des n der N-Gramme und des betrachteten Kontexts erreicht werden (vgl. Guthrie et al. 2006: 1225). Dies ist das Hauptargument für die Verwendung dieser Art von Sprachmodell. Denn das frühe Neuhochdeutsche und speziell wissenschaftliche Texte dieser Periode sind untererforscht und die Anzahl an passend aufbereiteten Korpora entsprechend gering, so dass die Vergrößerung der Trainingsdaten nicht ohne Weiteres möglich ist.

Für die Untersuchung in der vorliegenden Arbeit wurden fünf verschiedene Language Models auf allen verbleibenden medizinischen und theologischen Texten aus dem DTA (BBAW 2019) trainiert, die nicht in die Testdaten aufgenommen wurden (Anhang C, vgl. Tabelle 5.1). Die Trainingsdaten jeden Zeitabschnitts bilden die Grundlage zur Berechnung der Skipgram-

⁵ Der Unterschied zwischen eingebetteten und ausgelagerten Phrasen oder Relativsätzen wird bei Bigram-Modellen nur in deren unmittelbaren Umgebung deutlich und dürfte sich daher als zu gering erweisen, um einen Einfluss sichtbar zu machen (Kapitel 1).

Surprisalwerte in den Testkorpustexten jeden Zeitabschnitts. Pro Zeitabschnitt wurde also ein eigenes Sprachmodell trainiert.

Mithilfe eines Python-Skripts⁶ wurden die Lemmata aus den Texten gezogen und die Satzzeichen eliminiert, da deren Verwendung im frühen Neuhochdeutschen nicht standardisiert war und somit weder für die Informationsverteilung aussagekräftig ist noch über Autoren und Jahrhunderte verallgemeinert werden kann. Lemmata wurden aufgrund der geringen Menge zur Verfügung stehender Trainingsdaten des frühen Neuhochdeutschen gewählt. Zwar nehme ich dadurch den Verlust von relevanten Kasus- und Genusinformationen in Kauf, aber die Gefahr, keine gut nutzbaren Daten zu erhalten, wenn beispielsweise normalisierte Texte verwendet werden, ist groß genug, um dieses Vorgehen zu rechtfertigen. Alle Modellberechnungen werden mit dem SFB-internen LM-GUI des B4-Projekts durchgeführt (vgl. Kusmirek et al. 2023).

Das Tool dient hauptsächlich der erleichterten Bedienung bei der Berechnung der Language Models. Zunächst wurden die lemmatisierten, von Satzzeichen befreiten Trainingskorpora auf der Seite (vgl. Kusmirek et al. 2023) hochgeladen und entsprechend ihrer Veröffentlichungsdaten zu fünf Korpora zusammengefügt, um die Out-Of-Vocabulary-Ratio (OOV-Ratio) so gering wie möglich zu halten, da manche Begriffe erst zu späteren Zeitpunkten aufgetreten sind (vgl. für das Englische Degaetano-Ortlieb und Teich 2018). Die OOV-Ratio gibt an, wie groß der Anteil der Wörter ist, denen kein Surprisalwert zugeordnet werden kann, weil sie in den Trainingsdaten gar nicht vorkommen.⁷

Für die fünf Sprachmodelle wurden sowohl 2-Skip-Bigram- als auch Unigrammodelle in der Sektion „Language Models“ des GUIs trainiert. Im Anschluss wurden die in den Tabellen 4.2 und 4.3 aufgelisteten Texte in der lemmatisierten und von Satzzeichen befreiten Version ebenfalls im Tool (vgl. Kusmirek et al. 2023) hochgeladen. Einen Teil des Korpus habe ich nicht als eigenes Korpus im Tool hochgeladen, sondern aufgrund der besseren An-

⁶ Ich bedanke mich an dieser Stelle bei Lisa Schäfer für das Schreiben des Skripts.

⁷ In Kapitel 5.1 wurde auch das sogenannte Smoothing erwähnt. Dieses weist nur ungesehenen Kombinationen einen Wert zu. Wenn aber ein Wort in den Trainingsdaten gar nicht aufgetreten ist, kann ihm auch kein Wert zugeordnet werden.

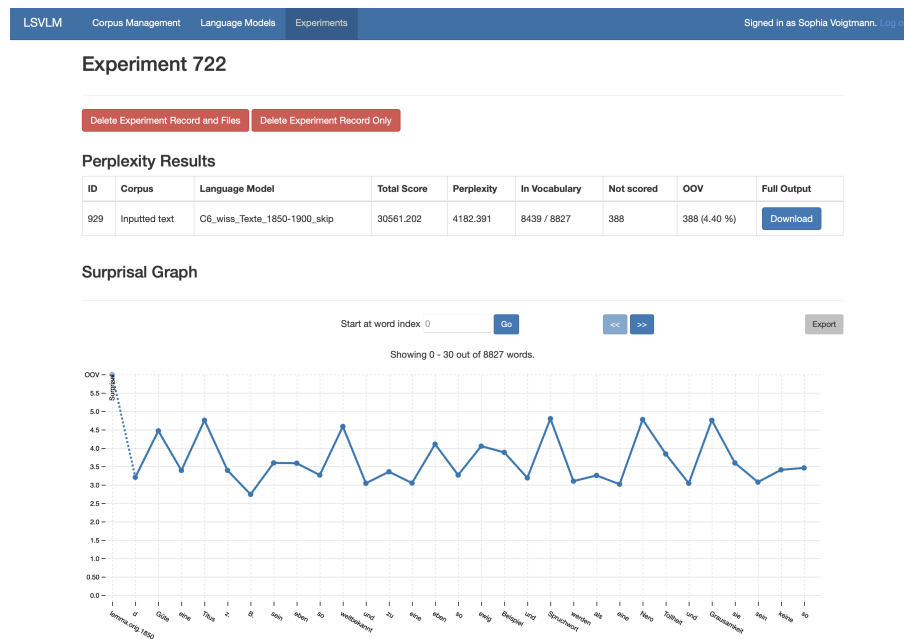


Abbildung 5.1.: Beispiel des Outputs aus dem Language-Model-Graphic-User-Interface (vgl. Kusmirek et al. 2023).

wendbarkeit für diese Texte über die Funktion der manuellen Texteingabe durch Kopieren der passenden Passagen hinzugefügt. Für das Testkorpus wurden in der Sektion „Experiments“ die Surprisalwerte ermittelt, was in Grafiken wie Abbildung 5.1 resultiert. Die so erhaltenen Werte wurden als Tabelle heruntergeladen und in einem nächsten Schritt zu den aus WebAnno gewonnenen Informationen (Kapitel 4.3) hinzugefügt und mit R (R Core Team 2018) für die weitere Berechnung nutzbar gemacht.⁸

Tabelle 5.1 zeigt, dass die OOV-Ratio in allen Modellen unter 15 %, in drei der fünf sogar unter 10 % liegt, was dafür spricht, dass ausreichend Werte erfasst wurden, um aussagekräftige Analysen anhand dieser Werte durchführen zu können.

⁸ Es ist an der Stelle zu betonen, dass nicht nur die Surprisalwerte der hier untersuchten Phrasen und Relativsätze berechnet wurden, sondern die aller Wörter des Satzes, in dem ein betrachtetes Element vorgekommen ist, auch wenn nur die als Extraposition annotierten Satzteile analysiert werden.

Zeitabschnitt	Trainingsdaten (in Tokens)	Testdaten (in To- kens)	OOV- Ratio
1650-1700	2107590	48169	8,93 %
1700-1750	1481259	39251	6 %
1750-1800	2572263	26325	14,72 %
1800-1850	998639	16757	6,28 %
1850-1900	1270561	29060	12,13 %

Tabelle 5.1.: Language Model mit Out-Of-Vocabulary (OOV)-Rate.

5.2. Summiertes und durchschnittliches Surprisal

Im letzten Kapitel 5.1 wurde beschrieben, wie die hier verwendeten Surprisalwerte, berechnet wurden.⁹ Diese Berechnungen führen jedoch nur zu Surprisalwerten pro Wort, was nicht ausreichend für die vorliegende Untersuchung ist, da ich mich mit der Informationsdichte von Konstituenten befasse. Das stellt insofern ein Problem dar, als die meisten Untersuchungen zu grammatischen Phänomenen und Informationsdichte sich nur mit punktuellen Veränderungen im Informationsprofil beschäftigen. Beispielhaft sind hier die Untersuchungen von Levy und Jaeger (2007) Frank und Jaeger (2008) und Jaeger (2010) heranzuführen. Auch sie untersuchen Relativsätze, befassen sich aber mit der Weglassbarkeit des Relativums in englischen Relativsätzen und der Surprisalveränderung auf einem nachfolgenden Wort. Bisherige Theorien und experimentelle Methoden, die Verarbeitungsschwierigkeiten messen, arbeiten also meist mit lokalen Phänomenen. Bigram-Sprachmodelle und bis zu einem gewissen Grad auch Skipgramme sind stark von einem eng definierten Kontext abhängig. Auch Lesezeitstudien, die Verzögerungen bei bestimmten einzelnen Wörtern messen, konzentrieren sich, vereinfacht gesagt, auf Probleme an einzelnen Stellen. Diese Methoden sind nicht gut geeignet, um den gesamten Verarbeitungsaufwand eines (Teil-)Satzes zu ermitteln. Da auch andere Faktoren für das Verständnis relevant sind, wie beispielsweise die parallele Verarbeitung grammatischer Strukturen oder die Einbeziehung verschiedener Quellen (vgl. z. B. Cutler 2008), ist es wichtig, ein Modell zu finden, das sich

⁹ Teile des in diesem Kapitel vorgestellten Materials flossen auch in die Veröffentlichungen Voigtmann und Speyer (2021; 2023) und Speyer und Voigtmann (2021a;b) ein.

dem Gesamtverarbeitungsaufwand annähert, aber auch für Korpusdaten verwendbar ist. Um die Verarbeitungsanforderungen ganzer Konstituenten zu berechnen, werden folglich die schon in Voigtmann und Speyer (2021), Speyer und Voigtmann (2021a;b), Voigtmann und Speyer (2022) und Voigtmann und Speyer (2023) erprobten Verfahren der *summierten* beziehungsweise *kumulativen* Surprisalwerte und *durchschnittlichen* Surprisalwerte zur Darstellung des Verarbeitungsaufwands größerer linguistischer Einheiten vorgeschlagen.

Die durchschnittlichen Surprisalwerte entsprechen dem arithmetischen Mittel. Die Surprisalwerte aller in der Nominal- beziehungsweise Präpositionalphrasen oder dem Relativsatz befindlichen Tokens werden addiert und dann durch ihre Anzahl geteilt. Bei den kumulativen Surprisalwerten wird nur die Addition vorgenommen.¹⁰ Die Berechnung dieser Größen ist also einfach. Schwieriger wird ihre Abgrenzung zum Faktor der Länge.

Je mehr Wörter in einem Satz vorhanden sind, desto mehr Informationen kann er enthalten. Je mehr Informationen vorhanden sind, desto mehr kognitive Kapazitäten werden für die Verarbeitung des Satzes verbraucht. Wenn aber gleichzeitig kognitive Kapazitäten auch für andere Verarbeitungsaufgaben eingesetzt werden, könnte ein dort auftretender Relativsatz oder eine besonders informative Phrase eine Überlastung der verfügbaren kognitiven Kapazitäten verursachen. In diesem Fall müsste die Kommunikation scheitern. Dieser Ansatz wird von den bekannten Theorien zur Extraposition von Relativsätzen von Hawkins (1992) und Gibson (1998) vertreten. Beide Ansätze beschränken sich auf die Messung der Komplexität anhand der Wortzahl. Der Ansatz der Informationsdichte von Shannon (1948) und neuere Forschungen

¹⁰ Die Idee hinter dem kumulativen Surprisalwert folgt der parallelen Verarbeitung von Informationen (vgl. z. B. McClelland und Elman 1986). Die gesamte Theorie der Informationsdichte von Shannon (1948) basiert auf dem inkrementellen Ansatz. Wörter werden nacheinander verarbeitet und die Wahrscheinlichkeit eines Wortes ergibt sich aus seinem Kontext. Deshalb vermute ich Folgendes: Die kognitiven Kapazitäten sind nach der Verarbeitung eines Wortes nicht sofort frei und auch nicht sofort wieder verfügbar. Stattdessen bilden sie eine Art Sockel, der mit jedem weiteren Wort in Abhängigkeit von dessen Surprisal größer wird. Erst wenn die Konstruktion abgeschlossen ist, wird die volle Verarbeitungskapazität wieder frei und der Füllprozess des Sockels kann auf niedrigem Niveau wieder beginnen. Auch wenn von Cuskley, Bailes und Wallenberg (2021) argumentiert wurde, dass mehr Wörter nicht automatisch zu mehr Informationen über bestimmte Wörter und damit vielleicht auch in der Gesamtmenge führen müssen, so ist doch davon auszugehen, dass die Addition von Surprisalwerten mit der Länge des untersuchten Materials korreliert. Je mehr Werte hinzugefügt werden, desto größer kann der kumulative Surprisalwert werden.

von Levy und Jaeger (2007), Levy (2008), Jaeger (2010) und beispielsweise Cuskley et al. (2021) zeigen jedoch, dass eine Zunahme der Länge nicht notwendigerweise mit einer signifikanten Zunahme der Information gleichzusetzen ist, wenn Information als die Vorhersagbarkeit eines Wortes im Kontext verstanden wird. Sowohl der unmittelbare Kontext eines Wortes als auch der erweiterte Kontext können die Wahrscheinlichkeit des Auftretens eines Wortes erhöhen. Dies wiederum würde den Informationsgehalt des betreffenden Wortes verkleinern und könnte schließlich dazu führen, dass der Gesamtinformationsgehalt des Satzes trotz einer höheren Anzahl von Wörtern sinkt (2).

- (2) a. Die Stadt wurde von Cäsar erobert.
 b. Die Stadt Rom wurde von Gaius Julius Cäsar erobert.

(2a) besteht aus fünf Wörtern. Ohne einen größeren Kontext ist der Informationsgehalt von *Cäsar* recht hoch. Wenn man nun an verschiedenen Stellen Wörter hinzufügt und damit die Länge des Satzes erhöht, verringert man gleichzeitig den Surprisalwert anderer Wörter. (2b) wird auf neun Wörter erweitert. Gleichzeitig sollten sowohl die Erwähnung *Roms* als auch die Erwähnung von Cäsars Vor- und Zunamen dafür sorgen, dass die Wahrscheinlichkeit von *Cäsar* mit dem vorangehenden *Gaius Julius* ansteigt und der negativ korrelierte Surprisalwert sinkt. Theoretisch, aber schwieriger zu beweisen, kann je nach verwendetem Language Model sogar die Erwähnung *Roms* dazu führen, dass dem vollständigen Namen ein niedrigerer Surprisalwert zugeordnet wird, da Cäsar und Rom eng miteinander verbunden sind. Die hinzugefügten Wörter können also dafür sorgen, dass der Satz durch die selektive Reduktion des Informationsgehalts leichter zu verarbeiten ist, obwohl er gleichzeitig länger geworden ist.

Dieser Effekt wurde unter anderem von Levy und Jaeger (2007) und in der Folge von Jaeger (2010) bei der Untersuchung von optionalen Elementen im Satz, beispielsweise der optionalen Verwendung von *that* als Relativsatzeinleiter, nachgewiesen (vgl. Jaeger 2005; Frank und Jaeger 2008; Jaeger 2010; 2011). In weniger erwartbaren Kontexten kann die Verwendung des Re-

lativpronomens dafür sorgen, dass ein überwältigend hoher Verarbeitungsaufwand auf dem erste Wort des Relativsatzes reduziert wird. Es kann also an dieser Stelle festgehalten werden, dass Informativität und Länge nicht zwangsläufig positiv miteinander korreliert sein müssen und eine getrennte Betrachtung der beiden sinnvoll ist.

Um den Einfluss der Länge auf den Verarbeitungsaufwand zusätzlich zu kontrollieren, muss der durchschnittliche Surprisalwert¹¹ berechnet werden. Eine Korrelation zwischen Länge und mittlerem Surprisal wird nicht mehr gefunden, die Länge wird praktisch ausgeklammert.

Da der Mittelwert stark von Ausreißern beeinflusst wird, könnte eine Auslagerung, die aus wenigen sehr überraschenden Wörtern besteht, einen hohen durchschnittlichen Surprisalwert haben, was einen höheren Verarbeitungsaufwand erzeugen und die Extraposition begünstigen sollte. Deshalb werde ich in dieser Arbeit ausschließlich den durchschnittlichen Surprisalwert einer Konstituente zur Berechnung ihrer Informativität verwenden.

Weder durchschnittliche noch kumulative Surprisal-Messungen wurden vor den oben erwähnten Untersuchungen zur Erklärung von Extraposition verwendet. Beide Methoden beziehen die gesamte Menge von Surprisalwerten ein, anstatt sich nur auf ein lokales Phänomen und die Wahrscheinlichkeitsveränderung eines Wortes an diesem Punkt zu konzentrieren.

5.3. Orality Score

Wie bereits in Kapitel 2.7 angedeutet, werden die in dieser Arbeit untersuchten Texte auch nach dem Grad ihrer konzeptionellen Mündlichkeit beziehungsweise Schriftlichkeit mithilfe des *Orality Scores* (vgl. Ortmann und Dipper 2024) eingeordnet. Dieser erlaubt es, „einen automatisch bestimmbar Score für Nähe und Distanz zu entwickeln, der anhand einfacher linguistischer Annotationen die Klassifizierung historischer Texte bezüglich ihrer konzeptionellen Mündlichkeit ermöglicht.“ (Ortmann und Dipper 2024: 20). Die folgenden Abschnitte beschreiben seine Berechnung.

¹¹ In den Tabellen wird das durchschnittliche Skipgram-Surprisal als *Skipgram_{mean}* bezeichnet.

Zunächst müssen Vorbereitungen getroffen werden. Ortmann und Dipper (2024) nutzen vor allem Merkmale, die auf Wörtern basieren. Deshalb ist die erste Voraussetzung, um die Scores bestimmen zu können, dass das Korpus mit STTS-POS-Tags (vgl. Schiller et al. 1999) versehen ist. Das wiederum bedingt die Bestimmbarkeit von Wortgrenzen mithilfe von Leerzeichen und Zeilenumbrüchen. Neben der Abgrenzung der Wörter werden Sätze mithilfe von Interpunktionszeichen identifiziert, was allerdings „hauptsächlich der automatischen Verarbeitbarkeit“ (Ortmann und Dipper 2024: 22) dient, da gerade bei sehr mündlichkeitsnahen Texten oder auch oralen Texten die Satzgrenzen nicht immer nach den Standards der Schriftlichkeit interpretierbar sind (vgl. Selting 2012).

Um die Genauigkeit der Merkmalermittlung zu steigern, werden Ellipsen rekonstruiert und vom Standarddeutschen abweichende Schreibweisen normalisiert, sofern diese Information bereits im zu untersuchenden Korpus vorhanden ist. Die Normalisierung erfolgt auf Wort- und Graphemebene, damit die Wortformen vergleichbar werden.¹²

Die Merkmale, die in die eigentliche Analyse einfließen, sind unten aufgeführt (vgl. Ortmann und Dipper 2024: 26). Es ist zu bemerken, dass einige der Features, die noch in Ortmann und Dipper (2019; 2020) als nützlich beschrieben wurden, hier verworfen wurden, um den Herausforderungen bei der Arbeit mit historischen Daten gerecht zu werden. Das trifft beispielsweise auf die Satzgrenzenproblematik zu, die besonders in den älteren Texten des DTA (BBAW 2019) stark zum Tragen kommt. Deshalb werden hier die durchschnittliche Satzlänge und ihr Median von der Analyse ausgenommen (vgl. Ortmann und Dipper 2024: 25). Die Genauigkeit dieser Merkmale ist stark von der Qualität der Annotation abhängig, wobei der geringste Einfluss von Annotationsfehlern auf die durchschnittliche Wortlänge angenommen wird (vgl. Ortmann und Dipper 2024: 27).

- Mittelwert der Wortlänge
- Anteil der Personalpronomen [...] der 1. Pers. Singular und Plural an

¹² Das gilt insbesondere für abweichende Schreibweisen von Personalpronomen der ersten Person Singular und Plural, da diese relevant für die weitere Analyse sind (vgl. Ortmann und Dipper 2019; 2020; 2024).

allen Wörtern

- Verhältnis von Vollverben [...] zu Nomen
- Anteil der Demonstrativpronomen [...] mit Lemmata „diese“ und „die“ welche als die kurze Form (*die*) realisiert sind
- Verhältnis subordinierender Konjunktionen [...] zu Vollverben
- Anteil der primären, Ein-Wort-Interjektionen wie *ach* an allen Wörtern.
- Anteil der Demonstrativpronomen [...] an allen Wörtern
- Anteil der Antwortpartikel [...] an allen Wörtern
- Anteil der Inhaltswörter [...] an allen Wörtern
- Anteil der Interrogativsätze¹³
- Anteil der Exklamativsätze
- Anteil der Sätze, die mit einer koordinierenden Konjunktion beginnen

In einem nächsten Schritt wird den Merkmalen ein „Information Gain (IG)“ (Ortmann und Dipper 2024: 29) zugeordnet. Je höher der Information Gain ist, „desto besser lässt sich Mündlichkeit anhand dieses Merkmals vorhersagen“ (Ortmann und Dipper 2024: 29). In der obigen Auflistung sind die Merkmale bereits abweichend von der Darstellung von Ortmann und Dipper (2024: 29) nach den IGs sortiert, die sich für die Evaluation der Methode anhand des Kasseler Junktionskorpus (vgl. Ágel und Hennig 2022) als am hilfreichsten erwiesen haben. Am meisten trägt die durchschnittliche Wortlänge zur Distinktion zwischen Mündlichkeit und Schriftlichkeit bei, am wenigsten der Anteil der Inhaltswörter. Aufgrund der Annotation im Kasseler Junktionskorpus wurden Features gestrichen, die sich auf Sätze beziehen. Alle verwendeten Features weisen zunächst jedoch nur auf die Zweiteilung zwischen konzeptionell mündlichen und schriftlichen Texten hin.

Um innerhalb einer in diesem Fall mündlichkeitsnahen Konzeption den Grad der Mündlichkeit bestimmen zu können, sind weitere Schritte nötig. Durch eine Korrelationsanalyse wurde herausgefunden, dass sich nur die folgenden Faktoren zur Bestimmung des Scores eignen: die gekürzten und normalen Demonstrativpronomen, das Verhältnis der Inhaltswörter an allen

¹³ Dieses Feature und die nachfolgenden Features wurden in die Gewichtung des Kajak Korpus (vgl. Ágel und Hennig 2022) nicht miteinbezogen.

Wörtern und das Verhältnis der Vollverben zu allen Nomen. Schließlich wird lineare Regression angewendet, „bei der das Modell von einem linearen Zusammenhang zwischen den Merkmalen des Textes und dem gegebenen Näherwert ausgeht. Das Modell besteht dann aus Koeffizienten, die die Gewichtung der einzelnen Merkmale angeben.“ (Ortmann und Dipper 2024: 31).

Basierend auf diesen Schritten kann schließlich der Orality Score berechnet werden. Dazu werden die Information Gain-Werte in Gewichtungen übertragen (vgl. Ortmann und Dipper 2024: 33), die grundsätzlich mit allen in der Liste aufgeführten Features verrechnet werden.¹⁴

Negative Gewichtungen werden für die durchschnittliche Wortlänge, den Anteil der mit einer koordinierenden Konjunktion beginnenden Sätze und den Anteil der Inhaltswörter an allen Wörtern verwendet, da diese weniger für eine konzeptionell mündliche als eine konzeptionell schriftliche Orientierung sprechen (vgl. Ortmann und Dipper 2024: 32). Die Daten werden schließlich „linear auf eine einheitliche Skala von 0 bis 1 abgebildet [...]. Minimum und Maximum werden hierbei für alle Textausschnitte gemeinsam bestimmt“ (Ortmann und Dipper 2024: 32f.). Es ist nicht möglich, absolute, allgemeingültige Werte für eine maximale oder minimale Mündlichkeit anzugeben (Ortmann und Dipper (2024), siehe Kapitel 2.7).

Der so skalierte Wert von jedem Merkmal wird mit der Gewichtung multipliziert. Der Orality Score ergibt sich schließlich aus der Summe dieser Werte (vgl. Ortmann und Dipper 2024: 33). Je höher diese ist, desto stärker wiegen die mit konzeptioneller Mündlichkeit verbundenen Merkmale. Das heißt, dass ein Text mit einem hohen Orality Score mündlichkeitsnäher ist als ein Text, der einen niedrigeren Score erhalten hat. Es wurde schon darauf hingewiesen, dass es kein absolutes Ende der Skala von Mündlichkeit und Schriftlichkeit gibt, auch wenn die Ausführungen von Koch und Oesterreicher (2007) das suggerieren (Kapitel 2.7). Deshalb kann nur die Aussage getroffen werden, dass ein Text mündlichkeitsnäher als ein anderer ist (vgl. Ortmann und Dipper 2024).

¹⁴ Es besteht allerdings immer die Möglichkeit, wie bei der Evaluation am Kassler Junktionskorpus (vgl. Ágel und Hennig 2022) gezeigt, einzelne Merkmale abzuwählen, wenn ein Korpus diese Merkmale nicht beinhaltet.

Der größte Vorteil von Ortmann und Dippers (2024) Orality Score ist dessen automatische Anwendbarkeit. Diese steht mit *COAST* (Conceptual Orality Analysis and Scoring Tool) online zur Verfügung.¹⁵ Die Anwendung basiert auf der Programmiersprache Python. Die Daten müssen im CoNLL-U- oder CoNLL-U Plus-Format vorliegen (vgl. Zeman, Hajič, Popel, Potthast, Straka, Ginter, Nivre und Petrov 2018).¹⁶ Die Auswahl der verschiedenen Features kann frei und ohne Vorgaben in Anpassung an das verwendete Korpus erfolgen. Auch die Gewichtung der einzelnen Features kann prinzipiell selbst angegeben werden. Wird sie nicht spezifiziert, werden die Werte aus Ortmann und Dipper (2024) verwendet. Als Output werden zwei Dateien erstellt; eine mit allen in das Modell eingespeisten Features und eine mit diesen Features und dem daraus resultierenden Orality Score.

Für die Daten dieser Arbeit wurden alle der möglichen Features zur Berechnung des Orality Scores einbezogen, da im Gegensatz zum Kassler Junktionskorpus (vgl. Ágel und Hennig 2022) beispielsweise auch Satzzeichen in ausreichendem Maß vorhanden sind, um die durchschnittliche Satzlänge oder den Anteil von Fragen zu verwenden. Natürlich stellt die Qualität der POS-Tags ein Problem für die Genauigkeit der Analyse dar, wie auch in Ortmann und Dipper (2020) im Hinblick auf das DTA (vgl. BBAW 2019) beschrieben wurde. Diese Ungenauigkeiten werden in Kauf genommen, da die Vorteile der Berechnung überwiegen und Methoden wie die von Ágel und Henning (2012) vorgeschlagene zu zeitintensiv für die hier verwendete Datenmenge gewesen wären.

In den bisherigen Teilabschnitten wurden die Berechnung der verwendeten Surprisalmaße und des Orality Scores beschrieben. Beide werden im weiteren Verlauf der Arbeit genutzt, um die in Kapitel 1.1 benannten Hypothesen zu testen. Wie das geschieht, beschreibt das folgende Kapitel.

¹⁵ Erreichbar ist das Tool und die Anwendungsbeschreibung über: <https://github.com/rubcompling/COAST> [letzter Zugriff: 16.12.2023].

¹⁶ Daten im CoNLL-U- beziehungsweise CoNLL-U Plus-Format liegen in einer durch Tabs separierten Tabelle vor. Jedes Token erhält eine eigene Zeile, weitere Annotationen stehen in separaten Spalten. Sätze werden durch Leerzeilen voneinander abgetrennt. Metainformationen zu den Texten können durch # in die Tabelle integriert werden. Während die verschiedenen Annotationsebenen bei CoNLL-U festgelegt sind, können sie bei CoNLL-U Plus von den Nutzenden des Tools selbst festgelegt gelegt und definiert werden (vgl. Ortmann und Dipper 2022; Zeman et al. 2018).

5.4. Statistische Analyseverfahren

Die betrachteten Phänomene werden mithilfe statistischer Tests im Programm R (R Core Team 2018; 2022) untersucht. Weil sich sowohl bei den Nominal- als auch bei den Präpositionalphrasen und bei den Relativsätzen die Variablen nur in ihrer Ausprägung, aber nicht in ihrer Art unterschieden, werden für alle Auslagerungen die gleichen Tests verwendet. Um Wiederholungen auf ein Minimum zu beschränken, werden die verwendeten Verfahren in diesem Kapitel dargestellt.

Deskriptive Statistik Zunächst wird daher mit einer Beschreibung der Variablen und ihrer Einordnung in die verschiedenen Skalen begonnen, weil sich hieraus Konsequenzen für den weiteren Umgang mit den Variablen ergeben. Unterschieden werden *kategoriale*, *ordinale*, *intervallskalierte* und *rationale* Variablen, wobei Gries (2021: 18f.) die letzten beiden zu einer Klasse der numerischen Variablen zusammenfasst.

Das niedrigste Skalenniveau ist das kategoriale. Daten dieses Niveaus lassen sich lediglich in bestimmte Kategorien unterteilen, die mathematisch bearbeitbare Werte darstellen (vgl. Gries 2021: 18). In meiner Arbeit fallen das Genre und der Relativsatztyp in diese Kategorie. Über Variablen dieses Skalenniveaus kann lediglich die Anzahl der Phänomene pro Kategorie angegeben werden (vgl. Gries 2021: 18). Werden mit R (R Core Team 2018; 2022) weitere Tests mit diesen Variablen berechnet, werden sie entweder mittels Dummy- (0 oder 1) oder Summenkodierung ($-0,5$ oder $0,5$) in numerische Variablen umgewandelt. Das Verfahren ändert nichts an ihrer Aussagekraft oder Bedeutung, ermöglicht aber die Verwendung in Regressionen, da die so kodierten Werte in den finalen Modellen als Haupteffekte auftreten, was bei den anderen Kodierungsarten nicht der Fall ist (vgl. Winter 2020: 145). Zudem kann die Regressionslinie, mit der ein Zusammenhang beobachtbar gemacht wird, nur bei numerischen Werten zwischen den beiden Ausprägungen der Variable gezogen werden (vgl. Winter 2020: 125ff.). Ich verwende für Genre und Relativsatztyp die Summenkodierungen. Für das Genre Medizin wird der Wert $-0,5$ verwendet, für Theologie $0,5$. Die Kodierung wird für alle

Analysen übernommen, da das Genre für jeden Text immer bestimmt wurde. Anders sieht es beim Typ der Relativsätze aus. Restriktive Relativsätze werden mit dem Wert $-0,5$ kodiert, appositive mit dem Wert $0,5$. Für Relativsätze, deren Typ unbestimmt blieb, wird ein Platzhalter (NA) eingesetzt. Die Variable des Relativsatztyps bleibt für sie also in der Regression unbeachtet, ohne dass die Sätze in der deskriptive Statistik aussortiert werden.

In die Kategorie der *ordinalskalierten* Variablen fallen die Zeitabschnitte der Texte, die sich aus den Veröffentlichungsjahren ergeben. Ordinalskalierte Daten lassen sich in eine Reihenfolge bringen (vgl. Gries 2021: 18). Die Zuordnung der Zeitabschnitte zu den ordinalen Variablen ist nicht gänzlich unkompliziert. Vom Prinzip her sind die Abstände zwischen den Zeitabschnitten gleich und messbar. Allerdings umfasst jeder Zeitabschnitt eine Vielzahl an verschiedenen Jahreszahlen mit unterschiedlichen Abständen zueinander (Tabelle 4.3). Deskriptiv werden auch hier vor allem die Anzahl der Texte beziehungsweise Auslagerungen pro Zeitabschnitt angegeben (siehe Tabellen 6.1, 6.12 und 6.21). Werden weitere Tests wie die Korrelationsanalyse in Kapitel 7.1.1 oder logistische Regressionen durchgeführt, wird der Zeitabschnitt numerisch kodiert. Die Ziffer 0 entspricht dann dem Zeitabschnitt 1650 bis 1700, die Ziffer 1 dem Abschnitt 1700 bis 1750, die Ziffer 2 dem Abschnitt 1750 bis 1800, die Ziffer 3 dem Abschnitt 1800 bis 1850 und die Ziffer 4 dem Abschnitt 1850 bis 1900. Im Fall der logistischen Regression dient diese Kodierung wie die Summenkodierung der Modellierung des kontinuierlichen Effekts der Zeit, der angenommen wird.¹⁷ Die Korrelation zwischen Zeitabschnitt und Orality Score, mit deren Hilfe überprüft wird, ob jüngere Texte tatsächlich schriftlichkeitsnäher sind, kann nur berechnet werden, wenn die Skala entsprechend kodiert ist (vgl. Gries 2021: 129f.).

Die anderen Variablen, sprich, die Länge der Auslagerungen, die Surprisalwerte und auch die Orality-Scores fallen unter die Kategorie der *numerischen* Variablen. Für sie ergibt die Ermittlung von Durchschnitt, Minimum und Maximum sowie der Quantile der Werte und Verhältnisse sinnvolle Er-

¹⁷ Die Kodierung der Zeitstufe weicht in Kapitel 7.1.1 von der hier beschriebenen ab, da mir die dort verwendete Kodierung einfacher lesbar erscheint. In diesem Kapitel wird die Zeitstufe 1650 bis 1700 als 1, 1700 bis 1750 als 2 und so weiter kodiert.

gebnisse (vgl. Gries 2021: 18). Statt einer anderen Kodierung der Werte können hier Transformationen notwendig werden, wie unten näher erläutert werden.

Zusammenfassend ist zunächst zu sagen, dass zwei kategoriale Variablen (Genre, Restriktivität der Relativsätze), eine ordinale Variable (Zeitabschnitt) und drei numerische Variablen (Länge, durchschnittliche Surprisalwerte, Orality Score) in die Hypothesentestung einfließen.

Vor der Hypothesentestung erfolgt die deskriptive Statistik. Hier werden Aussagen über die Verteilung der Daten getroffen. Dazu werden neben der Anzahl der Daten in absolute Zahlen und prozentualen Anteilen das *arithmetische Mittel* und die *Standardabweichung* der numerischen Variablen berechnet. Ersteres erlaubt Aussagen über die Ausprägung der Verteilung, während die Standardabweichung die Streuung der Werte um den Durchschnitt angibt (vgl. Winter 2020: 56). Beide Werte sind jedoch sehr stark von besonders hohen oder besonders niedrigen Werten beeinflusst. Robuster gegen diese Abweichungen aufgrund besonders hoher oder niedriger Werte ist der *Median*, der genau die Mitte zwischen allen Werten markiert (vgl. Winter 2020: 58). Wenn er und der Mittelwert nahe beieinander liegen, spricht das dafür, dass der Mittelwert nicht von wenigen extremen Werten in eine Richtung verschoben wird. Um solche Ausreißer besser einordnen zu können, werden zusätzlich auch die *Perzentile* angegeben, wenn die grafische Darstellung der Daten Hinweise darauf liefert, dass es innerhalb eines Datensatzes viele Ausreißer gibt (vgl. Winter 2020: 59f.). In jedem Datensatz gibt es vier Quartile, die nach 25 %-Schritten unterteilt sind. Beim ersten Quartil liegen also 25 % der Werte unter diesem Wert, der von diesem Quartil bezeichnet wird. Beim zweiten liegt die Hälfte aller Daten unter einem bestimmten Wert, beim dritten liegen 75 % aller Daten unter einem bestimmten Wert. Das vierte Quartil entspricht also dem Maximum aller Werte. Durch diese Angaben lassen sich Ausreißer also ebenfalls erkennen und die Daten beschreiben.

Inferenzstatistik In den vorherigen Abschnitten wurden bereits weitere statistische Tests wie die *Korrelation* nach Pearson und die *logistische Regression* genannt. Sie zählen zu den statistischen Modellen, die mathematisch den

Zusammenhang zwischen mindestens zwei Variablen beschreiben (vgl. Gries 2021: 134).

Eine *Korrelation* wird in Form des Korrelationskoeffizienten r zwischen zwei numerischen Variablen berechnet. Der Korrelationskoeffizient r variiert immer zwischen den Werten -1 (negativer Korrelation) und $+1$ (positiver Korrelation). Eine negative Korrelation bedeutet, dass ein Wert steigt, wenn der andere fällt; die positive Korrelation sagt hingegen ein Steigen oder Fallen beider Werte voraus. Je mehr der Korrelationskoeffizient sich -1 oder $+1$ annähert, desto eher kann man von einem (sehr starken) Zusammenhang dieser Variablen ausgehen. Ist r aber (nahe) 0 , kann nicht von einem Zusammenhang gesprochen werden (vgl. Gries 2021: 129).

Korrelationen lassen sich besonders gut grafisch darstellen. In der einfachsten Form wird dazu eine sogenannte Regressionslinie durch die Menge der Datenpunkte gezogen (vgl. Gries 2021: 130f.). Ist diese Linie eine Gerade, spricht man von linearer Regression (vgl. Gries 2021: 131). Die Linie wird über ihren Schnittpunkt mit der y -Achse im Nullpunkt (engl. *intercept*) und den Steigungsparameter (engl. *slope*), der angibt, um wie viel der Wert auf der y -Achse zu- oder abnimmt, wenn der Wert der x -Achse um 1 erhöht wird, bestimmt (vgl. Winter 2020: 71f.). Die Regressionslinie und generell die lineare Regression treffen eine Vorhersage über die Ausprägung einer Variable, wenn die andere verändert wird (vgl. Gries 2021: 130f.). Die Regressionslinie ist dann passend, wenn die gefundenen Datenpunkte einen möglichst niedrigen vertikalen Abstand zu der Regressionslinie, sogenannte Residuen, haben.

Bereits mehrfach fiel der Begriff der *Regression*, wenn es um die Möglichkeiten der Hypothesenüberprüfung geht.¹⁸ Die Regression bildet die hier hauptsächlich zur Hypothesentestung genutzte Methode. Die abhängige Variable bildet dabei die Position der Nominal- und Präpositionalphrasen beziehungsweise der Relativsätze, sodass sich die Verwendung einer binären logistischen Regression anbietet (vgl. Gries 2021: 319). Diese dient dabei nicht

¹⁸ Dabei geht es grundsätzlich darum, die Nullhypothese auszuschließen, die behauptet, dass es keinen Unterschied zwischen den Gruppen, hier also den ausgelagerten und eingebetteten Elementen, gibt (vgl. Winter 2020: 165f.). Die Alternativhypothese, die bestätigt werden soll, behauptet hingegen einen Zusammenhang, der nicht auf Zufall basiert (vgl. Winter 2020: 165f.).

nur dazu, den Zusammenhang einzelner Prädiktorvariablen und der abhängigen Variable, der Position, darzustellen, sondern auch dazu, herauszufinden, welche der diversen, oben genannten Prädiktorvariablen einen Effekt auf die Ausprägung der abhängigen Variable hat und wie dieser Effekt aussieht, der wieder grafisch mit Regressionslinien dargestellt werden kann.¹⁹

Bei der Regression kann es allerdings vorkommen, dass die Werte der unabhängigen Variablen verhältnismäßig weit von der Regressionslinie entfernt sind, ihre Residuen also zu unterschiedlich sind, sodass das Modell nicht konvergiert oder Zusammenhänge nicht korrekt erkennt. Das ist zum Beispiel der Fall, wenn sich die verschiedenen Werte zu sehr voneinander entscheiden (vgl. Gries 2021: 243f.). Für die vorliegende Arbeit ist das relevant, da die abhängige Variable (extrapoliert vs. *in situ*) dummy-gecodet wird. *Extrapoliert* erhält den Wert 0, *in situ* den Wert 1. Die anderen numerischen Variablen liegen allerdings in der Regel nicht innerhalb dieser Spanne, wie die deskriptiven Beschreibungen in den Tabellen 6.4, 6.14 oder 6.23 zeigen.

Aus diesem Grund werden nicht nur die kategorialen Prädiktoren verändert, sondern gegebenenfalls auch Transformationen an den numerischen Variablen durchgeführt (vgl. Gries 2021: 113), um herauszufinden, ob ein Zusammenhang zwischen der unabhängigen und der abhängigen Variable besteht. Dazu wird beispielsweise der Logarithmus auf der Basis von 10 für die numerisch skalierten Werte gebildet. Durch diese Transformation lässt sich die Spanne, innerhalb der die Werte liegen, verringern, sodass der Unterschied zu den kodierten Werten der abhängigen Variable kleiner wird. Ob die Transformation notwendig ist, wird durch das folgende Verfahren bestimmt.

Mithilfe von R (R Core Team 2018; 2022) und *ggplot2* (vgl. Wickham 2016) wird der Anteil von Sätzen abgebildet, die ausgelagert sind (*y*-Achse), wenn sie beispielsweise eine bestimmte Länge (*x*-Achse) haben. Dann wird erst ei-

¹⁹ Betrachtet man die genauen Daten, ist es häufig so, dass eine gerade Linie nicht die beste Abbildung für ein Verhältnis zwischen numerischen Variablen ist. Sichtbar wird das, wenn man die Regressionslinien beispielsweise mit *ggplot2* (vgl. Wickham 2016) erstellt und die Art der Regression nicht spezifiziert. Ein Beispiel für eine solche Linie findet sich in Grafik 7.2. Statt eine gerade Linie zu ziehen und große Residuen und daher eine schlechtere Modellanpassung in Kauf zu nehmen, wird die Regressionslinie so ausgeführt, dass alle Residuen einen möglichst geringen Abstand zur Linie haben. Dabei werden auch schlangenförmige Regressionslinien wie in Grafik 7.2 akzeptiert.

ne lineare Regressionslinie durch diese Punkte und anschließend eine Linie durch die Punkte gezogen, die die Residuen so gering wie möglich hält. Diese Linie wird zunächst aus den untransformierten Werten gebildet (Grafik 5.2).

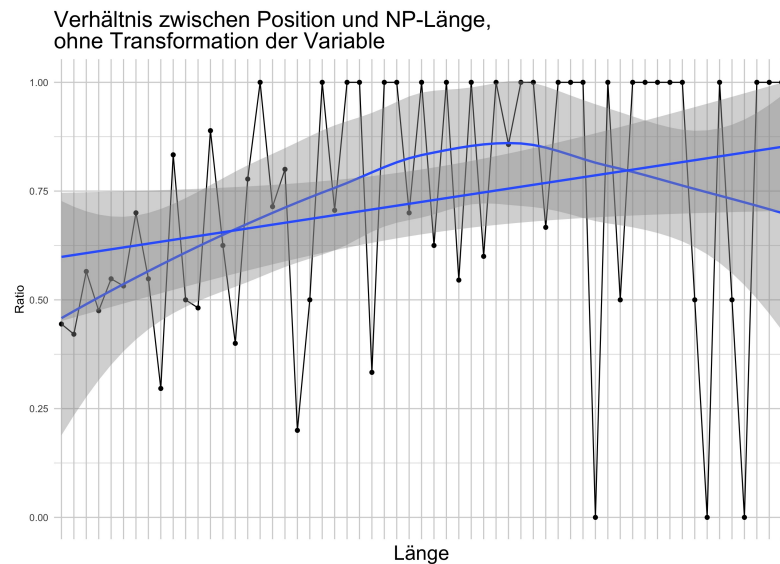


Abbildung 5.2.: Verhältnis zwischen der Position der Nominalphrasen und ihrer Länge. Die hellgrauen Bereiche um die Regressionslinien stellen das Konfidenzintervall dar.

Die Linien laufen nicht parallel, nähern sich einander aber recht gut an. Im nächsten Schritt wird die gleiche Grafik mit logarithmierten Längenwerten erstellt (Grafik 5.3). Zunächst soll bei Grafik 5.2 und Grafik 5.3 der Unterschied in den Linien bemerkt werden. Durch die visuelle Inspektion wird bewertet, ob die logarithmierten oder die natürlichen Werte näher an der linearen Regressionslinie liegen.²⁰ Nach dem gleichen Verfahren werden die numerischen Variablen, also die Surprisalwerte, der Orality Score und im Fall der Relativsätze die Distanz, überprüft.

Nach dieser Testung wird die eigentliche logistische Regression in R (vgl. R Core Team 2018; 2022) mithilfe der *glm*-Funktion durchgeführt (vgl. ‚glm‘, R Core Team 2022). Als abhängige Variable wird die Position genommen. Die unabhängigen Variablen bilden bei allen drei Elementen: Länge, durchschnittliches Skipgramsurprisal, Orality Score, Genre und Zeitstufe. Im Fall der Relativsätze wird auch noch der Typ und die Distanz einbezogen. Au-

²⁰ Für eine Bewertung und Einschätzung dieser beiden Grafiken siehe Kapitel 6.1.1.

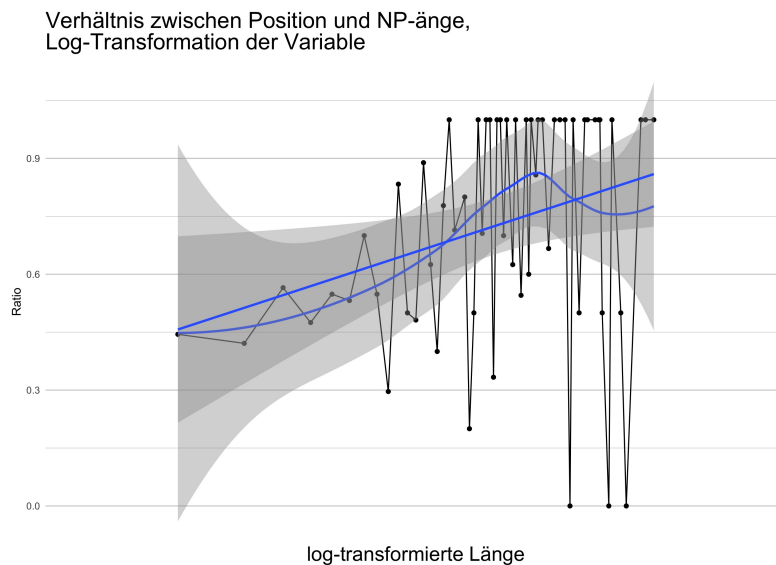


Abbildung 5.3.: Verhältnis zwischen der Position der Nominalphrasen und ihrer Länge nach Logtransformation.

ßerdem werden die Zwei-Wege-Interaktionen zwischen den Variablen berechnet.²¹ Die Familie des Modells wird als binomial spezifiziert, da die abhängige Variable kategorial ist und nur zwei Ausprägungen hat. Für die Interpretation der Ergebnisse Tabellen ist es auch wichtig die Richtung der einzelnen Werte zu verstehen. Da R (R Core Team 2018) automatisch die Ausprägung extraponiert als 0 und die *in situ*-Ausprägung als 1 kodiert, bedeutet ein negativer Estimate, dass Extraposition wahrscheinlicher wird, während ein positiver für eine Tendenz zur Einbettung spricht.

Modellauswahl Wie Gries (2021: 364ff.) darlegt, ergeben sich diverse Probleme aus einem Modell, das zahlreiche Haupteffekte und Interaktionen enthält. Sie erschweren nicht nur die Interpretation des Modells, sondern können auch dessen Aussagekraft verringern. Deshalb wende ich die rückwärts gerichtete Modellauswahl („backward model selection“, Gries 2021: 366) an. Bei diesem Vorgehen werden nacheinander je eine Interaktion beziehungsweise ein Haupteffekt ausgeschlossen. Ausschlaggebend für den Ausschluss ist der

²¹ Die Programmzeile in R (R Core Team 2018; 2022) sieht folgendermaßen aus: `glm(formula = Position ~ (durchschnittlicheSkipgram + Laenge + Orality + Genre + Zeitstufe(+Typ + Distanz))^2, family = binomial(), data = Datenset)`. Die Einklammerung der Werte *Typ* und *Distanz* bezieht sich darauf, dass diese nur für die Relativsätze und dort auch nicht in jeder Regression bestimmt wurden. Letzteres wird in Kapitel 6.3.5 erläutert.

p-Wert (vgl. Gries 2021: 366). Man beginnt mit den Interaktionen und schließt dort zunächst die mit dem höchsten nicht-signifikanten p-Wert aus (vgl. Gries 2021: 366). Um herauszufinden, ob der Ausschluss eines Prädiktors oder einer Interaktion tatsächlich zu einer Verbesserung des Modells geführt hat, wird ein *likelihood ratio*-Test mit der *anova*-Funktion in R (R Core Team 2018; 2022) durchgeführt. Diese Funktion erlaubt den Modellvergleich, da erfasst wird, wie gut das Modell die Daten erklärt (vgl. Winter 2020: 185).

Die Modellauswahl erfüllt also die Funktion des Plausibilitätsquotiententests (vgl. Winter 2020: 185, 261), wenn die beiden zu testenden Modelle in die Analyse einfließen und als Testparameter der Chi-Quadrat-Test genutzt wird. Wenn der p-Wert der *anova* nicht signifikant ist, erklärt das komplexere Modell die Daten nicht signifikant besser.

Die Modellauswahl wird so lange wiederholt, bis nur noch signifikante Effekte oder Haupteffekte, die an einer signifikanten Interaktion beteiligt sind, im Modell enthalten sind. Sobald *likelihood ratio*-Test mit *anova* einen signifikanten Unterschied zwischen den Modellen zeigt, ist der Prozess der rückwärtsgerichteten Modellauswahl abgeschlossen. Das finale Modell entspricht dann dem Modell vor dem Ausschluss des letzten Prädiktors und wird zur Interpretation der Werte und zum Ausschluss der Nullhypothesen verwendet.

Zusammenfassung Um den Einfluss verschiedener Variablen auf die Position der untersuchten Elemente bestimmen zu können, nutze ich verschiedene statistische Verfahren in R (R Core Team 2018; 2022).

Zu den Variablen zählen die kategorialen, summenkodierten Variablen des Genres und des Relativsatztyps, die ordinalskalierte Variable des Zeitabschnitts, die in eine numerische umgewandelt wird, und die per se numerischen Variablen der Länge, des durchschnittlichen Skipgram-Surprisals und bei Relativsätzen der Distanz. Alle Variablen werden zunächst deskriptiv betrachtet. Als Methoden der Inferenzstatistik werden Korrelationsanalysen für die Betrachtung der Orality Scores verwendet und logistische Regression, um den Einfluss der Variablen zu bestimmen. Dazu wird mit einem visuellen Verfahren überprüft, ob die numerischen Variablen logarithmiert werden

müssen. Dann wird eine binomiale, logistische Regression mit allen Variablen durchgeführt. Nacheinander werden einflussarme Interaktionen und Hauptprädiktoren durch eine rückwärtsgerichtete Modellauswahl ausgeschlossen, die auf einem Plausibilitätsquotiententests mithilfe von anova basiert. Die Interpretation der Werte und der Ausschluss der Nullhypothese erfolgt anhand des so kreierten finalen Modells.

Die Anwendung der hier beschriebenen Methoden und deren Ergebnisse werden in den folgenden Kapiteln präsentiert.

6. Ergebnisse zum Einfluss auf Extraposition

Das folgende Kapitel präsentiert die Ergebnisse der Korpusauswertung der Nominalphrasen (6.1), Präpositionalphrasen (6.2), Relativsätze (6.3) und der drei Phänomene zusammen (6.4). Es ist an dieser Stelle zu betonen, dass alle Berechnungen für die Nominal- als auch die Präpositionalphrasen auf den Phrasen erfolgen, deren Ermittlung in den Kapiteln 4.3.3 und 4.3.4 beschrieben wurde, also auf den sogenannten Minimalpaaren.¹ Im Fall der Relativsätze wurden alle im Korpus annotierten adjazenten und extraponierten Relativsätze zur Analyse genutzt.

Zunächst wird die deskriptive Statistik der Elemente vorgestellt. Ein besonderes Augenmerk liegt dabei auf der Anzahl der gefundenen Extrapositionen in ihren jeweiligen (Teil-)Korpora und den Ausprägungen der numerischen Variablen. Danach wird überprüft, ob Transformationen einzelner Variablen notwendig sind. Im Anschluss daran werden die logistischen Regressionen durchgeführt und die Ergebnisse ausgewertet. Hinweise darauf, ob die Nullhypothese ausgeschlossen werden kann, werden ebenfalls hier gegeben. Die Nullhypothese gilt als widerlegt, wenn der Einfluss der Skipgramwerte signifikant für Extraposition ist.

In Kapitel 6.5 erfolgt dann die Diskussion der gefundenen Ergebnisse. Dabei wird überprüft, inwiefern die Ergebnisse Evidenz für das Zutreffen von Hypothese H1 („Relativsätze und Phrasen mit hohen Surprisalwerten werden ausgelagert“) liefern können. Diese Hypothese unterteilt den Satz oberflächlich in zwei Verarbeitungseinheiten mit unterschiedlichen freien Verar-

¹ Folglich werden nur die händisch annotierten Nominal- und Präpositionalphrasen für die Vergleiche verwendet und nicht eine Gesamtheit aller eingebetteten Phrasen.

beitungskapazitäten. Die erste Verarbeitungseinheit umfasst den Bereich vom Vorfeld bis zum Mittelfeld und die zweite Verarbeitungseinheit das Nachfeld. Zum Zeitpunkt der Nachfeldverarbeitung wurden die meisten Informationen des Satzes inklusive der rechten Satzklammer verarbeitet, die in vielen Fällen erst das lexikalische Verb beinhaltet und somit Unsicherheiten über die Satzgliedfunktionen aller Konstituenten final auflöst. Daher bleibt im Nachfeld mehr Kapazität frei, um lexikalisch anspruchsvolles Material zu verarbeiten.

In Kapitel 6.5 werden außerdem Erklärungsansätze für Ergebnisse vorgeschlagen und diskutiert, die von Hypothese H1 abweichen und bei der Betrachtung der logistischen Regressionen aufgetreten sind.

6.1. Nominalphrasen

6.1.1. Nominalphrasen im gesamten Korpus

Im gesamten Korpus wurden 527 Nominalphrasen annotiert. Davon sind 306 (58 %) extrapониert und 221 (42 %) eingebettet.² Die Differenz von 85 Nominalphrasen zwischen den eingebetteten und ausgelagerten Nominalphrasen ergibt sich daraus, dass es nicht immer möglich war, die in Kapitel 4.3 beschriebenen Bedingungen zur Suche von Minimalpaaren einzuhalten.

Obwohl aktuelle Grammatiken Nominalphrasen im modernen Standarddeutschen als wenig extrapositionsfähig ansehen (Kapitel 3.1.1), zeigt sich in diesem Korpus keine Tendenz für die Abnahme nominaler Extraposition in jüngeren Texten, wie aus Tabelle 6.1 und Grafik 6.1 zu entnehmen ist. Sowohl in absoluten als auch relativen Zahlen, gemessen an der Gesamtmenge extrapонierter Nominalphrasen in meinem Korpus, wurden die meisten Nominalphrasen in den Texten von 1650 bis 1700 ausgelagert (n=98, 32 % aller ausgelagerten Nominalphrasen im Korpus). Im nächsten Abschnitt (1700 bis 1750) nimmt diese Zahl zwar ab (n=32, 10,5 %), steigt dafür von 1750 bis 1800 aber wieder an (n=71, 23,2 %). In der ersten Hälfte des 19. Jahrhunderts (1850 bis 1900) ist der absolute Wert ab geringsten (n=20, 6,5 %), bevor er von 1850 bis 1900 wieder ansteigt (n=85, 27,8 %). Insgesamt wirkt die Entwicklung über

² Von diesen 221 stehen 38 im Vorfeld (17,2%), alle anderen befinden sich im Mittelfeld.

den untersuchten Zeitraum hinweg wellenförmig (Tabelle 6.1, Spalte 3³). Für die untersuchten Genres zeigen die beiden rechten Spalten von Tabelle 6.1 allerdings ein gegenläufiges Bild. In den medizinischen Texten wurden im ersten (1650 bis 1700, 60 extraponierte Nominalphrasen) und letzten (1850 bis 1900, 68 extraponierte Nominalphrasen) Zeitabschnitt die meisten Phrasen gefunden, während die theologischen Texte eine Spitze an Extrapositionen von 1750 bis 1800 (63 extraponierte Nominalphrasen) erkennen lassen.

Zeitabschnitt	Gesamtzahl NP (extrap.)	Anteil an Sätzen in Prozent, gerundet	Medizinisch (extrap.)	Theologisch (extrap.)
1650 - 1700	183 (98)	1,26	105 (60)	78 (40)
1700 - 1750	57 (32)	0,25	14 (8)	43 (24)
1750 - 1800	98 (71)	1,19	14 (8)	83 (63)
1800 - 1850	30 (20)	0,25	23 (13)	7 (7)
1850 - 1900	159 (85)	0,45	130 (68)	29 (17)
gesamt	527 (306)	1,19	273 (157)	240 (151)

Tabelle 6.1.: Anzahl der Nominalphrasen nach Zeitabschnitt und Genre.

Während Tabelle 6.1 nur die Anzahl der Nominalphrasen nach Genre und Zeitabschnitt darstellt, zeigt Tabelle 6.2, welche Satzgliedfunktionen die Nominalphrasen in den einzelnen Zeitabschnitten erfüllen. Diese Funktionen wurden nachträglich händisch bestimmt. Es zeigt sich, dass die Satzgliedfunktion kein Kriterium gegen die Nachfeldfähigkeit einer Nominalphrase ist. Für keine der erfassten Funktionen kann eine lineare Zu- oder Abnahme über den untersuchten Zeitraum hinweg festgestellt werden. Nur die Appositionen werden von 1850 bis 1900 deutlich häufiger ausgelagert als in den früheren untersuchten Zeitstufen. Sie zählen zu den freien Angaben, werden also im Gegensatz zu Subjekten und Objekten nicht von der Verbvalenz gefordert. Es ist denkbar und auch im in Kapitel 6.5.1 zu diskutieren, dass dies einen Einfluss auf die Ergebnisse haben wird. Auch die nominalen Attribute im Nachfeld werden in Kapitel 6.1.4 einer separaten Analyse unterzogen.

³ Dieser Wert dient der Normalisierung aufgrund der unterschiedlichen Textgrößen in den verschiedenen Zeitstufen.

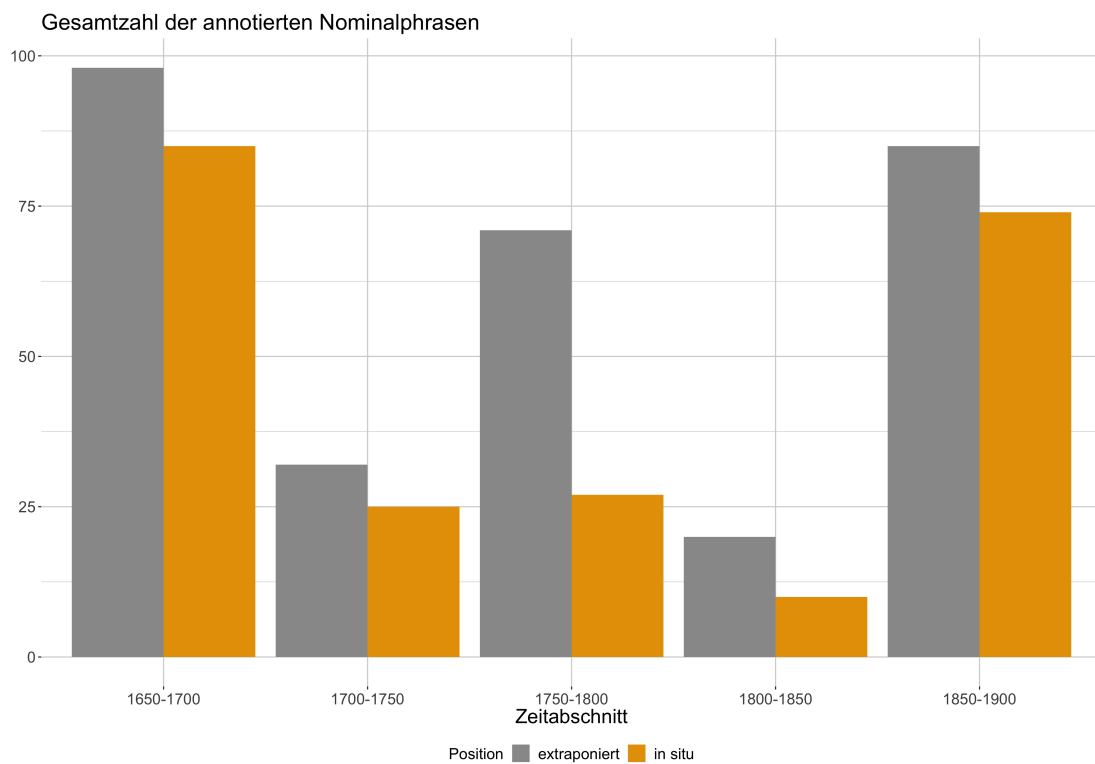


Abbildung 6.1.: Anzahl der annotierten Nominalphrasen pro Zeitabschnitt.

Zeitabschnitt	Subjekte	Prädikative	Akk.obj.	Dat.obj.	Gen.obj.	Attribute	Appositionen	Nachträge	sonstiges
1650 - 1700	26	16	43	11	1	17	2	0	3
1700 - 1750	6	4	11	3	1	0	5	1	0
1750 - 1800	19	12	10	1	0	8	2	7	6
1800 - 1850	1	0	3	0	0	1	2	10	0
1850 - 1900	20	1	36	12	5	7	27	3	2
gesamt	72	33	103	27	7	33	38	21	11

Tabelle 6.2.: Satzgliedfunktion der ausgelagerten Nominalphrasen pro Zeitabschnitt.

In Testung einbezogene Variablen

In Kapitel 5.4 ist ausführlich nachzulesen, welche statistischen Verfahren aus welchen Gründen für die vorliegende Untersuchung angewendet werden. Hier wird zunächst auf die in die Testung eingeflossenen Variablen eingegangen und werden deren mögliche Transformationen dargestellt.

Länge der Nominalphrase im Gesamtkorpus Die Länge aller Nominalphrasen variiert zwischen einem Wort und 144 Wörtern. Im Durchschnitt sind die Phrasen 16,54 Wörter lang und besitzen einen Median von 9 Wörtern. 75 % der Nominalphrasen sind weniger als 21 Wörter lang. Ihr Standardfehler beträgt 0,845 und ihre Varianz 19,39. Die hier genannten Werte beziehen sich auf alle Nominalphrasen, ohne eine Unterscheidung in ihrer Position zu treffen. Tabelle 6.3 stellt hingegen die Länge der *in situ*-Phrasen der Länge der extrapolierten Phrasen gegenüber.

	ausgelagerte NP	eingebettete NP
Minimum	1	1
Median	11,5	8
Durchschnitt	10,04	13,08
drittes Quartil	24	15
Maximum	144	108
Std.-Fehler	1,22	1,05
Varianz	21,36	15,66

Tabelle 6.3.: Länge der Nominalphrasen unterteilt in ausgelagerte und eingebettete Phrasen.

Im Schnitt sind die eingebetteten Phrasen etwas kürzer als ihre ausgelagerten Gegenstücke. 75 % der eingebetteten Phrasen sind weniger als 15 Wörter lang. Dieser Wert beträgt für die ausgelagerten Phrasen aber 24 Wörter. Die Varianzen und die Standardfehler unterscheiden sich nicht stark voneinander. Beide Grundgesamtheiten der untersuchten Ausprägungen der Nominalphrase sind also miteinander vergleichbar.

Im nächsten Schritt wird überprüft, ob die Werte der Länge transformiert werden müssen. Grafik 6.2 zeigt das Verhältnis der Häufigkeit der Auslagerung in Abhängigkeit der zunehmenden Länge, ohne dass die Variable

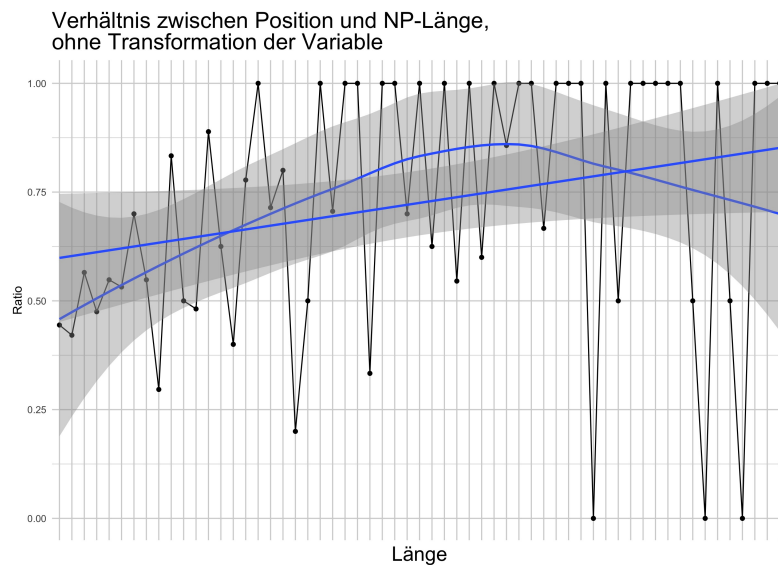


Abbildung 6.2.: Verhältnis zwischen Auslagerung und Länge der Nominalphrasen.

transformiert wurde. Auf der x -Achse wird die Länge der ausgelagerten Nominalphrasen abgebildet, auf der y -Achse die prozentuale Häufigkeit, mit der eine Variable dieser Länge ausgelagert wurde. Die Punkte der Grafik zeigen dabei, mit welcher Häufigkeit die entsprechenden Längen tatsächlich ausgelagert wurden. Die gerade Regressionslinie spiegelt den linearen Zusammenhang zwischen der Länge und der prozentualen Häufigkeit wider. Die gebogene Linie hält hingegen die Residuen möglichst klein. In der Grafik ist zu sehen, dass diese Linien nicht parallel verlaufen. Die angepasste Regressionslinie schneidet die lineare an zwei Punkten und liegt in den meisten Fällen entweder darunter oder darüber.

Für Grafik 6.3 wurden die Längenwerte auf der Basis 10 logarithmiert. Die Grafik selbst kann so gelesen werden wie Grafik 6.2: Die Länge wird auf der x -Achse verortet, der Anteil an Auslagerungen bei dieser Länge auf der y -Achse. Auch hier schneidet die angepasste Regressionslinie die lineare an zwei Stellen und divergiert von ihr in den meisten Fällen. Allerdings läuft sie länger parallel zu der linearen Regressionslinie als in der Grafik der untransformierten Variablen (Grafik 6.2). Aus diesem Grund werden die Werte der Länge für die logistische Regression logarithmiert.

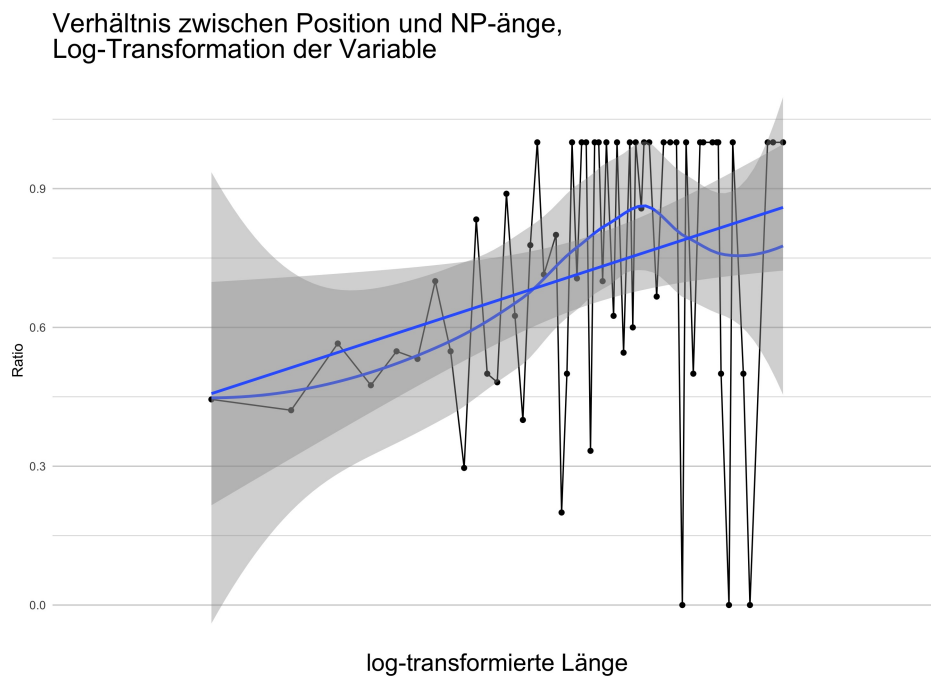


Abbildung 6.3.: Verhältnis zwischen Auslagerung und logarithmierter Länge der NP.

Durchschnittliches Skipgram-Surprisal der Nominalphrasen im Gesamtkorpus Die Skipgram-Surprisalwerte variieren zwischen 2,898 und 4,81 bei Betrachtung aller Nominalphrasen, unabhängig ihrer Position. Durchschnittlich beträgt das Skipgramsurprisal 3,72, der Median 3,688 und 75 % der Werte liegen unter 3,8. Insgesamt zeigt sich eine nicht allzu breite Streuung der Werte, was sich auch in Varianz und Standardfehler widerspiegelt. Beide liegen bei 0,012. Die Unterscheidung der durchschnittlichen Skipgramsurprisalwerte der Nominalphrasen anhand ihrer Position erfolgt in Tabelle 6.4.

	eingebettete NP	ausgelagerte NP
Minimum	2,898	2,993
Median	3,655	3,709
Durchschnitt	3,674	3,752
drittes Quartil	3,815	3,865
Maximum	4,519	4,801
Std.-Fehler	0,017	0,015
Varianz	0,26	0,27

Tabelle 6.4.: Durchschnittliche Skipgram-Surprisalwerte der Nominalphrasen unterteilt in ausgelagerte und eingebettete Phrasen.

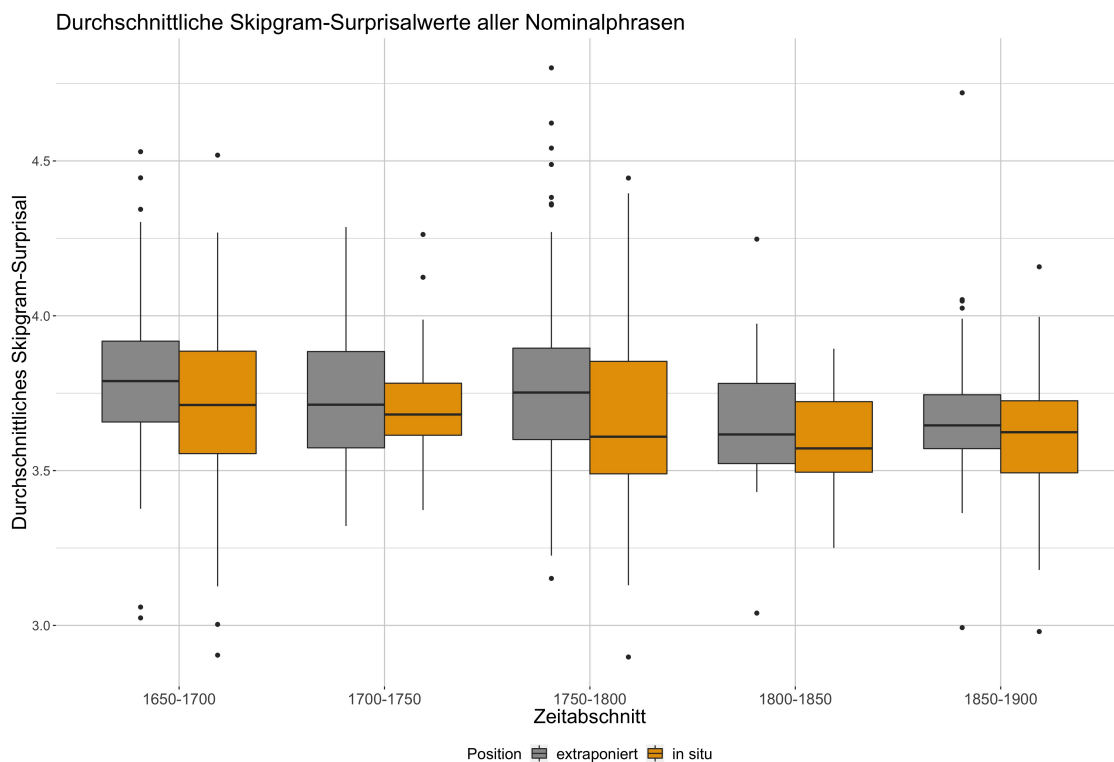


Abbildung 6.4.: Durchschnittliche Skipgram-Surprisalwerte der Nominalphrasen nach Zeitabschnitt.

Der Unterschied zwischen den durchschnittlichen Skipgramwerten der eingebetteten und ausgelagerten Phrasen wird erst im Nachkommabereich sichtbar. Die extraponierten Nominalphrasen besitzen marginal höhere Werte als ihre Gegenstücke in allen dargestellten Bereichen (vgl. Grafik 6.4).

Die gelben Balken der eingebetteten Nominalphrasen weisen in allen Zeitstufen niedrigere Werte auf als die grauen Balken der extraponierten Phrasen. Auch sind Ausreißer nach oben bei den eingebetteten Phrasen offensichtlich seltener, was sich auch in der recht kleinen Varianz und dem kleinen Standardfehler widerspiegelt (Tabelle 6.4). Bei den ausgelagerten Phrasen hingegen ist die Zahl der Ausreißer nach oben besonders von 1750 bis 1800 deutlich höher (Grafik 6.4). Auch die Unterschiede in den Medianlinien weisen darauf hin, dass die eingebetteten Nominalphrasen weniger überraschend zu sein scheinen als die extraponierten Nominalphrasen.

Um die Inferenzstatistik durchführen zu können, wird auch für die durchschnittlichen Skipgram-Surprisalwerte überprüft, ob sie logarithmiert werden

müssen. Betrachtet man Grafik 6.5, erkennt man kein klares Ergebnis hinsichtlich der durchschnittlichen Skipgramwerte. Die beiden Regressionslinien deuten aber darauf hin, dass der Anteil ausgelagerter Nominalphrasen mit einer Erhöhung der durchschnittlichen Skipgram-Surprisalwerte zunimmt. Zudem verlaufen die beiden Linien weitestgehend parallel, was dafür spricht, dass eine Logarithmierung nicht notwendig ist. Grafik 6.6 bestätigt das, da bei Logarithmierung keine bessere Annäherung der beiden Linien zu erkennen ist. Deshalb werden die durchschnittlichen Skipgram-Surprisalwerte für die logistische Regression, die auf dem gesamten Korpus berechnet wird, nicht transformiert. Die Werte des Orality Scores wurden mit dem gleichen Verfahren überprüft, wobei die Notwendigkeit einer Logarithmierung nicht festgestellt wurde.

Des Weiteren wurde überprüft, ob es eine signifikante Korrelation zwischen den durchschnittlichen Surprisalwerten und der Länge der extrapolierten Phrasen gibt. Dies ist nicht der Fall, wie der Korrelationskoeffizienten nach Pearson für die Länge und die durchschnittlichen Skipgram-Surprisalwerte ($r=-0,044$) zeigt. Der Korrelationskoeffizient wurde mit R (vgl. R Core Team 2022) berechnet. Die Länge und das durchschnittliche Surprisal können also als voneinander unabhängig betrachtet werden.

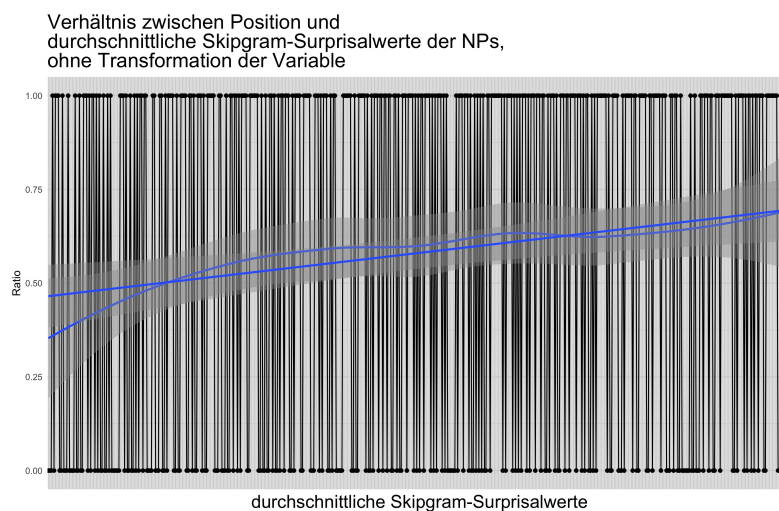


Abbildung 6.5.: Verhältnis zwischen der Position und den durchschnittlichen Skipgram-Surprisalwerten der NPs.

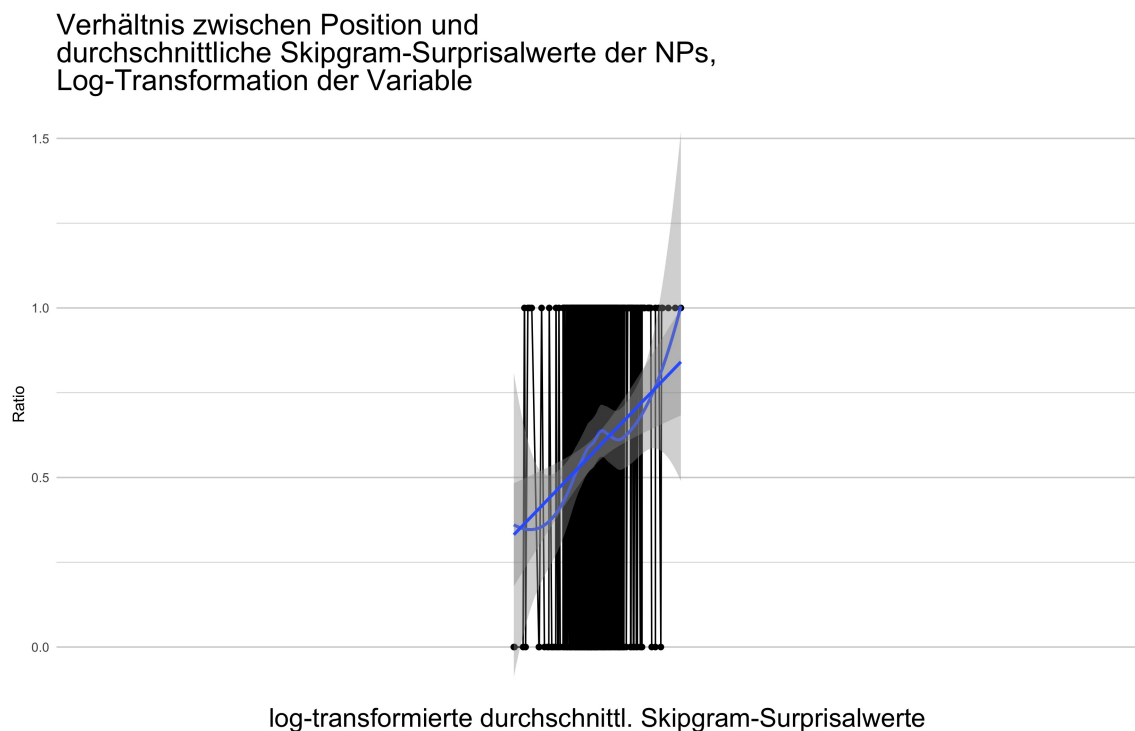


Abbildung 6.6.: Verhältnis zwischen der Position und den durchschnittlichen Skipgram-Surprisalwerten der NPs.

Logistische Regression der Nominalphrasen des gesamten Korpus

Die abhängige Variable der logistischen Regression ist die Position der Phrasen. Die unabhängigen Variablen bilden die logarithmierte Länge, die durchschnittlichen Skipgram-Surprisalwerte, der Orality Score, das summenkodierte Genre (Medizin: $-0,5$, Theologie: $0,5$) und die in numerische Variablen umgewandelten Zeitstufen. Die Berechnung erfolgt über eine generelle lineare logistische Regression mit R (glm, R Core Team 2022). Die Familie des Modells wird als binomial spezifiziert. Zusätzlich zum Einfluss der Hauptvariablen werden die Zweizegeinteraktionen zwischen ihnen berechnet. Danach erfolgt eine rückwärtsgewendete Modellauswahl basierend auf dem *likelihood ratio*-Test mit der *anova*-Funktion anhand des p-Wert-Kriteriums.

Das finale Modell der logistischen Regression in Tabelle 6.5 zeigt ein signifikantes Ergebnis der durchschnittlichen Skipgram-Surprisalwerte für Extraposition ($z=-2,444$, $p<0,05$). Dadurch ist die Nullhypothese widerlegt, denn der Faktor hat offensichtlich einen relevanten Einfluss auf die abhängige Va-

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	2,77	1,64	1,69	0,09	.
Skipgram _{mean}	-0,92	0,37	-2,44	0,01	*
Länge _{log}	-0,33	0,69	-0,48	0,63	
Orality Score	3,45	1,39	2,47	0,01	*
Genre	-2,36	0,92	-2,58	0,009	**
Zeitstufe	0,38	0,26	1,45	0,15	
Länge _{log} :Orality Score	-2,39	1,35	-1,77	0,08	.
Länge _{log} :Genre	1,81	0,97	1,87	0,06	.
Länge _{log} :Zeitstufe	-0,46	0,26	-1,75	0,08	.
Orality Score:Genre	2,50	0,83	3,00	0,003	**
Orality Score:Zeitstufe	-0,63	0,23	-2,68	0,007	**

Tabelle 6.5.: Logistische Regression aller im Korpus annotierten Nominalphrasen, finales Modell. Negative z-Werte sprechen für Extrapolation.

riable. Evidenz für das Zutreffen von Hypothese H1, die besagt, dass Phrasen mit hohen Surprisalwerten eher ausgelagert werden, wird folglich gefunden. Die Länge ist trotz ihres in der Literatur beschriebenen hohen Einfluss auf Extrapolation kein signifikanter Hauptprädiktor ($z=0,48$, $p=0,63$). Auch die Zeitstufe bildet keinen signifikanten Haupteffekt ab ($z=1,45$, $p=0,15$). Das Veröffentlichungsdatum eines Texts kann Extrapolation nicht erklären. Es scheint also keine lineare Präferenz für Auslagerung in alten beziehungsweise jungen Texten zu geben, wenn nur diese Variable betrachtet wird.

Einen höheren p-Wert als die durchschnittlichen Skipgramwerte erzielen allerdings das Genre ($z=-2,58$, $p<0,01$) und der Orality Score ($z=2,47$, $p<0,05$). Dabei deuten sie aber in gegenläufige Richtungen und sind zudem an Interaktionen beteiligt. Mündlichkeitsnähe sagt als Haupteffekt Einbettung voraus, Genre hingegen Auslagerung.

