



Korpuslinguistik und interdisziplinäre  
Perspektiven auf Sprache

Band **7**

**Sascha Wolfer**

**Verstehen und  
Verständlichkeit  
juristisch-  
fachsprachlicher Texte**

narr  
ranck  
e|atte  
mpto



Interdisciplinary Perspectives on Language  
Corpus Linguistics and  
Perspectives on Applied  
Linguistics and Foreign Language

# Verstehen und Verständlichkeit

Handbook of Applied Linguistics  
Volume 1: Applied Linguistics in the  
21st Century

# juristisch- fachsprachlicher Texte

Edited by  
Mark Baker, Mark Davies, John Hillier,  
Richard Kohn, Robert Kramlinger, John Leung,  
Michael McCarthy, Tony Mcenery, Anton Nij  
Muel Steen, John Swales, Heiko Zemanek

The book has been reviewed in a peer review  
process and is a peer reviewed journal.



**Korpuslinguistik und interdisziplinäre  
Perspektiven auf Sprache**

**Corpus Linguistics and  
Interdisciplinary Perspectives on Language**

**Bd./Vol. 7**

Herausgeber/Editorial Board:

Marc Kupietz, Harald Lungen, Christian Mair

Gutachter/Advisory Board:

Heike Behrens, Mark Davies, Martin Hilpert,  
Reinhard Köhler, Ramesh Krishnamurthy, Ralph Ludwig,  
Michaela Mahlberg, Tony McEnery, Anton Näf,  
Michael Stubbs, Elke Teich, Heike Zinsmeister

Die Bände der Reihe werden in einem Peer-review-  
Verfahren geprüft / CLIP is a peer reviewed series.

**Sascha Wolfer**

# **Verstehen und Verständlichkeit juristisch- fachsprachlicher Texte**

**narr** |  
VERLAG



Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de> abrufbar.

© 2017 · Narr Francke Attempto Verlag GmbH + Co. KG  
Dischingerweg 5 · D-72070 Tübingen

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt.  
Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne  
Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für  
Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und  
Verarbeitung in elektronischen Systemen.  
Gedruckt auf chlorfrei gebleichtem und säurefreiem Werkdruckpapier.

Internet: [www.narr.de](http://www.narr.de)  
E-Mail: [info@narr.de](mailto:info@narr.de)

Redaktion: Norbert Volz  
Layout: Sonja Tröster und Joachim Hohwieler  
Printed in Germany

ISSN 2191-9577  
ISBN 978-3-8233-8152-5

# Inhalt

<b>1.</b>	<b>Einleitung</b> .....	9
1.1	Gliederung.....	10
<b>2.</b>	<b>Textverarbeitung</b> .....	13
2.1	Kognitionswissenschaftliche Modelle des Textverstehens.....	17
2.2	Anaphernauflösung.....	22
<b>3.</b>	<b>Textverständlichkeit</b> .....	39
3.1	Verständlichkeitsoptimierung von Rechtssprache.....	39
3.2	Verständlichkeitsmodelle.....	50
3.3	Komplexitätsmerkmale der Rechtssprache.....	62
3.4	Müssen Rechtstexte verständlich sein?.....	66
<b>4.</b>	<b>Blickbewegungsmessung</b> .....	75
4.1	Blickbewegungsmaße beim Lesen.....	77
4.2	Lesekorpora.....	84
4.2.1	Der Zweck von Lesekorpora.....	85
4.2.2	Kontrolle von Störvariablen.....	87
<b>5.</b>	<b>Das Freiburg Legalese Reading Corpus (FLRC)</b> .....	91
5.1	Kompilierung des <i>Freiburg Legalese Reading Corpus</i> .....	91
5.1.1	Textauswahl und Gliederung des <i>Freiburg Legalese Reading Corpus</i> .....	91
5.1.2	Methode und Versuchsablauf.....	96
5.1.3	Fragen.....	100
5.2	Linguistische Oberflächenanalysen der verwendeten Texte.....	103
5.2.1	FLRCRef: Reformulierungskorpus.....	105
5.2.2	FLRCNat: Zeitungsartikel, Pressemitteilungen, Urteile.....	120
5.2.3	Referenzielle Ausdrücke.....	131
5.3	Vergleich des FLRC mit vorhandenen Lesekorpora.....	134
5.3.1	Einfache Kennzahlen.....	136
5.3.2	Verfügbare Variablen.....	140
5.4	Zusammenfassung.....	143

<b>6.</b>	<b>Verhaltensdaten</b> .....	145
6.1	Statistisches Vorgehen und Datenaufbereitung .....	145
6.2	Analyse der Verhaltensdaten .....	151
6.2.1	Lesedauer .....	151
6.2.2	Antwortperformanz .....	159
6.2.3	Antwortperformanz und Lesezeit .....	166
6.2.4	Zusammenfassung .....	172
<b>7.</b>	<b>Blickbewegungsdaten</b> .....	175
7.1	Analysen auf lexikalischer Ebene .....	177
7.1.1	<i>First fixation duration</i> (FFD) .....	178
7.1.2	<i>First-pass reading time</i> (FPRT) .....	180
7.1.3	<i>Regression path duration</i> (RPD) .....	185
7.1.4	<i>Total reading time</i> (TRT) .....	187
7.1.5	Zusammenfassung und Evaluation .....	190
7.2	Bi- und Trigramme .....	194
7.3	Analysen auf syntaktischer Ebene .....	200
7.3.1	Position im Satz und syntaktische Einbettung .....	200
7.3.2	Exkurs: Syntaktische Knoten als Analyseregionen .....	204
7.3.3	Relevanz auf Phrasenebene .....	205
7.3.4	Constraints als erklärendes Konzept: Relative Einbettungstiefe .....	209
7.3.5	Ähnlichkeitsbasierte Interferenz .....	216
7.3.6	Zusammenfassung .....	227
7.4	Analysen auf Textebene .....	229
7.4.1	Referenztyp .....	232
7.4.2	Referenzdistanz .....	233
7.4.3	Abstraktheit und zusätzliche Informationen .....	238
7.5	Auswirkungen der Reformulierungen auf die Rezeption .....	246
7.5.1	Nominalisierungen .....	248
7.5.2	Komplexe Nominalphrasen .....	254
7.5.3	Syntaktisch komplexe Texte .....	259
7.5.4	Fernabhängigkeiten: Dependenzdistanz .....	266

<b>8.</b>	<b>Implikationen</b> .....	281
8.1	Implikationen für die Grundlagenforschung .....	285
8.2	Angewandte Sprachwissenschaft .....	286
8.3	Schluss .....	289
<b>9.</b>	<b>Literatur</b> .....	291



# 1. Einleitung

Die französische Cour de Cassation, das höchste Gericht der ordentlichen Gerichtsbarkeit Frankreichs, das Urteile auf Rechtsfehler überprüft und gegebenenfalls aufhebt, begründet ihre Urteile nach dem Prinzip des *jugement à phrase unique*. Eine Gerichtsentscheidung besteht dabei aus einem einzigen Satz, der nach dem Muster „la cour ... casse“ (dt. der Hof ... hebt auf) aufgebaut ist. Begründet wird die Entscheidung über eingeschobene Nebensätze, die meist der Form „attendu que“ (dt. in Erwägung der Tatsache dass / unter Beachtung von) folgen oder über „considérant“ (dt. *berücksichtigend / unter Berücksichtigung von*) gebildet werden. Obwohl diese strenge Form gesetzlich nicht vorgeschrieben ist (Krefeld 1985), hält sich die Cour de Cassation noch immer streng an diesen Aufbau. Andere Gerichtsbarkeiten, wie die Cours d'Appel,<sup>1</sup> haben sich von dieser während der französischen Revolution kanonisierten Form abgewandt. Der Grund hierfür scheint auf der Hand zu liegen: Mit steigender Komplexität der Entscheidungen steigt die Anzahl der konjunktionalen Nebensatzkonstruktionen, die zwischen Subjekt (la cour) und Objekt (der aufzuhebenden Entscheidung) des Matrixsatzes gezwängt werden müssen. Im seit 1977 eingeführten „style nouveau“ werden Tatsachen, Verlauf des Prozesses sowie die Positionen der sich entgegenstehenden Parteien diskursiv verfasst. Tatsächlich war ein Ziel dieser Reform, die Urteile für die Zielgruppe der Rechtssuchenden verständlicher zu formulieren (Lashöfer 1992). Auch ÜbersetzerInnen tun sich offenbar mit der starren und formalisierten Einsatz-Form schwer. Die Form der *phrase unique* jedoch als „anachronistisches Kuriosum“ abzutun, wäre zu einfach (so auch Ballansat-Aebi 2000): Symbolisiert sie doch die inhaltliche Einheit des Urteils. Dass jedoch auch Standesbewusstsein der RichterInnen und Prestige mitspielen, wenn seitens der Justiz für die *phrase unique* argumentiert wird, ist in einigen Publikationen zum Thema mehr oder weniger deutlich zwischen den Zeilen zu lesen (so bspw. bei Ballansat-Aebi 2000; Krefeld 1985). Der Streitfall um die *phrase unique* scheint also stellvertretend für einen Konflikt zu stehen, der sich sicherlich nicht nur in Frankreich abspielt. Hier tritt das Problem besonders deutlich zu Tage, dass Fachsprachen mit ihren auf den ersten Blick unnötig verkomplizierenden Regeln und Traditionen oft ein schier unüberwindbares Verständnishindernis für Laien darstellen. Dies ist eine Erfahrung, die ExpertInnen im jeweiligen Gebiet so nicht teilen können. Oft stellt sich für sie die Kommunikation innerhalb der Fachsprache einfacher, effektiver und genauer dar. Dieser Konflikt

---

<sup>1</sup> Durch die Cours d'Appel (dt. Berufungsgerichtshof) kann über einen durch ein erstinstanzliches Gericht entschiedenen Fall erneut entschieden werden.



wird in der vorliegenden Arbeit dann relevant, wenn wir<sup>2</sup> uns die Frage stellen, ob Rechtstexte unter sprachlicher Hinsicht überhaupt optimierungsbedürftig sind. Hauptsächlich werden wir uns jedoch mit der Frage beschäftigen, wie sich ausgewählte linguistische Merkmale juristischer Fachtexte auf den Verstehensprozess auswirken.

Eines der Ziele dieser Arbeit ist es, die Konsequenzen von Reformulierungen zu untersuchen. Außerdem werde ich anhand eines Lesekorpus psycholinguistische Fragestellungen auf mehreren linguistischen Ebenen bearbeiten. Der Analyseprozess geht von der Wort- über die Satzebene bis hin zu textuellen Eigenschaften. Ein Analyseschritt ist dabei immer Voraussetzung für den nächsten. So sollen einige Quellen sprachlicher Komplexität systematisch in ihrer Auswirkung auf den Leseprozess untersucht werden. Außer natürlich-sprachlichen Texten beleuchte ich auch die Verarbeitung von gezielt reformulierten Textpassagen. Dabei werde ich analysieren, ob und welche Auswirkungen die Reformulierungen auf den Verstehensprozess haben.

Die Entscheidung, Urteile des Bundesverfassungsgerichts als Ausgangspunkt für die Analyse der Verständlichkeit von fachsprachlichen Texten zu verwenden, gründet sich demzufolge auf zwei Beobachtungen: 1) Texte aus dem Bereich der Rechtssprache können als prototypisch komplexe fachsprachliche Texte verstanden werden. 2) Im Feld der rechtssprachlichen Texte ist insbesondere den Entscheidungen des Bundesverfassungsgerichts hohe gesellschaftliche Relevanz beizumessen.

## 1.1 Gliederung

Im Laufe dieser in ihrem Wesen interdisziplinären Arbeit werden wir uns immer wieder zwischen verschiedenen Fachbereichen bewegen – insbesondere der Linguistik und der Kognitionswissenschaft. Die Arbeit gliedert sich im Folgenden in sechs Hauptkapitel. Kapitel 2 widmet sich der Textverarbeitung. Dort werden einige der einflussreichsten kognitionswissenschaftlichen Modelle des menschlichen Textverstehens vorgestellt. Dort wird auch speziell das Verstehen von Anaphern beleuchtet (2.2). Mit dem darauf folgenden Kapitel 3 bewegen wir uns zu Themen, die in der germanistischen Linguistik

---

<sup>2</sup> Das 'wir' kann in der vorliegenden Arbeit auch als 'ich' gelesen werden. Mit der 'wir'-Form möchte ich verdeutlichen, dass das dieser Dissertation zugrundeliegende Forschungsvorhaben innerhalb des psycholinguistischen Forschungslabors um Lars Konieczny in der Abteilung Kognitionswissenschaft an der Universität Freiburg verfolgt wurde. Somit sind immer Diskussionen, Ideensammlungen und Auseinandersetzungen mit vielen KollegInnen beteiligt. Alle Entwürfe der Annotationen, Studiendesigns, deren Implementierung sowie die statistischen Auswertungen habe ich federführend selbst angefertigt.

großen Einfluss haben. Zunächst werden drei Verständlichkeitsmodelle vorgestellt (3.2), die zumindest teilweise kognitionswissenschaftliche Textverarbeitungsmodelle aufgreifen. Im weiteren Verlauf von Kapitel 3 wird kurz vorgestellt, welche Komplexitätsmerkmale in juristischen Texten auf linguistischer Seite festgestellt wurden (3.3). Zum Abschluss von Kapitel 3 wird anhand ausgewählter Literatur ein Streitfall nachvollzogen (3.4): Müssen juristische Texte verständlich sein? Diese Frage ist für die vorliegende Arbeit insofern zentral, weil eine Verständlichkeitsoptimierung juristischer Texte hinfällig wäre, wenn es völlig klar wäre, dass sie gar nicht verständlich sein müssen. Ein großer Teil der in dieser Arbeit vorgestellten Ergebnisse wird allerdings auch unabhängig von der konkreten Textgattung relevant sein. Mit Kapitel 4 schwenkt die Arbeit zurück in die methodisch-kognitionswissenschaftliche Richtung: Es werden zunächst die eingesetzten Lesemaße vorgestellt (4.1). Dabei wird auch zu klären sein, wie die verschiedenen Maße interpretiert werden können. In Abschnitt 4.2 wird dann eine Vorstellung sogenannter Lesekorpora, Sammlungen von Blickbewegungen beim Lesen von Texten, folgen.

Die Kapitel 5, 6 und 7 sind die Kernkapitel dieser Dissertation. In Kapitel 5 wird das erhobene *Freiburg Legalese Reading Corpus* (FLRC) vorgestellt. Zunächst gehen wir auf den Aufbauprozess ein (5.1). Es wird zu klären sein, welches Design dem Korpus zugrunde liegt, wie die Texte ausgewählt wurden und wie die Blickbewegungsdaten der VersuchsteilnehmerInnen erhoben wurden. Abschnitt 5.2 kann für sich als ein kleines Ergebniskapitel gesehen werden: Hier werden die im FLRC verwendeten Texte anhand einiger korpuslinguistischer Maße einer Oberflächenanalyse<sup>3</sup> unterzogen. Dort wird gezeigt, wie sich bestimmte Teile des FLRC durch Reformulierung(en) verändern, wie sich Komplexität von einer linguistischen Ebene auf die andere verlagert und wie sich verschiedene Textgattungen innerhalb des FLRC unterscheiden. Im darauf folgenden Abschnitt 5.3 vergleichen wir den FLRC mit bereits vorhandenen Lese- oder Blickbewegungskorpora. In Kapitel 6 und 7 werden die Ergebnisse der vorliegenden Arbeit vorgestellt. In Kapitel 6 werden zunächst die verwendeten statistischen Methoden und das generelle Vorgehen bei der Modellanpassung beschrieben. Ergebnisse, die keine Analyse von Blickbewegungsdaten voraussetzen, werden im weiteren Verlauf des

---

<sup>3</sup> Das Wort 'Oberflächenanalyse' könnte für Verwirrung sorgen. Es wird nicht im Gegensatz zur sogenannten Texttiefenstruktur benutzt, die in Anlehnung an die Generative Transformationsgrammatik die abstrakte Struktur – van Dijk (1980) nennt dies Makrostruktur – eines Textes beschreibt (vgl. u.a. Schmidt 1976, S. 155ff.). Oberflächenanalyse soll vielmehr die Unterscheidung zwischen linguistischer Struktur und dem Verstehen derselbigen verdeutlichen. Hierzu gehört auch der Unterschied zwischen linguistischer Komplexität und Komplexität bezüglich der Rezeption.

Kapitels erarbeitet. Im Einzelnen werden dies die Gesamtlesedauer der Texte (6.2.1), die Antwortperformanz (6.2.2) sowie das Zusammenspiel der beiden Variablen (6.2.3) sein. In Kapitel 7, der sich der Analyse der Lese-/Blickbewegungsdaten widmet, werden wir uns zunächst durch die Ebenen der linguistischen Annotation vorarbeiten und entsprechende Effekte herausarbeiten. Effekte auf der lexikalischen Ebene bilden den Ausgangspunkt (7.1). Die Bi- und Trigramm-Ebene folgt (7.2), bevor Effekten auf syntaktischer Ebene ein relativ großer Raum eingeräumt wird (7.3). Effekte auf Textebene (7.4), genauer auf der Referenzebene, werden den Abschluss dieses Durchlaufs durch die linguistischen Ebenen bilden. Als Abschluss des Ergebniskapitels werden Rezeptionskonsequenzen der Reformulierungen (Abschnitt 7.5) herausgearbeitet. Die drei letztgenannten Abschnitte (syntaktische Ebene, textuelle Ebene, Konsequenzen der Reformulierungen) sind recht umfangreich und daher nochmals in sich unterteilt. Für die entsprechenden Gliederungen sei aber jeweils auf die Anfänge der entsprechenden Abschnitte verwiesen.

In Kapitel 8.1 werde ich die Ergebnisse der Reformulierungsstudien zusammenfassen und diskutieren. Es werden dann in Abschnitt 8.2 Implikationen für die Grundlagenforschung und die angewandte Sprachwissenschaft abgeleitet – an Stellen, an denen die Ergebnisse der vorliegenden Arbeit dies ermöglichen, werden auch konkrete (Re-)Formulierungsvorschläge gemacht. In Abschnitt 8.3 werden die Fäden aus den unterschiedlichen Fachbereichen und linguistischen Betrachtungsebenen wieder aufgenommen und verbunden.

## 2. Textverarbeitung

In diesem Kapitel wird aufgearbeitet, welche kognitiven Prozesse auf der Seite des Lesers/der Leserin wirken. Dabei werden verschiedene Faktoren wie linguistische Struktur, allgemeines Weltwissen und kognitive Beschränkungen eingeführt. Ein kurzes aufeinander aufbauendes Review der wichtigsten Theorien führt die relevanten Aspekte ein, die direkt oder indirekt in die später dargestellten Analysen und in die Struktur der Korpus-Annotation eingehen. Der Begriff Textverarbeitung wird dabei weitestgehend synonym mit dem Begriff Textverstehen verwendet. Das schließt maschinelle Textverarbeitung explizit aus und konzentriert sich auf die Repräsentationen und Prozesse, die relevant sind, wenn Menschen zusammenhängende Texte verstehen. Das heißt auch, dass in Abgrenzung zum folgenden Kapitel 3 die Verständlichkeit von Texten keinen oder nur sehr wenig Raum einnehmen wird. Aussagen über Verständlichkeit können zwar eventuell abgeleitet werden, spielen jedoch in den Quelltexten selbst eine eher untergeordnete Rolle.

Der Grundkonsens kognitionswissenschaftlicher Theorien zum Textverstehen besagt, dass der Prozess des Textverstehens **nicht** einfach die Übertragung von Bedeutungsinhalten vom Sender (SchreiberIn) zum Empfänger (LeserIn) darstellt. Gegeben diese sogenannte „Container-Metapher“ werden „Botschaften, sprachliche Ausdrücke, Bilder und elektronische Signale zu Containern für Bedeutungen, Ideen oder Dinge, die darin aufbewahrt, an ihren Bestimmungsort verfrachtet und anschließend wieder geöffnet werden“ (Krippendorf 1994, S. 26). Die/der EmpfängerIn muss diesen Behälter dann nur noch öffnen, den semantischen Gehalt daraus extrahieren und weiß dann genau, was die/der VerfasserIn des Textes an Bedeutung übertragen wollte. Eine Folge der Container-Metapher ist, dass nicht nur die Behälter selbst, sondern auch die darin enthaltene Information als physikalisch messbare Entität aufgefasst wird. Während das für den Container selbst (Papier bzw. die Schrift darauf, elektronisches Signal, Schall) noch offensichtlich zu sein scheint, liegt es allerdings für die zu übertragenden Inhalte selbst nicht unbedingt nahe. „Die Container-Metapher legt [...] nahe, Bedeutungen nicht viel anders zu analysieren als ein Geologe, der verschiedene Gesteinsbrocken in unterschiedliche Kästen sortiert.“ (ebd.). Krippendorf führt vor, wie diese Metapher versagen kann:

Wenn eine Botschaft Entitäten enthält, die jemand absichtlich dort hineingelegt hat, so sollte daraus folgen, daß ihr Empfänger sie genau so wieder entnimmt. Sollte er ihr aber etwas anderes entnehmen, so muß nach dieser Logik entwe-

der ein Fehler auf dem Übertragungsweg vorliegen, oder der Empfänger ist inkompetent, hinterhältig oder gar verrückt. Die Möglichkeit, daß die dieser Logik zugrundeliegende Metapher unangemessen ist, wird nur selten bei der Schuldzuweisung für solche Ungereimtheiten in Betracht gezogen. (ebd.)

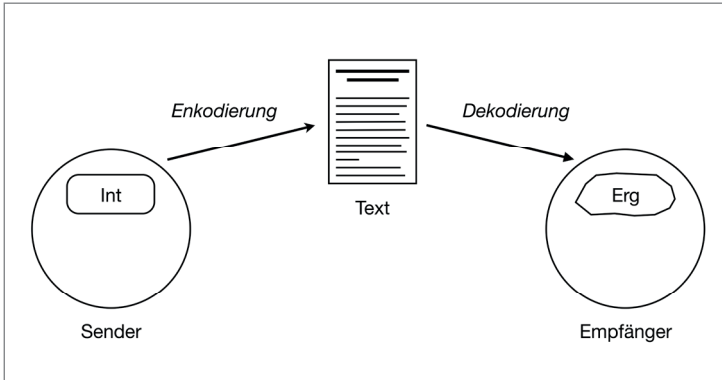


Abb. 2.1: Vereinfachtes kognitionswissenschaftliches Modell der textuellen Kommunikation (Int = intendierter Inhalt; Erg = Ergebnis des Verstehensprozesses).

Kognitionswissenschaftliche Textverstehensforschung postuliert daher ein komplexeres Bild. In Abbildung 2.1 wird die Auffassung, die in mehr oder weniger ausgeprägter Form von allen kognitionswissenschaftlichen Textverstehensmodellen vertreten wird, illustriert.

Die Grundidee ist hier, dass der Text das Produkt einer Transformation des vom Sender intendierten Inhalts darstellt. Während des Enkodierungsprozesses können Störsignale auftreten, die dazu führen, dass der Text nicht deckungsgleich mit dem intendierten Inhalt des Senders ist. Die potenziellen Quellen dieser Störsignale sind vielfältig: Editionsprozesse Dritter, Zeitdruck, Einschränkungen des Mediums oder sogar Eigenschaften der Sprache selbst, die dazu führen können, dass der Sender den eigentlich intendierten Inhalt nicht perfekt enkodieren kann. Die Textverstehensmodelle, die in diesem Abschnitt vorgestellt werden sollen, beschäftigen sich meist erst mit jenen Repräsentationen und Prozessen, die beim *Dekodierungsvorgang* beim Empfänger relevant sind. Dieser Dekodierungsvorgang wird als ein aktiver Konstruktionsvorgang gesehen, der einsetzt, wenn ein bereits fertig enkodierter Text bei der/dem EmpfängerIn ankommt. Produkt dieses Konstruktionsvorgangs ist eben nicht eine Repräsentation des Textinhalts, die deckungsgleich mit dem intendierten Inhalt des Senders ist. Das Ergebnis des Verstehensprozesses ist vielmehr etwas, das dem intendierten Inhalt optimalerweise gleichkommt, aber durch viele Einflussfaktoren während des Verste-

hensprozesses 'deformiert' werden kann. Insbesondere spielen eine Vielzahl individueller Variablen des Empfängers dabei eine Rolle.

Das kognitionswissenschaftliche Modell des Textverstehens in seiner allgemeinen Form ist inspiriert vom mathematischen Kommunikationsmodell (Shannon 1948; Weaver/Shannon 1949), das im deutschsprachigen Raum auch als Sender-Empfänger-Modell bezeichnet wird. Das Modell war ursprünglich für die mathematische Analyse der Kommunikationseffektivität in technisch realisierten Übertragungsverbindungen (beispielsweise einer Telefon- oder Telegrafenerleitung) gedacht. Es stellte sich dann heraus, dass das Modell in seiner allgemeinen Form auf eine ganze Reihe weiterer Aspekte anwendbar ist. Shannon (1948, S. 379) erkannte:

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. [Hervorh. i. Orig., S.W.]

Ein Kommunikationssystem besteht, Shannon (1948) folgend, aus fünf Teilen:

- 1) Der **Informationsquelle**, die eine Nachricht („message“, Shannon 1948, S. 380) oder eine Sequenz von Nachrichten produziert. Diese Nachricht soll an das empfangende Ende übertragen werden.
- 2) Dem **Transmitter**, der die Nachricht so bearbeitet, dass ein Signal daraus entsteht, das dafür geeignet ist, über den jeweiligen Kommunikationskanal transportiert zu werden. Im Bereich der Telefonie besteht diese Bearbeitung der Nachricht darin, dass Schalldruck in elektrische Signale umgewandelt wird.
- 3) Der **Kanal** übermittelt das Signal vom Transmitter zum Empfänger. Auf den Kanal wirkt Rauschen.
- 4) Der **Empfänger** empfängt das Signal und wandelt es gemäß den Transformationen, die der Transmitter angewandt hat, um. Dabei wird die Nachricht aus dem Signal rekonstruiert.
- 5) Der **Zielpunkt** („destination“, ebd., S. 381) ist die Person oder das Gerät, für die/das die Nachricht ursprünglich intendiert war.

Es scheint zunächst recht plausibel zu sein, die verschiedenen Komponenten des in Abbildung 2.1 illustrierten kognitionswissenschaftlichen Modells auf die Teile des Sender-Empfänger-Modells abzubilden. Das versuchen auch Shannon/Weaver (1963, S. 7), indem sie für gesprochene Sprache folgende Analogien ziehen:



In oral speech, the information source is the brain, the transmitter is the voice mechanism producing the varying sound pressure (the signal) which is transmitted through the air (the channel). [...] When I talk to you, my brain is the information source, yours the destination; my vocal system is the transmitter and your ear and associated eighth nerve is the receiver.

Doch wie weit tragen die Analogien? Betrachten wir das Konzept des Rauschens, das auf das Signal wirkt, während es über den Kanal übermittelt wird. Dieses Rauschen führt dazu, dass die Nachricht, die in Signalform beim Empfänger ankommt, verzerrt wird, Fehler oder zusätzliches Material enthält. Die empfangene Nachricht enthält somit – in informationstheoretischen Termini – zusätzliche Unsicherheit (vgl. Shannon/Weaver 1963, S. 18f.). Folgt man dieser Beschreibung, heißt das, dass die Nachricht nur auf ihrem Weg durch den Kanal 'verformt' werden kann. Die Analogie zu geschriebenem Text wäre beispielsweise das (physische) Verlorengehen von Material, also der Verlust eines Blatt Papiers oder ähnliches. Das kognitionswissenschaftliche Textverstehensmodell in seiner allgemeinen Form sagt hingegen, dass das meiste Potenzial zur Verformung der ursprünglich intendierten Nachricht im konstruktiven Verstehensprozess des Empfängers selbst liegt – also zu einem Zeitpunkt erfolgt, an dem im klassischen Sender-Empfänger-Modell die Nachricht bereits durch Rauschen verformt beim Empfänger angekommen ist. So kritisieren beispielsweise bereits Lakoff/Johnson (1980, S. 11f.) das von Reddy (1979)<sup>4</sup> „conduit metaphor“ (etwa: Leitungs-/Kanalmetapher) getaufte und an das Sender-Empfänger-Modell angelehnte Bild:

[T]he LINGUISTIC EXPRESSIONS ARE CONTAINERS FOR MEANING aspect of the CONDUIT metaphor entails that words and sentences have meanings in themselves, independent of any context or speaker. [...] The part of the metaphor that says LINGUISTIC EXPRESSIONS ARE CONTAINERS FOR MEANING entails that words (and sentences) have meanings [...] independent of contexts and speakers. These metaphors are appropriate in many situations – those where context differences don't matter and where all the participants in the conversation understand the sentences the same way. [Hervorh. i. Orig., S.W.]

Während des Textverstehens ist Kontext allerdings nahezu immer relevant, und dass die Interpretationsergebnisse verschiedener GesprächsteilnehmerInnen oder LeserInnen voneinander abweichen können, ist eine nahezu alltägliche Erfahrung. Jede Theorie des Textverstehens im Bereich der Kognitionswissenschaft beschäftigt sich mit jenen Faktoren, die dazu führen, dass unterschiedliche Individuen (oder gar identische Individuen zu unterschiedlichen Zeitpunkten oder in unterschiedlichen situativen Kontexten)

<sup>4</sup> Auch Reddy (1979) führt diese Metapher der Kritik halber ein.

ein und denselben Text anders verstehen, das heißt eine abweichende mentale Repräsentation des Gelesenen bilden. Eine der wichtigsten Variablen ist dabei unter anderem das dem Empfänger zur Verfügung stehende Welt- bzw. Hintergrundwissen.

## 2.1 Kognitionswissenschaftliche Modelle des Textverstehens

Die Rolle des Weltwissens wird unter anderem im *Construction-Integration-Modell* (vgl. Kintsch 1988, 1998) betont. Allen Modellen des Textverstehens im Bereich der Kognitionswissenschaft ist der Beschreibungsgegenstand gemein: Es werden Prozesse und Repräsentationen spezifiziert, die beim Textverstehen als relevant erachtet werden. Die Modelle unterscheiden sich dabei hauptsächlich in ihrem Schwerpunkt. So wird beim *Construction-Integration-Modell*, wie der Name schon vermuten lässt, Wert auf das Wechselspiel zweier Abläufe gelegt: Konstruktion, wobei insbesondere das Weltwissen ausschlaggebend ist und Integration, die über *Constraint Satisfaction*-Prozesse modelliert wird. Außerdem findet sich im *Construction-Integration-Modell* und dessen Vorläufern (vgl. Kintsch/van Dijk 1978) die Idee einer Drei-Ebenen-Repräsentation (Textoberfläche, Textbasis, Situationsmodell) des Textverstehens, die von der Mehrzahl der Textverstehensmodelle so übernommen wurde. Die Textoberfläche ist der tatsächliche Wortlaut des Textes. Der genaue Wortlaut ist zwar Ausgangspunkt beim Verstehen eines Textes, wird aber unter normalen Umständen meist schon im folgenden Satz wieder vergessen (vgl. Graesser/Millis/Zwaan 1997, S. 167).

- (1) a. Daniel hatte einen Termin mit Lars und war in die Stadt unterwegs.
- b. Da hatte er eine Panne.
- c. Nachdem er den Reifen geflickt hatte, fuhr er weiter.
- d. Leider kam er etwas zu spät.

Die Textbasis besteht – so u.a. bereits von Kintsch (1974) und Kintsch/van Dijk (1978) formuliert – aus Propositionen. Diese Prädikat-Argument-Konstruktionen sollen sprachunabhängig den semantischen Gehalt eines Satzes erfassen.<sup>5</sup> Beispiel 2 enthält für jeden der vier Sätze aus Beispiel 1 die darin enthaltenen Propositionen. Sicherlich wird die/der eine oder andere LeserIn dieser Arbeit mit den von mir extrahierten Propositionen nicht exakt übereinstimmen. Dies verdeutlicht bereits ein Problem des Ansatzes: Welche Textbasis soll genau zugrunde gelegt werden, wenn bereits zwei LeserInnen sich bei einem kurzen Text uneinig sind? Im Beispiel sind außerdem anaphorische

---

<sup>5</sup> Zur Notation von Propositionen siehe u.a. Kintsch/van Dijk (1978, S. 367).

Referenzen nicht aufgelöst. Es ist unklar, ob dies zulässig ist. Einerseits scheint es widersinnig, denn ein Pronomen in einer Proposition zu verwenden, scheint nicht im Sinne der Unabhängigkeit von sprachlichen Strukturen zu sein (Pronomina werden in Kintsch/van Dijk (1978) aufgelöst und Kintsch (1988, S. 165) schreibt explizit, dass nur „concepts and other propositions“ Argumente von Propositionen sein dürfen). Andererseits sind beim Auflösen von Pronomen bereits Inferenzprozesse beteiligt (siehe unten). Möchte man also eine Proposition so „roh“ wie möglich darstellen (wie sie also der Text der/dem LeserIn bereit stellt), sollte das Pronomen auch noch nicht aufgelöst sein. Weitere Unklarheit besteht in der Extraktion der Propositionen aus dem Quelltext. So setzen die Ausführungen von Kintsch (1988) zu dem Zeitpunkt ein, zu dem die Extraktion der Bedeutungseinheiten aus dem Text bereits abgeschlossen ist. Modelliert werden jene Prozesse, die wirken, wenn alle Propositionen fertig vorliegen. Wie jedoch genau der Extraktionsprozess abläuft (man könnte das semantisches oder propositionales Parsing nennen) wird nicht spezifiziert.