Die Ergebnisse der Interaktionen sind gemischt. Die Interaktion zwischen Länge und Genre ($z=1,86$, $p < 0,1$) zeigt: Lange Nominalphrasen aus theologischen Texten sind eher eingebettet. Die Interaktionen aus Länge und Orality Score ($z=-1,76$, $p<0,1$) lässt erkennen, dass lange Phrasen in mündlichkeitsnahen Texten eher ausgelagert und kurze Phrasen in mündlichkeitsnahen Texten eher eingebettet werden. Ebenfalls eine Beteiligung der Länge weist die

Interaktion zwischen ihr und der Zeitstufe auf ($z=-1,75$, $p<0,1$). Es wird deutlich, dass lange Phrasen, die aus jüngeren Texten stammen, eher ausgelagert werden.⁴ Die Interaktion aus Orality Score und Zeitstufe ($z=-2,68$, $p <0,05$) besagt Folgendes: Der Anteil von Auslagerungen nimmt in mündlichkeitsnahen Texten und jüngeren Texten über die Jahrhunderte hinweg zu. Die Interaktion aus dem Orality Score und dem Genre ($z=3,0$, $p<0,05$) lässt folgende Schlüsse zu: Nominalphrasen in mündlichkeitsnahen theologischen Texten werden deutlich häufiger eingebettet als Nominalphrasen in mündlichkeitsnahen medizinischen Texten. Dieses Ergebnis führt zu der Entscheidung, die beiden Genre in den Kapitel 6.1.2 und 6.1.3 voneinander getrennt zu betrachten.

6.1.2. Nominalphrasen in den medizinischen Texten

In den medizinischen Texten wurden insgesamt 157 (54,9 %) ausgelagerte und 129 (45,1 %) eingebettete Nominalphrasen gefunden ($n=286$). Tabelle 6.6 zeigt die deskriptiven Werte der Länge und des Skipgram-Surprisals und ist wie folgt zu lesen: Die ausgeklammerte Zahl gibt die Werte von eingebetteten und ausgelagerten Phrasen zusammen an, die eingeklammerten Zahlen unterteilen die Phrasen nach ihren Positionen. Dabei stehen die Werte der eingebetteten Nominalphrasen links des geraden Strichs und die der extrapolierten rechts davon. Wieder ist zu erkennen, dass die ausgelagerten Phrasen in der Regel etwas höhere Werte aufweisen als ihre *in situ*-Gegenstücke.

Um zu überprüfen, welcher Faktor die Position der Nominalphrase vorhersagt, wird wieder logistische Regression angewendet. Mit dem gleichen Vorgehen, wie im vorherigen Kapitel beschrieben, wird erneut visuell überprüft, ob eine Transformation der Variablen nötig ist: Das Ergebnis der Betrachtung ist, dass keine Transformationen nötig sind.

⁴ Anhang E.1.1 zeigt das Ergebnis einer logistischen Regression, in die nur Nominalphrasen, die aus vier Tokens bestehen, eingeflossen sind. Auch wenn die Länge vollständig kontrolliert wird, zeigt sich, dass die Position dieser 40 Nominalphrasen (19 extrapolierte und 21 *in situ*-Phrasen) vor allem durch deren Surprisalwerte vorausgesagt werden kann. Auch hier wird die Hypothese H1 bestätigt.

	Länge		Skipgram _{mean}	
Minimum	1	(1 1)	2,90	(2,90 2,99)
Median	9	(8 12)	3,67	(3,64 3,70)
Durchschnitt	14,95	(11,45 17,82)	3,71	(3,65 3,75)
drittes Quartil	21	(15 24)	3,85	(3,79 3,873)
Maximum	102	(48 102)	4,72	(4,52 4,72)
Standardfehler	0,89	(0,91 1,4)	0,02	(0,02 0,02)
Varianz	15,03	(10,3 17,52)	0,26	(0,26 0,26)

Tabelle 6.6.: Deskriptive Werte der Nominalphrasen aus medizinischen Texten. Links in der Klammer stehen die Werte der eingebetteten NPs, rechts die der ausgelagerten NPs.

Logistische Regression der Nominalphrasen des medizinischen Teilkorpus

Das Modell der logistischen Regression enthält die unabhängigen Variablen: durchschnittliche Skipgram-Surprisalwerte, Länge, Orality Score und Zeitstufe sowie die Zweibegeinteraktionen zwischen den einzelnen Faktoren. Das Genre wird aus dem Modell ausgeschlossen, da hier nur noch eine Ausprägung vorliegt. Wieder wird die rückwärtsgerichtete Modellauswahl mithilfe von anova (vgl. R Core Team 2022) angewendet. Das finale Modell ist in Tabelle 6.7 dargestellt und zeigt, dass Hypothese H1 bestätigt werden kann.

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	5,85	2,01	2,92	0,004	**
Skipgram _{mean}	-1,28	0,53	-2,40	0,016	*
Länge	-0,18	0,04	-4,64	<0,001	***
Orality Score	0,50	0,59	0,85	0,39	
Länge:Orality Score	-0,24	0,07	-3,59	0,0003	***

Tabelle 6.7.: Logistische Regression der im medizinischen Korpusteil annotierten Nominalphrasen, finales Modell. Negative z-Werte sprechen für Extraposition.

Phrasen mit hohen durchschnittlichen Skipgram-Surprisalwerten weisen auf Auslagerung hin ($z=-2,4$, $p<0,05$). Das gleiche gilt für Phrasen mit einer großen Länge, wobei dieser Wert einen niedrigeren p-Wert aufweist als der für die durchschnittlichen Surprisalwerte ($z=-4,64$, $p<0,001$). Der Orality Score ist der einzige verbleibende Hauptprädiktor, der keinen signifikanten Wert auf-

weist. Da er aber in Interaktion mit der Länge signifikant für Extraposition ist ($z=-3,59$, $p<0,01$), kann er nicht aus dem Modell ausgeschlossen werden. Diese Interaktion spricht dafür, dass lange Nominalphrasen in mündlichkeitsnahen Texten mit hoher Wahrscheinlichkeit ausgelagert werden, während kurze Nominalphrasen über alle Orality Scores hinweg eher eingebettet werden.

6.1.3. Nominalphrasen in den theologischen Texten

In den theologischen Texten wurden insgesamt 149 (61,8 %) extraponierte und 92 (38,2 %) eingebettete Nominalphrasen annotiert ($n=241$). Die deskriptiven Werte für die Skipgramwerte und die Länge finden sich in Tabelle 6.8.

	Länge		Skipgram _{mean}	
Minimum	1	(1 2)	2,89	(2,89 3,02)
Median	10	(8 10)	3,69	(3,67 3,72)
Durchschnitt	18,43	(15,37 20,32)	3,73	(3,70 3,75)
drittes Quartil	21	(15 24)	3,85	(3,83 3,86)
Maximum	144	(108 144)	4,80	(4,45 4,80)
Standardfehler	1,51	(2,17 2,03)	0,017	(0,03 0,02)
Varianz	23,43	(20,85 24,77)	0,27	(0,26 0,28)

Tabelle 6.8.: Deskriptive Werte der Nominalphrasen aus theologischen Texten. Links in der Klammer stehen die Werte der eingebetteten NPs, rechts die der ausgelagerten NPs.

Wieder haben die ausgelagerten Phrasen in allen Bereichen etwas höhere Werte als die eingebetteten. In der Hinsicht unterscheiden sich die theologischen also nicht von den medizinischen Texten. Anhand dieser Zahlen ist es aber nicht möglich zu unterscheiden, ob die Extraposition von der Länge, den durchschnittlichen Skipgramwerten oder den anderen Variablen vorausgesagt werden kann. Um das zu klären, wird im nächsten Abschnitt die logistische Regression durchgeführt.

Logistische Regression der Nominalphrasen des theologischen Teilkorpus

Die durchschnittlichen Skipgram-Surprisalwerte, die Länge, der Orality Score und die Zeitstufe der Nominalphrasen sowie deren Zweifacheinteraktionen

bilden die unabhängigen Variablen zur abhängigen Variable der Position. Das finale Modell (Tabelle 6.9) wurde mittels rückwärtsgerichteter Modellauswahl über anova (vgl. R Core Team 2022) ermittelt.

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	-9,13	7,11	-1,28	0,199	
Skipgram _{mean}	2,37	1,87	1,26	0,21	
Länge	- 0,28	0,167	-1,65	0,099	.
Orality Score	26,08	12,29	2,12	0,03	*
Zeitstufe	0,09	0,17	0,57	0,57	
Skipgram _{mean} :Länge	0,07	0,046	1,61	0,11	
Skipgram _{mean} :Orality Score	-6,82	3,23	-2,11	0,035	*
Länge:Zeitstufe	-0,04	0,015	-2,88	0,0039	**

Tabelle 6.9.: Logistische Regression der im theologischen Korpusteil annotierten Nominalphrasen, finales Modell. Negative z-Werte sprechen für Extrapolation.

Im Gegensatz zu den medizinischen Texten⁵ weisen die Skipgramwerte keinen signifikanten Einfluss auf Auslagerung auf ($z=1,26$, $p=0,21$). Ein Beleg für Hypothese H1 wird daher nicht gefunden. Auch die Länge erklärt Extrapolation nur tendenziell ($z=-1,65$, $p<0,1$). Signifikant für Einbettung ist stattdessen der Orality Score ($z=2,12$, $p<0,05$). Je mündlichkeitsnäher ein theologischer Text ist, desto eher werden die Nominalphrasen eingebettet.

Allerdings finden sich signifikante Interaktionen zwischen den Skipgram-Surprisalwerten und dem Orality Score ($z=-2,11$, $p<0,05$). Wenn eine Nominalphrase also in einem mündlichkeitsnahen Text vorkommt und zusätzlich hochinformativ ist, wird sie mit größerer Wahrscheinlichkeit ausgelagert. Im schriftlichkeitsnahen Korpusteil werden hochinformativ Nominalphrasen in diesem Register vermehrt eingebettet. Ein ebenfalls signifikantes Ergebnis liefert die Interaktion zwischen Länge und Zeitstufe ($z=-2,88$, $p<0,01$). Lange Nominalphrasen in jüngeren Texten stehen eher im Nachfeld. Diese Tendenz ist auch bei langen Nominalphrasen im ersten Zeitabschnitt (1650 bis 1700)

⁵ Der Effekt, dass bei den nach Genre getrennten Regressionen andere Prädiktoren signifikant sind als bei der Regression auf dem gesamten Korpus, entsteht, weil die Regressionen auf verschiedene Gesamtheiten von Werten, nämlich der Teilkorpora beziehungsweise des Gesamtkorpus, zurückgreifen und sich daraus unterschiedliche mathematische Datenbasen ergeben, die in unterschiedlichen Intercepts und Slopes resultieren und daher auch zu anderen Werten führen können (vgl. Winter 2020; Gries 2021).

festzustellen, allerdings werden hier auch lange Nominalphrasen etwas eher eingebettet als in den anderen vier Zeitabschnitten.

Zusammenfassend ist an dieser Stelle zu sagen, dass man einen deutlichen Unterschied zwischen den beiden Genres sieht. Während die medizinischen Texte einen starken Einfluss von informationellen Faktoren aufweisen, ist der Effekt in theologischen Texten nur bei hochinformativen Phrasen in mündlichkeitsnahen Texten zu erkennen. Das lässt vermuten, dass sich die beiden Genre getrennt voneinander entwickeln. Informationelle Aspekte der Auslagerung sind bei den theologischen Texten untergeordnet, während sie bei den medizinischen eine größere Bedeutung haben. Eine genauere Interpretation dieser Ergebnisse erfolgt in Kapitel 6.5.1.

6.1.4. Extraposition von attributiven Nominalphrasen

Dadurch, dass die Extraposition der Nominalphrasen auch in Zusammenhang mit der Stellung von attributiven Relativsätzen gesetzt werden soll, bietet es sich an, zu überprüfen, ob sich die attributiven Nominalphrasen anders verhalten als die gesamte Menge von Nominalphrasen. Allerdings wurden nur 23 extraponierte, attributive Nominalphrasen im Korpus gefunden: Neun attributive Nominalphrasen stehen im Zeitraum von 1650 bis 1700, sieben von 1750 bis 1800, eine von 1800 bis 1850 und sechs von 1850 bis 1900. Von 1700 bis 1750 wurden keine attributiven Nominalphrasen annotiert.

Über die `sample`-Funktion in R (vgl. R Core Team 2022) wurden aus den Zeitabschnitten, in denen attributive Nominalphrasen annotiert wurden, eingebettete Nominalphrasen zufällig ausgewählt, um einen Vergleich zu haben.⁶ Verglichen werden hier also die 23 ausgelagerten attributiven Nominalphrasen mit 23 eingebetteten Nominalphrasen. Deskriptiv ergeben sich für diese Phrasen folgende Werte (Tabelle 6.10). Auch hier scheinen die Werte der ausgelagerten Phrasen stets über denen der eingebetteten zu liegen. Eine Ausnahme bildet der Ausreißer von 108 Wörtern in der Länge der eingebetteten Nominalphrasen.

⁶ Bei der Annotation selbst wurde nicht zwischen attributiven und eigenständigen Phrasen unterschieden. Daher wurde dieser Umweg zur Minimalpaarfindung gewählt.

	Skipgram _{mean}		Länge	
Minimum	3,25	(3,25 3,39)	2	(2 2)
Median	3,66	(3,64 3,69)	9	(9 9)
Durchschnitt	3,69	(3,68 3,71)	16,44	(19,05 13,71)
drittes Quartil	3,78	(3,73 3,79)	21	(18,75 24)
Maximum	4,45	(4,45 4,11)	108	(108 43)
Standardfehler	1,59		0,036	

Tabelle 6.10.: Durchschnittliche Skipgram-Surprisalwerte und Länge der attributiven NP und ihrer Minimalpaare. Links in der Klammer stehen die Werte der eingebetteten NPs, rechts die der ausgelagerten NPs.

Regressionsanalyse der attributiven Nominalphrasen hinsichtlich der Extraposition Die logistische Regression erfolgte mit den gleichen Variablen, die nicht transformiert werden mussten, wie oben bereits angegeben. Das finale Modell ist in Tabelle 6.11 abgebildet.

	Estimate	Std.-Fehler	z-Wert	p-Wert
Intercept	24,99	15,50	1,61	0,11
Skipgram _{mean}	-7,05	4,28	-1,65	0,09
Länge	0,08	0,05	1,62	0,11
Zeitstufe	-6,70	4,46	-1,50	0,13
Skipgram _{mean} :Zeitstufe	1,94	1,25	1,56	0,12
Länge:Zeitstufe	-0,03	0,02	-1,73	0,08

Tabelle 6.11.: Logistische Regression der attributiven Nominalphrasen.

Bei den Hauptprädiktoren ist nur der Wert der Skipgram-Surprisal signifikant ($z=-1,65$, $p<0,1$). Weder die Länge noch die Zeitstufe haben als Haupteffekte einen Einfluss, der Orality Score wurde sogar ganz ausgeschlossen. Die Länge wird erst dann relevant, wenn der Text jüngeren Datums ist ($z=-1,73$, $p<0,1$). Je länger die Nominalphrase und je jünger der Text ist, aus dem sie stammt, desto wahrscheinlicher wird Auslagerung. Interessant ist an dieser Stelle, dass ein hoher Informationsgehalt von attributiven Nominalphrasen offenbar relevanter für die Position der Phrase ist als ihre Länge, was wiederum für eine Bestätigung der ersten Hypothese H1 spricht. Die Ergebnisse müssen dennoch mit Vorsicht betrachtet werden, weil die Anzahl der gefundenen

attributiven Nominalphrasen so gering ist und der Vergleich zwischen attributiven, extraponierten und eigenständigen, eingebetteten Phrasen erfolgt ist. Auf die generellen Unterschiede zwischen den attributiven und eigenständigen Nominalphrasen soll gesondert in Kapitel 6.5.1 eingegangen werden.

6.2. Präpositionalphrasen

6.2.1. Präpositionalphrasen im gesamten Korpus

Im gesamten Korpus wurden 1942 Präpositionalphrasen annotiert. Davon entfallen 1043 (53,7 %) Belege auf extraponierte und 899 (46,3 %) auf *in situ*-Präpositionalphrasen.⁷ Wie bereits in Kapitel 4.3.3 dargelegt, ergibt sich die unterschiedliche Anzahl an extraponierten und *in situ*-Phrasen durch die zu beachtenden Faktoren wie die vergleichbare Komplexität und die gleiche Präposition bei der Minimalpaarsuche. Davon unabhängig ist hingegen der Unterschied zwischen den medizinischen und theologischen Texten. In den medizinischen Texten finden sich 1,6mal so viele annotierte Phrasen (n=1214) wie in den theologischen Texten (n=728).

Betrachtet man die Anzahl der extraponierten Phrasen in den verschiedenen Zeitabschnitten, ergibt sich ein ähnliches Bild wie bei den Nominalphrasen. Die meisten extraponierten Präpositionalphrasen (n=239, 22,9 % aller ausgelagerten Präpositionalphrasen im Korpus) werden im ersten (1650 bis 1700) und im letzten Abschnitt (1850 bis 1900) gefunden (n=499, 47,8 %). Auch von 1750 bis 1800 ist sowohl die absolute (n=132, 12,7 %) als auch die relative Anzahl an Präpositionalphrasen gemessen an den Sätzen höher (Tabelle 6.12). Von 1700 bis 1750 wurden nur 103 ausgelagerte Präpositionalphrasen gefunden (9,9 %) und von 1800 bis 1850 sogar nur 70 (6,7 %). Insgesamt enthalten nie mehr als drei Prozent aller Sätze eine extraponierte Präpositionalphrase. Es handelt sich also um ein insgesamt niederfrequentes Phänomen. Gleichzeitig zeigt sich aber auch im Korpus ein Beleg dafür, dass Präpositionalphrasen häufiger ausgelagert werden als Nominalphrasen (vgl. z. B. Zifonun et al. 1997: 1651).

⁷ 80 *in situ*-Präpositionalphrasen stehen im Vorfeld (8,8 %). Der Rest findet sich im Mittelfeld.

Zeitraum	Gesamtzahl PP (extrap.)	Anteil Extrapo- sition an Sätzen in Prozent, gerundet	Medizinisch (extrap.)	Theologisch (extrap.)
1650 - 1700	433 (239)	3,08	275 (156)	158 (83)
1700 - 1750	184 (103)	0,81	37 (19)	147 (84)
1750 - 1800	231 (132)	2,22	20 (10)	211 (122)
1800 - 1850	121 (70)	0,96	64 (35)	57 (35)
1850 - 1900	973 (499)	2,63	818 (421)	155 (78)
gesamt	1942 (1043)	4,05	1214 (641)	728 (402)

Tabelle 6.12.: Anzahl der Präpositionalphrasen nach Zeitraum und Genre.

Auch die Entwicklung der beiden Phrasentypen ist hinsichtlich der Anzahl der ausgelagerten Phrasen pro Zeitraum sehr ähnlich. In beiden Fällen wirkt die Entwicklung wellenförmig, wie Grafik 6.7 erkennen lässt, die die absoluten Zahlen auf der y-Achse abbildet. Die zweite Spalte in Tabelle 6.12, die den Anteil von Sätzen mit extrapolierten Präpositionalphrasen an allen Sätzen im Korpus angibt, spiegelt die Wellenbewegung in Prozenten wider.⁸

Auch innerhalb der beiden Genre ist die Entwicklung nicht linear. In den medizinischen Texten wurden die meisten extrapolierten Phrasen von 1650 bis 1700 (n=156) und von 1850 bis 1900 (n=421) gefunden. In den Zeiträumen dazwischen wurden in den medizinischen Texten jeweils weniger als 50 Phrasen extrapoliert (Tabelle 6.12). Anders sieht das Bild bei den theologischen Texten aus. Bis auf die Zeiträume 1650 bis 1700 und 1850 bis 1900 enthalten die theologischen Texte immer mehr ausgelagerte Phrasen als die medizinischen, was mit der größeren Mündlichkeitsnähe dieser Texte zusammenhängen dürfte (Kapitel 7.1.1). Der Unterschied wird auch in Grafik 6.8 sichtbar, auch wenn hier nicht zwischen ausgelagerten und eingebetteten Phrasen unterschieden wird.

In einem ersten Schritt wurden attributive von eigenständigen Präposi-

⁸ Die hohe Anzahl an annotierten Präpositionalphrasen von 1850 bis 1900 ist vor allem dem Text von Ludwig zuzuschreiben (n=400 PP). Ohne den Text betrüge der prozentuale Anteil von Sätzen mit annotierten Präpositionalphrasen von 1850 bis 1900 nur 0,8%, was für eine Abnahme spricht.

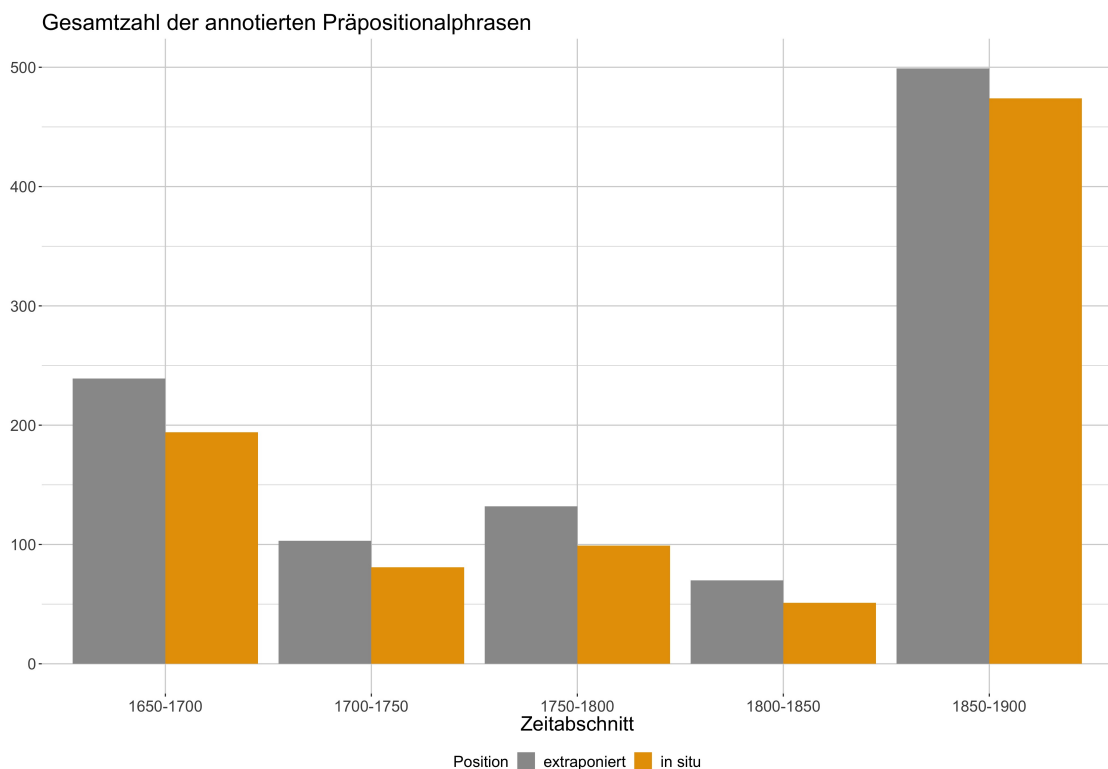


Abbildung 6.7.: Gesamtzahl der annotierten Präpositionalphrasen in absoluten Zahlen.

tionalphrasen unterschieden und dabei festgestellt, dass im gesamten Korpus nur 21 attributive, ausgelagerte Präpositionalphrasen existieren. Da die Unterscheidung erst in einem zweiten Schritt getroffen wurde, konnten nur 16 attributive eingebettete Präpositionalphrasen gefunden werden, die den in Kapitel 4.3.3 aufgelisteten Kriterien entsprechen. Von den insgesamt 37 attributiven Präpositionalphrasen wurden 22 in medizinischen und 15 in theologischen Texten gefunden. Im Zeitabschnitt 1750 bis 1800 wurden keine attributiven Präpositionalphrasen annotiert, obwohl in dem Zeitabschnitt insgesamt 132 Präpositionalphrasen im Nachfeld gefunden wurden, sodass nicht davon ausgegangen werden kann, dass dieses Ergebnis insgesamt an einer zu geringen Anzahl ausgelagerter Phrasen liegt.

In einem nächsten Schritt wurde für die Präpositionalphrasen die Unterscheidung getroffen, ob es sich bei ihnen um fakultative Ergänzungen, Objekte oder freie Angaben handelt. Die Auswertung und Zählung erfolgten manuell und werden in Grafik 6.9 dargestellt.

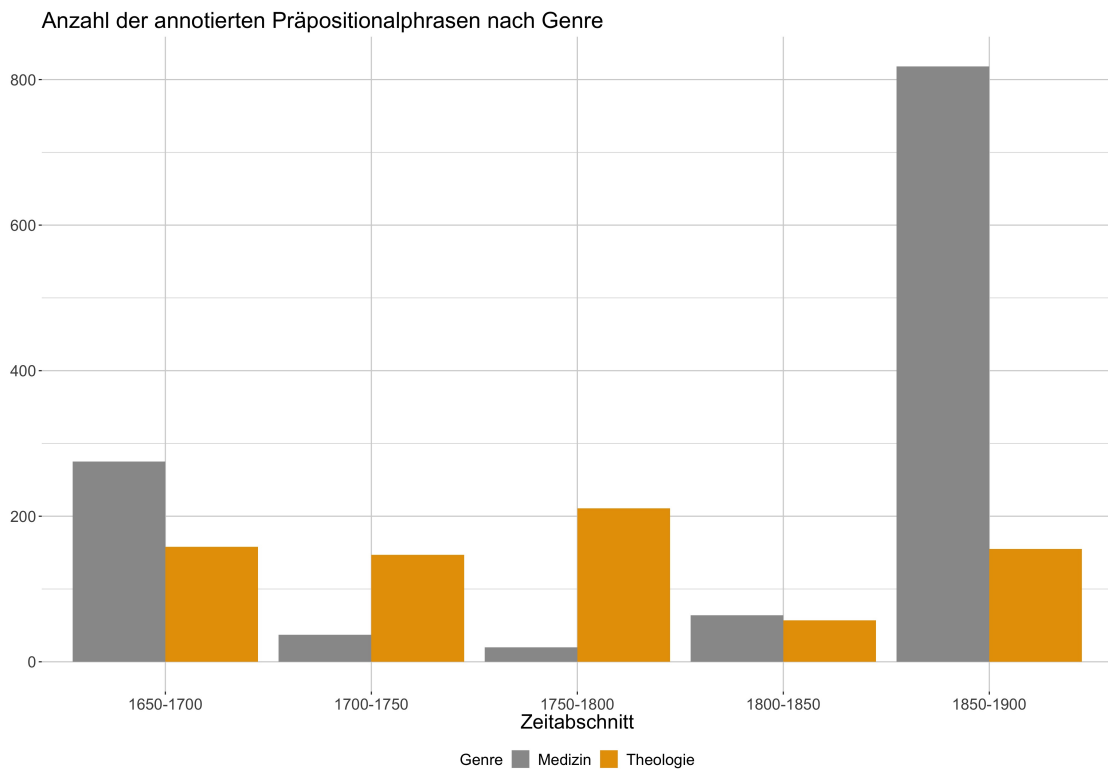


Abbildung 6.8.: Gesamtzahl der annotierten Präpositionalphrasen, unterteilt in die beiden Genre.

In allen Zeitabschnitten wurden sowohl Objekte als auch Ergänzungen als auch Angaben ausgelagert, ohne dass man konstant davon sprechen kann, dass ein Valenzstatus mit besonderer Häufigkeit ins Nachfeld gestellt wurde.

Wie auch im vorherigen Kapitel werden nacheinander verschiedene Untersuchungen an den Phrasen durchgeführt, die dazu dienen, herauszufinden, welche Faktoren einen Einfluss auf die Position einer Präpositionalphrase ausüben. Das Vorgehen entspricht dem bei den Nominalphrasen angewendeten und in Kapitel 5.4 beschriebenen Verfahren. Nach der Überprüfung auf nötige Transformationen werden alle Präpositionalphrasen im Gesamtkorpus deskriptiv beschrieben und dann mittels der logistischen Regression analysiert.

In Testung einbezogene Variablen

Länge der Präpositionalphrase im Gesamtkorpus Die Länge der Präpositionalphrasen in diesem Korpus variiert zwischen zwei und 225 Wörtern.⁹

⁹ Ein Beispiel für eine lange Präpositionalphrase im Nachfeld liefert (1).

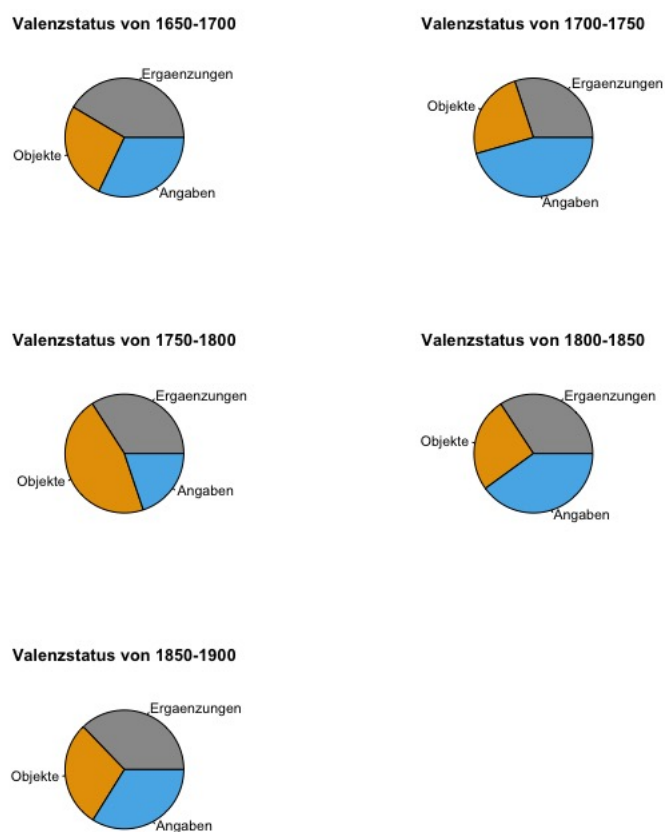


Abbildung 6.9.: Valenzstatus der Präpositionalphrasen nach Zeitabschnitten.

Offensichtlich ist die Phrase mit 225 Wörtern ein Extremfall. Im Durchschnitt sind ausgelagerte Phrasen 19,77 Wörter und im Median 14 Wörter lang. 75% aller ausgelagerten Phrasen sind kürzer als 26 Wörter (Tabelle 6.13). In allen aufgeführten deskriptiven Werten, vom Minimum abgesehen, sind die eingebetteten Präpositionalphrasen kürzer als ihre ausgelagerten Gegenstücke. Besonders stechen hier die Werte des dritten Quartils (eingebettet: 12 Wörter, ausgelagert: 26 Wörter) und des Maximums (eingebettet: 84 Wörter, ausgelagert: 225 Wörter) hervor. Erstere sind weniger als halb so lang wie die ausgelagerten Präpositionalphrasen, außerdem ist die längste eingebettete Präpositionalphrase nur ein Drittel so lang wie die längste ausgelagerte Präpositionalphrase. Interessant ist zudem die Streuung der Werte um den Mittelwert herum. Bei den eingebetteten Präpositionalphrasen zeigt sich eine viel kleinere Varianz ($s^2=134,51$) als bei den ausgelagerten Präpositionalphrasen ($s^2=415,04$), was auf eine hohe Anzahl von Ausreißern hinweist.

	eingebettete PP	ausgelagerte PP
Minimum	2	2
Median	8	14
Durchschnitt	11,28	19,77
drittes Quartil	12	26
Maximum	84	225
Std.-Fehler	0,39	0,63
Varianz	134,51	415,04

Tabelle 6.13.: Länge der Präpositionalphrasen unterteilt in eingebettete und ausgelagerte Phrasen.

Grafik 6.10 zeigt die vermutete große Anzahl an Ausreißern in allen Zeitabschnitten, aber auch bei den eingebetteten Präpositionalphrasen. Sowohl von 1650 bis 1700 als auch von 1850 bis 1900 ist die Anzahl von Ausreißern

- (1) Ich wil jetzo nicht reden von dem elend der Christlichen Kirchen/ in deroselben glieder/ welche unter den unrecht lehrenden/ in dem Babylonischen gefängnüß des Antichristischen Roms; unter der nicht nur so schweren Türckischen tyranney/ sondern auch theils unglaublicher unwissen heit/ teils vielen vermengten irrthumen/ insgesamt erschrocklichen ärgernüssen in griechen und den Morgen ländern/ und unter so vielen irrigen lehren anderer von dem Pabst zwar abgetretenen/ aber zu der reinigkeit der lehr nicht gekommenen/ gemeinden/ verborgen ligen/ und in höchster gefahr mit furcht und zittern ihre seligkeit würken müssen (Spener 1676, Satz 28)

bei den eingebetteten Phrasen größer als in den verbleibenden drei Zeitabschnitten. Die stark unterschiedlichen Varianzen erschweren allerdings die Vergleichbarkeit der beiden Phrasentypen. Die Wahrscheinlichkeit, dass die Variable transformiert werden muss, ist dadurch erhöht.

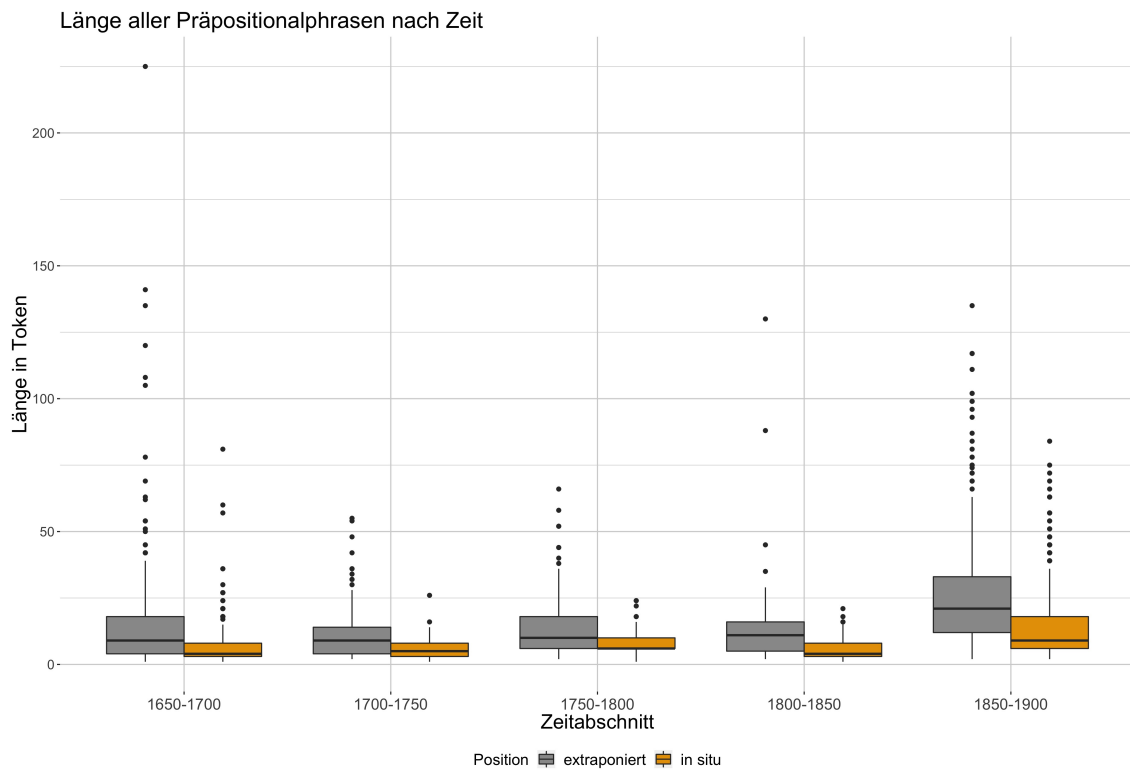


Abbildung 6.10.: Unterscheidung der Länge der Präpositionalphrasen zwischen eingebetteten und ausgelagerten in allen Zeitabschnitten.

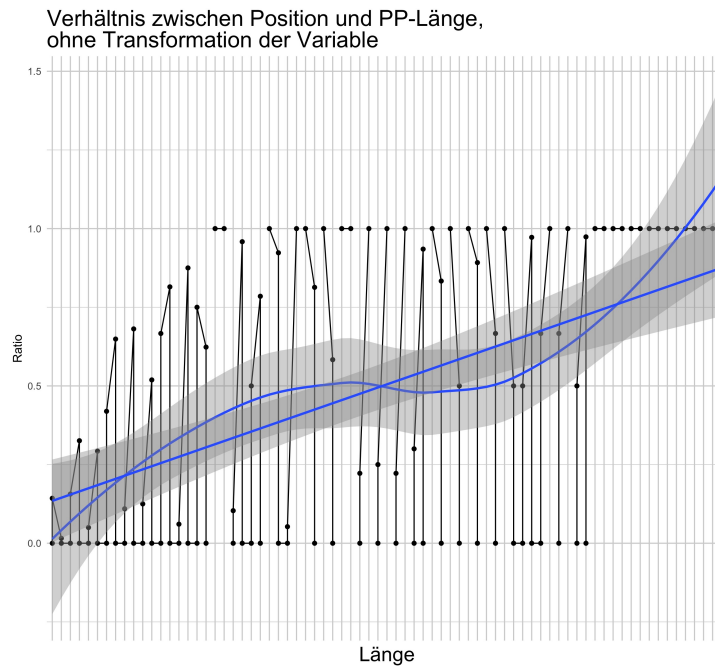


Abbildung 6.11.: Verhältnis zwischen Auslagerung und Länge der PP, ohne Transformation der Variable Länge.

Um diese Transformationen geht es im Folgenden. Grafik 6.11 zeigt das Verhältnis zwischen der Länge der Präpositionalphrase auf der x-Achse und dem Anteil der Phrasen, die bei einer bestimmten Länge ausgelagert sind, auf der y-Achse. Auslagerungen werden häufiger, je länger eine Präpositionalphrase ist. Dafür sprechen beide Regressionslinie. Die residuenverringere Linie schneidet die lineare Regressionslinie allerdings an drei Punkten und weist auch sonst stärkere Abweichungen auf, die nicht mehr vorkommen, wenn die Länge auf der Basis 10 logarithmiert wird (Grafik 6.12). Zwar verlaufen die Kurven in den Randbereichen nicht parallel zueinander, aber im Bereich der Mittleren Werte sind die Linien beinahe deckungsgleich, was dafür spricht, dass die Länge der Präpositionalphrasen bei der Analyse des Gesamtkorpus auf der Basis 10 logarithmiert werden.

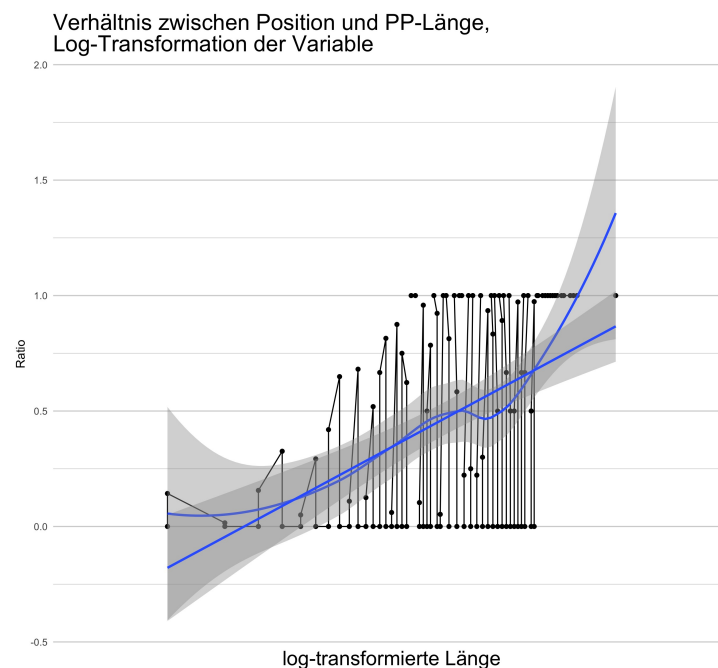


Abbildung 6.12.: Verhältnis zwischen Auslagerung und Länge der PP, mit Transformation der Variable Länge.

Durchschnittliches Skipgram-Surprisal der Nominalphrasen im Gesamtkorpus In Tabelle 6.14 sind die deskriptiven Skipgram-Surprisalwerte der eingebetteten und ausgelagerten Phrasen zu sehen. Die Unterschiede zwischen den Positionen sind marginal. Lediglich beim Minimum (eingebettet: 2,798, ausgelagert: 2,987) und dem Median (eingebettet: 3,62, ausgelagert: 3,65) liegen die Werte der eingebetteten unter denen der ausgelagerten Phrasen. Auch im Durchschnitt weisen sie einen niedrigeren Wert auf, allerdings erst auf der zweite Nachkommastelle. Es zeigt sich, dass 75 % der eingebetteten Phrasen unter dem Wert 3,8 liegen, während dieser Wert bei den exponierten Präpositionalphrasen 3,7 beträgt. Sowohl die Varianz als auch der Standardfehler sind gering und weisen nur kleine Unterschiede auf. Es ist also nicht von allzu vielen Ausreißern bei den Phrasen auszugehen. Dieser Eindruck wird auch von Grafik 6.13 gestützt. Lediglich im letzten Zeitabschnitt 1850 bis 1900 sind sowohl bei den eingebetteten als auch bei den ausgelagerten Präpositionalphrasen mehr als drei Ausreißer nach oben und nach unten zu sehen. Sowohl die Ähnlichkeit der Werte zueinander als auch die Nähe zu den dummy-codierten Werten der Position lassen vermuten, dass keine

Transformation notwendig wird, was ebenfalls überprüft wurde.

	eingebettete PP	ausgelagerte PP
Minimum	2,798	2,987
Median	3,619	3,647
Durchschnitt	3,646	3,662
drittes Quartil	3,802	3,773
Maximum	4,382	4,416
Std.-Fehler	0,39	0,63
Varianz	0,008	0,007

Tabelle 6.14.: Durchschnittliche Skipgram-Surprisalwerte der Präpositionalphrasen unterteilt in ausgelagerte und eingebettete Phrasen.

Die Logarithmierung der Skipgram-Surprisalwerte führt nicht zu einer verbesserten Abbildung der beiden Regressionslinien aufeinander. Betrachtet man Grafik 6.14, die den Anteil an Auslagerungen ins Verhältnis zu den durchschnittlichen Skipgram-Surprisalwerten setzt, zeigt sich keine starke positive Korrelationslinie, wie es bei der Länge der Fall war (Grafik 6.11). Die lineare Regressionslinie verläuft beinahe parallel zur x-Achse, auf der die durchschnittlichen Skipgramwerte abgebildet sind, und wird von der Regressionslinie, die die Residuen möglichst gering hält, an zwei Stellen geschnitten. Diese Linie zeichnet keinen klaren Zusammenhang zwischen steigenden durchschnittlichen Skipgramwerten und einer Auslagerungstendenz. Eher lässt sie den Schluss zu, dass Präpositionalphrasen mit besonders hohen durchschnittlichen Skipgramwerten zu einem niedrigeren Prozentsatz ausgelagert werden, was Hypothese H1 in Frage stellen würde. Um aber den Zusammenhang zwischen den beiden Variablen und der Position der Präpositionalphrasen besser erkennen zu können, wird im nächsten Kapitel die logistische Regression der Phrasen durchgeführt.

Davor wurde wie in Kapitel 6.1 mit R (vgl. R Core Team 2022) überprüft, ob die Länge und die durchschnittlichen Skipgram-Surprisalwerte korrelieren. Das ist nicht der Fall ($r = -0,037$).

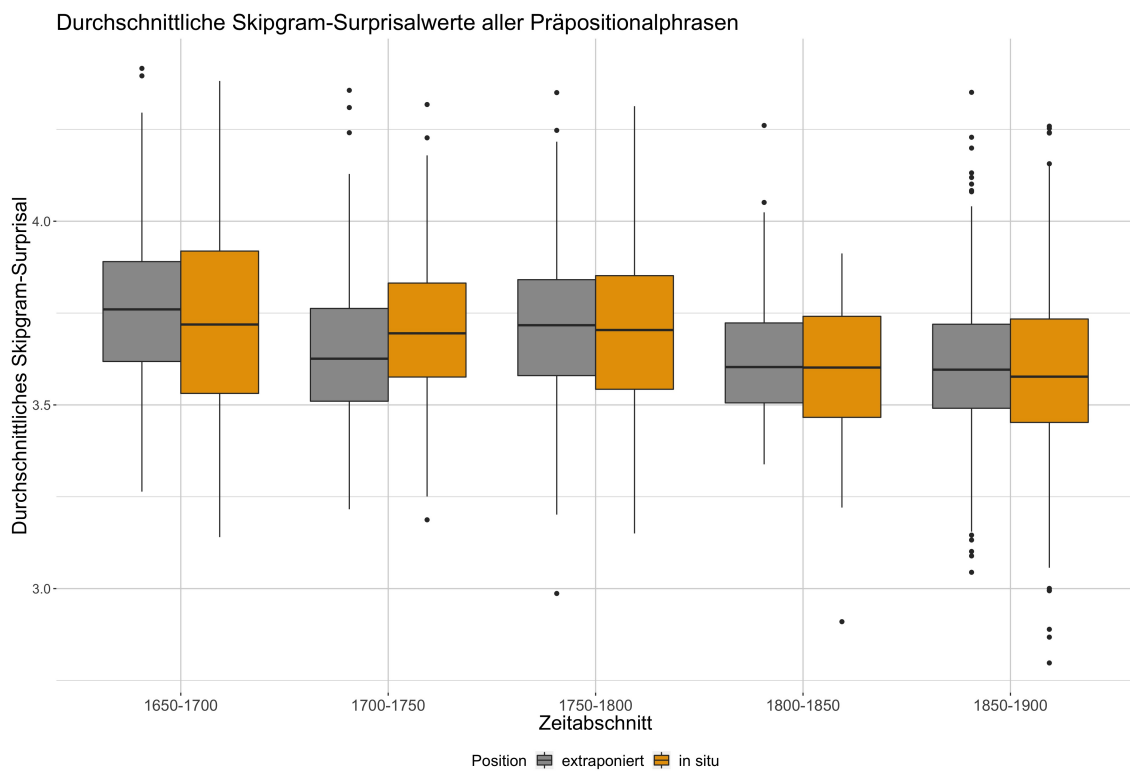


Abbildung 6.13.: Durchschnittliche Skipgram-Surprisalwerte der Präpositionalphrasen nach Zeitabschnitt.

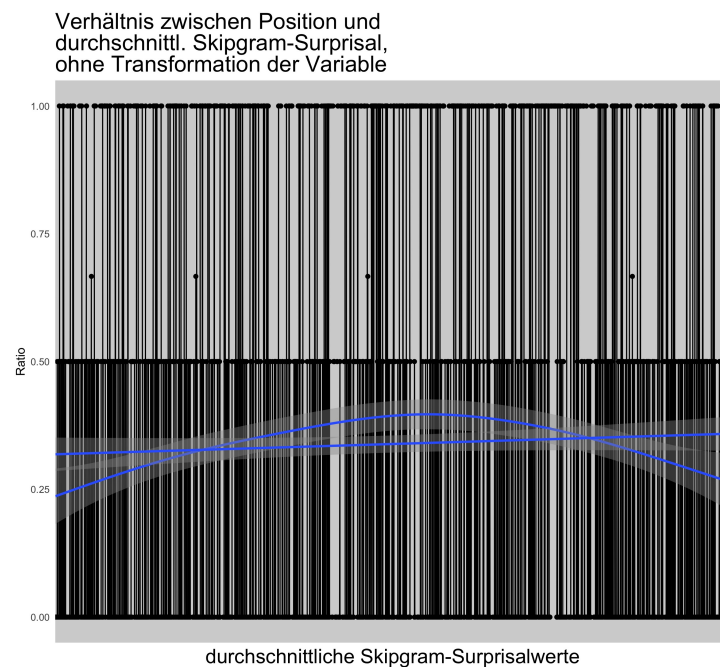


Abbildung 6.14.: Verhältnis zwischen der Position und den durchschnittlichen Skipgram-Surprisalwerten der PPs.

6.2.2. Logistische Regression der Präpositionalphrasen im Gesamtkorpus

Mithilfe der logistischen Regression soll ermittelt werden, welche der Variablen (durchschnittliches Skipgram-Surprisal, die logarithmierte Länge, Orality Score, Genre und Zeitstufe und den Zweifachinteraktionen) die Position einer Präpositionalphrase am besten erklären. Das Genre wurde summenkodiert (medizinische Texte: $-0,5$, theologische Texte: $0,5$), die Zeitstufe wurde in numerische Werte umgewandelt (vgl. Kapitel 5.4). Wie in Kapitel 5.4 beschrieben, wurde eine rückwärtsgerichtete Modellauswahl mittels eines *likelihood ratio*-Tests mit der *anova*-Funktion (vgl. R Core Team 2022) durchgeführt.

Das finale Modell (Tabelle 6.15) zeigt, dass die Hypothese H1 für die Präpositionalphrasen im Gesamtkorpus nicht bestätigt werden kann.

Die logarithmierte Länge der Präpositionalphrase erklärt die meisten Daten ($z=-15,23$, $p<0,001$). Je länger eine Phrase ist, desto wahrscheinlicher wird sie extraponiert. Der Effekt ist auch über alle Orality Scores hinweg zu sehen. Während lange Präpositionalphrasen allerdings immer häufiger extraponiert werden, werden mehr kurze Präpositionalphrasen in mündlichkeitsnahen Texten extraponiert, wie die Interaktion aus der Länge und dem Orality Score zeigt ($z=2,60$, $p<0,05$). Damit liefert die hiesige Untersuchung des Gesamtkorpus weitere Belege für die Nachfeldstellung langer Konstituenten (vgl. z.B. Behaghel 1932; Weber 2019). Die Informativität der Phrase hat keinen Einfluss auf die Nachfeldstellung.

Die Regressionsanalyse zeigt, dass die Zeitstufe mit Einbettung zusammenhängt ($z=2,03$, $p<0,05$). In jüngeren Texten wird die *in situ*-Stellung einer Präpositionalphrase wahrscheinlicher. Allerdings ist die Zeitstufe auch an der Interaktion mit der Mündlichkeitsnähe beteiligt ($z=-4,25$, $p<0,001$). In den mündlichkeitsnahen Texten ist es in jüngeren Texten wahrscheinlicher, ausgelagerte Präpositionalphrasen zu finden. Für schriftlichkeitsnahe Texte ergibt sich das genau gegenteilige Bild. Hier werden in den älteren Texten mehr Phrasen ausgelagert und in den jüngeren Texten mehr Präpositionalphrasen eingebettet. Allein ist Mündlichkeitsnähe nicht signifikant ($z=0,16$, $p=0,86$).

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	1,49	0,23	6,61	<0,001	***
Länge _{log}	-2,63	0,17	-15,23	<0,001	***
Orality Score	0,08	0,51	0,16	0,87	
Genre	-0,16	0,20	-0,79	0,43	
Zeitstufe	0,13	0,06	2,03	0,04	*
Länge _{log} :Orality Score	0,79	0,30	2,60	0,009	**
Orality Score:Genre	2,02	0,41	4,98	<0,001	***
Orality Score:Zeitstufe	-0,51	0,12	-4,25	<0,001	***

Tabelle 6.15.: Logistische Regression der Präpositionalphrasen im Gesamtkorpus, finales Modell. Negative z-Werte sprechen für Extraposition.

Die Präferenz für Auslagerung aus mündlichkeitsnahen Texten heraus erstreckt sich jedoch auch nicht auf Texte, die dem theologischen Genre angehören. In mündlichkeitsnäheren, theologischen Texten wird Einbettung wahrscheinlicher, wie die Interaktion zwischen dem Orality Score und dem Genre zeigt ($z=4,98$, $p<0,001$). In medizinischen Texten findet sich im mündlichkeitsnahen Register hingegen vermehrt Extraposition. Als eigenständiger Prädiktor übt folglich auch das Genre ($z=-0,792$, $p=0,43$) keinen Einfluss auf die Auslagerung aus.¹⁰

Im Folgenden werden die Teilkorpora wie bei der Betrachtung der Nominalphrasen betrachtet. Den Anfang macht die Analyse nach Genre. Danach wird überprüft, ob die attributiven Präpositionalphrasen andere Ergebnisse erzielen als alle Phrasen. Von einer separaten Analyse der eigenständigen Präpositionalphrasen wird abgesehen, da es aufgrund der geringen Zahl attributiver Phrasen unwahrscheinlich ist, dass diese das Gesamtergebnis maßgeblich beeinflussen.¹¹

¹⁰ Anhang E.1.2 referiert in Analogie zum Vorgehen bei den Nominalphrasen auf eine logistische Regression, die nur mit Präpositionalphrasen der Länge 6 durchgeführt wurde. Als Hauptprädiktor kann Informationsdichte die Position von Präpositionalphrasen auch in diesem Korpus aus 138 Präpositionalphrasen nicht erklären. Allerdings weist die Interaktion aus Genre und durchschnittlichen Skipgramwerten auf einen Unterschied zwischen den Teilkorpora hin.

¹¹ Eine separate Analyse der eigenständigen Präpositionalphrasen, allerdings ohne den Faktor der Mündlichkeitsnähe, wurde in Voigtmann und Speyer (2023) durchgeführt und unterstützt die Vermutung, dass sich das Ergebnis nicht entschieden von dem aller Präpositionalphrasen unterscheidet.

6.2.3. Präpositionalphrasen in den medizinischen Texten

Insgesamt 1214 der annotierten Präpositionalphrasen stammen aus medizinischen Texten. 641 (52,8 %) davon sind extraponiert und 573 (47,2 %) eingebettet. In die Analyse fließen neben dem Orality Score, der Zeitstufe auch die Werte der durchschnittlichen Skipgram-Surprisalwerte und der Länge ein. Deren deskriptive Werte werden in Tabelle 6.16 dargestellt.

	Länge		Skipgram _{mean}	
Minimum	2	(2 2)	2,798	(2,798 3,04)
Median	12	(9 16)	3,63	(3,62 3,64)
Durchschnitt	17,77	(12,91 22,1)	3,65	(3,65 3,66)
drittes Quartil	24	(15 30)	3,78	(3,82 3,78)
Maximum	135	(84 135)	4,396	(4,38 4,396)
Std.-Fehler	0,50	(0,54 0,78)	0,01	(0,01 0,01)
Varianz	304,91	(12,87 19,75)	0,054	(0,26 0,21)

Tabelle 6.16.: Deskriptive Werte der Präpositionalphrase aus medizinischen Texten. Links in der Klammer stehen die Werte der eingebetteten PPs, rechts die der ausgelagerten PPs.

Deskriptiv unterscheiden sich die eingebetteten von den ausgelagerten Phrasen in der Länge und in den durchschnittlichen Skipgramwerten. Im Minimum sind Phrasen in den verschiedenen Positionen gleich lang. Sonst sind die *in situ*-Präpositionalphrasen kürzer als die extraponierten Präpositionalphrasen. Der Unterschied wird mit bloßem Auge deutlich und könnte schon die Vermutung bestärken, dass Länge auch bei der Analyse dieses Teilkorpus einen größeren Einfluss haben wird.

Die Skipgram-Surprisalwerte zeigen hingegen nur marginale Unterschiede. Sowohl im Minimum als auch im Median als auch im arithmetischen Mittel und im Maximum liegen die Werte der eingebetteten Präpositionalphrasen unter denen der ausgelagerten. Im dritten Quartil zeigt sich das umgekehrte Bild. Hier ist der Wert der eingebetteten Präpositionalphrasen etwas höher (3,82) als der der ausgelagerten Präpositionalphrasen (3,78).

Logistische Regression der medizinischen Texte Wie in Kapitel 6.2.1 wurde überprüft, ob die numerischen Variablen zu transformieren sind. Nur die

Länge wurde auf der Basis von 10 logarithmiert.¹² Tabelle 6.17 zeigt das finale Modell und auch, dass keine Belege für Hypothese H1 geliefert werden, weil die Skipgram-Surprisalwerte aus der Analyse ausgeschlossen wurden.

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	1,77	0,23	7,62	<0,001	***
Länge _{log}	-2,99	0,23	-13,05	<0,001	***
Orality Score	-0,32	0,615	-0,53	0,59	
Zeitstufe	0,14	0,11	1,29	0,20	
Orality Score:Zeitstufe	-0,52	0,23	-2,24	0,03	*

Tabelle 6.17.: Logistische Regression der Präpositionalphrasen der medizinischen Texte, finales Modell. Negative z-Werte sprechen für Extrapolation.

Die unabhängige Variable der Länge weist auf Extrapolation hin ($z=-13,05$, $p<0,0001$). Je länger also eine Phrase ist, desto eher wird sie ins Nachfeld gestellt. Des Weiteren zeigt Tabelle 6.17 eine Interaktion zwischen der Mündlichkeitsnähe und der Zeitstufe ($z=-2,24$, $p<0,05$). Mit zunehmender Mündlichkeitsnähe wird die Extrapolation in allen Zeitstufen, vor allem in den jüngeren wahrscheinlicher.

Zwar finden diese Daten keinen Beleg für die Hypothese, dass die Auslagerung stark mit der lexikalischen Erwartbarkeit, berechnet durch Surprisal, korreliert. Trotzdem lassen sich verarbeitungsbezogene Erklärungen für dieses Ergebnis heranzuführen, wie in Kapitel 6.5.2 gezeigt wird. Interessant ist zudem die Interaktion zwischen dem Orality Score und der Zeitstufe, die für eine Veränderung im Verlauf der Jahrhunderte spricht, die die separate Betrachtung der fünf Zeitabschnitte lohnenswert macht (Kapitel 8).

6.2.4. Präpositionalphrasen in den theologischen Texten

Während in den medizinischen Texten über 1000 Präpositionalphrasen gefunden wurden, sind es in den theologischen insgesamt nur 728, die sich in

¹² Die abhängige Variable ist die Position, die unabhängigen Variablen setzen sich aus der logarithmierten Länge, den durchschnittlichen Skipgram-Surprisalwerten, der Zeitstufe und dem Orality Score zusammen. Da hier nur medizinische Texte analysiert werden, entfällt die Variable des Genres.

402 (55,2 %) extrapionierte und 326 (44,8 %) eingebettete Präpositionalphrasen teilen lassen. Tabelle 6.18 zeigt die deskriptiven Werte aller Phrasen aus den theologischen Texten.

	Länge		Skipgram _{mean}	
Minimum	2	(2 2)	2,889	(2,889 2,987)
Median	8	(6 10)	3,646	(3,63 3,657)
Durchschnitt	12,63	(8,4 16,06)	3,657	(3,646 3,666)
drittes Quartil	14	(10 18)	3,779	(3,764 3,786)
Maximum	225	(81 225)	4,416	(4,318 4,416)
Std.-Fehler	0,62	(0,45 1,04)	0,008	(0,01 0,01)
Varianz	283,73	(8,19 20,83)	0.05	(0,23 0,21)

Tabelle 6.18.: Deskriptive Werte der Präpositionalphrase aus theologischen Texten. Links in der Klammer stehen die Werte der eingebetteten PPs, rechts die der ausgelagerten PPs.

Das Bild, das sich auch schon für die medizinischen Texte abgezeichnet hat, setzt sich bei den theologischen Werken fort. Gerade bei der Länge zeigen sich deutliche Unterschiede zwischen den eingebetteten und ausgelagerten Präpositionalphrasen. Letztere weisen immer höhere Werte auf, wobei sie im Durchschnitt sogar doppelt so lang (16,06 Wörter) sind wie ihre *in situ*-Gegenstücke (8,4 Wörter). Die Betrachtung der Skipgram-Surprisalwerte zeigt ebenfalls, dass die ausgelagerten Präpositionalphrasen in den meisten Bereichen höhere Werte erzielen als die eingebetteten. Nur der Median weist einen niedrigeren Wert auf. Extrapionierte Präpositionalphrasen messen 3,657, eingebettete nur 3,63. Generell sind die Unterschiede wieder marginal.

Logistische Regression der theologischen Texte Das Vorgehen der logistischen Regression für die theologischen Texte entspricht dem der medizinischen. Die numerischen Variablen mussten nicht transformiert werden.¹³ Das finale Modell in Tabelle 6.19 unterstützt das Bild, das auch die Betrachtung der Präpositionalphrasen im Gesamt- und im medizinischen Teilkorpus gezeigt hat. Der Einfluss der durchschnittlichen Skipgramwerte auf die Position der betrachteten Phrasen beruht auf Zufall.

¹³ In das Modell fließen also die unabhängigen Variablen durchschnittliches Skipgram, Orality Score, Länge und Zeitstufe ein.

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	0,88	0,34	2,57	0,01	*
Länge	-0,11	0,042217	-2,55	0,01	*
Orality Score	-0,78	0,48	-1,63	0,10	
Zeitstufe	0,13	0,09	1,464	0,14	
Länge:Orality Score	0,12	0,07	1,75	0,08	.
Länge:Zeitstufe	-0,03	0,01	-3,56	0,0004	***

Tabelle 6.19.: Logistische Regression der Präpositionalphrasen der theologischen Texte, finales Modell. Negative z-Werte sprechen für Extrapolation.

Der Haupteffekt, der signifikant ist, ist die Länge ($z=-2,55$, $p<0,05$). Lange Präpositionalphrasen werden auch in theologischen Texten ausgelagert. In den theologischen Texten besteht zudem eine Interaktion zwischen der Länge und der Zeitstufe, die bei den medizinischen Texten komplett aus der Analyse ausgeschlossen wurde. Für die theologischen Texte bedeutet das, dass lange Phrasen in jüngeren Texten mit einer sehr hohen Wahrscheinlichkeit ins Nachfeld gestellt werden ($z=-3,56$, $p<0,001$), während kurze Präpositionalphrasen hingegen insgesamt eher eingebettet werden. Das lässt sich dahingehend interpretieren, dass sie für die Bewegung in jüngeren Texten unabhängig von deren Mündlichkeitsnähe auch den Ansporn durch die Länge der Phrasen benötigt, was wiederum die separate Betrachtung der Zeitstufen relevant werden lässt (Kapitel 8). Ebenfalls signifikant ist in den theologischen Texten die Interaktion zwischen Länge und Orality Score, die jedoch für Einbettung spricht ($z=1,75$, $p<0,1$). Wenn ein Text mündlichkeitsnäher wird und längere Phrasen beinhaltet, werden diese eher eingebettet. Kürzere Phrasen hingegen werden auch in mündlichkeitsnahen Texten eher eingebettet. Über alle Orality Scores hinweg werden aber lange Phrasen mit höherer Wahrscheinlichkeit ausgelagert.

6.2.5. Extraposition von attributiven Präpositionalphrasen

Im Korpus wurden nur 21 attributive, extraponierte Präpositionalphrasen gefunden.¹⁴ Ihnen stehen 16 eingebettete attributive Präpositionalphrasen mit passenden Präpositionen gegenüber, was insgesamt 37 Phrasen ergibt, die hier analysiert werden. Das erschwert es, die Ergebnisse zu verallgemeinern. Aufgrund der geringen Menge an zur Verfügung stehenden attributiven, extraponierten Präpositionalphrasen sind zuverlässige Aussagen an dieser Stelle allerdings kaum möglich.

Im Zeitabschnitt 1750 bis 1800 haben die Autoren keine attributiven Präpositionalphrasen ins Nachfeld gestellt. Von 1650 bis 1700 und von 1700 bis 1750 wurden je drei extraponierte Präpositionalphrasen gefunden. Diese Zahl verdoppelt sich quasi von 1800 bis 1850, wo sieben Präpositionalphrasen annotiert wurden, und erhöht sich im Abschnitt 1850 bis 1900 auf acht. Die Zahl der ausgelagerten attributiven Phrasen nimmt also auf niedrigem Niveau zu (Grafik 6.7). Unterteilt man das Teilkorpus nach Genre, wurden in medizinischen Texten 12 und in theologischen neun attributive, extraponierte Präpositionalphrasen gefunden.

Die mittleren Skipgramwerte haben einen Mittelwert von 3,59, einen Median von 3,6, einen Standardfehler von 0,029 und eine Standardabweichung von 0,17. Der Mittelwert der Länge ist 8,08, der Median ist 6, der Standardfehler 0,88 und die Standardabweichung 5,38. Obwohl die Verteilung der durchschnittlichen Surprisalwerte je nach Position der Präpositionalphrasen in Abbildung 6.15 weitestgehend klar erschien und nur im Zeitabschnitt 1700 bis 1750 die Skipgramwerte eingebetteten Phrasen über den ausgelagerten liegen, wird dennoch erneut auf die Untersuchung mittels einer logistischen Regression zurückgegriffen, bei der die Länge der Präpositionalphrasen und die Surprisalwerte logarithmiert wurden.

Aufgrund der geringen Anzahl von attributiven Phrasen und den hohen Standardfehlern muss das Modell (Tabelle 6.20) mit Vorsicht interpretiert werden. Wie schon bei den attributiven Nominalphrasen können die Skipgram-

¹⁴ Teile des in diesem Kapitel vorgestellten Materials flossen auch in die Veröffentlichung Voigtmann und Speyer (2023) ein.

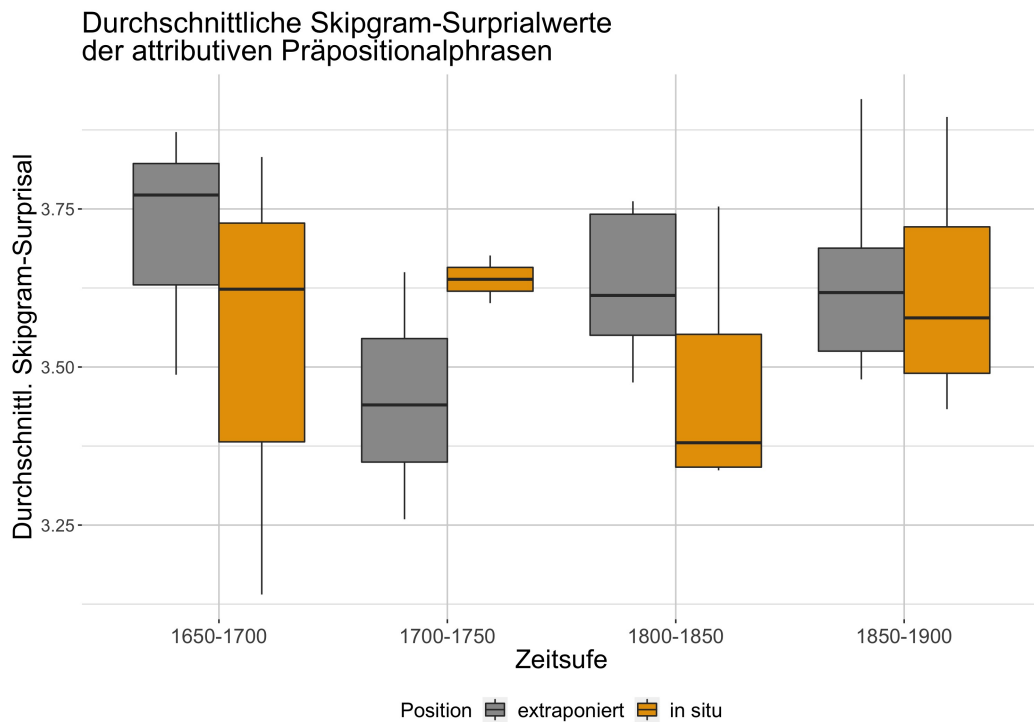


Abbildung 6.15.: Durchschnittliche Skipgram-Surprisalwerte nach Jahren, nur attributive Präpositionalphrasen.

Surprisalwerte nicht ausgeschlossen werden und liefern ein signifikantes Ergebnis für Auslagerung ($z=-1,95$, $p<0,1$). Zwar erklären sowohl die Länge ($z=2,12$, $p<0,05$) als auch die Zeitstufe mehr Datenpunkte ($z=-2,24$, $p<0,05$), dennoch kann die Nullhypothese durch den signifikanten Wert der Skipgram-Surprisalwerte widerlegt werden.

Aus dem Modell lassen sich des Weiteren folgende Schlüsse ziehen: Attributive Phrasen, die entweder lang oder informativ sind, werden mit höherer Wahrscheinlichkeit im Nachfeld platziert. Das gleiche gilt für Phrasen aus jüngeren Texten ($z=-2,24$, $p<0,05$) und für Phrasen aus mündlichkeitsnahen Texten ($z=-2,24$, $p<0,05$). Die Zugehörigkeit zum theologischen Genre erklärt Einbettung ($z=2,02$, $p<0,05$). Damit finden sich in auch bei bei den attributiven Präpositionalphrasen ähnliche Ergebnisse, die auch von der bisherigen Forschung beschrieben wurde (Kapitel 3.1). Wieder scheint eine große Länge und Informativität der Phrasen tendenziell für Einbettung zu sprechen ($z=2,09$, $p<0,05$), das gleiche gilt für Mündlichkeitsnähe und Informativität ($z=2,23$, $p<0,05$). Auch lange Phrasen in jüngeren Texten sprechen eher für

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	473,90	235,04	2,02	0,04	*
Skipgram _{mean}	-746,58	382,82	-1,95	0,05	.
Länge _{log}	-644,15	303,33	-2,12	0,03	*
Genre	47,35	23,44	2,02	0,04	*
Zeitstufe	-9,89	4,42	-2,24	0,03	*
Orality Score	-389,55	173,86	-2,24	0,03	*
Skipgram _{mean} :Länge _{log}	1050,927	503,79	2,09	0,04	*
Skipgram _{mean} :Orality Score	681,36	305,80	2,23	0,03	*
Länge _{log} :Genre	-24,348	15,23	-1,6	0,11	
Länge _{log} :Zeitstufe	8,72	4,25	2,05	0,04	*
Genre:Zeitstufe	-4,98	2,33	-2,14	0,03	*

Tabelle 6.20.: Logistische Regression der attributiven Präpositionalphrasen im Gesamtkorpus, finales Modell. Negative z-Werte sprechen für Extraposition.

Einbettung ($z=2,05$, $p<0,05$), wohingegen in jüngeren, theologischen Texten Phrasen wahrscheinlicher ausgelagert werden ($z=-2,14$, $p<0,05$).

Abschließend ist erneut zu betonen, dass es sich hier um eine Betrachtung von nur 37 Phrasen handelt. Um eindeutiger Aussagen treffen zu können, müsste das Gesamtkorpus erweitert werden, um mehr Beispiele für ausgelagerte, attributive Präpositionalphrasen zu finden. Das würde jedoch den Rahmen dieser Arbeit übersteigen und ist aufgrund der wenigen Korpora zu wissenschaftlichen Texten auch nicht ohne Weiteres möglich.

6.3. Relativsätze

In diesem Kapitel soll nun die letzte Gruppe von Extrapositionen betrachtet werden, die Relativsätze. Die von anderen Studien untersuchten Gründe für die Relativsatzextraposition (Kapitel 3.2) decken sich mit denen für die Nachfeldstellung von Nominal- und Präpositionalphrasen. Als zusätzliche Faktoren kommen außerdem die Restriktivität und die Distanz zwischen dem Antezedens und dem Nachfeld hinzu. Die Betrachtung der Distanz verursacht allerdings ein Problem: In Kapitel 4.3 wurde beschrieben, dass die Annotation der Distanz automatisch erfolgt und dass diese Werte manuell überprüft und,

wo nötig, korrigiert wurden. Bei dieser Überprüfung wurde allerdings bei einer Vielzahl der eingebetteten Relativsätze klar, dass diese gar nicht einfach ins Nachfeld ihres Matrixsatzes verschoben werden könnten, weil dadurch wie in (2a) Ambiguitäten in der Zuordnung des Antezedens entstünden oder wie in (2b) das Nachfeld schon von einem Nebensatz besetzt wird, was wiederum dessen Zuordnung zu seinem Matrixsatz erschweren könnte.

- (2) a. Blicke herab auf die Vorsätze, **die ich halten soll**, die Schwachheiten und Verkehrtheiten, **die nur anhaften**, die Versuchungen, **die mich bedrohen**. (Egger 1895: Satz 693)
- b. [...] sondern das vielfältige Sitzen und die dahero unterbleibende zur Gesundheit so nöthige Bewegung/ samt denen stetigen Lucubrationibus, **wodurch die LebensGeister geschwächet/ das humidum radicale vermindert und das calidum innatum verzehret wird/** seynd solche unvermeidliche Hindernüsse des gesunden Lebens/ [*daß meines Erachtens/ niemand mehr Ursache hat/ den zubrechlichen Leib vor den unzehlichen Arten der Kranckheiten zu verwahren/ als die Gelehrten worvon Herrn D. G. Franckens/ Königl. Dänischen LeibMedici gelehrte Oration de studiorum noxa ein unverwerfliches Zeugniuß ablegen kan*]_{NF}. (Abel 1699: Satz 4)

Die Wörter, die die beiden Relativsätze bei Extraposition hypothetisch überbrückt hätten, wurden dennoch gezählt und auch die in Analyse in Kapitel 6.3.5 aufgenommen, allerdings wird diese Variable aus den anderen Analysen ausgeschlossen, weil die Relativsätze in zu vielen Fällen aus oben genannten Gründen nicht ins Nachfeld hätten gestellt werden können, ohne andere Veränderungen im Satz notwendig zu machen, um dessen Grammatikalität beizubehalten.¹⁵

6.3.1. Relativsätze im gesamten Korpus

Insgesamt wurden 5332 Relativsätze zur Analyse verwendet. Von ihnen stehen 2626 (49,2 %) adjazent zu ihrem Antezedens und 2706 (50,3 %) im Nach-

¹⁵ Auch der Text von Ludwig wurde aus der Analyse der Relativsätze ausgeschlossen.

feld. Damit ist ihr Verhältnis weitestgehend ausgeglichen. Auf die Unterscheidung zwischen den beiden Genres trifft das hingegen nicht zu. 2871 (53,8 %) Relativsätze wurden in medizinischen Texten gefunden – davon sind 1547 (rund 53 %) extraponiert – 2461 (46,2 %) Sätze stammen aus dem theologischen Teilkorpus. Von ihnen sind 1159 (47,1 %) im Nachfeld positioniert.

Zeitabschnitt	Gesamtzahl RelS (extrap.)	Anteil Extraposition an Sätzen in Prozent, gerundet	Medizinisch (extrap.)	Theologisch (extrap.)
1650 - 1700	769 (408)	9,9	369 (202)	400 (206)
1700 - 1750	1435 (665)	11,3	659 (358)	776 (307)
1750 - 1800	1038 (425)	17,5	354 (125)	684 (300)
1800 - 1850	1175 (677)	16,1	988 (575)	187 (102)
1850 - 1900	915 (531)	8,2	501 (287)	414 (244)
gesamt	5332 (2706)	10	2871 (1547)	2461 (1159)

Tabelle 6.21.: Anzahl der Relativsätze nach Zeitabschnitt und Genre.

Weniger ausgeglichen ist die Verteilung der Relativsätze auf die verschiedenen Zeitabschnitte. Im Zeitabschnitt 1650 bis 1700 stehen 769 Relativsätze (14,4 %, aller ausgelagerten Relativsätze im Korpus), im Abschnitt von 1700 bis 1750 finden sich 1435 Relativsätze (26,9 %), 1750 bis 1800 1038 Sätze (19,5 %), von 1800 bis 1850 wurden 1175 Relativsätze annotiert (22 %) und im letzten Abschnitt schließlich 915 (17,2 %). Ihr prozentualer Anteil an allen Sätzen steigt im ersten Zeitabschnitt (1650 bis 1700) von rund 10 % auf bis 17 % im Zeitabschnitt 1750 bis 1800 an, bevor er wieder abnimmt und im letzten Zeitabschnitt (1850-1900) nur noch rund 8 % beträgt (siehe Tabelle 6.21, Spalte 3). Dabei ist zu beobachten, dass der Anteil von adjazenten und ausgelagerten Relativsätzen in jedem Zeitabschnitt zwischen 40 und 60 % schwankt. Sichtbar werden die Unterschiede, allerdings dargestellt in absoluten Zahlen, auch in Grafik 6.16.

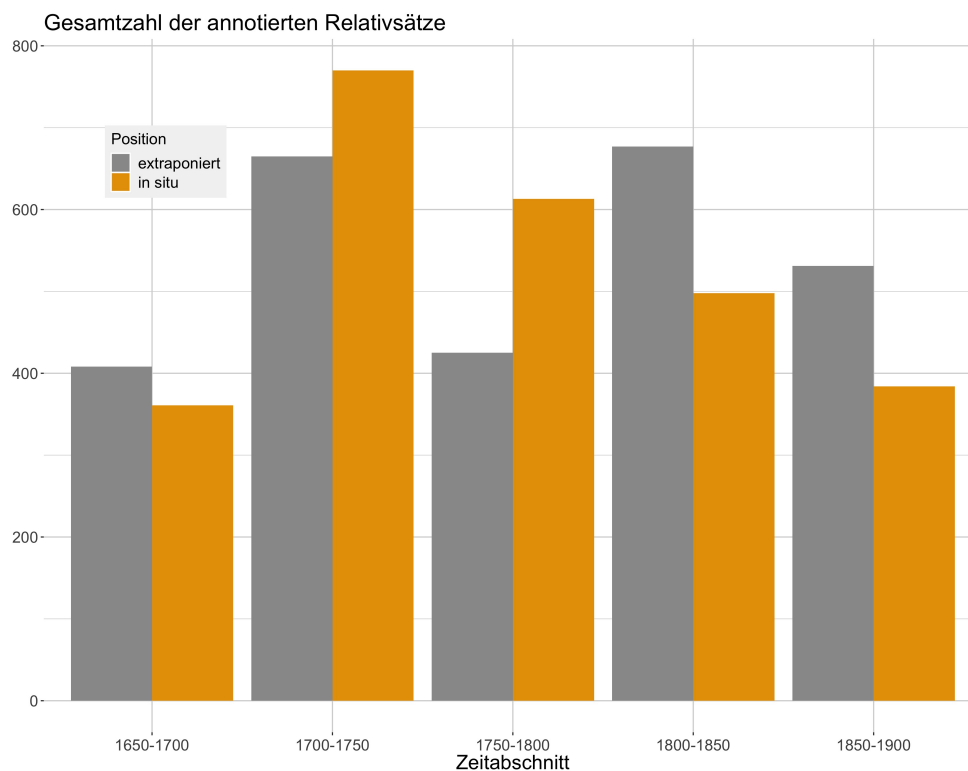


Abbildung 6.16.: Anzahl der Relativsätze pro Zeitabschnitt in absoluten Zahlen.

Tabelle 6.21 zeigt, dass mehr Relativsätze in den medizinischen als in den theologischen Texten gefunden wurde und das durch die Relativsätze aus dem 19. Jahrhundert erklärt werden kann. Von 1800 bis 1850 wurden fast neun mal so viele Relativsätze in medizinischen Texten ($n=988$) wie in theologischen ($n=187$) gefunden. Die vierte und fünfte Spalte von Tabelle 6.21 präsentieren in allen vorherigen Zeitabschnitten mehr Relativsätze in den theologischen Texten. Im Abschnitt 1750 bis 1800 sogar fast doppelt so viele in den theologischen ($n=684$) wie in den medizinischen ($n=354$).

Der Typ des Relativsatzes, also die Unterscheidung zwischen restriktiven und appositiven Relativsätzen, wurde nicht für jeden Relativsatz bestimmt (Kapitel 4.3). Daraus ergibt sich, dass der Typ von 1374 Relativsätzen im Gesamtkorpus nicht gegeben ist. Es werden also 1632 appositive und 2326 restriktive Relativsätze in den logistischen Regressionen analysiert. Von den eindeutig bestimmbar Relativsätzen überwiegt also der Anteil an restriktiven Relativsätzen. 50,6 % der restriktiven Relativsätze sind extrapositioniert ($n=1178$).

Bei den appositiven Relativsätzen sieht es fast genauso aus; von ihnen stehen 51,4 % im Nachfeld (n=839). Von den unbestimmten Relativsätzen entfallen 689 auf extrapolierte Relativsätze, was ebenfalls rund 50 % entspricht.

Im nachfolgenden Abschnitt werden die Variablen des durchschnittlichen Skipgram-Surprisal und der Länge für alle Phrasen betrachtet. Die Werte in den Tabellen entsprechen den Werten aller Relativsätze. Das inkludiert die Relativsätze, deren Typ unbestimmt und mit NA gekennzeichnet ist.

In Testung einbezogene Variablen

Länge der Relativsätze Wenn man sich mit der Position von Relativsätzen befasst, wird als Grund für deren Nachfeldstellung immer wieder die Länge angeführt (vgl. z.B. Uszkoreit et al. 1998). Daher soll auch hier auf sie eingegangen und untersucht werden, ob für die logistische Regression eine Transformation der Variable notwendig ist.

Tabelle 6.22 untermauert den Befund früherer Arbeiten: Die ausgelagerten Relativsätze sind in allen Werten länger als die adjazenten. Während die adjazenten Relativsätze im Schnitt 10,9 Wörter lang sind, beträgt dieser Wert bei den extrapolierten Relativsätzen 14,9 Wörter. Auch das dritte Quartil weist bei den ausgelagerten Sätzen einen höheren Wert auf, das heißt, 75 % aller Relativsätze haben weniger als 18 Wörter. Bei den eingebetteten Relativsätzen liegt der Wert bei nur 12 Wörter. Bemerkenswert ist, dass besonders lange Relativsätze eine Ausnahme zu bilden scheinen, obwohl gerade das 17. und 18. Jahrhundert für lange Sätze bekannt sind (vgl. z.B. Admoni 1990; Polenz et al. 2013). Die separate Analyse der einzelnen Zeitabschnitte wird Aufschluss darüber geben, ob es markante Unterschiede in der Länge der Relativsätze zwischen den einzelnen Zeitabschnitten gibt (Kapitel 8).

Auch Grafik 6.17 zeigt insgesamt ein recht eindeutiges Bild hinsichtlich der Längenunterschiede zwischen den eingebetteten und ausgelagerten Relativsätzen. Auf der y-Achse ist die Länge in Tokens abgebildet, auf der x-Achse sind die Relativsätze nach Zeitabschnitt aufgegliedert. Im Boxplot ist zu sehen, dass die ausgelagerten Relativsätze in allen Zeitstufen länger sind als die adjazenten Relativsätze. Auch die Ausreißer nach oben sind bei den extrapolierten

	eingebettete RelS	Ausgelagerte RelS
Minimum	1	2
Median	8	11
Durchschnitt	10,9	14,9
drittes Quartil	12	18
Maximum	147	150
Std.-Fehler	0,12	0,29
Varianz	146,33	295,11

Tabelle 6.22.: Länge der Relativsätze unterteilt in eingebettete und ausgelagerte Sätze.

nierten Sätzen häufiger und stärker. Schon diese großen Unterschiede lassen vermuten, dass eine Transformation der Werte notwendig ist. Tatsächlich ist das für die Länge der Fall. Ihre Werte werden für die logistische Regression auf der Basis von 10 logarithmiert.