- (2) a. P<sub>1</sub>: HABEN(DANIEL, TERMIN); P<sub>2</sub>: MIT(P<sub>1</sub>, LARS); P<sub>3</sub>: TRANSFER(DANIEL, STADT); P<sub>4</sub>: UND(P<sub>1</sub>, P<sub>3</sub>)  
 b. P<sub>5</sub>: HABEN(ER, PANNE)  
 c. P<sub>6</sub>: FLICKEN(ER, REIFEN); P<sub>7</sub>: WEITERFAHREN(ER); P<sub>8</sub>: NACH(P<sub>8</sub>, P<sub>6</sub>)  
 d. P<sub>9</sub>: VERSPÄTEN(ER); P<sub>10</sub>: LEIDER(P<sub>7</sub>)

Die Textbasis besteht jedoch nicht nur aus den Propositionen, die explizit in einem Text angelegt sind, sondern auch aus jenen, die aus dem Hintergrundwissen der LeserInnen hinzutreten müssen, um den Text kohärent zu repräsentieren. Dabei werden laut Kintsch (1988) in der ersten Phase – dem Konstruktionsprozess – unkontrolliert Propositionen aus dem Hintergrundwissen abgerufen (siehe hierzu auch die Ausführungen zu Inferenzen später in diesem Abschnitt) und zur Textbasis hinzugefügt. Dieser unkontrollierte Abrufprozess ist der *Resonance*-Metapher sehr ähnlich, die in ihrer ursprünglichen Ausprägung auf Tulving (1974) zurückgeht und etwas später von Ratcliff (1978) aufgegriffen wird (vgl. auch McKoon/Ratcliff 1998). Produkt der Konstruktionsphase kann ein „incoherent, potentially contradictory output“ (ebd., S. 164) sein, tatsächlich wird angenommen, dass dies in den allermeisten Fällen so ist. In der nächsten Phase, dem Integrationsprozess, werden dann potenzielle Widersprüche oder Inkohärenzen durch einen *Constraint Satisfaction*-Prozess aufgelöst, so dass am Ende eine Repräsentation der Textbasis steht, die keinerlei Widersprüche oder Inkohärenzen mehr aufweist. *Constraint Satisfaction* ist ein Konzept aus dem Forschungsbereich zur Künstlichen Intelligenz, wobei zu einer Menge gegebener Beschränkungen oder Bedingungen

ein Satz an Parametern gefunden wird, der alle oder möglichst viele der Beschränkungen bestmöglich erfüllt. Im *Construction-Integration*-Modell manifestieren sich die Bedingungen oder Einschränkungen in der Menge der Propositionen in der Textbasis sowie der Verbindungen zwischen ihnen. Die Aktivationsstärken der einzelnen Propositionen sind die Parameter, die während des Integrationsprozesses angepasst werden. Das Situationsmodell als oberste Repräsentationsebene des Textverstehens gilt als analoge, bildhafte Repräsentation des Verstandenen. Zwaan/Radvansky (1998) machen anhand eines Überblicks über die Forschung zu Situationsmodellen fünf situationale Dimensionen fest, die LeserInnen beim Aufbau eines Situationsmodells „überwachen“: Zeit, Raum, Kausalität, Intentionalität und Protagonisten/Objekte (ebd., S. 167f.). Eine Erweiterung nimmt Zwaan (2004) selbst vor, indem er mit seinem *Immersed Experiencer Framework* die Theorie der Situationsmodelle um Erkenntnisse aus der Forschung im Bereich der *Embodied Cognition*<sup>6</sup> erweitert.

Die Unterscheidung in Mikro- und Makroebene oder -struktur<sup>7</sup> eines Textes ist bereits den Vorläufern des *Construction-Integration*-Modells zuzuschreiben. Kintsch/van Dijk (1978) zeigen, wie anhand sogenannter Makrooperatoren,<sup>8</sup> oder Makroregeln, aus einer Textbasis die Kernaussage oder das Diskursthema herausgearbeitet werden kann. Dabei können Propositionen aus der Textbasis gelöscht werden, die weder Interpretationsbedingung für andere Propositionen sind, noch einer globaleren Struktur untergeordnet werden können („Deletion“/„Selection“, vgl. van Dijk 1980, S. 46f.). Durch Generalisierung können Sequenzen von Propositionen durch eine allgemeinere Makroproposition ersetzt werden. Durch Konstruktion wird eine Makroproposition eingeführt, die durch eine Sequenz von Propositionen definiert wird. Hierüber würde van Dijk (1980, S. 48) beispielsweise auch ein Script (Schank/Abelson 1977) operationalisieren. Makroregeln können zudem rekursiv auf schon bestehende Makropropositionen angewandt werden, was zur Bildung von Makropropositionen auf der nächsthöheren Abstraktionsebene führt. Außerdem fand das *Construction-Integration*-Modell Anwendung bei Rechenaufgaben (Kintsch/Greeno 1985; Kintsch 1988). Implementiert wurde es unter anderem,

---

<sup>6</sup> *Embodied Cognition* steht als Schlagwort für die Einbeziehung der körperlichen Wahrnehmung in die menschliche Kognition. Im Bereich des Sprachverstehens besagt die starke These, dass notwendige Bedingung für die Repräsentation sprachlicher Bedeutung die Aktivierung neuronaler Korrelate vorheriger Wahrnehmungs- und Interaktionsereignisse (oder Kombinationen dieser) ist. Für eine ausführliche Herleitung und die programmatische Argumentation siehe Barsalou (1999).

<sup>7</sup> Der Begriff der Makrostruktur geht auf Bierwisch (1965, zit. n. van Dijk 1977) zurück.

<sup>8</sup> Diese Makrooperatoren wurden nicht zuerst dort formuliert, sondern gehen unter anderem auf van Dijk (1977) zurück. Detailliert werden sie von van Dijk (1980) herausgearbeitet.

auch für die Modellierung von Inferenzprozessen (Schmalhofer/McDaniel/Keefe 2002). Kintsch (1998) zeigt außerdem, wie sein Modell mit der computerlinguistischen Methode *Latent Semantic Analysis* (Foltz/Kintsch/Landauer 1998; Landauer/Dumais 1994; Landauer/Foltz/Laham 1998) in Verbindung gebracht werden kann.

Ein großer Teil der Forschung zum Textverstehen beschäftigt sich mit Schlussfolgerungen, die während des Verstehensprozesses gezogen werden. Singer (2007) gibt einen ausführlichen Überblick und behandelt auch die vielfältigen Methoden, die in diesem Bereich eingesetzt werden. Ohne eine Masse an Schlussfolgerungen könnten Menschen Texte vermutlich gar nicht verstehen. Dies soll kurz anhand der kleinen Geschichte aus Beispiel 1 demonstriert werden.

Es wird uns nicht sehr viel Aufwand bereitet haben, diese Geschichte zu verstehen. Und doch haben wir, während wir diese vier Sätze gelesen haben, viele Schlussfolgerungen gezogen, die nötig waren, um ihre Bestandteile in einen sinnvollen Zusammenhang zu bringen. So beziehen sich wohl alle Nennungen von 'er' auf den in Satz 1a eingeführten Protagonisten Daniel (und nicht auf Lars, der als Antezedens ebenfalls zur Verfügung steht). Aus dem ersten Satz haben wir vermutlich, obwohl es nicht explizit genannt wurde, geschlossen, dass er aufgrund des Termins in die Stadt unterwegs ist. Offenbar handelt es sich bei der Panne um einen geplatzten Reifen und da Daniel ihn flicken kann, ist er vermutlich mit dem Fahrrad unterwegs. All dies sind sogenannte Brückeninferenzen,<sup>9</sup> da sie Schlussfolgerungen darstellen, die für das weitere Verständnis des Textes notwendig sind. Vermutlich haben wir zusätzlich einige Inferenzen gezogen, die nicht unbedingt für das weitere Verständnis des Textes notwendig sind. Diese sogenannten elaborativen oder erweiternden Inferenzen<sup>10</sup> (vgl. O'Brien et al. 1988; Potts/Keenan/Golding 1988; Singer 2007) könnten in der Geschichte aus Beispiel 1 beispielsweise sein, dass Daniel nicht erst in Laufweite von seinem Bestimmungsort eine Panne hatte. Daniel hat sich vermutlich auch über die Reifenpanne geärgert. Eventuell kennt jemand die in der Geschichte genannten Protagonisten näher und inferiert, dass Daniels Verspätung nicht tragisch ist, weil Lars höchstwahrscheinlich ebenfalls ein paar Minuten später kommt. Wir sehen also: Die Menge an potenziellen Inferenzen, die man beim Verstehen eines Textes ziehen **kann**, ist

<sup>9</sup> Der Ausdruck der *bridging inference* (vgl. u.a. Graesser et al. 1997; Singer et al. 1992; Singer 2007) geht auf Haviland/Clark (1974) zurück, die beim Lesen vom Aufbau einer „bridging structure“ (ebd., S. 518) sprechen.

<sup>10</sup> Singer (2007, S. 351) umschreibt elaborative Inferenzen meines Erachtens sehr treffend als „reasonable extrapolations of explicit text ideas that stem from world knowledge“.

riesig und sicherlich größer als die Menge an Inferenzen, die man ziehen **muss**. Wie viele Inferenzen gezogen werden, ist Bestandteil einer lange geführten Debatte in der kognitionswissenschaftlichen Textverstehensforschung. Es stehen sich Vertreter der minimalistischen und der konstruktivistischen Theorie gegenüber. Erstere postulieren, dass beim Verstehen eines Textes nur jene Inferenzen gezogen werden, die zum Verständnis des Textes notwendig sind (also lediglich reine Brückeninferenzen). McKoon/Ratcliff (1992, S. 440) nehmen an, dass

the only inferences that are encoded automatically during reading are those that are based on easily available information [...], and those that are required to make statements in the text locally coherent.

Verfechter der *constructionist hypothesis* hingegen vertreten die Position, dass viel mehr Inferenzen gezogen werden, also auch eine ganze Menge an elaborativen Inferenzen. So behaupten Graesser/Singer/Trabasso (1994, S. 371):

Readers potentially generate a rich variety of inferences when they construct a referential situation model of what the text is about. The proposed constructionist theory specifies that some, but not all, of this information is constructed under most conditions of comprehension.

Am wahrscheinlichsten ist wohl eine Mischform der beiden Ansätze. So unterscheidet Singer (2007, S. 351ff.) zwischen „highly constrained elaborative inferences“ und „predictive inferences“. Erstere sind dabei nicht unabdingbar für das Verstehen eines Textes, werden jedoch vom Inhalt sehr nahe gelegt. Wenn an Thanksgiving ein Vogel zerlegt wird, ist das zumindest in den USA ein sehr starker Hinweis darauf, dass dieser Vogel ein Truthahn ist. Es liegt nahe, dass LeserInnen relevante Eigenschaften von Objekten inferieren und diese in ihre Diskursrepräsentation einbauen. So konnten McKoon/Ratcliff (1988) zeigen, dass Testsätze wie ‘Tomatoes are red.’ dann schneller und eher korrekt verifiziert werden, wenn die Farbe der reifen Tomaten im vorhergehenden Kontext relevant erschien (beispielsweise weil sie Teil eines Stilllebens sind und ein Künstler mehrere Tage nach einer passenden Farbe gesucht hatte). Sätze wie ‘Tomatoes are round.’ hingegen wurden nur dann schneller und besser verifiziert, wenn der Kontext die Rundheit der Tomaten betonte (beispielsweise weil sie zuvor über den Boden gerollt wurden). Offenbar inferieren LeserInnen auch das Thema eines Satzes, wenn es nie explizit genannt wurde. So zeigten Till/Mross/Kintsch (1988), dass nach dem Satz „The townspeople were amazed to find that all the buildings had collapsed except the mint [= Münzprägestalt, S.W.]“ das konnotierte Wort „earthquake“ schneller verifiziert wurde als ein unrelatiertes Kontrollwort. Es ist zwar nicht völlig



ausgeschlossen, dass das dieser Arbeit zugrunde liegende Korpus für Inferenz-Fragestellungen aufbereitet werden kann, trotzdem wird diese Arbeit keine Stellung zum Theoriestreit Konstruktionisten vs. Minimalisten beziehen. Wir werden uns nun auf eine ganz bestimmte Art der Inferenz beschränken.

Dabei handelt es sich um eine spezielle Art der Brückeninferenz und eine Art der Schlussfolgerung, bei der es relativ unstrittig ist, dass sie beim normalen Lesen von Texten üblicherweise gezogen wird: die anaphorische Brückeninferenz. Während einer anaphorischen Brückeninferenz werden anaphorische Ausdrücke an ihr jeweiliges Antezedens gebunden. Diesem Vorgang soll hier ein eigener Abschnitt gewidmet werden, da er in der vorliegenden Arbeit ein zentrales Thema ist.

## 2.2 Anaphernauflösung

Die Auflösung von Anaphern ist eine zentrale Kompetenz beim Verstehen von zusammenhängenden Texten, denn nur über die korrekte Auflösung von referenziellen Bezügen können wir verstehen, wer etwas tut, mit wem oder was es getan wird und welche Gründe und Folgen bestimmte Ereignisse haben. Die Forschung in diesem Bereich ist deshalb auch sehr vielfältig. Ausgesprochen strukturelle Ansätze sollen in dieser Arbeit eher außen vor gelassen werden. Das soll natürlich die Relevanz von Ansätzen wie beispielsweise der entsprechenden Ausschnitte von Chomsky (1981) nicht schmälern. Die Wirkung der Bindungsprinzipien ist natürlich ein nicht zu unterschätzender Faktor bei der Anaphernauflösung. Sie sind aber eher im Sinne eines 'ganz oder gar nicht' als im Sinne eines 'einfacher oder schwieriger' zu interpretieren. Bindungsprinzipien machen eine binäre Aussage über die Möglichkeit einer Referenzanbindung. Interessanter für die vorliegende Arbeit sind jedoch Textverstehensmodelle, die explizit die kognitiven Prozesse beim Auflösen von rückwärtsreferenziellen Ausdrücken spezifizieren und somit Vorhersagen für einfachere oder komplexere Verarbeitung machen.

Ein einflussreiches Modell, das solche Ableitungen zulässt, ist das *Scenario Mapping and Focus*-Modell (vgl. Garrod/Sanford 1994; Gernsbacher/Foertsch 1999; Sanford/Garrod 1981, 1994, 1998), das sich aus dem Interesse von Garrod/Sanford (1977) an der Auflösung von Anaphern entwickelte. Den unterschiedlichen Schwerpunkt im Vergleich zum *Construction-Integration*-Modell fassen Gernsbacher/Foertsch (1999, S. 288) treffend zusammen:

For Kintsch and van Dijk, the primary question of interest during processing is "How does this new element change the scenario I am constructing?" For Sanford and Garrod, the primary question of interest is „Does this new element refer to something mentioned previously in the text, and if so, what?“

Eine zentrale Aussage des *Scenario Mapping and Focus*-Modells ist, dass im Text erwähnte Personen, Dinge, Ereignisse oder Relationen im Diskursfokus stehen können oder nicht (vgl. auch Grosz 1977). Der Diskursfokus wird dabei in expliziten und impliziten Fokus unterteilt. Jene Information, die Thema des aktuellen Satzes ist, ist hochaktiviert und augenblicklich abrufbar. Sie befindet sich im expliziten Fokus, also im Zentrum des Interesses. Im impliziten Fokus stehen all jene Informationen, die schwächer aktiviert sind. Diese Informationen können deshalb schwächer aktiviert sein, weil sie nicht mehr aktuelles Thema des Textes sind. Dass sie überhaupt noch aktiviert sind, liegt daran, dass sie gerade erst erwähnt wurden – wie in klassischen Thema-Rhema-Abfolgstrukturen (vgl. Dürscheid 2007, S. 181ff.) denkbar. Eine andere Möglichkeit ist, dass die Information im impliziten Fokus mit dem aktuellen Thema (das im expliziten Fokus steht) assoziativ eng verbunden ist. Information im impliziten Fokus ist zwar etwas weniger aktiviert als jene im expliziten Fokus, aber aktiver als Information, die gänzlich im Hintergrund steht und kann daher auch schneller abgerufen werden. Sanford/Garrod (1998) führen aus, dass im impliziten Fokus das augenblicklich relevante Szenario (Aktionen, Ereignisse usw.) verfolgt wird. Szenarios stellen *role slots* zur Verfügung, die durch *tokens*, die im expliziten Fokus stehen, gefüllt werden (vgl. ebd., S. 162f.). Beim Verstehen des Satzes „Daniel fährt nach Freiburg“ würde das Szenario 'irgendwo hinfahren' im impliziten Fokus stehen und unter anderem die *role slots* 'FahrerIn' und 'Fahrziel' zur Verfügung stellen. Auf diese werden die im expliziten Fokus stehenden *tokens* Daniel und Freiburg entsprechend abgebildet („gemappt“, daher der Name des Modells). Mit dieser Konzeption des Textverstehens berufen sich Sanford/Garrod (1998) u.a. auf die Arbeiten von Ericsson/Kintsch (1995) zum *long-term working memory*, denn sie sehen den impliziten Fokus als den „currently active part of long-term memory“ (Sanford/Garrod 1998, S. 163). Damit ist die Kapazität des impliziten Fokus grundsätzlich sehr groß, wohingegen die Kapazität des expliziten Fokus eingeschränkt ist. Außerdem greifen auch sie auf die Idee der *scripts* von Schank/Abelson (1977) zurück. In der ursprünglichen Ausformulierung des *Scenario Mapping and Focus*-Modells (vgl. Sanford/Garrod 1981) waren diese prototypischen Ereignisabläufe (wie beispielsweise ein Restaurantbesuch oder eine Gerichtsverhandlung) die häufigsten Beispiele für Szenarien. Anderson/Garrod/Sanford (1983) konnten zeigen, dass ein instantiiertes Szenario dazu führen kann, dass prototypisch relevante Akteure (KellnerInnen während des Restaurantbesuchs, RichterInnen während einer Gerichtsverhandlung usw.) ohne vorherige Nennung schnell verfügbar („accessible“) waren. Sie stehen im impliziten Fokus. Sobald ein Zeitsprung in der Geschichte darauf hindeutet, dass das Szenario nicht länger relevant ist, verschwinden die sogenannten „scenario-dependent entitites“ (ebd., S. 427) aus dem impli-

ziten Fokus und fallen somit in ihrer Verfügbarkeit ab – ein Muster, das für die allseits relevanten Hauptcharaktere einer Geschichte nicht zu beobachten war.