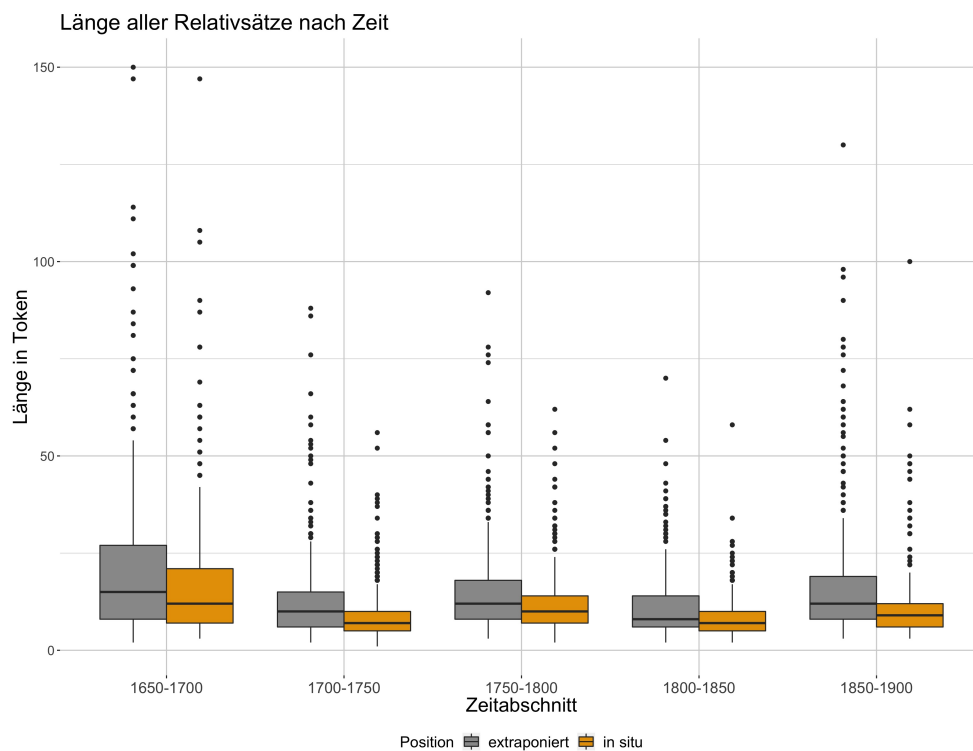


Abbildung 6.17.: Länge der eingebetteten und ausgelagerten Relativsätze nach Zeitstufen.

Durchschnittliches Skipgram-Surprisal der Relativsätze im Gesamtkorpus

Tabelle 6.23 gibt Aufschluss über die Unterschiede in den Skipgramwerten der adjazenten und extraponierten Relativsätze. Die Differenz zwischen den beiden Positionsausprägungen ist minimal. Die adjazenten Relativsätze haben ein durchschnittliches Skipgram-Surprisal von 3,645 und die ausgelagerten ein von 3,642. Der Unterschied wird also erst an der dritten Nachkommastelle sichtbar und der Wert ist für die ausgelagerten Relativsätze geringer als der bei den adjazenten. Der Median beider Positionen ist sogar gleich (3,637). Der Standardfehler (0,003) und die Varianz (0,03) zeigen ebenfalls eine geringe Streuung und bis zur vierten Nachkommastelle keinen Unterschied zwischen den eingebetteten und ausgelagerten Relativsätzen, sodass sie in der hier verwendeten Rundung sogar gleich erscheinen.

Auch Grafik 6.18 illustriert die marginalen Unterschiede. Auf deren y-Achse sind die durchschnittlichen Skipgramwerte der Relativsätze abgebildet, auf der x-Achse wird die Position der Relativsätze nach Zeitabschnitten unterschieden. Die in gelb dargestellten, eingebetteten Sätze weisen leicht höhere Werte auf als die grauen, extraponierten. Ob eine Transformation der Variablen nötig ist, wird durch die Überprüfung anhand der Kriterien, wie auch schon in den vorherigen Kapiteln angewendet wurden, bestätigt: Die durchschnittlichen Skipgram-Surprisalwerte werden nicht logarithmiert.

Auch besteht keine Korrelation zwischen der Länge und den durchschnittlichen Skipgram-Surprisalwerten nicht miteinander ($r=0,055$).

	eingebettete RelS	Ausgelagerte RelS
Minimum	2,909	3,215
Median	3,637	3,637
Durchschnitt	3,645	3,642
drittes Quartil	3,645	3,748
Maximum	4,318	4,359
Std.-Fehler	0,003	0,003
Varianz	0,03	0,03

Tabelle 6.23.: Durchschnittliche Skipgram-Surprisalwerte der Relativsätze unterteilt in eingebettete und ausgelagerte Relativsätze.

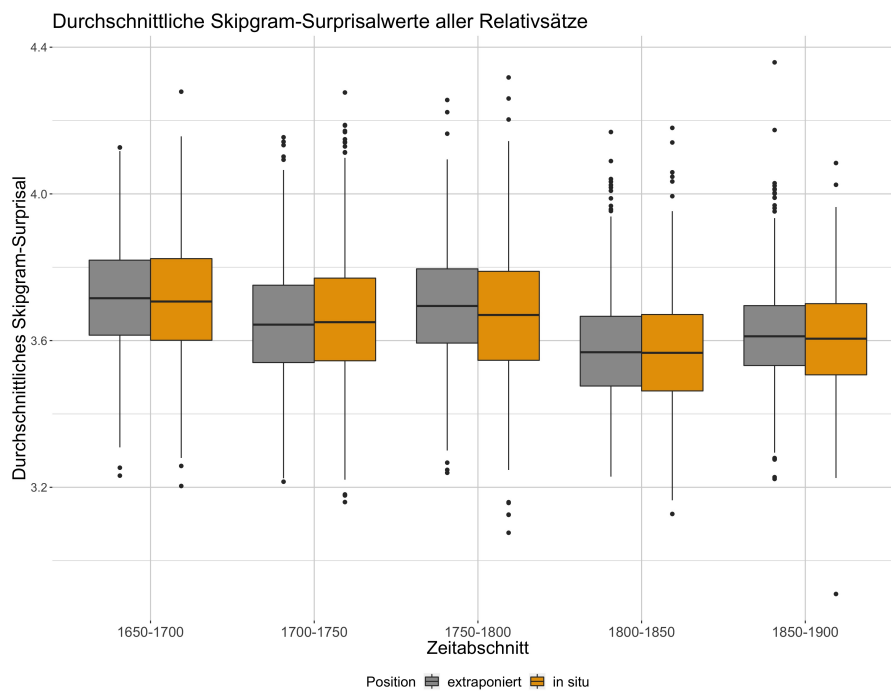


Abbildung 6.18.: Skipgram-Surprisalwerte der eingebetteten und ausgelagerten Relativsätze nach Zeitstufen.

6.3.2. Logistische Regression der Relativsätze im Gesamtkorpus

Wie bei den Nominal- und Präpositionalphrasen wird logistische Regression (glm, R Core Team 2018) genutzt, um herauszufinden, welche der Variablen den größten Einfluss auf die Nachfeldstellung von Relativsätzen ausübt. Als abhängige Variable gilt die Position, zu den unabhängigen Variablen gehören die logarithmierte Länge der Relativsätze, die durchschnittlichen Skipgram-Surprisalwerte, der Orality Score der einzelnen Texte, das Genre der Texte, der Typ der Relativsätze und die Zeitstufe. Sowohl das Genre (medizinische Texte: $-0,5$, theologische Texte: $0,5$) als auch der Relativsatztyp (restriktiv: $-0,5$, appositiv: $0,5$) werden summenkodiert, die Zeitstufe wurde in numerische Werte umgewandelt (Kapitel 5.4). Zwischen den Variablen wurden die Zweigeinteraktionen berechnet. Insgesamt wurden die Variablen und Interaktionen, die keinen Einfluss haben, durch eine rückwärtsgerichtete Modellauswahl aussortiert. Im finalen Modell, dargestellt in Tabelle 6.24, lässt sich erkennen, dass die Einflussfaktoren auf Extraposition diffus sind.

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	4,14	0,91	4,55	<0,001	***
Skipgram _{mean}	-0,34	0,21	-1,67	0,095	.
Länge _{log}	-2,90	0,42	-6,85	<0,001	***
Orality Score	-0,59	0,54	-1,18	0,27	
Genre	-1,26	0,45	-2,83	0,005	**
Zeitstufe	-0,33	0,13	-2,57	0,01	*
Typ	-0,003	0,07	-0,05	0,96	
Länge _{log} :Orality Score	1,3	0,45	2,90	0,004	**
Länge _{log} :Genre	0,53	0,32	1,66	0,09	.
Länge _{log} :Zeitstufe	0,22	0,12	1,83	0,07	.
Orality Score:Genre	2,49	0,27	9,18	<0,001	***
Orality Score:Zeitstufe	-0,35	0,11	-3,18	0,001	**
Genre:Zeitstufe	0,31	0,09	3,38	0,0007	***
Genre:Typ	0,43	0,14	3,11	0,002	**

Tabelle 6.24.: Logistische Regression aller Relativsätze im Gesamtkorpus, finales Modell. Negative z-Werte sprechen für Extraposition.

Für sich genommen sind nur die Mündlichkeitsnähe ($z=-1,098$, $p=0,27$) und der Relativsatztyp ($z=-0,045$, $p=0,94$) nicht relevant für die Position des Relativsatzes. Restriktivität macht Extraposition nicht wahrscheinlicher. Auch eine höhere Mündlichkeitsnähe liefert keinen Hinweis auf Extraposition.

Weitere Kriterien, die Einfluss auf die Position der Relativsätze ausüben, sind die durchschnittlichen Skipgram-Surprisalwerte ($z=-1,67$, $p<0,1$). Obwohl die Skipgram-Surprisalwerte nicht den größten Einfluss auf die Position von Relativsätzen haben, musste diese Variable nicht aus dem Modell ausgeschlossen werden, wie es bei den Präpositionalphrasen der Fall war (vgl. Tabelle 6.15). Die Hypothese H1 ist damit bestätigt: Relativsätze mit hohen Skipgram-Surprisalwerten werden tendenziell ausgelagert.

Die Zeitstufe indiziert Auslagerung ($z=-2,57$, $p<0,05$). Je jünger ein Text ist, desto wahrscheinlicher wird die Auslagerung von Relativsätzen. Dafür spricht auch das Genre ($z=-2,83$, $p<0,01$), wodurch beschrieben wird, dass in theologischen Texten Auslagerung wahrscheinlicher ist. Unter den eigenständigen Variablen ist die Länge erneut der signifikanteste Prädiktor ($z=-6,85$, $p<0,001$) für Extraposition. Konstituenten werden bei größerer Länge ins Nachfeld gesetzt.

Die Interaktion aus Länge und Genre ($z=1,66$, $p<0,1$) zeigt, dass lange Relativsätze in beiden Genres eher ausgelagert werden, wobei in theologischen Texten Adjazenz wahrscheinlicher ist. Auch in mündlichkeitsnahen Texten werden lange Relativsätze eher ausgelagert, vergleicht man sie mit schriftlichkeitsnahen Texten. Das besagt die Interaktion aus Orality Score und Länge ($z=2,9$, $p>0,01$). Interessant ist hier allerdings, dass sehr lange Relativsätze in den mündlichkeitsnächsten Texten etwas häufiger eingebettet werden. Einen ähnlichen Trend offenbart die Interaktion zwischen der logarithmierten Länge und der Zeitstufe ($z=1,83$, $p<0,1$). Am äußersten Rand der Skala, sprich bei sehr langen Relativsätzen, zeigt sich allerdings, dass diese in den Texten des 19. Jahrhunderts etwas häufiger eingebettet wurden als in den Texten des 17. Jahrhunderts.

Die Interaktion zwischen Orality Score und Genre ($z=9,18$, $p<0,001$) lässt folgende Schlüsse über die Daten zu: Relativsätze werden in mündlichkeitsnahen, theologischen Texten mit höherer Wahrscheinlichkeit eingebettet. Extraposition nimmt hingegen in jüngeren, mündlichkeitsnahen Texten zu, wie an der die Interaktion aus dem Orality Score und der Zeitstufe zeigt ($z=-3,18$, $p<0,001$) zu erkennen ist.

Theologische Texte aus jüngerer Zeit beinhalten eher adjazente Relativsätze, wie die Interaktion aus Genre und Zeitstufe ($z=3,38$, $p<0,001$) zeigt. Die letzte Interaktion, die zwischen Genre und der Restriktivität eines Relativsatzes ($z=3,11$, $p<0,01$), erlaubt die Aussage, dass Einbettung von Relativsätzen generell häufiger in theologischen Texten zu finden ist. Innerhalb der theologischen Texte werden restriktive Relativsätze eher adjazent platziert und appositive Relativsätze eher extraponiert. Im medizinischen Genre sieht das Bild genau umgekehrt aus: Restriktive Relativsätze werden hier eher extraponiert, appositive stehen mit höherer Wahrscheinlichkeit adjazent.

Um die Ergebnisse der Regression kurz zusammenzufassen: Für Auslagerung sprechen tendenziell die durchschnittlichen Skipgram-Surprisalwerte, wobei deren Einfluss eindeutig von dem der Länge überstiegen wird. Auch jüngere Texte machen es wahrscheinlicher, einen extraponierten Relativsatz zu finden. Außerdem werden Relativsätze dann eher ausgelagert, wenn sie

in mündlichkeitsnahen und gleichzeitig medizinischen Texten oder in schriftlichkeitsnahen, gleichzeitig theologischen Texten auftreten. Außerdem wird die Nachfeldstellung eines restriktiven Relativsatzes in medizinischen Texten wahrscheinlicher als im theologischen Genre. Hervorzuheben ist hier aber vor allem ein deutlich sichtbarer Einfluss der Zeit und der Relativsatzlänge.

Bedingt durch die großen Unterschiede zwischen den Genre wird in den nächsten beiden Abschnitten auf die Teilkorpora der beiden Genres eingegangen.

6.3.3. Relativsätze in den medizinischen Texten

Im medizinischen Subkorpora wurden 1547 (53 %) extraponierte und 1324 (47 %) adjazente Relativsätze gefunden. Einen größeren Unterschied gibt es bei der Restriktivität der Relativsätze: 745 von ihnen sind appositiv und fast doppelt so viele, nämlich 1403, sind restriktiv. Es bleiben 723 Relativsätze, deren Typ unbestimmt geblieben ist.

Die Verteilung auf die einzelnen Zeitabschnitt kann als ungleichmäßig beschrieben werden. Von 1650 bis 1700 wurden 369 Relativsätze annotiert, im nächsten Zeitabschnitt verdoppelt sich die Anzahl ($n=659$), bevor sie von 1750 bis 1800 wieder auf 354 fällt. In der ersten Hälfte des 19. Jahrhunderts verdreifacht sich der Anteil ($n=988$) und sinkt auch in der zweiten Hälfte des 19. Jahrhunderts wieder auf 501 Relativsätze. Die Entwicklung ist also keineswegs linear, was auch in die Interpretation der logistischen Regression hineinspielen kann. Dort werden auch die beiden numerischen Variablen der Länge und der durchschnittlichen Skipgram-Surprisalwerte einbezogen. Deren Kennwerte sind in Tabelle 6.25 dargestellt.

Erneut wird mit der Beschreibung der Länge begonnen, da deren Werte eindeutiger sind und das Bild bestätigen, das sich durch das gesamte Korpus und alle Elemente zieht: Die extraponierte Relativsätze sind länger als die adjazente Relativsätze. Das gilt sowohl für ihren Median (adjazent: 7, ausgelagert: 9 Wörter), als auch für das arithmetische Mittel (adjazent: 8,326, ausgelagert: 11,15 Wörter). Lediglich bei der höchsten gemessenen Zahl von Wörtern in einem Relativsatz liegt der Wert bei den eingebetteten Relativsätzen

	Länge		Skipgram _{mean}	
Minimum	1	(1 2)	2,909	(2,909 3,229)
Median	8	(7 9)	3,626	(3,627 3,625)
Durchschnitt	9,848	(8,326 11,15)	3,633	(3,636 3,631)
drittes Quartil	12	(10 14)	3,747	(3,751 3,743)
Maximum	58	(58 55)	4,279	(4,279 4,169)
Std.-Fehler	0,21		0,003	
Varianz	204,16		0,03	

Tabelle 6.25.: Deskriptive Werte der Relativsätze aus medizinischen Texten. Links in der Klammer stehen die Werte der eingebetteten Relativsätze, rechts die der ausgelagerten Relativsätze.

zen höher (58 Wörter) als bei den ausgelagerten (55 Wörter).

Wieder ist dieser Unterschied bei den Skipgram-Surprisalwerten weniger klar zu erkennen. Zwar liegt der kleinste gemessene Wert der eingebetteten Relativsätze (2,9) niedriger als bei den extrapolierten (3,23), aber sowohl der Median als auch das arithmetische Mittel zeigen einen viel geringeren Unterschied. Hier sind die eingebetteten Relativsätze (Median: 3,627, arithmetisches Mittel: 3,636) sogar informativer als ihre ausgelagerten Pendanten (Median: 3,625, arithmetisches Mittel: 3,631). Das Gleiche gilt für die Werte des dritten Quartils und des Maximums. Auch mit dieser Beobachtung setzt sich das Bild aus den vorherigen Untersuchungen fort: Die Unterschiede in den durchschnittlichen Skipgramwerten sind nur marginal und scheinen zum Teil auch der Hypothese zu widersprechen. Ob das tatsächlich der Fall ist, wird die nun folgende logistische Regression untersuchen, bei der nur die Länge auf der Basis von 10 logarithmiert wird.

Logistische Regression Die Position bildet die abhängige Variable. Die unabhängigen Variablen werden von den durchschnittlichen Skipgramwerten, der logarithmierten Länge, dem Orality Score, der Zeitstufe und dem Relativsatztyp gebildet. Letzterer ist summenkodiert. Restriktivität erhält den Wert -0,5 appositive Relativsätze den Wert 0,5. Zusätzlich werden die Zweigeinteraktionen zwischen den unabhängigen Variablen bestimmt und eine rückwärtsgerichtete Modellauswahl anhand der p-Werte mit anova (vgl. R

Core Team 2018) durchgeführt. Das finale Modell präsentiert Tabelle 6.26.

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	10,81	2,87	3,77	0,0002	***
Skipgram _{mean}	-1,85	0,74	-2,52	0,01	*
Länge _{log}	-3,62	0,72	-5,02	<0,001	***
Orality Score	-2,16	0,69	-3,13	0,002	**
Zeitstufe	-2,46	0,86	-2,88	0,004	**
Typ	-0,22	0,09	-2,29	0,02	*
Orality Score:Zeitstufe	-0,23	0,13	-1,72	0,08	.
Skipgram _{mean} :Zeitstufe	0,50	0,23	2,24	0,03	*
Länge _{log} :Orality Score	1,16	0,66	1,75	0,08	.
Länge _{log} :Zeitstufe	0,37	0,21	1,72	0,09	.

Tabelle 6.26.: Logistische Regression aller Relativsätze in den medizinischen Texten, finales Modell. Negative z-Werte sprechen für Extrapolation.

Zunächst lässt sich beobachten, dass die meisten Haupteffekte auf Extrapolation hinweisen, da alle Estimates hier negative Vorzeichen besitzen. Das bedeutet, dass appositive Relativsätze eher ausgelagert werden ($z=-2,29$, $p<0,5$)¹⁶, aber auch dass Relativsätze mit hohen Skipgram-Surprisalwerten wahrscheinlicher extrapoliert werden ($z=-2,52$, $p<0,05$). Damit wäre für dieses Teilkorpus die erste Hypothese, die vorschlägt, dass Elemente mit hohen Surprisalwerten ausgelagert werden, bestätigt, obwohl andere Faktoren einflussreicher sind. Zu ihnen gehören der Orality Score ($z=-3,13$, $p<0,01$), der besagt, dass Auslagerung durch die Mündlichkeitsnähe eines Texts erklärt werden kann, und die Zeitstufe ($z=-2,88$, $p<0,01$). Der Veröffentlichungszeitraum eines Texts übt ebenfalls einen Einfluss auf Auslagerung aus. Der stärkste Prädiktor ist aber wieder die Länge ($z=-5,02$, $p<0,001$), wodurch auch die Analyse dieses Teilkorpus in Einklang mit den vorherigen steht.

Die Länge darf als Haupteffekt interpretiert werden, da sie über alle Orality Scores hinweg einen starken Einfluss ausübt. Die Interaktionen aus Mündlichkeitsnähe und Orality Score ($z=1,75$, $p<0,1$) und aus Länge und Zeitstufe ($z=1,72$, $p<0,1$) besagen Folgendes. Treten lange Relativsätze in mündlichkeitsnahen Texten auf, werden sie eher eingebettet. Das gleiche gilt dann,

¹⁶ Grafik D.1 in Anhang D kann dieses Ergebnis veranschaulichen.

wenn ein Relativsatz lang ist und in einem jüngeren Text vorkommt. Die Interaktion zwischen dem Orality Score und Zeitstufe ($z=-1,72$, $p<0,1$) ist schwach signifikant. Je mündlichkeitsnäher und jünger ein Text ist, desto eher werden die Relativsätze ausgelagert.

Interessant ist allerdings auch die Interaktion zwischen der Zeitstufe und den durchschnittlichen Skipgramwerten ($z=2,24$, $p<0,05$). Über alle Zeitstufen hinweg, werden informative Relativsätze eher ausgelagert. In den jüngsten Texten werden informative Relativsätze aber eher eingebettet.

6.3.4. Relativsätze in den theologischen Texten

In den theologischen Texten wurden 1159 (47 %) extrapolierte und 1302 (53 %) adjazente Relativsätze annotiert. Anders als bei den medizinischen ist der Anteil der adjazenten also ein wenig höher. Appositiv sind 887 Sätze, 923 Sätze sind restriktiv und 651 Sätze bleiben unbestimmt. Erneut überwiegen also die restriktiven Sätze.

Die Verteilung auf die verschiedenen Zeitabschnitte ist noch weniger linear als bei den medizinischen Texten. Von 1650 bis 1700 wurden 400 Relativsätze in theologischen Texten gefunden, die Zahl steigt in der ersten Hälfte des 18. Jahrhunderts auf 776 Relativsätze und geht in der zweiten Hälfte des 18. Jahrhunderts langsam zurück ($n=684$), um dann von 1800 bis 1850 massiv abzufallen ($n=187$). 414 Relativsätze konnten im letzten Abschnitt (1850 bis 1900) annotiert werden.

	Länge		Skipgram _{mean}	
Minimum	2	(2 3)	3,076	(3,076 3,215)
Median	12	(10 14)	3,651	(3,649 3,653)
Durchschnitt	16,52	(13,51 19,89)	3,656	(3,655 3,657)
drittes Quartil	20	(16 24)	3,758	(3,764 3,754)
Maximum	150	(147 150)	4,359	(4,318 4,359)
Std.-Fehler	0,32		0,003	
Varianz	257,44		0,028	

Tabelle 6.27.: Deskriptive Werte der Präpositionalphrase aus theologischen Texten. Links in der Klammer stehen die Werte der eingebetteten Relativsätze, rechts die der ausgelagerten Relativsätze.

In die Inferenzstatistik fließen wieder die Länge und die durchschnittlichen Surprisalwerte ein, die in Tabelle 6.27 dargestellt sind. Die adjazenten Relativsätze sind im Schnitt kürzer als die extraponierten Relativsätze. Im Median beträgt der Unterschied vier Wörter, die *in situ* Relativsätze sind 10 und die extraponierten 14 Wörter lang, und im Durchschnitt ist der Unterschied noch ausgeprägter. Im Schnitt sind eingebettete Relativsätze 13,51 Wörter lang und die ausgelagerten 19,89. Im vorherigen Kapitel wurde schon bemerkt, dass die längsten Sätze in den theologischen Texten deutlich länger sind als die in den medizinischen Arbeiten. Im Maximum sind die ausgelagerten Relativsätze 150 Wörter lang und bei den eingebetteten 147 Wörter. Beides spricht für einen ausgesprochen komplexen, stark erweiterten Satz. Auch diese deskriptiven Werte weisen wieder darauf hin, dass die Länge wohl einen starken Effekt auf die Position der Relativsätze haben dürfte. Gleichzeitig lässt sich aufgrund der weiten Spanne vermuten, dass eine Logarithmierung der Variable nötig werden wird, was bestätigt werden kann.

Die Werte der durchschnittlichen Skipgram-Surprisal liegen erneut viel näher beieinander. Zwar liegt der Median der eingebetteten Relativsätze bei 3,649 und der der extraponierten über ihnen (3,653), aber der Unterschied ist marginal. Das trifft in noch größerem Maße auf den Durchschnitt der Relativsätze zu. Die eingebetteten haben hier einen Wert von 3,655 und die ausgelagerten einen von 3,657. Der Unterschied wird also erst in der dritten Nachkommastelle sichtbar. Es spricht also zumindest augenscheinlich wenig dafür, dass es einen tatsächlichen Unterschied zwischen den beiden Relativsatzpositionen gibt, der auf die durchschnittlichen Skipgram-Surprisalwerte zurückzuführen ist. Darin ähnelt diese Betrachtung denen aus den vorherigen Abschnitten. Auch deshalb ist es allerdings notwendig, die Werte genauer zu betrachten. Das geschieht im nächsten Abschnitt.

Logistische Regression Diese logistische Regression erfolgt analog zu der des medizinischen Teilkorpus.¹⁷

¹⁷ Die abhängige Variable bildet die Position der Relativsätze. Die unabhängigen Variablen sind die Skipgram-Surprisalwerte, die logarithmierte Länge, der Orality Score, die Zeitstufe und der Relativsatztyp. Letzterer ist summenkodiert. Restriktivität wurde als -0,5 kodiert, appositive Relativsätze erhielten die Kodierung 0,5. Auch die Zweibegeinteraktionen werden

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	726	2,09	3,47	0,0005	***
Skipgram _{mean}	-1,62	0,55	-2,95	0,003	**
Länge _{log}	-2,36	0,39	-6,09	<0,001	***
Orality Score	-8,36	3,62	-2,31	0,02	*
Zeitstufe	0,19	0,13	1,49	0,14	
Typ	0,22	0,09	2,25	0,03	*
Skipgram _{mean} :Orality Score	2,67	0,95	2,79	0,005	**
Länge _{log} :Orality Score	1,55	0,66	2,34	0,02	*
Orality Score:Zeitstufe	-0,56	0,26	-2,38	0,02	*

Tabelle 6.28.: Logistische Regression aller Relativsätze in den theologischen Texten, finales Modell. Negative z-Werte sprechen für Extrapolation.

Insgesamt weisen mehr Faktoren des finalen Modells in Tabelle 6.28 auf Einbettung hin. Zu ihnen zählt der Relativsatztyp ($z=2,25$, $p<0,5$). Hier deutet Restriktivität auf Auslagerung hin (Grafik D.2), was auch den Ergebnissen früherer Forschung zum Thema der Relativsatzextrapolation entspricht (Kapitel 3.2). Im medizinischen Teilkorpus war das genaue Gegenteil der Fall. Der Orality Score ($z=-2,31$, $p<0,05$), die durchschnittlichen Skipgram-Surprisalwerte ($z=-2,95$, $p<0,01$) und die Länge der Phrasen ($z=-6,09$, $p<0,001$) können die Auslagerung von Relativsätzen ebenfalls erklären. Die Skipgram-Surprisalwerte interagieren signifikant mit dem Orality Score ($z=2,79$, $p<0,01$). Hochinformativ Relativsätze in mündlichkeitsnahen Texten werden zudem mit hoher Wahrscheinlichkeit eingebettet. Anders sieht es bei der Interpretation der Länge aus. Auch sie interagiert mit dem Orality Score ($z=2,34$, $p<0,05$). In beiden Registern werden lange Relativsätze eher ausgelagert, wobei ihre Einbettung in sehr mündlichkeitsnahen Texten etwas wahrscheinlicher ist verglichen mit sehr schriftlichkeitsnahen Texten. Auch die Interaktion zwischen der Zeitstufe und dem Orality Score ($z=-2,38$, $p<0,05$) zeigt, dass ein Relativsatz eher in jüngeren und mündlichkeitsnäheren Texten im Nachfeld steht als in mündlichkeitsnahen, älteren Texten.

beachtet. Die Auswahl des finalen Modells (Tabelle 6.28) erfolgt über die rückwärtsgerichtete Modellauswahl mit anova (vgl. R Core Team 2018).

6.3.5. Extraposition von Relativsätzen unter Beachtung der Distanz

Am Anfang von Kapitel 6.3 wurde mit (2a) und (2b) gezeigt, warum es in zahlreichen Fällen gar nicht möglich wäre, den Relativsatz auszulagern, ohne Ambiguitäten zu schaffen. Für die Untersuchung der Distanz wird trotzdem davon ausgegangen, dass jeder Relativsatz als erstes Glied im Nachfeld stehen könnte, selbst wenn dieses bereits besetzt ist. Auch werden die Sätze nicht explizit ausgeschlossen, in denen das Kopfnomen bei Extraposition ambig würde.¹⁸ Tabelle 6.29 zeigt die Werte der Distanzen und differenziert nach adjazenten und extraponierten Relativsätzen. Der Unterschied ist offenkundig. Die Distanzen, die adjazente Relativsätze zu überbrücken hätten, sind viel größer als die, die die extraponierten tatsächlich überschreiten. Daher ist von einem Einfluss der Variable auszugehen.

	adjazent	extraponiert
Minimum	1	1
Median	5	1
Durchschnitt	6,25	1,96
drittes Quartil	8	2
Maximum	47	20

Tabelle 6.29.: Distanz zwischen Kopfnomen und Nachfeld der Relativsätze, unterschieden nach ihrer Position.

Um das festzustellen, wird wie in den vorherigen Teilkapiteln eine logistische Regression durchgeführt, die nun auch die Distanzen enthält. Die visuelle Überprüfung resultierte darin, dass die Distanz logarithmiert werden muss. Das Gleiche gilt für die Länge. Auch das restliche Vorgehen entspricht dem schon präsentierten.

Tabelle 6.30 zeigt das finale Modell nach der rückwärtsgerichteten Modellauswahl. Hypothese H1 kann in diesem Modell nicht bestätigt werden, weil alle Informationsdichte betreffenden Variablen und Interaktionen ausgeschlossen wurden.

¹⁸ Diese Überlegungen können auch anhand dieses Korpus in späteren Arbeiten aufgegriffen werden.

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	0,63	0,32	2,00	0,05	*
Distanz _{log}	3,33	0,37	9,14	<0,001	***
Länge _{log}	-2,15	0,17	-12,47	<0,001	***
Orality Score	0,62	0,53	1,16	0,25	
Genre	-1,48	0,54	-2,76	0,006	**
Zeitstufe	-0,34	0,07	-4,34	<0,001	***
Typ	-0,31	0,11	-2,69	0,007	**
Distanz _{log} :Zeitstufe	0,59	0,13	4,73	<0,001	***
Länge _{log} :Genre	0,87	0,33	2,599	0,009	**
Orality Score:Genre	2,16	0,37	5,78	<0,001	***
Orality Score:Zeitstufe	-0,28	0,15	-1,895	0,06	.
Orality Score:Typ	0,69	0,23	2,96	0,003	**
Genre:Zeitstufe	0,26	0,11	2,28	0,02	*

Tabelle 6.30.: Logistische Regression aller Relativsätze im Gesamtkorpus, inklusive Distanz, finales Modell. Negative z-Werte sprechen für Extraposition.

Neben der Länge ($z=-12,47$, $p<0,001$), der Zeitstufe ($z=-4,344$, $p<0,001$), dem Relativsatztyp ($z=-2,69$, $p<0,01$) und dem Genres ($z=-2,76$, $p<0,01$), die alle für Auslagerung sprechen, ist die Distanz tatsächlich einer der bedeutendsten einzelnen Prädiktoren ($z=9,14$, $p<0,001$). Je weiter ein Antezedens vom Nachfeld entfernt steht, desto wahrscheinlicher wird die Adjazenz des Relativsatzes. Insofern können die Beobachtung, die auch Uszkoreit et al. (1998) getroffen haben und die von Hawkins (1992) beschrieben wurde, bestätigt werden. Die Distanz interagiert auch signifikant mit der Zeitstufe ($z=4,73$, $p<0,001$). In jüngeren Texten signalisieren längere Distanzen Adjazenz.

Für sich genommen ist nur der Orality Score nicht signifikant für die Position von Relativsätzen ($z=1,16$, $p=0,25$). Er wird aber nicht aus dem Modell ausgeschlossen, da er in Interaktion mit dem Genre Einbettung erklären kann ($z=5,78$, $p<0,001$). Je mündlichkeitsnäher ein theologischer Text also ist, desto eher werden Relativsätze in Adjazenz realisiert. Wenn der Orality Score mit dem Relativsatztyp interagiert ($z=2,96$, $p<0,01$), wird Einbettung ebenfalls wahrscheinlicher. Für Auslagerung spricht hingegen die Interaktion zwischen der Mündlichkeitsnähe und der Zeitstufe ($z=-1,895$, $p<0,1$). Wenn die Mündlichkeitsnähe steigt und der Text jünger wird, wird Extraposition wahr-

scheinlicher. Die Interaktionen zwischen der Länge und dem Genre ($z=2,599$, $p<0,01$) und die zwischen dem Genre und der Zeitstufe ($z=2,275$, $p<0,015$) weisen auf Einbettung in theologischen Texten hin.

In den vorherigen Abschnitten wurden die Ergebnisse für die Nominal- und Präpositionalphrasen sowie Relativsätze vorgestellt, die sehr unterschiedliche Resultate hinsichtlich des Einflusses von Skipgram-Surprisalwerten auf die Position der Konstituenten aufweisen. Die Begründung für die getrennte Analyse liegt vor allem in der Unterschieden zwischen Nominal- und Präpositionalphrasen beziehungsweise im Satzstatus der Relativsätze begründet. Obwohl alle drei Phrasentypen ins Nachfeld extraponiert werden können, unterliegen sie doch verschiedenen, grammatischen Gegebenheiten, die in den Kapitel, 2.4, 2.5 und 2.6 beschrieben wurden. Im nächsten Kapitel wird von dem Vorgehen abgewichen.

6.4. Nominalphrasen, Präpositionalphrasen und Relativsätze

In diesem Kapitel werden die drei Phänomene nun zusammen betrachtet und mithilfe der logistischen Regression in R (vgl. R Core Team 2022) wird untersucht, ob sich eine Aussage darüber treffen lässt, ob Extraposition allgemein von den Variablen durchschnittliches Skipgram-Surprisal, Länge, Genre, Mündlichkeitsnähe und Zeitstufe beschreiben lässt.¹⁹ Das Verfahren zur Modellauswahl erfolgt wie in den vorherigen Kapiteln und wie in Kapitel 5.4 beschrieben. Auf eine erneute Darstellung der deskriptiven Statistik wird verzichtet, weil diese schon für die einzelnen Phänomene in den jeweiligen vorherigen Kapiteln gegeben wurde. Nach der rückwärtsgerichteten Modellauswahl mit anova (vgl. R Core Team 2022) bildet Tabelle 6.31 das finale Modell der Regression mit allen Bestandteilen ab.

Zunächst ist hier die große Anzahl hochsignifikanter Prädiktoren zu bemerken, die darauf schließen lässt, dass es nicht nur eine Variable gibt, die

¹⁹ Die für die Relativsätze typischen Variablen wie Typ und Distanz können für Nominal- und Präpositionalphrasen nicht erhoben werden und werden daher auch nicht in die hiesige Analyse aufgenommen.

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	4,43	0,79	5,58	<0,001	***
Skipgram _{mean}	-0,78	0,21	-3,66	<0,001	***
Länge _{log}	-1,79	0,07	-23,15	<0,001	***
Genre	-0,90	0,21	-4,25	<0,001	***
Orality Score	0,21	0,19	1,13	0,26	
Zeitstufe	-0,78	0,28	-2,75	0,006	**
Skipgram _{mean} :Zeitstufe	0,199	0,08	2,58	0,01	*
Länge _{log} :Genre	0,66	0,15	4,51	<0,001	***
Genre:Orality Score	2,39	0,19	12,91	<0,001	***
Genre:Zeitstufe	0,23	0,06	4,09	<0,001	***
Orality Score:Zeitstufe	-0,48	0,07	-7,47	<0,001	***

Tabelle 6.31.: Logistische Regression aller untersuchten Phänomene im Gesamtkorpus, finales Modell. Negative z-Werte sprechen für Extrapolation.

Extrapolation allgemein erklären kann. Der einzige, im Modell verbleibende Prädiktor, der nicht signifikant ist, ist der Orality Score ($z=1,13$, $p=0,25$). Die Mündlichkeitsnähe eines Texts ist also irrelevant. Einen kleinen Einfluss übt die Interaktion zwischen den durchschnittlichen Skipgram-Surprisalwerten und der Zeitstufe aus ($z=2,58$, $p<0,05$). Informative Elemente in jüngeren Texten stehen eher eingebettet. Etwas signifikanter ist die Zeitstufe ($z=-2,75$, $p<0,01$), die besagt, dass Extrapolation in jüngeren Texten wahrscheinlicher wird, wenn keine anderen Faktoren in die Betrachtung einfließen. Alle restlichen, im Modell verbleibenden Faktoren sind in hohem Maße signifikant. Unter den Einzelfaktoren sticht das durchschnittliche Skipgram-Surprisal hervor ($z=-3,66$, $p<0,001$), was unbeachtet der anderen Faktoren dafür spricht, dass die Informativität und die damit einhergehende Belastung der Verarbeitungskapazitäten eine Rolle für Extrapolation spielt.

So werden auch lange Elemente eher ausgelagert ($z=-23,15$, $p<0,001$), sowie Elemente aus theologischen Texten ($z=-4,25$, $p<0,001$). Auch jüngere Texte und ein höherer Orality Score, sprich größere Mündlichkeitsnähe, sprechen für Extrapolation. Dem gegenüber stehen die Interaktionen, die Einbettung voraussagen. Das ist der Fall, wenn lange Phrasen in theologischen Texten auftreten ($z=4,51$, $p<0,001$), wenn ein theologischer Text mündlichkeitsnah ist

($z=12,91$, $p<0,001$) und wenn ein theologischer Text jüngeren Datum ist ($z=4,1$, $p<0,001$). Vor allem das extralinguistische Kriterium Genre weist in dieser Betrachtung also in Abhängigkeit anderer Faktoren, die eigentlich eher für Auslagerung plädieren, eine Bevorzugung von eingebetteten Elementen auf.

Im Hinblick auf Hypothese H1 kann die Nullhypothese widerlegt werden, weil die durchschnittlichen Skipgram-Surprisalwerte dazu geeignet sind, die Position der verschiedenen Elemente in Übereinstimmung mit der Hypothese vorherzusagen: Je informativer sie sind, desto eher stehen sie im Nachfeld. Dennoch muss diese Regression mit einer gewissen Vorsicht betrachtet werden. Zum einen sind die drei Elemente, die hier gemeinsam in die Analyse eingeflossen sind, sehr unterschiedlich, wie zu Beginn des Kapitels bereits hervorgehoben wurde. Zum anderen sind die Daten nicht sehr ausbalanciert. Deshalb wird bei der nachfolgenden Interpretation vornehmlich auf die Einzelphänomene eingegangen und angeraten, diese Regression lediglich als Vervollständigung der Untersuchung zu betrachten.

6.5. Diskussion Hypothese 1: Surprisal und Auslagerung

Hypothese H1 lautet: „Relativsätze und Phrasen mit hohen Surprisalwerten werden ausgelagert.“ Die Idee dahinter ist, dass der Satz in zwei Verarbeitungseinheiten unterteilt ist. Die zweite Verarbeitungseinheit umfasst das Nachfeld. Da zu diesem Zeitpunkt die meisten Informationen des Satzes, inklusive der rechten Satzklammer, die in vielen Fällen erst das lexikalische Verb beinhaltet und somit Unsicherheiten über die Satzgliedfunktionen und den Valenzstatus aller Konstituenten final auflöst, verarbeitet wurden, bleibt im Nachfeld mehr Kapazität zur Verarbeitung von anspruchsvollem Material.

Um Belege für die Hypothese zu finden, wurde daher in den Kapitel 6.1, 6.2, 6.3 und 6.4 überprüft, ob die durchschnittlichen Skipgram-Surprisalwerte einen signifikanten Einfluss auf die Nachfeldstellung von Nominal- und Präpositionalphrasen sowie Relativsätze haben. Wurde ein Einfluss gefunden, gilt die Nullhypothese als widerlegt, auch wenn andere Faktoren niedrigere

p-Werte erhalten haben. Wie sich dieser Einfluss erklären lässt, wird ebenfalls in diesem Kapitel diskutiert. Davor steht jedoch eine kurze Wiederholung der beiden hauptsächlich zur Erklärung genutzten Verarbeitungsideen nach Gibson (1998) und Levy (2008) und der Variablen, mithilfe derer sie untersucht wurden. Gibson (1998) nimmt an, dass es ein Kontingent an Ressourcen zur Sprachverarbeitung im Arbeitsgedächtnis gibt, die sich in zwei Gruppen unterteilen lassen: Die Memory und Integration Cost (vgl. Gibson 1998: 9, 11).

Die Integration Cost wird mit der Surprisaltheorie in Zusammenhang gebracht. Diese beschreibt Levy (2008) als den Aufwand, der aufgewendet wird, wenn Vorhersagen über das nächste Element des Satzes getroffen und dieses ins Satzverständnis integriert werden. Die Integration Cost kann als vorwärtsgerichtet oder erwartungsbasiert im recht unmittelbaren Kontext beschrieben werden. Hohe Surprisalwerte beschreiben also beispielsweise, dass viele Ressourcen verbraucht werden, weil Erwartungen im spezifischen Kontext revidiert werden müssten.²⁰ Entlastung entsteht wiederum durch eine Veränderung des Kontexts oder die Verwendung von Wörtern mit niedrigeren Unigram-Surprisalwerten. Die Integration Cost wird in dieser Arbeit mit den Variablen der Surprisalwerte, des Relativsatztyps, der Satzgliedfunktion beziehungsweise des Valenzstatus und des Orality Scores assoziiert.

Die Variable der Surprisalwerte kann allerdings in zwei unterschiedliche Richtungen der Konstituentenposition weisen. Zum einen sprechen hohe Surprisalwerte, wie in Kapitel 1.1 argumentiert wurde, für Extraposition, weil davon ausgegangen wird, dass im Nachfeld mehr kognitive Kapazitäten für die Verarbeitung lexikalisch informativer Elemente zur Verfügung stehen und Hale (2001) zeigt, dass mehr kognitive Kapazitäten auch zur Verarbeitung hoher Surprisalwerte verbraucht werden. Prinzipiell ist aber ebenso denkbar, dass die Einbettung von Elementen trotz und mit hohen Surprisalwerten dafür sorgt, dass die nachfolgenden Elemente leichter verarbeitet werden, weil die entsprechenden Erwartungen für die folgenden Elemente geschaffen werden. Zur Veranschaulichung wiederhole ich die Beispiele (3) aus Kapitel 5.2.

²⁰ In Kapitel 2.1 wurde das beispielsweise an Beispiel (3) verdeutlicht („Sie dürfen die Braut jetzt küssen“). Würde man *Braut* durch *Katze* ersetzen, würden die kompletten Erwartungen durcheinander gebracht und somit viel Integration Cost verbraucht.

- (3) a. Die Stadt wurde von Cäsar erobert.
 b. Die Stadt Rom wurde von Gaius Julius Cäsar erobert.

Rom ist mit der Person Gaius Julius Cäsar assoziiert, wodurch angenommen werden kann, dass die Erwähnung des Stadtnamens die Integrationskosten auf dem Personennamen verringern kann.²¹

(3) ist lexikalischer Natur. Aber Rezipierende bauen auch Erwartungen auf grammatischen Konstruktionen auf, wodurch die *Satzgliedfunktion* und der *Valenzstatus* auch mit der Integration Cost verbunden sind. Besonders die Verarbeitung des Verbs erzeugt Erwartungen an weitere Satzglieder und deren Valenzstatus. In Kapitel 2.1 wurde das an Beispiel (4) („Ich habe dem Helfer gedankt“) verdeutlicht. Je weniger Wahlmöglichkeiten für diese Funktionen zur Verfügung stehen, desto schneller können sie verarbeitet werden. Daher kann die Mittelfeldstellung einer Konstituente die Entropie auf den Besetzungsmöglichkeiten der rechten Satzklammer verringern. Ähnlich funktioniert die Vorhersage des *Relativsatztyps*. Die in Kapitel 2.4.3 benannten Einleiter restriktiver Relativsätze machen einen folgenden Relativsatz erwartbarer. Auch die Diskursfunktionen bestimmter Elemente können Erwartungen erzeugen. Hier ist wieder die Restriktivität des Relativsatzes anzuführen. Wenn die Rezipierenden den Eindruck bekommen, dass ein (potenzielles) Antezedens weiterer Identifikation bedarf, rechnen sie auch mit einem Relativsatz.

Die Argumentation, warum der *Orality Score*, sprich die Mündlichkeitsnähe, Einfluss auf die Integration Cost haben sollte, ist schwieriger. Bei oraler Kommunikation ist es von Vorteil, wenn die Integration neuer Wörter möglichst schnell gelingt. Das kann zum einen Extraposition begünstigen. Zum anderen kann aber die Einbettung von Elementen andere erwartbarer macht, was bei oraler Kommunikation, von der auch konzeptionelle Mündlichkeit abgeleitet ist, den Vorteil der Extraposition überschreiben kann.

²¹ Das verwendete 2-Skip-Bigram-Modell (vgl. Guthrie et al. 2006) kann diese Zusammengehörigkeit, die für den menschlichen Leser einfach nachzuvollziehen ist, aber nicht darstellen. Das Modell berechnet bei Satz (3b) nur folgende Skipgramme: *Die – Stadt, Die – Rom, Stadt – Rom, Stadt – wurde, Rom – wurde, Rom – von, wurde – von, wurde – Gaius, von – Gaius, von – Julius, Gaius – Julius, Gaius – Cäsar, Julius – Cäsar, Julius – erobert, Cäsar – erobert*. *Rom* wird in dem Modell also nicht zur Vorhersage von *Cäsar* herangezogen werden. Trotz dieses Nachteils ist dieses Modell aus den in Kapitel 5.1 beschriebenen Gründen die beste Berechnungsart für Surprisal, die für die aktuelle Forschungsfrage zur Verfügung steht.

Für die Verarbeitung von Elementen ist es also zum einen bedeutsam, dass sie schnell und einfach in den bestehenden, lokalen Kontext integriert werden können, was der Integration Cost beschreibt. Es gilt aber zu bedenken, dass die Ressourcen für die Integration Cost als auch für die Memory Cost dabei aus dem gleichen Pool kognitiver Kapazität genommen werden. Wenn also mehr für die Integration neuer Wörter verwendet wird, steht weniger für die Memory Cost zur Verfügung und umgekehrt (vgl. Gibson 1998).

Die Memory Cost hängt zwar auch mit Erwartungen zusammen, ist aber mit Vergessen und Erinnern assoziiert. Sprachproduktion und -verständnis erfordern es, dass man sich an das erinnert, was in einem Satz oder Text bereits produziert wurde. Das Erinnern ist deshalb nötig, weil bestimmte Elemente Erwartungen hervorrufen. Je weiter das Element allerdings zurückliegt, desto weniger hilft es dabei, andere Elemente vorauszusagen. Laut Gibson (1998: 11) nimmt die Aktivierung von Erwartungen ab. In dieser Arbeit wird es nach Futrell et al. (2021) eher so formuliert, dass die Bestandteile vergessen werden. Somit ist dieser Blickwinkel eher rückwärtsgerichtet.²² Die Variablen, die diese Art von Verarbeitung darstellen, sind die Länge und die Distanz.²³ Die Auslagerung von langem Material verkürzt den Abstand zwischen den Satzklammern, unabhängig davon, welches Element ausgelagert wird. Die größere Distanz zwischen dem Antezedens und dem Relativsatz kann die korrekte Zuordnung des Relativsatzes zu einem Kopfnomen erschweren²⁴ oder den ganzen Relativsatz unerwarteter machen, weil beispielsweise passende Signalwörter wieder vergessen wurden oder er nur appositiv ist, Signalwörter also ganz fehlen können. Für die Variablen bedeutet das wiederum, dass eine große Länge der Konstituente insgesamt und eine kleine Distanz speziell zwischen Antezedens und Relativsatzbeginn für Extraposition sprechen sollten. Auch die Variable des Orality Scores ist relevant für die Memory Cost, wenn sie mit oraler Kommunikation verbunden ist. Denn in

²² In Voigtmann und Speyer (2022) wird außerdem beschrieben, dass Leerstellen in der Verarbeitung des Satzes freigehalten werden müssen, um das ausgelagerte Material an der passenden Position wieder in den Satz einzugliedern, was die Verarbeitung ebenfalls erschwert.

²³ Sie kann also auch mit Dependenzmaßen in Verbindung gebracht werden, die beispielsweise nach Collins (2014) genau wie die UID zumindest für englische Daten Wortstellungspräferenzen erklären können.

²⁴ Dieses Problem ist in flexionslosen Sprachen allerdings größer als im Deutschen.

der Flüchtigkeit dieses Mediums dürften kürzere Distanzen noch bedeutender werden (vgl. z. B. Horton und Gerrig 2002; 2005), weil gleichzeitig so viele andere Verarbeitungsmechanismen benötigt und angesprochen werden. Es ist zu erwarten, dass auch in konzeptionell mündlichen, aber medial graphischen Texten bewusst oder unbewusst versucht wird, die Gefahr des Vergessens zu reduzieren.

Kognitive Kapazitäten werden also durch Integration und Memory Cost konsumiert, die beide als Modell für Verarbeitungsaufwand dienen und die zur optimalen Kommunikation nicht übermäßig verbraucht werden dürfen. Die folgenden Abschnitte werden also zeigen, dass die Ergebnisse insgesamt mit Verarbeitungsschwierigkeiten, die durch Extraposition beziehungsweise Einbettung reduziert werden, erklärt werden können. Belege für Hypothese H1 finden sich dabei für:

- Nominalphrasen im Gesamtkorpus
- Nominalphrasen im medizinischen Teilkorpus
- attributiven Nominalphrasen
- attributiven Präpositionalphrasen
- Relativsätze im Gesamtkorpus
- Relativsätze im medizinischen Teilkorpus
- (Relativsätze im theologischen Teilkorpus)

Es muss aber bemerkt werden, dass die Skipgram-Surprisalwerte fast nie den stärksten Prädiktor bilden. Meisten kann die Länge mehr Positionsunterschiede erklären, was für eine stärkere Rücksichtnahme auf die Memory-Cost-Ressourcen spricht und deren Einfluss für das Deutsche hervorhebt. In den folgenden Kapiteln wird daher jeweils damit begonnen, die Ergebnisse der der logistischen Regression zu interpretieren und sie auf Memory beziehungsweise Integration Cost oder andere Verarbeitungsschwierigkeiten zu beziehen, die beispielsweise durch die Verwendung eines wissenschaftlichen Korpus' entstehen.

6.5.1. Nominalphrasen

Alle Nominalphrasen Im Gesamtkorpus weist die Variable der Skipgram-Surprisalwerte signifikant ($p < 0,05$) auf Auslagerung hin (Tabelle 6.5). Ohne andere Faktoren miteinzubeziehen, lässt sich anhand der Surprisalwerte vorhersagen, ob eine Nominalphrase eher eingebettet oder ausgelagert ist. Damit liefert dieses Ergebnis einen Beleg für das Zutreffen der ersten Hypothese. Allerdings bilden die Skipgram-Surprisalwerte nicht den stärksten Prädiktor. Besonders der extralinguistische Faktor Genre und der Orality Score sowie die Zeitstufe beeinflussen die Position der Nominalphrase mehr (Tabelle 6.5).

Sowohl als Einzelvariable als auch mit dem Genre interagierend weist der Orality Score eher auf Einbettung hin. Unabhängig von ihrer Informativität werden Nominalphrasen in mündlichkeitsnahen Texten also mit höherer Wahrscheinlichkeit eingebettet. Dieses Phänomen beobachtet beispielsweise auch Ebert (1980). Besonders deutlich wird das bei theologischen Texten, wie die Interaktion zwischen dieser Variable und dem Orality Score zeigt. In dieser Konstellation scheint es also bedeutender zu sein, den Satz mit der rechten Satzklammer abzuschließen, als andere zuvor beschriebene Verarbeitungserleichterungen wahrzunehmen.

Um das zu erklären, muss ein Blick auf die Bedeutung der rechten Satzklammer und der Funktion von Nominalphrasen geworfen werden. Die meisten Sätze im untersuchten Korpus weisen eine rechte Satzklammer auf, die das lexikalische Verb beherbergt. Erst mit dessen Verarbeitung wird auch die Valenz des verwendeten Verbs vollständig erkannt. Weil die Auslagerung die Distanz zwischen den Satzklammern reduziert, wird die rechte Satzklammer früher erkannt und kann somit auch früher verarbeitet werden als im Fall von Einbettung des ausgelagerten Elements. Dadurch löst sich sowohl die Entropie auf dem konkreten Verb als auch die Entropie auf dessen Aktanten auf. Die kognitiven Kapazitäten, die zuvor für die Integration Cost verwendet wurden (vgl. Gibson 1998; Levy 2008), werden wieder frei.

Nominalphrasen können mehr (obligatorische) Satzgliedfunktionen erfüllen als Präpositionalphrasen oder attributive Relativsätze. Tabelle 6.2 zeigt, dass auch alle Funktionen von extraponierten Nominalphrasen erfüllt wer-

den, obwohl Ramers (2006: 119) nur „umfangreich[e] oder adverbial[e]“ Nominalphrasen im modernen Standarddeutschen im Nachfeld sieht. Die Ressourcen zur Verarbeitung der lexikalischen Bedeutung und der Satzgliedfunktion werden aus demselben Pool wie die Ressourcen zur Funktionsverarbeitung der Nominalphrase geschöpft (vgl. Gibson 1998: 8). Wenn das Verb mit seiner Valenz erkannt wird und die Funktion der Nominalphrase im Nachfeld folglich eindeutig determiniert ist, muss keine Anstrengung auf die Funktionsvorhersage verwendet werden.²⁵ Mehr Kapazität kann folglich zur Verarbeitung der lexikalischen Information der Nominalphrase verwendet werden. Die Entrophiereduktion auf der Funktion garantiert den Verarbeitungsvorteil für Konstituenten mit hohen Surprisalwerten im Nachfeld.

Diese Verarbeitungserleichterung sollte auch für mündlichkeitsnahe Texte in besonderem Maße gelten. Hier müssen aber zwei Dinge erwähnt werden. Zum einen heißt konzeptionell mündlich nicht automatisch, dass die Texte auch oral vorgetragen und rezipiert werden. Im hier untersuchten Korpus ist sogar vom Gegenteil auszugehen. Zum anderen ist ein Blick auf die eingebetteten Phrasen, die der Untersuchung zugrunde liegen, notwendig. Über 80 % der eingebetteten Nominalphrasen wurden im Mittelfeld annotiert, allerdings war die lexikalische Füllung der rechten Satzklammer kein zwingendes Kriterium für die Annotation einer eingebetteten Phrase. Möglicherweise, wenn auch selten, wird das lexikalische Verb bereits in der linken Satzklammer verarbeitet, sodass die Satzgliedfunktionen an der Stelle schon bekannt ist. Die Verarbeitungserleichterung ähnelt dann dem, was für die extraponierten Nominalphrasen vorgeschlagen wurde.²⁶

²⁵ Gibson (1998: 14) argumentiert ähnlich dafür, dass keine Kapazitäten dafür aufgebraucht werden müssen, Prädikate vorherzusagen. Zudem finden Jaeger (2005; 2010; 2011) und Frank und Jaeger (2008) dafür auch Belege bei der *that*-Weglassung. Wenn das Relativum nicht realisiert wird, muss dessen Funktion auch zusätzlich zur Eigenbedeutung des ersten Worts des Relativsatzes verarbeitet werden, was mehr Ressourcen verbraucht.

²⁶ Hier muss aber auch darauf hingewiesen werden, dass bei den Nominalphrasen in Fall von lexikalischen Verben in der linken Satzklammer noch weitere Schwierigkeiten auftreten, die bei Extraposition ausgeschlossen werden können. Nach der Verarbeitung der linken Satzklammer ist zwar bekannt, welche Mitspieler das Verb zu sich nimmt, aber dennoch bestehen Stellungsfreiheiten im Mittelfeld (Kapitel 2.2.3 und z. B. Lenerz 1977), die es verkomplizieren, deren genauen Auftretenszeitpunkt vorherzusagen. Zudem können endozentrische Kasusmarkierungen auftreten, die die Verarbeitung wiederum kurzfristig erschweren, weil Unsicherheiten auf deren Zuordnung entstehen.

Dieser Effekt kann in der vorliegenden Arbeit allerdings nicht näher untersucht werden,

Zugleich ist aber nach Levy (2008) zu betonen, dass die Verarbeitung der rechten Satzklammer auch durch das Material vor der Klammer im Mittelfeld erleichtert werden kann. Das kann zum einen dadurch geschehen, dass die rechte Satzklammer lexikalisch erwartbarer wird, weil die davor stehenden Konstituenten schon ein Skript postulieren (vgl. z. B. Ostermann 2021), in dem nur noch wenige Verben möglich sind, oder aber weil die Kombination aus Satzgliedern den gleichen Effekt hat. Dieser Fall kann die Vorteile der Extraposition überschreiben und kann die Interaktion zwischen Orality Score und Genre erklären, die für theologische Texte bei größerer Mündlichkeitsnähe auf Einbettung verweist.

In diesen Texten scheint die kognitive Kapazität, die hinter der rechten Satzklammer frei wird, nicht so sehr benötigt zu werden, dass sie das Prestigekriterium (vgl. Takada 1998) des vollständigen, undurchbrochenen Satzrahmens überschreiben würde. Es lässt sich vermuten, dass die Texte aufgrund ihrer Mündlichkeitsnähe (Kapitel 1.1 und 2.7, Tabelle 7.1) insgesamt etwas leichter zu verarbeiten sind, weil sie die dazu notwendigen syntaktischen und lexikalischen Mechanismen bedienen. Dazu zählen auch die in Kapitel 5.3 beschriebenen Konzepte wie ein höherer Anteil von Pronomen an allen Wörtern.²⁷ Zugleich handelt es sich bei den Texten um theologische Texte. Die thematische Vielfalt der Werke ist begrenzter als bei den medizinischen. Die beschriebenen Konzepte sollten für die Autoren und ihr Publikum ebenfalls bekannter sein. Auf den Einfluss des Genres wird im Verlauf des Kapitels noch genauer eingegangen.

Der Orality Score interagiert signifikant mit der Zeitstufe ($p < 0,01$) und weist in dieser Kombination darauf hin, dass Extraposition in mündlichkeitsnahen Texten häufiger wird. Die vorliegende Untersuchung dokumentiert die Anfänge des in synchronen Studien belegten Trends, dass die Extraposition

weil diese Verarbeitungsmessung mindestens normalisierte oder anders kasusmarkierte Daten erfordern würden. Zum jetzigen Zeitpunkt sind solche Daten ein Desiderat und auch die entsprechende Aufarbeitung des vorliegenden Korpus würde eher in Datenknappheit und unsicheren Ergebnissen resultieren. Für zukünftige Projekte und Studien wäre es aber wünschenswert, sich diesem Problem anzunehmen.

²⁷ Möglicherweise verkürzt gerade dieser Aspekt auch die Distanz zwischen den Satzklammern und löst mögliche Unsicherheiten der Kasus durch das Flexionsparadigma der Pronomen schnellstmöglich auf. Gleichzeitig geben diese Pronomen aber weniger lexikalische Hinweise auf das Verb in der rechten Satzklammer.

von Nominalphrasen in jüngeren, mündlichkeitsnahen Texten häufiger vorkommt (vgl. z. B. Beneš 1979; Zahn 1991; Brooks 2006). Obwohl die zeitliche Entwicklung Gegenstand von Hypothese H3 ist, soll hier eine weitere Vermutung aufgestellt werden, die sich in die Bedeutung der Satzgliedfunktion der Nominalphrase einreicht. In der Vorrede zu Tabelle 6.2 wurde auf die Annotationsart der einzelnen Objekte hingewiesen und dabei bemerkt, dass Konstituenten auch dann als Subjekte oder Objekte annotiert wurden, wenn sie koordiniert sind und ihr erstes Glied im Mittelfeld des Matrixsatzes steht. Das ist vor allem in den Texten jüngeren Datums zu bemerken. Für die Verarbeitung bedeutet das, dass die Satzgliedfunktion bereits *vor* der rechten Satzklammer verarbeitet werden muss, und zugleich die Vorteile für die Vorhersage der rechten Satzklammer liefert. Zugleich können die im Nachfeld stehenden Nominalphrasen aufgrund der höheren Verarbeitungskapazität an dieser Stelle auch bei einer inhaltliche Komplexität leichter verarbeitet werden.

Zusammenfassend ist zu sagen, dass die Skipgramwerte der Nominalphrasen Rückschlüsse auf ihre Position zulassen, aber die Mündlichkeitsnähe und das Genre sowie der Publikationszeitpunkt teils in gegenseitiger Beeinflussung ebenfalls gute Vorhersagen über die Position der Nominalphrase machen können. Als Erklärung für die Ergebnisse wurde die Bedeutung der Satzklammer und die verschiedenen Satzgliedstatus der Nominalphrasen für die Verarbeitung hervorgehoben. Neben den beiden genannten Interaktionen zeigte sich auch das Genre eines Textes als guter Prädiktor für die Position der Nominalphrase. Auf diese Unterschiede soll im Folgenden bei der Betrachtung dieser beiden Teilkorpora näher untersucht werden.

Medizinisches Teilkorpus Auch im medizinischen Subkorpus weisen die Skipgram-Surprisalwerte auf einen signifikanten Einfluss auf die Position der Nominalphrasen hin ($p < 0,05$). Wie bei der Betrachtung des Gesamtkorpus zeigt sich, dass Hypothese H1 grundsätzlich bestätigt werden kann: Die Position der Nominalphrase ist anhand ihrer Informativität vorauszusagen.

Es ist allerdings wieder zu bemerken, dass die Länge als Hauptprädiktor ($p < 0,001$), mehr der gefundenen Auslagerungen erklären (Tabelle 6.7).

Für die medizinischen Texte sind die in Kapitel 3.1 erwähnten Gründe für Extraposition wie eine große Länge der Nominalphrase und die Mündlichkeitsnähe des Textes unabhängig vom Zeitabschnitt, in dem die Texte verfasst wurden, erfüllt. Bei der Betrachtung des Gesamtkorpus wurde dafür argumentiert, dass mehr kognitive Ressourcen durch die Integration Cost verbraucht wurden. Es überwiegt aufgrund der Bedeutung der Länge der gedächtnisbasierte, rückwärtsgerichtete Ansatz. Wie zu Beginn des Kapitels (6.5) geschildert, wird durch die Extraposition von langen Sätzen die Distanz zwischen den Satzklammern verkürzt. Dadurch wird das Vergessens des Satzbeginns unwahrscheinlicher. Auch die Satzklammer und die darin kodierete Verbvalenz steht entsprechend früher zur Verfügung und kann somit die Unsicherheiten auf der Satzgliedfunktion der extraponierten Nominalphrase früher reduzieren und die Verarbeitung informativer Elemente im Nachfeld erleichtern. Das alles sorgt dafür, dass weniger Memory Cost verbraucht wird. Neben den gedächtnisbasierten Variablen ist aber auch der Surprisalwert der Phrasen von Bedeutung. Es kann also gesagt werden, dass in den medizinischen Texten alle für Verarbeitung als relevant beschriebenen Ansätze angewendet werden.

Die Notwendigkeit dafür liegt im Genre begründet. Unter anderem besteht in medizinischen Texten die Notwendigkeit, komplexere und schwierigere Sachverhalte darzustellen, als es für theologische angenommen werden kann. Dass die Gegenstände, mit denen sich die theologische Texte befassen, begrenzt sind und gerade im untersuchten Zeitraum höchst alltagsnah für die Autoren und die Rezipierenden waren, wurde bereits angedeutet. Bei den medizinischen Texten ist vom Gegenteil auszugehen. Zum einen bietet der menschliche Körper eine breitere Anzahl von Themen. Das wird zusätzlich dadurch verstärkt, dass es in dieser Zeit zahlreiche neue Erkenntnisse, Behandlungsmethoden und wissenschaftlichen Fortschritt gab (vgl. z. B. Eckart 2021), die sprachlich dargestellt werden mussten. Diese Begebenheiten schaffen die Notwendigkeit neuer Begriffe. Neue Begriffe sollten schwerer zu verarbeiten sein als lange etablierte Wörter, die automatisch auch häufiger genutzt wurden.

Wenn also in der rechten Satzklammer ein seltenes oder sogar neues Verb steht, kann nicht davon ausgegangen werden, dass die Präsentation der obligatorischen Ergänzungen vor dem lexikalischen Verb dessen Verarbeitung erleichtert. Um zu überprüfen, ob die Elemente der rechten Satzklammer in den medizinischen Texten tatsächlich schwerer zu verarbeiten sind, wurden die Mittelwerte der Unigramsurprisal der rechten Satzklammern mithilfe eines Welch t-Tests in R (R Core Team 2022) verglichen. Die Werte der medizinischen Texte (3,34) weisen tatsächlich höhere Werte auf als die der theologischen Texte (3,28). Der Unterschied ist signifikant ($t=7,8$, $p<0,001$).²⁸

Das bedeutet wiederum, dass Sätze in medizinischen Texten insgesamt schwieriger zu verarbeiten sind und es daher notwendig ist, vorwärts- und rückwärtsgerichtete Prinzipien der optimalen Verarbeitung zu befolgen. Wenn die selteneren, vermutlich neuen Verben früher erkannt werden, kann die vom Verb abhängige Kernaussage des Satzes schneller verstanden werden. Aufgrund dieser sprachlichen Innovation und den damit verbundenen Schwierigkeiten, dürfte hier auch kein Vorteil für die Vorhersagbarkeit des Verbs in der rechten Satzklammer durch die Einbettung von Konstituenten entstehen (vgl. Levy 2008).²⁹ Dafür ist die Bedeutung der höheren, zur Verfügung stehenden kognitiven Kapazität hinter der rechten Satzklammer von größerer Bedeutung für die Verarbeitung des Materials, was die Signifikanz der Skipgram-Surprisalwerte erklärt.

Auch der Einfluss des Orality Scores in Interaktion mit der Länge lässt sich auf den Verarbeitungsvorteil durch die frühe Verarbeitung der rechten Satzklammer erklären, die bei konzeptioneller Mündlichkeit eine schnellere Verarbeitung ermöglicht.

²⁸ Obwohl für den Zeitabschnitt von 1650 bis 1700 die theologischen Texte im Trainingskorpus überwiegen, weil sich zu dieser Zeit erst etabliert, wissenschaftliche Texte auf Deutsch zu verfassen, kann insgesamt davon ausgegangen werden, dass das Trainingskorpus zwischen den beiden Genres ausgeglichen genug ist, sodass diese Unterschiede in den Unigram-Surprisalwerten der rechten Satzklammern nicht nur auf einen solchen Effekt zurückgeführt werden können (Anhang C).

²⁹ Die Überprüfung davon würde den Rahmen dieser Arbeit überschreiten, bietet sich aber für weitere Forschung an.

Theologisches Teilkorpus Im theologischen Teilkorpus kann kein unabhängiger Beleg für die Gültigkeit von Hypothese H1 gefunden werden. Die durchschnittlichen Skipgram-Surprisalwerte der Nominalphrasen sind nicht signifikant ($p=0,2$). Allerdings liegt im Modell neben einem signifikanten Haupteffekt des Orality Scores ($p<0,05$), der für Einbettung in mündlichkeitsnahen Texten spricht, eine signifikante Interaktion zwischen den Skipgram-Surprisalwerten und den Orality Scores vor ($p<0,05$). Die besagt, dass hohe Skipgramwerte in mündlichkeitsnahen Texten zur Vorhersage der Nominalphrasenextrapolation verwendet werden können. Diese freien Ressourcen gewinnen erst dann an Bedeutung, wenn das Nachfeld zu einem konzeptionell mündlichen Text gehört und Phrasen mit hohen Surprisalwerten dahin ausgelagert werden. Für dieses Teilkorpus zeigt sich also der in Hypothese H2b postulierte Zusammenhang zwischen Mündlichkeitsnähe und Informationsdichte (Kapitel 7.2). Dass für die schriftlichkeitsnahen, theologischen Texte ein genau gegenteiliger Effekt gefunden wird, bedarf allerdings Erklärungen.

Die theologischen Texte stellen insgesamt bekanntere Konzepte dar, die in geringerem Maß Innovationen abbilden müssen als die medizinischen und daher auch weniger Neuerungen im Vokabular benötigen. Ihre Inhalte sind präsenter für die Menschen im untersuchten Zeitraum. Daher dürften sie generell leichter verständlich gewesen sein und sich auch auf die Erwartungen der verwendeten Verben auswirken. Eine frühere Vorhersage eines bestimmten Satzgliedes wird so weniger relevant, wodurch keine doppelte Belastung aus Satzgliedstatus und lexikalischer Informativität einer Konstituente zustande kommt. Unterstützt werden diese Annahmen dadurch, dass die Uni-gramwerte der rechten Satzklammern in den theologischen Texten signifikant niedriger sind als die der medizinischen Texte.³⁰ Es kann also davon ausgegangen werden, dass die Platzierung einer Nominalphrase im Mittelfeld die Verarbeitung des restlichen Satzes erleichtert, weil sie die Integration Cost an der Stelle senkt, was wiederum erklären würde, warum auch hochinformativ Nominalphrasen zumindest in schriftlichkeitsnahen Texten eingebettet werden können.

³⁰ Die Werte der medizinischen Texte (3,34) weisen tatsächlich höhere Werte auf als die der theologischen Texte (3,28). Der Unterschied ist signifikant ($t=7,8$, $p<0,001$).

Zum anderen ist die Schreibpraxis der Theologen und die Autoren selbst als Erklärung heranzuziehen.³¹ Wie Tabelle 7.1 zeigt, sind die theologischen Texte generell mündlichkeitsnäher als die medizinischen, was auch an den Autoren liegen kann. In Anhang A werden die Autoren, deren Texte in das Korpus eingeflossen sind, kurz vorgestellt. Bei fünf bis sechs³² von 13 Theologen handelt es sich um Geistliche, die in ihrem Alltag vermutlich mehr konzeptionell mündliche Texte verfassten, die auch für die orale Perzeption gedacht waren. Diese Gewöhnung kann sich auch auf ihren Stil beim Schreiben von wissenschaftlichen Texten ausgeübt haben.

Allerdings verweist diese Argumentation auch auf den Unterschied zwischen Sprecher- und Hörerorientierung. Bisher wurde dafür argumentiert, dass Extraposition genutzt wird, um für die Rezipierenden besser verständliche Texte zu produzieren. Allerdings haben die Untersuchungen von Arnold et al. (2000) und Frank und Jaeger (2008) auch die Produktion und damit einhergehende Erleichterungen für die Sprechenden in den Blick genommen. Es kann auch hier angenommen werden, dass beide Parteien des Organonmodells (vgl. Bühler 1934) von Auslagerung profitieren. Der Vorteil für die Rezipierenden ist klar. Für die Produzierenden kann ein Nutzen angenommen werden, weil auch sie keine hohe Entropie mehr auf der Satzgliedfunktion des Elements hinter der rechten Satzklammer einkalkulieren müssen.

Anhand der sprachlichen Daten, speziell der Interaktion aus Länge und Zeitstufe ($p < 0,01$), ist anzunehmen, dass eine Veränderung im theologischen Schreiben über die Jahrhunderte stattgefunden hat, die durch „einen historischen Prozeß, in dem die Theologie sich Bedingungen anpaßt, die ihr von außen, durch die aus sich selbst begründete westeuropäische Aufklärung gestellt werden“ (Sparn 1985: 19) eingeleitet wurde. Das betrifft in der Theologie eher die Techniken, mit denen Themen wie die Bibelexigese betrachtet werden (vgl. z. B. Reventlow 1997). Es kann daher auch angenommen werden, dass die Theologen im 19. Jahrhundert mit der Entwicklung der Geisteswissenschaften im heutigen Sinne (vgl. z. B. Dilthey 1910) bewusster Stilistiken der schriftlich-

³¹ Da das Trainingskorpus zwischen medizinischen und theologischen Texten ausgeglichen ist, kann der Effekt nicht daher stammen (Anhang C).

³² Über Petrus Hanssen stehen keine Informationen zur Verfügung, daher kann nicht ausgeschlossen werden, dass er ebenfalls als Priester tätig war.

keitsnahen Sprache anwendeten, die bei den medizinischen Texten früher zu sehen sind. Dadurch deutet sich an, dass die vorhersagebasierten Verarbeitungserleichterungen in jüngeren Texten von den gedächtnisbasierten, die für die Extraposition langer Nominalphrasen argumentieren, abgelöst werden. Eine entsprechende Tendenz wird sich bei der Betrachtung von Hypothese H3 (Kapitel 8.4) bestätigen.

Attributive Nominalphrasen Bei der Betrachtung der attributiven Nominalphrasen wurde ein signifikanter Einfluss der durchschnittlichen Skipgram-Surprisalwerte gefunden ($p < 0,1$), der als Beleg für die Gültigkeit von Hypothese H1 zählt. In diesem Modell (Tabelle 6.11) ist sonst nur die Interaktion aus der Länge und der Zeitstufe signifikant (ebenfalls $p < 0,1$). Die beiden einflussreichen Prädiktoren können daher als Tendenzen zur Auslagerung verstanden werden (Tabelle 6.11). Bei den attributiven Nominalphrasen scheinen die gedächtnisbezogenen Einschränkungen also weniger schwer zu wiegen, als die in Kapitel 3.1 vorgestellten Grammatiken und Studien vermuten lassen. Hierbei ist aber auch ein Wechsel beobachtbar, der durch die signifikante Interaktion zwischen Länge und Zeitstufe signalisiert wird. In Kapitel 8.4 wird darauf näher eingegangen.

Hohe Surprisalwerte weisen auf Auslagerung hin. Surprisal bildet sogar die einzige nicht-interagierende Variable, die einen signifikanten Einfluss auf die Position der Nominalphrase aufweist. Als Erklärung lässt sich der Attributstatus dieser Nominalphrasen heranziehen. Durch diese Eigenschaft ist ihre Funktion im Satz determiniert, da Attribute nicht von der Verbvalenz gefordert werden. Stattdessen erfüllen sie die Funktion, nähere Informationen über ihr Antezedens zu geben oder dessen Identifikation zu gewährleisten. Unsicherheiten über die Funktion der Phrase bestehen zum Zeitpunkt ihrer Verarbeitung im Nachfeld also nicht mehr, da zu dem Zeitpunkt alle Mitspieler der Verbalhandlung erkannt und verarbeitet wurden. Die Unsicherheit bezüglich der Funktion fällt weg und die dadurch frei werdenden Kapazitäten können für Wörter mit höheren Surprisalwerten genutzt werden.

Über die Zeit hinweg ist eine Veränderung zu beobachten. Die Regression lässt den Schluss zu, dass in jüngeren Texten die Länge einen verstärkten Ein-

fluss auf Extraposition ausübt ($p < 0,1$). Es zeigt sich also ein erneuter Hinweis darauf, dass ein Wechsel von erwartungsbasierten hin zu gedächtnisbasierten Verarbeitungserleichterungen stattgefunden haben kann.

Zusammenfassung In den hier betrachteten Regressionsanalysen lässt sich also durchaus eine Evidenz dafür finden, dass Nominalphrasen mit hohen Skipgram-Surprisalwerten eher im Nachfeld als im Mittelfeld stehen. Problematisch ist allerdings, dass Surprisal nie am besten zur Voraussage der Position der Nominalphrasen geeignet ist. Lediglich bei den attributiven Nominalphrasen ist keine Variable ein besserer Prädiktor.

Als Argument für die Extraposition der Nominalphrasen wurde angeführt, dass nach Verarbeitung des Vollverbs keine Kapazitäten mehr für die Funktion der Nominalphrase verbraucht werden müssen und die Rezipierenden sich allein auf die lexikalische Schwierigkeit der Phrase konzentrieren können. Auf der anderen Seite zeigt sich aber auch ein verstärkter Einfluss extralinguistischer Faktoren. Vor allem das Genre ist hier anzuführen, das mit den unterschiedlichen Schreibpraxis der Autoren und deren Publikum verbunden wurde. Zudem unterscheiden sich die Themen der beiden Wissenschaftsrichtungen. Während der Gegenstand der Theologie bis heute kaum Veränderungen durchlaufen hat, müssen die Mediziner zahlreiche neue Erkenntnisse und Veränderungen in ihrer Zeit integrieren, was sich in der Verwendung seltenerer Verben widerspiegelt. In ihren Texten muss also davon ausgegangen werden, dass Integration und Memory Cost gleichermaßen beansprucht werden und die Texte daher so gestaltet sind, dass sie möglichst beide Ressourcen optimal nutzen.

Neben diesen Unterschieden deutet sich auch eine Veränderung über den untersuchten Zeitraum hinweg an, die explizit bei der Betrachtung von Hypothese H3 (Kapitel 8.4) untersucht wird.

6.5.2. Präpositionalphrasen

Alle Präpositionalphrasen Im Gegensatz zu den Ergebnissen der Nominalphrasen ist im Gesamtkorpus der Präpositionalphrasen durch die logistische

Regression kein Einfluss von Surprisalwerten auf Extraposition nachzuweisen. Die Skipgram-Surprisalwerte wurden auch aus den Interaktionen und aus dem gesamten Modell im Zuge der rückwärtsgerichteten Modellauswahl ausgeschlossen (Tabelle 6.15).

Der einflussreichste Einzelfaktor ist Länge ($p < 0,001$). Hier kommt also eine Verringerung der Memory Cost zum Tragen. Das frühe Erkennen der rechten Satzklammer scheint entscheidend für Extraposition zu sein, da längere Präpositionalphrasen mit höherer Wahrscheinlichkeit ausgelagert werden. Dadurch stehen die beiden Satzklammern näher beisammen, was die Gefahr reduziert, dass der Anfang des Satzes quasi vergessen wird, bevor die rechte Satzklammer verarbeitet werden konnte.

Neben dem reduzierten Vergessen des Satzanfangs kann auch der Valenzstatus von Präpositionalphrasen als Erklärung herangezogen werden. Im Gegensatz zu Nominalphrasen bilden Präpositionalphrasen seltener obligatorische Aktanten. In ihrer Funktion als freie Angaben oder fakultative Ergänzungen müssen sie aber nicht zwingend realisiert werden. Grafik 6.9 zeigt, dass diese weglassbaren Präpositionalphrasen fast immer häufiger im Nachfeld stehen als die Objekte. Eine Ausnahme bildet der Zeitabschnitt 1750 bis 1800, wo der Anteil der Objekte fast 50 % beträgt. Es ist daher fraglich, ob die frühere Stellung von fakultativen Ergänzungen und freien Angaben im Satz zu einem Verarbeitungsvorteil der rechten Satzklammer führt, der mit dem Vorteil von Objekten gänzlich vergleichbar ist. Der Vorteil für das Satzverständnis wird eher durch die früher erscheinende rechte Satzklammer erzeugt, da diese das mögliche Fehlen eines Objekts- oder einer Ergänzung deutlicher macht. Der Einfluss der Länge kann also darüber erklärt werden.

Gleichzeitig ist aber auch denkbar, dass das lexikalische Surprisal einen Einfluss auf die Extraposition von Präpositionalphrasen haben sollte, der allerdings aufgrund der Berechnungsart und der Form von Präpositionalphrasen nicht im Modell nachgewiesen werden kann.

Präpositionalphrasen bestehen zwingend aus einer Präposition und ihrem Komplement (Kapitel 2.6.1). Forschung darüber, wie viele Präpositionalphrasen neben der komplementären Nominalphrase auch einen Determi-

nierer im untersuchten Zeitraum zu sich nehmen, fehlt. Setzen sich die Beobachtungen von Waldenberger (2009) aber fort, sollte die Kombination im Frühneuhochdeutschen und frühen Neuhochdeutschen noch häufiger werden. Daraus resultiert aber ein Problem bei der Berechnung der Skipgram-Surprisalwerte auf der Basis von Lemmata. Denn erstens besteht die Präpositionalphrase dann auf jeden Fall aus zwei Funktionswörtern. Funktionswörter sind generell häufiger und daher erwartbarer. Ihr Surprisalwert ist niedriger als der von Inhaltswörtern (vgl. z. B. Degaetano-Ortlieb et al. 2019: 268). Zudem steht auch das Verb der rechten Satzklammer in seiner lemmatisierten Form, was das Auftreten der Präposition wiederum erwartbarer macht, als wenn das Language Model gleichzeitig abbilden könnte, dass die Präposition beispielsweise auf ein Partizip in der rechten Satzklammer folgt. Auch das kann den Surprisalwert auf den ersten beiden Wörtern der Präpositionalphrase reduzieren. Es ist zu erwarten, dass dieses Problem auch bei den eingebetteten Phrasen auftreten kann, was auch die sehr ähnlichen deskriptiven Werte der Minimalpaaren erklären kann (Tabelle 6.4). Der fehlende Einfluss der Surprisalwerte wäre demnach auf die Berechnungsart zurückzuführen.

Im Gegensatz zu diesem Language Model verarbeiten die menschlichen Rezipierenden aber sowohl die Wortform des Verbs als auch den Kasus des Determinierers. Außerdem steckt ein hohes Maß an Information in der Präposition, die beispielsweise als desemantisiert erkannt werden kann oder in ihrer Eigenbedeutung noch verarbeitet werden muss. Diese Informationen stellen wichtige Weichen für die weitere Verarbeitung dar, können mit aktuellen Mitteln in historischen Daten aufgrund der Datenknappheit für die Epoche aber nicht erfasst werden. Es muss aber davon ausgegangen werden, dass die Art der Surprisalberechnung einen großen Einfluss auf die Ergebnisse hat und die Verwendung von Daten, die beispielsweise auch die verschiedenen Kasus einbeziehen, zu anderen Ergebnissen führen würde.

Die Erfassungsart des einzigen anderen signifikanten Einzelprädiktors, der Zeitstufe ($p < 0,05$), sollte indes keinen bedeutenden Einfluss auf die Ergebnisse haben. Von den in Tabelle 6.15 dargestellten Prädiktoren weist die Zeitstufe das schwächste Signifikanzniveau auf. Dennoch ist es interessant,

dass sie auf Einbettung hinweist. Sie scheint zunächst also das Bild zu bestätigen, dass Präpositionalphrasen in älteren Texten häufiger ausgelagert wurden und das im Lauf der Zeit abnimmt. Der Zusammenhang mit anderen Faktoren wird durch die Interaktionen überprüft.

Die Interaktion zwischen der Zeitstufe und dem Orality Score ($p < 0,001$) weist auf Auslagerung hin, was auch bei der Analyse der Nominalphrasen gefunden wurde. In jüngeren, mündlichkeitsnahen Texten ist Extraposition wahrscheinlicher, was sich auch in der Darstellung von Grammatiken zu dem Thema (Kapitel 3.1) niederschlägt. Sowohl für die Nominalphrasen als auch die Präpositionalphrasen scheint eine Veränderung in der Schreib- und Gewöhnungspraxis von Autoren und deren Leserschaft stattgefunden zu haben. Die Auslagerung bewirkt eine Verkürzung der Satzklammerdistanz. Wenn man die beiden Faktoren der Länge und diese Interaktion betrachtet, dann erscheinen die gedächtnisbasierten Ansätze für die Verarbeitung von Präpositionalphrasen bedeutender zu sein beziehungsweise bedeutender zu werden.

Dennoch kann diese Schlussfolgerung aufgrund der signifikanten Interaktion des Orality Scores mit dem Genre ($p < 0,001$) nicht als allgemeingültig angesehen werden. Konzeptionell mündliche Texte, die gleichzeitig theologisch sind, weisen auf Einbettung von Präpositionalphrasen hin. Die Erklärung dafür beruft sich wieder auf die Schreibpraxis der Theologen und wird bei der separaten Betrachtung dieses Teilkorpus näher dargestellt.