Graesser et al. (1997, S. 174) beschreiben den Diskursfokus als eine Art „mental camera that zooms in on particular characters, objects, actions, events, and spatial regions“. Das ist eine gut vorstellbare Verdeutlichung des Konzeptes, doch sollte es dann auch möglich sein, diese mentale Kamera zu steuern. Zwei wichtige linguistische Fokusingstrumente („focusing devices“, Garrod/Freudenthal/Boyle 1994, S. 40) sind dabei die Neuigkeit einer Diskursentität („recency of mention“, ebd.) und die vorherige Topikalisierung. Dabei werden Pronomen vorzugsweise eingesetzt, um auf Entitäten zu referieren, die sich im expliziten Fokus befinden und aktuell das Topic des Satzes sind. Koreferenten eindeutigerer Formen (beispielsweise beschreibender Nominalphrasen) werden eher im impliziten Fokus gesucht (vgl. Anderson 1983; Garrod et al. 1994; Gernsbacher/Foertsch 1999; Sanford/Garrod 1989). Diese Aussagen sind so zu verstehen, dass die eben genannten Standardannahmen zwar verletzt werden dürfen, die Referenzauflösung aber länger dauern und in Einzelfällen sogar scheitern kann. Die Liste der Fokusingstrumente, die bereits in den Arbeiten von Sanford und Garrod genannt werden, ist lang – und die weiteren Modelle, die noch eingeführt werden sollen, erweitern diese Liste sukzessive. So führen Sanford/Garrod (1989) beispielsweise noch die Einführung mit einem Eigennamen, Agens-Status oder eine dominante Rolle als fokussierende Faktoren an (sie argumentieren außerdem, dass dies linguistische Eigenschaften sind, die stark auf die Hauptfigur in Narrationen hinweisen). So prominent seine Position im Modell auch immer sein mag, der Diskursfokus ist nicht der einzige Faktor, dem Sanford und Garrod maßgeblichen Einfluss auf den Prozess der Referenzauflösung zugestehen. Wie eben schon angedeutet, ist die sprachliche Form und Explizitheit der Anapher ebenfalls ausschlaggebend. Garrod et al. (1994, S. 41) führen drei häufige Formen in der Reihenfolge ihrer Explizitheit an:

**Pronomina:** Gernsbacher (1989) hält fest, dass Pronomen durch ihre Unterspezifiziertheit bzw. Nicht-Explizitheit oft zu mehreren potenziellen Antezedenzen passen. Die Informationen, die es der/dem LeserIn erlauben, die Anapher trotzdem zu binden, müssen oft erst erschlossen werden. Sie kommen definitiv nicht vom Pronomen selbst. Garrod et al. (1994) folgen ihr in dieser Argumentation.

**Definite Nominalphrasen:** Definite Nominalphrasen haben mehr semantischen Gehalt als Pronomen. Ein gebräuchliches Muster anaphorischer Referenz über definite Nominalphrasen ist die Verwendung des Oberbegriffs (in Bezug zum Antezedens). Beispiel 3 aus dem in dieser Arbeit vorgestellten

*Freiburg Legalese Reading Corpus* verdeutlicht dies. Dabei referiert „Beteiligungsrecht“ auf den „konstitutiven Parlamentsvorbehalt“ und ist Hyperonym von selbigem.

- (3) a. Die Antragsstellerin beruft sich auf den [konstitutiven Parlamentsvorbehalt]<sub>i</sub>.
- b. [Dieses Beteiligungsrecht]<sub>i</sub> hat ein hohes Gewicht, weil die Bundeswehr ein Parlamentssheer ist.

In der Besprechung weiterer Modelle und Studien, die Aussagen zur Auflösung von Anaphern treffen, werden noch mehr relevante Eigenschaften von Nominalphrasen-Referenzen herausgearbeitet.

**Wiederholte Namen:** Die Nennung des Eigennamens lässt die Figur, auf die referiert wird, stärker in der Diskursrepräsentation hervortreten als eine Rollenbeschreibung. Dabei kann es unter Umständen egal sein, wenn der Eigenname (beispielsweise „Frieder“) auf dieselbe Person referiert wie die Rollenzuschreibung (beispielsweise „der Philosoph“). Sanford/Moar/Garrod (1988) konnten zeigen, dass Sätze mit pronominalen Referenzen auf Figuren, die über Eigennamen eingeführt wurden, schneller gelesen werden als solche, in denen mit Rollenbeschreibungen referiert wird. Benennung scheint also (zumindest für narrativ geprägte Texte, in denen Hauptfiguren sehr wichtig sind) ein wichtiger Faktor in Sachen Fokuskontrolle zu sein.

Die laut dem *Scenario Mapping and Focus*-Modell dritte Komponente, die die Auflösung von referenziellen Ausdrücken beeinflusst, ist der sogenannte „pragmatic inference constraint“ (vgl. Garrod et al. 1994; Gernsbacher/Foertsch 1999). Dieser besagt, dass die/der LeserIn globale Kohärenz anstrebt. Das heißt, dass eine von den anderen Faktoren (Diskursfokus und Form der Anapher) nahegelegte Anbindung dann nicht aufgebaut werden kann, wenn dies der globalen Kohärenz der Geschichte widersprechen würde. Beispiel 4 verdeutlicht, wie die Anbindung eines Pronomens durch ein alternatives Satzende (Satz 4b) geändert werden kann (Beispiel angelehnt an Garrod et al. 1994, S. 39).

- (4) [Boris]<sub>1</sub> wollte [seinem Freund]<sub>2</sub> etwas Geld leihen.
- a. Weil er<sub>1</sub> gerade ziemlich pleite war, konnte er<sub>1</sub> sich das aber leider nicht leisten.
- b. Weil er<sub>2</sub> gerade ziemlich pleite war, brauchte er<sub>2</sub> das Geld wirklich dringend.

Die Anbindung des Pronomens *er* in Fortsetzung 4a an Boris sollte relativ problemlos funktionieren, da der Koreferent mit einem Eigennamen eingeführt wurde, in Subjektposition steht und die Agensrolle innehat. Boris steht

somit im expliziten Fokus, was die Referenzierung über ein Pronomen legitimiert. Fortsetzung 4b bereitet uns eventuell etwas mehr Schwierigkeiten, weil das Pronomen sich auf Boris' Freund bezieht – ein Schluss, der jedoch nur aufgrund des gesamten Satzes getroffen werden kann. Die pragmatischen Inferenzbeschränkungen stammen hier also von Kombinationen von Begebenheiten in der Welt (Pleite sein und kein Geld verleihen können oder Pleite sein und Geld benötigen).

Zusammenfassend: Das *Scenario Mapping and Focus*-Modell nimmt drei grundlegende Faktoren an, die die Auflösung von referenziellen Ausdrücken beeinflussen: Diskursfokus, Form der Anapher und pragmatische Inferenzconstraints. Eine Diskursentität kann im expliziten oder impliziten Fokus stehen, wobei der explizite Fokus einer strengen Ressourcenbeschränkung unterworfen ist.

Ariel (1988, 2000, 2001) konzentriert sich in ihrer *Accessibility Theory* ganz auf referenzielle Ausdrücke. Die Hauptthese der *Accessibility Theory* ist, dass die Zugänglichkeit<sup>11</sup> eines Diskurskonzepts (eines Referenten) mit der Form der Anapher korreliert („form-function correlation“, Ariel 2001, S. 32). Diese Korrelation ist von drei sich teilweise überlappenden Kriterien bestimmt: 1) **Informativität** („informativity“), d.h. der Menge an lexikalischer Information, die der anaphorische Ausdruck trägt, 2) **Rigidität** („rigidity“), d.h. der Möglichkeit, dass aufgrund der Form nur ein Referent gebunden werden kann und 3) **Abschwächung** („attenuation“), d.h. des phonologischen Ausmaßes der Anapher. Je informativer, konsistenter und unabgeschwächter die Form der Anapher ist, desto geringer zugänglich ist der Diskursreferent, auf den die Anapher referiert. Die Verwendung eines bestimmten referierenden Ausdrucks ist, so Ariel (2001, S. 29f.), eine Instruktion an die/den EmpfängerIn, eine bestimmte Information abzurufen, deren augenblickliche Zugänglichkeit in der Diskursrepräsentation zur Form des referierenden Ausdrucks passt. Ausdrücklich betont Ariel (ebd., S. 31), dass weniger der physikalische Kontext der Sprachsituation, sondern vielmehr die „discourse world“ (ebd.) ausschlaggebend dafür ist, wie hoch die Zugänglichkeit einer mentalen Repräsentation ist:

I have argued that it is the discursal rather than the physical salience of the entities involved which determines the degree of accessibility assigned to particular mental representations [...]. Although the physical context does affect the discourse model of the speakers, mental representations are a direct product of our discourse model only. (ebd.)

<sup>11</sup> *Zugänglichkeit* scheint mir eine angebrachte Übersetzung für den Begriff „accessibility“ zu sein. Auch Consten/Schwarz-Friesel (2009, S. 269) und Götz-Votteler (2008, S. 34) wählen diese Übersetzung.

Laut Ariel (2001, S. 32f.) definiert sich die Zugänglichkeit einer Diskursentität 1) über die Salienz der Diskursrepräsentation und 2) über die Beziehung zwischen Antezedens und referenziellem Ausdruck. Punkt 1 zieht unvermeidlich die Frage nach sich, was die Salienz einer Diskursrepräsentation beeinflusst. Hier werden globale Diskursthemen als hochsalient eingestuft. Lokale Diskursthemen sind ebenfalls salient, aber nicht mehr so stark. Nicht-Themen sind überhaupt nicht salient und somit auch nicht zugänglich. Auf Dialoge bezogen gesteht Ariel (2001) außerdem der/dem SprecherIn und der/dem AdressatIn einen hohen Grad an Zugänglichkeit zu, nicht aber dritten Personen. Sie bezieht sich außerdem auf das bereits oben angesprochene Konzept der „scenario-dependent entities“ (vgl. Anderson et al. 1983), also jener Diskursreferenten, die von bestimmten Situationsprototypen quasi automatisch bereitgestellt werden (KellnerInnen in Restaurants, ÄrztInnen in Krankenhäusern usw.). Ariel (2001, S. 33) bezeichnet diese als „frame induced entities“. Mit Bezug auf Arbeiten von u.a. O'Brien/Albrecht (1991) postuliert Ariel (2001) außerdem, dass Wettbewerber um die Rolle als Antezedens einen Einfluss auf die Salienz haben: Je mehr potenzielle Antezedenzen existieren, desto weniger zugänglich ist jedes einzelne von ihnen. Ohne diesen explizit zu erwähnen, nimmt die *Accessibility Theory* hier Bezug auf den *fan effect* (vgl. Anderson 1974) und die ebenfalls darauf aufbauende Theorie des *cue-based parsing* (vgl. Lewis/Vasishth 2005; Van Dyke 2007), die freilich nicht für die Anaphernauflösung bestimmt ist, deren Prinzipien aber darauf übertragbar sein sollten.

Dieser letzte Punkt, die Anzahl der potenziellen Wettbewerber, befindet sich meines Erachtens auf der Grenze zwischen den beiden Faktoren, die Ariel (2001) für die Zugänglichkeit der Diskursrepräsentation verantwortlich macht (Salienz der Diskursrepräsentation und die Antezedens-Anaphern-Beziehung, siehe oben). Im Grunde könnte man diesen Faktor ebenso gut der zweiten Kategorie zuschreiben, denn es geht letztlich darum, wie viele Antezedenzen zu den Abrufsignalen des anaphorischen Ausdrucks passen; also sind hier eindeutig sowohl das Antezedens als auch die Anapher selbst beteiligt. Ein Punkt, der eindeutig in die zweite Kategorie fällt, ist jedoch die Distanz zwischen Anapher und der letzten Nennung des Antezedens: „The larger the distance separating different mentions of the same mental entity, the lower the degree of accessibility with which the mental representation is entertained.“ (Ariel 2001, S. 33). Distanz ist dabei nicht unbedingt in Wörtern zu messen. Absatzgrenzen und abgeschlossene Zwischenhandlungen können ebenfalls Distanz erzeugen.



Ariel (2001) geht weiterhin auf die zahlreichen Interaktionen ein, die zwischen den eben genannten Faktoren bestehen. Diese können an dieser Stelle nicht alle referiert werden. Generell gilt, dass eine Kombination von Zugänglichkeitsfaktoren meist vorhersagemächtiger ist als einzelne Faktoren.

Ariel (2001) legt Wert darauf, dass Zugänglichkeitsfaktoren nicht exklusiv dafür verantwortlich sind, welcher referenzielle Ausdruck selektiert wird. Sie nennt einige Beispiele:

- Relevanzbasierte Auswahlkriterien können bei der Auswahl von referenziellen Ausdrücken helfen, die bezüglich ihrer Zugänglichkeit auf einer Stufe stehen.
- Sozio-kulturelle Faktoren können dazu führen, dass auf bestimmte gesellschaftliche Gruppen, so beispielsweise gezeigt für Frauen und Minderheiten (vgl. Mulkern 1996, S. 247f.), mit Ausdrücken referiert wird, die eine höhere Zugänglichkeit nahelegen (im o.g. Beispiel Vornamen) als laut der Zugänglichkeitsfaktoren erwartbar. Zum angesprochenen Beispiel zeigte Ariel (1990), dass Frauen und Araber in israelischen Zeitungsartikeln im Vergleich zu nicht-arabischen Männern eher mit Vornamen referenziert werden bzw. wurden. „Here the difference in conventional choice of single name for the different groups seems to convey the difference in the level of power or respect they have in society, rather than a difference in the degree of intimacy.“ (Mulkern 1996, S. 248).
- Stellvertretendes Sprechen kann dazu führen, dass der Zugänglichkeitsstatus des Antezedens in der mentalen Diskursrepräsentation der/des anderen „mitsimuliert“ wird. Ebenso scheint jenen Personen, deren Perspektive übernommen wird, eine höhere Zugänglichkeit zugesprochen zu werden (vgl. Ariel 1990, S. 203f.).
- Demonstrativa in Kombination mit einer Nominalphrase können prädikativ (im Gegensatz zu identifizierend) eingesetzt werden. Auf diese Weise eingesetzte Ausdrücke können die beim Empfänger bestehende Repräsentation der referierten Entität modifizieren. In diesem Falle werden Ausdrücke benutzt, die eine niedrigere Zugänglichkeit suggerieren als tatsächlich gegeben ist (vgl. Ariel 2001, S. 42).

All diese Faktoren, die über die in der *Accessibility Theory* beschriebenen Faktoren hinaus Einfluss auf die Wahl des referenziellen Ausdrucks haben, zeigen, 1) dass Zugänglichkeit nicht auf ein einziges linguistisches Prinzip heruntergebrochen werden kann und 2) dass die *Accessibility Theory* für sich nicht in Anspruch nehmen kann, exklusiv diese Wahl und deren Interpretation erklären zu können (vgl. ebd.). Diese Arbeit kann nicht allen von der *Accessibi-*

*lity Theory* eingeführten Konzepten gerecht werden. Außerdem kann auf kritische Stimmen nur verwiesen werden (vgl. z.B. Reboul 1997, die teilweise recht harsche Kritik anbringt). Der recht ausführliche Überblick, der hier gegeben wurde, zeigt, welche grundlegenden Prinzipien postuliert werden, dass Interaktionen zwischen diesen Prinzipien herrschen können und welche nicht von der Theorie abgedeckten Faktoren Einfluss nehmen können.