Obwohl konzeptionelle Mündlichkeit eher auf Auslagerung hinweisen sollte, ist sie hier mit Einbettung verbunden. Schon im Abschnitt zu den Nominalphrasen wurde argumentiert, dass die theologischen Texte bekanntere Konzepte präsentieren und frequentere Wörter nutzen. Insgesamt scheint der Verarbeitungsaufwand also geringer, als es für die medizinischen Texte zu vermuten ist. Daher gefährdet Einbettung das Satzverständnis hier nicht. Möglicherweise sind die Präpositionalphrasen in den mündlichkeitsnahen Texten insgesamt einfacher zu verarbeiten und benötigen daher die höheren Verarbeitungskapazitäten im Nachfeld nicht. Zugleich ist denkbar, dass die Einbettung der Phrasen eine Verarbeitungserleichterung der anderen Be-

standteile des Satzes gewährleistet, wie für Beispiel (3)³³ argumentiert wurde. Allerdings würde diese Erleichterung nur an wenigen Stellen im Satz zustande kommen, nämlich unmittelbar hinter der annotierten Präpositionalphrase, weshalb insgesamt von einem geringen Einfluss ausgegangen werden muss. Stattdessen kann auch herangezogen werden, dass das Genre stark mit Mündlichkeitsnähe korreliert (Tabelle 7.1) und man daher einen ähnlichen, auf die Inhalte des Genres zurückzuführenden Effekt hat, wie oben für die Interaktion aus dem Orality Score und dem Genre selbst skizziert wurde und auf den im Folgenden näher eingegangen wird.

Medizinisches Teilkorpus Auch im medizinischen Teilkorpus wurden die durchschnittlichen Skipgram-Surprisalwerte aus der logistischen Regression ausgeschlossen (Tabelle 6.17). Daher ist Hypothese H1 nicht zu bestätigen.

Der einflussreichste Faktor für Extraposition ist Länge ($p < 0,001$). Die Argumentation, warum lange Präpositionalphrasen in den medizinischen Texten eher ausgelagert werden, schließt sich fast nahtlos an die Ansätze an, die für die Extraposition der Nominalphrasen genutzt wurden. Die frühere Verarbeitung der Satzklammer sorgt für das frühere Erkennen der Verbvalenz. Bei den komplizierteren Inhalten der medizinischen Texte ist ein verkürzter Satzrahmen auch unabhängig von der lexikalischen Schwierigkeit hilfreich, die durch die Surprisalwerte gekennzeichnet wird und deren Einfluss vielleicht durch andere Language Models besser erkannt werden kann. Die Verkürzung des Satzrahmens verringert die Memory Cost und sorgt dafür, dass für die gesamte Verarbeitung des Satzes mehr Ressourcen frei sind, die in den medizinischen Texten auch genutzt werden müssen, weil neue Begriffe neue Wertigkeiten und Aktanten mit sich bringen, deren frühere Verarbeitung ebenfalls vorteilhaft ist.

Die Interaktion aus Mündlichkeitsnähe und Zeitstufe wurde ebenfalls schon bei den Nominalphrasen gesehen und besprochen. Auch in den me-

³³ Der Text des Beispiels lautet:

- 3 a) Die Stadt wurde von Cäsar erobert.
- b) Die Stadt Rom wurde von Gaius Julius Cäsar erobert.

dizinischen Texten, die generell eher schriftlichkeitsnah verfasst sind, ist ein höherer Anteil von Faktoren, die für Mündlichkeitsnähe sprechen, eine zuverlässige Möglichkeit, um die Position einer Präpositionalphrase zu erklären. Der Einfluss wird besonders in den jüngsten Texten noch verstärkt. Auch das deutet auf einen Sprachwandel hin, der in Kapitel 8.4 besprochen wird.

Theologisches Teilkorpus In den theologischen Texten wird wegen des Ausschlusses der Skipgramwerte (Tabelle 6.19) kein Beleg für den Einfluss von Informationsdichte auf die Extraposition von Präpositionalphrasen gefunden. Einflussreich ist wieder vor allem die Länge ($p < 0,05$). In jüngeren Texten steigt die Wahrscheinlichkeit, dass lange Phrasen extrapositioniert werden allerdings noch, wie die Interaktion die Interaktion aus Länge und Zeitstufe ($p < 0,001$) verdeutlicht. Zudem besteht eine Tendenz, lange Präpositionalphrasen in mündlichkeitsnahen Texten eher einzubetten ($p < 0,1$). Dieses Ergebnis der Interaktion aus Länge und Zeitstufe erklären sich durch die vermutete Annäherung im Schreiben der Theologen an einen (natur-)wissenschaftlicheren Stil. Die Tendenz, lange Präpositionalphrasen in mündlichkeitsnahen Texten einzubetten, kann auch auf Erklärungsansätze zurückgeführt werden, die oben bereits erwähnt wurden. Durch die bekannteren Inhalte der theologischen Texte, die in dieser Darstellungsform im Alltag der Menschen auch präsenter waren als die medizinischen Inhalte, kann der Effekt, dass Verarbeitungskapazitäten durch die vergrößerte Distanz zwischen den Satzklammern in den theologischen Texten zu stark ausgelastet wird, überschrieben werden. Deshalb führt Einbettung nicht zum Informationsverlust, obwohl in mündlichkeitsnahen Texten besondere Ansprüche an das Gedächtnis zu vermuten sind. Dennoch muss auf das schwache Signifikanzniveau verwiesen werden.

Attributive Präpositionalphrasen Weder die Betrachtung aller Präpositionalphrasen noch die separate Analyse nach Genre haben einen Einfluss von Skipgram-Surprisalwerten auf Extraposition gezeigt. Die Analyse der attributiven Präpositionalphrasen zeigt ein anderes Ergebnis, nämlich eine Tendenz dazu, dass sich die Nachfeldstellung von diesen Präpositionalphrasen durch ihre Informativität voraussagen lässt ($p < 0,1$). Zwar erklären die anderen Fak-

toren, die in Tabelle 6.20 aufgelistet sind, mehr Positionsvarianten. Aber erstens können die Skipgramwerte hier nicht ausgeschlossen werden und zweitens weisen sie in Interaktionen einen signifikanten p-Wert auf ($p < 0,05$). Für diese spezielle Gruppe von Präpositionalphrasen kann die erste Hypothese zumindest bestätigt werden.

Der deutlichste Unterschied ist klar die Funktion der Präpositionalphrasen. Es handelt sich um Attribute. Unsicherheiten über die Satzgliedfunktion besteht im Nachfeld nicht mehr, weil das Verb mit seinen Mitspielern an dieser Stelle schon verarbeitet wurde und schwächere, aber denkbare Vorteil für die Verarbeitung der rechten Satzklammer ebenfalls schon geleistet wurden. Daher können hinter der rechten Satzklammer wieder die freien Kapazitäten genutzt werden, um Präpositionalphrasen mit höherem Informationsgehalt zu verarbeiten.³⁴

Bei diesen Erklärungsansätzen ist dennoch zu beachten, dass nur 21 extraponierte attributive Phrasen im ganzen Korpus existieren und insgesamt keinen klaren Einfluss einzelner Faktoren auf die Stellung attributiver Präpositionalphrasen gefunden werden konnte. In Kapitel 6.2.5 wurde beschrieben, welcher Faktor für Einbettung und welcher für Extraposition spricht. Die Argumente für die Variablen Länge, Zeitstufe und Orality Score wurden an anderer Stelle schon gegeben. Der Fokus soll hier auf den Interaktionen mit den Skipgramwerten liegen. Die Skipgramwerte interagieren sowohl mit Länge, als auch mit Orality Score und deuten in beiden Interaktionen auf Einbettung hin (jeweils $p < 0,05$).

Lange, informative attributive Präpositionalphrasen werden also mit hoher Wahrscheinlichkeit eingebettet. Zu Beginn des Kapitels wurde argumen-

³⁴ Natürlich könnte es trotzdem sein, dass an dieser Stelle noch Präpositionalphrasen mit der Funktion einer freien Angabe treten. Insgesamt betrachtet stehen Phrasen dieser Funktion aber am seltensten im Nachfeld (Grafik 6.9). Eine Vermutung für dieses Ergebnis ist, dass diese Phrasen eher am linken Rand des Mittelfelds auftreten könnten und die zu überbrückende Distanz daher zu groß ist. Allerdings ist Forschung über die Position von Adverbialen im Mittelfeld selbst für synchrone Daten weitestgehend ein Desiderat (mit Ausnahme von Frey und Pittner 1998). Für den hier untersuchten Zeitraum fehlen derartige Studien vollkommen. Da aber bekannt ist, dass sie in ihrer Funktion als Rahmenbilder auch oft im Vorfeld stehen (vgl. Speyer 2010: 265), ließe sich argumentieren, dass sie nicht in hohem Maße an der Verarbeitung von Elementen im Mittelfeld beteiligt sind, sondern vor allem die Funktion erfüllen, die Verbzweitstellung des Deutschen zu sichern (Speyer 2010).

tiert, dass das Einfügen einer Phrase die Surprisalwerte auf dem nachfolgenden Material lexikalisch, aber nicht zwingend syntaktisch, reduzieren kann. Zudem wurde in Kapitel 5.2 begründet, warum mehr Wörter insgesamt auch zu einer Vereinfachung des Verständnisses führen können, obwohl dadurch die Distanz zwischen den Satzklammern verlängert wird. Da die attributiven Präpositionalphrasen im Schnitt aber auch kürzer sind als alle Präpositionalphrasen, sollte sich das nicht stark auswirken (Kapitel 6.2.5 und 6.2.1).

Prinzipiell ist der gleiche Effekt als Erklärung für die Interaktion aus Skipgramwerten und dem Orality Score heranzuziehen. Informative Phrasen, die kognitive Kapazität stärker verbrauchen sollten, werden an einer Stelle verarbeitet, die nach der Argumentation aus Kapitel 1.1 dafür speziell in mündlichkeitsnahen Texten schlecht geeignet ist. Hier ist nicht zu überprüfen, inwiefern die hohen Surprisalwerte im Kontext des Matrixsatzes tatsächlich schwerer zu verarbeiten sind als der Rest des Satzes, und somit überhaupt einen nach den Prinzipien der Uniform Information Density (vgl. Levy und Jaeger 2007) unerwünschten Ausschlag nach oben darstellen. Dafür spricht auch, dass 12 attributive Präpositionalphrasen aus medizinischen Texten stammen und nur neun aus theologischen, sodass grundsätzlich für eine schwerere Verarbeitung der Texte argumentiert werden kann.

Zudem ist auch der Attributstatus als Argument für Einbettung zu nennen. Bei den zuvor betrachteten Präpositionalphrasen handelte es sich fast ausschließlich um eigenständige Konstituenten. Daher zählt das Gesetz von Behaghel (1932), wonach Zusammengehöriges auch zusammenstehen sollte, für sie nicht, für die attributiven Präpositionalphrasen hingegen schon. Trotz ihrer hohen Informativität können die Phrasen besser verarbeitet werden, wenn sie adjazent zu ihrem Bezugswort stehen, weil sie sonst schlechter zugeordnet werden können und das Gedächtnis zu sehr belasten. Das kann wiederum in mündlichkeitsnahen Texten verstärkt zum Tragen kommen.

Zusammenfassung Bei den Präpositionalphrasen kommen vor allem die gedächtnisbasierten Verarbeitungsmechanismen als Erklärungsansätze in Frage, die die Memory Cost verringern. Außer bei der Analyse der attributiven Präpositionalphrasen zeigen die Skipgram-Surprisalwerte keinen (alleinigen)

Einfluss auf die Position der Präpositionalphrasen. Stattdessen ist die Länge der Präpositionalphrase am besten dazu geeignet, ihre Nachfeldstellung zu erklären. Dafür, dass das Surprisal keinen hohen Einfluss auf die Extraposition ausübt, wurde auch mit dessen Berechnungsart und dem Funktionswortcharakter der Präposition und des Artikels argumentiert. Zukünftige Untersuchungen zu diesem Thema sollten diese Problematik aufgreifen.

Was die vorliegende Untersuchung aber zeigt, ist die besondere Bedeutung kurzer Satzklammerdistanzen, die das frühere Verarbeiten des Verbs in der rechten Satzklammer ermöglichen. Dass dieser Faktor im medizinischen Teilkorpus ausgeprägter ist, wird wieder auf die Inhalte der Texte und die Veränderung in der Forschung im Zeitalter des frühen Neuhochdeutschen zurückgeführt, die ein schnelleres Verständnis der Sätze und auch des Erkennens, welchen Valenzstatus die Präpositionalphrase überhaupt einnehmen kann, vorteilhaft machen. Der Einfluss der Länge ist auch in den theologischen Texten erkennbar, allerdings lässt sich anhand dessen erst in jüngeren Texten sehr zuverlässig vorhersagen, wo die Präpositionalphrase platziert wird. Wieder wird hier zur Erklärung eine veränderte Schreibpraxis der Autoren herangezogen. Auch zeigt sich eine generelle Entwicklung dahin, dass in konzeptionell mündlichen Texten jüngeren Datums Auslagerung unabhängig von ihrer Länge oder Informativität häufiger wird.

Obwohl die attributiven Präpositionalphrasen als einzige einen Einfluss von Surprisalmessungen erkennen lassen, zeigen sich auch hier indirekt die Effekte der Distanz, allerdings der zwischen dem Kopfnomen und der ausgelagerten Präpositionalphrase. Sowohl in mündlichkeitsnahen Texten als auch bei langen Phrasen ist Einbettung wahrscheinlicher, was weiterer Analyse auf einer größeren Datenbasis zur Klärung des Befunds bedarf.

6.5.3. Relativsätze

Alle Relativsätze Auch die Relativsätze wurden dahingehend untersucht, ob Adjazenzstellung oder Extraposition durch die durchschnittlichen Skipgramwerte des Relativsatzes vorhergesagt werden können. Dabei zeigt Tabelle 6.24 eine Tendenz, dass hoch informative Relativsätze ins Nachfeld gestellt

werden ($p < 0,1$). Andere Faktoren erklären allerdings mehr Positionsdaten der Relativsätze. Trotzdem kann daraus geschlossen werden, dass Belege für die Korrektheit von Hypothese H1 zu finden sind.

Der einflussstärkste Einzelfaktor für Auslagerung ist die Länge ($p < 0,001$). Dieses Ergebnis wurde auch, wie Kapitel 3.2 darstellt, von zahlreichen anderen Studien bestätigt. Speziell bei den Relativsätzen konkurrieren zudem zwei verschiedene Distanzmaße miteinander: Die Distanz zwischen den Satzklammern, die auch bei den Nominal- und Präpositionalphrasen relevant ist, und die durch Einbettung maximiert wird, und die Distanz zwischen dem Relativsatzbeginn und dem Antezedens, die vergrößert wird, wenn der Relativsatz ausgelagert wird. Beide Fällen erklären die gedächtnisbasierten Verarbeitungsschwierigkeiten. Durch die Distanzstellung von einem langen Relativsatz zu seinem Kopfnomen wird die Gefahr, den Anfang des Matrixsatzes zu vergessen, verringert und die rechte Satzklammer wird früher gesehen, was in den Verarbeitungsvorteilen resultieren kann, die auch bei den Nominal- und Präpositionalphrasen illustriert wurden.

Die Ergebnisse in Tabelle 6.24 zeigen aber auch signifikante Interaktionen, wie die zwischen Orality Score und Genre ($p < 0,001$) sowie die Interaktion zwischen Genre und Zeitstufe ($p < 0,001$). Bei der ersten Interaktion (Orality Score und Genre) zeigt sich, dass Mündlichkeitsnähe in theologischen Texten für Adjazenz spricht. Bei mündlichkeitsnahen Texten wird aber davon ausgegangen, dass sie Prinzipien beachten, die sie auch besser verarbeitbar machen, wenn sie im oralen Medium rezipiert würden. Das führt zur Beobachtung, dass in oralen Texten Auslagerung frequenter ist. Für die Relativsätze ist der Faktor Mündlichkeitsnähe allein aber nicht signifikant ($p = 0,27$). Der relevante Unterschied muss in den theologischen Texten gesucht werden.

Schon bei den Nominal- und Präpositionalphrasen wurde für einen Unterschied zwischen den theologischen und den medizinischen Texten argumentiert. Zum einen ist die Thematik bei den theologischen alltagsrelevanter und konstanter als in den medizinischen Texten. Zum anderen wurde gezeigt, dass auch die rechten Satzklammern in den theologischen Texten wohl besser zu verarbeiten sind, weil sie mit frequenteren Wörtern besetzt sind. Die frühe-

re Verarbeitung der rechten Satzklammer bringt folglich in den theologischen Texten wieder einen geringeren Verarbeitungsvorteil. Diesen Verarbeitungs- oder Produktionsvorteil gewährleistet Adjazenz.

Diese Bedingung steht, meines Erachtens, in engem Zusammenhang mit der Funktion der Relativsätze. Relativsätze identifizieren ihr Bezugswort näher oder geben weitere Informationen darüber. Das wurde durch den Faktor des Typs überprüft, der mit dem Genre dahingehend interagiert, dass Einbettung in den theologischen Texten generell wahrscheinlicher ist und auch restriktive Relativsätze in diesem Genre eher adjazent stehen. Wenn das mit den bekannteren, einfacheren Themen der theologischen Texte und den frequenteren Wörtern in der rechten Satzklammer sowie der größeren Mündlichkeitsnähe in dem Genre zusammen gebracht wird, kann die Verarbeitung dadurch erleichtert werden, dass der Relativsatz unmittelbar nach seinem Antezedens verarbeitet wird.

Ein appositiver Relativsatz erfüllt nur die Funktion, Informationen über sein Antezedens zu geben. Dieses Antezedens selbst schafft aber keine hohe Erwartung daran, dass der Relativsatz (irgendwann) folgen muss. Die Entropie auf dem Element nach dem Antezedens ist also größer; es könnte zwar ein Relativsatz auftreten, es kann aber auch eine andere Konstituente folgen. Bei restriktiven Relativsätzen ist das nicht so. Es muss also davon ausgegangen werden, dass die Entropie auf dem Erscheinen eines restriktiven Relativsatzes etwas geringer ist, weil den Rezipierenden klar ist, dass das Kopfnomen nicht vollständig identifiziert wurde. Die Rezipierenden wissen aber auch, dass Relativsätze in ca. 50 % der Fälle extrapponiert stehen. Durch dieses Wissen wird es erleichtert, kognitive Kapazitäten für die Eingliederung des Relativsatzes frei zu halten. Aufgrund des weniger erwarteten Auftretens eines Relativsatzes nach einem identifizierten Kopfnomen, kann Extraposition die kognitiven Kapazitäten überlasten, wenn er Relativsatz im Nachfeld auftritt.

Unter den Voraussetzungen der theologischen Texte und deren teils angenommenen und teils überprüften sprachlichen Unterschieden zu den medizinischen Werken zeigt sich der Vorteil der Einbettung, wenn entweder der Relativsatztyp hinzu tritt oder die besonderen, zusätzlichen Anforderungen

an bessere Verarbeitung gestellt werden, die für mündlichkeitsnahe Texte beobachtet wurden. Die Einbettung von (appositiven) Relativsätzen in theologischen und mündlichkeitsnahen Texten erleichtert die Verarbeitung zudem auf beiden Ebenen: Die Rezipierenden würden den Relativsatz, stünde er im Nachfeld, unabhängig von dessen Informativität nicht mehr erwarten, wodurch Verarbeitungsschwierigkeiten entstehen. Zudem wäre durch die größere Distanz die Zuordnung und das rückwirkende Erkennen des korrekten Kopfnomens unter Umständen nicht mehr gewährleistet. Der Effekt scheint für theologische, jüngere Texte noch ausschlaggebender zu sein ($p < 0,001$).

Ein berechtigter Einwand ist die Frage danach, warum der Effekt nicht auch in medizinischen Texten beobachtbar ist. Zum einen lässt sich auch das auf die komplexeren Inhalte zurückführen, zum anderen auf die Argumentation bezüglich der selteneren, schwierigeren Satzklammer. Da die besprochenen Sachverhalte komplexer sind, kann die kürzere Distanz zwischen den Satzklammern hier schwerer wiegen als bei den theologischen Texten. Dies steht in Einklang mit den Beobachtungen, die für Nominal- und Präpositionalphrasen getroffen wurden. Durch die schwierigeren Themen kann eine zusätzliche Information über den Relativsatz trotz größerer Distanzen bei Extraposition für den Leser einen Vorteil bringen. Die Ergebnisse werden bei der Betrachtung der nach Genre getrennten Korpora vertieft interpretiert.

Was ebenfalls noch Erwähnung finden soll, ist die signifikante Interaktion aus Orality Score und Zeitstufe ($p < 0,05$), die auch für die beiden Phrasen beobachtet wurde. Auch diese deutet auf eine Veränderung in der Schreibpraxis über die Jahrhunderte hin (Kapitel 8).

Medizinisches Teilkorpus Im medizinischen Teilkorpus lässt sich der Einfluss von Skipgramwerten auf Extraposition ebenfalls nachweisen ($p < 0,05$). Ohne andere Faktoren einzubeziehen, bietet diese Variable einen signifikanten Hinweis darauf, wo der Relativsatz steht und liefert dadurch Evidenz für das Zutreffen von Hypothese H1: Hochinformativ Relativsätze stehen eher im Nachfeld, wo mehr Verarbeitungskapazität frei sind. Der Satzstatus der

Relativsätze scheint also einen Einfluss auf die Positionierung zu haben.³⁵

Trotzdem muss bedacht werden, dass Surprisal nicht der relevanteste Faktor für Auslagerung ist. Tabelle 6.26 zeigt klar, dass die Länge eines Relativsatzes seine Position am besten vorhersagen kann ($p < 0,001$) und der signifikanteste Prädiktor ist. Lange Relativsätze werden also ins Nachfeld gestellt. Dieses Ergebnis wurde in dieser Arbeit schon mehrfach gefunden und auf die dadurch verkürzte Distanz zwischen den Satzklammern zurückgeführt. Dass die frühere Verarbeitung der rechten Satzklammer in diesem Fall aber auch Erleichterungen für die Vorhersage der Funktion der Relativsätze bietet, ist auszuschließen, da nur attributive Relativsätze untersucht wurden. Das heißt, es liegen keine Unsicherheiten bezüglich ihrer Funktion vor.

Weitere signifikante Faktoren für Extraposition sind die Mündlichkeitsnähe und die Zeitstufe (je $p < 0,01$). Auch hier kann dafür argumentiert werden, dass Gedächtnisressourcen durch die Auslagerung geschont werden, weil die Satzklammern näher zusammen stehen. In Kapitel 6.5.1 wurde begründet, dass das mit den schwerer zu verarbeitenden Themen und selteneren Wörtern im medizinischen Genre zusammenhängt. Der Einfluss der Zeitstufe kann auch auf eine Veränderung innerhalb des Genres über die Jahrhunderte hinweisen, auf die bei der Betrachtung von Hypothese H3 eingegangen wird.

Während die Restriktivität des Relativsatzes im Gesamtkorpus nicht signifikant für Extraposition ist, zeigt sich im medizinischen Teilkorpus ein signifikantes Ergebnis ($p < 0,05$): Appositive Relativsätze werden mit höherer Wahrscheinlichkeit ausgelagert als restriktive, was der in Grammatiken vorgebrachten Tendenz zur Extraposition restriktiver Relativsätze widerspricht.

Bei den Interaktionen weist das Teilkorpus mit Ausnahme der Interaktion aus den Skipgram-Surprisalwerten und der Zeitstufe ($p < 0,05$) kaum signifikante Effekte auf. In jüngeren Texten scheint die höhere kognitive Kapazität für die Verarbeitung im Nachfeld nicht mehr so relevant zu sein. Einerseits kann das mit der Thematik und dem Stil der Autoren im 18. und 19. Jahrhundert zusammenhängen, die die oben beschriebene Verarbeitungserleichterung von Kopfnomen stärker nutzen, andererseits ist es möglich, dass der

³⁵ Interessant wäre es daher in späteren Arbeiten, zu überprüfen, ob Komplementsätze ebenfalls dieses Ergebnis erzielen würden.

Faktor der Distanz zwischen Antezedens und Relativsatz stärker zum Tragen kommt.

Die anderen beiden Interaktionen (Länge und Orality Score sowie Länge und Zeitstufe), die tendenziell Einbettung voraussagen ($p < 0,1$), lassen sich auch dadurch erklären, dass durch die Adjazenzstellung des Relativsatzes die Verarbeitung des Antezedens vereinfacht wird, da Erwartungen und Verarbeitung auch rückwärtsgerichtet angepasst werden können (vgl. Levy 2008). Die Interaktion des Orality Scores mit der Zeitstufe wurde bereits in den Kapiteln 6.5.1 und 6.5.2 besprochen. Nachfeldstellung ist in jüngeren Texten stärker mit Mündlichkeitsnähe assoziiert.

Theologisches Teilkorpus Auch im theologischen Teilkorpus ist der einflussreichste Faktor die Länge (Tabelle 6.28). Als Begründung für das Ergebnis werden die zuvor schon gegebenen Erklärungen bezüglich des verkürzten Satzrahmens herangezogen. Aber auch der Einfluss der Surprisalwerte kann als Haupteffekt nachgewiesen werden ($p < 0,01$). Allerdings besteht auch eine Interaktion aus Surprisal und Orality Score, die aber für Einbettung hochinformativer Relativsätze in mündlichkeitsnahen Texten spricht.

Wie auch beim medizinischen Subkorpus liefert das Ergebnis für die Surprisalwerte Evidenz für Hypothese H1. Es reiht sich also in die Argumentation ein, dass die freien Kapazitäten im Nachfeld dazu genutzt werden können, informative Relativsätze zu verarbeiten. Schwieriger mit dieser Überlegung zum Nachfeld zu vereinen, ist die Interaktion aus der Mündlichkeitsnähe und den hohen Skipgramwerten, die Einbettung erklärt. Bei der Betrachtung des Gesamtkorpus wurde dafür argumentiert, dass es bei theologischen Texten insgesamt von Vorteil sein kann, Relativsätze einzubetten, weil diese bei Extraposition an sich zu überraschend sein können, weil die Inhalte des Textes generell vertrauter und die zusätzlichen Informationen zur näheren Beschreibung der Kopfnomen daher weniger relevant sind (4).

- (4) Sie kann und soll aus der übernatürlichen Lebensquelle des neuen Adam, **welcher Christus ist**, neues, höheres Leben schöpfen [...] (Egger 1895: Satz 41)

Besonders bei Texten, die potenziell auch für orale Perzeption verständlich sein sollen, kann dieses Argument ausschlaggebend sein und auch die Interaktion zwischen der Länge und der Mündlichkeitsnähe ($p < 0,01$), die ebenfalls etwas mehr adjazente Relativsätze in sehr mündlichkeitsnahen Texten erkennen lässt, erklären. Zudem wurde schon argumentiert, dass die Verarbeitung der Satzklammern bei den theologischen Texten einen weniger starken Effekt zu haben scheint als in den medizinischen.

Weitere, aber schwächer signifikante p -Werte (jeweils $p < 0,05$) besitzen der Orality Score als Einzelfaktor, der für Auslagerung spricht, und der Typ des Relativsatzes. Letzteres ist wieder mit dem theologischen Genre zu erklären und damit, dass appositive Relativsätze im Fall von Extraposition zu überraschend sind, weil die darin gegebenen Sachverhalte nur zusätzliche Informationen liefern. Wieder verbinden sich hier vorwärts- und rückwärtsgerichtete Einflussfaktoren. Rückwirkend muss also das Kopfnomen in seiner Funktion als Antezedens erkannt werden, obwohl typische Hinweise, die für restriktive Relativsätze gefunden wurden, fehlen, was das Gedächtnis belastet. Ähnlich argumentiert Weber (2019) bei dem Unterschied in der Verarbeitung von der Extraposition von Präpositionalphrasen verglichen mit Relativsätzen. Zum anderen ist das generelle Auftreten des Relativsatzes an der Stelle nicht mehr zu erwarten, was für vorwärtsgerichtete Überraschung und Überlastung sprechen würde. In beiden Fällen wäre die erfolgreiche Kommunikation bei Extraposition gefährdet, weshalb Adjazenz zu bevorzugen ist.

Wieder ist hier auch die Interaktion aus Mündlichkeitsnähe und Zeitstufe signifikant ($p < 0,05$). Die verkürzte Distanz zwischen den Satzklammern wird also in mündlichkeitsnahen und jüngeren Texten auch dann relevant, wenn sie dem theologischen Genre angehören. Dieser Faktor liefert also auch hier einen Hinweis auf Sprachwandel (Kapitel 8).

Distanzen Warum die Distanz nicht von Anfang an in die Analyse eingefügt wurde, beschreibt Kapitel 4.3.2. Wenn die Distanz aber in die Regression aufgenommen wird, werden die Surprisalwerte ausgeschlossen (Tabelle 6.30).

Die Prädiktoren Länge, Distanz und Zeitstufe liefern gleich gute Vorhersagen über die Position der Relativsätze. Die Distanz und die Länge lassen

sich dabei auf die zu Beginn des Kapitels beschriebenen gedächtnisbasierten Verarbeitungsschwierigkeiten von Relativsätzen zurückführen. Lange Distanzen erschweren die Zuordnung des ausgelagerten Relativsatzes zu seinem Bezugswort und machen Vergessen wahrscheinlicher. Daher sprechen große Distanzen für Einbettung. Gleichzeitig vergrößert eine große Länge von Relativsätzen den Abstand zwischen den Satzklammern was ebenfalls für Verarbeitungsschwierigkeiten sorgt, weshalb lange Relativsätze trotzdem eher ausgelagert werden sollen. Generell wird Extraposition von Relativsätzen laut diesem Modell in jüngeren Zeitstufen wahrscheinlicher. Die Bedeutung der verkürzten Distanz zwischen Antezedens und Relativsatz gewinnt in moderneren Texten mehr an Bedeutung als in den älteren, was auch dafür spricht, dass die Entlastung des Gedächtnisses in jüngeren Texten bedeutender wird, worauf in Kapitel 8 genauer eingegangen wird. Die anderen signifikanten Werte unterscheiden sich nicht wesentlich von den Ergebnissen, die in Tabelle 6.24 gegeben wurden. Die Erklärung der Werte kann daher den vorherigen Kapitel entnommen werden.

Zusammenfassung Obwohl die Nullhypothese widerlegt wird, wenn die Distanzmessung aus der Regression ausgeschlossen wird, ist der Einfluss der Skipgram-Surprisalwerte auf Extraposition nur marginal und wird vor allem von der Länge der Relativsätze übertroffen. Auch hier zeigt sich aber auch ein Einfluss der Faktoren Genre und Mündlichkeitsnähe, die auf Adjazenz hinweisen. In Interaktionen ist zudem die Restriktivität der Relativsätze von Bedeutung. Sie spricht allerdings eher für Adjazenz, was die bisherige Forschung nicht vermutet.

Auch diese Ergebnisse sind damit zu erklären, dass zwischen der Funktion eines potenziell auszulagernden Elements und dessen lexikalischer Füllung unterschieden werden muss. Die Identifikationsfunktion des Bezugswort, die von den Relativsätzen erfüllt wird, kann erschwert werden, wenn das Auftreten des Relativsatzes generell unerwarteter wird. Das ist vor allem bei den appositiven Relativsätzen denkbar und erklärt, warum dieses Kriterium speziell bei den bekannteren theologischen Kopfnomen relevant ist. Auch hier zeigt sich also der Einfluss des allgemeinen Inhalts der untersuchten Tex-

te. Bei den medizinischen ist hingegen zu vermerken, dass auch die lexikalisch schwierige Verarbeitbarkeit der Relativsätze Einfluss auf ihre Position ausübt, weil – wie bei den Nominalphrasen beschrieben – die komplexeren Inhalte es erfordern, dass alle Verarbeitungskapazitäten optimal genutzt werden.

7. Ergebnisse zum Einfluss von Mündlichkeitsnähe

Die zweite Hypothese teilt sich in zwei Bausteine. Erstens soll überprüft werden, ob auch in diesem Korpus gezeigt werden kann, dass Extraposition in mündlichkeitsnahen häufiger als in schriftlichkeitsnahen Texten genutzt wird. Diese Vermutung liefert Hinweise darauf, dass die Nachfeldstellung tatsächlich mit besserer Verarbeitung assoziiert werden kann, weil diese für orale Texte noch wichtiger ist als für grafische. Zweitens soll überprüft werden, ob Surprisal in mündlichkeitsnahen Texten Extraposition stärker beeinflusst.

Dazu wird zunächst der Orality Score (vgl. Ortman und Dipper 2024) für jeden Text und der Anteil von Sätzen mit Extrapositionen an allen Sätzen bestimmt. Dann wird mithilfe einer Korrelationsanalyse der Zusammenhang zwischen beiden Werten überprüft. In einem anschließenden Schritt wurden die Texte in zwei Gruppen geteilt; eine eher mündlichkeitsnahe und eine eher schriftlichkeitsnahe. Für beide Gruppen wurde logistische Regression durchgeführt, die den Einfluss der Skipgram-Surprisalwerte auf Extraposition zeigen soll. Eine Trennung nach den verschiedenen ausgelagerten Phänomenen wurde hier nicht vorgenommen, da Extraposition trotz der Unterschiede zwischen Nominal- und Präpositionalphrasen sowie Relativsätzen als Ganzes betrachtet werden soll.

7.1. Mündlichkeitsnähe

7.1.1. Orality Scores im Korpus

Der Orality Score ist eine der unabhängigen Variablen, deren Einfluss in allen Modellen überprüft wird. Seine Berechnung erfolgt wie in Kapitel 5.3 beschrieben und inkludiert alle dort genannten Variablen. Eine Abwahl einzelner Faktoren wurde nicht getroffen und die Gewichtung der Faktoren erfolgte nach den in Ortman und Dipper (2024) vorgestellten Maßen. Um die Daten einfacher interpretierbar zu machen, wird an dieser Stelle wieder auf zwei Faktoren hingewiesen: Die Orality Scores sind keine absoluten Werte, weil konzeptionelle Mündlichkeit und konzeptionelle Schriftlichkeit keine festgelegten Endpunkte einer allgemeingültigen Skala sind, auf der sich einzelne Texte näher an dem einen oder näher an dem anderen Pol positionieren. Es ist nur möglich zu sagen, dass ein Text verglichen mit einem anderen Text mehr Merkmale aufweist, die für Mündlichkeitsnähe definiert wurden. Je höher der Orality Score eines Texts ist, desto größer ist der Anteil von Merkmalen der konzeptionellen Mündlichkeit.¹

Um bestimmen zu können, welche statistischen Verfahren mit diesem Wert durchgeführt werden dürfen, muss zunächst die Skala der Variable definiert werden (vgl. Gries 2021). Wie bereits erwähnt, sind die Texte nur miteinander und nicht gegen eine allgemein gültige Skala vergleichbar. Es gibt keinen festen Nullpunkt für Mündlichkeitsnähe. Die Orality Scores können also in eine bestimmte Reihenfolge gebracht werden, die einen Text als ‚am mündlichkeitsnächsten‘ und einen als ‚am schriftlichkeitsnächsten‘ bezeichnen sowie die zwischen den Polen stehenden Texte ordnen. Diese Eigenschaft würde den Orality Score auf einer Ordinalskala verorten (vgl. Gries 2021: 18).

Für die Texte des in dieser Arbeit verwendeten Korpus ergeben sich folgende Orality Scores, die in Tabelle 7.1 aufgeführt sind. Am Anfang der Tabelle stehen die schriftlichkeitsnahen Texte, am Ende die mündlichkeitsnahen.

¹ Obwohl die Formulierung ungenauer ist als die eben genannte, werden Texte mit einem hohen Orality Score dennoch als mündlichkeitsnah oder konzeptionell mündlich und solche mit einem niedrigen Orality Score als schriftlichkeitsnah beziehungsweise konzeptionell schriftlich bezeichnet.

Autor	Orality Score	Zeit	Genre
Carus	-0,631	1800-1850	Medizin
Ludwig	-0,552	1850-1900	Medizin
Koch	-0,471	1850-1900	Medizin
Strauss	-0,404	1800-1850	Theologie
Hasak	-0,333	1850-1900	Theologie
Gall	-0,281	1750-1800	Medizin
Kraepelin	-0,249	1850-1900	Medizin
Purmann	-0,038	1650-1700	Medizin
Reil	0,032	1800-1850	Medizin
Braeuner	0,096	1700-1750	Medizin
Daumer	0,120	1850-1900	Theologie
Hanssen	0,290	1700-1750	Theologie
Abel	0,338	1650-1700	Medizin
Rotth	0,459	1650-1700	Theologie
Pahl	0,461	1750-1800	Theologie
Niviandts	0,572	1700-1750	Theologie
Spener	0,599	1650-1700	Theologie
Bengel	0,627	1750-1800	Theologie
Benner	0,704	1700-1750	Theologie
Egger	0,732	1850-1900	Theologie
Unzer	0,792	1700-1750	Medizin
Loehe	1,097	1800-1850	Theologie

Tabelle 7.1.: Orality Scores im Korpus.

Der Text mit den meisten mündlichkeitsnahen Merkmalen stammt von Loehe. Er gehört dem Genre Theologie an und wurde im Zeitraum 1800 bis 1850 verfasst. Am anderen Ende der Skala steht ein Text von Carus, der ebenfalls zwischen 1800 bis 1850 entstanden ist, aber aus dem medizinischen Korpusenteil stammt. Diese Beobachtung weist bereits auf das hin, was auch die vierten Spalte von Tabelle 7.1 erkennen lässt: Die medizinischen Texte sind mit wenigen Ausnahmen eher schriftlichkeitsnah, während die theologischen Texte vermehrt mündlichkeitsnah sind. Der Eindruck wird durch die Berechnung des arithmetischen Mittels bestärkt. Im Durchschnitt ist der Orality Score der medizinischen Texte $-0,096$, der der theologischen Texte $0,41$.²

² Ob die theologischen Texte tatsächlich auch mehr im oralen Medium rezipiert wurden, lässt sich nicht mehr rekonstruieren. Es kann aber davon ausgegangen werden, dass es Zuhörenden leichter gefallen sein dürfte, die theologischen Texte hörend zu verarbeiten als die medizinischen.

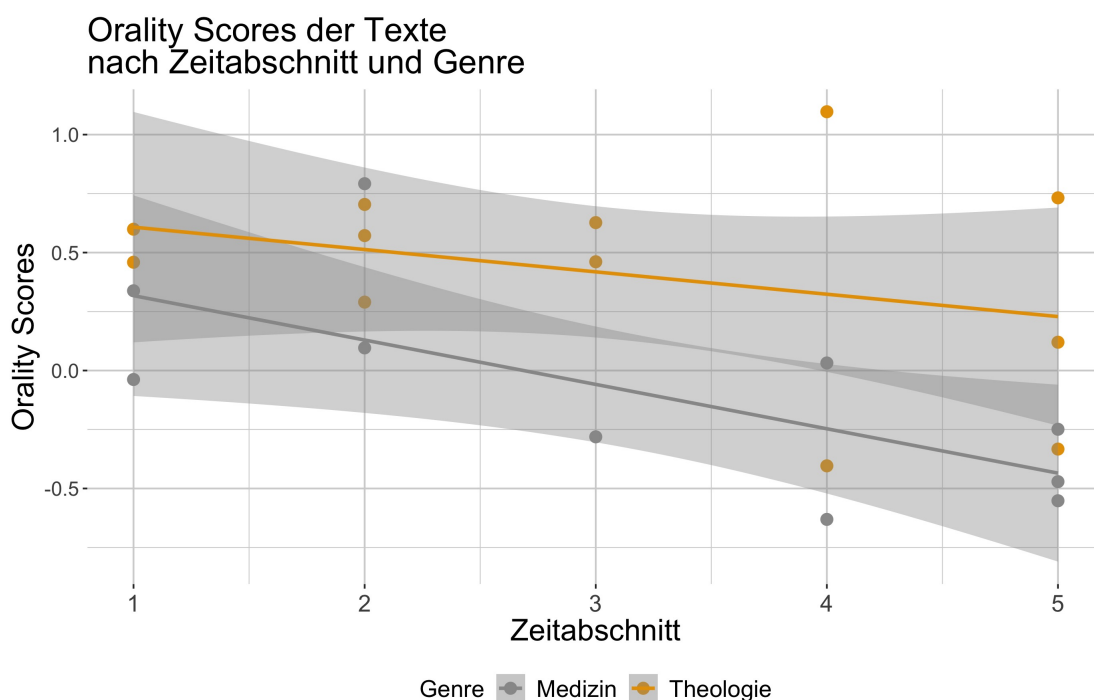


Abbildung 7.1.: Orality Scores unterteilt nach Zeit und Genre der Texte im Korpus. 1 steht für den Zeitabschnitt „1650-1700“, 2 für den Zeitabschnitt „1700-1750“ usw.

Den Unterschied zwischen den medizinischen und theologischen Texten stellt auch die Grafik 7.1 dar. Die gelben Punkte bilden die theologischen Texte ab und liegen in den meisten Zeitabschnitten über den Werten der medizinischen Texte, die in grau gehalten sind. Zusätzlich illustriert die Grafik, dass die Orality Scores über die betrachteten Jahrhunderte hinweg, die durch die Zahlen 1 bis 5 gekennzeichnet sind, wobei 1 für den Zeitraum von 1650 bis 1700 und 5 für den Zeitraum von 1850 bis 1900 steht, abnehmen. Sowohl die medizinischen als auch die theologischen Texte nähern sich immer mehr dem konzeptionell schriftlichen Bereich an. Die Tendenz ist für die theologischen Texte schwächer, wie an der flach abfallenden gelben Regressionslinie zu erkennen ist. Für die medizinischen Texte ist der Effekt hingegen etwas ausgeprägter, weil der Abfall der grauen Regressionslinie stärker ist. Sie setzt aber auch bereits bei einem niedrigeren Niveau im Zeitabschnitt 1650 bis 1700 an und entfernt sich über die Jahrhunderte weiter von der Regressionslinie der theologischen Texte. Die Veränderung hin zur Schriftlichkeitsnähe vollzieht sich in den medizinischen Texten also früher und deutlicher. Der zu beobach-

tende Effekt lässt sich auch anhand der durchschnittlichen Orality Scores pro Zeitabschnitt ablesen (Tabelle 7.2). Auch Tabelle 7.2 ist geordnet. Sie zeigt zuerst die schriftlichkeitsnahen Texte und dann die mündlichkeitsnahen Texte.

Zeit	Orality Score
1850-1900	-0,1255
1800-1850	0,0235
1750-1800	0,2690
1650-1700	0,3395
1700-1750	0,4908

Tabelle 7.2.: Orality Scores nach Zeitabschnitt.

Auch hier ist die Tendenz erkennbar: Den niedrigsten Orality Score erhalten die Texte aus dem Zeitabschnitt 1850 bis 1900, den höchsten Orality Score die Texte aus dem Abschnitt 1700 bis 1750. Im ersten Zeitabschnitt (1650-1700) hingegen sind die Texte tendenziell ein wenig schriftlichkeitsnäher.

Von dieser Abweichung der linearen Reihenfolge der Zeitabschnitte abgesehen, ist das Bild eindeutig. Je jünger ein Text ist, desto schriftlichkeitsnäher ist er. Als Erklärung dafür wird die veränderte Schreibpraxis herangezogen (Kapitel 1.1). Während die frühen wissenschaftlichen Texte noch dem Briefstil ähneln, der von Koch und Oesterreicher (2007) und Selting (2012) näher am Pol konzeptioneller Mündlichkeit verortet wird, wenden sich die späteren Texte mehr Prinzipien der Schriftlichkeit zu.

Aufgrund der Datenlage lässt sich nicht sagen, ob die Autoren der im Korpus verwendeten Texte die Konzeption mehr aufgrund einer Orientierung an die Rezipierenden oder sich selbst verfasst haben. Da alle Texte jedoch veröffentlicht wurden, ist davon auszugehen, dass die Autoren zumindest ein Publikum im Sinn hatten (vgl. Koch und Oesterreicher 2007). Durch die angenommene Rezipierendenorientierung ergeben sich in Kombination mit den höheren Mündlichkeitswerten der Texte einige Erwartungen hinsichtlich des Zusammenhangs zwischen den untersuchten Auslagerungen, die zur Beantwortung der zweiten Hypothese beitragen.

7.1.2. Mündlichkeitsnähe und Auslagerungsfrequenz

Zahlreiche Autoren (darunter Altmann 1981) betrachten Nachfeldbesetzung als Phänomen der Mündlichkeitsnähe. In Kapitel 1.1 wird vorgeschlagen zu untersuchen, ob die beobachteten Auslagerungen mit ihrem Vorkommen in mündlichkeitsnahen Texten zusammenhängen.³ Ein Weg, sich dem vor den jeweiligen Regressionsanalysen zu nähern, ist, sich die Frequenzen der Auslagerungen in Abhängigkeit des Orality Scores anzusehen (Grafik 7.2).

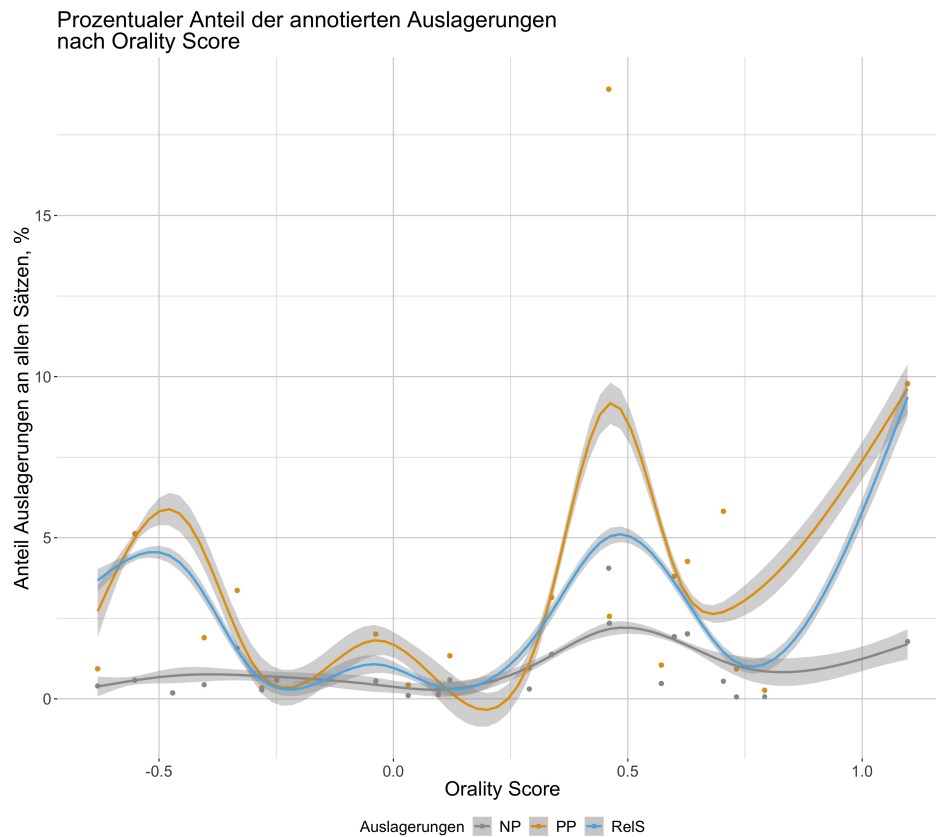


Abbildung 7.2.: Anzahl der Auslagerungen nach Orality Scores, relative Werte gemessen an der Gesamtzahl der Sätze im jeweiligen Text.

Auf der x-Achse sind die Orality Scores aus den Texten abgebildet. Die Punkte markieren die jeweiligen Anteile der extrapolierten Phrasen an allen Sätzen in dem Text, zu dem der Orality Score berechnet wurde. Graue Punkte stehen für Nominalphrasen, gelbe für Präpositionalphrasen und blaue für Relativsätze. Zudem sind in der Grafik die Korrelationslinien zwischen den ein-

³ Die zur Debatte stehende Hypothese lautet: (H2a) Auslagerung wird verstärkt in mündlichkeitsnahen Texten verwendet.

zelen Anteilen an Auslagerungen zu sehen. Die Linien dienen vor allem der visuellen Führung durch die Grafik, um die Zusammenhänge zwischen den Anteilen der Auslagerung und der Veränderung in den Orality Scores besser sichtbar zu machen. Die farbliche Kodierung entspricht der Auslagerungsart; grau steht für Nominalphrasen, gelb für Präpositionalphrasen und blau für Relativsätze. Die hellgrauen Bereiche um die Regressionslinien weisen auf das Konfidenzintervall hin (vgl. Wickham 2016).

Die Grafik gibt Folgendes zu erkennen: Der Anteil an Extrapositionen von Präpositionalphrasen und Relativsätzen ist sowohl bei den sehr schriftlichkeitsnahen als auch bei den sehr mündlichkeitsnahen Texten des hier verwendeten Korpus hoch. Es kann also schwerlich von einem linearen, allgemeinen Zusammenhang zwischen der Mündlichkeitsnähe der Texte und dem Anteil an Extrapositionen gesprochen werden. Dazu sind die Schwankungen innerhalb der Präpositionalphrasen und der Relativsätze zu groß. Lediglich der Verlauf der Regressionslinie der Nominalphrasen erlaubt den Schluss, dass es hier auch eine sichtbare Tendenz für einen lineareren Zusammenhang zwischen Nachfeldbesetzung und Mündlichkeitsnähe zu geben scheint.

Da die Orality Scores pro Text bestimmt wurden, spiegeln sich bei der Betrachtung der einzelnen Zusammenhänge zwischen dem Anteil der Extrapositionen in einem Text und dessen Orality Score auch Stilpräferenzen einzelner Autoren wider. Besonders am Ausschlag der Präpositionalphrasen, der die Kurve stark nach oben steigen lässt, ist dies sichtbar. Dennoch ist zu beobachten, dass der Anteil an Extrapositionen für alle drei untersuchten Phänomene mit zunehmender Mündlichkeitsnähe steigt. Die abgebildete Tendenz am äußersten Bereich der konzeptionellen Schriftlichkeit am linken Rand der x-Achse suggeriert zudem weniger ausgelagerte Nominalphrasen in schriftlichkeitsnahen Texten. In den Texten, die weniger klar einer Konzeption zugeordnet werden können, bestehen zudem mehr Freiheiten hinsichtlich der Position der untersuchten Phrasen und Relativsätze, worauf die Schwankungen im Bereich von -0,25 bis 0,5 hindeuten. In einem nächsten Schritt wird nun überprüft, ob sich lineare Tendenzen zwischen dem Anteil an Auslagerungen und der Konzeption des Textes erkennen lassen (Grafik 7.3).

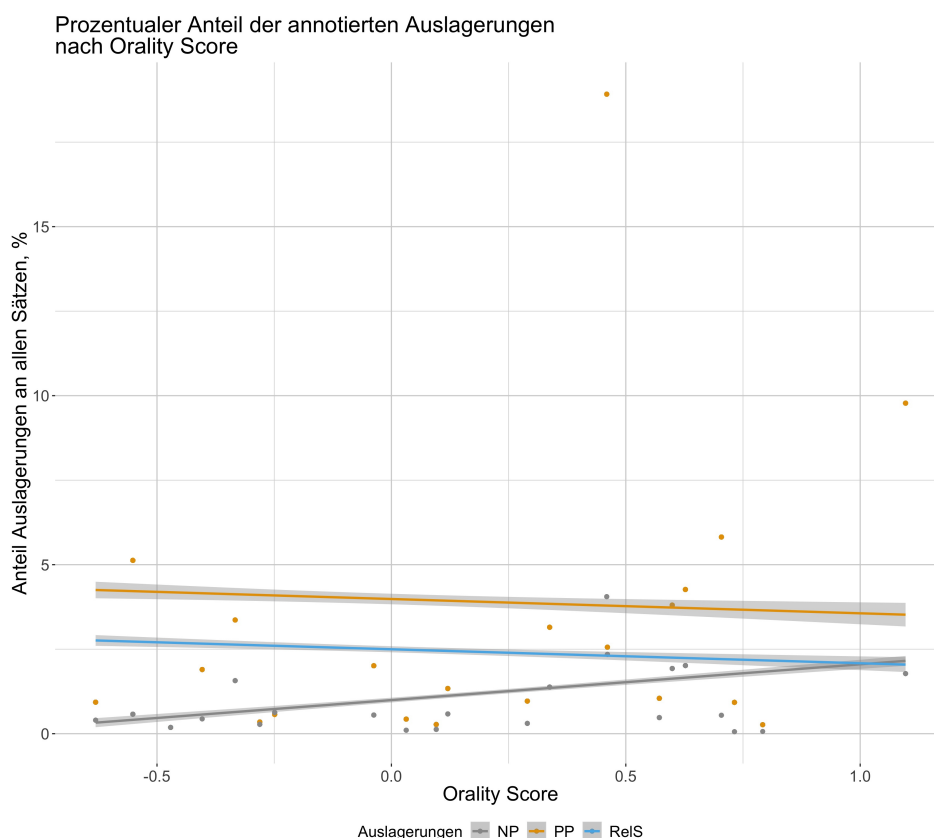


Abbildung 7.3.: Anzahl der Auslagerungen nach Orality Scores, relative Werte gemessen an der Gesamtzahl der Sätze im jeweiligen Text, inklusive linearer Regressionslinien.

Grafik 7.2 und Grafik 7.3 stellen die gleichen Informationen zur Verfügung, unterscheiden sich aber in den Regressionslinien. Während Grafik 7.2 die tatsächliche Regressionslinie zwischen den gefundenen Werten präsentiert, nehmen die hier gezeigten Linien einen linearen Zusammenhang zwischen den Orality Scores und dem Anteil an Auslagerungen an. Sie spiegeln eine mögliche Korrelation zwischen beiden Werten (vgl. Gries 2021: 131).

Die Regressionslinien der Präpositionalphrasen und der Relativsätze verlaufen fast parallel zur x-Achse. Dies deutet darauf hin, dass es keine Korrelation zwischen dem Anteil dieser Auslagerungsarten und der Mündlichkeitsnähe der Texte gibt, in denen sie gefunden wurden. In der Tat weist der Korrelationstest sowohl für die Präpositionalphrasen ($r=-0,26$), als auch für die Relativsätze ($r=-0,03$) keinen Zusammenhang auf, da beide Werte sehr nah an 0 sind (vgl. Gries 2021: 138). Besonders interessant ist an dieser Stelle das

Vorzeichen. Ein negatives Vorzeichen bedeutet, dass eine Variable zunimmt, die andere jedoch abnimmt (Gries (2021: 138f.), Winter (2020: 89)). Im Fall eines positiven Vorzeichens werden beide Variablen größer (vgl. Gries 2021; Winter 2020). Für den vorliegenden Fall bedeutet das, dass der Anteil der ausgelagerten Relativsätze und Präpositionalphrasen an allen Sätzen des jeweiligen Texts *geringer* wird, je mündlichkeitsnäher der entsprechende Text ist. Für beide Phänomene kann Hypothese H2a also nicht bestätigt werden.

Anders sieht es im Fall der Nominalphrasen aus. Die graue Regressionslinie in Grafik 7.3 weist eine Steigung auf, je mündlichkeitsnäher die Texte werden. Das bedeutet, der Anteil an ausgelagerten Nominalphrasen steigt, je mehr Merkmale konzeptioneller Mündlichkeit ein Text enthält. Auch der Korrelationseffizient ($r=0,6$) ist positiv und ausreichend, um im Fall dieser Phrasen Hypothese H2a bestätigen zu können.

7.1.3. Surprisal und Orality

Im vorherigen Kapitel wurde gezeigt, dass eine Zunahme der Extraposition in Abhängigkeit der Mündlichkeitsnähe eines Textes nicht generell zunimmt. Eine positive Korrelation wurde lediglich für die Nominalphrasen gefunden, was Evidenz für das Zutreffen von Hypothese H2a liefern kann. Neben der Frage nach dem Zusammenhang zwischen der Extrapositionshäufigkeit und dem Orality Score, befasst sich Hypothese H2 auch damit, ob sich eine Bedeutungszunahme für die Variable der Informationsdichte in mündlichkeitsnahen Texten finden lässt (Hypothese H2b). Dieser Frage kann sich auf zwei Arten genähert werden. Zum einen können die Interaktionen aus Skipgram-Surprisalwerten und den Orality Scores betrachtet werden. Das erfolgte in den Kapiteln 6.1, 6.2 und 6.3 in den Gesamt- und Subkorpora, sofern die Interaktion überhaupt signifikant für die Position der untersuchten Elemente ist. Zum anderen können auch Regressionsanalysen aller Phänomene getrennt nach Gruppen aus Orality Scores vorgenommen werden.

Dazu werden die Präpositionalphrasen, Nominalphrasen und Relativsätze zusammengenommen und über die `subset`-Funktion in R (vgl. R Core Team 2022) in zwei Gruppen geteilt. Die Grenze bildet der Orality Score 0,29, der

genau in der Mitte von Tabelle 7.1 liegt. Die Texte, die Orality Scores kleiner als 0,29 aufweisen, werden dem schriftlichkeitsnahen Pol zugeordnet. Die Texte, deren Orality Score größer als 0,29 ist, gelten als mündlichkeitsnah.⁴

Mündlichkeitsnahes Subkorporum Im mündlichkeitsnahen Subkorporum stehen 296 Nominalphrasen, 809 Präpositionalphrasen und 2143 Relativsätze, insgesamt fallen also 3248 Datenpunkte an. Von denen sind 1714 ausgelagert und 1534 eingebettet. Das theologische Genre überwiegt mit 2471 betrachteten Elementen, während dem medizinischen nur 777 Datenpunkte zugeordnet werden. Die Verteilung auf die Zeitstufen ist Tabelle 7.3 zu entnehmen.

Zeitstufe	1650-1700	1700-1750	1750-1800	1800-1850	1850-1900
Datenpunkte	1106	831	983	97	231

Tabelle 7.3.: Auslagerungen pro Zeitabschnitt aus den mündlichkeitsnahen Texten.

Das arithmetische Mittel der durchschnittlichen Skipgramwerte beträgt 3,68. Ihr kleinster Wert beläuft sich auf 2,898 und ihr höchster auf 4,801. 75 % der Werte sind kleiner als 3,79. Die Länge der Phänomene ist durchschnittlich 14,6 Tokens lang. Das Minimum der Werte ist ein Wort, das Maximum 225. Der Wert des dritten Quartils beträgt 18 Tokens.

In das Modell der logistischen Regression fließen, wie in Kapitel 5.4 beschrieben, die unabhängigen Variablen durchschnittliches Skipgram, logarithmierte Länge und das Genre und die Zeitstufe ein. Die abhängige Variable bildet natürlich die Position. Nach Durchführung der Modellauswahl ergibt sich folgendes Regressionsmodell (Tabelle 7.4).

Die Tabelle zeigt, dass die durchschnittlichen Skipgram-Surprisal einen signifikanten Effekt auf Auslagerung haben ($z=-2,3$, $p<0.05$). Dieser Wert bestätigt die Hypothese. Die betrachteten Texte liegen alle im mündlichkeits-

⁴ Das bedeutet, dass die Texte Purmann (1680), Bräuner (1714), Hanssen (1731), Gall (1791), Carus (1820), Reil (1803), Strauß (1835), Daumer (1895), Hasak (1893), Koch (1878) und Ludwig (1856) als konzeptionell schriftlichkeitsnah gezählt werden. Die Texte Abel (1699), Rotth (1692), Spener (1676), Benner (1739), Niviandts (1708), Unzer (1746), Bengel (1751), Pahl (1799), Löhe (1847) und Egger (1895) sind in das mündlichkeitsnahe Subkorporum eingeflossen .

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	2,97	0,69	4,29	<0,001	***
Skipgram _{mean}	-0,40	0,18	-2,23	0,03	*
Länge _{log}	-1,79	0,16	-11,09	<0,001	***
Genre	-0,04	0,29	-0,15	0,88	
Zeitstufe	-0,17	0,04	-4,46	<0,001	***
Länge _{log} :Genre	1,29	0,33	3,96	<0,001	***

Tabelle 7.4.: Logistische Regression aller Elemente in den mündlichkeitsnahen Texten.

nahen Spektrum und weisen dann extrapolierte Phrasen oder Relativsätze auf, wenn diese informativ sind. Allerdings weisen sowohl die logarithmierte Länge als auch die Zeitstufe höhere Signifikanzniveaus auf und sind daher bedeutendere Faktoren für die Positionierung der untersuchten Elemente. Die Länge ($z=-11,1$, $p<0,001$) weist dabei auf Auslagerung ebenso hin wie die Zeitstufe ($z=-4,46$, $p<0,001$). Lange Phrasen und Relativsätze werden also eher ins Nachfeld gestellt. Das gleiche gilt für die Phänomene, wenn sie in jüngeren Texten auftreten. Auch dann wird Nachfeldstellung wahrscheinlicher. Im Gegensatz dazu zeigt die Interaktion zwischen Länge und Genre ($z=3,96$, $p<0,001$), dass lange Phrasen in theologischen Texten sehr wahrscheinlich eingebettet werden.

Schriftlichkeitsnahes Subkorpus 6347 Datenpunkte wurden im schriftlichkeitsnahen Subkorpus gefunden. Sie teilen sich in 231 Nominalphrasen, 1133 Präpositionalphrasen und 4983 Relativsätze ein. 3086 Datenpunkte wurden ausgelagert, 3261 eingebettet. Auf die beiden Genre verteilt, finden sich hier, wie zu erwarten, 5340 Datenpunkte im medizinischen Genre, während nur 1007 Datenpunkte in den als theologisch klassifizierten Texten vorliegen. Die Verteilung der Annotationen auf die Zeitabschnitte zeigt Tabelle 7.5.

Zeitstufe	1650-1700	1700-1750	1750-1800	1800-1850	1850-1900
Datenpunkte	298	877	396	1250	3526

Tabelle 7.5.: Auslagerungen pro Zeitabschnitt aus den schriftlichkeitsnahen Texten.

Die durchschnittlichen Skipgramwerte variieren zwischen 2,79 und 4,72. Ihr Durchschnitt beträgt 3,625 und der Wert des dritten Quartils 3,728. Die Länge der Datenpunkte schwankt zwischen einem und 153 Tokens, mit einem Durchschnitt von 17,41 Tokens. 75 % der Daten sind kürzer als 22 Tokens.

Auch für die Elemente im schriftlichkeitsnahen Subkorpus wurde eine logistische Regression durchgeführt, die den oben beschriebenen Prinzipien folgt. Das daraus resultierende finale Modell ist in Tabelle 7.6 dargestellt.

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	4,02	0,89	4,53	<0,001	***
Skipgram _{mean}	-0,28	0,22	-1,26	0,21	
Länge _{log}	-3,04	0,35	-8,67	<0,001	***
Genre	4,04	1,64	2,46	0,01	*
Zeitstufe	-0,37	0,08	-4,63	<0,001	***
Skipgram _{mean} :Genre	-0,81	0,44	-1,82	0,07	.
Länge _{log} :Zeitstufe	0,38	0,08	4,95	<0,001	***
Genre:Zeitstufe	-0,33	0,05	-6,06	<0,001	***

Tabelle 7.6.: Logistische Regression aller Elemente in den schriftlichkeitsnahen Texten.

Im Gegensatz zur Regression für den mündlichkeitsnahen Text fällt hier auf, dass die durchschnittlichen Skipgram-Surprisalwerte der auszulagernden Elemente keinen signifikanten Einfluss auf ihre Position haben ($z=-1,26$, $p=0,21$). Dieses Ergebnis liefert bereits Evidenz für den in Hypothese H2b vorgeschlagenen Unterschied zwischen mündlichkeitsnahen und schriftlichkeitsnahen Texten. Die Länge ($z=-8,67$, $p<0,001$) sagt Auslagerung ebenso voraus wie die Zeitstufe ($z=-4,63$, $p<0,001$). Es besteht zudem eine Interaktion aus Genre und Zeitstufe ($z=-6,1$, $p<0,001$). Das bedeutet, dass lange Phrasen und Relativsätze in jüngeren und theologischen Texten mit hoher Wahrscheinlichkeit im Nachfeld zu finden sind, in medizinischen Texten werden sie hingegen eher eingebettet. Die Interaktion aus Länge und Zeitstufe ($z=4,95$, $p<0,001$) sagt aus, dass in jüngeren Texten lange Phrasen eher eingebettet werden. Weitere signifikante Ergebnisse liefert das Genre der Texte ($z=2,46$, $p<0,05$), das besagt, dass Einbettung in theologischen Texten wahrscheinlicher wird, und die Interaktion zwischen Genre und durchschnittlichem Skipgram-Surprisal ($z=-$

1,82, $p < 0,1$). Wenn das Element in diesen Texten ein hohes durchschnittliches Skipgram-Surprisal aufweist, steht es tendenziell eher im Nachfeld. In den medizinischen Texten scheint die Frequenz von Einbettung und Auslagerung hingegen von den Surprisalwerten unbeeinflusst zu sein.

7.2. Diskussion Hypothese 2: Mündlichkeitsnähe und Extraposition

Die zweite Hypothese teilt sich in zwei Teilhypothesen.

- (H2a) Auslagerung wird verstärkt in mündlichkeitsnahen Texten verwendet.
- (H2b) In Texten, die mündlichkeitsnäher sind, ist der Einfluss von hohen Surprisalwerten größer als in schriftlichkeitsnahen Texten.

Der erste Teil ist eher deskriptiver Natur und soll überprüfen, ob dieses Postulat auch im vorliegenden, eindeutig medial schriftlichen Korpus gilt. Die zweite ist mehr auf die allgemeine Hypothese der Verarbeitbarkeit bezogen. Man sollte erwarten, dass in oraler und auch konzeptionell mündlicher Kommunikation besonders großer Wert darauf gelegt wird, kognitive Kapazitäten nicht überzustrapazieren. Ob das der Fall ist, wurde mithilfe der Regressionsanalyse in Kapitel 7.1.3 betrachtet. Hier sollen nun das Ergebnis diskutiert werden, dass Mündlichkeitsnähe tatsächlich einen größeren Einfluss von Informationsdichte auf Auslagerung ausübt.

7.2.1. Hypothese 2a: Häufigkeit von Extraposition

Wie zuvor bereits beschrieben, wird in der ersten Hypothese zur Mündlichkeitsnähe untersucht, ob Extraposition verstärkt in mündlichkeitsnahen Texten verwendet wird. Ein wenig problematisch an dieser Hypothese ist die Tatsache, dass die beiden Genres stark mit den verschiedenen Ausprägungen von Mündlichkeitsnähe korrelieren (Tabelle 7.1). Von den schriftlichkeitsnahen Texten zählen nur drei von elf zum theologischen Genre. Bei den mündlichkeitsnahen Texten finden sich sogar nur zwei medizinische Texte, die Merkmale von konzeptionell mündlichkeitsnahen Texten aufweisen. Zu-

dem muss darauf hingewiesen werden, dass bei den drei schriftlichkeitsnahen Texten zwei jüngeren Datums sind und sie auch den Text von Hasak inkludieren, der mehr architektonisch als theologisch ist. Vollkommen losgelöst vom Genre lassen sich die Konzeptionen der Texte also nicht betrachten. Das bietet aber auch einen Hinweis auf den Unterschied zwischen natur- und geisteswissenschaftlichen Texten.

Bei den theologischen Texten muss von einem Einfluss der Predigten, die die Theologen häufig verfasst haben, ausgegangen werden. In diesen wurden religiöse Inhalte für die breite Bevölkerung verständlich gemacht. Die Predigten und auch andere religiöse Texte wie Traktate oder Lebensberichte religiöser Personen mussten, um viele Menschen erreichen zu können, auch bei oraler Rezeption verstanden werden, weil die Analphabetismusquote im untersuchten Zeitraum noch hoch war (vgl. z. B. Elspaß 2005). Diese Hörerorientierung bedingt also, dass die Texte auch konzeptionell mündlichkeitsnäher verfasst wurden.

Zudem ist zu vermuten, dass die Theologen nicht in Sprachreflexion geschult waren, sodass sie in ihren wissenschaftlich(er)en Publikationen nicht von diesem Stil zu einem schriftlichkeitsnäheren gewechselt sind. Außerdem haben theologische Themen im untersuchten Zeitraum eine viel größere Bedeutung im Leben der Menschen eingenommen, als das heutzutage der Fall ist. Daher waren die verwendeten, speziell mit Religion in Verbindung zu setzenden Wörter auch bekannter und belasteten das Gedächtnis weniger stark. Zudem können zwar verschiedene Bereiche religiösen Lebens in den Texten verarbeitet werden, aber auch die waren weniger starken Veränderungen über die Jahrhunderte unterworfen. Infolgedessen ist auch die Notwendigkeit neuer Wörter oder der Darstellung neuer, komplexer Sachverhalte in diesen Werken unwahrscheinlicher. Außerdem ist anzunehmen, dass die Theologen auch mehr mit dem Lateinischen vertraut waren, sodass auch eine Gewöhnung an lange Dependenz (vgl. z. B. Gulordava und Merlo 2015) durch den Kontakt mit dem Lateinischen angenommen werden kann. Beachtet man allerdings die Erkenntnisse von Frank et al. (2016), dürfte den Autoren die unterschiedliche Grammatik dennoch bewusst gewesen sein und

der Einfluss des Lateinischen weniger stark ins Gewicht fallen und weniger zufällig genutzt worden sein.

Dieser Argumentation folgend prägt der Einfluss der rezipierten (lateinischen) Texte die Sprachpraxis weniger als das Verfassen der eigenen Texte und eine Gewöhnung an diesen Stil. Die gleichen Argumente können auch für die gegenteiligen Beobachtungen im medizinischen Teilkorpus herangezogen werden. Auch damals gehörte das Verfassen von Texten, die für die Allgemeinheit verständlich waren, nicht zu den Kernaufgaben von Medizinern. Ihr Adressatenkreis bestand aus anderen Medizinern, bei denen eine hohe Analphabetenquote ausgeschlossen werden kann. Eine Hörerorientierung, die diesen Faktor aufgreift, kann hier folglich ausgeschlossen werden. Dazu kommen die größeren Veränderungen und Entdeckungen in der Medizin (vgl. z. B. Eckart 2021), die neue Wörter notwendig machen und aufgrund ihrer Neuheit auch schwerer zu verarbeiten waren.

Trotz dieser Unterschiede, die in gewissem Maß auch über den ganzen untersuchten Zeitraum präsent zu bleiben scheinen, offenbart Grafik 7.1 in beiden Genres eine Entwicklung hin zum schriftlichkeitsnahen Diskurs. Es kann davon ausgegangen werden, dass sich der schon bestehende, schriftlichkeitsnahe Stil der Mediziner festigt und erste Hinweise auf die heute vorliegenden Eigenschaften wissenschaftlichen Schreibens liefert, die aber systematisch und empirisch meines Wissens nicht gut untersucht sind. Stärker macht sich aber bemerkbar, dass sich auch die theologischen Texte einem schriftlichkeitsnahen Stil annähern. Ein Argument dafür ist die zunehmende Professionalisierung der Ausbildung von Theologen und ein gesteigertes Bewusstsein für verschiedene Schreibanlässe und Stile. Auch die Tatsache, dass sich die Geisteswissenschaften in den letzten untersuchten Zeitabschnitten als eigenständiger wissenschaftlicher Bereich herauszubilden begannen (vgl. z. B. Dilthey 1910), kann darauf hinweisen, dass eine Entwicklung stattgefunden hat, die sich auch im Stil (geistes-)wissenschaftlicher Texte widerspiegelt.

Bezug zu Auslagerung und Mündlichkeitsnähe Unabhängig von der Zuordnung zum Genre kann die Untersuchung der Häufigkeiten der verschiedenen Extrapolationen in mündlichkeitsnahen Texten untersucht werden. Für

Präpositionalphrasen ($r=-0,26$) und Relativsätze ($r=-0,03$) ist kein Zusammenhang zwischen Orality Score und Anzahl der Auslagerungen zu erkennen. Das negative Vorzeichen suggeriert sogar eher eine Tendenz zur Einbettung für die beiden Phänomene, die auch von den Regressionsanalysen bestätigt wird. Zugleich sind die Korrelationskoeffizienten aber so nah an 0, dass kein Zusammenhang belegt wird. Nur die Nominalphrasen weist eine positive, relevante Korrelation ($r=0,6$) zwischen Auslagerung und Mündlichkeitsnähe eines Textes auf. Nur für sie kann Hypothese H2a also bestätigt werden. Warum die Auslagerung von Nominalphrasen die für mündlichkeitsnahen Texte relevanten Verarbeitungsvorteile, vor allem bezogen auf die Satzgliedfunktionen der Nominalphrasen, die sich essenziell von denen der Präpositionalphrasen und Relativsätze unterscheidet, bringen kann, wurde bei der Analyse der Nominalphrasen in Kapitel 6.5.1 beschrieben.

7.2.2. Hypothese 2b: Mündlichkeitsnähe und Surprisal

Hypothese H2b fragt, ob der Einfluss von Surprisalwerten auf Extraposition in mündlichkeitsnahen Texten größer ist als in schriftlichkeitsnahen Texten. Dazu wurden zwei Regressionen mit der abhängigen Variable Position und anhand aller annotierten Datenpunkte durchgeführt. Es wurde nicht zwischen Nominal-, Präpositionalphrasen und Relativsätzen unterschieden.

Tabelle 7.4 zeigt die Ergebnisse der Regression der mündlichkeitsnahen Texte, Tabelle 7.6 die des schriftlichkeitsnahen Subkorpus. Die Betrachtung der Surprisalwerte bestätigt die Hypothese. Die durchschnittlichen Skipgram-Surprisalwerte sind in den mündlichkeitsnahen Texten signifikant ($p<0,05$), aber nicht in den schriftlichkeitsnahen ($p=0,20$). In mündlichkeitsnahen Texten kann die Position der hier untersuchten Elemente also mithilfe der Skipgramwerte erklärt werden, in den schriftlichkeitsnahen Texten ist das nicht der Fall. Problematisch ist, dass mit dieser Methode nicht gezeigt werden kann, ob der Einfluss sukzessive abnimmt.

Bei der Betrachtung der beiden Regressionen ist aber auch hervorzuheben, dass in beiden Fällen andere Faktoren, vornehmlich die Länge und die Zeitstufe (je $p<0,001$), mehr Datenpunkte erklären sind als die Skipgramwer-

te. Auch das zeigt, dass im Deutschen die Schonung von Gedächtnisressourcen bei der Verarbeitung und das damit verbundene Verkürzen des Satzrahmens von größerer Bedeutung ist als die Verarbeitungsschwierigkeiten, die sich durch Surprisalwerte vorhersagen lassen. Zusätzlich stützen auch diese Ergebnisse den Eindruck, dass über die Jahrhunderte ein Wechsel in der Schreibpraxis zwischen theologischen und medizinischen Texten stattgefunden hat, der im folgenden Kapitel behandelt wird.

8. Ergebnisse zur diachronen Entwicklung von Extraposition

Die dritte Hypothese der vorliegenden Arbeit untersucht, ob der Einfluss von Informationsdichte auf Auslagerungen über die Zeit geringer wird. Speyer (2015a) und Schildt (1976) finden in ihren Untersuchungen einen abnehmenden Einfluss von Informationsstruktur auf Extraposition. Informationsstruktur und Informationsdichte lassen sich assoziieren, weshalb vermutet wird, dass sich die Einflussabnahme auch auf Informationsdichte übertragen lässt. In dem in Kapitel 6.5 angedeuteten Rahmen würde das also bedeuten, dass die vorwärtsgerichteten, vorhersagebasierten Verarbeitungsansätze weniger bedeutsam werden und von anderen, beispielsweise den gedächtnisbasierten, Ansätzen abgelöst werden.

Dazu wurden die Korpora der Nominalphrasen, Präpositionalphrasen und Relativsätze in die fünf untersuchten Zeitabschnitte unterteilt. Für jeden Zeitabschnitt und jedes Phänomen wurde dann eine eigene Regressionsanalyse durchgeführt. Eine Beschreibung der wichtigsten Ergebnisse der Regressionsanalysen erfolgt in den Kapiteln 8.1, 8.2 und 8.3. In Tabelle 8.23 sind die Ergebnisse dieser Regressionsanalysen zusammengefasst.

8.1. Nominalphrasen in den verschiedenen Zeitabschnitten

In Kapitel 6.1 wurde das Gesamtkorpus und die Teilkorpora bestehend aus den medizinischen und theologischen Texten analysiert. Bei der Betrachtung des gesamten Korpus (Tabelle 6.5) hat sich gezeigt, dass die Zeitstufe kein

signifikanter Haupteffekt für die Position der Nominalphrase ist. Wie zu erwarten, wird es also nicht wahrscheinlicher, ausgelagerte Nominalphrasen zu sehen, wenn der betrachtete Text jüngeren Datums ist. Die Häufigkeit extraponyierter Nominalphrasen sagt jedoch nichts über den Einfluss von Informationsdichte auf Auslagerung in den entsprechenden Zeitstufen aus.

Die Tabellen 8.1 und 8.2 zeigen die deskriptiven Werte der Länge und der Skipgram-Surprisalwerte in den verschiedenen Zeitabschnitten. In allen Zeitabschnitten ist die durchschnittliche Länge der ausgelagerten Nominalphrasen etwas länger als bei den eingebetteten Phrasen. Besonders deutlich wird der Unterschied von 1750 bis 1800. Hier sind die ausgelagerten Phrasen im Schnitt mehr als doppelt so lang wie die eingebetteten, was allerdings mit der großen Variation der Werte im vierten Quartil zu erklären ist. Auch im 19. Jahrhundert ist der Unterschied der durchschnittlichen Länge der Nominalphrasen hervorzuheben.

Die durchschnittlichen Skipgramwerte ähneln einander erneut stark. Die Differenz zeigt sich in der Regel erst hinter dem Komma, von 1700 bis 1750 und von 1850 bis 1900 sogar erst in der zweiten Nachkommastelle. Um den Einfluss der verschiedenen Positionen erfassen zu können, werden fünf logistische Regressionsanalysen pro Phänomen durchgeführt. Die abhängige Variable ist die Position. Die unabhängigen Variablen werden von den durchschnittlichen Skipgram-Surprisalwerten, der Länge, dem Orality Score und dem Genre gebildet. Die Zeitstufe muss ausgeschlossen werden, weil sie jeweils nur noch in einer Ausprägung vorliegt. Für die numerischen Variablen wurde mittels visueller Inspektion ermittelt, ob sie logarithmiert werden müssen. Das Verfahren wurde beispielsweise in Kapitel 6.1.1 und in Kapitel 5.4 ausführlich beschrieben.

Zeitabschnitt	Position	Faktor	Minimum	Median	Durchschnitt	drittes Quartil	Maximum
1650-1700	in situ	Länge	1	6	14,76	15	108
1650-1700	in extraponiert	Länge	1	10	19,38	23,25	132
1700-1750	in situ	Länge	2	6	6,8	8	18
1700-1750	extraponiert	Länge	2	9,50	10,22	13,25	30
1750-1800	in situ	Länge	1	4	6,37	15	22
1750-1800	extraponiert	Länge	2	8	16,68	20	144
1800-1850	in situ	Länge	4	9	10,2	14	20
1800-1850	extraponiert	Länge	1	13,50	15,45	20,25	57
1850-1900	in situ	Länge	2	12	16,11	24	48
1850-1900	extraponiert	Länge	2	18	24,78	233	102

Tabelle 8.1.: Deskriptive Statistik der Länge der annotierten Nominalphrasen, in Zeitabschnitte eingeteilt.

Zeitabschnitt	Position	Faktor	Minimum	Median	Durchschnitt	drittes Quartil	Maximum
1650-1700	in situ	Surprisal	2,904	3,712	3,721	3,886	4,519
1650-1700	extraponiert	Surprisal	3,024	3,789	3,804	3,918	4,530
1700-1750	in situ	Surprisal	3,373	3,681	3,713	3,782	4,263
1700-1750	extraponiert	Surprisal	3,321	3,713	3,736	3,884	4,286
1750-1800	in situ	Surprisal	2,898	3,609	3,674	3,853	4,445
1750-1800	extraponiert	Surprisal	3,152	3,752	3,803	3,895	4,801
1800-1850	in situ	Surprisal	3,251	3,572	3,593	3,723	3,894
1800-1850	extraponiert	Surprisal	3,040	3,617	3,661	3,782	4,247
1850-1900	in situ	Surprisal	2,980	3,624	3,619	3,726	4,158
1850-1900	extraponiert	Surprisal	2,993	3,646	3,676	3,745	4,720

Tabelle 8.2.: Deskriptive Statistik der Skipgram-Surprisalwerte der annotierten Nominalphrasen, in Zeitabschnitte eingeteilt.

Extrapolation von Nominalphrasen im Zeitraum 1650 bis 1700 Im ersten Zeitabschnitt (1650 bis 1700) wurden 98 extrapolierte und 85 eingebettete Nominalphrasen annotiert. Die numerischen Variablen mussten nicht transformiert werden. Die rückwärtsgerichtete Modellauswahl mit anova (vgl. R Core Team 2022) führt zum in Tabelle 8.3 dargestellten Modell.

	Estimate	Std.-Fehler	z-Wert	p-Wert
Intercept	18,22	1688,43	0,01	0,99
Skipgram _{mean}	-11,93	6,14	-1,94	0,05 .
Länge	-1,36528	0,79	-1,74	0,08 .
Orality Score	-48,39	2818,83	-0,02	0,99
Genre	-42,18	3376,48	-0,01	0,99
Skipgram _{mean} :Länge	0,05	0,05	0,86	0,39
Skipgram _{mean} :Orality Score	24,17	13,11	1,84	0,07 .
Skipgram _{mean} :Genre	-6,40	3,92	-1,63	0,10
Länge:Orality Score	2,27	1,63	1,39	0,16
Länge:Genre	-0,35	0,45	-0,79	0,43
Orality Score:Genre	116,23	5636,49	0,02	0,98

Tabelle 8.3.: Logistische Regression der Nominalphrasen des Zeitabschnitts 1650 bis 1700.

Die Hauptprädiktoren der durchschnittlichen Skipgramwerte ($z=-1,94$, $p<0,1$) und der Länge ($z=-1,7$, $p<0,1$) weisen eine Tendenz zur Auslagerung auf. Hochinformativ und lange Nominalphrasen werden demnach mit hoher Wahrscheinlichkeit ins Nachfeld gestellt. Die meisten Interaktionen in diesem Modell sind nicht signifikant. Lediglich die zwischen den durchschnittlichen Skipgramwerten und dem Orality Score weist ein signifikantes Ergebnis für Einbettung auf ($z=1,8$, $p<0,1$). Sehr informative Phrasen werden in sehr mündlichkeitsnahen Texten also eher eingebettet, während informative Phrasen aus Texten mit anderen Orality Scores eher ausgelagert stehen.

Extrapolation von Nominalphrasen im Zeitraum 1700 bis 1750 Im Zeitraum von 1700 bis 1750 wurden nur 32 extrapolierte Nominalphrasen und 25 eingebettete Nominalphrasen annotiert, weshalb die Ergebnisse mit Vorsicht zu betrachten sind. Nur für die Variable der Länge ist eine Logarithmierung

notwendig. Mithilfe der rückwärtsgerichteten Modellauswahl mit anova (vgl. R Core Team 2022) wurde das finale Modell (Tabelle 8.4) ermittelt.

	Estimate	Std.-Fehler	z-Wert	p-Wert
Intercept	2,46	4,51	0,55	0,59
Skipgram _{mean}	-0,27	1,195	-0,22	0,82
Länge _{log}	-2,04	1,05	-1,94	0,05

Tabelle 8.4.: Logistische Regression der Nominalphrasen des Zeitabschnitts 1700 bis 1750.