Almor (1999) nimmt unter anderem auch Bezug auf Ariels *Accessibility Theory* und stellt diese in Beziehung zur Relevanztheorie (vgl. Sperber/Wilson 1995; Wilson/Sperber 2005), die in ihrer allgemeinsten Form behauptet, dass in der Sprachverarbeitung normalerweise ein Gleichgewicht zwischen kognitiver Funktion und Verarbeitungskomplexität angestrebt wird. Sollten SprecherInnen dieses Gleichgewicht stören, so ist die Störung selbst als informationstragendes Element zu verstehen. Genau diesen Grundgedanken würde Ariel ebenfalls unterschreiben, denn die *Accessibility Theory* kann als Ausgestaltung dieses Gedankens für den Einsatz und die Verarbeitung von Anaphern gesehen werden. Almor (1999) möchte nun unter anderem mit seiner *Informational Load Hypothesis* dort einsetzen, wo sowohl Sperber/Wilson (1995) als auch Ariel (1990) seiner Meinung nach aufhören: „[N]either theory provides an explicit account of the psychological mechanisms that underlie cognitive function or computational cost.“ (Almor 1999, S. 749). Der Prozess der Anaphernverarbeitung stellt, so Almor, einen Optimierungsprozess dar. Verarbeitungskosten sollten immer eine Diskursfunktion erfüllen. Anders ausgedrückt: Der Verarbeitungsaufwand von Anaphern sollte ihre funktionelle Berechtigung widerspiegeln. Konkret heißt das: Anaphern sollten dann schneller gelesen werden, wenn die Kosten, die mit der Anapher verbunden sind, berechtigt sind. Mit dieser Argumentation bezieht sich Almor (1999) auf die Quantitätsmaxime (vgl. Grice 1975, S. 45f.), die besagt, dass ein Beitrag so informativ wie nötig, jedoch nicht informativer sein sollte – gegeben den aktuellen Stand des Austauschs. Die *Informational Load Hypothesis* kontrastiert dabei nicht nur pronominale mit über Nominalphrasen realisierte Referenzen. Auch innerhalb der Gruppe von Nominalphrasen wird unterschieden nach Diskursfunktion und/oder Kosten. Beispiel 5, das an Almor (1999, S. 749) angelehnt ist, verdeutlicht dies.

- (5) Ein Rotkehlchen fraß eine Frucht.
  - a. Das Rotkehlchen schien sehr zufrieden.
  - b. Der kleine nasse Vogel schien sehr zufrieden.
  - c. Der Vogel schien sehr zufrieden.
  - d. Es schien sehr zufrieden.

Die Referenzen in den Fortsetzungen unterscheiden sich in ihrer Allgemeinheit/Spezifität und sind somit mit verschiedenen Kosten verbunden. Fortsetzung 5d ist allgemeiner als Fortsetzung 5c, welche wiederum allgemeiner ist als die Fortsetzungen 5a und 5b. Fortsetzung 5b hingegen enthält die einzige Bezugnahme, die neue Information hinzufügt.<sup>12</sup> Auf diesen Unterscheidungscharakter zwischen verschiedenen Arten von referierenden Nominalphrasen legt Almor besonderen Wert. Almor (1999, S. 749) argumentiert, dass nur im Rahmen der *Informational Load Hypothesis* zwischen verschiedenen referierenden Nominalphrasen unterschieden werden könne. Das ist meines Erachtens nicht haltbar. Ariel (2001, S. 32) führt das oben beschriebene Konzept der Abschwächung („attenuation“) ein. Zumindest auf dieser Dimension lassen sich auch in Ariels Terminologie die Fortsetzungen 5a bis 5c unterscheiden. Auch das Kriterium der Informativität würde diesen Unterschied erfassen. Allerdings konzentriert sich Ariel (1988, S. 82) bei Markern mit höherer Informativität nicht auf das Hinzufügen von Information, sondern darauf, ob sie „good search guide[s]“ für den Abruf des Antezedens sind: „The more information it [= the Accessibility marker] imparts the better retriever it is.“

Die beiden Hauptfunktionen von Anaphern sieht Almor darin, das Antezedens zu identifizieren und dessen Diskursrepräsentation mit Informationen anzureichern. Die Verarbeitungskomplexität anaphorischer Ausdrücke soll dabei anhand dreier interagierender Faktoren vorhergesagt werden: dem Diskursfokus, der Menge an Information, die die Anapher neu hinzufügt, sowie dem *informational load* des Anapher-Antezedens-Paars. Die Interaktion dieser drei Faktoren stellt Almor (1999, S. 753) wie folgt dar:

The ILH (= Informational Load Hypothesis) implies that if an anaphor does not add new information about its referent, then the smaller the informational load of the anaphor-antecedent pair, the easier it is to process [...] as long as this informational load is sufficient for identifying the antecedent.

Dementsprechend würde ein anaphorischer Ausdruck dann schwieriger zu verarbeiten sein, wenn 1) das Antezedens im Fokus steht, 2) über eine Anapher referenziert wird, die einen hohen *informational load* hat und die 3) keine neue Information über das Antezedens hinzufügt. Dieses Beispiel deutet schon an, dass eine Art Rechtfertigungsmechanismus für Anapher-Antezedens-Paare mit hohem *informational load* greift. Dieser Rechtfertigungszwang wird abgeleitet aus der oben angesprochenen Quantitätsmaxime (vgl. Grice 1975, S. 45f.).

<sup>12</sup> Es sei denn, es soll mit der Referenz „der Vogel“ (Fortsetzung 5c) auf generelle Eigenschaften der Kategorie Vogel verwiesen werden. In diesem Fall würde auch diese Referenz neue Information aktivieren (vgl. Almor 1999, S. 749).

Wie wird der *informational load* eines Anaphern-Antezedens-Paares operationalisiert? Almor (1999, S. 751f.) definiert dies unter Zuhilfenahme des Konzepts der konzeptuellen Distanz („C-difference“), dem Abstand zwischen den semantischen Repräsentationen der Anapher und dem Antezedens. Exemplifiziert wird diese konzeptuelle Distanz an Kategoriennamen und Instanzen der Kategorien (Taxonomie-Verhältnisse). Konkret: Gegeben sei das Antezedens  $N$  'der Vogel'. Referiert nun der weniger generelle (ergo spezifischere) anaphorische Ausdruck  $P$  (beispielsweise  $P_1$ : 'das Rotkehlchen' oder  $P_2$ : 'das verletzte Rotkehlchen') auf dieses Antezedens, liegt die *C-difference*  $CD(P, N)$  im positiven Bereich. Außerdem gilt:  $CD(P_1, N) < CD(P_2, N)$ . Die *C-difference* kann auch negative Werte annehmen. Das ist der Fall, wenn  $P$  genereller ist als das Antezedens  $N$ . Im Beispiel könnte dies 'das Tier' ( $P_3$ ) oder 'die Kreatur' ( $P_4$ ) sein. Dabei gilt hier:  $CD(P_3, N) > CD(P_4, N)$ .  $CD(P_4, N)$  liegt also weiter im negativen Bereich als  $CD(P_3, N)$ .  $CD(P, N) = 0$  kann nur im Falle einer wiederholenden Anapher gelten.  $|CD|$  steigt nicht nur mit der steigenden Anzahl zu traversierender Taxonomiebaum-Knoten zwischen  $P$  und  $N$ , sondern kann auch auf einer Ebene variieren. Almor (1999, S. 751) schlägt vor, die konzeptuelle Distanz über Verifikationszeiten für Aussagen in der Form ' $N$  ist ein  $P$ ' bzw. ' $P$  ist ein  $N$ ' zu operationalisieren. Es würde demnach erwartet, dass  $CD('die Amsel', 'der Vogel') < CD('die Riesentrappe', 'der Vogel')$ . Auf einer Taxonomie der Vögel befinden sich Amseln und Riesentrappen zwar gleich weit entfernt von Vögeln, jedoch würde die Aussage 'die Amsel ist ein Vogel' in Mitteleuropa höchstwahrscheinlich schneller verifiziert werden als 'die Riesentrappe ist ein Vogel'.

Anhand der *C-difference* zwischen Anapher und Antezedens kann nun der *informational load* bestimmt werden. Im Grunde findet eine direkte Übertragung der konzeptuellen Distanz statt: Der *informational load* ist eine monoton steigende Funktion der konzeptuellen Distanz. Demnach muss für jede zwei Anaphern-Antezedens-Paare gelten: Wenn  $CD(P_1, N_1) > CD(P_2, N_2)$ , dann ist der *informational load* (IL) für Paar 1 ebenfalls größer als für Paar 2.

Auf zwei grundlegende Einschränkungen weist Almor (1999, S. 752) selbst hin: 1) Wenn sowohl Anapher als auch Antezedens unterschiedlich sind ( $P_1 \neq P_2$  und  $N_1 \neq N_2$ ) kann anhand der *Informational Load Hypothesis* in ihrer augenblicklichen Form der *informational load* zweier Anapher-Antezedens-Paare nicht verglichen werden. 2) Außerdem können nicht systematisch Differenzhypothesen für Anaphern abgeleitet werden, die neue Information über den Koreferenten hinzufügen. Der Grund dafür liegt in einer fehlenden Quantifizierung der Menge an neuer Information. Und wie in 1) eben gezeigt, ist der *informational load* lediglich vergleichend, aber nicht absolut quantifiziert. Ein kurzes Beispiel zeigt das Problem. Es sei  $N = 'ein Vogel'$ ,  $P_1 = 'die Riesentrappe'$ ,  $P_2 = 'die Amsel'$ . Dann gilt:  $CD(P_1, N) > CD(P_2, N)$ . Es würde demnach erwartet, dass die Aussage 'die Riesentrappe ist ein Vogel' länger verifiziert wird als 'die Amsel ist ein Vogel'. In der Realität ist es jedoch umgekehrt: 'die Amsel ist ein Vogel' wird schneller verifiziert als 'die Riesentrappe ist ein Vogel'.

trappe' und  $P_2 = \text{'die Amsel'}$ . In diesem Fall könnte der *informational load* mit dem Hinzufügen neuer Information kovariieren. Es würde gelten:  $IL(P_1, N) > IL(P_2, N)$ , aber ebenso könnte die Identifizierung von  $N$  als eine Riesentrappe ( $P_1$ ) mehr neue Information hinzufügen als die Identifizierung als Amsel ( $P_2$ ). Die *Informational Load Hypothesis* macht aber, wie zuvor ausgeführt, gegensätzliche Vorhersagen für den Effekt der beiden Faktoren auf die Verarbeitungskomplexität der Anapher: „[H]igher informational load may make the anaphor harder to process as long as the antecedent is identifiable, but more new information may make it easier to process because the high informational load would be functional.“ (ebd.). Eine Vorhersage über den kombinierten Effekt ist damit nicht mehr möglich. Dazu wäre eine genaue Messmethode nötig, die *informational load* und die Menge neu hinzugefügter Information unabhängig voneinander quantifizieren müsste.

Die *Informational Load Hypothesis* kann als Erweiterung der *Accessibility Theory* in eine bestimmte Richtung gesehen werden. Über das Konzept des *informational loads* sollen Antezedens-Anapher-Paare bezüglich der Aufwändigkeit bei ihrer Verarbeitung vergleichbar gemacht werden. Die Grundidee ist, dass die Verarbeitung des referenziellen Ausdrucks dann schwierig ist, wenn ein hoher *informational load* nicht gerechtfertigt ist. Einfacher sollten referenzielle Ausdrücke dann sein, wenn entweder ein hoher *informational load* gerechtfertigt ist oder ein niedriger *informational load* (das Extrembeispiel sind Pronomen) in Kontexten vorliegt, in denen die Diskursentität, auf die referenziert wird, sehr salient ist und keine zusätzliche Information hinzugefügt werden soll. Im Folgenden wird noch eine Auswahl an Studien vorgestellt, die sich allesamt mit der Auflösung von Anaphern beschäftigen und dabei unterschiedliche Schwerpunkte setzen. Dabei wird auf bereits eingeführte Modelle und gemeinsame Modellvorstellungen Bezug genommen.

O'Brien/Duffy/Myers (1986) konzentrieren sich auf referierende Nominalphrasen und argumentieren anhand dreier Experimente, dass referierende Nominalphrasen die Arbeitsgedächtnisrepräsentation des entsprechenden Antezedens reinstanciieren. Sie schließen dies aus kürzeren Reaktionszeiten auf eine *probe recognition task*<sup>13</sup> (Experiment 1) sowie verkürzten Benennungslatenzen des Antezedens (Experiment 2 und 3) im Vergleich zu drei Kontrollbedingungen. Damit nehmen sie ältere Untersuchungen zu anaphorischen Inferenzen (vgl. McKoon/Ratcliff 1980) wieder auf, die eine *probe recognition*

<sup>13</sup> In *probe recognition tasks* wird den VersuchsteilnehmerInnen ein Wort (hier das Antezedens) präsentiert, nachdem sie einen Abschnitt gelesen haben. Die VersuchsteilnehmerInnen müssen dann entscheiden, ob das Wort im eben gelesenen Abschnitt vorkam oder nicht.

*task* einsetzen, und versuchen, einige potenzielle Störvariablen zu kontrollieren. Eine dieser Störvariablen ist die Distanz zwischen Antezedens und referierendem Ausdruck – eine Variable, die auch in die Analysen der vorliegenden Arbeit eingehen wird.

O'Brien et al. (1997) nehmen in ihrer Untersuchung zur Verarbeitung expliziter Anaphern die *Resonance*-Metapher auf. Nach McKoon/Gerrig/Greene (1996, S. 919f.) lässt sich diese wie folgt umschreiben:

[I]nformation retrieval is a fast, passive process by which cues in short-term memory interact with all information in long-term memory in parallel [...]. This fast, easy process accesses all of the information in memory, but the degree to which any specific cue in short-term memory evokes any specific piece of information in long-term memory depends on how strong an association exists in memory between the two, so that some things in memory will be evoked to a greater degree than others.

Diese Idee kann nun auf die Diskursverarbeitung übertragen werden. Die Stärke der Assoziation wird dabei über die Menge an *featural overlap* operationalisiert. Wenn nun eine Anapher (egal ob Pronomen oder volle Nominalphrase) gelesen wird, werden alle potenziellen Antezedenzien über den *Resonance*-Prozess aktiviert. Je mehr Eigenschaften das jeweilige Antezedens mit dem anaphorischen Ausdruck teilt („*featural overlap*“, O'Brien et al. 1997, S. 3), desto höher die Aktivierung des Antezedens. Je höher die Aktivierung des potenziellen Antezedens, desto höher die Wahrscheinlichkeit, dass es als tatsächliches Antezedens an den anaphorischen Ausdruck angebunden wird. Je expliziter nun der anaphorische Ausdruck wird, desto schneller sollte das korrekte Antezedens angebunden werden können. Pronomen befinden sich auf diesem Explizitheitskontinuum am unteren Ende – der Grad an *featural overlap* ist hier minimal. Eine weitere Hypothese, die aus der *Resonance*-Metapher abgeleitet werden kann, ist folgende: Stehen zwischen Anapher und Antezedens ein oder mehrere potenzielle Mitkandidaten, die ebenfalls einen hohen *featural overlap* mit dem anaphorischen Ausdruck haben, sollte die Verarbeitung der Anapher erschwert sein. Corbett (1984) konnte exakt diesen Effekt bereits im Rahmen eines Lesezeit-Paradigmas zeigen, bei dem in der Mitte des Bildschirms Textabschnitte wortweise präsentiert wurden. Die Präsentation des jeweils nächsten Wortes wurde durch einen Tastendruck der Versuchsteilnehmenden initiiert. Wiederum muss an dieser Stelle auf die Theorie des *cue-based parsing* (Lewis/Vasishth 2005; Van Dyke 2007) verwiesen werden, das ganz ähnliche Mechanismen in der Domäne der Verb-Nomenkongruenz annimmt. Wie oben besprochen verweist Ariel (2001) auf diesen Interferenzeffekt bei der Anaphernauflösung.



Einige der hier zitierten Ansätze (vgl. Almor 1999; Sanford/Garrod 1981) beziehen sich auf die Architektur des menschlichen (verbalen) Arbeitsgedächtnisses (Baddeley 1992; Caplan/Waters 1999; Just/Carpenter 1992). Drei Eigenschaften sind hierbei entscheidend: 1) Kapazitätsbeschränkung, 2) Interferenzprozesse und 3) Aktivierungszерfall. Aus den Kapazitätsbeschränkungen des Arbeitsgedächtnisses wird von Almor (1999) das Optimierungsprinzip hinsichtlich Kosten und Nutzen eines anaphorischen Ausdrucks abgeleitet: Wenn die Kapazität schon stark beschränkt ist, muss der vorhandene Platz auch effektiv genutzt werden – so die Grundidee. Im *Construction-Integration*-Modell (Kintsch 1988) bestimmt die Arbeitsgedächtniskapazität, wie viele Textbasis-Propositionen in den nächsten Verarbeitungsschritt übernommen werden. Kleinere oder größere Arbeitsgedächtniskapazität kann in manchen Fällen zu einer Änderung der Interpretation führen. In den allermeisten Fällen jedoch fallen dadurch die am schwächsten aktivierten Propositionen aus der Diskursrepräsentation heraus. „Überlebt“ eine schwach aktivierte Proposition aufgrund höherer Arbeitsgedächtniskapazität, hat dies für das Endprodukt der Interpretation meist nur geringe Folgen.