Zwei Dinge sind hier zu bemerken: Die einzigen Variablen, die im Modell übrigblieben, sind intralinguistisch mit Verarbeitung assoziiert. Zudem weisen die beiden Estimates auf Auslagerung hin, obwohl nur der Prädiktor der Länge signifikant für Auslagerung ist ($z=-1,94$, $p<0,1$). Lange Phrasen werden eher extrapoliert. Dahingehend unterscheidet sich dieses Modell schon von dem von 1650 bis 1700.

Extrapolation von Nominalphrasen im Zeitraum 1750 bis 1800 Im dritten zu betrachtenden Zeitabschnitt (1750 bis 1800) ist der Unterschied in der Anzahl aus extrapolierten und eingebetteten Phrasen am größten. 71 ausgelagerte Phrasen wurden bestimmt, während nur 21 eingebettete annotiert wurden. Keine der Variablen muss transformiert werden. Die rückwärtsgerichtete Modellauswahl resultiert in dem in Tabelle 8.5 präsentierten Modell.

	Estimate	Std.-Fehler	z-Wert	p-Wert
Intercept	6,05	2,79	2,17	0,03 *
Länge	-2,23	1,15	-1,93	0,05 .
Genre	30,63	14,583	2,10	0,04 *
Orality Score	-35,04	16,12	-2,17	0,03 *
Länge:Genre	-11,237	6,04	-1,86	0,06 .
Länge:Orality Score	12,40	6,65	1,87	0,06 .

Tabelle 8.5.: Logistische Regression der Nominalphrasen des Zeitabschnitts 1750 bis 1800.

Der Ausschluss der Skipgram-Surprisalwerte weist bereits darauf hin, dass Hypothese H1 nicht bestätigt werden kann. In diesem Zeitabschnitt zäh-

len das Genre ($z=2,1$, $p<0,05$) und der Orality Score ($z=2,17$, $p<0,05$) zu den bedeutendsten Hauptprädiktoren. Gehört ein Text in der zweiten Hälfte des 18. Jahrhunderts dem theologischen Genre an, werden Nominalphrasen eher eingebettet. Ist der Text mündlichkeitsnah, erhöht sich die Wahrscheinlichkeit für Auslagerung. Ebenfalls für Extraposition spricht die Variable der Länge ($z=-1,9$, $p<0,1$). Lange Phrasen werden eher ausgelagert.

Die Interaktion zwischen Länge und Genre ($z=-1,86$, $p<0,1$) weist auf Extraposition langer Phrasen in theologischen Texten und auf Einbettung langer Phrasen in medizinischen Texten hin. Auch in theologischen Texten werden lange Phrasen eher ausgelagert. Anders sieht es in mündlichkeitsnahen Texten aus. Die Interaktion aus Länge und dem Orality Score ($z=1,86$, $p<0,1$) sagt Einbettung von langen Phrasen in konzeptionell mündlichen Texten voraus, während lange Phrasen in konzeptionell schriftlichen Texten eher eingebettet werden.

Extraposition von Nominalphrasen im Zeitraum 1800 bis 1850 Wie im Zeitraum von 1700 bis 1750 ist auch in diesem Zeitabschnitt (1800 bis 1850) die Aussagekraft der Betrachtungen eingeschränkt. Insgesamt wurden nur 30 Nominalphrasen annotiert. 20 davon befinden sich im Nachfeld, zehn sind eingebettet. Die numerischen Variablen müssen nicht logarithmiert werden und die rückwärtsgerichtete Modellauswahl erlaubt den Ausschluss aller Interaktionen. Das daraus resultierende Modell wird in Tabelle 8.6 dargestellt.

	Estimate	Std.-Fehler	z-Wert	p-Wert
Intercept	0,09	0,79	0,11	0,91
Länge	-0,11	0,06	-1,75	0,08 .
Orality	-2,47	1,32	-1,87	0,06 .

Tabelle 8.6.: Logistische Regression der Nominalphrasen des Zeitabschnitts 1800 bis 1850.

Sowohl die Länge ($z=-1,75$, $p<0,1$) als auch der Orality Score ($z=-1,87$, $p<0,1$) sind signifikant für Auslagerung. Das bedeutet, dass längere Nominalphrasen mit höherer Wahrscheinlichkeit ausgelagert werden. Das gleiche gilt für Nominalphrasen, die aus mündlichkeitsnahen Texten stammen. Zum

einen scheint sich die Entwicklung fortzusetzen, dass die Skipgramwerte an Bedeutung verlieren. Evidenz für Hypothese H1 lässt sich daher nicht finden. Gleichzeitig deutet das darauf hin, dass Hypothese H3 bestätigt werden kann.

Extrapolation von Nominalphrasen im Zeitraum 1850 bis 1900 Von 1850 bis 1900 wurden 85 ausgelagerte und 74 eingebettete Nominalphrasen annotiert. Um die Regressionsanalyse erfolgreich durchführen zu können, wurden die Skipgram-Surprisalwerte logarithmiert. Das finale Modell findet sich in Tabelle 8.7.

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	-5,37	7,26	-0,74	0,46	
Skipgram _{mean, log}	8,31	12,62	0,66	0,51	
Länge	-0,05	0,02	-3,40	0,0007	***
Orality	-0,48	0,98	-0,49	0,62	
Genre	-29,34	14,51	-2,02	0,04	*
Skipgram _{mean, log} :Genre	52,13	25,23	2,07	0,04	*
Orality:Genre	6,56	1,93	3,398	0,0007	***

Tabelle 8.7.: Logistische Regression der Nominalphrasen des Zeitabschnitts 1850 bis 1900.

Das Modell zeigt zwei signifikante Hauptprädiktoren. Der bedeutendste ist die Länge ($z=-3,4$, $p<0,001$). Hier bestätigen sich die Ergebnisse aus früheren Studien (Kapitel 3.1.2). Nominalphrasen von großer Länge werden ins Nachfeld gestellt. Auf Nachfeldstellung in theologischen Texten deutet der signifikante Wert des Genres hin ($z=-2,02$, $p<0,5$).

Die Interaktion zwischen den Skipgram-Surprisalwerten und dem Genre ($z=2,07$, $p<0,05$) zeigt: Wenn eine Nominalphrase mit einem hohen, durchschnittlichen Skipgram-Surprisal in einem theologischen Text vorkommt, steht die Phrase eher eingebettet. In die Richtung weist auch die Interaktion zwischen dem Orality Score und dem Genre ($z=3,39$, $p<0,001$). Die Zugehörigkeit einer Phrase zu einem mündlichkeitsnahen Text, der gleichzeitig theologisch ist, macht Einbettung wahrscheinlicher.

Diese Ergebnisse festigen das Bild, dass die Bedeutung der Informationsdichte für die Position von Nominalphrasen über die Jahrhunderte hinweg

abnimmt. Nur im ersten Zeitabschnitt (1650 bis 1700) ergab die Betrachtung ein signifikantes Ergebnis für den Zusammenhang zwischen Nominalphrasen mit hohen Surprisalwerten und Extraposition. Seit dem 18. Jahrhundert ist dieser Prädiktor nicht mehr signifikant und wird vor allem vom Faktor Länge abgelöst. Dieser Wechsel spricht dafür, dass Hypothese H3 zutreffend ist und tatsächlich eine Veränderung über die Jahrhunderte hinweg stattfindet. Ob sich das auch in bei der Betrachtung der Präpositionalphrasen zeigen wird, die sich nicht analog zu den Nominalphrasen bei der Betrachtung des Gesamtkorpus und der Teilkorpora verhalten (Kapitel 6.2), klärt der nächste Abschnitt.

8.2. Präpositionalphrasen in den verschiedenen Zeitabschnitten

Dieses Unterkapitel befasst sich mit der Analyse der Teilkorpora der Präpositionalphrasen, die in die fünf Zeitabschnitte unterteilt wurden. An dieser Stelle werden zunächst wie im Kapitel zu den Nominalphrasen die deskriptiven Werte der fünf Zeitabschnitte gegeben (Tabellen 8.8 und 8.9).

Der Trend der Beobachtungen aus den Teilkorpora der Nominalphrasen setzt sich fort. Die Unterschiede zwischen der Länge (Tabellen 8.8) der eingebetteten und ausgelagerten Präpositionalphrasen ist deutlich sichtbar in den Kategorien Median, Durchschnitt und drittes Quartil. Die eingebetteten Präpositionalphrasen sind im Schnitt halb so lang wie die ausgelagerten. Daher ist zu vermuten, dass sich auch hier vor allem ein starker Effekt der Länge auf die Position der Präpositionalphrasen zeigen dürfte. Eine klare Entwicklung zeigt sich bei der Länge ebenfalls nicht, obwohl der letzte Zeitabschnitt 1850 bis 1900 dadurch hervorsteht, dass die eingebetteten Phrasen länger sind als in den anderen Zeitabschnitten.

Betrachtet man die Skipgram-Surprisalwerte (Tabelle 8.9), fällt keine klare Tendenz auf. Die ausgelagerten und eingebetteten Phrasen bleiben in den einzelnen Zeitabschnitten weitestgehend gleich. Es zeigt sich aber, dass die extrapponierten Präpositionalphrasen außer im Abschnitt 1700 bis 1750 im Schnitt

einen höheren durchschnittlichen Skipgram-Surprisalwert als ihre eingebetteten Gegenpaare aufweisen. Das trifft auch in geringerem Maß auf den Median zu.

Im Folgenden werden wieder die logistischen Regressionen (vgl. R Core Team 2022) durchgeführt. Als unabhängige Variablen werden die Länge, die durchschnittlichen Skipgram-Surprisalwerte der Phrasen, das Genre und der Orality Score genommen, wobei für jeden Abschnitt überprüft wird, ob eine Logarithmierung einzelner oder aller numerischen Variablen notwendig ist. Die abhängige Variable wird von der Position gebildet.

Zeitabschnitt	Position	Faktor	Minimum	Median	Durchschnitt	drittes Quartil	Maximum
1650-1700	in situ	Länge	2,000	4,000	7,732	8,000	81,000
1650-1700	extraponiert	Länge	2,00	9,00	16,13	18,00	225,00
1700-1750	in situ	Länge	2,000	5,000	6,457	8,000	26,000
1700-1750	extraponiert	Länge	2,00	10,00	13,94	18,00	66,00
1750-1800	in situ	Länge	2,000	6,000	7,697	10,000	24,000
1750-1800	extraponiert	Länge	2,00	10,00	13,94	18,00	66,00
1800-1850	in situ	Länge	2	4	6	8	21
1800-1850	extraponiert	Länge	2,00	11,00	14,34	16,00	130,00
1850-1900	in situ	Länge	2,00	9,00	14,87	18,00	84,00
1850-1900	extraponiert	Länge	2,0	21,0	25,4	33,0	135,0

Tabelle 8.8.: Deskriptive Statistik der Länge der annotierten Präpositionalphrasen, in Zeitabschnitte eingeteilt.

Zeitabschnitt	Position	Faktor	Minimum	Median	Durchschnitt	drittes Quartil	Maximum
1650-1700	in situ	Surprisal	3,140	3,719	3,727	3,919	4,382
1650-1700	extraponiert	Surprisal	3,264	3,760	3,761	3,890	4,416
1700-1750	in situ	Surprisal	3,187	3,695	3,706	3,832	4,318
1700-1750	extraponiert	Surprisal	3,216	3,626	3,658	3,763	4,356
1750-1800	in situ	Surprisal	3,150	3,704	3,706	3,852	4,313
1750-1800	extraponiert	Surprisal	2,987	3,717	3,717	3,841	4,350
1800-1850	in situ	Surprisal	2,910	3,602	3,589	3,741	3,912
1800-1850	extraponiert	Surprisal	3,338	3,603	3,634	3,723	4,261
1850-1900	in situ	Surprisal	2,798	3,577	3,597	3,734	4,259
1850-1900	extraponiert	Surprisal	3,044	3,596	3,605	3,720	4,351

Tabelle 8.9.: Deskriptive Statistik der Skipgram-Surprisalwerte der annotierten Präpositionalphrasen, in Zeitabschnitte eingeteilt.

Extrapolation von Präpositionalphrasen im Zeitraum von 1650 bis 1700 Im ersten Teilabschnitt (1650 bis 1700), wurden 239 ausgelagerte und 194 eingebettete Präpositionalphrasen gefunden. Die Länge wird logarithmiert. Das finale Modell folgt in Tabelle 8.10.

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	29,94	10,87	2,76	0,006	**
Skipgram _{mean}	-51,22	18,67	-2,74	0,006	**
Länge _{log}	-23,31	9,54	-2,44	0,01	*
Genre	16,92	9,06	1,87	0,06	.
Orality Score	-27,84	17,23	-1,62	0,11	
Skipgram _{mean} :Länge _{log}	32,41	16,15	2,01	0,04	*
Skipgram _{mean} :Genre	-38,74	15,86	-2,44	0,01	*
Skipgram _{mean} :Orality	51,48	29,34	1,76	0,08	.
Länge _{log} :Orality Score	4,17	1,95	2,14	0,03	*
Genre:Orality Score	12,17	4,24	2,87	0,004	**

Tabelle 8.10.: Logistische Regression der Präpositionalphrasen im Zeitabschnitt 1650-1700, finales Modell. Negative z-Werte sprechen für Extrapolation.

Die durchschnittlichen Skipgram-Surprisalwerte zählen zu den signifikanten Prädiktoren ($z=-2,74$, $p<0,01$). Das bedeutet, dass Phrasen mit hohen durchschnittlichen Skipgramwerten sehr wahrscheinlich ins Nachfeld gestellt werden. Die Länge ($z=-2,44$, $p<0,05$) hat ebenfalls einen Einfluss auf die Auslagerung der Phrasen, allerdings erklärt sie weniger Datenpunkte als die Surprisalwerte. Bei den Hauptvariablen findet sich auch ein signifikantes Ergebnis des Genres ($z=1,87$, $p<0,1$). Gehört ein Text dem theologischen Genre an, wird Nachfeldstellung der Phrasen wahrscheinlicher. Allerdings ist der Einfluss schwächer als der der von Skipgram-Surprisal und Länge.

Auch in den Interaktionen finden sich signifikante Werte, die zum Teil in unterschiedliche Richtungen weisen. Die Interaktion zwischen den Skipgramwerten und dem Genre ($z=-2,44$, $p<0,05$) zeigt, dass Auslagerungen im theologischen Genre bei Phrasen mit hohen Skipgramwerten wahrscheinlicher sind als im medizinischen Genre. Ebenfalls interessant ist die Interaktion aus dem Genre und dem Orality Score ($z=2,87$, $p<0,01$): Einbettung wird in theologischen Texten dann wahrscheinlicher, wenn der Text gleichzeitig mehr

Merkmale gesprochener Sprache enthält. Gehört eine der annotierten Präpositionalphrasen einem sehr mündlichkeitsnahen, medizinischen Text an, steht sie eher im Nachfeld.

Im Zusammenhang mit der Mündlichkeitsnähe steht auch die Interaktion aus ihr und der Länge ($z=2,14$, $p<0,05$). Sehr lange Präpositionalphrasen aus sehr mündlichkeitsnahen Texten werden verglichen mit sehr langen Phrasen aus schriftlichkeitsnahen Texten etwas eher eingebettet.

Schwach signifikant ist auch die Interaktion zwischen den Surprisalwerten und dem Orality Score ($z=1,7$, $p<0,1$). Einbettung wird in mündlichkeitsnahen Texten hingegen auch bei informativen Phrasen wahrscheinlicher.

Die Interaktion von Skipgramwerten mit Länge ($z=2,01$, $p<0,05$) weist darauf hin, dass eine lange und zugleich schwer zu verarbeitende Präpositionalphrase eingebettet wird. Dieses Ergebnis wird in Kapitel 8.4 ausführlicher besprochen. An dieser Stelle soll vorerst nur hervorgehoben werden, dass die Skipgram-Surprisalwerte die einflussreichste unabhängige Variable dieser Analyse darstellen.

Extrapolation von Präpositionalphrasen im Zeitraum von 1700 bis 1750 Im zweiten Zeitabschnitt (1700 bis 1750) wurden 103 extrapolierte und 81 eingebettete Präpositionalphrasen gefunden. Nur die Länge muss logarithmiert werden. Nach dem Ausschluss aller Interaktionen, des Genres und des Orality Scores bleiben in diesem Modell nur die unabhängigen Variablen der durchschnittlichen Skipgram-Surprisalwerte und die der Länge (Tabelle 8.11).

	Estimate	Std.-Fehler	z-Wert	p-Wert
Intercept	-2,85	2,60	-1,09	0,27
Skipgram _{mean}	1,19	0,71	1,68	0,09 .
Länge	-2,16	0,53	-4,08	<0,001 ***

Tabelle 8.11.: Logistische Regression der Präpositionalphrasen im Zeitabschnitt 1700-1750, finales Modell. Negative z-Werte sprechen für Extrapolation.

Im Kontrast zu den Ergebnissen des Zeitabschnitts von 1650 bis 1700 weisen die Skipgram-Surprisalwerte auf Einbettung hin ($z=1,68$, $p<0,1$). Hoch

informative Phrasen werden eingebettet. Der signifikanteste Prädiktor ist die Länge ($z=-4,08$, $p<0,001$). Die Länge der Phrase erhöht die Wahrscheinlichkeit der Auslagerung.

Extrapolation von Präpositionalphrasen im Zeitraum von 1750 bis 1800 In der zweiten Hälfte des 18. Jahrhunderts (1750 bis 1800), wurden 132 ausgelagerte und 99 eingebettete Präpositionalphrasen gefunden. Auch in diesem Abschnitt wurde die Länge logarithmiert. Die rückwärtsgerichtete Modellauswahl resultiert schließlich im Modell in Tabelle 8.12.

	Estimate	Std.-Fehler	z-Wert	p-Wert
Intercept	2,69	2,19	1,23	0,22
Skipgram _{mean}	-0,17	0,57	-0,29	0,77
Länge	-2,55	0,56	-4,59	<0,001 ***

Tabelle 8.12.: Logistische Regression der Präpositionalphrasen im Zeitabschnitt 1750-1800, finales Modell. Negative z-Werte sprechen für Extrapolation.

Die einzige Variable, die ein signifikantes Ergebnis erzielt, ist die Länge ($z=-4,59$, $p<0,001$). Wieder werden also lange Phrasen mit hoher Wahrscheinlichkeit ins Nachfeld gestellt. Hypothese H1 kann nicht bestätigt werden. Im Vergleich zum Ergebnis des vorherigen Zeitabschnitts (1700 bis 1750) deutet sich der zu Beginn des Kapitels erwähnte Einflusswechsel hin zu gedächtnisbasierten Verarbeitungsvoraussetzungen an.

Extrapolation von Präpositionalphrasen im Zeitraum von 1800 bis 1850 Im Vergleich zum 18. Jahrhundert nimmt die Zahl der ausgelagerten Präpositionalphrasen und ihrer eingebetteten Gegenstücke in der ersten Hälfte des 19. Jahrhunderts ab. Von 1800 bis 1850 wurden nur noch 70 ausgelagerte und 51 eingebettete Präpositionalphrasen, die den Annotationskriterien entsprechen, annotiert. Ihre Länge wurde auf der Basis von zehn logarithmiert.

Das finale Modell in Tabelle 8.13 zeigt zahlreiche signifikante Variablen und Interaktionen. Begonnen wird mit den Variablen , die für Extrapolation sprechen. Im Gegensatz zu den vorherigen Zeitabschnitten, können die

durchschnittlichen Skipgram-Surprisalwerte als Haupteffekt Extraposition erklären ($z=-2,15$, $p<0,05$). Das bedeutet, dass hoch informative Phrasen in diesem Zeitabschnitt ins Nachfeld bewegt werden. Ebenfalls für Auslagerung sprechen die Länge ($z=-2,11$, $p<0,05$) und die Mündlichkeitsnähe ($z=-2,17$, $p<0,05$).

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	34,33	14,78	2,32	0,02	*
Skipgram _{mean}	-8,68	4,04	-2,15	0,03	*
Länge _{log}	-41,06	19,49	-2,11	0,04	*
Genre	24,50	13,89	1,76	0,08	.
Orality Score	-23,13	10,68	-2,17	0,03	*
Skipgram:Länge	10,16	5,33	1,91	0,06	.
Skipgram:Genre	-6,299	3,74	-1,69	0,09	.
Skipgram:Orality Score	6,09	2,92	2,09	0,04	*
Länge:Genre	-3,46	1,92	-1,799	0,07	.

Tabelle 8.13.: Logistische Regression der Präpositionalphrasen im Zeitabschnitt 1800-1850, finales Modell. Negative z-Werte sprechen für Extraposition.

Die Interaktion zwischen den Skipgramwerten und dem Genre ($z=-1,69$, $p<0,1$) ist wie folgt zu deuten: Die Wahrscheinlichkeit, dass hochinformative Präpositionalphrasen extrapositioniert werden, steigt, wenn der Text theologisch ist. Die Interaktion zwischen der Länge und dem Genre ($z=-1,799$, $p<0,1$) erlaubt die Interpretation der Länge als Haupteffekt; je länger eine Phrase ist, desto eher wird sie in beiden Genres ausgelagert, wobei lange Phrasen in medizinischen Texten etwas wahrscheinlicher eingebettet werden.

Auch die Interaktion aus den Skipgramwerten und der Länge ($z=1,91$, $p<0,1$) zeigt: Phrasen, die entweder lang oder informativ sind, werden zwar ausgelagert, treffen die beiden Faktoren jedoch auf eine Variable zu, wird diese eher eingebettet. Zu einem ähnlichen Ergebnis kommt man bei der Betrachtung der Interaktion aus den Skipgram-Surprisalwerten und dem Orality Score ($z=2,09$, $p<0,05$). Präpositionalphrasen mit mittleren Surprisalwerten werden allerdings eher in mündlichkeitsnahen Texten eingebettet.

Extrapolation von Präpositionalphrasen im Zeitraum von 1850 bis 1900 Im letzten Zeitabschnitt steigt die Zahl der annotierten Phrasen aufgrund ihres häufigen Vorkommens in dem Werk von Ludwig (1856) stark an. Insgesamt wurden 499 ausgelagerte und 474 eingebettete Phrasen annotiert. Für die logistische Regression wurde erneut die Länge logarithmiert. Das Ergebnis der rückwärtsgerichteten Modellauswahl wird in Tabelle 8.14 präsentiert.

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	0,75	0,45	1,66	0,09	.
Länge _{log}	-2,53	0,35	-7,25	<0,001	***
Genre	1,91	0,77	2,49	0,01	*
Orality Score	-3,88	0,91	-4,26	<0,001	***
Länge _{log} :Orality Score	1,13	0,63	1,78	0,07	.
Genre:Orality Score	7,06	1,46	4,82	<0,001	***

Tabelle 8.14.: Logistische Regression der Präpositionalphrasen im Zeitabschnitt 1850-1900, finales Modell. Negative z-Werte sprechen für Extrapolation.

Erneut zeigt die Regression kein einheitliches Bild hinsichtlich der Tendenz zu Einbettung beziehungsweise Auslagerung. Die niedrigsten p-Werte weisen hier die Länge ($z=-7,25$, $p<0,001$) und die Mündlichkeitsnähe ($z=-4,26$, $p<0,001$) auf. Lange Präpositionalphrasen werden also eher ausgelagert, genauso wie Präpositionalphrasen aus mündlichkeitsnahen Texten. Damit reiht sich auch dieses Jahrhundert in die Betrachtung ein, dass es nicht ausschlaggebend ist, wie informativ die Phrase ist, sondern wie lang sie ist. Auch die Interaktion aus der Länge und dem Orality Score ($z=1,78$, $p<0,1$) zeigt den Einfluss der Länge über alle Orality Scores hinweg, wobei Einbettung in sehr mündlichkeitsnahen Texten insgesamt noch wahrscheinlicher ist als in schriftlichkeitsnahen Texten. Dieses Ergebnis spiegelt das Genre ($z=2,49$, $p<0,05$) hingegen nicht wider. Präpositionalphrasen werden in theologischen und in theologischen, mündlichkeitsnahen Texten eher eingebettet, wie auch die Interaktion zwischen Genre und Orality Score ($z=4,82$, $p<0,001$) zeigt.

8.3. Relativsätze in den verschiedenen Zeitabschnitten

Wie bereits in den Kapiteln zur Analyse der Nominal- und Präpositionalphrasen, werden im Folgenden die Teilkorpora der Relativsätze der verschiedenen Zeitabschnitte untersucht. Das Kapitel beginnt mit der Deskription der Werte. Die Tabellen 8.16 und 8.17 geben das Minimum, den Median, den Durchschnitt, das dritte Quartil und das Maximum der Länge beziehungsweise der Skipgramwerte der Relativsätze an. Die Zahlen der restriktiven und appositiven beziehungsweise unbestimmten Relativsätze werden in Tabelle 8.15 wiedergegeben. Dabei sieht man, dass außer im Zeitabschnitt 1750 bis 1800 immer mehr restriktive Relativsätze als appositive annotiert wurden.

Zeitabschnitt	1650-1700	1700-1750	1750-1800	1800-1850	1850-1900
restriktiv	422	722	316	701	209
appositiv	233	394	448	460	127
unbestimmt	133	351	286	35	593

Tabelle 8.15.: Anzahl der restriktiven, appositiven und unbestimmten Relativsätze in den Zeitabschnitten.

Der Median und der Durchschnitt der Länge der ausgelagerten Relativsätze ist in allen Zeitabschnitten größer als die Werte der adjazenten Relativsätze. Betrachtet man den Median, zeigt sich, dass dieser Unterschied in den Zeitabschnitten 1650 bis 1700, 1700 bis 1750 und 1850 bis 1900 drei Wörter ausmacht und im Abschnitt 1750 bis 1800 zwei. Bemerkenswert ist die geringe Differenz im Abschnitt 1800 bis 1850, die nur ein Wort beträgt und auch im Durchschnitt kleiner ist als in den anderen Zeitabschnitten. Von 1800 bis 1850 sind die *in situ* Relativsätze 8,5 Wörter lang und die ausgelagerten Relativsätze rund 11 Wörter lang. Das macht nur einen Unterschied von 2,5 Wörtern. In den anderen Zeitabschnitten ist die Differenz deutlich größer. Von 1650 bis 1700 beträgt sie sogar fünf Wörter, von 1700 bis 1750 und von 1750 bis 1800 jeweils vier und im letzten Abschnitt wieder rund fünf Wörter. Nur im ersten Zeitabschnitt (1650 bis 1700) sind die extrapolierten Relativsätze im Schnitt mehr als zwanzig Wörter lang, wobei in dieser Länge nicht von einer klaren

Entwicklung gesprochen werden kann, da alle anderen Werte immer wieder leicht schwanken. Diese erste Beobachtung, die auch von Grafik 6.17 verstärkt wird, ist, dass die Länge sich auch bei der Analyse nach Zeitabschnitten als robuster Prädiktor erweisen dürfte, weil die Differenzen dieser Werte, mit der Ausnahme des Zeitabschnitts 1800 bis 1850, deutlich sind.

Anhand der Werte der durchschnittlichen Skipgram-Surprisalwerte und auch anhand von Grafik 6.23 ist es schwerer, Unterschiede zwischen den eingebetteten und ausgelagerten Relativsätzen zu erkennen, weil diese nur minimal sind. Der Median der eingebetteten Phrasen liegt in den Zeitabschnitten 1650 bis 1700, 1750 bis 1800, 1800 bis 1850 und 1850 bis 1900 unterhalb des Medians der ausgelagerten Phrasen, wobei dieser Unterschied frühestens in der zweiten Nachkommastelle sichtbar wird. Im Abschnitt 1700 bis 1750 ist der Median der eingebetteten Relativsätze höher als der der ausgelagerten. Das ist auch beim Durchschnitt der Werte der Fall.

Interessant ist allerdings, dass sich in einem sehr feinen Rahmen tatsächlich eine Entwicklung in den Surprisalwerten auf niedrigem Niveau erkennen lässt. Im ersten Zeitabschnitt (1650 bis 1700) liegen die Skipgramsurprisalwerte deutlich höher als im letzten Zeitabschnitt. Die Abnahme erfolgt nicht linear: Von 1750 bis 1800 steigen die Werte vorübergehend wieder an. Doch sind sie, wenn man das 19. Jahrhundert als Ganzes betrachtet, niedriger als im 17. Jahrhundert. Dennoch erfolgt von 1850 bis 1900 wieder ein kleiner Anstieg. Anhand der vorliegenden Daten kann keine Aussage darüber getroffen werden, ob das ein Ausreißer ist oder eine neue Entwicklung einleitet.

In den folgenden Abschnitten werden wieder Regressionen mit R (vgl. R Core Team 2018) durchgeführt, die alle nach dem gleichen Muster ablaufen. Die abhängige Variable ist die Position, die unabhängigen Variablen bilden die durchschnittlichen Skipgram-Surprisalwerte, die Länge, der Orality Score, das Genre der Texte und der Relativsatztyp, inklusive aller Zweiwegeninteraktionen. Wie zuvor auch wurde überprüft, ob Transformationen der numerischen Variablen nötig sind. Erfolgte Transformationen werden in den Modelltabellen kenntlich gemacht. Dargestellt und beschrieben werden nur die finalen Modelle.

Zeitabschnitt	Position	Faktor	Minimum	Median	Durchschnitt	drittes Quartil	Maximum
1650-1700	in situ	Länge	3,0	12,0	17,4	21,0	147,0
1650-1700	extrapoliert	Länge	2,00	15,00	22,11	27,00	150,00
1700-1750	in situ	Länge	2,000	7,000	8,453	10,000	56,000
1700-1750	extrapoliert	Länge	2,00	10,00	12,53	15,00	88,00
1750-1800	in situ	Länge	2,00	10,00	11,89	14,00	62,00
1750-1800	extrapoliert	Länge	3,00	12,00	15,51	18,00	92,00
1800-1850	in situ	Länge	2,000	7,000	8,472	10,000	58,000
1800-1850	extrapoliert	Länge	2,00	8,00	10,97	14,00	70,00
1850-1900	in situ	Länge	3,00	9,00	11,25	12,00	100,00
1850-1900	extrapoliert	Länge	3,00	12,00	16,82	19,00	130,00

Tabelle 8.16.: Deskriptive Statistik der Länge der annotierten Relativsätze, in Zeitabschnitte eingeteilt.

Zeitabschnitt	Position	Faktor	Minimum	Median	Durchschnitt	drittes Quartil	Maximum
1650-1700	in situ	Surprisal	3,204	3,707	3,716	3,824	4,279
1650-1700	extrapponiert	Surprisal	3,232	3,716	3,721	3,819	4,127
1700-1750	in situ	Surprisal	3,160	3,650	3,663	3,770	4,276
1700-1750	extrapponiert	Surprisal	3,215	3,644	3,646	3,751	4,154
1750-1800	in situ	Surprisal	3,076	3,670	3,667	3,789	4,318
1750-1800	extrapponiert	Surprisal	3,240	3,694	3,698	3,796	4,256
1800-1850	in situ	Surprisal	3,127	3,567	3,574	3,671	4,180
1800-1850	extrapponiert	Surprisal	3,229	3,568	3,575	3,666	4,169
1850-1900	in situ	Surprisal	2,909	3,605	3,601	3,701	4,084
1850-1900	extrapponiert	Surprisal	3,223	3,612	3,617	3,695	4,359

Tabelle 8.17.: Deskriptive Statistik der Skipgram-Surprisalwerte der annotierten Relativsätze, in Zeitabschnitte eingeteilt.

Extrapolation von Relativsätzen im Zeitraum von 1650 bis 1700 Das Modell aus dem ersten Zeitabschnitt (1650 bis 1700) zeigt die multiplen Faktoren auf, die in diesem Zeitabschnitt die Stellung von Relativsätzen beeinflussen (Tabelle 8.18). Zu den Variablen, die nicht zur Vorhersage der Extrapolation geeignet sind, zählt das durchschnittliche Skipgramsurprisal ($z=0,11$, $p=0,91$).

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	2,48	2,59	0,96	0,34	
Skipgram _{mean}	0,07	0,64	0,11	0,91	
Länge	-0,052	0,02	-2,95	0,003	**
Orality Score _{log}	6,81	2,86	2,38	0,02	*
Genre	-9,07	4,29	-2,11	0,03	*
Typ	-13,75	4,91	-2,80	0,005	**
Skipgram _{mean} :Genre	2,09	1,14	1,83	0,07	.
Skipgram _{mean} :Typ	2,30	1,20	1,91	0,06	.
Länge:Genre	0,07	0,04	1,84	0,07	.
Länge:Typ	0,04	0,01	3,33	0,0008	***
Orality Score _{log} :Typ	-12,67	5,71	-2,22	0,03	*
Genre:Typ	2,70	1,32	2,05	0,04	*

Tabelle 8.18.: Logistische Regression aller Relativsätze im Zeitabschnitt 1650-1700, finales Modell. Negative z-Werte sprechen für Extrapolation.

Bei den Hauptvariablen deutet die Mündlichkeitsnähe auf Einbettung hin ($z=2,38$, $p<0,05$). Alle anderen Faktoren bilden Extrapolation ab, so auch das Genre Theologie ($z=-2,11$, $p<0,05$). Ebenfalls gute Voraussagen über die Position des Relativsatzes treffen seine Länge ($z=-2,95$, $p<0,01$) und sein Typ ($z=-2,8$, $p<0,01$), die beide für Extrapolation sprechen.

Nur tendenziell signifikante Interaktionen finden sich zwischen den Skipgramwerten und dem Genre ($z=1,83$, $p<0,1$), was bedeutet, dass informative Relativsätze aus theologischen Texten eher eingebettet werden als die aus medizinischen Texten, Skipgramwerten und der Restriktivität ($z=1,91$, $p<0,1$), was trotz einer großen Variabilität eine höhere Wahrscheinlichkeit für die Adjazenz restriktiver, informativer Relativsätze zeigt, und die Interaktion zwischen der Länge und dem Genre ($z=1,84$, $p<0,1$). Obwohl lange Relativsätze in beiden Genres eher ausgelagert werden, ist der Anteil adjazenter Rela-

tivsätze in theologischen Texten höher. Erklärungsvorschläge dazu werden in Kapitel 8.4 diskutiert.

Signifikant ist auch die Interaktion zwischen dem Relativsatztyp und dem Genre eines Texts ($z=2,05$, $p<0,5$). Wenn ein Relativsatz appositiv ist und in einem theologischen Text steht, wird er wahrscheinlicher eingebettet. Ausgelagert werden hingegen appositive Relativsätze in mündlichkeitsnäheren Texten, wie die Interaktion zwischen dem Orality Score und dem Relativsatztyp zeigt ($z=-2,22$, $p<0,05$). Ausgelagert werden hingegen appositive Relativsätze in mündlichkeitsnäheren Texten. Lange Phrasen, die gleichzeitig restriktiv sind, werden eher adjazent zu ihrem Bezugsnomen positioniert, wie die Interaktion zwischen der Länge und dem Relativsatztyp ($z=3,33$, $p<0,001$) erkennen lässt. Trotzdem stehen kürzere Relativsätze eher adjazent.

Extrapolation von Relativsätze im Zeitraum von 1700 bis 1750 Auch im finalen Modell (Tabelle 8.19) des Zeitraums 1700 bis 1750 sind die durchschnittlichen Skipgram-Surprisalwerte nicht signifikant ($z=1,031$, $p=0,30$). Das Gleiche gilt für den Relativsatztyp ($z=1,125$, $p=0,26$). Laut den Haupteffekten stehen Relativsätze in theologischen Texten eher im Nachfeld ($z=-2,06$, $p<0,5$). Besonders relevant für die Nachfeldstellung sind die Länge des Relativsatzes ($z=-9,68$, $p<0,001$) und die Mündlichkeitsnähe eines Textes ($z=-5,66$, $p<0,001$), was sich mit der Literatur und vorherigen Beobachtungen dieser Arbeit deckt.

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	1,59	1,57	1,01	0,31	
Skipgram _{mean}	0,43	0,42	1,03	0,30	
Länge _{log}	-2,59	0,27	-9,68	<0,001	***
Orality Score	-1,43	0,25	-5,66	<0,001	***
Genre	-6,43	3,13	-2,06	0,04	*
Typ	0,16	0,14	1,13	0,26	
Skipgram _{mean} :Genre	1,81	0,85	2,13	0,03	*
Länge _{log} :Genre	2,00	0,54	1,86	0,06	.
Genre:Typ	0,49	0,28	1,73	0,08	.

Tabelle 8.19.: Logistische Regression aller Relativsätze im Zeitabschnitt 1700-1750, finales Modell. Negative z-Werte sprechen für Extrapolation.

Schwach signifikant sind die Interaktionen zwischen Länge und Genre ($z=1,86$, $p<0,1$), die Vorhersagen darüber trifft, dass lange, theologische Relativsätze eher adjazent stehen, obwohl Extraposition grundsätzlich häufiger bei langen Relativsätzen vorkommt. Ebenfalls tendenziell signifikant ist die Interaktion zwischen dem Genre und dem Relativsatztyp ($z=1,73$, $p<0,1$). Theologische Texte weisen insgesamt eher Relativsatzadjazenz auf, was sich noch leicht verstärkt, wenn der Relativsatz appositiv ist. Letztere wird auch von der signifikanten Interaktion zwischen den durchschnittlichen Skipgram-Surprisalwerten mit dem Genre der Texte erhöht ($z=2,13$, $p<0,5$). Hochinformativ Relativsätze aus theologischen Texten stehen eher adjazent.

Extraposition von Relativsätze im Zeitraum von 1750 bis 1800 Verglichen mit den vorherigen Zeitabschnitten ist das Modell aus der zweiten Hälfte des 18. Jahrhunderts kürzer (Tabelle 8.20). Das durchschnittliche Skipgram-Surprisal wurde im Zuge der rückwärtsgerichteten Modellauswahl ausgeschlossen.

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	1,46	0,33	4,42	<0,001	***
Länge _{log}	-2,68	0,35	-7,63	<0,001	***
Genre	-10,97	1,14	-9,65	<0,001	***
Orality Score	12,72	1,33	9,57	<0,001	***
Typ	0,99	0,42	2,40	0,02	*
Genre:Typ	7,33	1,93	3,80	0,0001	***
Orality Score:Typ	-7,63	2,19	-3,48	0,0005	***

Tabelle 8.20.: Logistische Regression aller Relativsätze im Zeitabschnitt 1750-1800, finales Modell. Negative z-Werte sprechen für Extraposition.

Eine signifikante Vorhersage zur Position eines Relativsatzes wird vom Relativsatztyp gebildet ($z=2,4$, $p<0,05$). Restriktive Relativsätze stehen damit eher adjazent. Sowohl die Länge ($z=-7,63$, $p<0,001$), als auch das Genre ($z=-9,65$, $p<0,001$) sprechen für Extraposition. Allein ist der Orality Score ein Prädiktor für Einbettung bei großer Mündlichkeitsnähe ($z=9,57$, $p<0,001$). Die Interaktion aus der Mündlichkeitsnähe und dem Relativsatztyp ($z=-3,48$,

$p < 0,001$) bestätigt, dass theologische Texte Extraposition signifikant häufiger aufweisen. Auch die Interaktion zwischen Genre und dem Relativsatztyp ($z = 3,80$, $p < 0,001$) zeigt Folgendes: Theologische Relativsätze werden über beide Typen hinweg häufiger eingebettet, wobei sich der Effekt für restriktive Relativsätze verstärkt. Der starke Einfluss der Länge ist jedoch ein wiederkehrendes Motiv, das sich auch im nächsten Zeitabschnitt fortsetzt.

Extraposition von Relativsätze im Zeitraum von 1800 bis 1850 Auch in der ersten Hälfte des 19. Jahrhunderts wurde die Informationsdichte im Lauf der Modellauswahl ausgeschlossen (Tabelle 8.21).

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	0,43	0,14	3,07	0,002	**
Länge	-0,07	0,01	-5,95	<0,001	***
Orality Score	-1,51	0,35	-4,31	<0,001	***
Typ	-0,25	0,13	-1,88	0,06	.
Genre	0,44	0,17	2,52	0,01	*
Länge:Orality Score	0,06	0,03	1,88	0,06	.
Orality Score:Genre	1,75	0,38	4,59	<0,001	***

Tabelle 8.21.: Logistische Regression aller Relativsätze im Zeitabschnitt 1800-1850, finales Modell. Negative z-Werte sprechen für Extraposition.

Für Auslagerung sprechen erneut die Länge der Relativsätze ($z = -5,95$, $p < 0,001$), der Orality Score ($z = -4,31$, $p < 0,001$), der Auslagerung in mündlichkeitsnahen Texten wahrscheinlicher macht, und die Restriktivität ($z = -1,88$, $p < 0,1$). Appositive Relativsätze werden folglich eher ausgelagert. Nicht auf Auslagerung, sondern auf Adjazenz deutet das Genre ($z = 2,52$, $p < 0,05$) hin: Theologische Texte zeigen eher adjazente Relativsätze.

Die Interaktionen zwischen der Länge und dem Orality Score ($z = 1,88$, $p < 0,1$) bestätigt, dass es Einbettung in den mündlichkeitsnähesten Texten gibt. Die Interaktion zwischen dem Orality Score und dem Genre ($z = 4,59$, $p < 0,001$) zeigt: Extraposition wird in medizinischen Texten wahrscheinlicher vor allem, wenn diese mündlichkeitsnäher sind. Die relevantesten Faktoren sind also wieder die Länge und die Mündlichkeitsnähe, was mit den Funden der Lite-

ratur (Kapitel 3.2) übereinstimmt und vielfach in dieser Arbeit als bedeutsam benannt wurden.

Extrapolation von Relativsätzen im Zeitraum von 1850 bis 1900 Das finale Modell des letzten Zeitabschnitts (Tabelle 8.22) ähnelt dem des vorherigen Abschnitts. Auch hier wurde die Informationsdichte ausgeschlossen und auch in diesem Modell spricht das Genre für Einbettung ($z=2,774$, $p<0,001$). Die drei verbleibenden Hauptvariablen Länge, Orality Score und Relativsatztyp dienen als Prädiktoren für Auslagerung, wobei wieder die Länge und die Mündlichkeitsnähe als die bedeutendsten Faktoren gelten. Im Detail heißt das Folgendes:

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	1,46	0,27	5,48	<0,001	***
Länge _{log}	-1,88	0,28	-6,71	<0,001	***
Orality Score	-2,81	0,74	-3,79	0,0002	***
Typ	-0,23	0,13	-1,79	0,07	.
Genre	0,49	0,18	2,77	0,005	**
Länge _{log} :Orality Score	1,97	0,76	2,61	0,009	**
Orality Score:Genre	1,84	0,39	4,74	<0,001	***

Tabelle 8.22.: Logistische Regression aller Relativsätze im Zeitabschnitt 1850-1900, finales Modell. Negative z-Werte sprechen für Extrapolation.

Ein langer Relativsatz steht mit hoher Wahrscheinlichkeit im Nachfeld ($z=-6,71$, $p<0,001$). Ebenso stehen Relativsätze in mündlichkeitsnahen Texten mit höherer Wahrscheinlichkeit im Nachfeld ($z=-3,79$, $p<0,001$).

Der Relativsatztyp kann die Position zwar auch erklären, aber zu einem geringeren Grad ($z=-1,79$, $p<0,1$). Dennoch lässt sich auch dieser Wert so interpretieren, dass appositive Relativsätze eher ausgelagert werden. All diese Ergebnisse wurden auch so schon von anderen Forschenden gefunden (Kapitel 3.2). Die anderen Prädiktoren sagen Einbettung voraus. Theologische Texte ($z=2,77$, $p<0,01$) erhöhen die Wahrscheinlichkeit, adjazente Relativsätze zu finden. Die Interaktion zwischen der Länge des Relativsatzes und dem Orality Score ($z=2,61$, $p<0,01$) zeigt, dass lange Relativsätze eher extrapoliert

werden, es im mündlichkeitsnähesten Text aber auch Adjazenz gibt. Auch die Interaktion zwischen dem Orality Score und dem Genre deutet auf Einbettung hin ($z=4,74$, $p<0,001$). Wenn also ein Text zugleich die Kriterien von konzeptioneller Mündlichkeit und der Zugehörigkeit zum theologischen Genre erfüllt, macht er das Auftreten von Relativsätzen in adjazenter Stellung zu ihrem Antezedens wahrscheinlicher. Die Diskussion dieser und der vorherigen Ergebnisse erfolgt im nächsten Kapitel (8.4).

8.4. Diskussion Hypothese 3: Diachrone Perspektive von Extraposition

Die dritte Hypothese der vorliegenden Arbeit lautet: „Über die Zeit wird der Einfluss der Informationsdichte auf Auslagerung geringer.“

Die Kapitel 8.1, 8.2 und 8.3 untersuchen, welche verschiedenen Faktoren pro Zeitabschnitt die Position der Elemente signifikant voraussagen können, indem pro Zeitabschnitt (1650 bis 1700, 1700 bis 1750, 1750 bis 1800, 1800 bis 1850 und 1850 bis 1900) eine eigene logistische Regression durchgeführt wurde. Die Ergebnisse sind in Tabelle 8.23 zusammengefasst.

Insgesamt kann die in Skipgram-Surprisalwerten gemessene Informationsdichte Auslagerung höchstens in den älteren Texten des Zeitabschnitts 1650 bis 1700 zuverlässig prognostizieren. Das ist allerdings auch nur bei den Nominal- und Präpositionalphrasen der Fall. Bei der Betrachtung der Relativsätze in den fünf untersuchten Zeitabschnitten konnte der Einfluss der Skipgram-Surprisalwerte nie als Haupteffekt nachgewiesen werden.

Dennoch gelten die Ergebnisse als Evidenz für das Zutreffen von Hypothese H3. Gründe für die Abnahme des Einfluss von Informationsdichtemaßen und die scheinbar gleichzeitige Zunahme des Gewichts der Länge werden in den nachfolgenden Abschnitten diskutiert.

	Zeitstufe	Skipgram	Länge	max. Signifikanz
NP	1650-1700	$z=-1,94$ $p<0,1$	$z=-1,74$ $p<0,1$	Länge, Skipgram
NP	1700-1750	$z=-0,22$ $p=0,82$	$z=1,8$ $p<0,1$	Länge
NP	1750-1800	/	$z=-1,932$ $p<0,1$	Genre, OS
NP	1800-1850	/	$z=-1,750$ $p<0,1$	Länge, OS
NP	1850-1900	$z=0,66$ $p=0,51$	$z=-3,40$ $p<0,001$	Länge, OS:Genre
PP	1650-1700	$z=-2,74$ $p<0,01$	$z=-2,44$ $p<0,05$	Skipgram, Genre:OS
PP	1700-1750	$z=1,68$ $p<0,1$	$z=-4,08$ $p<0,001$	Länge
PP	1750-1800	$z=-0,29$ $p=0,77$	$z=-4,59$ $p<0,001$	Länge
PP	1800-1850	$z=-2,15$ $p<0,05$	$z=-2,11$ $p<0,05$	Skipgram, Länge, Skipgram:OS
PP	1850-1900	/	$z=-7,25$ $p<0,001$	Länge, OS Score, Genre:OS
ReIS	1650-1700	$z=0,11$ $p=0,91$	$z=-2,95$ $p<0,001$	Länge:Typ
ReIS	1700-1750	$z=1,03$ $p=0,3$	$z=-9,69$ $p<0,001$	OS, Länge
ReIS	1750-1800	/	$z=-7,63$ $p<0,001$	Länge, Genre, OS, Genre:Typ, OS:Typ
ReIS	1800-1850	/	$z=-5,95$ $p<0,001$	Länge, OS, OS:Genre
ReIS	1850-1900	/	$z=-6,71$ $p<0,001$	Länge, OS, OS:Genre

Tabelle 8.23.: Zusammenfassung der Regressionsergebnisse der verschiedenen Zeitstufen. Aus Platzgründen wurden der Orality Score als „OS“ abgekürzt. „/“ bedeutet, dass die Variable aus der Regression ausgeschlossen wurde.

Nominalphrasen Ob Nominalphrasen unabhängig von ihrem Surprisalwert eher in älteren oder jüngeren Texten ausgelagert werden, wurde mithilfe der Variable Zeitstufe untersucht (Kapitel 6.1). Diese ist weder im Gesamtkorpus noch im medizinischen oder im theologischen Teilkorpus signifikant für Extraposition. Somit wird eine lineare Einflussnahme dieser Zeitstufe für die Extrapositionshäufigkeit dieser Phrasenkategorie ausgeschlossen.

Tabelle 8.23 erlaubt eine schnelle Betrachtung der Signifikanzniveaus der Skipgram-Surprisalwerte der Nominalphrasen pro Zeitabschnitt. Als einzelner Faktor kann Surprisal Extraposition nur im ersten Zeitabschnitt (1650 bis 1700) erklären. Der Einfluss der Informationsdichte liegt hier auf dem gleichen Niveau wie der Einfluss des Haupteffekts der Länge (je $p < 0.1$). Im ersten Zeitabschnitt sind die Signifikanzniveaus mit $p < 0,1$ generell eher gering. Die einzelnen Variablen sprechen höchstens für eine Tendenz der Extraposition. Länge und Informativität der Phrasen können ihre Position also gleichermaßen vorhersagen. Dabei ist aber zu bemerken, dass die Nominalphrasen mit hohen Surprisalwerten in dem Text mit dem höchsten Orality Score etwas eher eingebettet werden. In den anderen Texten werden hochinformativ Phrasen aber in Übereinstimmung mit Hypothese H1 eher ausgelagert.

Die frühe Verarbeitung der rechten Satzklammer spielt folglich eine Rolle (vgl. Kapitel 6.5.1), weil dann die Satzgliedfunktion der Nominalphrase im Nachfeld bekannt ist und keine weiteren Verarbeitungsprobleme hervorruft. Gleichzeitig stehen die kognitiven Kapazitäten früher für die Verarbeitung lexikalisch anspruchsvoller Phrasen zur Verfügung. Ist der Text allerdings mündlichkeitsnah, werden Nominalphrasen eher eingebettet. Es ist zu vermuten, dass die Einbettung der Nominalphrasen die weiteren Bestandteile des Satzes leichter verarbeitbar macht und – bedenkt man die Tatsache, dass die meisten mündlichkeitsnahen Texte auch theologische sind – die frühere Verarbeitung der rechten Satzklammer vielleicht aufgrund des Genres weniger verarbeitungsrelevant werden lässt (Kapitel 6.5).

Ab dem Abschnitt 1700 bis 750 kann der Skipgram-Surprisalwert Extraposition nicht mehr vorhersagen. Zwischen 1750 und 1850 wird die Variable komplett aus der Regression ausgeschlossen. Im Abschnitt von 1850 bis 1900

erklärt sie aber in Interaktion mit dem Genre Einbettung in theologischen Texten ($p < 0,05$). Umgekehrt heißt das, dass die Auslagerung von Nominalphrasen in medizinischen Texten auch von ihrer Informativität abhängt.

Insgesamt ist also zu bemerken, dass die Bedeutung von Surprisalwerten durch den Einfluss von Länge abgelöst wird. Diese Ablösung erfolgt aber keinesfalls linear, denn im Zeitabschnitt 1750 bis 1800 ist nicht die Länge mit $p < 0,1$ unter den einflussreichsten Faktoren, sondern das Genre und der Orality Score, was mit einer Veränderung des wissenschaftlichen Schreibens in diesen Zeitabschnitten zusammenhängen dürfte. Vierhaus (1985: 11) beschreibt für das 18. Jahrhundert einen Wandel in der wissenschaftlichen Praxis und des Schreibstils:

Doch wandelte sich der gelehrte Stil; er versachlichte sich in dem Maße, wie die kritische Prüfung tradierter und die Erschließung neuer Kenntnisse, die Mitteilung gesicherten Wissens und die Beschreibung erprobter Methoden, um sie zur Diskussion zu stellen in den Vordergrund traten.

Gleichzeitig betont Vierhaus (1985: 11) aber auch, dass sich dieser Wandel nicht linear vollzogen hat, was auch die hier gefundene nicht-lineare Entwicklung der Einflussfaktoren erklären kann, was Klein (2011) ebenfalls bestätigt.

Ab 1800 ist also die Länge der einflussreichste Prädiktor für Auslagerung. Das weist auf eine Bedeutungszunahme des gedächtnisbasierten Ansatzes hin, die wiederum bedingt, dass die erwartungsbasierten Verarbeitungsbedingungen an Relevanz verloren haben. Folglich kann davon gesprochen werden, dass Evidenz für das Zutreffen von Hypothese H3 bei der Betrachtung der Nominalphrasen gefunden wurde.

Präpositionalphrasen Wenn nur die Häufigkeit von Auslagerungen unabhängig von ihrer Informationsdichte in die Analyse einbezogen wird, findet sich im Gesamtkorpus der Präpositionalphrasen das Ergebnis, dass Einbettung in jüngeren Texten wahrscheinlicher ist. In älteren Texten überwiegt folglich Extraposition (Tabelle 6.15). Auch bei den attributiven Präpositionalphrasen ist Auslagerung in jüngeren Texten wahrscheinlicher. Bei der Betrachtung dieser Teilkorpora der Präpositionalphrasen zeigt sich also eine lineare Ent-

wicklungstendenz. Darin kann ein Hinweis darauf liegen, dass der Attributstatus für die Veränderung der Position in jüngeren Texten relevant ist und einen Ansatzpunkt für die Unterschiede zwischen Präpositionalphrasen und Nominalphrasen liefert.

Betrachtet man jedoch den Einfluss der Informationsdichte auf Präpositionalphrasenextraposition in den einzelnen Zeitabschnitten mithilfe von Tabelle 8.23, zeigt sich kein Unterschied zwischen den Präpositionalphrasen und den Nominalphrasen im ersten Abschnitt. Von 1650 bis 1700 bilden die Skipgramwerte den signifikantesten Prädiktor für Extraposition ($p < 0,01$),¹ vor der Länge ($p < 0,05$). Die höheren kognitiven Kapazitäten, die im Nachfeld zur Verfügung stehen, scheinen für die Verarbeitung der Präpositionalphrasen also besonders relevant zu sein. Dadurch wird zwar auch die Distanz zwischen den Satzklammern kürzer gehalten, doch die Bedeutung davon ist nicht so groß wie in den späteren Abschnitten.

Im Abschnitt von 1700 bis 1750 ist Länge am bedeutendsten für Auslagerung ($p < 0,001$). Die Skipgramwerte in dieser Periode weisen sogar eher auf Einbettung hin ($p < 0,1$). Es ist denkbar, dass die Einbettung informativer Phrasen die Verarbeitung nachfolgender Konstituenten vereinfacht. Allerdings wirkt dieser Abschnitt (1700 bis 1750) eher wie eine Anomalie.

Ab dem 19. Jahrhundert ist Länge der erklärungsstärkste Prädiktor für die Extraposition von Präpositionalphrasen und weist wie auch bei den Nominalphrasen darauf hin, dass die Bedeutung gedächtnisbasierter Verarbeitungsmethoden überwiegt. Allerdings erreichen die Skipgramwerte von 1800 bis 1850 das gleiche Signifikanzniveau wie die Länge ($p < 0,05$). Offenbar wurden in diesem Zeitraum sowohl die rückwärts- als auch die vorwärtsgerichteten Prinzipien zur Verarbeitungserleichterung genutzt.² Zum einen ist der Satzrahmen kurz zu halten, um eine frühere Verarbeitung der rechten Satzklammer zu gewährleisten, zum anderen müssen aber auch die freien Kapa-

¹ Die Interaktion mit dem Genre zeigt allerdings auch, dass vor allem hochinformativ präpositionalphrasen in theologischen Texten wahrscheinlicher extrapositioniert werden als in medizinischen. Möglicherweise trifft in der medizinischen Literatur eine Verarbeitungserleichterung für andere Elemente im Satz zu.

² Möglicherweise zeigt sich hier aber auch der wellenförmige Sprachwandel, den Degaetano-Ortlieb und Teich (2018: 30) fürs Englische beschreiben.

zitäten hinter der rechten Satzklammer genutzt werden, um schwierigere Präpositionalphrasen zu verarbeiten. Möglicherweise haben die Autoren und die Rezipierenden die in diesem Abschnitt behandelten Themen komplexer gefunden, sodass die Stellungsfreiheiten unbewusst noch stärker genutzt wurden, um optimales Verständnis zu garantieren.

Im Abschnitt 1850 bis 1900 ändern sich die einflussreichen Prädiktoren aber. Die Skipgramwerte wurden aus dem Modell ausgeschlossen. Die Länge bildet den signifikantesten Prädiktor für Auslagerung ($p < 0,001$). Das Ergebnis spiegelt die Beobachtungen wider, die für synchrone Daten getroffen wurden (Kapitel 3). Mündlichkeitsnähe macht Extraposition ebenso wahrscheinlicher wie die große Länge einer Phrase.

Zwar ist das Ergebnis weniger linear als bei den Nominalphrasen, trotzdem zeigt sich ein deutlicher Unterschied zwischen den signifikanten Prädiktoren hin zu den eher gedächtnisbasierten. Das spricht insgesamt für eine Bestätigung von Hypothese H3.

Relativsätze Bei den Relativsätzen ist Auslagerung in jüngeren Texten wahrscheinlicher ($p < 0,05$), wie im Gesamtkorpus an der Variable Zeitstufe zu erkennen ist. Das Gleiche gilt in medizinischen Texten und bei der Inklusion der Distanzen in die Regression. Aber damit sind noch keine Aussagen über den Einfluss von Surprisal in den einzelnen Zeiten getroffen.

Welchen Einfluss die Surprisalwerte auf Extraposition von Relativsätzen in welchem Zeitabschnitt haben, wird in Tabelle 8.23 und in Kapitel 8.3 dargestellt. Surprisalwerte sind in keinem Zeitabschnitt als Einzelfaktor signifikant für die Stellung des Relativsatzes. Nur im ersten Zeitabschnitt interagieren sie mit dem Genre und dem Typ der Relativsätze. Hochinformativ Relativsätze werden folglich nur im medizinischen Korpus extraponiert. Die Adjazenzstellung, auch von informativen Relativsätzen, kann in theologischen Texten und bei appositiven Sätzen deshalb von Vorteil sein, weil sie im Nachfeld in ihrer Form, aber nicht wegen ihres lexikalischen Gehalts, zu überraschend sind. Das gleiche Argument ist für die Interaktion zwischen Länge und dem Relativsatztyp anzuführen.

Im Abschnitt von 1700 bis 1750 ist die Interaktion aus Skipgram und

Genre signifikant für Einbettung hochinformativer Relativsätze in theologischen Texten ($p < 0,05$). Bedeutend sind auch die Einzelfaktoren Länge und Orality für Auslagerung (je $p < 0,001$). Auf den Genreunterschied wurde in Kapitel 6.5 eingegangen. Durch die bekannteren Inhalte und Verben in der rechten Satzklammer in theologischen Texten entfallen die Verarbeitungsherausforderungen, die in den medizinischen Texten vorliegen, und erlauben auch die Adjazenzstellung von Relativsätzen, ohne Verarbeitungsprobleme durch die erhöhte Distanz der Satzklammern zu riskieren. Ab 1750 werden die Skipgram-Surprisalwerte aus den Modellen ausgeschlossen und Länge wird konstant zu einem der bedeutendsten Faktoren für Extraposition und ist jeweils hochsignifikant ($p < 0,001$).

Es ist also zu erkennen, dass auch hier eine Abnahme des Einflusses von Informationsdichte stattfindet, wenn auch nicht in dem Umfang wie bei den Nominal- und Präpositionalphrasen. Es zeigt sich der bedeutendere Einfluss von gedächtnisbasierten Verarbeitungsansätzen der satzförmigen Relativsätze, der bei diesen generell deutlicher zu Tage tritt als bei den Phrasen.

Da die in den ersten beiden Zeitabschnitten aber in Interaktionen signifikant sind, lässt sich auch für dieses Phänomen die Hypothese H3 bestätigen.

Allgemeine Überlegungen zum Einflusswandel In allen Phänomenen kann die Hypothese, die eine Abnahme der Relevanz von Informationsdichte auf Extraposition sieht, bestätigt werden. Es lässt sich beobachten, dass es wichtiger zu werden scheint, die Memory Cost gering zu halten. Folglich ist ein Einflusswandel von vorhersagebasierten Faktoren zu gedächtnisbasierten Verarbeitungsmodellen erkennbar. Daraus resultiert aber die Frage, wie dieser Einflusswandel zustande kommt.

Bei den Nominalphrasen werden die durchschnittlichen Surprisalwerte über den betrachteten Zeitraum geringer. Das gilt allerdings sowohl für die eingebetteten als auch für die ausgelagerten Phrasen und der Unterschied ist eher klein. Die Erklärung, dass es am verwendeten Vokabular und dessen Kombination liegt, kann also nicht zutreffen. Stattdessen ist davon auszugehen, dass sich die wissenschaftliche Sprache vor allem im 19. Jahrhundert insgesamt optimiert hat und sich sowohl im Diskurstyp als auch im Stil ver-

ändert, wie Degaetano-Ortlieb et al. (2019: 277) auch fürs Englische darlegen. Fürs Deutsche argumentiert Kretzenbacher (1995: 24f.) für eine ähnliche Entwicklung, die durch die sprachliche Normierung seitens Klopstocks und Adelung im 18. Jahrhundert eingeleitet und gefestigt wird.

Die vorliegenden Daten stützen die These des Wandels im Deutschen auch empirisch. Die Autoren, deren Texte in das verwendete Korpus eingeflossen sind, nutzten die Stellungsfreiheiten des Deutschen offenbar so, dass auch verarbeitungsaufwendigere Phrasen unabhängig von ihrer speziellen Position besser zu verarbeiten sind.³ Es ist also denkbar, dass sich die Informativität der eingebetteten und ausgelagerten Phrasen einander annähern und daher kein durch Regressionsanalysen erkennbarer Unterschied hinsichtlich der Position zu sehen ist. In diesem Fall wird es für die Stellung der Phrasen weniger relevant, dass im Nachfeld mehr kognitive Kapazitäten zur Verfügung stehen, wie in Kapitel 1.1 beschrieben. Diese zusätzliche Kapazität muss nicht mehr für die Verarbeitung verwendet werden und kann aus den bedeutenden Faktoren für die phrasale Auslagerung ausgeschlossen werden.

Das bedeutet wiederum, dass es nicht daran liegt, dass sich die kognitiven Fähigkeiten der Autoren und ihrer Leserschaft über den untersuchten Zeitraum hinweg verändern, sie umgangssprachlich ‚schlauer‘ werden, wodurch sie hochinformativ Phrasen auch im Mittelfeld besser verstehen kön-

³ Zum jetzigen Zeitpunkt kann darüber eher spekuliert werden. Zwar gibt es durch die Arbeit von Cuskley et al. (2021) eine Möglichkeit, die gleichförmige Informationsverteilung über einen Satz zu berechnen. Das bedeutet grundsätzlich, dass man überprüfen kann, ob ein Satz ein gleichmäßigeres Informationsprofil im Sinne der UID (vgl. Levy und Jaeger 2007) erhält, wenn eine Phrase entweder ausgelagert oder eingebettet wird.

Dazu könnte man in Anlehnung an in Ortman (2023) dargestellte Verfahren ein Variantenkorpus erzeugen, in dem die ausgelagerten Phrasen eingebettet würden, und im Anschluss mithilfe der DORM-Werte (‘Deviation of the Rolling Mean’, vgl. Cuskley et al. 2021) gezeigt würde, ob das den Satz schwerer zu verarbeiten macht. Allerdings ist man bei der Erzeugung der Variante damit konfrontiert, dass es zu viele Unsicherheiten im Deutschen darüber gibt, wo die entsprechende Phrase im Mittelfeld stehen würde. Diese genaue Position ist aber essenziell, um die Veränderung im Informationsprofil hervorzurufen und kann daher aufgrund des aktuellen Forschungsstands nicht durchgeführt werden.

Das umgekehrte und durchaus machbare Verfahren, eingebettete Phrasen auszulagern, kann eher ein verzerrtes Ergebnis liefern, da Phrasen grundsätzlich häufiger eingebettet auftreten und daher auch an dieser Position besser zu verarbeiten sein sollten, wenn sie ursprünglich dort stehen. Wenn weitere Forschung zur Wortstellung im Deutschen es aber ermöglicht, die Ursprungsposition von Auslagerungen besser zu bestimmen, ist das hier skizzierte Verfahren dazu geeignet, um die Erklärungsansätze der Arbeit, die auf der besseren Verarbeitung des Restsatzes durch Einbettung beruhen und wie eben beschrieben ein insgesamt gleichförmigeres Informationsprofil neuerer Texte vermuten, auf eine empirische Belegbasis zu stellen.

nen. Stattdessen haben die Autoren gelernt, auf lexikalischer Ebene effizienter zu kommunizieren.⁴ Besonders deutlich wird das für die Nominalphrasen, die insgesamt auch den höchsten Einfluss von Informationsdichte von allen untersuchten Phänomenen, gemessen in Skipgram, aufweisen.

Der Bedeutungsgewinn der Länge ist außerdem im zeitlichen Verlauf zu sehen. Bei der Analyse der verschiedenen Regressionen wurde bereits ausführlich darauf hingewiesen, dass durch die grundsätzlich mögliche Einbettung von im Korpus ausgelagertem Material der Abstand zwischen den Satzklammern vergrößert würde. Dadurch wird das Gedächtnis beansprucht, da das lexikalische Verb des deutschen Satzes erst zu einem sehr späten Zeitpunkt verarbeitet werden kann, bis dahin aber sehr viele mögliche Satzalternativen im Gedächtnis behalten werden müssen und Abhilfe erst durch die Verarbeitung der rechten Satzklammer erfolgen kann.⁵

Das hier verwendete Korpus erlaubt Schlüsse darüber, dass es zwischen 1650 und 1900 wichtiger wurde, auf dieses Kriterium zu achten. Natürlich kann das daran liegen, dass die Autoren, wie oben beschrieben, bei der lexikalischen Gestaltung ihrer Sätze geschickter wurden, aber auf dieses grammatische Kriterium weniger Rücksicht nehmen konnten. Es ist aber auch denkbar, dass hier ein verarbeitungsbezogener Sprachwandel stattgefunden hat, der aufgrund der stiefmütterlichen Behandlung des frühen Neuhochdeutschen in der aktuellen Forschung bisher nicht erkannt wurde.

Dass die Bedeutung des Satzrahmens sich erst im frühen Neuhochdeutschen final festigte, ist aufgrund der Tatsache auszuschließen, dass vollständige Satzrahmen schon im Mittelhochdeutschen in mehr als der Hälfte der

⁴ Prinzipiell ist denkbar, dass sich trotzdem noch Unterschiede auf der grammatischen oder pragmatischen Ebene ergeben, die aber mit dem vorliegenden Datenset und den Mitteln, die zum Zeitpunkt der Korpuserstellung zur Verfügung standen, nicht überprüft werden konnten.

⁵ Natürlich kann hier berechtigt eingeworfen werden, dass man die Abstände zwischen den Satzklammern hätte berücksichtigen und als Faktor in die Regressionsanalyse aufnehmen können. Allerdings gibt es keinen Wert, wie viele Wörter zwischen den Satzklammern stehen können, bevor ihre Distanz zu groß ist, und auch hängt die kleinstmögliche Distanz zwischen den Satzklammern von zahlreichen Faktoren ab. Darunter zählt auch die Verbvalenz, die bedingt, welche Mitspieler an der Verbalhandlung obligatorisch sind, aber auch die Verwendung von beispielsweise Satzadverbialen und ähnlichem.

Außerdem konnten Speyer und Lemke (2017) den die Bedeutung des Matrixsatzes nicht für alle von ihnen erhobenen Daten nachweisen.

Sätze aufgetreten sind (vgl. z. B. Paul 2007) und die feste Etablierung des Rahmens übereinstimmend während des Frühneuhochdeutschen angenommen wird (vgl. z. B. Admoni 1990; Dal und Eroms 2014). Damit stehen den Autoren ausreichend viele Textvorbilder zur Verfügung, die die typische, deutsche Aufteilung des Verbs auf beide Klammern aufweisen, die auch in ihre Schreibpraxis vorzufinden sind.

Auch die Argumentation, dass sich die Autoren, vor allem die Theologen und frühen Mediziner, vom Lateinischen beeinflussen lassen (vgl. z. B. Voigtmann und Speyer 2021), ist schwer haltbar. Frank et al. (2016) konnten zeigen, dass Sprecher durchaus in der Lage sind, zwischen den syntaktischen Strukturen ihrer Mutter- und der erlernten Fremdsprachen zu differenzieren. Die Ergebnisse dieser Studie sind allerdings insofern in Zweifel zu ziehen (vgl. Kapitel 2.1), dass die Probandengruppe keine guten Rückschlüsse auf die Allgemeinbevölkerung oder auf die Autoren dieses Korpus zulässt. Zudem galt das Lateinische bis ins 16. Jahrhundert hinein als prestigeträchtig (vgl. Klein 2011: 37). Klein (2011: 39) schreibt dem Deutschen seit dem 16. Jahrhundert einen Prestigegewinn zu, der aber erst „gegen Ende des 17. Jahrhunderts nachdrückliche Formen annehmen konnte“, also dem Zeitraum, zu dem die vorliegende Untersuchung ansetzt. Der wohl abnehmende Einfluss des Lateinischen kann also höchstens für die Texte des 17. Jahrhunderts angenommen werden. Denkbar – und ebenfalls durch zukünftige Studien zu überprüfen – ist stattdessen, dass die Gewöhnung an einen bestimmten Stil, beispielsweise den von Predigten im Fall der Theologen, einen größeren Einfluss darauf ausübt, wie man auch in anderen Situationen schreibt. Diese angenommene Gewöhnung wurde in den obigen Diskussionen zur Erklärung der Differenzen zwischen Mediziner und Theologen herangezogen (Kapitel 6.5 und 7.2).

Die Veränderung über die Zeit hinweg kann also auch mit der Professionalisierung des Schreibens in beiden Genres einhergehen. Arbeiten von Konopka (1996) und Takada (1998), die auch diese Zeit aber unter hier nicht näher beschriebenen Gesichtspunkten betrachtet haben, zeigen, dass lange Distanzen zwischen Satzklammern im 17. und 18. Jahrhundert nichts Ungewöhnliches waren. Wird also der Ansatz des Language Dependent Structural

Forgetting (vgl. Vasishth, Suckow, Lewis und Kern 2010; Frank et al. 2016; Futrell et al. 2021) zur Erklärung herangezogen, kann davon ausgegangen werden, dass sowohl die Schreibenden als auch Rezipierenden dieser Jahrhunderte an diese Strukturen gewöhnt waren und es ihnen deshalb leichter gefallen ist, sie zu verarbeiten. Allerdings wird in der Untersuchung der Grammatiken aus dem 18. Jahrhundert (vgl. Konopka 1996) darauf hingewiesen, dass die lange Distanz zwischen der linken und rechten Satzklammer von zeitgenössischen Schreibenden bereits kritisiert und als archaisch wahrgenommen wurde. Eine allgemeine, frühere Verarbeitung der rechten Satzklammer scheint also bereits im 18. Jahrhundert bevorzugt zu werden. An dieser Stelle ist die Verarbeitung, wie in Kapitel 6.5.1, angedeutet sowohl vorwärts- als auch rückwärtsgerichtet vorteilhaft. Zum einen wird das Gedächtnis entlastet, zum anderen können fehlende Mitspieler an der Verbalhandlung ohne Unsicherheiten erkannt und identifiziert werden. Selbst bei den Präpositionalphrasen wird nach Verarbeitung des lexikalischen Materials in der rechten Satzklammer der Valenzstatus der Auslagerung erkannt, das heißt, die Rezipierenden sehen, ob es sich um eine freie Angabe handelt.⁶ Lediglich für die Relativsätze gilt der vorwärtsgerichtete Ansatz nicht, weil sie grundsätzlich nur die attributive Funktion aufweisen. Allerdings verbrauchen sie durch ihren Satzstatus grundsätzlich mehr Ressourcen als die Nominal- und Präpositionalphrasen.

Wenn also schon im 18. Jahrhundert eine Verkürzung der Mittelfelder gefordert wurde, kann es prestigeträchtig geworden sein, die Distanz zwischen den Satzklammern kürzer zu halten. Prestige gilt als einer der Hauptgründe für Sprachwandel (vgl. z. B. Mattheier 1981) und kann dafür gesorgt

⁶ Bei den Präpositionalphrasen ist im Korpus ebenfalls eine Veränderung in den Präpositionen von Ergänzungen und Objekten zu erkennen, die in der Forschung noch keinen mir bekannten Niederschlag gefunden hat.

Es würde den Rahmen dieser Arbeit übersteigen, sich gesondert damit zu befassen, jedoch fiel bei der Annotation und den nachfolgenden manuellen Analysen immer wieder auf, dass Präpositionalobjekte im 17. und 18. Jahrhundert mit anderen Präpositionen gebildet wurden, als es heute der Fall ist. Eine Erklärung dafür kann hier nicht gegeben werden, doch es ist zu vermuten, dass diese Veränderung auch eine Rolle für den signifikanten Einfluss von Surprisalwerten von 1800 bis 1850 verantwortlich ist, weil eine bis dahin nicht verwendete Präposition plötzlich ein Objekt einleitet, das dann im Nachfeld steht. In diesen Fällen würde sich auch ein höherer Surprisalwert auf dieser Präposition bemerkbar machen, der dann den relevanten Einfluss dieser Variable erklären könnte.

haben, dass sich dieses Phänomen auch an anderen Stellen durchsetzt und langsam Gewöhnungen an lange Abhängigkeiten früherer Zeitstufen und des Lateinischen weniger stark zum Tragen kommen. Gewöhnung kann nach Futrell et al. (2021) dafür sorgen, dass Verarbeitungserleichterungen stattfinden. Wenn sich Schreibende und Rezipierende folglich an kurze Mittelfelder gewöhnt haben, geht mit diesen erfolgreiche Kommunikation einher. Damit kann dieser Vorteil den der Extraposition von hochinformativem Material überschreiben, sodass auch Material ausgelagert wird, das im Mittelfeld gut zu verarbeiten gewesen wäre. Selbst für die Theologen, die zum einen mit archaischen Texten vertraut und zum anderen häufig in Kontakt mit dem Lateinischen und dessen langen Abhängigkeiten gekommen sind, ist dieser Einfluss zu vermuten, auch wenn er langsamer stattgefunden zu haben scheint als im medizinischen Teilkorpus.⁷

Ein berechtigter Einwand an der hier vorgebrachten Analyse und der Argumentation für eine Professionalisierung des Schreibens, die mit einem veränderten Wissenschaftsverständnis der Theologie und einer Gewöhnung an kurze Satzrahmen einhergeht, ist folgender: Tabellen 7.1 und 7.2 zeigen, dass Schriftlichkeitsnähe der Texte mit abnehmendem Alter korreliert, weshalb zu vermuten ist, dass sich nicht so sehr ein Sprachwandel als einfach ein Wandel vom mündlichkeitsnahen Stil, in dem die Informationsdichte nachweislich einen größeren Effekt auf Extraposition im Allgemeinen hat (Tabelle 7.4), zu schriftlichem Stil ist (Tabelle 7.6). Bereits in der Einleitung (Kapitel 1.1) wurde der Zusammenhang aus zunehmender Schriftlichkeitsnähe und dem Publikationszeitpunkt der Texte angemerkt. Um ausschließen zu können, dass es sich bei dem beschriebenen Zusammenhang nicht wirklich nur um einen Wechsel des Registers handelt, müssten weitere Textsorten, die entweder schriftlichkeits- oder mündlichkeitsnah sind, von 1650 bis 1900 untersucht werden, um zu überprüfen, ob der Effekt dort ebenfalls in ähnlicher Ausprägung gefunden würde.

Allerdings kann die Zunahme der Schriftlichkeitsnähe in den jüngeren

⁷ Zudem argumentiert Polenz und Moulin (2021) dafür, dass schon im Frühneuhochdeutschen andere Sprachen wie das Französische und Englische einen größeren Einfluss gewinnen und das Lateinische in Teilen ablösen.

Texten ebenso Zeichen für zunehmende Professionalisierung darstellen wie der oben beschriebene Ausgleich der Surprisalwerte zwischen ausgelagerten und eingebetteten Phrasen, die als Professionalisierung des Stils betrachtet wurden. Degaetano-Ortlieb et al. (2019) argumentieren in ihrer Untersuchung von englischer Wissenschaftssprache ebenfalls für einen derartigen Effekt. Auch werden die Unterschiede zwischen den theologischen und medizinischen Texten nicht dadurch unterminiert, dass Professionalisierung im Schreiben mit der Verwendung schriftlichkeitsnaher Sprachmittel einhergeht.

Auch der beschriebene Wandel hin zur Schonung der Gedächtnisressourcen kann als unabhängig von der konzeptionellen Schriftlichkeit und Mündlichkeit (vgl. Koch und Oesterreicher 2007) gesehen werden, weil ausreichend viele synchrone Arbeiten (vgl. z. B. Beneš 1979; Zahn 1991; Brooks 2006) gezeigt haben, dass Auslagerung in mündlichkeitsnahen Texten frequent auftritt und dort mit Gedächtnisproblemen assoziiert werden kann.

Zusammenfassend ist also Folgendes zu sagen: Meine Untersuchung weist einen Bedeutungswandel von vorhersagte- zu gedächtnisbasierten Verarbeitungsansätzen in wissenschaftlichen Texten nach, der mit einer gesteigerten Professionalisierung des Schreibstils und einer generellen Optimierung des wissenschaftlichen Schreibstils erklärt wurde, der auch dadurch Ausdruck findet, dass die Texte jüngeren Datums mehr schriftlichkeitsnahe Merkmale aufweisen. Um allgemeine Aussagen über das frühe Neuhochdeutsche zu treffen, müssten natürlich weitere Texte und Genres untersucht werden.

9. Zusammenfassung

Die vorliegende Arbeit geht der übergeordneten Fragestellung nach, ob Extraposition zur Verarbeitungserleichterung eines Satzes dienen kann. Dazu untersucht sie, inwiefern Informationsdichte als Erklärung für die optionale Nachfeldstellung von Nominalphrasen, Präpositionalphrasen und Relativsätzen im frühen Neuhochdeutschen (hier verstanden als Zeitraum von 1650 bis 1900) verantwortlich gemacht werden kann.

Ausgangspunkt dieser Fragestellung ist das Phänomen im Deutschen, dass die oben genannten Phrasen und attributiven Relativsätze zwar wesentlich häufiger im Vor- oder Mittelfeld eines Satzes auftreten, aber auch im Nachfeld stehen können (vgl. z. B. Zifonun et al. 1997; Wöllstein 2010; 2014). Es ist davon auszugehen, dass die Extraposition merkmalsgetrieben ist. Grammatiken und bisherige empirische Studien zu diesem Thema führen Auslagerung unter anderem auf den Faktor der Länge der ausgelagerten Elemente und den Faktor der Mündlichkeitsnähe der Sätze beziehungsweise Texte zurück, in denen das Phänomen beobachtet wird (Kapitel 3).

Meine Arbeit verfolgt den neuen Ansatz, Auslagerung mithilfe von Informationsdichte zu erklären. Informationsdichte kann nach Shannon (1948) und Hale (2001) so verstanden werden, dass der Verarbeitungsaufwand, gemessen als Surprisal, mit der Auftretenswahrscheinlichkeit eines Wortes in einem bestimmten Kontext zusammenhängt (Kapitel 2.1). Ändert man den Kontext, ändert sich auch der Verarbeitungsaufwand (vgl. z. B. Levy und Jaeger 2007; Levy 2008). Ob die Verarbeitung erfolgreich ist, hängt aber auch davon ab, ob ausreichend kognitive Kapazitäten dafür zur Verfügung stehen (vgl. z. B. Levy und Jaeger 2007; Levy 2008). Empirische Untersuchungen zu dem Thema, ob verschiedene Kapazitäten für verschiedene Positionen des deutschen Satzes angenommen werden können, sind zum aktuellen Zeitpunkt ein Desi-

derat (zu den theoretischen Überlegungen siehe Vinckel-Roisin 2012; Speyer 2015a). Dennoch ist begründet zu vermuten, dass das Nachfeld im Deutschen Vorteile für die Verarbeitung bringen kann.

Eine der bedeutendsten Charakteristiken des Deutschen ist die Verteilung des Prädikats auf die beiden Satzklammern, die durch ein mit Konstituenten besetztes Mittelfeld voneinander getrennt sind (Kapitel 2.2). In den meisten Fällen tritt das lexikalische Verb, das die Satzaussage und die Verbalenz trägt, erst in der rechten Satzklammer auf. Auch wenn die vorherigen Wörter und Konstituenten sowohl in ihrer Bedeutung als auch in ihrer syntaktischen Funktion Vorhersagen über den Inhalt der rechten Satzklammer erlauben, kann trotzdem davon ausgegangen werden, dass der Satz erst mit ihrer Verarbeitung vollständig verstanden werden kann, weil viele Informationen essenziell vom lexikalischen Verb abhängen. Zahlreiche Verarbeitungsressourcen werden bis zur rechten Satzklammer folglich dafür verwendet, neue Wörter in die Struktur zu integrieren und von ihnen ausgehende Vorhersagen beizubehalten oder zu verwerfen, ohne gleichzeitig zu vergessen, was bereits verarbeitet wurde. Die Kosten während des Satzverständnisses können nach Gibson (1998) in einen vorwärtsgerichteten, erwartungsbasierten Aufwand (Integration Cost) und einen rückwärtsgerichteten, gedächtnisbasierten Aufwand (Memory Cost) unterteilt werden. Letzteres kann im Vergessen von Informationen, die früh im Satz standen, resultieren, wodurch keine Vorhersagen mehr auf Basis dieser Wörter getroffen werden können, was wiederum die Verarbeitung anderer Bestandteile im Satz erschwert. Beides kann durch den Zeitpunkt der Präsentation der rechten Satzklammer beeinflusst werden, da sich an dem Punkt beispielsweise Unsicherheiten darüber auflösen, welche Mitspieler noch an der Verbalhandlung beteiligt sein müssen. Daraus lässt sich schließen, dass mehr Verarbeitungskapazitäten im Nachfeld im Vergleich zum Mittelfeld frei sind, um lexikalisch verarbeitungsaufwändige Elemente zu verstehen.

Daher stelle ich als Hypothese auf, dass Nominal- und Präpositionalphrasen sowie Relativsätze dann ausgelagert werden, wenn ihr durchschnittlicher Surprisalwert hoch ist (Kapitel 1.1). Zugleich muss aber auch den Ergebnissen

bisheriger Studien Rechnung getragen werden, die Extraposition vor allem im Zusammenhang mit der Länge der ausgelagerten Elemente betrachtet (Kapitel 3). Daher wird dieser Faktor auch in die Untersuchungen einbezogen. Ist die Länge einflussreicher als die Surprisalwerte, ist davon auszugehen, dass die vermuteten höheren Kapazitäten im Nachfeld weniger ausschlaggebend sind als die frühere Präsentation der rechten Satzklammer.

Neben dem Einfluss der Länge wird in den bisher erschienenen empirischen Studien zu Extraposition auch der Einfluss der Mündlichkeitsnähe eines Texts aufgeführt (Kapitel 2.7). Diese mache Auslagerung wahrscheinlicher (vgl. Averintseva-Klisch 2007; Imo 2015; Ortmann und Dipper 2019). Bisher fehlen aber auch Studien, die untersuchen, ob der Einfluss von Informationsdichte in mündlichkeitsnahen Texten größer ist als in schriftlichkeitsnahen Texten, wenn beide im grafischen Medium präsentiert werden (mit Ausnahme von Ortmann 2023). Auch dieser Frage wird eine eigene Hypothese gewidmet. Abschließend wird untersucht, ob Sprachwandel anhand der Einflussfaktoren sichtbar wird. Die Studien von Schildt (1976) und Speyer (2015a) zeigen nämlich, dass informationsstrukturelle Kriterien im Laufe des Frühneuhochdeutschen weniger relevant für Nachfeldstellung werden. Informationsstruktur und Informationsdichte können nach aktuellem Forschungsstand nicht vollständig korreliert werden. Trotzdem wird mit Hypothese 3 untersucht, ob der Einfluss von Informationsdichte auf Auslagerung im Laufe des untersuchten Zeitabschnitts geringer wird. Ich verfolge in dieser Arbeit also diese Hypothesen:

- (H1) Relativsätze und Phrasen mit hohen Surprisalwerten werden ausgelagert.
- (H2a) Auslagerung wird verstärkt in mündlichkeitsnahen Texten verwendet.
- (H2b) In Texten, die mündlichkeitsnäher sind, ist der Einfluss von hohen Surprisalwerten größer als in schriftlichkeitsnahen Texten.
- (H3) Über die Zeit wird der Einfluss der Informationsdichte auf Auslagerung geringer.

Um Belege für die Korrektheit der Hypothesen zu finden, untersuche ich ein Korpus aus wissenschaftlichen Texten des frühen Neuhochdeutschen (Kapitel 4.1). Die Wahl wissenschaftlicher Texte erfolgte, da die vorliegende Arbeit im Kontext des Sonderforschungsbereichs 1102 verfasst wurde und es Teil des Projekts war, auch das wissenschaftliche Genre zu untersuchen. Neben dieser eher praktischen Motivation für die Auswahl des Korpus wird auch ein Forschungsdesiderat mit der Wahl behoben. Arbeiten, die Wissenschaftssprache in den Fokus nehmen, sind meines Wissens nach auch für das moderne Standarddeutsche eine Seltenheit und für frühere Zeitabschnitte fehlen sie fast vollkommen, obwohl es seit dem siebzehnten Jahrhundert wissenschaftliche Aufsätze und Monografien auf Deutsch gibt. Die Ergebnisse dieser Arbeit können also zum Anlass genommen werden, der Textsorte in weiterführenden Arbeiten mehr Aufmerksamkeit zu schenken.

Das verwendete Korpus setzt sich aus medizinischen und theologischen Texten zusammen, die arbiträr aus den entsprechenden Kategorien des Deutschen Textarchivs (vgl. BBAW 2019) gewählt wurden. Die Genres wurden als Repräsentanten für natur- und geisteswissenschaftliche Fächer und Schreibtraditionen ausgewählt. In diesen Texten wurden manuell extrapolierte Nominalphrasen und Präpositionalphrasen exhaustiv annotiert (Kapitel 4.3). Für jede dieser Phrasen wurde ein eingebettetes Vergleichselement gesucht, wobei das wichtigste Kriterium für diese Auswahl eine vergleichbare Komplexität der Phrase, die auch mit einer ähnlichen Länge einhergeht, war. Gerade die manuelle Wahl einer Vergleichsphrase birgt das Risiko, Phrasen zu wählen, von denen die annotierende Person glaubt, dass sie zu einer Bestätigung der Hypothese führen können. Durch die gegebenen Kriterien wurde versucht dem entgegenzuwirken. Dafür wurde auch in Kauf genommen, dass nicht immer die gleiche Anzahl extrapolierte und *in situ*-Phrasen gefunden werden konnte.¹

Das Problem schwierig zu erreichender Objektivität entfällt bei der Annotation der Relativsätze. Hier wurden alle attributiven adjazenten und extra-

¹ Ein nachträglicher Vergleich der hier annotierten eingebetteten Phrasen mit zu fällig ausgewählten *in situ*-Nominal- und Präpositionalphrasen zeigte insgesamt, dass diese Phrasen im Vorfeld und Mittelfeld kürzer sind als extrapolierte Nominal- und Präpositionalphrasen.

ponierten Relativsätze untersucht.² Auch die Satzklammern wurden in den Sätzen manuell annotiert, in denen ein Phänomen gefunden wurde. Handelte es sich bei der Auslagerung um einen Relativsatz, wurden außerdem die Kopfnomen und der Relativsatztyp, das heißt seine Restriktivität, bestimmt. Automatisch wurden die Länge der ausgelagerten Phänomene und ihre Skipgram-Surprisalwerte bestimmt (Kapitel 5.1). Halbautomatisch erfolgte die Bestimmung der Distanz zwischen dem Relativsatzanfang und dem Antezedens, da die vollautomatische Bestimmung bei der hypothetischen Distanz, über die ein adjazenter Relativsatz bewegt werden müsste, zu fehleranfällig ist. Neben diesen die Phänomene betreffenden Annotationen wurde der Orality Score pro Text mithilfe von COAST (Ortmann und Dipper 2024) bestimmt. Der Score gibt Aufschluss über die Mündlichkeitsnähe eines Textes (Kapitel 5.3). Außerdem wurden die Texte entsprechend ihrer Einordnung im DTA in medizinische und theologische geteilt. Anhand ihres Veröffentlichungsdatums wurde das Korpus zudem in fünf Zeitabschnitte (1650 bis 1700, 1700 bis 1750, 1750 bis 1800, 1800 bis 1850, 1850 bis 1900) untergliedert.

Zur Messung von Informationsdichte wurden Skipgram-Surprisalwerte auf Lemmata (vgl. Guthrie et al. 2006) verwendet, um das Problem der Datenknappheit im Untersuchungszeitraum zu umgehen (Kapitel 5.1). Alternativen, wie den Kontext für die Surprisalwertberechnung zu verkleinern oder ein größeres Trainingskorpus zu kreieren, konnten nicht verwendet werden, ohne entweder die Qualität der Aussage über die Werte zu verschlechtern, da nur möglichst große n-gram-Modelle eine gute Repräsentation von Sprache ermöglichen (vgl. z. B. Manning und Schütze 2001), oder an die Grenze der vorhandenen Daten in diesem Zeitraum zu stoßen. Als Trainingskorpora wurden alle nicht in das Testkorpus integrierten medizinischen und theologischen Texte des DTA aus den entsprechenden Zeitabschnitten verwendet (Anhang C). Die Berechnung der Skipgram-Surprisalwerte erfolgte über ein SFB-internes Tool (vgl. Kusmirek et al. 2023).

Bisherige korpusbasierte Studien, die sich mit syntaktischer Variation be-

² Zusätzlich wurde eine Kategorie der ambigen Relativsätze geschaffen. Diese betrifft Relativsätze, die adjazent am äußeren rechten Rand eines Mittelfeldes stehen, das keine rechte Satzklammer aufweist. Sie wurden aus der Analyse ausgeschlossen.

fassen, schauen meistens auf punktuelle Veränderungen im Surprisalprofil eines Satzes. Dieses Vorgehen ist für diese Arbeit deshalb nicht gut geeignet, weil ich die Surprisalwerte der gesamten Auslagerung einbeziehen möchte. Daher werden hier zwei neuartige Berechnungsmethoden präsentiert: das kumulative und durchschnittliche Surprisal (Kapitel 5.2).³ Für die hiesige Analyse wurde das kumulative Surprisal ausgeschlossen, da die Korrelation mit der Länge zu groß ist. Stattdessen wird nur das durchschnittliche Surprisal über das arithmetische Mittel aller Surprisalwerte der Auslagerung berechnet.

Die eben genannten Variablen (Skipgram-Surprisalwerte, Länge, Genre der Texte, Orality Score der Texte, gegebenenfalls Relativsatztyp und Distanz und ihre Zweibegeinteraktionen) wurden mithilfe von logistischer Regression auf ihren Einfluss auf die Position der jeweiligen Elemente untersucht (Kapitel 6). Um Belege für Hypothese 1 zu finden, wurden die Elemente zunächst im Gesamtkorpus untersucht. Da sich an dieser Stelle aber bereits ein Einfluss des Genres zeigte, wurden zusätzlich nach Genre getrennte Korpora als Datengrundlage für eine weitere logistische Regression genutzt.

Die Ergebnisse der Untersuchung unterscheiden sich zwischen den betrachteten Phänomenen und vor allem auch zwischen den Teilkorpora (Kapitel 6.5). Für alle übereinstimmend ist eine Entwicklung dahingehend zu erkennen, dass Extraposition in jüngeren und mündlichkeitsnahen Texten wahrscheinlicher wird. Diese Aussage ist allerdings unabhängig von den Eigenschaften der Auslagerung. Deren Einfluss divergiert stark zwischen den einzelnen Phänomenen.

Für die Nominalphrasen kann Hypothese 1 weitestgehend bestätigt werden. Sowohl im Gesamtkorpus als auch im medizinischen Teilkorpus ist der Einfluss der Surprisalwerte signifikant und erlaubt die Vorhersage der Nachfeldstellung in Übereinstimmung mit der Hypothese. Allerdings kann dieses Ergebnis nicht für das theologische Teilkorpus gefunden werden. Außerdem erklären die Skipgram-Surprisalwerte nicht alle Daten im Gesamtkorpus. Der Einfluss des Genres war auch ausschlaggebend für die getrennte Analyse

³ Vorherige Anwendungen dieser Methoden stehen ebenfalls im Kontext des SFB-Projekts und wurden von Speyer und Lemke (2017), Speyer und Voigtmann (2021a;b), Voigtmann und Speyer (2021) und Voigtmann und Speyer (2023) vorgestellt.

nach Textsorten. Entgegen den Ergebnissen bisheriger Untersuchungen weist Mündlichkeitsnähe in den logistischen Regressionen, die zur Testung dieser Hypothese verwendet wurden, allerdings oft auf Einbettung hin. Insgesamt zeigt sich für die Nominalphrasen also der vermutete Effekt, dass die freie Kapazität hinter der rechten Satzklammer dafür genutzt werden (muss), kompliziertere Nominalphrasen zu verarbeiten. Hier interagieren die freiwerdenden Kapazitäten durch das vollständige Erfassen der rechten Satzklammer und des darin stehenden lexikalischen Verbs mit der früheren Präsentation eben dieses Satzteils. Die frühe Verarbeitung des Verbs ist also substantiell, weil es die Hauptlast der Satzaussage und den Valenzstatus trägt, sprich, der Lexikoneintrag des Verbs erkannt wurde (vgl. Gibson 1998). Die Rezipierenden wissen dann, welche Mitspieler das Verb verlangt und ob diese schon verarbeitet wurden. Auch Unsicherheiten über das Verb selbst werden an dieser Stelle behoben, wodurch die Integration Cost verringert wird. Mit der Verarbeitung der rechten Satzklammer ist den Rezipierenden also bewusst, ob und gegebenenfalls welche Argumente noch fehlen. Wenn dann im Nachfeld eine weitere Konstituente erkannt wird, entfällt die Schwierigkeit, deren Funktion vorhersagen zu müssen. Auch der Kasus der Nominalphrase wird in den allermeisten Fällen unmittelbar nach der Satzklammer verarbeitet, weil er auf dem Determinierer sichtbar wird. Die freie, kognitive Kapazität kann einzig auf die lexikalische Verarbeitung konzentriert werden, die mithilfe der Skipgram-Surprisalwerten gemessen wird. Für die Nominalphrasen ist der Einfluss der Skipgramwerte daher also erklärt. Es muss an dieser Stelle natürlich bemerkt werden, dass der Verarbeitungseinfluss durch die Satzgliedfunktionserkennung zumindest in dieser Arbeit und mit den hier verwendeten Daten nicht nachgewiesen werden kann. Eine weitere Korpusuntersuchung, die nicht auf Lemmata basiert, könnte dieses Desiderat beheben, wenn mehr Daten aus dem frühen Neuhochdeutschen zur Verfügung stünden, um normalisierte Daten zu verwenden. Alternativ könnten die Satzgliedfunktionen annotiert werden und mit ihrer Hilfe Analysen angestellt werden. Aktuell liegen nach meinem Kenntnisstand jedoch keine Korpora vor, die diese Ansprüche erfüllen, weshalb hier qualitativ argumentiert wurde.

Die Besetzungsvariationen der rechten Satzklammer erklären unter anderem auch den Unterschied zwischen den medizinischen und theologischen Texten. Bei einer nachträglichen Untersuchung der Unigram-Surprisalwerte der rechten Satzklammern zeigte sich, dass die medizinischen Texte seltenere Verben verwenden als die theologischen Texte. Das unterstützt die gesamte Annahme, dass in den medizinischen Texten auch seltenere und neuere Inhalte besprochen wurden, was auch durch die Darstellung von beispielsweise Eckart (2021) unterstützt wird. Im betrachteten Zeitraum fanden diverse Veränderungen in der Medizin statt. Unter anderem wurde der Einfluss chemikalischer Prozesse als bedeutsam wahrgenommen und Naturwissenschaftler sowie Mediziner begannen Experimente zum Erkenntnisgewinn zu nutzen (vgl. z. B. Eckart 2021). Das hohe Maß an neuen Ideen, Konzepten und Erkenntnissen in diesem Genre bedingt auch die Verwendung neuer Worte, die wiederum neue Valenzen mit sich bringen. Für die Rezipierenden ist es also im medizinischen Kontext umso wichtiger, frühzeitig die rechte Satzklammer zu erkennen und mehr kognitive Kapazitäten darauf verwenden zu können, die im Nachfeld stehenden Konstituenten zu verarbeiten.

In den theologischen Texten ist dieser Effekt weniger stark ausgeprägt. Zum einen kann es daran liegen, dass die Texte nach heutigen Maßstäben weniger wissenschaftlich sind als die medizinischen. Experimente können naturgemäß nicht durchgeführt werden und auch die Themen der Texte haben sich höchstens marginal verändert. Naturwissenschaftliches Denken floss in ihre Gestaltung sicherlich nicht mit ein, was sich auch darin spiegelt, dass die Unigram-Surprisalwerte der rechten Satzklammern in diesem Teilkorpus signifikant niedrigere Werte aufweisen als die der medizinischen Texte. Es ist also davon auszugehen, dass auch die Valenzstatus bekannter und die Sätze daher einfacher als Sätze aus den medizinischen Texten zu verarbeiten waren. Das kann auch mit den alltagsnäheren und -relevanteren Inhalten zusammenhängen. Außerdem dürfte die Schreibpraxis der Autoren einen Einfluss auf die Gestaltung der Sätze ausüben. Bei den Medizinern ist begründet zu vermuten, dass sie vornehmlich für andere Mediziner, also ein sehr homogenes Publikum schreiben. Bei den Theologen hingegen ist anzunehmen, dass ihr

Publikum heterogener ist und sie auch mehr mündlichkeitsnahe, oral rezipierte Texte verfasst haben. Die Gewöhnung an diesen Schreibstil kann auch dazu beitragen, dass er sich in den mehr wissenschaftlichen Werken niederschlägt. Die Gestaltung der Texte ließe sich folglich sowohl der Hörer- als auch der Sprecherorientierung zuweisen. Zudem lässt sich bei der Analyse der theologischen Teilkorpora erkennen, dass auch dieses Genre über den untersuchten Zeitraum hinweg schriftlichkeitsnäher und im weitesten Sinne des Wortes naturwissenschaftlicher wird. An dieser Stelle wären weitere Untersuchungen anderer naturwissenschaftlicher und geisteswissenschaftlicher Texte wünschenswert, um die Gültigkeit dieser Ergebnisse auch darin überprüfen und im Idealfall nachweisen zu können.

Die Bedeutung der Satzklammer für die Verarbeitung des Satzes wird auch für die Präpositionalphrasen als Erklärung herangezogen. In den drei untersuchten Korpora wurden für Hypothese H1 keine Belege gefunden, da die Werte der Skipgram-Surprisal nicht signifikant sind. Stattdessen wird übereinstimmend mit früheren Forschungsergebnissen vor allem ein Einfluss der Länge sichtbar. Wenn lange Präpositionalphrasen ausgelagert werden, verkürzt sich im Vergleich zur Einbettung dieser Phrasen, die auch immer eine Option ist, der Satzrahmen. Die oben skizzierten Vorteile kommen also auch für dieses Phänomen zum Tragen. Auch der Unterschied zwischen den theologischen und den medizinischen Texten wurde erneut zur Argumentation hinzugefügt. Zum einen wurde damit argumentiert, dass die Präpositionalphrasen weniger Funktionen im Satz erfüllen können, als das für Nominalphrasen der Fall ist. Die Unsicherheit darüber, welche Funktion noch fehlt, ist also insgesamt kleiner, sodass auch generell weniger kognitive Ressourcen darauf verwendet werden. Das würde wiederum bedeuten, dass auch mehr kognitive Kapazitäten für das lexikalische Verständnis der Präpositionalphrase aufgewendet werden können.

Es ist aber genauso denkbar, dass die Präposition selbst und der nachfolgende Determinierer für die Berechnung der Surprisalwerte und für die Unterscheidung zwischen den ausgelagerten und eingebetteten Phrasen problematisch sind. Beide Wörter sind Funktionswörter. Das bedeutet, dass sie per

se hochfrequent sind, was noch dadurch verstärkt wird, dass die Berechnung der Surprisalwerte auf Lemmata erfolgte, weshalb Kasusunterschiede im Artikel auch wegfallen. Die Surprisalwerte auf diesen beiden Wörtern dürften also auch Anteil daran haben, dass die Unterschiede zwischen den eingebetteten und ausgelagerten Präpositionalphrasen in dieser Variable nicht so ausgeprägt sind, dass sich ein Effekt im Modell beobachten ließe. Wiederum wäre es hier wünschenswert, normalisierte Daten erheben zu können, ohne auf das Problem der Datenknappheit zu stoßen.

Gleichzeitig lässt sich der signifikanteste Prädiktor, die Länge, auch auf die Memory Cost (vgl. Gibson 1998) beziehen. Würden die langen Konstituenten eingebettet stehen, würde sich die Distanz zwischen den Satzklammern zu stark vergrößern, sodass Konstituenten, die am Satzanfang gestanden haben, vergessen würden beziehungsweise nicht mehr zur Vorhersageerleichterung von anderen Elementen im Satz genutzt werden können. Die Verkürzung des Satzrahmens kann insgesamt für die Präpositionalphrasen als ausschlaggebender erachtet werden und erfordert in Anbetracht der Ergebnisse auch nicht den Umweg über die Valenzstatus der Präpositionalphrasen. Auch zeigt sich im Fall der Präpositionalphrasen wieder der Unterschied zwischen den beiden untersuchten Genres. Aufgrund der schwierigen und weniger alltagsrelevanten Themen und Wörter der medizinischen Texte kommt der Satzrahmenverkürzung durch Extraposition eine größere Bedeutung zu, weil angenommen werden kann, dass sie die kognitiven Kapazitäten noch stärker beanspruchen als die theologischen Texte. Im theologischen Teilkorpus gewinnt die Distanzverringerng des Satzrahmens erst in jüngeren Texten an Bedeutung, was wieder auf einen Wechsel im Schreibstil der Theologen und vielleicht auf eine Festigung des wissenschaftlichen Stils generell hindeutet.

Auch bei den Relativsätzen erklärt die Länge mehr Daten, als Surprisalwerte es tun, obwohl letztere ebenfalls tendenziell signifikant sind. Zudem ist der Einfluss der Länge in den medizinischen Texten größer als in den theologischen Texten. Daher sehe ich einen weiteren Hinweis darauf, dass die Divergenz zwischen den beiden Genres und den wissenschaftlichen Praktiken als vereinigender Faktor bei der Analyse angesehen werden kann. Alle Ergeb-

nisse weisen darauf hin, dass die Verkürzung des Satzrahmens bei den untersuchten Elementen entscheidend für die optimale Gestaltung deutscher Sätze zu sein scheint. Er überschreibt sogar die vermutete Vergrößerung der kognitiven Kapazitäten, die für das Nachfeld besprochen wurde. Zugleich wurde über die Verarbeitungsanforderungen in Abhängigkeit von der Bekanntheit und Frequenz der verwendeten Worte dafür argumentiert, dass diese auch einen Einfluss darauf haben, wie optimal die zur Verfügung stehenden kognitiven Kapazitäten verwendet werden müssen.

Die Analyse der Relativsätze hat zudem gezeigt, dass der Relativsatztyp mit dem Genre interagiert. Zum einen zeigt das erneut die Relevanz der differenzierten Betrachtung der beiden Genre. Zum anderen lässt sich das Ergebnis, dass Relativsätze in theologischen Texten eher adjazent stehen, aber auch als Hinweis auf die allgemeine Erwartbarkeit eines Relativsatzes sehen. Während die appositiven Relativsätze in den medizinischen Texten eher ausgelagert werden, stehen sie in den theologischen eher adjazent. Anders gesagt, stehen restriktive Relativsätze in den medizinischen Texten also eher adjazent. Für sie kann eine allgemeine Verarbeitungserleichterung dadurch angenommen werden, dass ein Antezedens vollständig identifiziert wird, was der Funktion von restriktiven Relativsätzen entspricht. Dafür wird auch eine Verlängerung des Satzrahmens in Kauf genommen, wobei anzumerken ist, dass die Länge ein erklärungsstärkerer Prädiktor als der Relativsatztyp ist. Dass appositive Relativsätze folglich in diesem Korpus erfolgreich ins Nachfeld gestellt werden können, liegt wieder an den Inhalten der medizinischen Texte. Zusätzliche Informationen über komplizierte Kopfnomen sollten auch an dieser Position noch gut aufgenommen und für das Verständnis integriert werden können. Aufgrund der nicht-signifikanten Interaktion zwischen den Skipgramwerten und der Restriktivität, ist dieses Ergebnis unabhängig davon, ob die Skipgramwerte des Relativsatzes hoch sind oder nicht.

Bei den theologischen Texten weisen abgesehen von den Surprisalwerten und der Länge mehr Faktoren auf Einbettung hin. An entsprechender Stelle wurde dafür argumentiert, dass die Extraposition in theologischen Texten unter dem Faktor der Mündlichkeitsnähe schlechter bewertet wird, weil die

Trennung der Relativsätze von ihren Kopfnomen dafür sorgen kann, dass diese generell nicht mehr erwartet werden. Wieder kommt als Erklärung hierfür der einfachere, alltagsrelevantere und weniger Veränderungen unterworfenen Inhalt der Texte in Frage. Wenn die Kopfnomen insgesamt erwartbarer sind, sind zusätzliche Informationen über sie nicht so relevant für ihre Verarbeitung. Daher ist anzunehmen, dass die Relativsätze nicht aufgrund ihres Inhalts unerwartbarer sind, sondern generell nicht mehr als notwendig erachtet werden und ihre Extraposition deshalb unterlassen wird. Wiederum müsste dieses Argument mithilfe eines besseren Language Models untermauert werden, das den Relativsatz als Konstituente im gleichzeitigen Kontext der inhaltlich-lexikalischen Erwartungen an den Satz vorhersagen kann. Zum aktuellen Zeitpunkt ist ein solches Modell ein Desiderat, weshalb auch hier ausschließlich qualitativ argumentiert werden muss.

Die Ergebnislage zum Einfluss des Orality Scores ist in den betrachteten Regressionen eher diffus. Zwar wurde dafür argumentiert, dass er eher für Auslagerung sprechen sollte, allerdings weist er in den Regressionen immer wieder auf Einbettung hin. Vor allem in den theologischen Korpora wird dieser Einfluss sichtbar. An den entsprechenden Stellen wurde damit argumentiert, dass die Einbettung der Phrasen oder Relativsätze andere Elemente im Satz erwartbarer machen kann, obwohl der Satzrahmen dadurch verlängert wird. Auch wurde dahingehend argumentiert, dass die Vertrautheit mit dem Lateinischen besonders bei den Theologen und in den älteren Texten noch größer war, weshalb längere Satzrahmen in der theologischen Textsorte ein kleineres Problem darstellen sollten, weil sowohl die Schreibenden als auch die Rezipierenden an längere Abhängigkeiten gewöhnt waren.

Auf der anderen Seite ist denkbar, dass man hier eine klarere Unterscheidung zwischen konzeptioneller und medialer Mündlichkeit hätte treffen müssen. Die Eigenschaften konzeptioneller Mündlichkeit wurden oft auch aus Texten genommen, die oral perzipiert wurden. Durch das flüchtige orale Medium werden die kognitiven Kapazitäten natürlich anders gefordert als im grafischen Medium. Letzteres ist aber die Art, wie die hier vorliegenden Texte wahrgenommen wurden. Es kann also davon ausgegangen werden, dass

dieser Unterschied auch den Autoren bewusst war und das einen Einfluss auf die Auslagerungspraktiken der untersuchten Phrasen hat, obwohl mehr Merkmale verwendet werden, die als typisch für konzeptionell und medial mündliche Texte in diesen Fällen galten.

Zusätzlich zur Gesamtheit der Nominal- und Präpositionalphrasen wurden die attributiven Nominal- und Präpositionalphrasen separat analysiert. Problematisch dabei ist, dass jeweils weniger als fünfzig Belege gefunden wurden, was die Aussagekraft der Ergebnisse naturgemäß stark einschränkt. Für die Nominalphrasen kann die erste Hypothese bestätigt werden: Surprisalwerte bilden den stärksten Prädiktor für Auslagerung. Zusätzlich dazu zeigt die Interaktion aus Länge und Zeitstufe, dass sich dies in jüngeren Texten geändert haben dürfte. Als Erklärung wurde angeführt, dass die Funktion der Phrase im Nachfeld nicht mehr verarbeitet werden muss und zusätzlich eine Verarbeitungserleichterung für das Prädikat gewährleistet sein kann, weil das Antezedens der attributiven Nominalphrase im Mittelfeld verarbeitet wurde. Auch für die attributiven Präpositionalphrasen weisen die Skipgramwerte signifikant auf Extraposition hin, allerdings ist dieser Wert nur schwach signifikant, während die anderen Werte deutlich bessere Vorhersagen über die Position der Präpositionalphrasen machen. Auch die Interaktionen zwischen den Surprisalwerten und beispielsweise dem Orality Score weisen signifikantere Werte auf und sagen Einbettung voraus. Das lässt darauf schließen, dass weniger informative Phrasen ins Nachfeld extraponiert werden, weil sie für die Verarbeitung beziehungsweise Identifikation des Antezedens weniger relevant sind und daher eher von ihm getrennt stehen können. Daher würde sich diese Argumentation in die zu den Relativsatztypen einreihen, wo erklärt wurde, welchen Einfluss die Auftretenswahrscheinlichkeit eines Relativsatzes als (Teil-)Satzglied auf dessen Position ausübt. Trotzdem muss hervorgehoben werden, dass sowohl die Analyse der Nominalphrase als auch die der Präpositionalphrasen, die Attributstatus besitzen, schwer als allgemeingültig zu betrachten ist, weil die Datenmenge so gering ist.

Zu Hypothese 1 ist abschließend zu sagen, dass sie zwar nicht für alle untersuchten Auslagerungsphänomene bestätigt werden kann, aber bei den

Nominalphrasen (außer denen in theologischen Texten), den attributiven Präpositionalphrasen und den Relativsätzen, wenn der Faktor Distanz zwischen Antezedenz und Relativsatzanfang nicht in die Analyse einfließt, dennoch als Erklärung für Auslagerung dienen kann. Auch in den Fällen, in denen andere Faktoren, wie die Länge oder die Zugehörigkeit zum theologischen oder medizinischen Teilkorpus, mehr Datenpunkte erklären können, konnte die Diskussion in Kapitel 6.5 als auch ihre Zusammenfassung hier aber deutlich machen, dass die Freiheit in der Stellung von Nominalphrasen, Präpositionalphrasen und Relativsätzen auf Prinzipien zur Verarbeitungserleichterung zurückgeführt werden kann. Diese richtet sich allerdings oft weniger nach der lexikalischen Vorhersagbarkeit, als dass sie eine Mischung aus Funktions- und Bedeutungsvorhersagen widerspiegelt, die allerdings von den verwendeten Surprisalwerten nicht gemessen werden können und für deren Bestimmung in historischen, deutschen Daten meines Wissens bisher kein Mittel existiert.

Essenziell für die Verarbeitung des Satzes ist die frühe Verarbeitung der rechten Satzklammer, was sich besonders durch den starken Einfluss der Länge auf die Position der auslagerbaren Konstituenten zeigt. Daneben wurde auch über die beiden Verarbeitungskosten, die von Gibson (1998) vorgestellt wurden, argumentiert. Auch ich nehme an, dass es nur einen begrenzten Pool kognitiver Kapazität gibt, dessen Mittel optimal ausgenutzt werden sollten. Entweder wird ein Satz also so gestaltet, dass die Vorhersage von Material erleichtert wird, oder es greifen Prinzipien, die gegen das Vergessen von Elementen sprechen, die eben diese Vorhersageerleichterung unterstützen können. Gleichzeitig spielt es eine Rolle, wie kompliziert die Inhalte der Texte sind und ob die Rezipierenden und Schreibenden vertraut mit den Inhalten sind. Das bildet wiederum eine Brücke zu der gemessenen lexikalischen Bedeutung und auch zu den von Futrell et al. (2021) vorgestellten Gewöhnungseffekten, obwohl die bei der Betrachtung dieser Hypothese noch einen eher lexikalischen und keinen strukturellen Effekt besitzen. Letzterer wird für die Interpretation der Ergebnisse von Hypothese 3 relevant. Davor steht allerdings die Betrachtung von Hypothese 2.

Um Belege für den ersten Teil der zweiten Hypothese⁴ zu finden, wurde für die drei Phänomene jeweils die Korrelation zwischen dem Anteil der Sätze mit Auslagerung an allen Sätzen und dem Orality Score der Texte berechnet (Kapitel 7). Eine positive, bedeutende Korrelation wurde nur für Nominalphrasen gefunden. Weder die Präpositionalphrasen noch die Relativsätze stehen hinsichtlich ihrer Extraponierbarkeit in einem Zusammenhang zur Mündlichkeitsnähe eines Textes. Die Hypothese kann also auch nur für Nominalphrasen bestätigt werden. Daraus ist zu schlussfolgern, dass für sie besondere Bedingungen bei der Extraposition gelten.

Für den zweiten Teil der zweiten Hypothese⁵ wurde das Korpus für alle Phänomene zusammengenommen und in zwei Teile entsprechend ihres Orality Scores geteilt. Als Trennwert wurde der mittlere Wert aus Tabelle 7.1 genutzt. Für die beiden Subkorpora wurde mithilfe der logistischen Regression überprüft, welche der Variablen Extraposition am besten vorhersagen kann. Mir ist bewusst, dass die Zweiteilung des Korpus einer groben Trennung entspricht und eine feingliedrigere Unterteilung nach den verschiedenen Orality Scores ebenfalls eine Möglichkeit gewesen wäre, um die Hypothese zu überprüfen. Allerdings wäre diese Unterteilung aufgrund der Tatsache, dass es keine festen Pole für Mündlichkeitsnähe gibt, ebenso willkürlich erfolgt wie die Zweiteilung, die zudem besser operationalisierbar war. Hypothese H2b wurde mit diesem Vorgehen bestätigt: In schriftlichkeitsnahen Texten sind Skipgram-Surprisalwerte nicht zur Vorhersage von Extraposition geeignet. In mündlichkeitsnahen Werken stellen sie zwar nicht den besten Prädiktor dar, können die Position der untersuchten Elemente aber erklären.

Auch die Belege für Hypothese 3 wurden mittels logistischer Regression gefunden (Kapitel 8). Die Korpora der Nominalphrasen, der Präpositionalphrasen und der Relativsätze wurden nach den oben genannten Zeitabschnitten unterteilt und für jeden Zeitabschnitt wurde eine eigene Regression durchgeführt, die zeigen konnte, ob und wie viel Skipgram-Surprisalwerte zur Positionsvorhersage beitragen können. Für alle drei Phänomene wurde

⁴ H2a „Auslagerung wird verstärkt in mündlichkeitsnahen Texten verwendet.“

⁵ H2b „In Texten, die mündlichkeitsnäher sind, ist der Einfluss von hohen Surprisalwerten größer als in schriftlichkeitsnahen Texten.“

gezeigt, dass die Skipgram-Surprisalwerte im letzten Zeitraum nicht mehr zur Vorhersage von Extraposition genutzt werden können. Bei den Nominal- und Präpositionalphrasen war die Informativität der Auslagerung zumindest noch von 1650 bis 1700 einer der signifikantesten Prädiktoren. Die Präpositionalphrasen weisen zudem im Zeitabschnitt 1800 bis 1850 einen signifikanten Einfluss von Skipgramwerten auf, was bei den Relativsätzen zu keinem Zeitpunkt unabhängig von anderen Faktoren gefunden werden konnte. Insgesamt sehe ich Hypothese 3 daher als bestätigt an und habe in Kapitel 8.4 dafür argumentiert, dass ein Einflusswandel von den vorhersagebasierten hin zu den gedächtnisbasierten Verarbeitungsschwierigkeiten vollzogen wird, was durch den zunehmenden Einfluss der Länge signalisiert wird.

Zum anderen kann die abnehmende Bedeutung der Surprisalwerte für eine bessere Kodierung der betrachteten Konstituenten sprechen. Dadurch würden sich, wie auch bei den Präpositionalphrasen bereits besprochen, die Surprisalwerte der extrapolierten und *in situ*-Phrasen so annähern, dass der Unterschied zu gering wird, um einen Effekt hinsichtlich der Position zu generieren. Zum anderen darf aber auch der Einfluss der Länge als zunehmend wichtigster Einzelprädiktor nicht vernachlässigt werden. Es wird also bedeutender, die Memory Cost möglichst zu verringern, indem der Satzrahmen eher kurzgehalten wird. Der punktuelle, starke Vorteil für die Integration Cost, der für die Nachfeldstellung hochinformativer Phrasen angenommen wurde, entfällt spätestens in den Texten des 19. Jahrhunderts. Dieser Effekt kann auch durch die Gewöhnung an kurze Satzrahmen, die bereits im 18. Jahrhundert als prestigeträchtig galten (vgl. Konopka 1996; Takada 1998), hervorgerufen worden sein. Die Bevorzugung dieser Stilistik sorgt zusätzlich dafür, dass derartige Konstruktionen durch strukturelle Gewöhnung an sie schneller und besser verarbeitet wurden.

Dieses Prinzip kann als allgemeingültig für die medizinischen und theologischen Autoren und deren jeweilige Publika gelten. Allerdings lässt sich wieder feststellen, dass die Theologen diese Entwicklung später vollzogen haben als die Mediziner, was vor allem an den Interaktionen zwischen Länge und Zeitstufe in diesen Teilkorpora sichtbar wird. Als Argument hierfür wur-

de wieder herangeführt, dass sich die Professionalisierung in diesem wissenschaftlichen Bereich später vollzogen hat als bei den Medizinern. In beiden Fällen zeigt sich dennoch die Abnahme des Einflusses von Surprisalwerten auf Extraposition und damit eine Bestätigung der dritten Hypothese.

Obwohl nicht für alle vorgestellten Hypothesen Belege gefunden werden konnten, gelingt es der vorliegenden Arbeit nachzuweisen, dass sich die Extraponierbarkeit von Nominalphrasen, Präpositionalphrasen und Relativsätzen auf Verarbeitungsvorteile für den Gesamtsatz beziehungsweise ihren Matrixsatz zurückführen lässt. Damit leistet sie methodische Pionierarbeit für diesen Erklärungsansatz im Deutschen und verwendet Methoden, namentlich die durchschnittlichen Skipgram-Surprisalwerte, die außerhalb des Forschungsprojekts, mit dem die Arbeit assoziiert ist, noch nicht verwendet wurden. Außerdem befasst sich die Arbeit mit wissenschaftlichen Texten und dem frühen Neuhochdeutschen. Für beide fehlt systematische, empirische Forschung besonders mit einer Perspektive des Sprachwandels weitestgehend. Das vorliegende Dissertationsprojekt leistet dabei einen Beitrag dazu, erste Ansätze für die Unterschiede zwischen natur- und geisteswissenschaftlichen Texten und für unterschiedliche Verarbeitungsansprüche je nach Inhalt und Komplexität eines Textes zu finden. Auf der Ebene der Verarbeitung argumentiert sie ferner für das Zutreffen eines gemeinsamen Verarbeitungskontingents, aus dem sowohl erinnerungsbezogene als auch vorhersagebezogene Kapazitäten geschöpft werden, die im Verlauf des Jahrhunderts unterschiedlich stark für die Verarbeitung verwendet werden. Auch das hängt mit Veränderungen in der Sprache und der Schreibpraxis im frühen Neuhochdeutschen zusammen, die bisher auch nicht ausreichend im Fokus der Sprachwissenschaft standen.

Im Verlauf der Arbeit konnte beziehungsweise musste allerdings immer wieder darauf hingewiesen werden, dass der Fokus auf diese untererforschten Gebiete Probleme mit sich bringt. Datenknappheit, die die Verwendung von Lemmata nötig macht, obwohl sich während der Analysen gezeigt hat, dass mindestens normalisierte Daten zu bevorzugen wären, und die niederqualitative Aufbereitung der vorhandenen Korpora sind nur zwei der Probleme, die

eine quantitativ-empirische Untermauerung der Argumente zum aktuellen Zeitpunkt erschweren. Auch fehlen Anknüpfungspunkte für die Betrachtung wissenschaftlicher Sprache, die nicht auf präskriptiven Werken zum Verfassen wissenschaftlicher Texte im modernen Standarddeutschen beruhen, ebenso wie bereits bestehende Analysen zum Sprachwandel in diesem Jahrhundert. Für letzteren findet die vorliegende Arbeit Hinweise, die als Anreiz dafür gesehen werden können, dass sich weitere Arbeiten mit dem Zeitabschnitt oder wissenschaftlichen Texten im frühen Neuhochdeutschen zu beschäftigen.

Literaturverzeichnis

Abel, H. K. (1699), *Wohlerfahrner Leib-Medicus Der Studenten, welcher So wohl allen auf Schulen Gymnasiis und Universitäten Lebenden oder auf Reisen begriffenen gelehrten Personen/ als auch allen Menschen insgemein die nöthigsten Reguln und herrlichsten Artzeneyen mittheilet/ Krafft deren sie nicht allein die Gesundheit nechst Gott erhalten/ sondern auch die zugestossenen Kranckheiten abwenden und vertreiben können*, Groschuff, Leipzig.

Admoni, V. G. (1970), *Der deutsche Sprachbau. 3. durchgesehene und erweiterte Aufl.*, Beck'sche Verlagsbuchhandlung, München.

Admoni, V. G. (1990), *Historische Syntax des Deutschen*, Niemeyer, Tübingen.

Ágel, V. (2000), *Syntax des Neuhochdeutschen bis zur Mitte des 20. Jahrhunderts*, in W. Besch, A. Betten, O. Reichmann und S. Sonderegger, eds, 'Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung', De Gruyter, Berlin, pp. 1855–1903.

Ágel, V. und Hennig, M. (2022), 'Kasseler Junktionskorpus (Version 1.1)'. Justus-Liebig-Universität Gießen.

URL: <https://doi.org/10.34644/laudatio-dev-nCQsCnMB7CArCQ9CDmun>

Ágel, V. und Henning, M. (2012), *Überlegungen zur Theorie und Praxis des Nähe- und Distanzsprechens*, in V. Ágel und M. Henning, eds, 'Zugänge zur Grammatik der gesprochenen Sprache', Vol. 269 of *Reihe Germanistische Linguistik*, Niemeyer, Tübingen.

Altmann, H. (1981), *Formen der Herausstellung im Deutschen: Rechtsversetzung, Linksversetzung, freies Thema und verwandte Konstruktionen*, Linguistische Arbeiten; 106, Niemeyer, Tübingen.

- Andresen, M. und Zinsmeister, H. (2019), *Korpuslinguistik*, Narr.
URL: <https://elibrary.narr.digital/book/99.125005/9783823392262>
- Arnold, J., Losongco, A., Wasow, T. und Ginstrom, R. (2000), 'Heaviness vs. Newness: The Effects of Structural Complexity and Discourse Status on Constituent Ordering', *Language* 76, 28–55.
- Askedal, J. O. (1986a), 'Zur vergleichenden Stellungsfelderanalyse von Verbal-sätzen und nicht-verbalen Satzgliedern im Deutschen.', *Deutsch als Fremdsprache: Zeitschrift zur Theorie und Praxis des Deutschunterrichts für Ausländer* 23(5), 269–273.
- Askedal, J. O. (1986b), 'Über Stellungsfelder und Satztypen im Deutschen', *Deutsche Sprache* 14, 193–223.
- Auer, P. (1991), 'Vom Ende deutscher Sätze', *Zeitschrift für germanistische Linguistik* 19(2), 141–157.
URL: <https://doi.org/10.1515/zfgl.1991.19.2.141>
- Averintseva-Klisch, M. (2007), Anaphoric properties of german right dislocation, in M. Schwarz-Friesel, M. Consten und M. Knees, eds, 'Anaphors in Text: Cognitive, formal and applied approaches to anaphoric reference', number 86 in 'Studies in Language Companion Series', John Benjamins Publishing Company, Amsterdam / Philadelphia, pp. 165–182.
- Aylett, M. und Turk, A. (2004), 'The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech', *Language and Speech* 47(1), 31–56.
- Bader, J. (2002), 'Schriftlichkeit und Mündlichkeit in der Chat-Kommunikation', *Networx* 29, 1–145.
- Bayer, J. (1984), 'COMP in Bavarian Syntax', *Linguistic Review* 3, 209–274.
- BBAW (2019), 'Deutsches Textarchiv. Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache'. Berlin-Brandenburgische Akademie der Wissenschaften; <http://www.deutschestextarchiv.de/>.

- Bech, G. (1955), *Studien über das deutsche Verbum infinitum*, Vol. 1 of *Historisk-filologiske meddelelser*; 35,2, Munksgaard, Kopenhagen.
- Behaghel, O. (1932), *Deutsche Syntax: eine geschichtliche Darstellung*, Vol. 4: Wortstellung, Periodenbau of *Germanische Bibliothek*, Winter, Heidelberg.
- Beneš, E. (1979), 'Die Ausklammerung im Deutschen als grammatische Norm und als stilistischer Effekt', *Muttersprache. Zeitschrift zur Pflege und Erforschung der deutschen Sprache* 78(10), 289–298.
- Bengel, J. A. (1751), *Abriß der so genannten Brüdergemeine, in welchem die Lehre und die ganze Sache geprüft, das Gute und Böse dabey unterschieden, und insonderheit die Spangenbergische Declaration erläutert wird*, Metzler, Stuttgart.
- Benner, J. H. (1739), *Christliches Bedencken von dem vorsetzlichen Meineid, wodurch das Recht des Nechsten wissentlich gekräncket wird Ob, und wie ein solcher Meineid von Gott vergeben werde?*, Brönner, Frankfurt am Main, Leipzig.
- Bennett, P., Durrell, M., Ensslin, A., Scheible, S. und Whitt, R. (2007), 'GerManC (Version 1.0)'. University of Manchester.
URL: <https://doi.org/10.34644/laudatio-dev-9SX7DnMB7CArCQ9CBBj>
- Bergmann, R., Moulin, C. und Ruge, N. (2019), *Alt- und Mittelhochdeutsch: Arbeitsbuch zur Grammatik der älteren deutschen Sprachstufen und zur deutschen Sprachgeschichte*, UTB; 3534, 10., überarbeitete und korrigierte auflage edn, Vandenhoeck & Ruprecht, Göttingen.
- Bhatt, C. (2005), *Die syntaktische Struktur der Nominalphrase im Deutschen*, Studien zur deutschen Grammatik; 38, unveränd. nachdr. der ausg. von 1990 edn, Stauffenburg, Tübingen.
- Biber, D. (2006), *University Language. A corpus-based study of spoken and written registers*, number v. 23 in 'Studies in Corpus Linguistics', John Benjamins Publishing Co, Amsterdam.
- Bies, A. (1996), *Syntax and discourse factors in early new high german: Evidence for verb-final word order*, Master's thesis, University of Pennsylvania.

- Birkner, K. (2008), *Relativ(satz)konstruktionen im gesprochenen Deutsch: syntaktische, prosodische, semantische und pragmatische Aspekte*, Linguistik - Impulse und Tendenzen; 28, de Gruyter, Berlin.
- Borter, A. (1982), *Syntaktische Klammerbildung in Notkers Psalter*, Das Althochdeutsche von St. Gallen; 7, de Gruyter, Berlin.
- Brandt, M., Reis, M., Rosengren, I. und Zimmermann, I. (1992), Satztyp, Satzmodus und Illokution. 3. Satzstruktur und Satztyp, in I. Rosengren, ed., 'Satz und Illokution', Max Niemeyer Verlag, Berlin, New York, pp. 4–33.
- Braunmüller, K. (1982), *Syntaxtypologische Studien zum Germanischen*, Tübinger Beiträge zur Linguistik; 197, Narr, Tübingen.
- Breindl, E. (2006), Präpositionalphrasen, in V. Ágel, ed., 'Dependenz und Valenz: ein internationales Handbuch der zeitgenössischen Forschung', Vol. 2 of *Handbücher zur Sprach- und Kommunikationswissenschaft*, de Gruyter, Berlin, Boston.
- Brinkmann, H. (1971), *Die deutsche Sprache: Gestalt und Leistung*, 2., neubearbeitete und erweiterte auflage edn, Schwann, Düsseldorf.
- Brooks, T. (2006), *Untersuchungen zur Syntax in oberdeutschen Drucken des 16. - 18. Jahrhunderts*, Schriften zur deutschen Sprache in österreich; 36, Lang, Frankfurt am Main.
- Brugmann, K. und Cohn, L. (1913), *Handbuch der klassischen Altertumswissenschaft*, Vol. Bd. 2, Abt. 1: Griechische Grammatik: Lautlehre, Stammbildungs- und Flexionslehre, Syntax, 4., verm. aufl. edn, Beck, München.
- Bräuner, J. J. (1714), *Pest-Büchlein: Oder Kurtzer/ doch gründlicher Unterricht Von der jetzmalen über Teutschland schwebenden gefährlichen Seuche der Pestilenz, In welchem wohlmeynend gezeiget wird/ Woher solche kommet/ wie sie ansteckend und fortgepflanztet/ auch wie sich [...] welche mit solchen Pest-inficirten Patienten umgehen müssen/ auch sonst jederman darbey zu verhalten habe: Nicht weniger auch/ wie man solche Seuche von sich durch Gottes Gnade abwenden/ und darwider*

glücklich präserviren und curiren kann Alles nach der sicherest und bewährtesten Methode [...] treulich entworffen/ und zu jedermans Nutzen und Gebrauch zum Druck befördert, Hocker, Frankfurt am Main.

Bühler, K. (1934), *Sprachtheorie: die Darstellungsfunktion der Sprache*, UTB für Wissenschaft, 3. aufl., ungekürzter neudr. d. ausg. jena, fischer, 1934 edn, Lucius und Lucius, Stuttgart. 1999.

Büring, D. und Hartmann, K. (1995), All right!, in U. Lutz und J. Pafel, eds, 'On Extraction and Extraposition in German', Vol. 11 of *Linguistik Aktuell/Linguistics Today*, John Benjamins Publishing Company, Amsterdam / Philadelphia, pp. 179–212.

Carus, C. G. (1820), *Lehrbuch der Gynäkologie. Bd. 1*, Fleischer, Leipzig.

Chen, S. F. und Goodman, J. (1999), 'An empirical study of smoothing techniques for language modeling', *Computer Speech & Language* **13**(4), 359–394.

Chomsky, N. (1975), *The logical structure of linguistic theory*, Plenum Press, New York, NY.

Collins, M. X. (2014), 'Information Density and Dependency Length as Complementary Cognitive Models', *Journal of Psycholinguistic Research* **43**(5), 651–681.

Coniglio, M., Linde, S. und Ruetten, T. (2017), 'Relative Clauses in Old High German: A Corpus-Based Statistical Investigation of Their Syntactic and Information-Structural Properties', *Journal of Germanic Linguistics* **29**(2), 101–146.

Cowan, N. (2010), 'The Magical Mystery Four: How Is Working Memory Capacity Limited, and Why?', *Current Directions in Psychological Science* **19**(1), 51–57.

Cuskley, C., Bailes, R. und Wallenberg, J. (2021), 'Noise resistance in communication: Quantifying uniformity and optimality', *Cognition* **214**.

- Cutler, A. (2008), 'The abstract representations in speech processing', *The Quarterly Journal of Experimental Psycholog* **61**(11), 1601–1619.
- Dal, I. und Eroms, H.-W. (2014), *Kurze deutsche Syntax auf historischer Grundlage*, Sammlung kurzer Grammatiken germanischer Dialekte, 4. Aufl. edn, de Gruyter, Berlin, Boston.
- Daumer, G. F. (1895), *Die dreifache Krone Rom's. Versuch einer neuen Beleuchtung und Charakterisirung des römisch-katholischen Priester- und Kirchenthums, namentlich was dessen elementare und principielle Inhaltsbestimmungen und deren vorläufige Begründung und Erscheinung in vorchristlicher Zeit und Welt betrifft*, Aschendorff, Münster.
- Degaetano-Ortlieb, S., Kermes, H., Khamis, A. und Teich, E. (2019), An Information-Theoretic Approach to Modeling Diachronic Change in Scientific English, in C. Suhr, ed., 'From Data to Evidence in English Language', Research, Brill, Leiden, Niederlande.
- Degaetano-Ortlieb, S. und Teich, E. (2018), Using relative entropy for detection and analysis of periods of diachronic linguistic change, in 'In Proceedings of the 2nd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature at COLING2018', Santa Fe, NM, USA. ACL. https://www.uni-saarland.de/fileadmin/upload/lehrstuhl/teich/degaetano/latech2018_final.pdf.
- Demske-Neumann, U. (1990), Charakteristische Strukturen von Satzgefügen in den Zeitungen des 17. Jahrhunderts, in A. Betten, ed., 'Neuere Forschungen zur historischen Syntax des Deutschen', Niemeyer, Tübingen, pp. 239–252.
- Demske, U. (2001), *Merkmale und Relationen: diachrone Studien zur Nominalphrase des Deutschen*, *Studia linguistica Germanica*; 56, repr. edn, de Gruyter, Berlin, Boston.
- Demske, U. (2019), 'Referenzkorpus Frühneuhochdeutsch: Baumbank.UP'. Universität Potsdam: Institut für Germanistik; <https://hdl.handle.net/11022/0000-0007-EAF7-B>.

- Dilthey, W. (1910), *Der Aufbau der geschichtlichen Welt in den Geisteswissenschaften*, Suhrkamp-Taschenbuch Wissenschaft ; 354, 4. Aufl., 1993, erstdruck 1910 edn, Suhrkamp, Frankfurt am Main. Herausgeber: Riedel, Manfred.
- Donaldson, B. C. (1993), *A grammar of Afrikaans*, Mouton grammar library ; 8, Mouton de Gruyter, Berlin, Boston.
- Dora, C. (1994), *Augustinus Egger von St. Gallen: 1833 - 1906; ein Bischof zwischen Kulturkampf, sozialer Frage und Modernismustreit*, St.-Galler Kultur und Geschichte; 23, Staatsarchiv und Stiftsarchiv, St. Gallen.
- Doyle, G. und Frank, M. C. (2015), 'Shared common ground influences information density in microblog texts', *Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL* pp. 1587–1596.
- Drach, E. (1937), *Grundgedanken der deutschen Satzlehre*, Diesterweg, Frankfurt a.M.
- Drozd, L. und Seibicke, W. (1973), *Deutsche Fach- und Wissenschaftssprache. Bestandaufnahme - Theorie - Geschichte*, Brandstetter, Wiesbaden.
- Dürscheid, C. (1989), *Zur Vorfeldbesetzung in deutschen Verbzweit-Strukturen*, Wissenschaftlicher Verlag Trier, Trier.
- Ebert, R. P. (1980), 'SOCIAL AND STYLISTIC VARIATION IN EARLY NEW HIGH GERMAN WORD ORDER: THE SENTENCE FRAME (>SATZRAHMEN<9)', 102(Jahresband), 357–398.
URL: <https://doi.org/10.1515/bgsl.1980.1980.102.357>
- Ebert, R. P. (1986), *Historische Syntax des Deutschen*, Vol. 2: 1300 - 1750 of *Germanistische Lehrbuchsammlung*; 6, Lang, Bern [u.a.].
- Ebert, R. P., Reichmann, O., Solms, H.-J. und Wegera, K.-P. (1993), *Frühneuhochdeutsche Grammatik*, Sammlung kurzer Grammatiken germanischer Dialekte, Niemeyer, Tübingen.

- Eckart, W. U. (2021), *Geschichte, Theorie und Ethik der Medizin*, Springer eBook Collection, 9. edition 2021. edn, Springer Berlin Heidelberg, Berlin, Heidelberg.
- Egger, A. (1895), *Der christliche Vater in der modernen Welt, Erbauungs- und Gebetbuch*, unbekannt, unbekannt.
- Eichinger, L. M. (1995), Syntaktischer Wandel und Verständlichkeit: zur Serialisierung von Sätzen und Nominalgruppen im frühen Neuhochdeutschen, in H. Weinrich und H. L. Kretzenbacher, eds, 'Linguistik der Wissenschaftssprache', Forschungsbericht; 10, de Gruyter, Berlin.
- Eisenberg, P. und Thieroff, R. (2020), *Grundriss der deutschen Grammatik : Der Satz*, Springer eBook Collection, 5th ed. 2020. edn, J.B. Metzler, Stuttgart.
- Elspaß, S. (2005), *Sprachgeschichte von unten: Untersuchungen zum geschriebenen Alltagsdeutsch im 19. Jahrhundert*, Reihe germanistische Linguistik; 263, Niemeyer, Tübingen.
- Engel, U. (1970), Studie zur Geschichte des Satzrahmens und seiner Durchbrechung, in E. Eggers und S. Neumann, eds, 'Studie zur Syntax des heutigen Deutsch: Paul Grebe zum 60. Geburtstag (Sprache der Gegenwart; 6)', Schwann, Düsseldorf.
- Engel, U. (2009), *Syntax der deutschen Gegenwartssprache*, Grundlagen der Germanistik, 4., völlig neu bearb. Aufl. edn, Schmidt, Berlin.
- Erben, J. (2000a), Syntax des Frühneuhochdeutschen, in W. Besch, A. Betten, O. Reichmann und S. Sonderegger, eds, 'Sprachgeschichte: Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung', Vol. 2, De Gruyter, chapter 114, pp. 1584–1594.
- Erben, J. (2000b), Syntax des Neuhochdeutschen bis zur Mitte des 20. Jahrhunderts, in W. Besch, A. Betten, O. Reichmann und S. Sonderegger, eds, 'Sprachgeschichte: Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung', Vol. 2, De Gruyter, Berlin, Boston, chapter 131, pp. 1855–1903.

- Eroms, H.-W. (1991), Valenzgebundene präpositionalkonstruktionen im deutschen, in R. Harweg, S. Kishitani und M. Scherner, eds, 'Die deutsche Sprache. Gestalt und Leistung. Hennig Brinkmann in der Diskussion. Zu seinem neunzigsten Geburtstag.', Nodus Publikationen, Münster, pp. 39–54.
- Fischer, A. (1987), Das Genitivobjekt und damit konkurrierende Objekte nach Verben in Leipziger Frühdrucken, in J. Schildt, ed., 'Zum Sprachwandel in der deutschen Literatursprache des 16. Jhs', Akademie-Verlag, Berlin, pp. 267–324.
- Fisseni, B. (2017), 'Das Bonner Frühneuhochdeutschkorpus (FnhdC)'. <http://www.korpora.org/FnhdC/>.
- Fleischer, J. (2005), 'Relativsätze in den Dialekten des Deutschen: Vergleich und Typologie', *Linguistik Online* 24(3), 171–186.
- Fortmann, C. und Frey, W. (1997), Konzeptuelle Struktur und Grundabfolge der Argumente, in F.-J. d'Avis und U. Lutz, eds, 'Zur Satzstruktur im Deutschen', Vol. 90 of *Arbeitspapiere des SFB 340*, Universität Stuttgart, Tübingen, Stuttgart, Tübingen, pp. 143–170.
- Francis, E. J. und Michaelis, L. A. (2017), 'When relative clause extraposition is the right choice, it's easier', *Language and Cognition* 9(2), 332–370.
- Francis, E. und Michaelis, L. (2014), *Why move? How weight and discourse factors combine to predict relative clause extraposition in English*, Oxford University Press, pp. 70–87.
- Frank, A. F. und Jaeger, F. (2008), 'Speaking Rationally: Uniform Information Density as an Optimal Strategy for Language Production', *Proceedings of the 30th Annual Meeting of the Cognitive Science Society* pp. 939–944.
- Frank, S. L., Trompenaars, T. und Vasishth, S. (2016), 'Cross-Linguistic Differences in Processing Double-Embedded Relative Clauses: Working-Memory Constraints or Language Statistics?', *Cognitive Science* 40(3), 554–578.
- Frey, W. (2004), 'A medial topic position for German', *Linguistische Berichte* 198, 153–190.

- Frey, W. und Pittner, K. (1998), 'Zur Positionierung der Adverbiale im deutschen Mittelfeld', *Linguistische Berichte* **176**, 489–534.
- Fries, N. (1991), Präpositionen und Präpositionalphrasen im Deutschen, im Altgriechischen und im Neugriechischen, in S. Olsen und G. Fanselow, eds, 'DET, COMP und INFL. Zur Syntax funktionaler Kategorien und grammatischer Funktionen', Max Niemeyer Verlag, Berlin, New York, pp. 73–82.
URL: <https://doi.org/10.1515/9783111353838.73>
- Fritsch, W. J. (1990), *Gestalt und Bedeutung der deutschen Relativsätze*, Uni-Druck, München.
- Futrell, R., Gibson, E. und Levy, R. (2021), 'Lossy-Context Surprisal: An Information-Theoretic Model of Memory Effects in Sentence Processing', *Cognitive Science* **44**(3), 1–54.
- Gall, F. J. (1791), *Philosophisch-Medicinische Untersuchungen über Natur und Kunst im kranken und gesunden Zustande des Menschen*, Gräffer, Wien.
- Gallmann, P. (2016), Der Satz, in A. Wöllstein, K. Kunkel-Razum, F. Münzberg und S. Ripp, eds, 'Der Duden in zwölf Bänden: das Standardwerk zur deutschen Sprache', 9., vollständig überarbeitete und aktualisierte auflage edn, Vol. Band 4: Duden - Die Grammatik, Dudenverlag, Berlin.
- Genzel, D. und Charniak, E. (2002), Entropy Rate Constancy in Text, in 'Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics', Association for Computational Linguistics, Philadelphia, Pennsylvania, USA, pp. 199–206.
URL: <https://aclanthology.org/P02-1026>
- Genzel, D. und Charniak, E. (2003), Variation of Entropy and Parse Trees of Sentences as a Function of the Sentence Number, in 'Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing', pp. 65–72.
URL: <https://aclanthology.org/W03-1009>

- Gibson, E. (1998), 'Linguistic complexity: Locality of syntactic dependencies', *COGNITION* **68**, 1–76.
- Gibson, E., Futrell, R., Piantadosi, S., Dautriche, I., Mahowald, K., Bergen, L. und Levy, R. (2019), 'How Efficiency Shapes Human Language', *Trends in Cognitive Sciences* **23**(5), 389–407.
- Glück, H. (2000), *Metzler-Lexikon Sprache*, 2., überarb. und erw. Aufl. edn, Metzler, Stuttgart.
- Glück, H. und Rödel, M. (2016), *Metzler Lexikon Sprache*, J.B. Metzler Verlag, Stuttgart.
- Gries, S. T. (2021), *Statistics for Linguistics with R. A Practical Introduction*, De Gruyter Mouton, Berlin, Boston.
URL: <https://doi.org/10.1515/9783110718256>
- Griesbach, H. (1986), *Neue deutsche Grammatik*, 1. dr. edn, Langenscheidt, Berlin [u.a.].
- Gulordava, K. und Merlo, P. (2015), Diachronic Trends in Word Order Freedom and Dependency Length in Dependency-Annotated Corpora of Latin and Ancient Greek, in 'Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)', Uppsala University, Uppsala, Sweden, Uppsala, Sweden, pp. 121–130.
URL: <https://aclanthology.org/W15-2115>
- Gundel, J., Hedberg, N. und Zacharski, R. (1993), 'Cognitive Status and the form of Referring Expressions in Discourse', *Language* **69**(2), 274–307.
- Guthrie, D., Allison, B., Liu, W., Guthrie, L. und Wilks, Y. (2006), 'A Closer Look at Skip-gram Modelling', *Proceedings of the Fifth International Conference on Language Resources and Evaluation*. pp. 1222–1225.
- Hahn, M., Degen, J. und Futrell, R. (2021), 'Modeling word and morpheme order in natural language as an efficient trade-off of memory and surprisal', *Psychol Rev.* **128**(4), 726–756.

- Haider, H. (1988), 'Die Struktur der deutschen Nominalphrase', *Zeitschrift für Sprachwissenschaft* 7(1), 32–59.
- Haider, H. (1992), Branching and Discharge, in 'Arbeitspapiere des SFB 340', Vol. 23, Universität Stuttgart, Tübingen, Stuttgart, Tübingen.
- Hale, J. (2001), 'A probabilistic Early parser as a psycholinguistic model', *Proceedings of the second meeting of the North American chapter of the Association for Computational Linguistics (NAACL'01)* pp. 159–166.
- Hanssen, P. (1731), *Achtzig erläuterte Grund-Fragen. in welchen Die Lehre der Evangelisch-Lutherischen Kirche von dem Mittler-Ampt Jesu Christi und dessen Einfluß zu einem heiligen Leben wider ihre Feinde überhaupt, insonderheit aber wider den beruffenen Johann Conrad Dippel, der sich zum Beweis seiner wunderlichen Gemüths-Neigung bishero Christianum Democritum genannt, solchergestalt befestiget, daß die dahin gehörige Warheiten nach Vernunft und Schrift auseinander hergeleitet, folglich in einer nohtwendigen Verbindung und unüberwindlichen Gewißheit dargestellt werden: Denen, die mit dergleichen schädlichen Lehren zum Argen versucht worden, zur Warnung; schwachen Seelen zur Befestigung und neu-angehenden Predigern, um denen heutigen Zweiffelmachern desto gründlicher begegnen zu können, zu mehrerm Nachsinnen mitgetheilet, s.e., Lübeck, Hamburg.*
- Hasak, M. (1893), *Die Predigtkirche im Mittelalter*, Wilhelm Ernst & Sohn, Berlin.
- Hawkins, J. A. (1992), 'Syntactic Weight Versus Information Structure in Word Order Variation', *Informationsstruktur und Grammatik. Linguistische Berichte Sonderhefte* 4, 196–219.
- Hawkins, J. A. (2009), An efficiency theory of complexity and related phenomena, in 'Language Complexity as an Evolving Variable', Geoffrey Sampson, David Gil, Peter Trudgill, pp. 252–268.
- Hegendüs, I. (2012), Wie kann nächstes diachron untersucht werden?, in V. Ágel und M. Henning, eds, 'Zugänge zur Grammatik der gesprochenen Sprache', Vol. 269 of *Reihe Germanistische Linguistik*, Niemeyer, Tübingen.

- Helbig, G. und Buscha, J. (2005), *Deutsche Grammatik: Ein Handbuch für den Ausländerunterricht*, neubearb., nachdr. edn, Langenscheidt, Berlin.
- Henderson, J. (2004), Lookahead in Deterministic Left-Corner Parsing, in 'Proceedings of the Workshop on Incremental Parsing: Bringing Engineering and Cognition Together', IncrementParsing '04, Association for Computational Linguistics, USA, pp. 26—33.
- Hentschel, E. und Weydt, H. (2013), *Handbuch der deutschen Grammatik*, De Gruyter-Studium, De Gruyter, Berlin, Boston.
- Hinterhölzl, R. (2010), Zur Herausbildung der Satzklammer im Deutschen, in 'Historische Textgrammatik und Historische Syntax des Deutschen', De Gruyter, Berlin, Boston, pp. 121–138.
- Hoberg, U. (1981), *Die Wortstellung in der geschriebenen deutschen Gegenwartssprache*, Heutiges Deutsch I/10, Hueber, München.
- Hoffmann, L. (2019), *Deutsche Syntax: Ansichten und Aussichten*, Jahrbuch; 1991, Jahrbuch des Instituts für Deutsche Sprache; 1991, de Gruyter, Berlin, Boston.
- URL:** <https://www.degruyter.com/doc/cover/9783110622447.jpg>; <https://www.degruyter.com/cover/covers/9783110622447.jpg>
- Horton, W. und Gerrig, R. (2002), 'Speakers' experiences and audience design: knowing when and knowing how to adjust utterances to addressees', *Journal of Memory and Language* 47(2), 589–606.
- Horton, W. und Gerrig, R. J. (2005), 'The impact of memory demands on audience design during language production', *Cognition* 96(2), 127–142.
- Höhle, T. N. (1982), Beiträge zur deutschen Grammatik: Gesammelte Schriften von Tilman N. Hohle, in M. Reis, F. Richter, S. Müller und T. N. Höhle, eds, 'Classics in Linguistics', zweite durchgesehene auflage, 2019 edn, Classics in linguistics; 5, Language Science Press, Berlin.
- Höllein, D. (2019), *Präpositionalobjekt vs. Adverbial: Die semantischen Rollen der Präpositionalobjekte*, De Gruyter, Berlin, Boston.

- Höskuldur, T. (2007), *The Syntax of Icelandic*, Cambridge Syntax Guides, Cambridge University Press, Cambridge.
- Imo, W. (2015), Nachträge im Spannungsfeld von Medialität, Situation und interaktionaler Funktion, in H. Vinckel-Roisin, ed., 'Das Nachfeld im Deutschen', deGruyter, Berlin.
- Inaba, J. (2007), *Die Syntax der Satzkomplementierung. Zur Struktur des Nachfeldes im Deutschen*, Studia Grammatica, Akademie Verlag, Berlin.
- Jacobs, J. (1988), 'Probleme der freien Wortstellung im Deutschen', *Sprache und Pragmatik* 5, 8–37.
- Jaeger, F. (2005), Optional that indicates production difficulty: Evidence from disfluencies, in 'Proceedings of DiSS'05, Disfluency in Spontaneous Speech Workshop Proceedings of DiSS'05, Disfluency in Spontaneous Speech Workshop', pp. 1–7.
- Jaeger, F. (2010), 'Redundancy and reduction: Speakers manage syntactic information density.', *Cognitive Psychology* 61(1), 23–62.
- Jaeger, F. (2011), 'Corpus-based Research on Language Production: Information Density and Reducible Subject Relatives', *Language from A Cognitive Perspective: Grammar, Usage, and Processing - Studies in Honor of Tom Wasow* pp. 161–197.
- Johnson, K. (1988), 'Clausal Gerunds, the ECP, and Gouvernement', *Linguistic Inquiry* 19, 583–610.
- Jurafsky, D. (2003), Probabilistic modeling in psycholinguistics: Linguistic comprehension and production, in R. Bod, J. Hay und S. Jannedy, eds, 'Probabilistic linguistics', MIT Press, Cambridge, MA, pp. 1–50.
- Kathol, A. und Pollard, C. (1995), 'Extrapolation via Complex Domain Formation', *ANNUAL MEETING- ASSOCIATION FOR COMPUTATIONAL LINGUISTICS* 33, 174–180.

- Keller, F. (2004), 'The entropy rate principle as a predictor of processing effort: An evaluation against eye-tracking data', *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 317–324.
- Klein, T. und Dipper, S. (2016), *Handbuch zum Referenzkorpus Mittelhochdeutsch*, Bochumer Linguistische Arbeitsberichte, 19, Bochum.
URL: <https://linguistics.rub.de/forschung/arbeitsberichte/19.pdf>
- Klein, W. P. (2011), Deutsch statt Latein! Zur Entwicklung der Wissenschaftssprache in der frühen Neuzeit, in W. Eins, H. Glück und S. Pretscher, eds, 'Wissen schaffen - Wissen kommunizieren', Harrassowitz-Verlag, Wiesbaden, pp. 35–47.
- Kneser, R. und Ney, H. (1995), 'Improved backing-off for M-gram language modeling', *1995 International Conference on Acoustics, Speech, and Signal Processing* 1, 181–184.
- Koch, P. und Oesterreicher, W. (2007), 'Schriftlichkeit und kommunikative Distanz Schriftlichkeit und kommunikative Distanz', *Zeitschrift für germanistische Linguistik* 35(3), 346–375.
- Koch, R. (1878), *Untersuchung über die Aetiologie der Wundinfectionskrankheiten*, Vogel, Leipzig.
- Konopka, M. (1996), *Strittige Erscheinungen der deutschen Syntax im 18. Jahrhundert*, Reihe germanistische Linguistik; 173, Niemeyer, Tübingen.
- Kraepelin, E. (1892), *Ueber die Beeinflussung einfacher psychischer Vorgänge durch einige Arzneimittel. Experimentelle Untersuchungen*, Fischer, Jena.
- Kretzenbacher, H. L. (1995), Wie durchsichtig ist die Sprache der Wissenschaft, in H. L. Kretzenbacher und H. Weinrich, eds, 'Linguistik der Wissenschaftssprache', de Gruyter, Berlin, New York, pp. 15–39.
- Krifka, M. (2008), 'Basic notions of information structure', *Acta Linguistica Hungarica* 55(3-4), 243–276.

- Kusmirek, A., Greenberg, C., Oualil, Y. und Klakow, D. (2023), 'Lm-gui: A graphical user interface for n-gram language models', <https://github.com/uds-lsv/languagemodels>.
- Lehmann, C. (1984), *Der Relativsatz: Typologie seiner Strukturen: Theorie seiner Funktionen: Kompendium seiner Grammatik*, Language Universals Series; Volume 3, Gunter Narr Verlag, Tübingen.
- Lemnitzer, L. und Zinsmeister, H. (2015), *Korpuslinguistik: eine Einführung*, Narr Studienbücher, 3., überarbeitete und erweiterte auflage edn, Narr Francke Attempto, Tübingen.
- Lenerz, J. (1977), *Zur Abfolge nominaler Satzglieder im Deutschen*, Narr, Tübingen.
- Levy, R. (2008), 'Expectation-based syntactic comprehension', *Cognition* **106**(3), 1126–1177.
- Levy, R., Fedorenko, E., Breen, M. und Gibson, E. (2012), 'The processing of extraposed structures in English', *Cognition* **122**(1), 12–36.
URL: <https://www.sciencedirect.com/science/article/pii/S0010027711002010>
- Levy, R. und Jaeger, T. F. (2007), Speakers optimize information density through syntactic reduction, in B. Schölkopf, J. Platt und T. Hofmann, eds, 'Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference', The MIT Press, pp. 849–856.
- Light, C. (2011), The information structure of subject extraposition in Early New High German, in S. Müller, ed., 'Proceedings of the HPSG 2011 Conference', Department of Linguistics, CSLI Publications, University of Washington, pp. 315 – 326.
- Ludwig, K. F. W. (1856), *Lehrbuch der Physiologie des Menschen*, Winter, Heidelberg, Leipzig.
- Lutz, L. (1981), *Zum Thema "Thema": Einführung in die Thema-Rhema-Theorie*, Hamburger Arbeiten zur Linguistik und Texttheorie: 1, Hamburger Buchagentur, Hamburg.

- Löhe, J. K. W. (1847), *Prediget das Evangelium aller Creatur! Predigt am Missionsfeste zu Nürnberg den 17. Juni 1847*, Joh. Phil. Raw'sche Buchhandlung, Nürnberg.
- Lötscher, A. (1972), Some Problems Concerning Standard German Relative Clauses, in P. M. Peranteau, J. N. Levi und G. C. Phares, eds, 'Papers from the Relative Clause Festival', Chicago Linguistic Society, Chicago, pp. 47–58.
- Manning, C. und Schütze, H. (2001), *Foundations of Statistical Natural Language Processing*, MIT Press, Cambridge, MA.
- Marillier, J.-F. (1993), Integration vs. Extraposition von Relativsätzen, in J.-F. Marillier, ed., 'Satzanfang – Satzende. Syntaktische, semantische und pragmatische Untersuchungen zur Satzabgrenzung und Extraposition im Deutschen', Narr, Tübingen, pp. 219–232.
- Mattheier, K. J. (1981), 'Wege und Umwege zur neuhochdeutschen Schriftsprache', *Zeitschrift für germanistische Linguistik* 9(3), 274–307.
- McClelland, J. L. und Elman, J. L. (1986), 'The TRACE model of speech perception', *Cognitive Psychology* 18(1), 1–86.
- Meibauer, J., Steinbach, M. und Altmann, H., eds (2013), *Satztypen des Deutschen*, De Gruyter, Berlin, Boston.
URL: <https://doi.org/10.1515/9783110224832>
- Miller, G. A. (1956), 'The magical number seven, plus or minus two: Some limits on our capacity for processing information', *Psychological Review* 63(2), 81—97.
- Modrián-Horváth, B. (2019), Analytische Verbform, in S. J. Schierholz, ed., 'Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK) Online', De Gruyter, Berlin, Boston.
URL: [https://www.degruyter.com/database/WSK/entry/wsk_id_wsk_aartikel_aartikel20823/html](https://www.degruyter.com/database/WSK/entry/wsk_id_wsk_artikel_aartikel20823/html)
- Morlicchio, E. (1991), 'Der Satzrahmen: Die Länge des Satzes und die Mitteilungsperspektive', *Università degli Studi di Napoli L'Orientale: CINECA IRIS*
- .

Musan, R. (2002), 'Informationsstrukturelle Dimensionen im Deutschen. Zur Variation der Wortstellung im Mittelfeld', *Zeitschrift für germanistische Linguistik* 30, 198–221.

Müller, S. (2011), Yet another paper about partial verb phrase fronting in German, in 'Research Report', Research report; 97,7, Saarländische Universitäts- und Landesbibliothek, Saarbrücken. In: Kaiserslautern; Saarbrücken: DFKI, 1997.

Mürmann, M. (2014), *Wahrscheinlichkeitstheorie und stochastische Prozesse*, Springer Spektrum, Berlin.

Nationalbibliothek, D. (2022a), 'Abel, Heinrich Kaspar. Katalog der Deutschen Nationalbibliothek'.

URL: <https://d-nb.info/gnd/124786979>

Nationalbibliothek, D. (2022b), 'Bengel, Johann Albert. Katalog der Deutschen Nationalbibliothek'.

URL: <https://d-nb.info/gnd/118655477>

Nationalbibliothek, D. (2022c), 'Benner, Johannes Hermann. Katalog der Deutschen Nationalbibliothek'.

URL: <https://d-nb.info/gnd/116121254>

Nationalbibliothek, D. (2022d), 'Bräuner, Johann Jacob. Katalog der Deutschen Nationalbibliothek'.

URL: <https://d-nb.info/gnd/124536557>

Nationalbibliothek, D. (2022e), 'Carus, Carl Gustav. Katalog der Deutschen Nationalbibliothek'.

URL: <https://d-nb.info/gnd/118519352>

Nationalbibliothek, D. (2022f), 'Daumer, Georg Friedrich. Katalog der Deutschen Nationalbibliothek'.

URL: <https://d-nb.info/gnd/118678957>

Nationalbibliothek, D. (2022g), 'Gall, Franz Joseph. Katalog der Deutschen Nationalbibliothek'.

URL: <https://d-nb.info/gnd/118537245>

Nationalbibliothek, D. (2022h), 'Hasak, Max. Katalog der Deutschen Nationalbibliothek'.

URL: <https://d-nb.info/gnd/117503002>

Nationalbibliothek, D. (2022i), 'Katalog der Deutschen Nationalbibliothek'.

URL: <https://portal.dnb.de/opac.htm>

Nationalbibliothek, D. (2022j), 'Koch, Robert. Katalog der Deutschen Nationalbibliothek'.

URL: <https://d-nb.info/gnd/118564064>

Nationalbibliothek, D. (2022k), 'Kraepelin, Emil. Katalog der Deutschen Nationalbibliothek'.

URL: <https://d-nb.info/gnd/118565915>

Nationalbibliothek, D. (2022l), 'Ludwig, Karl Friedrich Wilhelm. Katalog der Deutschen Nationalbibliothek'.

URL: <https://d-nb.info/gnd/118729470>

Nationalbibliothek, D. (2022m), 'Löhe, Johann Konrad Wilhelm. Katalog der Deutschen Nationalbibliothek'.

URL: <https://d-nb.info/gnd/118728784>

Nationalbibliothek, D. (2022n), 'Niviandts, Friedrich. Katalog der Deutschen Nationalbibliothek'.

URL: <https://d-nb.info/gnd/140752897>

Nationalbibliothek, D. (2022o), 'Pahl, Johann Gottfried von. Katalog der Deutschen Nationalbibliothek'.

URL: <https://d-nb.info/gnd/10038384X>

Nationalbibliothek, D. (2022p), 'Purmann, Gottfried. Katalog der Deutschen Nationalbibliothek'.

URL: <https://d-nb.info/gnd/124546420>

Nationalbibliothek, D. (2022q), 'Reil, Johannes Christian. Katalog der Deutschen Nationalbibliothek'.

URL: <https://d-nb.info/gnd/118599224>

Nationalbibliothek, D. (2022r), 'Rotth, Albrecht Christian. Katalog der Deutschen Nationalbibliothek'.

URL: <https://d-nb.info/gnd/115378510>

Nationalbibliothek, D. (2022s), 'Spener, Philipp Jacob. Katalog der Deutschen Nationalbibliothek'.

URL: <https://d-nb.info/gnd/118616099>

Nationalbibliothek, D. (2022t), 'Strauß, David. Katalog der Deutschen Nationalbibliothek'.

URL: <https://d-nb.info/gnd/118619055>

Nationalbibliothek, D. (2022u), 'Unzer, Johann August . Katalog der Deutschen Nationalbibliothek'.

URL: <https://d-nb.info/gnd/118963600>

Niviandts, F. (1708), *Güldenenes Schwert. Womit Wider jetziger Zeit ein einreissende Ketzereyen bevorab wider den Lutheranismum und Calvinismum*, Selbstverlag, Köln.

Nübling, D. (2016), Die nicht flektierbaren Wortarten, in A. Wöllstein, K. Kunkel-Razum, F. Münzberg und S. Ripp, eds, 'Der Duden in zwölf Bänden: das Standardwerk zur deutschen Sprache', 9., vollständig überarbeitete und aktualisierte auflage edn, Vol. Band 4: Duden - Die Grammatik, Dudenverlag, Berlin.

Nübling, D., Dammel, A., Duke, J. und Szczepaniak, R. (2013), *Historische Sprachwissenschaft des Deutschen Eine Einführung in die Prinzipien des Sprachwandels*, Narr, Tübingen.

URL: <https://elibrary.narr.digital/book/99.125005/9783823378235>

O'Grady, G. (2016), Given/New: What do the terms refer to?, in B. Clark und J. A. Hita, eds, 'The dynamicity of communication below, around and above

the clause', Vol. 9 (1) of *English Text Construction*, John Benjamins Publishing Company, pp. 9–32.

Ortmann, K. (2020), Automatic topological field identification in (historical) german texts, in 'Proceedings of the The 4th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL).', Barcelona, Spain (online), pp. 10—18.

Ortmann, K. (2023), Computational Methods for Investigating Syntactic Change: Automatic Identification of Extraposition in Modern and Historical German, PhD thesis, Ruhr-Universität Bochum, Bochum.

Ortmann, K. und Dipper, S. (2019), Variation between Different Discourse Types: Literate vs. Oral, in 'Proceedings of the Sixth Workshop on NLP for Similar Languages, Varieties and Dialects', Association for Computational Linguistics, Ann Arbor, Michigan, pp. 64–79.

URL: <https://aclanthology.org/W19-1407>

Ortmann, K. und Dipper, S. (2020), Automatic Orality Identification in Historical Texts, in 'Proceedings of the 12th Language Resources and Evaluation Conference', European Language Resources Association, Marseille, France, pp. 1293–1302. <https://aclanthology.org/2020.lrec-1.162>.

Ortmann, K. und Dipper, S. (2022), 'Coast (conceptual orality analysis and scoring tool)', <https://github.com/rubcompling/COAST>.

Ortmann, K. und Dipper, S. (2024), Nāhetexte automatisch erkennen: Entwicklung eines linguistischen Scores für konzeptionelle Mündlichkeit in historischen Texten, in W. Imo und J. Wesche, eds, 'Sprechen und Gespräche in historischer Perspektive. Sprach- und literaturwissenschaftliche Zugänge', Studien zu Literaturwissenschaft und Linguistik, J.B. Metzler, Berlin, pp. 17–36.

Ostermann, S. (2021), Script knowledge for natural language understanding, PhD thesis, Saarland University, Saarbrücken.