Interferenz-Prozesse werden meist auf eine Weise angenommen, wie sie prominent von Anderson (1974) in Form des *fan effects* formalisiert wurden. Je mehr mögliche Abrufkandidaten (= Antezedenzen) zu einem Abrufhinweis (= anaphorischer Ausdruck) passen, desto länger dauert es, bis der korrekte Abrufkandidat tatsächlich abgerufen wird. Nicol/Swinney (2003, S. 72f.) umschreiben diesen Prozess so, dass alle potenziellen Abrufkandidaten in Betracht gezogen werden, von denen in einem weiteren Schritt unpassende eliminiert werden. Das ist – je nach Formalisierung – weitestgehend überführbar in die Beschreibung im Sinne des *fan effects*. Nicht kompatibel wäre eine Konzeption, in der alle möglichen Antezedenzen zunächst abgerufen werden, um dann nach und nach wieder eliminiert zu werden. Im Sinne der dem *fan effect* zugrundeliegenden Architektur würde letzten Endes tatsächlich nur **ein** Antezedens abgerufen. Dieser Vorgang dauert nur umso länger, je mehr Kandidaten vorliegen.

Zusätzlich zu Kapazitätseinschränkungen und Interferenz durch mehrere Abrufkandidaten werden meist Zerfallsprozesse im Arbeitsgedächtnis angenommen. Diese führen dazu, dass die Aktivierung und damit die Abrufwahrscheinlichkeit von Elementen sinkt, bei denen der letzte Zugriff weiter in der Vergangenheit liegt. Effekte der Distanz zwischen Antezedens und anaphorischem Ausdruck werden oft auf diese Weise erklärt (vgl. O'Brien et al. 1997).<sup>14</sup>

<sup>14</sup> Duffy/Rayner (1990, S. 111f.) weisen darauf hin, dass dabei eventuelle Interaktionen mit *topic shifts* beachtet werden sollten. Sie zeigten in Blickbewegungsstudien, dass die Distanz keinen

Es ist leicht denkbar, dass Zerfalls- und Interferenzprozesse interagieren. Wenn ein korrektes Antezedens eine größere Oberflächendistanz zum anaphorischen Ausdruck als ein Konkurrent hat, ist die Aktivierung dieses korrekten Antezedens sozusagen „doppelt“ herabgesetzt: Weil der Konkurrent interferiert und weil die Aktivierung der Antezedens-Diskursrepräsentation bereits zu gewissem Maße zerfallen ist. Auch in der vorliegenden Arbeit wird die Oberflächendistanz zwischen referenziellem Ausdruck und Antezedens eine Rolle in den Analysen spielen. Interferenz und Zerfall werden traditionellerweise als zwei unterschiedliche Prozesse aufgefasst und formalisiert. Inwieweit sie auch funktionell (und nicht nur auf Effektebene) eng miteinander verzahnt sind, liegt nicht im Fokus der vorliegenden Arbeit, wird aber beispielsweise von Altmann/Gray (2002) beleuchtet.

Zu den bis hier vorgestellten Einflussfaktoren auf die Anaphernauflösung kommen noch einige andere hinzu, die hier nicht alle im Detail beschrieben werden können. So konnten beispielsweise Weldle (2011) sowie Weldle et al. (2010) zeigen, dass sich lokale syntaktische Kohärenzen (vgl. Konieczny et al. 2009; Tabor/Galantucci/Richardson 2004) auf die Verarbeitung von anaphorischen Ausdrücken (dort Pronomina und Reflexiva) auswirken. Bindungsprinzipien, die innerhalb der lokalen syntaktischen Kohärenz wirken, scheinen vorübergehend mit den Bindungsprinzipien zu interferieren, die von der globalen Satzinterpretation nahegelegt werden. In sprachkontrastiven Studien (vgl. Hemforth et al. 2010) wurde untersucht, wie sich Anbindungspräferenzen in verschiedenen Sprachen unterscheiden. Sturt (2003) zeigt, dass Genderstereotype starke und sehr frühe Einflussfaktoren in der Anaphernauflösung sein können. Kaiser (2010, 2011) zeigt Einflüsse von korrigierendem/kontrastivem Fokus und deren Interaktion mit anderen Faktoren. Weiterhin scheint auch die Frequenz des Antezedens (vgl. van Gompel/Majid 2004) und ob dieses negiert ist oder nicht (vgl. Cook/Myers/O'Brien 2005) einen Einfluss auf die Geschwindigkeit der Anaphernauflösung zu haben.

Die obigen Ausführungen legen nahe, dass Referenzauflösung (also die Herstellung von Koreferenz auf Ebene der Diskursrepräsentation) maßgeblich von fünf Hauptfaktoren beeinflusst wird, die untereinander interagieren. Diese Hauptfaktoren werden wiederum von Unterfaktoren beeinflusst, die dann meist klar einem der Hauptfaktoren zuordenbar sind. Die folgende Auflistung soll einen Überblick über diese komplexen Zusammenhänge vermitteln. Dabei werden all jene Faktoren als Hauptfaktoren bezeichnet, die laut der eingeführten Literatur einen unmittelbaren Einfluss auf die Ana-

---

Einfluss mehr auf die Lesezeit pro Buchstabe und die *first-pass reading time* der Anapher hat, wenn zwischen Antezedens und anaphorischem Ausdruck kein *topic shift* stattfindet.



phernauflösung haben. Faktoren, deren Einfluss über einen Hauptfaktor vermittelt scheint, werden dementsprechend als Unterfaktoren klassifiziert.

**Hauptfaktor 1: Die Zugänglichkeit des Antezedens** ist beeinflusst 1) vom Fokusstatus des Antezedens, welcher wiederum abhängig sein kann vom instantiierten Szenario, 2) von der semantischen Rolle des Antezedens, 3) von der Topikalisierung des Antezedens, 4) von der Anzahl und Passung potenzieller Mitbewerber um den Antezedensstatus, wobei auch Stereotype, beispielsweise bezüglich Gender, greifen können, 5) von dem Zerfall der Diskursrepräsentation (welcher wiederum abhängig ist von der Dauer des letzten Zugriffs) und 6) von der Frequenz des Antezedens.

**Hauptfaktor 2: Die Oberflächenform der Anapher** ist beeinflusst 1) von den Eigenschaften des Antezedens,<sup>15</sup> 2) sozio-kulturellen Faktoren, 3) davon, ob die Anapher zusätzliche Information zur Diskursrepräsentation des Antezedens hinzufügen soll (prädikativer Einsatz) und 4) von der Zugänglichkeit des Antezedens (Hauptfaktor 1). Letzteres ist die Hauptaussage der *Accessibility Theory*.

**Hauptfaktor 3: Der featural overlap zwischen Anapher und Antezedens** ist beeinflusst 1) von den semantischen Eigenschaften des Antezedens, 2) von den semantischen Eigenschaften der Anapher, 3) von der linguistischen Oberflächenform des Antezedens und 4) von der Form der Anapher (Hauptfaktor 2).

**Hauptfaktor 4: Die Rechtfertigung der anaphorischen Form** ist abhängig 1) vom *informational load* des Antezedens-Anaphern-Paares, welcher sich aus der konzeptuellen Distanz zwischen Antezedens und Anapher ergibt, 2) davon, ob die Anapher prädikativ eingesetzt wird, 3) von der Zugänglichkeit des Antezedens (Hauptfaktor 1) und 4) von der Form der Anapher (Hauptfaktor 2).

**Hauptfaktor 5: Einflüsse globaler Diskurskohärenz** sind abhängig vom Weltwissen der rezipierenden Person und dem semantischen Gehalt des Satzes (siehe Beispiel 4).

Die vielfältigen und recht unübersichtlichen Interaktionen zwischen den Hauptfaktoren zeigen sich daran, dass in fast allen Punkten andere Hauptfaktoren als beitragende Elemente auftauchen. Abbildung 2.2 macht einerseits deutlich, wie komplex die Zusammenhänge sind und zeigt außerdem, dass insbesondere die Antezedens-Zugänglichkeit (Hauptfaktor 1), die Oberflä-

<sup>15</sup> Das ist offensichtlich. Die Anapher muss – soweit das für die entsprechende Anapher relevant ist – beispielsweise in Genus und Numerus mit dem Antezedens übereinstimmen.

chenform der Anapher (Hauptfaktor 2) und die Rechtfertigung der anaphorischen Form (Hauptfaktor 3) untereinander stark verwoben sind.

In der vorliegenden Arbeit werden wir uns bei unseren Analysen auf einige wenige der genannten Faktoren beschränken müssen. Dies ist allein schon dadurch gegeben, dass einige Phänomene (beispielsweise negierte Antezedenzen) im vorgestellten Lesekorpus entweder gar nicht oder nur extrem selten vorkommen. Andererseits fehlen teilweise die entsprechenden linguistischen Annotationsebenen. Dies ist keine Einschränkung des Korpus selbst. In zukünftigen Studien, die anhand des *Freiburg Legalese Reading Corpus* durchgeführt werden könnten, müssten entsprechende Annotationsebenen hinzugefügt werden.

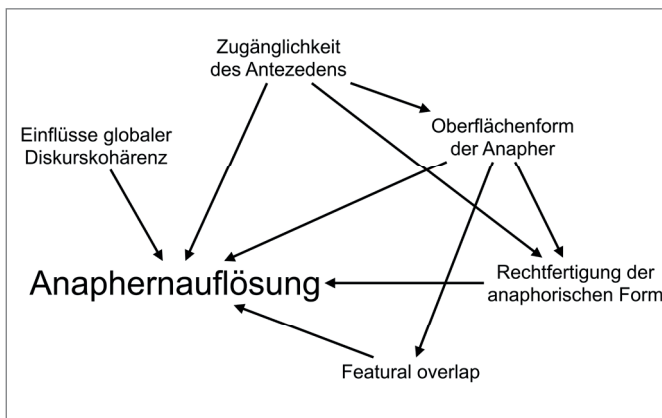


Abb. 2.2: Zusammenhänge zwischen den fünf Hauptfaktoren für die Anaphernauflösung. Unterfaktoren sind nicht im Schaubild aufgeführt.

Wollen wir aus den in diesem Abschnitt eingeführten kognitionswissenschaftlichen Modellen des Textverstehens direkte Aussagen zur Verständlichkeit von Texten ziehen, stehen wir vor einem Problem. Solche Positionierungen sind schlichtweg nicht im Skopus dieser Modelle: Aussagen, die nach dem Muster funktionieren, dass ein Text einfacher/schwieriger wird, wenn bestimmte sprachliche Muster verändert werden, sind in den Quelltexten so nicht vorhanden. Wenn überhaupt Vorhersagen solcher Art aus den vorgestellten Modellen abgeleitet werden können, dann würden diese fast immer auf einer feinkörnigeren Ebene als dem vollständigen Text operieren. So wäre es beispielsweise denkbar, für einen Text einige Verarbeitungszyklen gemäß des *Construction-Integration-Modells* zu simulieren (mit den im entsprechenden Abschnitt genannten Einschränkungen). Dann könnten in gewissen Grenzen Vorhersagen gemacht werden, welche Propositionen am Ende des

Textes als Hauptaussagen (Makropropositionen) im mentalen Diskursmodell der Leserin/des Lesers enkodiert werden, oder welche Inferenzen wann im Verständnisprozess notwendig oder wahrscheinlich werden (für einen Modellierungsversuch in diese Richtung vgl. Schmalhofer et al. 2002). Der Hauptgrund für die mangelnde Überführbarkeit von kognitionswissenschaftlichen Textverstehensmodellen in Vorhersagen über die Verständlichkeit eines Textes liegt in der theoretischen Ausrichtung dieser Modelle. Meist sind schlicht keine „Übersetzungsmechanismen“ vorhanden, die die im Modell formalisierten Konzepte in vorhersagende Variablen überführen, die sich auf die Verständlichkeit eines Texts beziehen. Es fehlt die *linking hypothesis*, die es ermöglichen würde, bestimmte Texteigenschaften mit Textkomplexität in Verbindung zu bringen. Diesen fehlenden Verbindungsstücken zwischen Modellprozessen und Textverständlichkeit stehen auf der anderen Seite zahlreiche Instrumente gegenüber, die den individuellen Verstehensprozess vorhersagen. Das ist die eigentliche theoretische Ausrichtung der Modelle: Beschreibungsobjekt ist das verstehende Individuum, nicht der verstandene Text. So kann beispielsweise simuliert werden, ob Weltwissen für das Verstehen eines Textes besonders wichtig ist, bzw. wann welche Elemente aus dem Weltwissen des verstehenden Individuums aktiviert werden. Andere Analysen simulieren, welchen Einfluss eine geminderte Arbeitsgedächtniskapazität auf den Verstehensprozess hat<sup>16</sup> oder wann im Laufe des Verstehensvorgangs die verfügbaren „Zwischenspeicher“ überlastet sind (vgl. Kintsch/Greeno 1985). Einige dieser Fragen beziehen sich natürlich auch auf den Text selbst. Doch nochmals: Primärer Bezugspunkt bleibt das verstehende Individuum. Einige andere Ansätze aus der Kognitionswissenschaft und Instruktionspsychologie bieten mehr Anwendungsbezug. In dieser Richtung sind beispielsweise die Forschungsarbeiten aus dem Bereich der Text-Bild-Integration bzw. des multimedialen Lernens zu nennen (Mayer 1997, 2005a, 2005b; Schnotz 2005; Schnotz/Bannert 1999, 2003; Schnotz et al. 2002; Sweller 2005). In den dort ausgearbeiteten Modellen zum Text-Bild-Verstehen ist der Anwendungscharakter sehr ausgeprägt. Da in den empirischen Untersuchungen der vorliegenden Arbeit keine Bilder zum Einsatz kamen, soll darauf aber nicht weiter eingegangen werden. Stattdessen werde ich im folgenden Abschnitt drei Vertreter von Verständlichkeitsmodellen aus der germanistischen Linguistik vorstellen. Dabei werden sich Prinzipien, die im aktuellen Abschnitt vorgestellt wurden, teilweise wiederfinden.

<sup>16</sup> Ein Parameter des *Construction-Integration*-Modells ist hierfür besonders geeignet. Die Standardannahme ist, dass von Verarbeitungszyklus  $i$  zu  $i+1$  die vier am stärksten aktivierten Propositionen übernommen werden (vgl. Kintsch 1988, S. 176). Tapiero/Denhrière (1995) variierten u.a. diesen Wert, um verschiedene Hypothesen bezüglich einer schrittweisen Elaboration der Textrepräsentation gegeneinander zu testen.

### 3. Textverständlichkeit

Die Bereiche Textverarbeitung und Textverständlichkeit haben zunächst erstaunlich wenig Überlappungspunkte. Die kognitionswissenschaftliche Forschung zum Textverstehen ist – wie am Ende des vorherigen Abschnitts ausgeführt wurde – sehr grundlagenorientiert. Rückschlüsse oder Vorhersagen für die Verständlichkeitsforschung sind die Ausnahme. Textverständlichkeitsmodelle, wie sie aus Schnittbereichen von (Instruktions-)Psychologie und Linguistik hervorgebracht wurden, stellen aber einige Ansatzpunkte zur Verfügung, die in diesem Kapitel aufgegriffen werden. Einige dieser Modelle greifen auch Elemente der im vorherigen Abschnitt eingeführten kognitionswissenschaftlichen Textverstehensmodelle auf. Den Abschluss dieses Kapitels bildet eine Zusammenfassung der linguistischen Komplexitätsmerkmale von Rechtstexten (3.3) sowie die für diese Arbeit überaus relevante Diskussion, ob und für wen Rechtssprache verständlich sein muss (3.4). Zunächst soll aber kurz ein Überblick gegeben werden, welche Bemühungen es in Deutschland und einigen anderen ausgewählten Ländern bereits hinsichtlich der Verständlichkeitsoptimierung von Rechtssprache gibt. Es schließt sich eine kurze Besprechung des Konzepts der *plain language* an.

#### 3.1 Verständlichkeitsoptimierung von Rechtssprache

Juristische Texte, denen gemeinhin das Klischee anhaftet, für Laien besonders unzugänglich zu sein, eine Auffassung, die auch von JuristInnen geteilt wird, wie Eichhoff-Cyrus/Strobel (2009) zeigen, bilden in der vorliegenden Arbeit den Ausgangspunkt der Analyse. Juristische Texte bieten sich nicht nur wegen den damit verbundenen meist negativen Einstellungen (vgl. beispielsweise Eichhoff-Cyrus/Antos/Schulz 2009) als Startpunkt an. Rechtstexte sind aufgrund ihrer besonderen linguistischen Oberflächenmerkmale als Prototyp schwer verständlicher Texte geeignet. Hier seien nur gelegentlich exzessive Einbettungstiefen genannt. Auch muss Texten aus dem Bereich der Rechtssprache eine zentrale gesellschaftliche Relevanz beigemessen werden. Rechtssprache ist hier als umfassender Sammelbegriff zu verstehen, der sowohl die Sprache der rechtlichen Normen (Gesetze, Verordnungen, Kommentare usw.) als auch die Verwaltungssprache, die zur Umsetzung von Gesetzen gebraucht wird, umfasst (vgl. Eichhoff-Cyrus/Strobel 2009). Die Grenzen zwischen den verschiedenen Bereichen der Rechtssprache sind fließend, denn Behörden sind im Schriftwechsel mit den BürgerInnen an die Rechtsgrundlagen gebunden. Da ist es nicht verwunderlich, dass sprachliche Merkmale

häufig aus den zu Grunde liegenden Normen übernommen werden, um die Gefahr einer abweichenden Deutungsmöglichkeit zu vermeiden. Einige Verwaltungen rühmen sich damit, besonders bürgernahe oder -freundliche Schreiben zu verfassen, in manchen Fällen wurde auch wissenschaftlicher Rat hinzugezogen. So kooperieren beispielsweise der Bund und das Germanistische Institut der Ruhr-Universität Bochum bei dem Projekt IDEMA (Internetdienst für eine moderne Amtssprache). Dieses Projekt hat u.a. eine Datenbank „Verständliche Verwaltungssprache“ hervorgebracht, die MitarbeiterInnen des Bunds durch Formulierungstipps, ein Online-Wörterbuch, Glossare sowie Vorher-Nachher-Textbeispiele beim Verfassen von verständlichen und zugleich rechtssicheren Texten unterstützt. Diese Datenbank ist Teil der von der Bundesregierung beschlossenen Strategie „Moderne Verwaltung“, deren Ziel eine leistungsfähigere, serviceorientierte und wirtschaftlichere Verwaltung ist. Doch solche Maßnahmen sind in Deutschland bisher noch die Ausnahme bzw. spielen bei der Erstellung eines Gesetzes, Urteils oder Bürgerschreibens eine untergeordnete Rolle. Bei dem oben genannten Programm „Moderne Verwaltung“, das Teil des Regierungsprogramms „Vernetzte und transparente Verwaltung“ ist, wird die Verständlichkeit von Texten nicht explizit behandelt – am ehesten fällt dieses Thema unter die Stichwörter Transparenz und Bürgernähe, die Teil des Programms sind.

Im deutschen Gesetzgebungsprozess war die Verständlichkeitsoptimierung von Gesetzestexten lange kein wichtiges Thema. In jüngster Zeit scheint sich auf diesem Gebiet einiges zu ändern. Die einzige Instanz, an die sich entsprechende Stellen zur Verständlichkeitsorientierung von Gesetzestexten wenden konnten, war lange Zeit der Redaktionsstab der Gesellschaft für deutsche Sprache (GfdS) beim Deutschen Bundestag. Die Ausführungen im entsprechenden Abschnitt der Geschäftsordnung des Deutschen Bundestags sind jedoch recht unverbindlich gehalten.

Ein beim Bundestag eingerichteter oder angesiedelter Redaktionsstab soll *auf Beschluss des federführenden Ausschusses* einen Gesetzentwurf auf sprachliche Richtigkeit und Verständlichkeit prüfen und *bei Bedarf* Empfehlungen an den Ausschuss richten. Der federführende Ausschuss *kann* den Redaktionsstab im gesamten Verlauf seines Beratungsverfahrens hinzuziehen und um Prüfung bitten. Dies gilt insbesondere für die Prüfung von Änderungsanträgen, deren Annahme zu erwarten ist. Darüber hinaus bietet der Redaktionsstab auch sonstige sprachliche Beratung an [meine Hervorhebungen, S.W.]. (§ 80a GO-BT)

Alle kursiv gedruckten Formulierungen vermitteln den Eindruck, dass es im Ermessen der beteiligten Ausschüsse liegt, ob der Redaktionsstab in seiner beratenden Tätigkeit hinzugezogen wird. Hinzu kommt, dass die sprachliche

Überarbeitung von Gesetzes- oder Verordnungsentwürfen nur eine von vielen Aufgaben ist, für die der Redaktionsstab zuständig ist. Er bietet außerdem eine Sprachberatung an, in deren Rahmen Mitgliedern des Bundestags und Bundesrats persönlich Auskünfte „bei allen sprachlichen Zweifelsfällen“ (Internetseite des Redaktionsstabs)<sup>17</sup> erteilt werden. Auch die Abfassung von Gutachten in Zusammenarbeit mit der GfdS gehört zu den Aufgaben des Redaktionsstabs beim Deutschen Bundestag.

Ein zweiter Redaktionsstab, der Redaktionsstab Rechtssprache, nahm 2009 seine Arbeit auf. Er ist beim Bundesministerium der Justiz angesiedelt und unterstützt die Bundesministerien bei ihrer Arbeit an Gesetzen und Verordnungen. Dieser zweite Redaktionsstab ist direktes Ergebnis des Projekts „Verständliche Gesetze“. Vier Sprachwissenschaftlerinnen erprobten im Rahmen des zweijährigen Projekts, wie Sprachberatung in den Gesetzgebungsprozess integriert werden kann. Unter anderem wurde das neue Wohngeldgesetz (WoGG) und Versorgungsausgleichgesetz (VersAusglG) im Rahmen des Projekts in Sachen Verständlichkeit optimiert. Die Deutsche Gesellschaft für Gesetzgebung zeichnete das Bundesministerium der Justiz für das VersAusglG mit dem „Preis für gute Gesetzgebung“ aus (vgl. Bundesministerium der Justiz 2011), dabei wurde unter anderem die klare Struktur des Gesetzes hervorgehoben.

Zwei Redaktionsstäbe arbeiten inzwischen also nahe der höchsten Ebenen der deutschen Gesetzgebungsorgane. Dabei wird der eben genannte Redaktionsstab Rechtssprache zu einem früheren Zeitpunkt einbezogen als der Redaktionsstab beim Deutschen Bundestag. Die Gemeinsame Geschäftsordnung der Bundesministerien sieht die Sprachprüfung in der Rechtsprüfung in der ersten ministeriellen Phase der Gesetzgebung vor (§ 46 GGO). Hieran ist unter anderem der Redaktionsstab Rechtssprache beteiligt. Es ist jedoch lediglich eine Prüfung „in rechtssystematischer und rechtsförmlicher Hinsicht“ (§ 46 Absatz 1 GGO) vorgesehen. Das Stichwort Verständlichkeit findet hier keinen Eingang in die Formulierung der Geschäftsordnung. Nach dem Kabinettsbeschluss zum Gesetzesentwurf beginnt dann die parlamentarische Phase des Gesetzgebungsverfahrens und der Redaktionsstab am Deutschen Bundestag „übernimmt“. Die Regelungen sehen hier bisher lediglich eine beratende Tätigkeit vor (siehe oben). Auch wenn inzwischen weite Teile des Gesetzgebungsverfahrens von Redaktionsstäben flankiert sind, kann von einer festen, institutionalisierten Einbindung der Verständlichkeitsprüfung in den Gesetzgebungsprozess nach wie vor nicht die Rede sein.

---

<sup>17</sup> <http://gfds.de/ueber-die-gfds/redaktionsstab-im-bundestag/> (Stand: 16.3.2015).



Andere Länder Europas zeigen sich innovativer: Insbesondere die Schweiz tut sich diesbezüglich hervor. So ist dort „begleitende Sprachberatung im Gesetzgebungsverfahren institutionalisiert“ (Frank 2007). Im eidgenössischen Sprachdienst, der dem Bundesrat unterstellt ist, werden die amtlichen Gesetze des Bundes sprachlich verbessert. Er sorgt dafür, dass „die Erlasse des Bundes auf allen Stufen [...] präzise, klar, widerspruchsfrei und so einfach wie möglich formuliert sind“.<sup>18</sup> Der Sprachdienst ist sehr früh in den Prozess der Gesetzesformulierung integriert. Es werden vor allem Gesetze und Abstimmungserläuterungen (wo nötig in Kooperation zwischen deutscher, französischer und italienischer Abteilung) redigiert. Außerdem erstellen die etwa 50 MitarbeiterInnen des Sprachdienstes Leitfäden zur Sprachverwendung, übersetzen Texte und überprüfen deutsche Übersetzungen völkerrechtlicher Verträge, schreiben Reden und stehen den Angestellten der Bundesverwaltung in sprachlichen Fragen zur Seite. Markus Nussbaumer, Sektionschef des deutschen Sprachdienstes der Bundeskanzlei wies 2007 bei der Tagung „Verständlichkeit als Bürgerrecht?“ der Gesellschaft für deutsche Sprache und der Universität Halle-Wittenberg darauf hin, dass er verständliche Gesetzestexte für möglich und wichtig hält. Die Relevanz speist sich dabei insbesondere durch das „Vertrauen der Bevölkerung in die Rechtsauslegung“ (Frank 2007). Der Sprachdienst arbeitet im Rahmen der verwaltungsinternen Redaktionskommission (VIRK), die die „Funktion der ersten Leserin“ hat, in die „politischen Kämpfe nicht involviert ist und unvoreingenommen, von außen, an die Texte herantritt“ (Nussbaumer 2008, S. 303). Die VIRK greift oftmals „sehr tief in die Texte ein“ (ebd.). Nussbaumer (2008) betont, dass dies insbesondere deshalb möglich ist, weil die VIRK zu einem Zeitpunkt mit der Textbearbeitung beginnen kann, wenn die inhaltliche Ausgestaltung der Texte noch „im Fluss“ (ebd.) ist. Das heißt, die inhaltlichen Diskussionen und Kompromissfindung sind noch im Gange und die Kommission ist nicht bloßer Vollstrecker bereits gefasster Beschlüsse.

Schweden gilt, ähnlich wie die Schweiz, als Vorreiter in Sachen klarer Gesetzgebung und dem damit verbundenen institutionellen Vorgehen. Dort arbeiten Rechtsberater und SprachexpertInnen paarweise an der Revision von Gesetzesentwürfen. Dieses Vorgehen ist bereits seit den späten 1970ern Usus. Ein wichtiger Meilenstein ist der Administrative Procedure Act (1986),<sup>19</sup> in dem es in Abschnitt 7 heißt: „The authority shall aim at expressing itself in an easily understandable way. The authority shall also by other means make

<sup>18</sup> [www.bk.admin.ch/org/bk/00332/04825/index.html?lang=de](http://www.bk.admin.ch/org/bk/00332/04825/index.html?lang=de) (Stand: 16.3.2015).

<sup>19</sup> Der Administrative Procedure Act ist abrufbar unter [www.government.se/sb/d/5805/a/64892](http://www.government.se/sb/d/5805/a/64892) (Stand: 16.3.2015).



matters easy for the people with whom it deals.“ Später findet sich im Rahmen des Language Act (2009)<sup>20</sup> unter einigen Regelungen zu Landessprache, Minoritätssprache(n) und dem Umgang mit der schwedischen Gebärdensprache der Abschnitt 11: „The language of the public sector is to be cultivated, *simple and comprehensible* [meine Hervorhebung, S.W.].“ Von diesen rechtlichen Eckpfeilern aus fand die Verständlichkeit von Regierungstexten, die für die Öffentlichkeit bestimmt sind, Einzug in die Verwaltungsverordnungen. Heute gilt: „No new legislation can be sent to the printers without the approval of the Secretariat for Legal and Linguistic Draft Revision.“ (Hasselrot 2012). Die Stelle ist verantwortlich für die Qualitätskontrolle der kompletten schwedischen Gesetzgebung – auf Parlaments- und Regierungsebene. Das Secretariat for Legal and Linguistic Draft Revision beschäftigt vier LinguistInnen, die gemeinsam mit juristischen Experten jedes Gesetz unter die Lupe nehmen. Jedes zu erlassende Gesetz muss zwingend das Sekretariat durchlaufen. Verständlichkeitstests in Abhängigkeit der LeserInnen-Expertise sind nicht Teil des Optimierungsprozesses (Anne-Marie Hasselrot, persönliche E-Mail-Kommunikation vom 5. Februar 2013). Hasselrot (2012) weist auf einen interessanten Aspekt hin, der die Folgen von verständlichen Reformulierungen in der Rechtssprache betrifft. Die einfache Feststellung, dass „legislative language influences official language at all levels“ (Hasselrot 2012, o.S.) deutet darauf hin, dass man von einer Art Ausbreitungs- oder Durchsickerungseffekt sprechen kann. Finden einfache, verständlichere Formulierungen erst Eingang in Gesetze, werden diese Formulierungen beispielsweise auch in Verwaltungsschreiben an Bürger übernommen. Den allgemeinen Mechanismus kann man sich folgendermaßen vorstellen: Verwaltungsbeschäftigte, die Schreiben an Bürger verfassen, befinden sich in einem Spannungsfeld. Einerseits sollen (wollen?) sie Bürgern verständlich vermitteln, was von ihnen aufgrund welcher Rechtsgrundlagen verlangt wird. Andererseits strebt die/der VerwaltungsmitarbeiterIn auch die korrekte Wiedergabe der Rechtsgrundlage an. Eine eigenmächtige Reformulierung der entsprechenden Verordnungen stellt dabei ein Risiko dar, weil ein/e Nicht-RechtsexpertIn explizit geregelte Sachverhalte eventuell durch eine Reformulierung verfälschen könnte. Die sicherste Alternative ist aus dieser Perspektive, die Originalformulierung aus der Rechtsgrundlage zu übernehmen. Schröder/Würdemann (2008) formulieren dies explizit: „Um rechtlich abgesichert zu sein, übernehmen die Mitarbeiter in den Verwaltungen häufig die Gesetzesformulierungen direkt in ihre Schreiben.“ (S. 326). Wenn nun die Rechtsgrundlage an sich schon durch Sprach- und RechtsexpertInnen auf Verständlichkeit (bei gleich blei-

---

<sup>20</sup> Eine englische Übersetzung des Language Acts ist abrufbar unter [www.government.se/sb/d/8317/a/138133](http://www.government.se/sb/d/8317/a/138133) (Stand: 16.3.2015).

bender Rechtssicherheit) hin optimiert wurde, fällt diese Unsicherheit weg. Anders formuliert: Bei Übernahme der Originalformulierung ist nun die Wahrscheinlichkeit höher, dass beide Ziele erfüllt werden können – Verständlichkeit für die/den BürgerIn und Rechtssicherheit für die/den Verwaltungsbeschäftigte/n. Diese Argumentation kann von Verwaltungsschreibern aus auf alle Texte erweitert werden, bei denen Rechtsgrundlagen den Ausgangspunkt bilden.

Auch in den Vereinigten Staaten gibt es Bemühungen, die Interaktion des Staates mit den Bürgern verständlicher zu gestalten. Am 1. Juni 1998 veröffentlichte der damalige US-Präsident Bill Clinton ein Memorandum, das Behörden dazu anhielt, in klarer Sprache („Plain Language“) zu schreiben. Seit 2004 sollen Internetseiten von US-Regierungsbehörden ebenfalls in Plain Language verfasst sein. Eine der neuesten Entwicklungen ist der Plain Writing Act (2010),<sup>21</sup> den US-Präsident Barack Obama 2010 unterschrieb. Darin werden Behörden dazu verpflichtet, in allen neu verfassten oder überarbeiteten Dokumenten und Formularen Plain Language zu verwenden. Die Überprüfung oder Optimierung von Texten aller Art bleibt dabei behördenintern. Laut Plain Writing Act muss ein/e MitarbeiterIn jeder Behörde beauftragt werden, die Einhaltung der Plain Language zu überwachen. Andere MitarbeiterInnen müssen geschult werden. Außerdem muss behördenintern ein Ablauf etabliert werden, der dafür sorgt, dass auch zukünftig die Prinzipien der Plain Language eingehalten werden. Jährliche Berichte an den Kongress sind vorgeschrieben. In den Gesetzgebungsprozess ist die Optimierung von Textverständlichkeit nicht zwingend integriert. Das Office of the Federal Register (OFR) redigiert und archiviert Rechtsdokumente.<sup>22</sup> Allerdings gilt: „Currently, the OFR does not have the legal authority to review agency documents for plain language.“ (Amy Bunk, Director of Legal Affairs and Policy, OFR, persönliche E-Mail-Kommunikation vom 30.10.2012). Es wird jedoch betont: „Although, we support plain language and recommend using pl [= plain language, S.W.] techniques to agency staff.“ (ebd.). Es ergibt sich ein Gesamtbild rechtlicher Regelungen bezüglich Plain Language in den USA, die eher auf Selbstverpflichtung und individuelle Motivation ausgerichtet sind als auf verpflichtende Implementierung im Gesetzgebungs- oder Formulierungsprozess. Von sehr gewissenhafter Erfüllung der Plain-Language-Kriterien bis hin zu beiläufig redigierten Texten scheint also alles möglich zu sein. Momentan

<sup>21</sup> Der Plain Writing Act ist abrufbar unter [www.gpo.gov/fdsys/pkg/PLAW-111publ274/html/PLAW-111publ274.htm](http://www.gpo.gov/fdsys/pkg/PLAW-111publ274/html/PLAW-111publ274.htm) (Stand: 19.5.2017).

<sup>22</sup> Dies schließt nicht den United States Code, das Gesetzbuch der Vereinigten Staaten, mit ein. Dieses wird vom Law Revision Counsel des Repräsentantenhauses kodifiziert.

befindet sich der Plain Regulations Act (2012)<sup>23</sup> im Gesetzgebungsverfahren der USA. Dieser soll die Maßnahmen, die im Plain Writing Act (2010) festgelegt sind, auf Regeln aller Art – inkl. Gesetzestexte – ausweiten.