- Oubouzar, E. (1990), Zur Ausbildung des bestimmten Artikels im Althochdeutschen, in Y. Desportes, ed., 'Althochdeutsch, Syntax und Semantik', Centre d'Études Linguistiques Jacques Goudet, Lyon, pp. 71–87.
- Pafel, J. (2009), 'Zur linearen Syntax des deutschen Satzes', *Linguistische Berichte* 2009(217), 37–79.
- Pahl, J. G. v. (1799), *Leben und Thaten des ehrwürdigen Paters Simpertus*, Claß in Heilbronn, Heilbronn.
- Panhuis, D. G. J. (2011), *The communicative perspective in the sentence: a study of Latin word order*, Studies in language companion series (SLCS), John Benjamins Publishing Company, Amsterdam.
- Paranhos Zitterbart, J. (2002), *Zur korrelativen Subordination im Deutschen*, Max Niemeyer Verlag, Berlin, New York.
URL: <https://doi.org/10.1515/9783110911657>
- Pate, J. K. und Goldwater, S. (2015), 'Talkers account for listener and channel characteristics to communicate efficiently', *Journal of Memory and Language* 78, 1–17.
- Paul, H. (2007), *Mittelhochdeutsche Grammatik*, Sammlung kurzer Grammatiken germanischer Dialekte, 25. Aufl. edn, Niemeyer, Tübingen. Herausgeber: Thomas Klein, Hans-Joachim Solms, Klaus-Peter Wegera, Ingeborg Schrobler, Ingeborg und Heinz-Peter Prell.
- Petkova-Kessanlis, M. (2015), Nachfeldbesetzung und ihre kommunikative Funktion in wissenschaftlichen Texten, in H. Vinckel-Roison, ed., 'Das Nachfeld im Deutschen. Theorie und Empirie', de Gruyter, Berlin, Boston, pp. 211–288.
- Pittner, K. und Berman, J. (2021), *Deutsche Syntax: ein Arbeitsbuch*, Narr Studienbücher, 7., überarbeitete und erweiterte Auflage edn, Narr Francke Attempto, Tübingen.
URL: <http://meta.narr.de/u1/9783823394112.jpg>

- Polenz, P. v. (2020), *Geschichte der deutschen Sprache*, De Gruyter Studium, 11., überarbeitete auflage von norbert richard wolf edn, De Gruyter, Berlin, Boston.
- Polenz, P. v. und Moulin, C. (2021), *Deutsche Sprachgeschichte vom Spätmittelalter bis zur Gegenwart. Einführung, Grundbegriffe. 14. bis 16. Jahrhundert*, Vol. 1, De Gruyter, Berlin, Boston.
- Polenz, P. v., Moulin, C. und Harion, D. (2013), *Deutsche Sprachgeschichte vom Spätmittelalter bis zur Gegenwart; Band II, 17. und 18. Jahrhundert*, De Gruyter Studienbuch, 2. ed. edn, De Gruyter, Berlin.
- Poschmann, C. und Wagner, M. (2016), 'Relative clause extraposition and prosody in German', *Natural Language and Linguistic Theory* **34**(3), 1021–1066.
- Presch, G. (1977), *Syntaktische Diskontinuität: Linearität als grammatisches Prinzip und als Problem sprachlicher Rezeption*, Germanistische Texte und Studien; 3, Olms, Hildesheim [u.a.].
- Prince, E. (1981), Toward a Taxonomy of Given – New Information, in P. Cole, ed., 'Radical pragmatics', Academic Pr., New York, pp. 223–256.
- Purmann, G. (1680), *Der rechte und wahrhafftige Feldscher Oder/ Die rechte und warhafftige Feldschers-Kunst*, Hynitzsch, Halberstadt.
- 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/>
- R Core Team (2022), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: <https://www.R-project.org/>
- Ramers, K. H. (2006), 'Topologische Felder: Nominalphrase und Satz im Deutschen', *Zeitschrift für Sprachwissenschaft* **25**(1), 95–128.
URL: <https://doi.org/10.1515/zfs.2006.003>

- Rauth, P. (2020), *Ditransitive Konstruktionen im Deutschen. Geschichte und Steuerung der Objektfolge im Mittelfeld*, Stauffenburg, Tübingen.
- Rehm, G. (2002), Schriftliche Mündlichkeit in der Sprache des World Wide Web, in A. Ziegler und C. Dürscheid, eds, 'Kommunikationsform E-Mail', Stauffenburg, Tübingen, p. 263–308.
- Reil, J. C. (1803), *Rhapsodien über die Anwendung der psychischen Curmethode auf Geisteszerrüttung*, Curt, Halle (Saale).
- Reis, M. (1997), Zum syntaktischen Status unselbständiger Verbzweit-Sätze, in C. Dürscheid, ed., 'Sprache im Fokus', Niemeyer, Tübingen, pp. 121–144.
- Repp, S. und Struckmeier, V. (2020), *Syntax: eine Einführung*, Einführungen in die Sprachwissenschaft Springer eBook Collection, J.B. Metzler, Berlin.
- Reventlow, H. (1997), *Epochen der Bibelauslegung*, Beck, München.
- Richter, G. (1985), 'Einige Anmerkungen zur Norm und Struktur des gesprochenen Deutsch', *Deutsch als Fremdsprache* 22(3), 149–153.
- Roark, B. (2001), 'Probabilistic top-down parsing and language modeling', *Computational Linguistics* 27(2), 249–276.
- Roth, A. C. (1692), *Eylfertiges Bedencken über M. August Hermann Franckens/Pastoris zu Glauche vor Halle/Seine Schutz-Predigt/ Ob Er Durch dieselbe seinen Zweck/ den er auf dem Titul gedachter Predigt/ berühret hat/ erlanget oder nicht?. Auf Begehren gestellet von einem Diener GOTTes in Halle/ An einen Seiner Beicht-Kinder, unbekannt, Halle.*
- Sabel, J. (2000), 'Das Verbstellungsproblem im Deutschen: Synchronie und Diachronie', *Deutsche Sprache* 28, 74–99.
- Sabel, J. und Mamoru, S. (2005), *The Free Word Order Phenomenon: Its Syntactic Sources and Diversity*, number Vol. 69 in 'Studies in Generative Grammar', De Gruyter Mouton, Berlin, Boston.
- URL: <https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=210956&lang=desite=ehost-live>

- Sahel, S. (2015), Zur Ausklammerung von Relativsätzen und Vergleichspräsen im frühen Neuhochdeutschen, in H. Vinckel-Roisin, ed., 'Das Nachfeld im Deutschen', deGruyter, Berlin.
- Sapp, C. D. (2014), 'Extraposition in Middle and New High German', *The Journal of Comparative Germanic Linguistics* 17(2), 129–156.
- Schallert, O. (2014), *Zur Syntax der Ersatzinfinitivkonstruktion: Typologie und Variation*, Studien zur deutschen Grammatik; Bd. 87, Stauffenburg-Verlag, Tübingen.
- Schierholz, S. J. und Uzonyi, P., eds (2022), *Band 2 Syntax*, De Gruyter, Berlin, Boston.
URL: <https://doi.org/10.1515/9783110698527>
- Schildt, J. (1976), Zur Ausbildung der Satzklammer, in G. Kettmann und J. Schildt, eds, 'Zur Ausbildung der Norm in der deutschen Literatursprache auf der syntaktischen Ebene', Akademie-Verlag.
- Schiller, A., Teufel, S., Stöckert, C. und Thielen, C. (1999), 'Guidelines für das Tagging deutscher Textcorpora mit STTS (Kleines und großes Tagset)'.
<http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf>.
- Schlipphak, K. (2008), *Erwerbsprinzipien der deutschen Nominalphrase: Erwerbsreihenfolge und Schemata; die Interaktion sprachlicher Aufgabenbereiche*, Perspektiven germanistischer Linguistik; 1, Ibidem-Verlag, Stuttgart.
- Schwitalla, J. (2006), *Gesprochenes Deutsch: eine Einführung*, Grundlagen der Germanistik; 33, 3., neu bearb. Aufl. edn, E. Schmidt, Berlin.
- Schwitalla, J. und Tiittula, L. (2009), *Mündlichkeit in literarischen Erzählungen: Sprach- und Dialoggestaltung in modernen deutschen und finnischen Romanen und deren Übersetzungen*, Stauffenburg Linguistik; 48, Stauffenburg-Verlag, Tübingen.
- Selting, M. (2012), 'Grammatik des gesprochenen Deutsch' im Rahmen der Interaktionalen Linguistik, in V. Ágel und M. Hennig, eds, 'Zugänge zur

- Grammatik der gesprochenen Sprache', Vol. 269 of *Reihe Germanistische Linguistik*, Niemeyer, Tübingen.
- Shannon, C. E. (1948), 'A Mathematical Theory of Communication', *The Bell System Technical Journal* 27(3), 379–423.
- Shannon, T. (1992), Towards an adequate characterization of relative clause extraction in modern German, in I. Rauch, ed., 'On Germanic linguistics. Issues and methods', Vol. 68 of *Trends in linguistics: Studies and monographs*, de Gruyter, Berlin, Boston, pp. 253–282.
- Sikos, L., Greenberg, C., Drenhaus, H. und Crocker, M. (2017), 'Information density of encodings: The role of syntactic variation in comprehension', *Proceedings of the 39th Annual Conference of the Cognitive Science Society* .
- Sonderegger, S. (1979), *Grundzüge deutscher Sprachgeschichte: Diachronie des Sprachsystems*, Vol. 1: Einführung, Genealogie, Konstanten, de Gruyter, Berlin, Boston.
- Sparn, W. (1985), Vernünftiges Christentum. über die geschichtliche Aufgabe der theologischen Aufklärung im 18. Jahrhundert in Deutschland, in R. Vierhaus, ed., 'Wissenschaften im Zeitalter der Aufklärung: Aus Anlaß des 250jährigen Bestehens des Verlages Vandenhoeck & Ruprecht', Vandenhoeck & Ruprecht, Göttingen, pp. 18–57.
- Spener, P. J. (1676), *Pia Desideria. Oder Hertzliches Verlangen/ Nach Gottgefälliger Besserung der wahren Evangelischen Kirchen/ sampt einigen dahin einfältig abzweckenden Christlichen Vorschlägen Philipp Jacob Speners. Sampt angehengten Zweyer Christlichen Theologorum darüber gestellten/ und zu mehrer auffbauung höchst-dienlichen Bedencken*, Zunner; Fritgen, Frankfurt am Main.
- Speyer, A. (2008a), 'Doppelte Vorfelddbesetzung im heutigen Deutschen und im Frühneuhochdeutschen', *Linguistische Berichte* 216, 455–485.
- Speyer, A. (2008b), German vorfeld-filling as constraint interaction, in A. Benz und P. Kühnlein, eds, 'Constraints in Discourse', *Pragmatics & Beyond New*

Series 172, John Benjamins Publishing Company, Amsterdam / Philadelphia, pp. 267—290.

Speyer, A. (2009), 'Das Vorfeldranking und das Vorfeld-es', *Linguistische Berichte Heft* **219**.

Speyer, A. (2010), Filling the Vorfeld in spoken and written discourse, in S.-K. Tanskanen, M.-L. Helasvuo, M. Johansson und M. Raitaniemi, eds, 'Discourses in Interaction', *Pragmatics & Beyond New Series* 203, John Benjamins Publishing Company, Amsterdam / Philadelphia, pp. 263–290.

Speyer, A. (2011), 'Die Freiheit der Mittelfeldabfolge im Deutschen. Ein modernes Phänomen', *Beiträge zur Geschichte der deutschen Sprache und Literatur* **133**(1), 14–31.

Speyer, A. (2015a), 'Auch früher wollte man informieren – Zum Einfluss der Informationsstruktur auf die Syntax in der Geschichte des Deutschen', *Zeitschrift für germanistische Linguistik* **43**(3), 485–515.

Speyer, A. (2015b), 'Informationsdichte als Faktor für Einbettung. Eine sprachhistorische Studie an Kausalsätzen', *Sprachwissenschaft* **40**, 453–485.

Speyer, A. und Lemke, R. (2017), 'Information Density as a Factor for Variation in the Embedding of Relative Clauses', *ArXiv: arXiv:1705.06457* pp. 1–10. Letzter Zugriff: 02.03.2023.

Speyer, A. und Voigtmann, S. (2021a), Factors for the integration of causal clauses in the history of German, in L. Jedrzejowski und C. Fleczonek, eds, 'On the Variation of Causal Clauses', Buske, Hamburg.

Speyer, A. und Voigtmann, S. (2021b), Informationelle Bedingungen für die Selbständigkeit kausaler Satzaussagen. Eine diachrone Sichtweise, in R. Külpmann und R. Finkbeiner, eds, 'Selbständigkeit von Satzaussagen und Satzwertigkeit von Äußerungen', Buske, Hamburg, pp. 177–206.

Stolcke, A. (2002), SRILM-an extensible language modeling toolkit, in 'Seventh international conference on spoken language processing'.

- Strauß, D. (1835), *Das Leben Jesu, kritisch bearbeitet*, Osiander, Tübingen.
- Stutterheim, C. v. (2011), Hörerorientierung in der Kommunikation. Psycholinguistische Evidenz pro und contra, in A. Lehr, M. Kammerer, K.-P. Konerding, A. Storrer, C. Thimm und W. Wolski, eds, 'Sprache im Alltag. Beiträge zu neuen Perspektiven in der Linguistik. Herbert Ernst Wiegand zum 65. Geburtstag gewidmet', De Gruyter, Berlin, Boston, pp. 467–484.
URL: <https://doi.org/10.1515/9783110880380.467>
- Takada, H. (1998), *Grammatik und Sprachwirklichkeit von 1640 - 1700: zur Rolle deutscher Grammatiker im schriftsprachlichen Ausgleichsprozeß*, Reihe Germanistische Linguistik; 203, Niemeyer, Tübingen.
- Tily, H. und Piantadosi, S. (2009), Refer efficiently: Use less informative expressions for more predictable meanings, in 'Proceedings of the workshop on the production of referring expressions: Bridging the gap between computational and empirical approaches to reference.', pp. 1–8.
- Tipler, P. A., Mosca, G. und Wagner, J. (2015), *Physik für Wissenschaftler und Ingenieure*, Lehrbuch, Springer Spektrum, Berlin.
- Tomczyk-Popínska, E. (1987), 'Linguistische Merkmale der deutschen gesprochenen Standardsprache', *Deutsche Sprache: Zeitschrift für Theorie, Praxis, Dokumentation* 15(3), 336–375.
- Uhmann, S. (1993), Das Mittelfeld im Gespräch, in S. Reis, ed., 'Wortstellung und Informationsstruktur', Max Niemeyer Verlag, Berlin, New York, pp. 313–354.
- Unzer, J. A. (1746), *Gedanken vom Einfluß der Seele in ihren Körper*, Hemmerde, Halle.
- Uszkoreit, H., Brants, T., Duchier, D., Krenn, B., Konieczny, L., Oepen, S. und Skut, W. (1998), 'Studien zur performanzorientierten Linguistik. Aspekte der Relativsatzextraposition im Deutschen', *Kognitionswissenschaft* 7(2), 129–133.

- van Os, M., Kray, J. und Demberg, V. (2021), 'Mishearing as a Side Effect of Rational Language Comprehension in Noise', *Frontiers in Psychology* **12**, 1–17.
- van Son, R. und Pols, L. (2003), 'How efficient is speech?', *Proceedings (Instituut voor Fonetische Wetenschappen, Universiteit van Amsterdam)* **25**, 171–184.
- Vasishth, S., Suckow, K., Lewis, R. L. und Kern, S. (2010), 'Short-term forgetting in sentence comprehension: Crosslinguistic evidence from verb-final structures', *Language and Cognitive Processes* **25**, 533–567.
- Vierhaus, R. (1985), Einleitung, in R. Vierhaus, ed., 'Wissenschaften im Zeitalter der Aufklärung: Aus Anlaß des 250jährigen Bestehens des Verlages Vandenhoeck & Ruprecht', Vandenhoeck & Ruprecht, Göttingen.
- Vinckel, H. (2006), 'Die diskursstrategische Bedeutung des Nachfelds im Deutschen. Eine Untersuchung anhand politischer Reden der Gegenwartssprache', *Zeitschrift für Rezensionen zur germanistischen Sprachwissenschaft* **1**.
- Vinckel-Roisin, H. (2012), 'Das Nachfeld im Deutschen: Rechte Satzperipherie und Diskurstopik-Auszeichnung', *Studia Linguistica* **31**, 143–163.
- Voigtmann, S. (2022), Informational aspects of the extraposition of relative clauses, in M. Coniglio, C. de Bastiani und N. Catasso, eds, 'Language Change at the Interfaces. Intrasentential and intersentential phenomena', John Benjamins Publishing Company, Amsterdam / Philadelphia.
- Voigtmann, S. und Speyer, A. (2021), 'Information density and the extraposition of german relative clauses', *Frontiers in Psychology* pp. 1–18. <https://doi.org/10.3389/fpsyg.2021.650969>.
- Voigtmann, S. und Speyer, A. (2022), Information density as a factor for syntactic variation in early new high german, in 'Proceedings of Linguistic Evidence 2020', Tübingen, Germany.
- Voigtmann, S. und Speyer, A. (2023), 'Where to place a phrase?', *Journal of Historical Syntax. Proceedings of the 22nd Diachronic Generative Syntax (DiGS)*

Conference 7(6-19).

URL: <https://doi.org/10.18148/HS/2023.V7I6-19.151>

Waldenberger, S. (2009), *Präpositionen und Präpositionalphrasen im Mittelhochdeutschen*, Studien zur mittelhochdeutschen Grammatik; 3, Niemeyer, Tübingen.

Ward, N. und Vega, A. (2009), 'Looking for Entropy Rate Constancy in Spoken Dialog', *Departmental Technical Reports* 46.

Wasow, T. (1997), 'Remarks on grammatical weight', *Language Variation and Change* 9, 81–105.

Weber, S. (2018), Nominal Modification in Language Production: Extraposition of Prepositional Phrases in German, PhD thesis, Goethe Universität, Frankfurt am Main.

Weber, S. (2019), 'The Acceptability of Extraposition of PPs out of NP in German', *Proceedings of Linguistic Evidence 2018: Experimental Data Drives Linguistic Theory* 27, 63–84.

Welke, K. (2011), *Valenzgrammatik des Deutschen. Eine Einführung*, de Gruyter, Berlin, Boston.

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

URL: <https://ggplot2.tidyverse.org>

William Revelle (2024), *psych: Procedures for Psychological, Psychometric, and Personality Research*, Northwestern University, Evanston, Illinois. R package version 2.4.1.

URL: <https://CRAN.R-project.org/package=psych>

Winter, B. (2020), *Statistics for linguists: an introduction using R*, Routledge, Taylor and Francis Group, New York. <https://www.gbv.de/dms/bowker/toc/9781138056091.pdf>.

- Wöllstein, A. (2010), *Topologisches Satzmodell*, Kurze Einführungen in die germanistische Linguistik; 8, Winter, Heidelberg.
- Wöllstein, A. (2014), *Topologisches Satzmodell*, 2 edn, Winter, Heidelberg.
- Wöllstein, A., Heilmann, A., Stepan, P. und Vikner, S. (1997), *Deutsche Satzstruktur: Grundlagen der syntaktischen Analyse*, Stauffenburg Einführungen, Stauffenburg, Tübingen.
- Yimam, S. M., Gurevych, I., de Castilho, R. E. und Biemann, C. (2013), WebAnno: A Flexible, Web-based and Visually Supported System for Distributed Annotations, in 'Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (System Demonstrations) (ACL 2013)', pp. 1–6.
- Zahn, G. (1991), *Beobachtungen zur Ausklammerung und Nachfeldbesetzung in gesprochenem Deutsch*, Erlanger Studien; 93, Palm und Enke, Erlangen.
- Zeman, D., Hajič, J., Popel, M., Potthast, M., Straka, M., Ginter, F., Nivre, J. und Petrov, S. (2018), CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, in 'Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies', pp. 1–21.
- Zifonun, G., Hoffmann, L. und Strecker, B. (1997), *Grammatik der deutschen Sprache*, Niemeyer, Berlin.
- Zipf, G. K. (1965), *The psycho-biology of language: an introduction to dynamic philology*, The MIT Press paperback series; 38, M.I.T. Press, Cambridge, Mass.

A. Informationen zu den Autoren

- Bei **Heinrich Kaspar Abel** handelt es sich um einen Arzt aus Schleusingen im heutigen Thüringen. Er arbeitete am königlichen polnischen und kursächsischen Hof als Medicus. Über seine Lebensdaten ist nichts bekannt; als Wirkungsdaten werden die Jahre 1692 bis 1740 angegeben. Durch die Arbeit als Hofmedicus ist davon auszugehen, dass er Kontakte zu Menschen aus verschiedenen Regionen des deutschsprachigen Raums hatte und zudem des Lateinischen kundig war (vgl. Nationalbibliothek 2022a).
- **Johann Albert Bengel** war promovierter Theologe, Pietist und Bibelwissenschaftler aus Schwaben. Er lebte von 1687 bis 1752 und unternahm Studienreisen nach Halle und Nürnberg, kam also mit den dortigen Sprechgewohnheiten in Berührung und war Klosterpräzeptor in Denkendorf im heutigen Badenwürttemberg. 1741 war er Probst von Herbrechtingen und 1749 Konsistorialrat in Stuttgart. Er zählt zu den pietistischen ‚Schwabenvätern‘ (vgl. Nationalbibliothek 2022b).
- Auch **Johannes Hermann Benner** war ein evangelischer Theologe und Professor an der Justus-Liebig-Universität und Kirchenrat in Gießen (Hessen) tätig war. Er wurde 1699 geboren und starb 1782. Weitere Details über sein Leben sind nicht bekannt (vgl. Nationalbibliothek 2022c).
- Die Informationslage über den Arzt **Johann Jacob Bräuner** ist noch geringer. Es ist bekannt, dass er 1647 geboren wurde und zumindest eine Zeit lang in Frankfurt am Main tätig war (vgl. Nationalbibliothek 2022d).
- Über den Arzt, Naturwissenschaftler, Maler, Philosophen, Psychologen und Schriftsteller **Carl Gustav Carus**, der von 1789 bis 1869 lebte, gibt es hingegen deutlich mehr Informationen. Unter anderem war er Leibarzt des sächsischen Königs, Vorsitzender der Medizinalabteilung im

Sächsischen Ministerium des Innern, erwarb neben seinem medizinischen auch den philologischen Dokortitel und den Titel des des ‚Ritters des königlich sächsischen Civil-Verdienstordens‘. Sein Wirkungsort war vor allem in Sachsen. Trotzdem ist allein aufgrund der Berufe und der beiden Dokortitel davon auszugehen, dass Carus auf eine Vielzahl von Quellen zurückgreifen konnte, auch wenn nicht ersichtlich ist, ob er weitgereist ist (vgl. Nationalbibliothek 2022e).

- **Georg Friedrich Daumer** war Schriftsteller, Lehrer und Philosoph. Bekannt ist, dass er 1875 in Würzburg starb und entweder 1800 oder 1805 in Nürnberg geboren wurde. Weitere Informationen über sein Leben fehlen (vgl. Nationalbibliothek 2022f).
- Der Schweizer **Augustinus Egger** wurde 1833 in St. Gallen geboren und starb dort 1906. Er war in seinem Geburtsort erst als Priester und dann als Bischof tätig. Er wird als Autor von 139 Schriften gelistet (vgl. Dora 1994).
- **Franz Joseph Gall**, Begründer der Phrenologie und Arzt, wurde 1758 in Tiefenbronn geboren und starb 1828 in Montrouge. Als sein Wirkungsort wird Wien genannt, wobei er zur Verbreitung der Phrenologie durch Europa reiste und sich schließlich in der Nähe von Paris für seinen Lebensabend niederließ. Diese Tour durch Europa fand allerdings nach der Veröffentlichung des hier genutzten Werks statt (vgl. Nationalbibliothek 2022g).
- Über **Petrus Hanssen** stehen keine Informationen zur Verfügung.
- Ebenso dürftig ist die Informationslage über **Max Hasak**. Er fällt insofern aus dem Rahmen der bisher erwähnten Autoren, dass er Baumeister statt Philosoph, Geistlicher oder Arzt war. Dennoch wird sein hier genutztes Werk bei den theologischen Texten des DTA geführt. Sein Wirkungsort war Berlin (vgl. Nationalbibliothek 2022h).
- Das Problem, zu wenige Informationen zu haben, stellt sich hingegen bei **Robert Koch** nicht. 1843 wurde er in Clausthal geboren, forschte und arbeitete vor allem in Berlin und starb 1910 in Baden-Baden. Er war Bakteriologe, Arzt, Autor und 1905 Medizin-nobelpreisträger. Koch

gilt als Begründer der modernen Bakteriologie und Mikrobiologie in Deutschland. Er bildet den prominentesten Autor im Korpus (vgl. Nationalbibliothek 2022j).

- Der Psychiater und Arzt **Emil Kraepelin** ist ebenfalls nicht unbekannt. Er wurde 1856 in Neustrelitz geboren und starb 1926 in München. Als Wirkungsorte werden Tartu und Heidelberg genannt. Von den Autoren, über die mir mehr Informationen vorliegen, ist der der am weitesten gereiste. Unter anderem war er von 1886 bis 1889 Professor der Psychiatrie an der Universität Dorpat, dem heutigen Tartu in Estland, bevor er 1891 nach Heidelberg wechselte und dann nach München weiterzog, wo er von 1904 bis 1922 lehrte und 1917 die Deutsche Forschungsanstalt für Psychiatrie gründete. Bekannt wurde Kraepelin als Erfinder des Klassifikationssystems für psychische Störungen und als einer der ersten, die experimentelle Methoden in der Psychiatrie nutzten (vgl. Nationalbibliothek 2022k).
- Wieder bei den Theologen ist **Johann Konrad Wilhelm Löhe** zu verorten. Er wurde 1808 in Fürth in Bayern geboren und starb 1872 in Neuendettelsau, war Pfarrer und evangelischer Theologe. Seit 1837 war er Pfarrer in Neuendettelsau und wurde wegen der Gründung eines Mutterhauses für Diakonissen als fränkischer Diakonissenvater bekannt (vgl. Nationalbibliothek 2022m).
- Sein Zeitgenosse ist der nächste Autor im Korpus, der Professor der Physiologie **Karl Friedrich Wilhelm Ludwig**. Er lebte von 1816 bis 1895, wobei ein anderslautendes Sterbejahr 1893 ist. Während seiner Laufbahn war er an den Universitäten von Marburg, Wien, Leipzig und Zürich tätig, kam also mit verschiedenen deutschen Varietäten in Berührung (vgl. Nationalbibliothek 2022l).
- **Friedrich Niviands** wirkte von 1701 bis 1717 in Düsseldorf, wo er als Buchhändler arbeitete und 1701 zum Katholizismus übertrat. Weitere Lebensdaten sind zu ihm nicht zu finden. Seine Schrift *Güldenes Schwert* gilt als polemische Schrift mit theologischer Thematik (vgl. Nationalbibliothek 2022n).

- Die Angabe zu den Lebensdaten von **Gottfried Matthäus Purmann** variiert. Er wurde entweder 1649 geboren und starb 1711 oder er lebte von 1648 bis 1721. Geboren wurde er in Lubin (Polen) und starb in Breslau (Polen). Als Beruf findet sich zu ihm die Angabe, dass er als Chirurg und Stadtarzt in Halberstadt und Breslau sowie als Regimentsfeldscher gearbeitet hat (vgl. Nationalbibliothek 2022p).
- Ebenfalls zu den Theologen zählt **Johann Gottfried von Pahl**. Er wurde 1768 in Aalen im heutigen Baden-Württemberg geboren und starb 1839 in Stuttgart. Als Pfarrer, Historiker, Politiker, Prälat und Memoirenschreiber wirkte er vor allem in Neubronn in der Gemeinde Abtsgmünd (vgl. Nationalbibliothek 2022o).
- Der entweder 1758 oder 1759 in Rhaude geborene und 1813 in Halle an der Saale gestorbene **Johann Christian Reil** war Arzt, Anatom und Autor. Er wird als Internist gelistet, befasste sich jedoch auch mit psychologischen und psychiatrischen Themen (vgl. Nationalbibliothek 2022q).
- **Albrecht Christian Rotth**, geboren 1651 in Ottenhausen oder Greußen, und gestorben 1701 in Leipzig, wirkte in Jena, Halle an der Saale und Leipzig. Er war evangelischer Theologe und Pädagoge und Gegner des Pietismus. Er studierte in Jena, machte dort 1675 seinen Magister, arbeitete anschließend als Hauslehrer und an der Universität. 1683 wurde er zum Konrektor am Luther Gymnasium in Halle, bevor er sechs Jahre später (1689) zum Pfarrer ordiniert wurde. Seit 1692 war er Subdiakon in Leipzig und an der dortigen Universität Privatdozent. Zwei Jahre vor seinem Tod, also seit 1699, wurde er zum Diakon. Das Gebiet um Leipzig und Halle an der Saale hat er zeit seines Lebens offensichtlich nicht verlassen (vgl. Nationalbibliothek 2022r).
- Auch bei **Philipp Jacob Spener** handelt es sich um einen Theologen und evangelischen Pfarrer. Er wird als Gründer des lutherischen Pietismus bezeichnet. Spener wurde 1635 in Rappoltsweiler geboren und verstarb 1705 in Berlin. Er war Gymnasiast in Colmar, hat in Straßburg studiert und dort 1653 seinen Magister gemacht. Seit 1663 war er Freiprediger am Straßburger Münster, bevor er 1666 als Pfarrer nach Frankfurt am

Main wechselte. Sein Weg führte ihn 1686 weiter als Oberhofprediger nach Dresden und schließlich 1691 als Propst nach Berlin. Hier starb er schließlich auch (vgl. Nationalbibliothek 2022s).

- Der letzte Theologe auf der Liste ist **David Strauß**. Er lebte von 1808 bis 1874 in Ludwigsburg, war neben seinem Beruf als evangelischer Theologe auch Schriftsteller und Philosoph. Sein bekanntestes Werk *Das Leben Jesu, kritisch bearbeitet* ist das hier verwendete (vgl. Nationalbibliothek 2022t).
- Der letzte im Korpus verwendete Autor ist **Johann August Unzer**. Der Arzt, Psychologe und Schriftsteller wurde 1727 in Halle an der Saale geboren und starb 1799 in Altona, was auch als sein Wirkungsort angegeben wird (vgl. Nationalbibliothek 2022u).
- Was sich schon in den kurzen Biografien zeigt, ist die Heterogenität der Texte, aber auch Probleme darin, wirkliche Aussagen über die Sprachkontakte der Autoren zu treffen. Die Gruppe der Autoren ist, wie zu erwarten, recht homogen. Es handelt sich um klassisch gebildete Männer, von denen auszugehen ist, dass sie Latein und in Teilen Altgriechisch beherrschten. Die von ihnen im DTA repräsentierten Texte sind oftmals auch ihre wichtigsten Werke, die redaktionelle Prozesse durchlaufen haben dürften.

B. Liste der für die Analysen verwendeten Sätze

An dieser Stelle findet sich die Liste der Sätze, die in die Analyse einbezogen wurden. Die Sätze sind nach den Autoren (Tabellen 4.2 und 4.3) unterteilt und weisen die Satznummer auf, die den Sätzen im GitHub Repository (<https://github.com/rubcompling/C6Samples>) zugeordnet wurden. Die Sortierung erfolgt alphabetisch.

Benner_1	Benner_67	Benner_133	Daumer_21	Daumer_87	Daumer_154
Benner_2	Benner_68	Benner_134	Daumer_22	Daumer_89	Daumer_155
Benner_3	Benner_69	Benner_135	Daumer_23	Daumer_90	Daumer_156
Benner_4	Benner_70	Benner_136	Daumer_24	Daumer_91	Daumer_157
Benner_5	Benner_71	Benner_137	Daumer_25	Daumer_92	Daumer_158
Benner_6	Benner_72	Benner_138	Daumer_26	Daumer_93	Daumer_159
Benner_7	Benner_73	Benner_139	Daumer_27	Daumer_94	Daumer_160
Benner_8	Benner_74	Benner_140	Daumer_28	Daumer_95	Daumer_161
Benner_9	Benner_75	Benner_141	Daumer_29	Daumer_96	Daumer_162
Benner_10	Benner_76	Benner_142	Daumer_30	Daumer_97	Daumer_163
Benner_11	Benner_77	Benner_143	Daumer_31	Daumer_98	Daumer_164
Benner_12	Benner_78	Benner_144	Daumer_32	Daumer_99	Daumer_165
Benner_13	Benner_79	Benner_145	Daumer_33	Daumer_100	Daumer_166
Benner_14	Benner_80	Benner_146	Daumer_34	Daumer_101	Daumer_167
Benner_15	Benner_81	Benner_147	Daumer_35	Daumer_102	Daumer_168
Benner_16	Benner_82	Benner_148	Daumer_36	Daumer_103	Daumer_169
Benner_17	Benner_83	Benner_149	Daumer_37	Daumer_104	Daumer_170
Benner_18	Benner_84	Benner_150	Daumer_38	Daumer_105	Daumer_171
Benner_19	Benner_85	Benner_151	Daumer_39	Daumer_106	Daumer_172
Benner_20	Benner_86	Benner_152	Daumer_40	Daumer_107	Daumer_174
Benner_21	Benner_87	Benner_153	Daumer_41	Daumer_108	Daumer_175
Benner_22	Benner_88	Benner_154	Daumer_42	Daumer_109	Daumer_176
Benner_23	Benner_89	Benner_155	Daumer_43	Daumer_110	Daumer_177
Benner_24	Benner_90	Benner_156	Daumer_44	Daumer_111	Daumer_178
Benner_25	Benner_91	Benner_157	Daumer_45	Daumer_112	Daumer_179
Benner_26	Benner_92	Benner_158	Daumer_46	Daumer_113	Daumer_180
Benner_28	Benner_93	Benner_159	Daumer_47	Daumer_114	Daumer_181
Benner_29	Benner_94	Benner_160	Daumer_48	Daumer_115	Daumer_182
Benner_30	Benner_95	Benner_161	Daumer_49	Daumer_116	Daumer_183
Benner_31	Benner_96	Benner_162	Daumer_50	Daumer_117	Daumer_184
Benner_32	Benner_97	Benner_163	Daumer_51	Daumer_118	Daumer_185
Benner_33	Benner_98	Benner_164	Daumer_52	Daumer_119	Daumer_186
Benner_34	Benner_99	Benner_165	Daumer_53	Daumer_120	Daumer_187
Benner_35	Benner_100	Benner_166	Daumer_54	Daumer_121	Daumer_188
Benner_36	Benner_101	Benner_167	Daumer_55	Daumer_122	Daumer_189
Benner_37	Benner_102	Benner_168	Daumer_56	Daumer_123	Daumer_190
Benner_38	Benner_103	Benner_169	Daumer_57	Daumer_124	Daumer_191
Benner_39	Benner_104	Benner_170	Daumer_58	Daumer_125	Daumer_192
Benner_40	Benner_105	Benner_171	Daumer_59	Daumer_126	Daumer_193
Benner_41	Benner_106	Benner_172	Daumer_60	Daumer_127	Daumer_194
Benner_42	Benner_107	Benner_173	Daumer_61	Daumer_128	Daumer_195
Benner_43	Benner_108	Benner_174	Daumer_62	Daumer_129	Daumer_196
Benner_44	Benner_109	Benner_175	Daumer_63	Daumer_131	Daumer_198
Benner_45	Benner_110	Benner_176	Daumer_64	Daumer_132	Daumer_199
Benner_46	Benner_111	Benner_177	Daumer_65	Daumer_133	Daumer_200
Benner_47	Benner_112	Daumer_1	Daumer_66	Daumer_134	Daumer_201
Benner_48	Benner_113	Daumer_2	Daumer_67	Daumer_135	Daumer_203
Benner_49	Benner_114	Daumer_3	Daumer_68	Daumer_136	Daumer_204
Benner_50	Benner_115	Daumer_4	Daumer_69	Daumer_137	Daumer_205
Benner_51	Benner_116	Daumer_5	Daumer_70	Daumer_138	Daumer_206
Benner_52	Benner_117	Daumer_6	Daumer_71	Daumer_139	Daumer_207
Benner_53	Benner_118	Daumer_7	Daumer_73	Daumer_140	Daumer_208
Benner_54	Benner_119	Daumer_8	Daumer_74	Daumer_141	Daumer_209
Benner_55	Benner_121	Daumer_9	Daumer_75	Daumer_142	Daumer_210
Benner_56	Benner_122	Daumer_10	Daumer_76	Daumer_143	Daumer_211
Benner_57	Benner_123	Daumer_11	Daumer_77	Daumer_144	Daumer_212
Benner_58	Benner_124	Daumer_12	Daumer_78	Daumer_145	Daumer_213
Benner_59	Benner_125	Daumer_13	Daumer_79	Daumer_146	Daumer_214
Benner_60	Benner_126	Daumer_14	Daumer_80	Daumer_147	Daumer_215
Benner_61	Benner_127	Daumer_15	Daumer_81	Daumer_148	Daumer_216
Benner_62	Benner_128	Daumer_16	Daumer_82	Daumer_149	Daumer_217
Benner_63	Benner_129	Daumer_17	Daumer_83	Daumer_150	Daumer_218
Benner_64	Benner_130	Daumer_18	Daumer_84	Daumer_151	Daumer_219
Benner_65	Benner_131	Daumer_19	Daumer_85	Daumer_152	Daumer_220
Benner_66	Benner_132	Daumer_20	Daumer_86	Daumer_153	Daumer_221

Daumer_222	Daumer_288	Daumer_354	Hasak_40	Hasak_117	Loehe_58
Daumer_223	Daumer_289	Daumer_355	Hasak_41	Hasak_118	Loehe_59
Daumer_224	Daumer_290	Daumer_356	Hasak_42	Hasak_119	Loehe_60
Daumer_225	Daumer_291	Daumer_357	Hasak_43	Hasak_120	Loehe_61
Daumer_226	Daumer_292	Daumer_358	Hasak_44	Hasak_121	Loehe_62
Daumer_227	Daumer_293	Daumer_359	Hasak_45	Hasak_122	Loehe_63
Daumer_228	Daumer_294	Daumer_360	Hasak_46	Hasak_123	Loehe_64
Daumer_229	Daumer_295	Daumer_362	Hasak_47	Hasak_124	Loehe_65
Daumer_230	Daumer_296	Daumer_363	Hasak_48	Hasak_126	Loehe_66
Daumer_231	Daumer_297	Daumer_364	Hasak_49	Hasak_127	Loehe_67
Daumer_233	Daumer_298	Daumer_365	Hasak_50	Hasak_128	Loehe_68
Daumer_234	Daumer_299	Daumer_366	Hasak_51	Hasak_129	Loehe_69
Daumer_235	Daumer_300	Daumer_367	Hasak_52	Hasak_130	Loehe_70
Daumer_236	Daumer_301	Daumer_368	Hasak_53	Loehe_2	Loehe_71
Daumer_237	Daumer_302	Daumer_369	Hasak_54	Loehe_3	Loehe_72
Daumer_238	Daumer_303	Daumer_370	Hasak_55	Loehe_4	Loehe_73
Daumer_239	Daumer_304	Daumer_371	Hasak_56	Loehe_5	Loehe_74
Daumer_240	Daumer_305	Daumer_372	Hasak_57	Loehe_6	Loehe_75
Daumer_241	Daumer_306	Daumer_373	Hasak_59	Loehe_7	Loehe_76
Daumer_242	Daumer_307	Daumer_374	Hasak_60	Loehe_8	Loehe_78
Daumer_243	Daumer_308	Daumer_375	Hasak_61	Loehe_9	Loehe_79
Daumer_244	Daumer_309	Daumer_376	Hasak_62	Loehe_12	Loehe_81
Daumer_245	Daumer_310	Daumer_377	Hasak_63	Loehe_13	Loehe_82
Daumer_246	Daumer_311	Daumer_378	Hasak_64	Loehe_14	Loehe_83
Daumer_247	Daumer_312	Daumer_379	Hasak_66	Loehe_15	Loehe_84
Daumer_248	Daumer_313	Daumer_380	Hasak_68	Loehe_16	Loehe_85
Daumer_249	Daumer_314	Daumer_381	Hasak_69	Loehe_17	Loehe_86
Daumer_250	Daumer_315	Daumer_382	Hasak_70	Loehe_18	Loehe_87
Daumer_251	Daumer_316	Daumer_383	Hasak_71	Loehe_19	Loehe_88
Daumer_252	Daumer_317	Hasak_1	Hasak_72	Loehe_20	Loehe_89
Daumer_253	Daumer_318	Hasak_2	Hasak_73	Loehe_21	Loehe_91
Daumer_254	Daumer_319	Hasak_3	Hasak_75	Loehe_22	Loehe_92
Daumer_255	Daumer_320	Hasak_4	Hasak_77	Loehe_23	Loehe_94
Daumer_256	Daumer_322	Hasak_5	Hasak_78	Loehe_24	Loehe_95
Daumer_257	Daumer_323	Hasak_6	Hasak_79	Loehe_25	Loehe_96
Daumer_258	Daumer_324	Hasak_7	Hasak_80	Loehe_26	Loehe_97
Daumer_259	Daumer_325	Hasak_8	Hasak_82	Loehe_27	Loehe_98
Daumer_260	Daumer_326	Hasak_9	Hasak_83	Loehe_28	Loehe_99
Daumer_261	Daumer_327	Hasak_11	Hasak_85	Loehe_29	Loehe_101
Daumer_262	Daumer_328	Hasak_12	Hasak_86	Loehe_30	Loehe_103
Daumer_263	Daumer_329	Hasak_13	Hasak_87	Loehe_31	Loehe_104
Daumer_264	Daumer_330	Hasak_14	Hasak_88	Loehe_33	Loehe_105
Daumer_265	Daumer_331	Hasak_15	Hasak_90	Loehe_34	Loehe_106
Daumer_266	Daumer_332	Hasak_17	Hasak_91	Loehe_35	Loehe_107
Daumer_267	Daumer_333	Hasak_18	Hasak_92	Loehe_36	Loehe_108
Daumer_268	Daumer_334	Hasak_19	Hasak_94	Loehe_37	Loehe_109
Daumer_269	Daumer_335	Hasak_20	Hasak_96	Loehe_38	Loehe_110
Daumer_270	Daumer_336	Hasak_21	Hasak_97	Loehe_39	Loehe_111
Daumer_271	Daumer_337	Hasak_22	Hasak_98	Loehe_40	Loehe_112
Daumer_272	Daumer_338	Hasak_24	Hasak_99	Loehe_41	Loehe_113
Daumer_273	Daumer_339	Hasak_25	Hasak_100	Loehe_42	Loehe_114
Daumer_274	Daumer_340	Hasak_26	Hasak_101	Loehe_43	Loehe_115
Daumer_275	Daumer_341	Hasak_27	Hasak_102	Loehe_44	Pahl_1
Daumer_276	Daumer_342	Hasak_28	Hasak_104	Loehe_45	Pahl_3
Daumer_277	Daumer_343	Hasak_29	Hasak_105	Loehe_46	Pahl_4
Daumer_278	Daumer_344	Hasak_30	Hasak_107	Loehe_47	Pahl_5
Daumer_279	Daumer_345	Hasak_31	Hasak_108	Loehe_49	Pahl_6
Daumer_280	Daumer_346	Hasak_32	Hasak_109	Loehe_50	Pahl_7
Daumer_281	Daumer_347	Hasak_33	Hasak_110	Loehe_51	Pahl_8
Daumer_282	Daumer_348	Hasak_34	Hasak_111	Loehe_52	Pahl_9
Daumer_283	Daumer_349	Hasak_35	Hasak_112	Loehe_53	Pahl_10
Daumer_284	Daumer_350	Hasak_36	Hasak_113	Loehe_54	Pahl_11
Daumer_285	Daumer_351	Hasak_37	Hasak_114	Loehe_55	Pahl_12
Daumer_286	Daumer_352	Hasak_38	Hasak_115	Loehe_56	Pahl_13
Daumer_287	Daumer_353	Hasak_39	Hasak_116	Loehe_57	Pahl_14

Pahl_15	Pahl_88	Pahl_157	Pahl_223	Pahl_294	Pahl_365
Pahl_17	Pahl_89	Pahl_158	Pahl_224	Pahl_295	Pahl_366
Pahl_18	Pahl_90	Pahl_159	Pahl_225	Pahl_296	Pahl_367
Pahl_19	Pahl_91	Pahl_160	Pahl_226	Pahl_297	Pahl_368
Pahl_20	Pahl_92	Pahl_161	Pahl_227	Pahl_298	Pahl_369
Pahl_21	Pahl_93	Pahl_162	Pahl_228	Pahl_299	Pahl_372
Pahl_22	Pahl_94	Pahl_163	Pahl_229	Pahl_300	Pahl_373
Pahl_23	Pahl_95	Pahl_164	Pahl_230	Pahl_301	Pahl_374
Pahl_24	Pahl_96	Pahl_165	Pahl_231	Pahl_302	Pahl_375
Pahl_25	Pahl_97	Pahl_166	Pahl_232	Pahl_303	Pahl_376
Pahl_26	Pahl_98	Pahl_167	Pahl_233	Pahl_304	Pahl_377
Pahl_27	Pahl_99	Pahl_168	Pahl_234	Pahl_305	Pahl_378
Pahl_28	Pahl_100	Pahl_169	Pahl_235	Pahl_306	Pahl_379
Pahl_29	Pahl_101	Pahl_170	Pahl_236	Pahl_307	Pahl_380
Pahl_30	Pahl_103	Pahl_171	Pahl_237	Pahl_308	Pahl_381
Pahl_31	Pahl_104	Pahl_172	Pahl_238	Pahl_309	Pahl_382
Pahl_32	Pahl_105	Pahl_173	Pahl_239	Pahl_310	Pahl_383
Pahl_33	Pahl_106	Pahl_174	Pahl_240	Pahl_311	Pahl_384
Pahl_34	Pahl_107	Pahl_175	Pahl_241	Pahl_312	Pahl_385
Pahl_35	Pahl_108	Pahl_176	Pahl_242	Pahl_313	Pahl_386
Pahl_37	Pahl_109	Pahl_177	Pahl_243	Pahl_314	Pahl_387
Pahl_38	Pahl_110	Pahl_178	Pahl_244	Pahl_315	Pahl_388
Pahl_39	Pahl_111	Pahl_179	Pahl_245	Pahl_316	Pahl_390
Pahl_40	Pahl_112	Pahl_180	Pahl_246	Pahl_317	Pahl_391
Pahl_41	Pahl_113	Pahl_181	Pahl_247	Pahl_318	Pahl_392
Pahl_42	Pahl_114	Pahl_182	Pahl_248	Pahl_319	Pahl_393
Pahl_43	Pahl_115	Pahl_183	Pahl_249	Pahl_320	Pahl_394
Pahl_44	Pahl_117	Pahl_184	Pahl_250	Pahl_322	Pahl_395
Pahl_45	Pahl_118	Pahl_185	Pahl_251	Pahl_323	Pahl_396
Pahl_46	Pahl_119	Pahl_186	Pahl_253	Pahl_324	Pahl_397
Pahl_47	Pahl_120	Pahl_187	Pahl_254	Pahl_325	Pahl_398
Pahl_48	Pahl_121	Pahl_188	Pahl_256	Pahl_326	Pahl_399
Pahl_49	Pahl_122	Pahl_190	Pahl_258	Pahl_327	Pahl_400
Pahl_50	Pahl_123	Pahl_191	Pahl_259	Pahl_328	Pahl_401
Pahl_51	Pahl_124	Pahl_192	Pahl_260	Pahl_329	Pahl_402
Pahl_52	Pahl_125	Pahl_193	Pahl_262	Pahl_330	Pahl_404
Pahl_53	Pahl_126	Pahl_194	Pahl_264	Pahl_331	Pahl_405
Pahl_55	Pahl_127	Pahl_195	Pahl_265	Pahl_333	Pahl_406
Pahl_58	Pahl_128	Pahl_196	Pahl_266	Pahl_334	Pahl_407
Pahl_59	Pahl_129	Pahl_197	Pahl_267	Pahl_335	Pahl_408
Pahl_60	Pahl_130	Pahl_198	Pahl_268	Pahl_336	Pahl_409
Pahl_61	Pahl_131	Pahl_199	Pahl_269	Pahl_337	Pahl_410
Pahl_62	Pahl_132	Pahl_200	Pahl_270	Pahl_338	Pahl_412
Pahl_63	Pahl_133	Pahl_201	Pahl_271	Pahl_340	Pahl_413
Pahl_65	Pahl_134	Pahl_202	Pahl_272	Pahl_341	Pahl_414
Pahl_66	Pahl_135	Pahl_203	Pahl_273	Pahl_342	Pahl_415
Pahl_67	Pahl_136	Pahl_204	Pahl_274	Pahl_343	Pahl_416
Pahl_68	Pahl_137	Pahl_205	Pahl_275	Pahl_344	Pahl_417
Pahl_69	Pahl_140	Pahl_206	Pahl_276	Pahl_346	Pahl_419
Pahl_70	Pahl_141	Pahl_207	Pahl_277	Pahl_348	Pahl_420
Pahl_71	Pahl_142	Pahl_208	Pahl_278	Pahl_349	Pahl_421
Pahl_72	Pahl_143	Pahl_209	Pahl_279	Pahl_350	Pahl_422
Pahl_73	Pahl_144	Pahl_210	Pahl_280	Pahl_351	Pahl_423
Pahl_74	Pahl_145	Pahl_211	Pahl_281	Pahl_352	Pahl_424
Pahl_75	Pahl_146	Pahl_212	Pahl_282	Pahl_354	Pahl_425
Pahl_76	Pahl_147	Pahl_213	Pahl_283	Pahl_355	Pahl_427
Pahl_77	Pahl_148	Pahl_214	Pahl_284	Pahl_356	Pahl_428
Pahl_78	Pahl_149	Pahl_215	Pahl_285	Pahl_357	Pahl_429
Pahl_79	Pahl_150	Pahl_216	Pahl_286	Pahl_358	Pahl_430
Pahl_81	Pahl_151	Pahl_217	Pahl_287	Pahl_359	Pahl_432
Pahl_83	Pahl_152	Pahl_218	Pahl_289	Pahl_360	Pahl_434
Pahl_84	Pahl_153	Pahl_219	Pahl_290	Pahl_361	Pahl_435
Pahl_85	Pahl_154	Pahl_220	Pahl_291	Pahl_362	Pahl_436
Pahl_86	Pahl_155	Pahl_221	Pahl_292	Pahl_363	Pahl_438
Pahl_87	Pahl_156	Pahl_222	Pahl_293	Pahl_364	Pahl_439

Pahl_440	Niviands_37	Niviands_112	Niviands_185	Niviands_257	Niviands_329
Pahl_441	Niviands_38	Niviands_113	Niviands_186	Niviands_258	Niviands_330
Pahl_442	Niviands_39	Niviands_114	Niviands_187	Niviands_259	Niviands_331
Pahl_443	Niviands_40	Niviands_115	Niviands_188	Niviands_260	Niviands_332
Pahl_444	Niviands_41	Niviands_116	Niviands_189	Niviands_261	Niviands_333
Pahl_445	Niviands_43	Niviands_117	Niviands_190	Niviands_262	Niviands_334
Pahl_447	Niviands_44	Niviands_118	Niviands_191	Niviands_263	Niviands_335
Pahl_448	Niviands_46	Niviands_119	Niviands_192	Niviands_264	Niviands_336
Pahl_449	Niviands_47	Niviands_120	Niviands_193	Niviands_265	Niviands_337
Pahl_450	Niviands_48	Niviands_121	Niviands_194	Niviands_266	Niviands_338
Pahl_451	Niviands_49	Niviands_122	Niviands_196	Niviands_268	Niviands_339
Pahl_452	Niviands_50	Niviands_123	Niviands_197	Niviands_269	Niviands_340
Pahl_453	Niviands_51	Niviands_124	Niviands_198	Niviands_270	Niviands_341
Pahl_454	Niviands_52	Niviands_125	Niviands_199	Niviands_271	Niviands_342
Pahl_455	Niviands_53	Niviands_126	Niviands_200	Niviands_272	Niviands_343
Pahl_456	Niviands_54	Niviands_127	Niviands_201	Niviands_273	Niviands_344
Pahl_457	Niviands_55	Niviands_128	Niviands_202	Niviands_274	Niviands_345
Pahl_458	Niviands_56	Niviands_129	Niviands_203	Niviands_276	Niviands_346
Pahl_459	Niviands_57	Niviands_130	Niviands_204	Niviands_277	Niviands_347
Pahl_460	Niviands_58	Niviands_131	Niviands_205	Niviands_278	Niviands_348
Pahl_461	Niviands_59	Niviands_132	Niviands_207	Niviands_279	Niviands_349
Pahl_462	Niviands_60	Niviands_134	Niviands_208	Niviands_280	Niviands_350
Pahl_463	Niviands_61	Niviands_135	Niviands_210	Niviands_281	Niviands_351
Pahl_464	Niviands_62	Niviands_137	Niviands_211	Niviands_282	Niviands_352
Pahl_465	Niviands_63	Niviands_138	Niviands_212	Niviands_283	Niviands_353
Pahl_466	Niviands_64	Niviands_139	Niviands_213	Niviands_284	Niviands_354
Pahl_467	Niviands_65	Niviands_140	Niviands_214	Niviands_285	Niviands_355
Pahl_468	Niviands_66	Niviands_141	Niviands_215	Niviands_287	Niviands_356
Pahl_469	Niviands_68	Niviands_142	Niviands_217	Niviands_288	Niviands_357
Niviands_1	Niviands_70	Niviands_143	Niviands_218	Niviands_289	Niviands_358
Niviands_2	Niviands_71	Niviands_144	Niviands_219	Niviands_290	Niviands_359
Niviands_3	Niviands_72	Niviands_145	Niviands_220	Niviands_291	Niviands_360
Niviands_4	Niviands_73	Niviands_146	Niviands_221	Niviands_292	Niviands_361
Niviands_5	Niviands_74	Niviands_147	Niviands_223	Niviands_293	Niviands_362
Niviands_6	Niviands_75	Niviands_148	Niviands_224	Niviands_294	Niviands_363
Niviands_7	Niviands_76	Niviands_150	Niviands_226	Niviands_295	Niviands_364
Niviands_8	Niviands_77	Niviands_152	Niviands_227	Niviands_296	Niviands_365
Niviands_9	Niviands_78	Niviands_153	Niviands_228	Niviands_297	Niviands_366
Niviands_10	Niviands_79	Niviands_154	Niviands_229	Niviands_299	Niviands_367
Niviands_11	Niviands_81	Niviands_155	Niviands_230	Niviands_300	Niviands_368
Niviands_12	Niviands_83	Niviands_156	Niviands_231	Niviands_301	Niviands_369
Niviands_13	Niviands_84	Niviands_157	Niviands_232	Niviands_302	Niviands_371
Niviands_14	Niviands_85	Niviands_160	Niviands_234	Niviands_303	Niviands_372
Niviands_15	Niviands_86	Niviands_162	Niviands_235	Niviands_304	Niviands_373
Niviands_16	Niviands_87	Niviands_163	Niviands_236	Niviands_305	Niviands_374
Niviands_17	Niviands_88	Niviands_164	Niviands_237	Niviands_306	Niviands_375
Niviands_18	Niviands_89	Niviands_166	Niviands_238	Niviands_307	Niviands_376
Niviands_19	Niviands_90	Niviands_167	Niviands_239	Niviands_308	Niviands_377
Niviands_20	Niviands_91	Niviands_168	Niviands_240	Niviands_309	Niviands_378
Niviands_21	Niviands_93	Niviands_169	Niviands_241	Niviands_310	Niviands_379
Niviands_22	Niviands_94	Niviands_170	Niviands_242	Niviands_311	Niviands_380
Niviands_23	Niviands_95	Niviands_171	Niviands_243	Niviands_312	Niviands_381
Niviands_24	Niviands_96	Niviands_172	Niviands_244	Niviands_313	Niviands_382
Niviands_25	Niviands_97	Niviands_173	Niviands_245	Niviands_314	Hanssen_1
Niviands_26	Niviands_98	Niviands_174	Niviands_246	Niviands_316	Hanssen_2
Niviands_27	Niviands_99	Niviands_175	Niviands_247	Niviands_317	Hanssen_3
Niviands_28	Niviands_100	Niviands_176	Niviands_248	Niviands_318	Hanssen_4
Niviands_29	Niviands_103	Niviands_177	Niviands_249	Niviands_319	Hanssen_5
Niviands_30	Niviands_104	Niviands_178	Niviands_250	Niviands_322	Hanssen_6
Niviands_31	Niviands_105	Niviands_179	Niviands_251	Niviands_323	Hanssen_7
Niviands_32	Niviands_107	Niviands_180	Niviands_252	Niviands_324	Hanssen_8
Niviands_33	Niviands_108	Niviands_181	Niviands_253	Niviands_325	Hanssen_9
Niviands_34	Niviands_109	Niviands_182	Niviands_254	Niviands_326	Hanssen_10
Niviands_35	Niviands_110	Niviands_183	Niviands_255	Niviands_327	Hanssen_11
Niviands_36	Niviands_111	Niviands_184	Niviands_256	Niviands_328	Hanssen_12

Hanssen_443	Hanssen_508	Rotth_28	Strauss_22	Strauss_89	Carus_4
Hanssen_444	Hanssen_509	Rotth_29	Strauss_23	Strauss_90	Carus_5
Hanssen_445	Hanssen_510	Rotth_30	Strauss_24	Strauss_91	Carus_6
Hanssen_446	Hanssen_511	Rotth_31	Strauss_25	Strauss_92	Carus_7
Hanssen_447	Hanssen_512	Rotth_32	Strauss_26	Strauss_93	Carus_8
Hanssen_448	Hanssen_513	Rotth_33	Strauss_27	Strauss_94	Carus_9
Hanssen_449	Hanssen_514	Rotth_34	Strauss_28	Strauss_95	Carus_10
Hanssen_450	Hanssen_515	Rotth_35	Strauss_29	Strauss_96	Carus_11
Hanssen_451	Hanssen_517	Rotth_36	Strauss_30	Strauss_97	Carus_12
Hanssen_452	Hanssen_519	Rotth_37	Strauss_31	Strauss_98	Carus_13
Hanssen_453	Hanssen_520	Rotth_39	Strauss_32	Strauss_99	Carus_14
Hanssen_454	Hanssen_521	Rotth_40	Strauss_33	Strauss_100	Carus_15
Hanssen_455	Hanssen_523	Rotth_41	Strauss_34	Strauss_101	Carus_16
Hanssen_456	Hanssen_524	Rotth_42	Strauss_35	Strauss_102	Carus_17
Hanssen_457	Hanssen_525	Rotth_43	Strauss_36	Strauss_103	Carus_18
Hanssen_458	Hanssen_526	Rotth_44	Strauss_37	Strauss_104	Carus_19
Hanssen_459	Hanssen_527	Rotth_45	Strauss_38	Strauss_105	Carus_20
Hanssen_460	Hanssen_528	Rotth_46	Strauss_39	Strauss_106	Carus_21
Hanssen_461	Hanssen_529	Rotth_47	Strauss_40	Strauss_107	Carus_22
Hanssen_462	Hanssen_530	Rotth_48	Strauss_41	Strauss_108	Carus_23
Hanssen_463	Hanssen_531	Rotth_50	Strauss_42	Strauss_109	Carus_24
Hanssen_464	Hanssen_532	Rotth_51	Strauss_43	Strauss_110	Carus_25
Hanssen_465	Hanssen_534	Rotth_52	Strauss_44	Strauss_111	Carus_26
Hanssen_466	Hanssen_535	Rotth_53	Strauss_45	Strauss_112	Carus_27
Hanssen_467	Hanssen_536	Rotth_54	Strauss_46	Strauss_113	Carus_28
Hanssen_468	Hanssen_537	Rotth_55	Strauss_47	Strauss_114	Carus_29
Hanssen_469	Hanssen_538	Rotth_56	Strauss_48	Strauss_115	Carus_30
Hanssen_470	Hanssen_539	Rotth_57	Strauss_49	Strauss_116	Carus_31
Hanssen_471	Hanssen_540	Rotth_58	Strauss_50	Strauss_117	Carus_32
Hanssen_472	Hanssen_541	Rotth_59	Strauss_51	Strauss_118	Carus_33
Hanssen_473	Hanssen_542	Rotth_60	Strauss_52	Strauss_119	Carus_34
Hanssen_474	Hanssen_543	Rotth_61	Strauss_53	Strauss_122	Carus_35
Hanssen_475	Hanssen_544	Rotth_62	Strauss_54	Strauss_123	Carus_36
Hanssen_476	Hanssen_545	Rotth_63	Strauss_55	Strauss_124	Carus_37
Hanssen_477	Hanssen_546	Rotth_64	Strauss_56	Strauss_125	Carus_38
Hanssen_478	Hanssen_547	Rotth_65	Strauss_57	Strauss_126	Carus_39
Hanssen_479	Hanssen_548	Rotth_66	Strauss_58	Strauss_127	Carus_40
Hanssen_480	Hanssen_549	Rotth_67	Strauss_59	Strauss_128	Carus_41
Hanssen_481	Hanssen_550	Rotth_68	Strauss_60	Strauss_129	Carus_42
Hanssen_482	Rotth_1	Rotth_69	Strauss_61	Strauss_130	Carus_43
Hanssen_483	Rotth_2	Rotth_70	Strauss_62	Strauss_131	Carus_44
Hanssen_484	Rotth_3	Rotth_71	Strauss_63	Strauss_132	Carus_45
Hanssen_485	Rotth_4	Rotth_72	Strauss_64	Strauss_133	Carus_46
Hanssen_486	Rotth_5	Rotth_73	Strauss_65	Strauss_134	Carus_47
Hanssen_487	Rotth_7	Rotth_74	Strauss_66	Strauss_135	Carus_49
Hanssen_488	Rotth_8	Strauss_1	Strauss_67	Strauss_136	Carus_50
Hanssen_489	Rotth_9	Strauss_2	Strauss_68	Strauss_137	Carus_51
Hanssen_490	Rotth_10	Strauss_3	Strauss_69	Strauss_138	Carus_52
Hanssen_491	Rotth_11	Strauss_4	Strauss_70	Strauss_139	Carus_53
Hanssen_492	Rotth_12	Strauss_6	Strauss_71	Strauss_141	Carus_54
Hanssen_493	Rotth_13	Strauss_7	Strauss_72	Strauss_142	Carus_55
Hanssen_494	Rotth_14	Strauss_8	Strauss_73	Strauss_143	Carus_56
Hanssen_495	Rotth_15	Strauss_9	Strauss_74	Strauss_145	Carus_57
Hanssen_496	Rotth_16	Strauss_10	Strauss_75	Strauss_146	Carus_58
Hanssen_497	Rotth_17	Strauss_11	Strauss_76	Strauss_147	Carus_59
Hanssen_498	Rotth_18	Strauss_12	Strauss_77	Strauss_148	Carus_60
Hanssen_499	Rotth_19	Strauss_13	Strauss_78	Strauss_149	Carus_61
Hanssen_500	Rotth_20	Strauss_14	Strauss_79	Strauss_150	Carus_62
Hanssen_501	Rotth_21	Strauss_15	Strauss_80	Strauss_151	Carus_63
Hanssen_502	Rotth_22	Strauss_16	Strauss_81	Strauss_152	Carus_64
Hanssen_503	Rotth_23	Strauss_17	Strauss_82	Strauss_153	Carus_65
Hanssen_504	Rotth_24	Strauss_18	Strauss_85	Strauss_154	Carus_67
Hanssen_505	Rotth_25	Strauss_19	Strauss_86	Carus_1	Carus_68
Hanssen_506	Rotth_26	Strauss_20	Strauss_87	Carus_2	Carus_69
Hanssen_507	Rotth_27	Strauss_21	Strauss_88	Carus_3	Carus_70

Carus_71	Carus_145	Carus_213	Unzer_19	Unzer_108	Unzer_193
Carus_72	Carus_146	Carus_214	Unzer_20	Unzer_109	Unzer_197
Carus_73	Carus_147	Carus_215	Unzer_22	Unzer_111	Unzer_198
Carus_74	Carus_148	Carus_216	Unzer_24	Unzer_113	Unzer_199
Carus_75	Carus_149	Carus_217	Unzer_25	Unzer_115	Unzer_200
Carus_76	Carus_150	Carus_218	Unzer_26	Unzer_116	Unzer_201
Carus_78	Carus_151	Carus_219	Unzer_28	Unzer_118	Unzer_204
Carus_79	Carus_152	Carus_220	Unzer_29	Unzer_119	Unzer_206
Carus_81	Carus_153	Carus_221	Unzer_30	Unzer_120	Unzer_207
Carus_82	Carus_154	Carus_222	Unzer_31	Unzer_121	Unzer_208
Carus_84	Carus_155	Carus_223	Unzer_33	Unzer_123	Unzer_209
Carus_85	Carus_156	Carus_224	Unzer_34	Unzer_124	Unzer_210
Carus_86	Carus_157	Carus_226	Unzer_35	Unzer_125	Unzer_212
Carus_88	Carus_158	Carus_227	Unzer_36	Unzer_126	Unzer_213
Carus_89	Carus_159	Carus_228	Unzer_37	Unzer_127	Unzer_214
Carus_90	Carus_160	Carus_229	Unzer_38	Unzer_128	Unzer_215
Carus_91	Carus_162	Carus_230	Unzer_39	Unzer_130	Unzer_216
Carus_92	Carus_163	Carus_231	Unzer_43	Unzer_131	Unzer_218
Carus_93	Carus_164	Carus_232	Unzer_45	Unzer_133	Unzer_220
Carus_94	Carus_165	Carus_233	Unzer_46	Unzer_134	Unzer_221
Carus_95	Carus_166	Carus_234	Unzer_47	Unzer_135	Unzer_222
Carus_96	Carus_167	Carus_235	Unzer_49	Unzer_136	Unzer_224
Carus_97	Carus_168	Carus_236	Unzer_50	Unzer_137	Unzer_225
Carus_98	Carus_169	Carus_237	Unzer_51	Unzer_138	Unzer_227
Carus_99	Carus_170	Carus_238	Unzer_53	Unzer_139	Unzer_228
Carus_100	Carus_171	Carus_239	Unzer_54	Unzer_140	Unzer_229
Carus_101	Carus_172	Carus_240	Unzer_56	Unzer_141	Unzer_230
Carus_102	Carus_173	Carus_241	Unzer_57	Unzer_142	Unzer_233
Carus_103	Carus_175	Carus_244	Unzer_59	Unzer_143	Unzer_234
Carus_105	Carus_176	Carus_245	Unzer_60	Unzer_144	Unzer_235
Carus_106	Carus_177	Carus_246	Unzer_61	Unzer_145	Unzer_236
Carus_107	Carus_178	Carus_247	Unzer_62	Unzer_146	Unzer_238
Carus_108	Carus_180	Carus_248	Unzer_65	Unzer_147	Unzer_239
Carus_109	Carus_181	Carus_249	Unzer_67	Unzer_148	Unzer_240
Carus_110	Carus_182	Carus_250	Unzer_69	Unzer_149	Unzer_241
Carus_111	Carus_183	Carus_251	Unzer_70	Unzer_150	Unzer_242
Carus_112	Carus_184	Carus_252	Unzer_71	Unzer_151	Unzer_243
Carus_114	Carus_185	Carus_253	Unzer_73	Unzer_152	Unzer_244
Carus_115	Carus_186	Carus_254	Unzer_74	Unzer_153	Unzer_245
Carus_116	Carus_187	Carus_255	Unzer_75	Unzer_154	Unzer_246
Carus_117	Carus_188	Carus_256	Unzer_76	Unzer_155	Unzer_247
Carus_118	Carus_189	Carus_257	Unzer_80	Unzer_157	Unzer_248
Carus_119	Carus_190	Carus_258	Unzer_82	Unzer_158	Unzer_249
Carus_120	Carus_191	Carus_259	Unzer_83	Unzer_159	Unzer_250
Carus_122	Carus_192	Carus_260	Unzer_84	Unzer_161	Unzer_251
Carus_123	Carus_193	Carus_261	Unzer_85	Unzer_163	Unzer_252
Carus_124	Carus_194	Carus_262	Unzer_86	Unzer_165	Unzer_253
Carus_125	Carus_195	Carus_263	Unzer_87	Unzer_166	Unzer_254
Carus_126	Carus_196	Carus_264	Unzer_88	Unzer_167	Unzer_255
Carus_127	Carus_197	Carus_265	Unzer_90	Unzer_168	Unzer_256
Carus_129	Carus_198	Carus_266	Unzer_91	Unzer_169	Unzer_257
Carus_130	Carus_199	Carus_267	Unzer_93	Unzer_170	Unzer_258
Carus_131	Carus_200	Unzer_1	Unzer_94	Unzer_171	Unzer_259
Carus_132	Carus_201	Unzer_2	Unzer_95	Unzer_175	Unzer_260
Carus_134	Carus_202	Unzer_3	Unzer_96	Unzer_178	Unzer_261
Carus_135	Carus_203	Unzer_4	Unzer_97	Unzer_179	Unzer_262
Carus_136	Carus_204	Unzer_5	Unzer_98	Unzer_180	Unzer_263
Carus_137	Carus_205	Unzer_7	Unzer_99	Unzer_182	Unzer_264
Carus_138	Carus_206	Unzer_8	Unzer_100	Unzer_185	Unzer_265
Carus_139	Carus_207	Unzer_9	Unzer_102	Unzer_186	Unzer_268
Carus_140	Carus_208	Unzer_10	Unzer_103	Unzer_188	Unzer_269
Carus_141	Carus_209	Unzer_11	Unzer_104	Unzer_189	Unzer_270
Carus_142	Carus_210	Unzer_12	Unzer_105	Unzer_190	Unzer_271
Carus_143	Carus_211	Unzer_14	Unzer_106	Unzer_191	Unzer_274
Carus_144	Carus_212	Unzer_16	Unzer_107	Unzer_192	Unzer_276

Unzer_277	Unzer_358	Unzer_445	Reil_58	Reil_129	Reil_198
Unzer_278	Unzer_359	Unzer_446	Reil_59	Reil_130	Reil_199
Unzer_280	Unzer_361	Unzer_447	Reil_60	Reil_131	Reil_200
Unzer_281	Unzer_362	Unzer_448	Reil_61	Reil_132	Reil_201
Unzer_282	Unzer_363	Unzer_449	Reil_62	Reil_133	Reil_202
Unzer_283	Unzer_364	Unzer_452	Reil_63	Reil_134	Reil_203
Unzer_284	Unzer_365	Unzer_453	Reil_64	Reil_135	Reil_204
Unzer_285	Unzer_366	Unzer_454	Reil_65	Reil_136	Reil_205
Unzer_286	Unzer_367	Unzer_456	Reil_66	Reil_137	Reil_206
Unzer_287	Unzer_368	Unzer_457	Reil_67	Reil_138	Reil_207
Unzer_288	Unzer_369	Unzer_458	Reil_68	Reil_139	Reil_208
Unzer_289	Unzer_370	Unzer_459	Reil_70	Reil_140	Reil_209
Unzer_290	Unzer_371	Unzer_460	Reil_71	Reil_141	Reil_210
Unzer_291	Unzer_372	Unzer_461	Reil_72	Reil_142	Reil_211
Unzer_292	Unzer_373	Reil_1	Reil_73	Reil_143	Reil_212
Unzer_293	Unzer_374	Reil_3	Reil_74	Reil_144	Reil_213
Unzer_294	Unzer_375	Reil_4	Reil_75	Reil_145	Reil_214
Unzer_295	Unzer_376	Reil_5	Reil_76	Reil_146	Reil_215
Unzer_297	Unzer_378	Reil_6	Reil_77	Reil_147	Reil_216
Unzer_299	Unzer_379	Reil_7	Reil_78	Reil_148	Reil_217
Unzer_300	Unzer_380	Reil_8	Reil_79	Reil_149	Reil_218
Unzer_304	Unzer_382	Reil_9	Reil_80	Reil_150	Reil_220
Unzer_306	Unzer_383	Reil_10	Reil_81	Reil_151	Reil_221
Unzer_307	Unzer_387	Reil_11	Reil_82	Reil_152	Reil_222
Unzer_308	Unzer_388	Reil_12	Reil_83	Reil_153	Reil_223
Unzer_311	Unzer_389	Reil_13	Reil_84	Reil_154	Reil_224
Unzer_312	Unzer_390	Reil_14	Reil_85	Reil_155	Reil_225
Unzer_313	Unzer_391	Reil_15	Reil_86	Reil_156	Reil_226
Unzer_314	Unzer_392	Reil_16	Reil_87	Reil_157	Reil_227
Unzer_315	Unzer_393	Reil_17	Reil_88	Reil_158	Reil_228
Unzer_316	Unzer_394	Reil_18	Reil_89	Reil_159	Reil_229
Unzer_317	Unzer_395	Reil_19	Reil_90	Reil_160	Reil_230
Unzer_320	Unzer_396	Reil_20	Reil_91	Reil_161	Reil_231
Unzer_321	Unzer_397	Reil_22	Reil_92	Reil_164	Reil_232
Unzer_323	Unzer_398	Reil_24	Reil_93	Reil_165	Reil_233
Unzer_324	Unzer_399	Reil_25	Reil_94	Reil_166	Reil_234
Unzer_325	Unzer_400	Reil_26	Reil_96	Reil_167	Reil_235
Unzer_326	Unzer_401	Reil_28	Reil_97	Reil_168	Reil_237
Unzer_327	Unzer_404	Reil_29	Reil_98	Reil_169	Reil_238
Unzer_328	Unzer_405	Reil_30	Reil_99	Reil_170	Reil_239
Unzer_329	Unzer_406	Reil_31	Reil_100	Reil_171	Reil_240
Unzer_330	Unzer_407	Reil_32	Reil_101	Reil_172	Reil_241
Unzer_331	Unzer_409	Reil_33	Reil_102	Reil_173	Reil_242
Unzer_332	Unzer_410	Reil_34	Reil_103	Reil_174	Reil_243
Unzer_333	Unzer_412	Reil_35	Reil_104	Reil_175	Reil_244
Unzer_334	Unzer_413	Reil_36	Reil_105	Reil_176	Reil_245
Unzer_336	Unzer_414	Reil_37	Reil_107	Reil_177	Reil_246
Unzer_338	Unzer_415	Reil_39	Reil_108	Reil_178	Reil_247
Unzer_339	Unzer_416	Reil_40	Reil_109	Reil_179	Reil_248
Unzer_340	Unzer_417	Reil_41	Reil_110	Reil_180	Reil_249
Unzer_341	Unzer_418	Reil_42	Reil_111	Reil_181	Reil_250
Unzer_342	Unzer_419	Reil_43	Reil_112	Reil_183	Reil_251
Unzer_343	Unzer_420	Reil_44	Reil_113	Reil_184	Reil_252
Unzer_344	Unzer_421	Reil_45	Reil_114	Reil_185	Reil_253
Unzer_347	Unzer_422	Reil_46	Reil_115	Reil_186	Reil_254
Unzer_348	Unzer_423	Reil_47	Reil_116	Reil_187	Reil_255
Unzer_349	Unzer_427	Reil_48	Reil_119	Reil_188	Reil_256
Unzer_350	Unzer_431	Reil_49	Reil_120	Reil_189	Reil_257
Unzer_351	Unzer_432	Reil_50	Reil_121	Reil_190	Reil_258
Unzer_352	Unzer_433	Reil_51	Reil_123	Reil_191	Reil_259
Unzer_353	Unzer_434	Reil_52	Reil_124	Reil_192	Reil_260
Unzer_354	Unzer_439	Reil_53	Reil_125	Reil_193	Reil_262
Unzer_355	Unzer_440	Reil_55	Reil_126	Reil_194	Reil_263
Unzer_356	Unzer_441	Reil_56	Reil_127	Reil_195	Reil_264
Unzer_357	Unzer_443	Reil_57	Reil_128	Reil_196	Reil_265

Reil_266	Reil_334	Reil_400	Reil_469	Reil_539	Reil_606
Reil_267	Reil_335	Reil_401	Reil_470	Reil_540	Reil_607
Reil_268	Reil_336	Reil_402	Reil_471	Reil_541	Reil_608
Reil_269	Reil_337	Reil_403	Reil_472	Reil_542	Reil_609
Reil_270	Reil_338	Reil_404	Reil_473	Reil_543	Reil_610
Reil_271	Reil_339	Reil_405	Reil_474	Reil_544	Reil_611
Reil_272	Reil_340	Reil_406	Reil_475	Reil_546	Reil_612
Reil_273	Reil_341	Reil_407	Reil_476	Reil_547	Reil_613
Reil_274	Reil_342	Reil_408	Reil_477	Reil_548	Reil_614
Reil_275	Reil_343	Reil_409	Reil_479	Reil_549	Reil_615
Reil_276	Reil_344	Reil_410	Reil_480	Reil_550	Reil_616
Reil_277	Reil_345	Reil_411	Reil_481	Reil_551	Reil_617
Reil_278	Reil_346	Reil_412	Reil_482	Reil_552	Reil_618
Reil_279	Reil_347	Reil_413	Reil_483	Reil_553	Reil_619
Reil_280	Reil_348	Reil_415	Reil_484	Reil_554	Reil_621
Reil_281	Reil_349	Reil_416	Reil_486	Reil_555	Reil_622
Reil_282	Reil_350	Reil_417	Reil_487	Reil_557	Reil_623
Reil_283	Reil_351	Reil_418	Reil_488	Reil_558	Reil_624
Reil_285	Reil_352	Reil_419	Reil_489	Reil_559	Reil_625
Reil_286	Reil_353	Reil_420	Reil_490	Reil_560	Reil_626
Reil_287	Reil_354	Reil_421	Reil_491	Reil_561	Reil_627
Reil_288	Reil_355	Reil_422	Reil_492	Reil_562	Reil_628
Reil_289	Reil_356	Reil_423	Reil_493	Reil_563	Reil_629
Reil_290	Reil_357	Reil_425	Reil_494	Reil_564	Reil_630
Reil_291	Reil_358	Reil_426	Reil_495	Reil_565	Reil_631
Reil_293	Reil_359	Reil_427	Reil_496	Reil_566	Reil_632
Reil_294	Reil_360	Reil_428	Reil_497	Reil_567	Reil_633
Reil_295	Reil_361	Reil_429	Reil_498	Reil_568	Reil_634
Reil_296	Reil_362	Reil_430	Reil_499	Reil_569	Reil_635
Reil_297	Reil_363	Reil_431	Reil_500	Reil_570	Reil_636
Reil_298	Reil_364	Reil_432	Reil_501	Reil_571	Reil_637
Reil_299	Reil_366	Reil_433	Reil_502	Reil_572	Reil_638
Reil_300	Reil_367	Reil_434	Reil_503	Reil_573	Reil_639
Reil_301	Reil_368	Reil_435	Reil_504	Reil_574	Reil_640
Reil_302	Reil_369	Reil_436	Reil_506	Reil_575	Reil_641
Reil_303	Reil_370	Reil_437	Reil_507	Reil_576	Reil_642
Reil_304	Reil_371	Reil_438	Reil_508	Reil_577	Reil_643
Reil_305	Reil_372	Reil_439	Reil_509	Reil_578	Reil_644
Reil_306	Reil_373	Reil_440	Reil_510	Reil_579	Reil_645
Reil_307	Reil_374	Reil_441	Reil_511	Reil_580	Reil_646
Reil_308	Reil_375	Reil_442	Reil_512	Reil_581	Reil_647
Reil_309	Reil_376	Reil_443	Reil_514	Reil_582	Reil_648
Reil_310	Reil_377	Reil_444	Reil_515	Reil_583	Reil_649
Reil_311	Reil_378	Reil_445	Reil_516	Reil_584	Reil_650
Reil_312	Reil_379	Reil_446	Reil_517	Reil_585	Reil_651
Reil_313	Reil_380	Reil_447	Reil_518	Reil_586	Reil_653
Reil_315	Reil_381	Reil_448	Reil_519	Reil_587	Reil_654
Reil_316	Reil_382	Reil_449	Reil_520	Reil_588	Reil_655
Reil_317	Reil_383	Reil_451	Reil_522	Reil_589	Reil_656
Reil_318	Reil_384	Reil_453	Reil_523	Reil_590	Reil_657
Reil_319	Reil_385	Reil_454	Reil_524	Reil_591	Reil_658
Reil_320	Reil_386	Reil_455	Reil_525	Reil_592	Reil_659
Reil_321	Reil_387	Reil_456	Reil_526	Reil_593	Reil_660
Reil_322	Reil_388	Reil_457	Reil_527	Reil_594	Reil_661
Reil_323	Reil_389	Reil_458	Reil_528	Reil_595	Reil_662
Reil_324	Reil_390	Reil_459	Reil_529	Reil_596	Reil_663
Reil_325	Reil_391	Reil_460	Reil_530	Reil_597	Reil_664
Reil_326	Reil_392	Reil_461	Reil_531	Reil_598	Reil_665
Reil_327	Reil_393	Reil_462	Reil_532	Reil_599	Reil_666
Reil_328	Reil_394	Reil_463	Reil_533	Reil_600	Reil_667
Reil_329	Reil_395	Reil_464	Reil_534	Reil_601	Reil_668
Reil_330	Reil_396	Reil_465	Reil_535	Reil_602	Reil_669
Reil_331	Reil_397	Reil_466	Reil_536	Reil_603	Reil_670
Reil_332	Reil_398	Reil_467	Reil_537	Reil_604	Reil_671
Reil_333	Reil_399	Reil_468	Reil_538	Reil_605	Reil_672

Reil_673	Reil_741	Reil_809	Reil_879	Reil_949	Reil_1026
Reil_674	Reil_742	Reil_810	Reil_880	Reil_950	Reil_1027
Reil_675	Reil_743	Reil_811	Reil_881	Reil_951	Reil_1028
Reil_676	Reil_744	Reil_812	Reil_883	Reil_952	Reil_1029
Reil_677	Reil_745	Reil_813	Reil_884	Reil_953	Reil_1030
Reil_678	Reil_746	Reil_815	Reil_885	Reil_954	Reil_1031
Reil_679	Reil_747	Reil_816	Reil_886	Reil_955	Reil_1032
Reil_680	Reil_748	Reil_817	Reil_888	Reil_956	Reil_1033
Reil_681	Reil_749	Reil_818	Reil_889	Reil_957	Reil_1034
Reil_682	Reil_750	Reil_819	Reil_890	Reil_959	Reil_1035
Reil_683	Reil_752	Reil_820	Reil_891	Reil_960	Reil_1036
Reil_684	Reil_753	Reil_821	Reil_892	Reil_961	Reil_1037
Reil_685	Reil_754	Reil_823	Reil_893	Reil_962	Reil_1038
Reil_686	Reil_755	Reil_824	Reil_894	Reil_963	Reil_1039
Reil_687	Reil_756	Reil_825	Reil_895	Reil_964	Reil_1040
Reil_688	Reil_758	Reil_826	Reil_896	Reil_965	Reil_1041
Reil_689	Reil_759	Reil_827	Reil_897	Reil_967	Reil_1042
Reil_690	Reil_760	Reil_828	Reil_898	Reil_968	Reil_1043
Reil_691	Reil_761	Reil_829	Reil_899	Reil_969	Reil_1044
Reil_692	Reil_762	Reil_830	Reil_901	Reil_970	Reil_1046
Reil_693	Reil_763	Reil_831	Reil_902	Reil_971	Reil_1047
Reil_694	Reil_764	Reil_832	Reil_903	Reil_972	Reil_1048
Reil_695	Reil_765	Reil_833	Reil_904	Reil_973	Reil_1049
Reil_696	Reil_766	Reil_834	Reil_905	Reil_974	Reil_1050
Reil_697	Reil_767	Reil_835	Reil_906	Reil_975	Reil_1051
Reil_698	Reil_768	Reil_836	Reil_907	Reil_977	Reil_1053
Reil_699	Reil_769	Reil_839	Reil_908	Reil_979	Reil_1054
Reil_700	Reil_770	Reil_840	Reil_909	Reil_980	Reil_1055
Reil_701	Reil_771	Reil_841	Reil_910	Reil_981	Reil_1057
Reil_702	Reil_772	Reil_842	Reil_912	Reil_983	Reil_1058
Reil_703	Reil_773	Reil_843	Reil_913	Reil_984	Reil_1059
Reil_704	Reil_774	Reil_844	Reil_914	Reil_986	Reil_1061
Reil_706	Reil_775	Reil_845	Reil_915	Reil_987	Reil_1062
Reil_707	Reil_776	Reil_846	Reil_916	Reil_988	Reil_1063
Reil_708	Reil_777	Reil_847	Reil_917	Reil_989	Reil_1064
Reil_709	Reil_778	Reil_848	Reil_918	Reil_990	Reil_1065
Reil_710	Reil_779	Reil_849	Reil_919	Reil_994	Reil_1066
Reil_711	Reil_780	Reil_850	Reil_920	Reil_995	Reil_1067
Reil_712	Reil_781	Reil_851	Reil_921	Reil_996	Reil_1068
Reil_713	Reil_782	Reil_852	Reil_922	Reil_997	Reil_1069
Reil_714	Reil_783	Reil_853	Reil_924	Reil_998	Reil_1070
Reil_716	Reil_784	Reil_854	Reil_925	Reil_999	Reil_1071
Reil_717	Reil_785	Reil_855	Reil_926	Reil_1000	Reil_1073
Reil_718	Reil_786	Reil_856	Reil_927	Reil_1002	Reil_1074
Reil_719	Reil_787	Reil_857	Reil_928	Reil_1003	Reil_1075
Reil_720	Reil_788	Reil_858	Reil_929	Reil_1004	Reil_1076
Reil_721	Reil_789	Reil_860	Reil_930	Reil_1006	Reil_1077
Reil_722	Reil_790	Reil_861	Reil_931	Reil_1007	Reil_1079
Reil_723	Reil_791	Reil_862	Reil_932	Reil_1008	Reil_1080
Reil_724	Reil_792	Reil_863	Reil_933	Reil_1009	Reil_1081
Reil_725	Reil_793	Reil_864	Reil_934	Reil_1010	Reil_1082
Reil_726	Reil_794	Reil_865	Reil_935	Reil_1011	Reil_1083
Reil_727	Reil_795	Reil_866	Reil_936	Reil_1012	Reil_1085
Reil_728	Reil_796	Reil_867	Reil_937	Reil_1013	Reil_1086
Reil_729	Reil_797	Reil_868	Reil_938	Reil_1014	Reil_1087
Reil_730	Reil_798	Reil_869	Reil_939	Reil_1015	Reil_1088
Reil_731	Reil_799	Reil_870	Reil_940	Reil_1016	Reil_1089
Reil_732	Reil_800	Reil_871	Reil_941	Reil_1017	Reil_1090
Reil_734	Reil_801	Reil_872	Reil_942	Reil_1018	Reil_1092
Reil_735	Reil_802	Reil_873	Reil_943	Reil_1019	Reil_1093
Reil_736	Reil_803	Reil_874	Reil_944	Reil_1021	Reil_1094
Reil_737	Reil_804	Reil_875	Reil_945	Reil_1022	Reil_1095
Reil_738	Reil_805	Reil_876	Reil_946	Reil_1023	Reil_1096
Reil_739	Reil_806	Reil_877	Reil_947	Reil_1024	Reil_1098
Reil_740	Reil_807	Reil_878	Reil_948	Reil_1025	Reil_1099

Reil_1100	Reil_1180	Reil_1247	Purmann_41	Purmann_125	Purmann_202
Reil_1101	Reil_1181	Reil_1248	Purmann_42	Purmann_127	Purmann_203
Reil_1102	Reil_1182	Reil_1249	Purmann_43	Purmann_128	Purmann_204
Reil_1104	Reil_1183	Reil_1250	Purmann_45	Purmann_129	Purmann_205
Reil_1105	Reil_1184	Reil_1251	Purmann_48	Purmann_130	Purmann_206
Reil_1107	Reil_1185	Reil_1252	Purmann_49	Purmann_131	Purmann_207
Reil_1108	Reil_1186	Reil_1254	Purmann_55	Purmann_132	Purmann_208
Reil_1109	Reil_1187	Reil_1255	Purmann_56	Purmann_133	Purmann_209
Reil_1111	Reil_1188	Reil_1256	Purmann_59	Purmann_134	Purmann_210
Reil_1112	Reil_1189	Reil_1257	Purmann_62	Purmann_135	Purmann_211
Reil_1115	Reil_1190	Reil_1258	Purmann_63	Purmann_136	Purmann_212
Reil_1116	Reil_1191	Reil_1259	Purmann_65	Purmann_137	Purmann_213
Reil_1117	Reil_1192	Reil_1260	Purmann_68	Purmann_139	Purmann_215
Reil_1119	Reil_1193	Reil_1261	Purmann_70	Purmann_140	Purmann_216
Reil_1121	Reil_1194	Reil_1262	Purmann_71	Purmann_141	Purmann_219
Reil_1122	Reil_1195	Reil_1263	Purmann_72	Purmann_142	Purmann_223
Reil_1123	Reil_1196	Reil_1264	Purmann_73	Purmann_143	Purmann_225
Reil_1124	Reil_1197	Reil_1265	Purmann_74	Purmann_144	Purmann_226
Reil_1125	Reil_1198	Reil_1266	Purmann_75	Purmann_147	Purmann_227
Reil_1126	Reil_1199	Reil_1267	Purmann_77	Purmann_148	Purmann_229
Reil_1127	Reil_1200	Reil_1268	Purmann_78	Purmann_150	Purmann_232
Reil_1128	Reil_1201	Reil_1269	Purmann_79	Purmann_151	Purmann_233
Reil_1129	Reil_1202	Reil_1270	Purmann_80	Purmann_152	Purmann_234
Reil_1131	Reil_1203	Reil_1271	Purmann_81	Purmann_153	Purmann_236
Reil_1133	Reil_1204	Reil_1272	Purmann_82	Purmann_154	Purmann_237
Reil_1134	Reil_1205	Reil_1273	Purmann_83	Purmann_156	Purmann_238
Reil_1135	Reil_1206	Reil_1274	Purmann_84	Purmann_158	Purmann_240
Reil_1136	Reil_1207	Reil_1275	Purmann_85	Purmann_159	Purmann_241
Reil_1137	Reil_1208	Reil_1276	Purmann_86	Purmann_160	Purmann_242
Reil_1138	Reil_1210	Reil_1277	Purmann_87	Purmann_161	Purmann_243
Reil_1139	Reil_1211	Reil_1278	Purmann_88	Purmann_162	Purmann_244
Reil_1140	Reil_1212	Reil_1279	Purmann_89	Purmann_163	Purmann_245
Reil_1141	Reil_1213	Reil_1280	Purmann_90	Purmann_164	Purmann_246
Reil_1142	Reil_1214	Reil_1281	Purmann_91	Purmann_165	Purmann_247
Reil_1143	Reil_1215	Reil_1282	Purmann_92	Purmann_166	Purmann_250
Reil_1144	Reil_1216	Reil_1283	Purmann_93	Purmann_167	Purmann_251
Reil_1145	Reil_1217	Reil_1284	Purmann_94	Purmann_168	Purmann_252
Reil_1146	Reil_1218	Reil_1285	Purmann_95	Purmann_169	Purmann_253
Reil_1147	Reil_1219	Reil_1286	Purmann_96	Purmann_170	Purmann_254
Reil_1148	Reil_1220	Purmann_1	Purmann_97	Purmann_171	Purmann_255
Reil_1149	Reil_1221	Purmann_2	Purmann_98	Purmann_173	Purmann_256
Reil_1150	Reil_1222	Purmann_3	Purmann_99	Purmann_174	Purmann_257
Reil_1151	Reil_1223	Purmann_5	Purmann_100	Purmann_175	Purmann_258
Reil_1153	Reil_1224	Purmann_6	Purmann_101	Purmann_176	Purmann_259
Reil_1154	Reil_1225	Purmann_7	Purmann_102	Purmann_177	Purmann_260
Reil_1155	Reil_1226	Purmann_8	Purmann_103	Purmann_178	Purmann_261
Reil_1156	Reil_1227	Purmann_9	Purmann_104	Purmann_179	Purmann_262
Reil_1157	Reil_1228	Purmann_10	Purmann_105	Purmann_180	Purmann_263
Reil_1158	Reil_1229	Purmann_11	Purmann_106	Purmann_181	Purmann_264
Reil_1159	Reil_1230	Purmann_13	Purmann_107	Purmann_183	Purmann_265
Reil_1160	Reil_1232	Purmann_14	Purmann_109	Purmann_184	Purmann_266
Reil_1161	Reil_1233	Purmann_15	Purmann_111	Purmann_186	Purmann_267
Reil_1162	Reil_1234	Purmann_16	Purmann_112	Purmann_187	Purmann_268
Reil_1163	Reil_1235	Purmann_17	Purmann_113	Purmann_188	Purmann_270
Reil_1165	Reil_1236	Purmann_19	Purmann_114	Purmann_190	Purmann_272
Reil_1166	Reil_1237	Purmann_20	Purmann_115	Purmann_191	Purmann_274
Reil_1169	Reil_1238	Purmann_22	Purmann_116	Purmann_192	Purmann_275
Reil_1171	Reil_1239	Purmann_23	Purmann_117	Purmann_193	Purmann_276
Reil_1172	Reil_1240	Purmann_24	Purmann_118	Purmann_194	Purmann_278
Reil_1174	Reil_1241	Purmann_25	Purmann_119	Purmann_195	Purmann_279
Reil_1175	Reil_1242	Purmann_26	Purmann_120	Purmann_196	Purmann_280
Reil_1176	Reil_1243	Purmann_28	Purmann_121	Purmann_197	Purmann_282
Reil_1177	Reil_1244	Purmann_35	Purmann_122	Purmann_198	Purmann_283
Reil_1178	Reil_1245	Purmann_37	Purmann_123	Purmann_200	Purmann_286
Reil_1179	Reil_1246	Purmann_38	Purmann_124	Purmann_201	Purmann_288

Kraepelin_444	Kraepelin_519	Kraepelin_596	Koch_20	Koch_102	Koch_174
Kraepelin_445	Kraepelin_521	Kraepelin_597	Koch_21	Koch_103	Koch_175
Kraepelin_446	Kraepelin_522	Kraepelin_600	Koch_22	Koch_104	Koch_176
Kraepelin_447	Kraepelin_523	Kraepelin_601	Koch_25	Koch_105	Koch_180
Kraepelin_448	Kraepelin_524	Kraepelin_602	Koch_26	Koch_106	Koch_181
Kraepelin_449	Kraepelin_525	Kraepelin_604	Koch_27	Koch_107	Koch_185
Kraepelin_450	Kraepelin_526	Kraepelin_607	Koch_28	Koch_108	Koch_187
Kraepelin_451	Kraepelin_527	Kraepelin_608	Koch_29	Koch_109	Koch_188
Kraepelin_452	Kraepelin_529	Kraepelin_610	Koch_30	Koch_110	Koch_189
Kraepelin_453	Kraepelin_530	Kraepelin_612	Koch_31	Koch_111	Koch_190
Kraepelin_454	Kraepelin_532	Kraepelin_613	Koch_32	Koch_113	Koch_191
Kraepelin_455	Kraepelin_533	Kraepelin_617	Koch_33	Koch_114	Koch_192
Kraepelin_456	Kraepelin_535	Kraepelin_618	Koch_34	Koch_115	Koch_193
Kraepelin_457	Kraepelin_536	Kraepelin_619	Koch_35	Koch_117	Koch_194
Kraepelin_458	Kraepelin_537	Kraepelin_620	Koch_36	Koch_118	Koch_196
Kraepelin_459	Kraepelin_538	Kraepelin_621	Koch_37	Koch_119	Koch_197
Kraepelin_460	Kraepelin_539	Kraepelin_622	Koch_39	Koch_120	Koch_198
Kraepelin_461	Kraepelin_540	Kraepelin_623	Koch_40	Koch_122	Koch_199
Kraepelin_462	Kraepelin_541	Kraepelin_624	Koch_41	Koch_123	Koch_200
Kraepelin_463	Kraepelin_542	Kraepelin_626	Koch_42	Koch_124	Koch_201
Kraepelin_465	Kraepelin_543	Kraepelin_628	Koch_43	Koch_125	Koch_202
Kraepelin_466	Kraepelin_544	Kraepelin_629	Koch_44	Koch_126	Koch_203
Kraepelin_467	Kraepelin_547	Kraepelin_630	Koch_45	Koch_127	Koch_204
Kraepelin_468	Kraepelin_548	Kraepelin_632	Koch_46	Koch_128	Koch_205
Kraepelin_469	Kraepelin_549	Kraepelin_634	Koch_47	Koch_129	Koch_206
Kraepelin_470	Kraepelin_550	Kraepelin_635	Koch_48	Koch_130	Koch_207
Kraepelin_471	Kraepelin_551	Kraepelin_641	Koch_49	Koch_131	Koch_209
Kraepelin_473	Kraepelin_552	Kraepelin_643	Koch_50	Koch_132	Koch_210
Kraepelin_474	Kraepelin_553	Kraepelin_645	Koch_51	Koch_133	Koch_211
Kraepelin_475	Kraepelin_557	Kraepelin_649	Koch_53	Koch_134	Koch_212
Kraepelin_476	Kraepelin_558	Kraepelin_650	Koch_54	Koch_136	Koch_214
Kraepelin_478	Kraepelin_559	Kraepelin_653	Koch_56	Koch_137	Koch_215
Kraepelin_479	Kraepelin_560	Kraepelin_654	Koch_57	Koch_138	Koch_216
Kraepelin_480	Kraepelin_561	Kraepelin_655	Koch_61	Koch_139	Koch_218
Kraepelin_481	Kraepelin_562	Kraepelin_657	Koch_62	Koch_140	Koch_221
Kraepelin_482	Kraepelin_563	Kraepelin_660	Koch_63	Koch_141	Koch_222
Kraepelin_483	Kraepelin_564	Kraepelin_661	Koch_64	Koch_142	Koch_223
Kraepelin_484	Kraepelin_565	Kraepelin_662	Koch_65	Koch_143	Koch_224
Kraepelin_485	Kraepelin_567	Kraepelin_664	Koch_67	Koch_144	Koch_225
Kraepelin_488	Kraepelin_568	Kraepelin_665	Koch_68	Koch_145	Koch_226
Kraepelin_489	Kraepelin_569	Kraepelin_667	Koch_69	Koch_146	Koch_227
Kraepelin_490	Kraepelin_570	Kraepelin_669	Koch_70	Koch_147	Koch_228
Kraepelin_491	Kraepelin_572	Kraepelin_670	Koch_71	Koch_148	Koch_229
Kraepelin_492	Kraepelin_573	Kraepelin_672	Koch_72	Koch_149	Koch_232
Kraepelin_493	Kraepelin_574	Kraepelin_673	Koch_73	Koch_150	Koch_234
Kraepelin_494	Kraepelin_575	Kraepelin_674	Koch_74	Koch_151	Koch_235
Kraepelin_495	Kraepelin_576	Kraepelin_675	Koch_75	Koch_152	Koch_236
Kraepelin_496	Kraepelin_577	Kraepelin_676	Koch_76	Koch_153	Koch_237
Kraepelin_497	Kraepelin_578	Kraepelin_677	Koch_77	Koch_155	Koch_238
Kraepelin_498	Kraepelin_579	Koch_1	Koch_78	Koch_156	Koch_239
Kraepelin_499	Kraepelin_580	Koch_2	Koch_80	Koch_157	Koch_241
Kraepelin_500	Kraepelin_581	Koch_4	Koch_81	Koch_158	Koch_242
Kraepelin_501	Kraepelin_582	Koch_5	Koch_84	Koch_159	Koch_243
Kraepelin_502	Kraepelin_583	Koch_6	Koch_85	Koch_160	Koch_244
Kraepelin_506	Kraepelin_584	Koch_8	Koch_86	Koch_161	Koch_245
Kraepelin_507	Kraepelin_585	Koch_9	Koch_87	Koch_162	Koch_246
Kraepelin_509	Kraepelin_586	Koch_10	Koch_89	Koch_163	Koch_247
Kraepelin_510	Kraepelin_587	Koch_11	Koch_91	Koch_164	Koch_248
Kraepelin_511	Kraepelin_588	Koch_12	Koch_92	Koch_167	Koch_250
Kraepelin_512	Kraepelin_589	Koch_13	Koch_93	Koch_168	Koch_251
Kraepelin_513	Kraepelin_590	Koch_14	Koch_94	Koch_169	Koch_252
Kraepelin_514	Kraepelin_591	Koch_15	Koch_95	Koch_170	Koch_253
Kraepelin_515	Kraepelin_592	Koch_16	Koch_97	Koch_171	Koch_254
Kraepelin_516	Kraepelin_593	Koch_17	Koch_99	Koch_172	Koch_255
Kraepelin_518	Kraepelin_595	Koch_19	Koch_101	Koch_173	Koch_256

Koch_257	Gall_71	Gall_144	Gall_214	Gall_292	Gall_368
Koch_259	Gall_73	Gall_145	Gall_215	Gall_293	Gall_369
Koch_260	Gall_74	Gall_146	Gall_217	Gall_294	Gall_370
Koch_261	Gall_75	Gall_147	Gall_218	Gall_295	Gall_371
Koch_262	Gall_76	Gall_148	Gall_219	Gall_296	Gall_372
Koch_263	Gall_77	Gall_149	Gall_221	Gall_297	Gall_374
Koch_264	Gall_78	Gall_150	Gall_223	Gall_299	Gall_375
Gall_1	Gall_79	Gall_151	Gall_228	Gall_300	Gall_376
Gall_2	Gall_80	Gall_152	Gall_229	Gall_301	Gall_377
Gall_4	Gall_81	Gall_153	Gall_230	Gall_302	Gall_378
Gall_5	Gall_82	Gall_154	Gall_232	Gall_304	Gall_379
Gall_6	Gall_83	Gall_155	Gall_233	Gall_305	Gall_380
Gall_7	Gall_84	Gall_156	Gall_234	Gall_307	Gall_381
Gall_9	Gall_85	Gall_157	Gall_235	Gall_308	Gall_382
Gall_11	Gall_86	Gall_158	Gall_236	Gall_309	Gall_383
Gall_12	Gall_87	Gall_159	Gall_237	Gall_310	Gall_384
Gall_13	Gall_88	Gall_160	Gall_238	Gall_311	Gall_385
Gall_14	Gall_90	Gall_161	Gall_239	Gall_313	Gall_386
Gall_15	Gall_91	Gall_162	Gall_240	Gall_314	Gall_387
Gall_16	Gall_94	Gall_163	Gall_241	Gall_315	Gall_388
Gall_17	Gall_95	Gall_165	Gall_242	Gall_316	Gall_389
Gall_19	Gall_96	Gall_166	Gall_243	Gall_319	Gall_390
Gall_20	Gall_97	Gall_168	Gall_244	Gall_320	Gall_392
Gall_21	Gall_98	Gall_169	Gall_246	Gall_322	Gall_393
Gall_22	Gall_99	Gall_170	Gall_247	Gall_323	Gall_394
Gall_23	Gall_100	Gall_171	Gall_248	Gall_324	Gall_395
Gall_24	Gall_101	Gall_172	Gall_249	Gall_325	Gall_396
Gall_25	Gall_102	Gall_173	Gall_250	Gall_326	Gall_397
Gall_26	Gall_103	Gall_176	Gall_251	Gall_327	Gall_398
Gall_27	Gall_104	Gall_177	Gall_252	Gall_328	Gall_399
Gall_28	Gall_105	Gall_178	Gall_253	Gall_329	Gall_400
Gall_29	Gall_106	Gall_179	Gall_255	Gall_330	Gall_401
Gall_30	Gall_107	Gall_180	Gall_256	Gall_331	Gall_402
Gall_31	Gall_108	Gall_181	Gall_257	Gall_332	Gall_403
Gall_32	Gall_109	Gall_183	Gall_258	Gall_333	Gall_404
Gall_33	Gall_110	Gall_184	Gall_259	Gall_334	Gall_405
Gall_34	Gall_111	Gall_185	Gall_260	Gall_335	Gall_406
Gall_36	Gall_112	Gall_186	Gall_261	Gall_336	Gall_407
Gall_37	Gall_113	Gall_187	Gall_262	Gall_338	Gall_408
Gall_39	Gall_114	Gall_188	Gall_263	Gall_339	Gall_409
Gall_40	Gall_115	Gall_189	Gall_264	Gall_340	Gall_410
Gall_42	Gall_116	Gall_190	Gall_265	Gall_341	Gall_413
Gall_43	Gall_117	Gall_191	Gall_266	Gall_343	Gall_414
Gall_44	Gall_118	Gall_192	Gall_267	Gall_344	Gall_415
Gall_45	Gall_119	Gall_193	Gall_268	Gall_345	Gall_416
Gall_46	Gall_120	Gall_194	Gall_270	Gall_346	Gall_417
Gall_47	Gall_121	Gall_195	Gall_271	Gall_347	Gall_418
Gall_48	Gall_122	Gall_196	Gall_272	Gall_348	Gall_419
Gall_49	Gall_123	Gall_197	Gall_273	Gall_349	Gall_420
Gall_50	Gall_124	Gall_198	Gall_274	Gall_350	Gall_422
Gall_52	Gall_125	Gall_199	Gall_275	Gall_351	Gall_423
Gall_53	Gall_126	Gall_200	Gall_276	Gall_352	Gall_424
Gall_54	Gall_127	Gall_201	Gall_279	Gall_353	Gall_425
Gall_55	Gall_130	Gall_202	Gall_280	Gall_354	Gall_426
Gall_56	Gall_133	Gall_203	Gall_281	Gall_355	Gall_428
Gall_57	Gall_134	Gall_204	Gall_282	Gall_356	Gall_429
Gall_59	Gall_135	Gall_205	Gall_283	Gall_357	Gall_430
Gall_62	Gall_136	Gall_206	Gall_284	Gall_359	Gall_431
Gall_63	Gall_137	Gall_207	Gall_285	Gall_360	Gall_432
Gall_65	Gall_138	Gall_208	Gall_286	Gall_361	Gall_433
Gall_66	Gall_139	Gall_209	Gall_287	Gall_362	Gall_434
Gall_67	Gall_140	Gall_210	Gall_288	Gall_364	Gall_435
Gall_68	Gall_141	Gall_211	Gall_289	Gall_365	Gall_436
Gall_69	Gall_142	Gall_212	Gall_290	Gall_366	Gall_437
Gall_70	Gall_143	Gall_213	Gall_291	Gall_367	Gall_438

Gall_439	Gall_517	Gall_593	Abel_58	Abel_144	Abel_221
Gall_440	Gall_518	Gall_595	Abel_59	Abel_145	Abel_222
Gall_441	Gall_519	Gall_596	Abel_61	Abel_146	Abel_223
Gall_442	Gall_520	Gall_597	Abel_62	Abel_147	Abel_224
Gall_443	Gall_521	Gall_598	Abel_64	Abel_148	Abel_226
Gall_445	Gall_522	Gall_599	Abel_65	Abel_149	Abel_227
Gall_447	Gall_523	Gall_601	Abel_66	Abel_150	Abel_228
Gall_448	Gall_524	Gall_602	Abel_67	Abel_151	Abel_229
Gall_450	Gall_525	Gall_604	Abel_68	Abel_152	Abel_230
Gall_451	Gall_526	Gall_605	Abel_69	Abel_154	Abel_231
Gall_452	Gall_527	Gall_606	Abel_70	Abel_155	Abel_232
Gall_453	Gall_528	Gall_607	Abel_71	Abel_156	Abel_234
Gall_454	Gall_529	Gall_609	Abel_72	Abel_158	Abel_236
Gall_455	Gall_530	Gall_610	Abel_73	Abel_159	Abel_237
Gall_456	Gall_531	Gall_611	Abel_75	Abel_160	Abel_238
Gall_458	Gall_533	Gall_612	Abel_76	Abel_161	Abel_239
Gall_459	Gall_534	Gall_613	Abel_77	Abel_163	Abel_240
Gall_460	Gall_535	Gall_614	Abel_78	Abel_164	Abel_242
Gall_461	Gall_537	Gall_615	Abel_81	Abel_167	Abel_243
Gall_462	Gall_538	Gall_616	Abel_83	Abel_168	Abel_244
Gall_464	Gall_539	Gall_617	Abel_84	Abel_170	Abel_245
Gall_465	Gall_540	Gall_618	Abel_86	Abel_171	Abel_246
Gall_467	Gall_541	Gall_619	Abel_87	Abel_172	Abel_247
Gall_468	Gall_542	Gall_620	Abel_88	Abel_173	Abel_248
Gall_470	Gall_543	Gall_621	Abel_91	Abel_174	Abel_249
Gall_471	Gall_544	Abel_1	Abel_94	Abel_175	Abel_250
Gall_472	Gall_545	Abel_2	Abel_95	Abel_176	Abel_251
Gall_473	Gall_546	Abel_3	Abel_96	Abel_177	Abel_252
Gall_474	Gall_547	Abel_4	Abel_97	Abel_178	Abel_254
Gall_475	Gall_548	Abel_6	Abel_98	Abel_180	Abel_255
Gall_477	Gall_549	Abel_7	Abel_100	Abel_181	Abel_256
Gall_478	Gall_550	Abel_9	Abel_101	Abel_182	Abel_257
Gall_479	Gall_551	Abel_10	Abel_102	Abel_183	Abel_258
Gall_481	Gall_552	Abel_12	Abel_103	Abel_184	Abel_259
Gall_482	Gall_553	Abel_13	Abel_104	Abel_185	Abel_260
Gall_483	Gall_554	Abel_14	Abel_105	Abel_187	Abel_261
Gall_484	Gall_555	Abel_16	Abel_107	Abel_188	Abel_262
Gall_485	Gall_556	Abel_17	Abel_108	Abel_189	Abel_263
Gall_486	Gall_557	Abel_18	Abel_109	Abel_190	Abel_265
Gall_487	Gall_558	Abel_19	Abel_111	Abel_191	Abel_267
Gall_488	Gall_559	Abel_21	Abel_112	Abel_192	Abel_269
Gall_489	Gall_560	Abel_22	Abel_113	Abel_193	Abel_271
Gall_490	Gall_561	Abel_24	Abel_114	Abel_194	Abel_272
Gall_491	Gall_563	Abel_25	Abel_115	Abel_195	Abel_273
Gall_492	Gall_565	Abel_26	Abel_116	Abel_196	Abel_274
Gall_493	Gall_566	Abel_27	Abel_118	Abel_197	Abel_275
Gall_494	Gall_567	Abel_28	Abel_119	Abel_198	Abel_276
Gall_495	Gall_568	Abel_29	Abel_121	Abel_199	Abel_277
Gall_496	Gall_571	Abel_30	Abel_122	Abel_200	Abel_280
Gall_497	Gall_572	Abel_31	Abel_123	Abel_201	Abel_281
Gall_499	Gall_573	Abel_34	Abel_124	Abel_202	Abel_283
Gall_500	Gall_575	Abel_35	Abel_125	Abel_203	Abel_286
Gall_501	Gall_577	Abel_38	Abel_127	Abel_204	Abel_287
Gall_502	Gall_580	Abel_40	Abel_128	Abel_205	Abel_288
Gall_503	Gall_581	Abel_42	Abel_129	Abel_207	Abel_290
Gall_504	Gall_582	Abel_44	Abel_130	Abel_208	Abel_291
Gall_505	Gall_583	Abel_45	Abel_132	Abel_209	Abel_292
Gall_506	Gall_584	Abel_46	Abel_133	Abel_210	Abel_293
Gall_508	Gall_585	Abel_48	Abel_135	Abel_211	Abel_294
Gall_509	Gall_586	Abel_49	Abel_136	Abel_212	Abel_296
Gall_511	Gall_587	Abel_50	Abel_138	Abel_214	Abel_297
Gall_512	Gall_588	Abel_53	Abel_139	Abel_215	Abel_298
Gall_514	Gall_589	Abel_55	Abel_140	Abel_217	Abel_299
Gall_515	Gall_591	Abel_56	Abel_141	Abel_218	Abel_300
Gall_516	Gall_592	Abel_57	Abel_143	Abel_220	Abel_304

Abel_306	Abel_407	Abel_497	Braeuner_47	Braeuner_153	Braeuner_265
Abel_307	Abel_408	Abel_499	Braeuner_48	Braeuner_155	Braeuner_266
Abel_308	Abel_409	Abel_500	Braeuner_49	Braeuner_156	Braeuner_268
Abel_312	Abel_410	Abel_501	Braeuner_50	Braeuner_157	Braeuner_269
Abel_313	Abel_413	Abel_502	Braeuner_51	Braeuner_158	Braeuner_270
Abel_314	Abel_414	Abel_503	Braeuner_57	Braeuner_159	Braeuner_271
Abel_316	Abel_416	Abel_505	Braeuner_58	Braeuner_160	Braeuner_272
Abel_317	Abel_418	Abel_507	Braeuner_59	Braeuner_163	Braeuner_273
Abel_318	Abel_419	Abel_509	Braeuner_61	Braeuner_164	Braeuner_274
Abel_319	Abel_420	Abel_510	Braeuner_62	Braeuner_165	Braeuner_275
Abel_320	Abel_421	Abel_511	Braeuner_63	Braeuner_172	Braeuner_276
Abel_322	Abel_422	Abel_512	Braeuner_68	Braeuner_173	Braeuner_277
Abel_323	Abel_424	Abel_513	Braeuner_69	Braeuner_174	Braeuner_278
Abel_324	Abel_425	Abel_514	Braeuner_70	Braeuner_175	Braeuner_279
Abel_327	Abel_426	Abel_515	Braeuner_71	Braeuner_183	Braeuner_280
Abel_328	Abel_433	Abel_516	Braeuner_72	Braeuner_184	Braeuner_281
Abel_331	Abel_434	Abel_517	Braeuner_73	Braeuner_185	Braeuner_282
Abel_335	Abel_435	Abel_518	Braeuner_74	Braeuner_189	Braeuner_283
Abel_336	Abel_436	Abel_519	Braeuner_75	Braeuner_190	Braeuner_284
Abel_337	Abel_437	Abel_521	Braeuner_77	Braeuner_192	Braeuner_285
Abel_338	Abel_438	Abel_522	Braeuner_79	Braeuner_195	Braeuner_286
Abel_339	Abel_439	Abel_523	Braeuner_80	Braeuner_196	Braeuner_287
Abel_341	Abel_440	Abel_524	Braeuner_82	Braeuner_197	Braeuner_288
Abel_342	Abel_441	Abel_525	Braeuner_83	Braeuner_199	Braeuner_290
Abel_344	Abel_443	Abel_526	Braeuner_85	Braeuner_201	Braeuner_292
Abel_345	Abel_444	Abel_527	Braeuner_88	Braeuner_202	Braeuner_294
Abel_348	Abel_445	Abel_528	Braeuner_92	Braeuner_203	Braeuner_295
Abel_351	Abel_446	Abel_530	Braeuner_93	Braeuner_205	Braeuner_296
Abel_353	Abel_447	Abel_531	Braeuner_94	Braeuner_206	Braeuner_297
Abel_354	Abel_448	Abel_532	Braeuner_95	Braeuner_207	Braeuner_299
Abel_357	Abel_449	Braeuner_1	Braeuner_96	Braeuner_208	Braeuner_301
Abel_358	Abel_450	Braeuner_2	Braeuner_100	Braeuner_209	Braeuner_302
Abel_359	Abel_452	Braeuner_3	Braeuner_102	Braeuner_212	Braeuner_303
Abel_360	Abel_453	Braeuner_4	Braeuner_103	Braeuner_217	Braeuner_304
Abel_361	Abel_454	Braeuner_5	Braeuner_104	Braeuner_220	Braeuner_305
Abel_363	Abel_455	Braeuner_6	Braeuner_105	Braeuner_221	Braeuner_306
Abel_364	Abel_456	Braeuner_7	Braeuner_107	Braeuner_223	Braeuner_307
Abel_366	Abel_457	Braeuner_8	Braeuner_108	Braeuner_224	Braeuner_308
Abel_367	Abel_458	Braeuner_9	Braeuner_110	Braeuner_225	Braeuner_309
Abel_368	Abel_459	Braeuner_10	Braeuner_111	Braeuner_227	Braeuner_310
Abel_370	Abel_460	Braeuner_11	Braeuner_112	Braeuner_229	Braeuner_311
Abel_374	Abel_461	Braeuner_13	Braeuner_113	Braeuner_230	Braeuner_314
Abel_377	Abel_463	Braeuner_14	Braeuner_116	Braeuner_231	Braeuner_318
Abel_379	Abel_464	Braeuner_16	Braeuner_117	Braeuner_232	Braeuner_322
Abel_380	Abel_465	Braeuner_17	Braeuner_118	Braeuner_233	Braeuner_323
Abel_381	Abel_467	Braeuner_18	Braeuner_119	Braeuner_234	Braeuner_324
Abel_382	Abel_468	Braeuner_19	Braeuner_120	Braeuner_236	Braeuner_325
Abel_384	Abel_469	Braeuner_20	Braeuner_122	Braeuner_237	Braeuner_328
Abel_385	Abel_470	Braeuner_21	Braeuner_123	Braeuner_238	Braeuner_329
Abel_386	Abel_471	Braeuner_23	Braeuner_124	Braeuner_239	Braeuner_330
Abel_388	Abel_473	Braeuner_24	Braeuner_125	Braeuner_240	Braeuner_333
Abel_389	Abel_477	Braeuner_25	Braeuner_126	Braeuner_244	Braeuner_334
Abel_390	Abel_479	Braeuner_27	Braeuner_129	Braeuner_245	Braeuner_336
Abel_391	Abel_480	Braeuner_28	Braeuner_130	Braeuner_248	Braeuner_337
Abel_393	Abel_482	Braeuner_30	Braeuner_137	Braeuner_249	Braeuner_338
Abel_394	Abel_485	Braeuner_32	Braeuner_138	Braeuner_250	Braeuner_339
Abel_395	Abel_486	Braeuner_34	Braeuner_139	Braeuner_252	Braeuner_340
Abel_397	Abel_488	Braeuner_35	Braeuner_140	Braeuner_254	Braeuner_341
Abel_398	Abel_489	Braeuner_36	Braeuner_141	Braeuner_256	Braeuner_342
Abel_400	Abel_490	Braeuner_37	Braeuner_142	Braeuner_257	Braeuner_345
Abel_401	Abel_491	Braeuner_38	Braeuner_143	Braeuner_259	Braeuner_346
Abel_402	Abel_492	Braeuner_39	Braeuner_144	Braeuner_261	Braeuner_347
Abel_403	Abel_494	Braeuner_40	Braeuner_145	Braeuner_262	Braeuner_349
Abel_405	Abel_495	Braeuner_43	Braeuner_147	Braeuner_263	Braeuner_350
Abel_406	Abel_496	Braeuner_44	Braeuner_151	Braeuner_264	Braeuner_351

Braeuner_352	Braeuner_441	Blumenbach_	Ludwig_83	Ludwig_166	Ludwig_240
Braeuner_354	Braeuner_442	449	Ludwig_84	Ludwig_167	Ludwig_241
Braeuner_355	Braeuner_443	Blumenbach_	Ludwig_85	Ludwig_168	Ludwig_242
Braeuner_359	Braeuner_444	557	Ludwig_86	Ludwig_169	Ludwig_244
Braeuner_361	Braeuner_445	Ludwig_1	Ludwig_87	Ludwig_170	Ludwig_245
Braeuner_362	Braeuner_446	Ludwig_2	Ludwig_88	Ludwig_171	Ludwig_246
Braeuner_363	Braeuner_447	Ludwig_3	Ludwig_90	Ludwig_172	Ludwig_247
Braeuner_365	Braeuner_448	Ludwig_4	Ludwig_91	Ludwig_173	Ludwig_248
Braeuner_366	Braeuner_449	Ludwig_5	Ludwig_93	Ludwig_174	Ludwig_249
Braeuner_367	Braeuner_450	Ludwig_6	Ludwig_94	Ludwig_175	Ludwig_250
Braeuner_368	Braeuner_452	Ludwig_7	Ludwig_95	Ludwig_177	Ludwig_251
Braeuner_372	Braeuner_453	Ludwig_8	Ludwig_96	Ludwig_178	Ludwig_252
Braeuner_373	Braeuner_454	Ludwig_9	Ludwig_97	Ludwig_179	Ludwig_253
Braeuner_374	Braeuner_455	Ludwig_10	Ludwig_100	Ludwig_181	Ludwig_254
Braeuner_375	Braeuner_456	Ludwig_11	Ludwig_101	Ludwig_182	Ludwig_255
Braeuner_376	Braeuner_457	Ludwig_12	Ludwig_102	Ludwig_183	Ludwig_256
Braeuner_377	Braeuner_458	Ludwig_13	Ludwig_103	Ludwig_184	Ludwig_257
Braeuner_378	Braeuner_459	Ludwig_14	Ludwig_104	Ludwig_185	Ludwig_258
Braeuner_379	Braeuner_460	Ludwig_15	Ludwig_105	Ludwig_186	Ludwig_260
Braeuner_380	Braeuner_461	Ludwig_16	Ludwig_107	Ludwig_187	Ludwig_261
Braeuner_381	Braeuner_464	Ludwig_17	Ludwig_108	Ludwig_189	Ludwig_263
Braeuner_382	Braeuner_466	Ludwig_18	Ludwig_109	Ludwig_190	Ludwig_264
Braeuner_383	Braeuner_467	Ludwig_20	Ludwig_111	Ludwig_191	Ludwig_265
Braeuner_384	Braeuner_468	Ludwig_22	Ludwig_113	Ludwig_192	Ludwig_266
Braeuner_385	Braeuner_469	Ludwig_23	Ludwig_114	Ludwig_193	Ludwig_270
Braeuner_386	Braeuner_470	Ludwig_24	Ludwig_115	Ludwig_194	Ludwig_271
Braeuner_387	Braeuner_472	Ludwig_25	Ludwig_117	Ludwig_195	Ludwig_272
Braeuner_388	Braeuner_473	Ludwig_27	Ludwig_119	Ludwig_196	Ludwig_273
Braeuner_389	Braeuner_474	Ludwig_29	Ludwig_120	Ludwig_197	Ludwig_276
Braeuner_390	Braeuner_475	Ludwig_31	Ludwig_121	Ludwig_198	Ludwig_279
Braeuner_391	Braeuner_476	Ludwig_32	Ludwig_122	Ludwig_199	Ludwig_280
Braeuner_392	Braeuner_477	Ludwig_34	Ludwig_126	Ludwig_200	Ludwig_283
Braeuner_393	Braeuner_478	Ludwig_35	Ludwig_127	Ludwig_201	Ludwig_284
Braeuner_394	Braeuner_479	Ludwig_36	Ludwig_128	Ludwig_202	Ludwig_285
Braeuner_395	Braeuner_480	Ludwig_38	Ludwig_130	Ludwig_203	Ludwig_286
Braeuner_399	Braeuner_481	Ludwig_39	Ludwig_131	Ludwig_204	Ludwig_287
Braeuner_401	Braeuner_482	Ludwig_42	Ludwig_132	Ludwig_206	Ludwig_289
Braeuner_402	Braeuner_483	Ludwig_43	Ludwig_133	Ludwig_207	Ludwig_290
Braeuner_403	Braeuner_484	Ludwig_44	Ludwig_134	Ludwig_208	Ludwig_291
Braeuner_404	Braeuner_485	Ludwig_45	Ludwig_135	Ludwig_209	Ludwig_292
Braeuner_405	Braeuner_486	Ludwig_46	Ludwig_137	Ludwig_210	Ludwig_293
Braeuner_409	Braeuner_487	Ludwig_47	Ludwig_138	Ludwig_212	Ludwig_294
Braeuner_411	Braeuner_488	Ludwig_48	Ludwig_139	Ludwig_213	Ludwig_295
Braeuner_412	Braeuner_489	Ludwig_49	Ludwig_140	Ludwig_214	Ludwig_296
Braeuner_413	Braeuner_490	Ludwig_50	Ludwig_141	Ludwig_215	Ludwig_297
Braeuner_414	Braeuner_491	Ludwig_51	Ludwig_142	Ludwig_216	Ludwig_298
Braeuner_416	Braeuner_492	Ludwig_52	Ludwig_144	Ludwig_217	Ludwig_299
Braeuner_417	Braeuner_493	Ludwig_54	Ludwig_145	Ludwig_219	Ludwig_300
Braeuner_419	Braeuner_494	Ludwig_58	Ludwig_146	Ludwig_220	Ludwig_301
Braeuner_420	Braeuner_495	Ludwig_59	Ludwig_147	Ludwig_221	Ludwig_302
Braeuner_421	Braeuner_496	Ludwig_60	Ludwig_149	Ludwig_222	Ludwig_303
Braeuner_424	Braeuner_497	Ludwig_62	Ludwig_150	Ludwig_223	Ludwig_304
Braeuner_425	Braeuner_498	Ludwig_64	Ludwig_151	Ludwig_224	Ludwig_305
Braeuner_426	Braeuner_499	Ludwig_66	Ludwig_152	Ludwig_225	Ludwig_306
Braeuner_427	Braeuner_500	Ludwig_67	Ludwig_153	Ludwig_226	Ludwig_307
Braeuner_428	Braeuner_501	Ludwig_68	Ludwig_154	Ludwig_227	Ludwig_308
Braeuner_429	Braeuner_503	Ludwig_69	Ludwig_155	Ludwig_229	Ludwig_309
Braeuner_430	Braeuner_504	Ludwig_72	Ludwig_156	Ludwig_230	Ludwig_310
Braeuner_431	Braeuner_506	Ludwig_75	Ludwig_157	Ludwig_231	Ludwig_311
Braeuner_433	Braeuner_507	Ludwig_76	Ludwig_158	Ludwig_233	Ludwig_312
Braeuner_434	Braeuner_508	Ludwig_77	Ludwig_159	Ludwig_234	Ludwig_313
Braeuner_435	Braeuner_509	Ludwig_79	Ludwig_161	Ludwig_235	Ludwig_314
Braeuner_436	Braeuner_510	Ludwig_80	Ludwig_162	Ludwig_236	Ludwig_315
Braeuner_437	Braeuner_511	Ludwig_81	Ludwig_164	Ludwig_237	Ludwig_316
Braeuner_439	Braeuner_512	Ludwig_82	Ludwig_165	Ludwig_239	Ludwig_317

Ludwig_318	Ludwig_396	Ludwig_477	Ludwig_559	Ludwig_640	Ludwig_735
Ludwig_319	Ludwig_397	Ludwig_478	Ludwig_560	Ludwig_641	Ludwig_736
Ludwig_320	Ludwig_398	Ludwig_479	Ludwig_561	Ludwig_642	Ludwig_739
Ludwig_321	Ludwig_399	Ludwig_480	Ludwig_562	Ludwig_643	Ludwig_740
Ludwig_323	Ludwig_400	Ludwig_481	Ludwig_563	Ludwig_644	Ludwig_741
Ludwig_324	Ludwig_402	Ludwig_482	Ludwig_564	Ludwig_645	Ludwig_742
Ludwig_325	Ludwig_403	Ludwig_483	Ludwig_565	Ludwig_647	Ludwig_744
Ludwig_326	Ludwig_404	Ludwig_484	Ludwig_566	Ludwig_649	Ludwig_745
Ludwig_329	Ludwig_405	Ludwig_486	Ludwig_567	Ludwig_650	Ludwig_747
Ludwig_331	Ludwig_406	Ludwig_487	Ludwig_568	Ludwig_652	Ludwig_748
Ludwig_333	Ludwig_408	Ludwig_488	Ludwig_570	Ludwig_654	Ludwig_749
Ludwig_334	Ludwig_409	Ludwig_489	Ludwig_571	Ludwig_655	Ludwig_750
Ludwig_335	Ludwig_410	Ludwig_490	Ludwig_574	Ludwig_656	Ludwig_751
Ludwig_336	Ludwig_411	Ludwig_491	Ludwig_575	Ludwig_657	Ludwig_752
Ludwig_337	Ludwig_412	Ludwig_492	Ludwig_576	Ludwig_658	Ludwig_753
Ludwig_338	Ludwig_413	Ludwig_493	Ludwig_579	Ludwig_659	Ludwig_754
Ludwig_339	Ludwig_414	Ludwig_495	Ludwig_580	Ludwig_661	Ludwig_755
Ludwig_341	Ludwig_415	Ludwig_496	Ludwig_581	Ludwig_662	Ludwig_756
Ludwig_343	Ludwig_417	Ludwig_497	Ludwig_582	Ludwig_663	Ludwig_757
Ludwig_344	Ludwig_418	Ludwig_498	Ludwig_583	Ludwig_664	Ludwig_758
Ludwig_345	Ludwig_419	Ludwig_499	Ludwig_584	Ludwig_665	Ludwig_759
Ludwig_346	Ludwig_420	Ludwig_500	Ludwig_586	Ludwig_666	Ludwig_760
Ludwig_347	Ludwig_422	Ludwig_502	Ludwig_588	Ludwig_670	Ludwig_762
Ludwig_348	Ludwig_424	Ludwig_503	Ludwig_589	Ludwig_671	Ludwig_763
Ludwig_349	Ludwig_426	Ludwig_508	Ludwig_590	Ludwig_673	Ludwig_765
Ludwig_350	Ludwig_428	Ludwig_511	Ludwig_591	Ludwig_674	Ludwig_766
Ludwig_351	Ludwig_429	Ludwig_512	Ludwig_594	Ludwig_675	Ludwig_767
Ludwig_352	Ludwig_430	Ludwig_513	Ludwig_595	Ludwig_677	Ludwig_768
Ludwig_353	Ludwig_431	Ludwig_514	Ludwig_596	Ludwig_679	Ludwig_769
Ludwig_354	Ludwig_432	Ludwig_517	Ludwig_597	Ludwig_680	Ludwig_772
Ludwig_355	Ludwig_433	Ludwig_518	Ludwig_598	Ludwig_682	Ludwig_773
Ludwig_356	Ludwig_436	Ludwig_519	Ludwig_599	Ludwig_683	Ludwig_774
Ludwig_357	Ludwig_437	Ludwig_520	Ludwig_600	Ludwig_684	Ludwig_775
Ludwig_358	Ludwig_440	Ludwig_521	Ludwig_601	Ludwig_685	Ludwig_776
Ludwig_359	Ludwig_441	Ludwig_522	Ludwig_602	Ludwig_686	Ludwig_777
Ludwig_360	Ludwig_443	Ludwig_523	Ludwig_603	Ludwig_687	Ludwig_779
Ludwig_361	Ludwig_444	Ludwig_524	Ludwig_604	Ludwig_688	Ludwig_780
Ludwig_362	Ludwig_445	Ludwig_525	Ludwig_605	Ludwig_690	Ludwig_781
Ludwig_363	Ludwig_446	Ludwig_526	Ludwig_606	Ludwig_691	Ludwig_783
Ludwig_364	Ludwig_447	Ludwig_527	Ludwig_607	Ludwig_692	Ludwig_784
Ludwig_365	Ludwig_448	Ludwig_528	Ludwig_608	Ludwig_693	Ludwig_785
Ludwig_366	Ludwig_449	Ludwig_529	Ludwig_609	Ludwig_695	Ludwig_786
Ludwig_367	Ludwig_450	Ludwig_530	Ludwig_610	Ludwig_696	Ludwig_787
Ludwig_369	Ludwig_451	Ludwig_531	Ludwig_611	Ludwig_697	Ludwig_788
Ludwig_370	Ludwig_452	Ludwig_532	Ludwig_612	Ludwig_699	Ludwig_789
Ludwig_371	Ludwig_453	Ludwig_535	Ludwig_613	Ludwig_700	Ludwig_790
Ludwig_372	Ludwig_454	Ludwig_536	Ludwig_615	Ludwig_702	Ludwig_792
Ludwig_373	Ludwig_455	Ludwig_537	Ludwig_616	Ludwig_703	Ludwig_793
Ludwig_374	Ludwig_456	Ludwig_538	Ludwig_617	Ludwig_704	Ludwig_794
Ludwig_378	Ludwig_457	Ludwig_539	Ludwig_619	Ludwig_707	Ludwig_795
Ludwig_380	Ludwig_458	Ludwig_540	Ludwig_620	Ludwig_708	Ludwig_798
Ludwig_381	Ludwig_459	Ludwig_541	Ludwig_623	Ludwig_709	Ludwig_799
Ludwig_382	Ludwig_460	Ludwig_542	Ludwig_624	Ludwig_710	Ludwig_800
Ludwig_383	Ludwig_461	Ludwig_543	Ludwig_625	Ludwig_711	Ludwig_801
Ludwig_384	Ludwig_462	Ludwig_545	Ludwig_626	Ludwig_712	Ludwig_803
Ludwig_385	Ludwig_463	Ludwig_547	Ludwig_627	Ludwig_716	Ludwig_804
Ludwig_386	Ludwig_464	Ludwig_548	Ludwig_628	Ludwig_719	Ludwig_806
Ludwig_388	Ludwig_467	Ludwig_549	Ludwig_630	Ludwig_725	Ludwig_807
Ludwig_389	Ludwig_468	Ludwig_550	Ludwig_631	Ludwig_726	Ludwig_808
Ludwig_390	Ludwig_469	Ludwig_551	Ludwig_632	Ludwig_727	Ludwig_809
Ludwig_391	Ludwig_470	Ludwig_552	Ludwig_633	Ludwig_728	Ludwig_810
Ludwig_392	Ludwig_471	Ludwig_554	Ludwig_634	Ludwig_730	Ludwig_812
Ludwig_393	Ludwig_473	Ludwig_555	Ludwig_635	Ludwig_731	Ludwig_813
Ludwig_394	Ludwig_474	Ludwig_557	Ludwig_636	Ludwig_732	Ludwig_815
Ludwig_395	Ludwig_475	Ludwig_558	Ludwig_638	Ludwig_734	Ludwig_816

Ludwig_2689	Ludwig_2775	Ludwig_2854	Egger_15	Egger_92	Egger_170
Ludwig_2690	Ludwig_2776	Ludwig_2855	Egger_16	Egger_94	Egger_171
Ludwig_2692	Ludwig_2778	Ludwig_2858	Egger_18	Egger_95	Egger_172
Ludwig_2693	Ludwig_2779	Ludwig_2859	Egger_19	Egger_96	Egger_173
Ludwig_2694	Ludwig_2781	Ludwig_2860	Egger_20	Egger_98	Egger_175
Ludwig_2699	Ludwig_2783	Ludwig_2861	Egger_22	Egger_99	Egger_176
Ludwig_2700	Ludwig_2784	Ludwig_2862	Egger_23	Egger_101	Egger_177
Ludwig_2701	Ludwig_2785	Ludwig_2863	Egger_24	Egger_103	Egger_178
Ludwig_2702	Ludwig_2786	Ludwig_2864	Egger_26	Egger_104	Egger_179
Ludwig_2703	Ludwig_2787	Ludwig_2865	Egger_27	Egger_105	Egger_180
Ludwig_2705	Ludwig_2788	Ludwig_2866	Egger_28	Egger_106	Egger_181
Ludwig_2706	Ludwig_2789	Ludwig_2868	Egger_29	Egger_107	Egger_182
Ludwig_2708	Ludwig_2790	Ludwig_2870	Egger_30	Egger_108	Egger_183
Ludwig_2710	Ludwig_2792	Ludwig_2871	Egger_31	Egger_110	Egger_184
Ludwig_2711	Ludwig_2793	Ludwig_2872	Egger_32	Egger_111	Egger_185
Ludwig_2712	Ludwig_2794	Ludwig_2873	Egger_33	Egger_112	Egger_186
Ludwig_2715	Ludwig_2795	Ludwig_2874	Egger_34	Egger_113	Egger_187
Ludwig_2717	Ludwig_2796	Ludwig_2875	Egger_35	Egger_114	Egger_188
Ludwig_2718	Ludwig_2797	Ludwig_2876	Egger_36	Egger_115	Egger_189
Ludwig_2720	Ludwig_2798	Ludwig_2877	Egger_37	Egger_116	Egger_190
Ludwig_2721	Ludwig_2800	Ludwig_2878	Egger_38	Egger_117	Egger_191
Ludwig_2724	Ludwig_2801	Ludwig_2880	Egger_39	Egger_118	Egger_192
Ludwig_2727	Ludwig_2802	Ludwig_2881	Egger_40	Egger_119	Egger_193
Ludwig_2728	Ludwig_2803	Ludwig_2882	Egger_41	Egger_120	Egger_194
Ludwig_2729	Ludwig_2804	Ludwig_2883	Egger_42	Egger_121	Egger_196
Ludwig_2730	Ludwig_2805	Ludwig_2885	Egger_43	Egger_122	Egger_197
Ludwig_2731	Ludwig_2806	Ludwig_2886	Egger_44	Egger_123	Egger_198
Ludwig_2732	Ludwig_2807	Ludwig_2887	Egger_45	Egger_124	Egger_199
Ludwig_2733	Ludwig_2808	Ludwig_2889	Egger_46	Egger_125	Egger_201
Ludwig_2734	Ludwig_2809	Ludwig_2890	Egger_48	Egger_126	Egger_202
Ludwig_2735	Ludwig_2810	Ludwig_2891	Egger_49	Egger_128	Egger_204
Ludwig_2736	Ludwig_2811	Ludwig_2892	Egger_50	Egger_129	Egger_205
Ludwig_2737	Ludwig_2812	Ludwig_2893	Egger_51	Egger_130	Egger_206
Ludwig_2738	Ludwig_2813	Ludwig_2895	Egger_53	Egger_131	Egger_207
Ludwig_2739	Ludwig_2814	Ludwig_2896	Egger_54	Egger_132	Egger_208
Ludwig_2740	Ludwig_2815	Ludwig_2897	Egger_55	Egger_134	Egger_209
Ludwig_2741	Ludwig_2816	Ludwig_2898	Egger_56	Egger_135	Egger_210
Ludwig_2742	Ludwig_2818	Ludwig_2899	Egger_57	Egger_136	Egger_211
Ludwig_2743	Ludwig_2819	Ludwig_2900	Egger_58	Egger_137	Egger_212
Ludwig_2744	Ludwig_2820	Ludwig_2902	Egger_59	Egger_138	Egger_213
Ludwig_2745	Ludwig_2821	Ludwig_2903	Egger_60	Egger_139	Egger_214
Ludwig_2748	Ludwig_2822	Ludwig_2904	Egger_61	Egger_140	Egger_215
Ludwig_2749	Ludwig_2824	Ludwig_2905	Egger_62	Egger_141	Egger_216
Ludwig_2750	Ludwig_2825	Ludwig_2906	Egger_63	Egger_142	Egger_217
Ludwig_2752	Ludwig_2826	Ludwig_2907	Egger_64	Egger_143	Egger_218
Ludwig_2753	Ludwig_2827	Ludwig_2908	Egger_65	Egger_144	Egger_219
Ludwig_2754	Ludwig_2828	Ludwig_2909	Egger_66	Egger_146	Egger_220
Ludwig_2755	Ludwig_2829	Ludwig_2910	Egger_68	Egger_148	Egger_222
Ludwig_2756	Ludwig_2830	Ludwig_2911	Egger_69	Egger_149	Egger_223
Ludwig_2757	Ludwig_2831	Ludwig_2913	Egger_70	Egger_150	Egger_224
Ludwig_2758	Ludwig_2832	Ludwig_2915	Egger_71	Egger_151	Egger_225
Ludwig_2760	Ludwig_2833	Ludwig_2916	Egger_72	Egger_152	Egger_227
Ludwig_2761	Ludwig_2834	Ludwig_2917	Egger_74	Egger_154	Egger_228
Ludwig_2762	Ludwig_2835	Egger_1	Egger_75	Egger_155	Egger_229
Ludwig_2763	Ludwig_2836	Egger_2	Egger_76	Egger_156	Egger_230
Ludwig_2764	Ludwig_2841	Egger_3	Egger_77	Egger_158	Egger_231
Ludwig_2766	Ludwig_2842	Egger_5	Egger_79	Egger_159	Egger_232
Ludwig_2767	Ludwig_2843	Egger_6	Egger_81	Egger_160	Egger_233
Ludwig_2768	Ludwig_2844	Egger_7	Egger_82	Egger_162	Egger_234
Ludwig_2769	Ludwig_2846	Egger_8	Egger_84	Egger_163	Egger_235
Ludwig_2770	Ludwig_2847	Egger_9	Egger_86	Egger_164	Egger_236
Ludwig_2771	Ludwig_2848	Egger_10	Egger_87	Egger_165	Egger_237
Ludwig_2772	Ludwig_2849	Egger_12	Egger_89	Egger_166	Egger_238
Ludwig_2773	Ludwig_2852	Egger_13	Egger_90	Egger_168	Egger_239
Ludwig_2774	Ludwig_2853	Egger_14	Egger_91	Egger_169	Egger_241

Egger_242	Egger_313	Bengel_32	Bengel_104	Bengel_171	Bengel_240
Egger_243	Egger_315	Bengel_33	Bengel_105	Bengel_172	Bengel_241
Egger_244	Egger_316	Bengel_34	Bengel_106	Bengel_173	Bengel_242
Egger_245	Egger_319	Bengel_35	Bengel_107	Bengel_174	Bengel_243
Egger_246	Egger_320	Bengel_36	Bengel_108	Bengel_175	Bengel_244
Egger_247	Egger_321	Bengel_37	Bengel_109	Bengel_176	Bengel_245
Egger_248	Egger_323	Bengel_38	Bengel_110	Bengel_177	Bengel_246
Egger_249	Egger_324	Bengel_39	Bengel_111	Bengel_178	Bengel_247
Egger_250	Egger_325	Bengel_40	Bengel_112	Bengel_179	Bengel_248
Egger_251	Egger_326	Bengel_41	Bengel_113	Bengel_180	Bengel_249
Egger_252	Egger_327	Bengel_42	Bengel_114	Bengel_181	Bengel_250
Egger_253	Egger_328	Bengel_43	Bengel_116	Bengel_182	Bengel_251
Egger_254	Egger_329	Bengel_44	Bengel_117	Bengel_183	Bengel_252
Egger_255	Egger_330	Bengel_45	Bengel_118	Bengel_184	Bengel_253
Egger_256	Egger_331	Bengel_47	Bengel_119	Bengel_185	Bengel_254
Egger_257	Egger_332	Bengel_49	Bengel_120	Bengel_186	Bengel_255
Egger_258	Egger_333	Bengel_50	Bengel_121	Bengel_188	Bengel_256
Egger_259	Egger_334	Bengel_51	Bengel_122	Bengel_189	Bengel_257
Egger_260	Egger_335	Bengel_52	Bengel_123	Bengel_190	Bengel_258
Egger_261	Egger_337	Bengel_53	Bengel_124	Bengel_191	Bengel_259
Egger_262	Egger_338	Bengel_55	Bengel_125	Bengel_192	Bengel_260
Egger_263	Egger_343	Bengel_56	Bengel_126	Bengel_193	Bengel_261
Egger_264	Egger_345	Bengel_57	Bengel_127	Bengel_194	Bengel_262
Egger_265	Egger_347	Bengel_58	Bengel_128	Bengel_195	Bengel_263
Egger_266	Egger_348	Bengel_59	Bengel_129	Bengel_196	Bengel_264
Egger_267	Egger_349	Bengel_60	Bengel_130	Bengel_197	Bengel_265
Egger_268	Egger_350	Bengel_61	Bengel_131	Bengel_198	Bengel_266
Egger_269	Egger_351	Bengel_62	Bengel_132	Bengel_199	Bengel_267
Egger_271	Egger_354	Bengel_63	Bengel_133	Bengel_200	Bengel_268
Egger_273	Egger_355	Bengel_64	Bengel_134	Bengel_201	Bengel_269
Egger_274	Egger_356	Bengel_65	Bengel_135	Bengel_202	Bengel_270
Egger_275	Egger_357	Bengel_66	Bengel_136	Bengel_203	Bengel_271
Egger_276	Egger_358	Bengel_68	Bengel_137	Bengel_204	Bengel_272
Egger_277	Egger_359	Bengel_69	Bengel_138	Bengel_205	Bengel_273
Egger_279	Egger_360	Bengel_71	Bengel_139	Bengel_206	Bengel_274
Egger_280	Bengel_1	Bengel_72	Bengel_140	Bengel_207	Bengel_275
Egger_281	Bengel_2	Bengel_73	Bengel_141	Bengel_208	Bengel_276
Egger_282	Bengel_3	Bengel_74	Bengel_142	Bengel_209	Bengel_277
Egger_283	Bengel_4	Bengel_75	Bengel_143	Bengel_210	Bengel_278
Egger_285	Bengel_5	Bengel_76	Bengel_144	Bengel_211	Bengel_279
Egger_286	Bengel_6	Bengel_77	Bengel_145	Bengel_212	Bengel_280
Egger_287	Bengel_7	Bengel_79	Bengel_146	Bengel_213	Bengel_281
Egger_288	Bengel_8	Bengel_80	Bengel_147	Bengel_214	Bengel_282
Egger_289	Bengel_9	Bengel_81	Bengel_148	Bengel_215	Bengel_283
Egger_290	Bengel_10	Bengel_82	Bengel_149	Bengel_216	Bengel_284
Egger_291	Bengel_11	Bengel_84	Bengel_150	Bengel_217	Bengel_285
Egger_292	Bengel_12	Bengel_85	Bengel_151	Bengel_218	Bengel_286
Egger_293	Bengel_13	Bengel_86	Bengel_152	Bengel_220	Bengel_287
Egger_294	Bengel_14	Bengel_87	Bengel_153	Bengel_221	Bengel_288
Egger_295	Bengel_15	Bengel_88	Bengel_154	Bengel_222	Bengel_289
Egger_296	Bengel_16	Bengel_89	Bengel_156	Bengel_223	Bengel_290
Egger_297	Bengel_17	Bengel_90	Bengel_157	Bengel_224	Bengel_291
Egger_298	Bengel_18	Bengel_91	Bengel_158	Bengel_225	Bengel_292
Egger_299	Bengel_19	Bengel_92	Bengel_159	Bengel_226	Bengel_293
Egger_300	Bengel_20	Bengel_93	Bengel_160	Bengel_227	Bengel_294
Egger_301	Bengel_21	Bengel_94	Bengel_161	Bengel_228	Bengel_296
Egger_302	Bengel_22	Bengel_95	Bengel_162	Bengel_229	Bengel_297
Egger_305	Bengel_23	Bengel_96	Bengel_163	Bengel_230	Bengel_298
Egger_306	Bengel_24	Bengel_97	Bengel_164	Bengel_231	Bengel_299
Egger_307	Bengel_25	Bengel_98	Bengel_165	Bengel_232	Bengel_300
Egger_308	Bengel_26	Bengel_99	Bengel_166	Bengel_233	Bengel_301
Egger_309	Bengel_27	Bengel_100	Bengel_167	Bengel_236	Bengel_302
Egger_310	Bengel_28	Bengel_101	Bengel_168	Bengel_237	Bengel_303
Egger_311	Bengel_29	Bengel_102	Bengel_169	Bengel_238	Bengel_304
Egger_312	Bengel_30	Bengel_103	Bengel_170	Bengel_239	Bengel_305

Bengel_717	Spener_55	Spener_126	Spener_195	Spener_272	Spener_342
Bengel_718	Spener_56	Spener_127	Spener_196	Spener_273	Spener_343
Bengel_719	Spener_57	Spener_128	Spener_197	Spener_274	Spener_344
Bengel_720	Spener_58	Spener_129	Spener_198	Spener_275	Spener_345
Bengel_721	Spener_59	Spener_130	Spener_199	Spener_276	Spener_346
Bengel_722	Spener_60	Spener_131	Spener_200	Spener_277	Spener_347
Bengel_723	Spener_61	Spener_132	Spener_201	Spener_278	Spener_349
Bengel_725	Spener_62	Spener_133	Spener_202	Spener_279	Spener_350
Bengel_726	Spener_63	Spener_134	Spener_203	Spener_281	Spener_351
Bengel_727	Spener_64	Spener_135	Spener_204	Spener_282	Spener_352
Bengel_728	Spener_65	Spener_137	Spener_205	Spener_283	Spener_353
Bengel_729	Spener_66	Spener_138	Spener_206	Spener_284	Spener_354
Bengel_730	Spener_67	Spener_139	Spener_207	Spener_285	Spener_355
Bengel_731	Spener_69	Spener_141	Spener_208	Spener_286	Spener_356
Bengel_732	Spener_70	Spener_142	Spener_209	Spener_287	Spener_357
Bengel_733	Spener_71	Spener_143	Spener_210	Spener_288	Spener_359
Bengel_734	Spener_72	Spener_144	Spener_211	Spener_289	Spener_360
Bengel_735	Spener_74	Spener_145	Spener_212	Spener_291	Spener_361
Bengel_736	Spener_75	Spener_146	Spener_217	Spener_292	Spener_362
Bengel_737	Spener_76	Spener_147	Spener_218	Spener_294	Spener_363
Spener_1	Spener_77	Spener_148	Spener_219	Spener_295	Spener_364
Spener_2	Spener_78	Spener_149	Spener_220	Spener_296	Spener_365
Spener_3	Spener_79	Spener_151	Spener_221	Spener_297	Spener_366
Spener_4	Spener_81	Spener_152	Spener_222	Spener_298	Spener_367
Spener_5	Spener_82	Spener_154	Spener_223	Spener_299	Spener_368
Spener_6	Spener_83	Spener_155	Spener_224	Spener_300	Spener_369
Spener_7	Spener_84	Spener_156	Spener_225	Spener_301	Spener_370
Spener_8	Spener_85	Spener_157	Spener_226	Spener_302	Spener_371
Spener_9	Spener_86	Spener_158	Spener_227	Spener_304	Spener_372
Spener_12	Spener_87	Spener_159	Spener_228	Spener_305	Spener_373
Spener_14	Spener_88	Spener_160	Spener_229	Spener_306	Spener_374
Spener_15	Spener_89	Spener_161	Spener_230	Spener_307	Spener_375
Spener_17	Spener_91	Spener_162	Spener_231	Spener_308	Spener_376
Spener_18	Spener_92	Spener_163	Spener_233	Spener_309	Spener_377
Spener_19	Spener_93	Spener_164	Spener_234	Spener_310	Spener_378
Spener_21	Spener_94	Spener_165	Spener_235	Spener_311	Spener_379
Spener_23	Spener_95	Spener_166	Spener_237	Spener_312	Spener_380
Spener_24	Spener_96	Spener_167	Spener_238	Spener_313	Spener_381
Spener_25	Spener_98	Spener_168	Spener_239	Spener_314	Spener_382
Spener_26	Spener_99	Spener_169	Spener_240	Spener_315	Spener_383
Spener_27	Spener_100	Spener_170	Spener_241	Spener_316	Spener_384
Spener_28	Spener_101	Spener_171	Spener_244	Spener_317	Spener_385
Spener_29	Spener_102	Spener_172	Spener_245	Spener_318	Spener_386
Spener_30	Spener_103	Spener_173	Spener_247	Spener_319	Spener_387
Spener_31	Spener_104	Spener_174	Spener_249	Spener_320	Spener_388
Spener_32	Spener_105	Spener_175	Spener_250	Spener_321	Spener_389
Spener_33	Spener_106	Spener_176	Spener_251	Spener_322	Spener_390
Spener_34	Spener_107	Spener_177	Spener_252	Spener_323	Spener_391
Spener_35	Spener_108	Spener_178	Spener_254	Spener_324	Spener_392
Spener_38	Spener_109	Spener_179	Spener_255	Spener_325	Spener_393
Spener_39	Spener_110	Spener_180	Spener_256	Spener_326	Spener_394
Spener_40	Spener_111	Spener_181	Spener_257	Spener_327	Spener_395
Spener_41	Spener_112	Spener_182	Spener_258	Spener_328	Spener_396
Spener_42	Spener_113	Spener_183	Spener_260	Spener_329	Spener_397
Spener_43	Spener_114	Spener_184	Spener_261	Spener_330	Spener_398
Spener_44	Spener_115	Spener_185	Spener_262	Spener_331	Spener_400
Spener_45	Spener_116	Spener_186	Spener_263	Spener_332	Spener_401
Spener_46	Spener_117	Spener_187	Spener_264	Spener_333	Spener_402
Spener_47	Spener_118	Spener_188	Spener_265	Spener_334	Spener_403
Spener_48	Spener_119	Spener_189	Spener_266	Spener_336	Spener_404
Spener_49	Spener_121	Spener_190	Spener_267	Spener_337	Spener_405
Spener_50	Spener_122	Spener_191	Spener_268	Spener_338	Spener_406
Spener_51	Spener_123	Spener_192	Spener_269	Spener_339	Spener_407
Spener_52	Spener_124	Spener_193	Spener_270	Spener_340	Spener_409
Spener_53	Spener_125	Spener_194	Spener_271	Spener_341	Spener_410

Spener_411	Spener_482	Purmann_231	Abel_429	Ludwig_982
Spener_412	Spener_483	Purmann_281	Abel_481	Ludwig_1070
Spener_413	Spener_484	Purmann_313	Abel_498	Ludwig_1120
Spener_414	Spener_485	Purmann_314	Braeuner_29	Ludwig_1137
Spener_416	Spener_486	Purmann_341	Braeuner_31	Ludwig_1314
Spener_417	Spener_487	Purmann_351	Braeuner_33	Ludwig_1318
Spener_419	Spener_489	Kraepelin_51	Braeuner_64	Ludwig_1331
Spener_420	Spener_490	Kraepelin_84	Braeuner_67	Ludwig_1344
Spener_421	Spener_491	Kraepelin_198	Braeuner_76	Ludwig_1368
Spener_422	Spener_492	Kraepelin_221	Braeuner_84	Ludwig_1372
Spener_423	Spener_493	Kraepelin_226	Braeuner_97	Ludwig_1423
Spener_424	Spener_494	Kraepelin_277	Braeuner_109	Ludwig_1432
Spener_425	Spener_495	Kraepelin_363	Braeuner_128	Ludwig_1513
Spener_426	Spener_498	Kraepelin_406	Braeuner_132	Ludwig_1527
Spener_427	Pahl_16	Kraepelin_464	Braeuner_133	Ludwig_1556
Spener_429	Pahl_82	Kraepelin_528	Braeuner_152	Ludwig_1558
Spener_430	Pahl_353	Kraepelin_554	Braeuner_166	Ludwig_1564
Spener_431	Niviands_80	Kraepelin_656	Braeuner_167	Ludwig_1636
Spener_432	Niviands_82	Kraepelin_658	Braeuner_168	Ludwig_1689
Spener_433	Niviands_133	Koch_55	Braeuner_171	Ludwig_1820
Spener_434	Niviands_315	Koch_96	Braeuner_176	Ludwig_1919
Spener_435	Hanssen_94	Koch_179	Braeuner_177	Ludwig_1936
Spener_436	Hanssen_122	Koch_182	Braeuner_194	Ludwig_1961
Spener_437	Hanssen_216	Koch_184	Braeuner_198	Ludwig_1964
Spener_438	Hanssen_259	Koch_230	Braeuner_210	Ludwig_2016
Spener_439	Hanssen_284	Gall_58	Braeuner_214	Ludwig_2018
Spener_440	Hanssen_379	Gall_64	Braeuner_267	Ludwig_2091
Spener_441	Hanssen_392	Gall_129	Braeuner_312	Ludwig_2129
Spener_442	Hanssen_403	Gall_342	Braeuner_317	Ludwig_2158
Spener_443	Strauss_121	Gall_412	Braeuner_321	Ludwig_2159
Spener_444	Unzer_13	Gall_574	Braeuner_326	Ludwig_2164
Spener_445	Unzer_23	Gall_608	Braeuner_364	Ludwig_2175
Spener_447	Unzer_27	Abel_5	Braeuner_371	Ludwig_2209
Spener_448	Unzer_42	Abel_15	Braeuner_408	Ludwig_2227
Spener_449	Unzer_48	Abel_20	Braeuner_410	Ludwig_2256
Spener_450	Unzer_78	Abel_32	Braeuner_432	Ludwig_2268
Spener_451	Unzer_101	Abel_63	Braeuner_502	Ludwig_2289
Spener_452	Unzer_132	Abel_85	Braeuner_513	Ludwig_2478
Spener_453	Unzer_266	Abel_89	Ludwig_28	Ludwig_2496
Spener_454	Unzer_310	Abel_106	Ludwig_53	Ludwig_2497
Spener_455	Unzer_318	Abel_110	Ludwig_78	Ludwig_2509
Spener_456	Unzer_385	Abel_117	Ludwig_148	Ludwig_2595
Spener_457	Unzer_424	Abel_153	Ludwig_160	Ludwig_2612
Spener_458	Unzer_444	Abel_162	Ludwig_238	Ludwig_2658
Spener_459	Reil_182	Abel_165	Ludwig_267	Ludwig_2714
Spener_460	Reil_197	Abel_169	Ludwig_269	Ludwig_2716
Spener_461	Reil_219	Abel_179	Ludwig_277	Ludwig_2839
Spener_462	Reil_261	Abel_186	Ludwig_278	Ludwig_2845
Spener_463	Reil_733	Abel_206	Ludwig_322	Ludwig_2884
Spener_464	Reil_822	Abel_225	Ludwig_340	Egger_127
Spener_465	Reil_859	Abel_253	Ludwig_434	Egger_157
Spener_466	Reil_978	Abel_264	Ludwig_573	Egger_174
Spener_467	Reil_1005	Abel_270	Ludwig_646	Bengel_314
Spener_468	Reil_1078	Abel_284	Ludwig_651	Spener_68
Spener_469	Reil_1106	Abel_289	Ludwig_705	Spener_213
Spener_470	Reil_1110	Abel_301	Ludwig_713	Spener_246
Spener_471	Reil_1120	Abel_302	Ludwig_746	Spener_335
Spener_472	Purmann_12	Abel_303	Ludwig_771	Spener_358
Spener_473	Purmann_46	Abel_305	Ludwig_835	Spener_399
Spener_476	Purmann_54	Abel_315	Ludwig_853	Spener_428
Spener_477	Purmann_69	Abel_343	Ludwig_887	Ludwig_907
Spener_478	Purmann_155	Abel_349	Ludwig_916	
Spener_479	Purmann_220	Abel_355	Ludwig_917	
Spener_480	Purmann_222	Abel_356	Ludwig_947	
Spener_481	Purmann_224	Abel_399	Ludwig_949	

C. Texte im Trainingskorpus

Autor	Kurztitel	Jahr	Zeitabschnitt
Spener	Friede	1686	1650-1700
Siegemund	Unterricht	1690	1650-1700
Schupp	Schriften	1663	1650-1700
Santa	Judas (3)	1692	1650-1700
Santa	Judas (1)	1686	1650-1700
Geiger	Schütz	1672	1650-1700
Dannhauer	Catechismus 11	1678	1650-1700
Dannhauer	Catechismus 10	1673	1650-1700
Dannhauer	Catechismus 5	1654	1650-1700
Dannhauer	Catechismus 4	1653	1650-1700
Bürger	Canditatus	1692	1650-1700
Spener	Bedencken	1702	1700-1750
Woyt	Gazophylacium	1737	1700-1750
Wesenigk	Polysigma	1702	1700-1750
Szentivanyi	Ketzereylehren	1703	1700-1750
Spalding	Bestimmung	1749	1700-1750
Seip	Gesundbrunnen	1717	1700-1750
Pomet	Meterialist	1717	1700-1750
Pertsch	Recht	1721	1700-1750
Benner	Herrnhuterey (4)	1748	1700-1750
Benner	Herrnhuterey (2)	1746	1700-1750
Benner	Herrnhuterey (1)	1746	1700-1750

Arnold	Cyprian	1700	1700-1750
Sailer	Selbstmord	1785	1750-1800
Sailer	Prediger	1791	1750-1800
Münter	Bekehren	1772	1750-1800
Humboldt	Galvanische	1797	1750-1800
Hufeland	Leben	1797	1750-1800
Haller	Anfangsgründe (8)	1776	1750-1800
Haller	Anfangsgründe (7)	1775	1750-1800
Haller	Anfangsgründe (5)	1772	1750-1800
Haller	Anfangsgründe (4)	1768	1750-1800
Haller	Anfangsgründe (3)	1766	1750-1800
Haller	Anfangsgründe (2)	1762	1750-1800
Haller	Anfangsgründe (1)	1759	1750-1800
Blumenbach	Physiologie	1795	1750-1800
Blumenbach	Physiologie	1792	1750-1800
Weiss	Starrkrampf	1824	1800-1850
Strauss	Jesus (2)	1836	1800-1850
Schubert	Symbolik	1814	1800-1850
Müller	Gesichterscheinung	1826	1800-1850
Hahnemann	Organon	1810	1800-1850
Griesinger	Psychische	1845	1800-1850
Carus	Psyche	1846	1800-1850
Carus	Gynäkologie	1820	1800-1850
Burdach	Propädeutik	1800	1800-1850
Blumenbach	Knochen	1807	1800-1850
Blumenbach	Anatomie	1805	1800-1850
Weismann	Naturzüchtung	1893	1850-1900
Vogt	Briefe	1851	1850-1900
Virchow	Cellularpathologie	1858	1850-1900
Siebold	Süßwasserfische	1858	1850-1900
Schleiden	Menschengeschlecht	1863	1850-1900

Roux	Kampf	1881	1850-1900
Pflüger	Rückenmann	1853	1850-1900
Mendel	Pflanzenhybriden	1866	1850-1900
Löhe	Predigt	1873	1850-1900
Löhe	Entgegenkommen	1857	1850-1900
Löhe	Abendmahlzucht	1853	1850-1900
Hering	Lichtsinn	1878	1850-1900
Häckel	Plastidule	1876	1850-1900
Fuhlrott	Neanderthaler	1865	1850-1900
Egger	Vater	1895	1850-1900
Ebbinghaus	Gedächtnis	1885	1850-1900
Daumer	Krone	1859	1850-1900
Berlin	Wortblindheit	1887	1850-1900

Tabelle C.1.: Trainingskorpus

D. Zusätzliche Grafiken



Abbildung D.1.: Anteil der appositiven (0,5) und restriktiven (-0,5) Relativsätze an den ausgelagerten Relativsätzen im medizinischen Teilkorpus. Der z-Wert der logistischen Regression (Tabelle 6.26) ist negativ.



Abbildung D.2.: Anteil der appositiven (0,5) und restriktiven (-0,5) Relativsätze an den ausgelagerten Relativsätzen im theologischen Teilkorpus. Der z-Wert der logistischen Regression (Tabelle 6.28) ist positiv.

E. Zusätzliche Regressionsanalysen

E.1. Regressionsanalysen bei gleicher Länge

E.1.1. Nominalphrasen

Für die Berechnung dieser Regression (Tabelle E.1) wurden nur Nominalphrasen einbezogen, die vier Wörter lang sind, weil das der häufigsten Länge ($n=40$) im untersuchten Korpus entspricht. Von den 40 Nominalphrasen sind 19 extraponiert und 18 eingebettet. 22 stammen aus medizinischen Texten, 18 aus theologischen. Die Verteilung auf die Zeitabschnitte ist folgende:

- 1650-1700: 14
- 1700-1750: 3
- 1750-1800: 10
- 1800-1850: 4
- 1850-1900: 9

	Estimate	Std.-Fehler	z-Wert	p-Wert
Intercept	12,479	5,196	2,402	0,0163 *
Skipgram _{mean}	-3,511	1,420	-2,472	0,0134 *
Genre	-20,345	11,197	-1,817	0,0692 .
Orality Score	3,208	1,679	1,910	0,0561 .
Skipgram _{mean} :Genre	4,665	2,819	1,655	0,0980 .

Tabelle E.1.: Logistische Regression aller im Korpus annotierten Nominalphrasen der Länge 4, finales Modell.

E.1.2. Präpositionalphrasen

In die in Tabelle E.2 dargestellte Regression flossen nur Präpositionalphrasen mit 6 Tokens ein (n=138). Je 68 Phrasen sind extraponiert beziehungsweise eingebettet, wobei die eingebetteten Phrasen über die Sample-Funktion (vgl. R Core Team 2022) aus einem Pool von 145 *in situ*-Phrasen gezogen wurden. 55 stammen aus medizinischen Texten, 81 aus theologischen. Die Verteilung auf die Zeiten sieht wie folgt aus:

- 1650-1700: 24
- 1700-1750: 13
- 1750-1800: 43
- 1800-1850: 5
- 1850-1900: 51

	Estimate	Std.-Fehler	z-Wert	p-Wert
Intercept	-4,7439	3,0246	-1,568	0,1168
Skipgram _{mean}	0,7541	0,7585	0,994	0,3201
Genre	12,7717	5,6694	2,253	0,0243 *
Orality Score	3,6193	2,0608	1,756	0,0790 .
Zeitstufe	0,5627	0,2586	2,176	0,0296 *
Skipgram _{mean} :GenreSC	-3,6836	1,5114	-2,437	0,0148 *
Orality Score:Zeitstufe	-1,2037	0,4934	-2,439	0,0147 *

Tabelle E.2.: Logistische Regression der im Korpus annotierten Präpositionalphrasen der Länge 6, finales Modell.

E.2. Relativsätze ohne Variable Typ

Die Regression wurde mit R (R Core Team 2022) durchgeführt und der Text von Ludwig (1856) zur besseren Vergleichbarkeit mit Tabelle 6.24 erneut abgeschlossen.

Auch wenn sich die Werte der Estimates damit die z- und p-Werte der Ergebnisse natürlicherweise ändern, da nun auch die Relativsätze miteinbezogen wurden, deren Relativsatztyp unbestimmt geblieben ist, sind fast die gleichen Variablen relevant für Extrapolation wie auch in Tabelle 6.24.

	Estimate	Std.-Fehler	z-Wert	p-Wert	
Intercept	3,945	0,71	5,55	<0,0001	***
Skipgram _{mean}	-0,34	0,18	-1,95	0,05	.
Länge _{log}	-2,56	0,256	-9,998	<0,0001	***
Genre	-0,74	0,31	-2,36	0,02	*
Orality Score	-1,25	0,43	-2,93	0,003	**
Zeitstufe	-0,32	0,096	-3,29	0,001	***
Länge _{log} :Genre	0,53	0,28	1,898	0,06	.
Länge _{log} :Orality Score	1,19	0,396	3,01	0,003	**
Länge _{log} :Zeitstufe	0,15	0,09	1,67	0,09	.
Genre:Orality Score	2,17	0,24	9,12	<0,0001	***
Genre:Zeitstufe	0,28	0,07	3,96	<0,0001	***
Orality Score:Zeitstufe	-0,29	0,09	-3,09	0,002	**

Tabelle E.3.: Logistische Regression aller Relativsätze im Gesamtkorpus *exklusive der Variable des Relativsatztyps*, finales Modell. Negative z-Werte sprechen für Extraposition.

Für Auslagerung sprechen: Die durchschnittlichen Skipgram-Surprisalwerte, die Länge, das Genre, die Zeitstufe und – hier liegt der größte Unterschied zu der Regressionanalyse, die den Relativsatztyp einbezieht – der Orality Score sowie die Interaktion aus dem Orality Score und der Zeitstufe.

Für Einbettung sprechen: die Interaktion aus Länge und Genre, aus Länge und Orality Score, aus Länge und Zeitstufe, aus Genre und Orality Score und aus Genre und Zeitstufe.

Daraus ist zu schlussfolgern, dass der Ausschluss der Relativsätze ohne Typbestimmung aus der logistischen Regression in Kapitel 6.3 zulässig war und das Ergebnis als stabil angesehen werden kann.

Danksagung

An dieser Stelle ist es an der Zeit, mich bei all denen zu bedanken, die mich beim Schreiben dieser Dissertation unterstützt haben.

An erster Stelle gilt mein Dank meinem Doktorvater, Professor Dr. Augustin Speyer, für die Betreuung der Arbeit, für die Unterstützung und den konstruktiven Austausch. Ich habe unsere Gespräche immer sehr motivierend und inspirierend gefunden. Ich danke auch Professorin Dr. Stefanie Dipper, meiner Zweitgutachterin, nicht nur für besagtes Zweitgutachten, sondern auch für die konstruktive Zusammenarbeit im Projekt und die Bereitstellung von WebAnno als Annotationstool. Ebenfalls möchte ich mich bei Professor Dr. Ingo Reich für das Drittgutachten zur Arbeit.

Des Weiteren möchte ich mich bei meinen Kollegen der Abteilung *Neuere deutsche Sprachwissenschaft* bedanken, besonders bei Dr. Jessica Schmidt, Lisa Schäfer, Dr. Robin Lemke, Dr. Philipp Rauth, Julia Hertel und Jenny Diener sowie bei Constanze Göbel und Ella. Ein großer Dank geht auch an die beiden Projekt-HiWis Nils Dörr und Katharina Wilhelm, die mich bei der Annotation unterstützt haben. Außerdem geht mein Dank erneut Richtung Ruhr-Universität Bochum, nämlich an Dr. Katrin Ortmann.

Diese Arbeit wäre natürlich nicht möglich gewesen, ohne die Unterstützung meiner Freunde und meiner Familie. Dafür möchte ich mich bei meinem Mann Johannes Mischo und bei Sara Zimmermann bedanken. Last, but definitely not least möchte ich mich bei meinen Eltern, Klaus Voigtmann und Ursula Kunz-Voigtmann für alles bedanken – auch dafür, dass ihr bereit wart, eure Namen für meine Beispiele bereitzustellen!

Selbständigkeitserklärung

Hiermit versichere ich, dass ich die Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die anderen Werken dem Wortlaut oder dem Sinn nach entnommen sind, wurden unter Angabe der Quellen als Entlehnung kenntlich gemacht. Bei Zeichnungen, Skizzen und Plänen sowie bildlichen und grafischen Darstellungen ist angegeben, ob sie selbständig gefertigt, nach eigenen Angaben oder durch andere ausgeführt oder übernommen worden sind.

Saarbrücken,

Unterschrift: _____