Doch was sollen wir eigentlich konkret unter Plain Language verstehen? Historisch wird das Plain Language Movement mit der Verbraucherbewegung in den USA nach dem Zweiten Weltkrieg in Verbindung gebracht (vgl. Masson/Waldron 1994). Wenn es um Plain Language in US-Regierungsdokumenten geht, kann vor 1970 jedoch nur von vereinzelt Bemühungen gesprochen werden (vgl. Locke 2004). Das erste Dokument, „to appear entirely in plain English“ (ebd.) ist wahrscheinlich ein 1977 veröffentlichtes Regierungsdokument, das die Regeln für die Benutzung öffentlicher Funkfrequenzen regelt (Citizens Band Radio Rules). Seitdem haben sich mehrere US-Präsidenten für den Gebrauch klarer Sprache eingesetzt. Seit dem oben genannten präsidentialen Memorandum von Bill Clinton (vgl. Clinton 1998) breitete sich klare Sprache als Wert beim Abfassen von Regierungsdokumenten weiter aus. Als vorläufiger Endpunkt der offiziellen Regelungen gilt der oben bereits erwähnte Plain Writing Act (2010).

Der Begriff Plain Language ist fest mit dem Ziel der Kommunikation verknüpft. Sprache ist in diesem Sinne dann klar, wenn die Zielgruppe die von ihr benötigte Information findet, das Gefundene versteht und es so anwenden kann, dass Ihre Bedürfnisse befriedigt werden. „No one technique defines plain language. Rather, plain language is defined by results – it is easy to read, understand, and use“, so auf der zentralen Internetseite des Plain Language Action and Information Networks.<sup>24</sup> Trotzdem gibt es zahlreiche Richtlinien und Beispiele, die zeigen, wie Plain Language realisiert werden kann – auch von offizieller Seite in Form der Federal Plain Language Guidelines (2011). Zum Prinzip der Plain Language gehören demnach nicht nur Formulierungsprinzipien auf verschiedensten Ebenen (Aktiv statt Passiv, kurze Sätze, nur ein Thema pro Absatz, Beispiele einsetzen usw.), sondern auch Hinweise darauf, wie mit multiplen Adressatenkreisen umgegangen werden kann. Außerdem findet sich in den über 100 Seiten langen Guidelines ein Kapitel, das ausschließlich das Schreiben von Internetveröffentlichungen abhandelt. Ein kurzer Abschnitt über das Testen von (Re-)Formulierungen, der ebenfalls in den Guidelines enthalten ist, bleibt eher qualitativ und interviewbasiert.

Der Erfolg von Reformulierungen, die aufgrund der Leitlinien klarer Sprache entstanden, wurde u.a. von Masson/Waldron (1994) empirisch überprüft. Sie

---

<sup>23</sup> Der Plain Regulations Act ist in seiner derzeitigen Form abrufbar unter <http://centerforplainlanguage.org/wp-content/uploads/2012/01/Plain-Regulations-Act.pdf> (Stand: 16.3.2015).

<sup>24</sup> [www.plainlanguage.gov/whatisPL/index.cfm](http://www.plainlanguage.gov/whatisPL/index.cfm) (Stand: 16.3.2015).

reformulierten vier Dokumente, die relevant für die Öffentlichkeit sein können: eine Hypothek, einen Kaufvertrag, ein Bankdarlehen sowie eine Pachtverlängerung. Von jedem Dokument wurden drei reformulierte Versionen angefertigt. Im ersten Schritt wurden alle veralteten Begriffe ersetzt („indenture“ → „agreement“) oder entfernt. Die Satzstruktur wurde so nah wie möglich am Original belassen. Diesen Schritt zählen die AutorInnen nicht zu den nach Plain-Language-Kriterien reformulierten Texten. Im zweiten Schritt wurden schwierige Wörter durch einfachere ersetzt, lange Sätze wurden in kürzere geteilt und Passivstrukturen wurden in Aktivstrukturen umgewandelt. Außerdem wurden alle Referenzen auf die beteiligten Parteien in Personalpronomen der ersten und zweiten Person Singular umgewandelt. „I“ referierte nun immer auf den Leser, der stets die Rolle des Konsumenten einnahm. „You“ referierte immer auf jene Partei, die die jeweilige Übereinkunft mit dem Leser schloss (die Bank, den Vermieter usw.). Die Abfolge der im Text enthaltenen Aussagen wurde parallel zum Original gehalten, soweit es mit den Reformulierungen vereinbar war. Im letzten Reformulierungsschritt wurden zusätzlich juristische Fachbegriffe durch einfachere Terminologie ersetzt oder im Text erklärt. Die VersuchsteilnehmerInnen hatten nach dem Lesen des Dokuments sowie dem wiederholten Lesen eines ausgewählten Abschnitts zwei Aufgaben: Sie sollten den gelesenen Ausschnitt paraphrasieren und je vier Fragen beantworten, die sich auf das komplette Dokument bezogen. Zusätzlich wurde die Lesezeit für das komplette Dokument und den ausgewählten Abschnitt gemessen.

Die durchschnittliche Lesegeschwindigkeit (gemessen in Wörtern pro Minute) stieg lediglich im Übergang vom zweiten zum dritten Reformulierungsschritt signifikant an. Es scheint also für die Lesegeschwindigkeit von Vorteil zu sein, Fachbegriffe gegen einfachere, alltäglichere Wörter auszutauschen oder die Fachbegriffe zu erklären. Zunächst überraschend scheint es, dass die Reformulierungen auf der Satzebene (Übergang von erster zu zweiter Reformulierungsstufe) keinen Anstieg der Lesegeschwindigkeit nach sich zu ziehen scheinen. Die Paraphrasierungen der VersuchsteilnehmerInnen wurden auf Vollständigkeit hin bewertet. Dabei wurde eingeschätzt, wie viele Propositionen des gelesenen Abschnittes jede/r VersuchsteilnehmerIn in ihrer/seiner Paraphrasierung einschloss. Insgesamt zeigte sich ein Vorteil der anhand der Plain-Language-Kriterien reformulierten Texte (Reformulierungsstufen 2 und 3) versus dem Original und den Reformulierungen, in denen lediglich veraltete Wörter ersetzt oder entfernt wurden. Beim Paraphrasieren dieser ersten Reformulierungsstufe schnitten die Versuchsteilnehmer/innen aber immer noch besser ab als bei den Originalen. Außerdem fällt auf, dass in 11 von 13 Fällen, in denen die TeilnehmerInnen angaben, keine Paraphrasierung

erstellen zu können, sie zuvor einen Abschnitt aus einem Originaltext oder einem Text der Reformulierungsstufe 1 lasen. Die Analyse der Antworten auf die gestellten Fragen zeigte, dass Fragen nach Plain-Language-Texten häufiger korrekt beantwortet wurden (84% korrekte Antworten) als Fragen nach Originaltexten oder Texten, in denen lediglich veraltete Begriffe ausgetauscht wurden (69% korrekte Antworten). Ein Vorteil für jene Texte, in denen zusätzlich rechtliche Fachbegriffe erklärt wurden, konnte nicht festgestellt werden. Auch der Anteil korrekter Begründungen für die Antworten auf die Fragen stieg für jene Texte an, in denen Plain Language verwendet wurde. Aufgrund ihrer Ergebnisse schließen Masson/Waldron (1994, S. 78):

[V]ery little is achieved by simply removing archaic terms and legalese. Although these changes may make the document aesthetically more pleasing, to make real gains in reader comprehension, revisions must attack the complexity of language and syntax in a more radical way.

Gemeint ist damit der Übergang von erster zu zweiter Reformulierungsstufe, also jener Übergang, bei dem tatsächlich in die syntaktische Struktur der Sätze eingegriffen wurde und gezielt Pronomen eingeführt wurden, die eindeutig auf eine der beiden Vertragsparteien referieren. Leider kann der isolierte Einfluss dieser beiden Maßnahmen im Nachhinein nicht auseinandergelassen werden. Bei bereits nach Plain-Language-Kriterien reformulierten Texten sehen die AutorInnen keinen Vorteil mehr, Fachbegriffe noch weiter zu erklären oder durch alltagstaugliche Begriffe zu ersetzen. In der Tat legen ihre Daten dies nahe. Weder Paraphrasierungs- noch Antwortperformanz verbesserten sich im Übergang zwischen zweiter und dritter Reformulierungsstufe. Es lässt sich nur spekulieren, warum dies der Fall war. Die AutorInnen argumentieren in Richtung grundlegender Missverständnisse der Laien gegenüber juristischen Fachbegriffen und ungenauem rechtlichem Vorwissen, das sich aus „folk theories of the law“ (ebd., S. 79) speist. Diese falschen Auffassungen über rechtliche Konzepte könnten auch nicht durch alltagssprachliche Erklärungen ausgehebelt werden, so die AutorInnen (vgl. auch Tannen/Wallat 1986, die sehr ähnliche Ergebnisse in der Kommunikation zwischen KinderärztInnen und Eltern berichten).

Ein entscheidender Punkt, der sowohl von deutschen als auch von schweizerischen Beteiligten betont wird, ist der Zeitpunkt, ab dem Sprachdienste (oder vergleichbare Institutionen) in den Formulierungsprozess von Gesetzen einbezogen werden. Die Ergebnisse der Gesetzesformulierung sind demnach besser, je früher Redaktionsstäbe in den Prozess mit einbezogen werden. Dies hat mehrere Gründe. So verdeutlicht Markus Nussbaumer das größte Plus des Schweizer Modells:

Bereits bei der Entstehung eines Gesetzes wird von Anfang an der zentrale Sprachdienst [...] herangezogen, um die Perspektive der künftigen Leser bzw. Nutzer einzubringen. Denn eine spätere Überarbeitung wäre nur eine sprachlichen [sic!] Oberflächenbehandlung und könnte an der inhaltlichen Unverständlichkeit nichts mehr ändern! Dank der rechtzeitigen interdisziplinären Zusammenarbeit wird der Gehalt des Textes gesichert und zugleich die sprachliche Komplexität reduziert. Nutzen aus dieser Arbeit ziehen sowohl die Verfasser als auch die Adressaten der Gesetze (juristische Fachleute, Vermittler und Bürger); der Umgang mit den Gesetzestexten wird einfacher. (Wein/Kison 2007, o.S.)

Auch der deutsche Redaktionsstab Rechtssprache argumentierte auf seiner Internetseite<sup>25</sup> für eine frühzeitige Beteiligung. Optimale Ergebnisse seien nur bei einer „frühzeitigen Zusammenarbeit“ garantiert. Wenn der Entwurf erst im Rahmen der Rechtsprüfung beim Redaktionsstab eingereicht wird, seien „notwendige sprachliche Änderungen oftmals nur noch schwer möglich“. Die Hinweise des Rechtsstabs fließen in diesem Fall lediglich in die Stellungnahme des Bundesministeriums für Justiz ein. Die damalige Bundesjustizministerin Sabine Leutheusser-Schnarrenberger betonte: „Die besten Ergebnisse gibt es, wenn der Redaktionsstab in einem frühen Entwurfsstadium zu Rate gezogen wird.“ (Bundesministerium der Justiz 2011). Nussbaumer (2001, S. 118) sieht die Hauptursachen für schwer verständliche Rechtstexte hauptsächlich in einer „zu große[n] Verstricktheit der Autoren in die Sache, politische[n] Rücksichtnahmen“ und „mangelhafter Beherrschung der Textsorte“. Damit sind die Formulierungsprobleme nicht grundlegend andere als bei jedem Schreiben. Nussbaumer weist aber darauf hin, dass sie „angesichts der Spezifik der Produktion von Gesetzestexten“ stärker wirken. Er fordert den Blick von außen, „durch einen unverstellten Leser“ (ebd.). Diese Perspektive nehme die verwaltungsinterne Redaktionskommission ein. Außer der frühen Beteiligung der interdisziplinär zusammengesetzten Kommission (siehe oben) am Formulierungsprozess betont Nussbaumer noch einen anderen Punkt. So fordert er für die Kommission auch das Recht, „bewusst materielle Setzungen in Frage“ zu stellen, und „normative Konzepte“ umzustoßen, um „andere, einfachere, eingängigere“ (ebd., S. 119) vorzuschlagen. „Wo der materielle Gehalt a priori Tabu ist, ist richtig verstandene Gesetzesredaktion nicht möglich.“ (ebd.). Das ist überraschend, zeigt es doch, wie weit die Befugnisse der Redaktionskommission in der Schweiz gehen. Juristische Laien dürfen offensichtlich den formal-juristischen Gehalt der ihnen vorgelegten Entwürfe zumindest im Dialog mit den RechtsexpertInnen hinterfragen. Oft führt dieses Vorgehen, so

<sup>25</sup> Die Internetseite, von der die folgenden Zitate stammen, ist nicht mehr online. Unter <http://archive-de.com/page/876100/2012-12-07/http://www.gfds.de/redaktionsstab-rechtssprache/> konnte aber Ende 2014 noch eine gespeicherte Version eingesehen werden.



Nussbaumer (2001) weiter, zu einer letztendlichen Version, die weder der Vorlagefassung der juristischen Fachleute noch der ersten überarbeiteten Fassung der LinguistInnen entspricht.

Verwaltungstexte, insbesondere Schreiben, die direkt an die BürgerInnen gerichtet sind, sind durch ihren unmittelbaren Kontakt mit den Betroffenen und den damit verbundenen konkreten Maßnahmen als gesellschaftlich relevant zu bewerten. Gesetze regeln auf abstrakterer Ebene das gesellschaftliche Zusammenleben. Besondere gesellschaftliche Relevanz – zum Teil auf noch abstrakterer Ebene – besitzt das Bundesverfassungsgericht, dessen RichterInnen in ihrer Rolle als „Verfassungshüter“ (so beispielsweise Welt-Online vom 7. September 2011)<sup>26</sup> die Aufgabe haben, über die Einhaltung des Grundgesetzes zu wachen. Das Gericht selbst sieht sich in der Pflicht „der freiheitlich-demokratischen Grundordnung Ansehen und Wirkung zu verschaffen“ und ist sich im Klaren darüber, dass seine Arbeit „auch politische Wirkung“ hat (so auf der eigenen Internetpräsenz).<sup>27</sup> Diese Wirkung und Verantwortung erwächst nicht zuletzt daraus, dass die Entscheidungen des Bundesverfassungsgerichts unanfechtbar und bindend für alle übrigen Staatsorgane sind. Es liegt daher auf der Hand, dass Entscheidungen, die das Bundesverfassungsgericht trifft, gesellschaftlich höchstrelevant sind und nahezu immer alle Menschen, die in Deutschland leben, betreffen.<sup>28</sup> Daraus erwächst die Pflicht, diese Entscheidungen adäquat zu vermitteln – nur wer Entscheidungen versteht, kann sich ein Bild davon machen, was sie für das eigene Leben bedeuten. Im Extremfall betreffen die Entscheidungen einzelne Menschen sogar direkt.<sup>29</sup> Dabei wäre es geradewegs fatal, wenn die Entscheidung nicht adäquat vermittelt würde. Roman Herzog formulierte es in der Süddeutschen Zeitung vom 9. Juli 1999 folgendermaßen: „Was nicht zu verstehen ist, kann weder auf Verständnis noch auf Befolgung hoffen.“

Doch wie müssen Texte formuliert sein, damit sie einfacher zu verstehen sind? WissenschaftlerInnen aus der angewandten Sprachwissenschaft und

---

<sup>26</sup> [www.welt.de/wirtschaft/article13589311/Verfassungshueter-richten-ueber-Griechenland-Hilfen.html](http://www.welt.de/wirtschaft/article13589311/Verfassungshueter-richten-ueber-Griechenland-Hilfen.html) (Stand: 16.3.2015).

<sup>27</sup> [www.bundesverfassungsgericht.de/DE/Das-Gericht/Aufgaben/aufgaben\\_node.html](http://www.bundesverfassungsgericht.de/DE/Das-Gericht/Aufgaben/aufgaben_node.html) (Stand: 16.3.2015).

<sup>28</sup> Das Bundesverfassungsgericht entscheidet natürlich nicht nur über staatsinterne Angelegenheiten. Ein Beispiel aus unserem Korpus, bei dem darüber entschieden wurde, ob der Bundestag an der Entscheidung über einen AWACS-Einsatz im Luftraum der Türkei beteiligt werden muss, macht dies deutlich. Das ändert aber nichts an der Relevanz, die das Gericht für die Gesellschaft hat.

<sup>29</sup> Dies ist bei einem Beispiel aus dem hier verwendeten Korpus der Fall. Hier wurde über die Rechtmäßigkeit diverser Urteile vorinstanzlicher Gerichte entschieden, die die Rechtmäßigkeit des Aufenthalts eines konkreten Ausländers in Deutschland betrafen.



der Instruktionspsychologie haben hier mehrfach versucht, Impulse zu geben. Grundlage sind praktisch immer wissenschaftliche Modelle, die aufgrund einer bestimmten Datenlage Empfehlungen für verständliches Schreiben aussprechen. Drei solcher Verständlichkeitsmodelle oder -konzeptionen werden im folgenden Abschnitt vorgestellt. Ich beschränke mich dabei auf Modelle, die in der germanistischen Sprachwissenschaft besonderen Einfluss hatten.

### 3.2 Verständlichkeitsmodelle

Das Hamburger Verständlichkeitskonzept nach Langer/Schulz von Thun/Tausch (1974, 2006) hat großen Einfluss in der Forschung zu Textverständlichkeit und insbesondere in der angewandten Linguistik. Weiterhin werden einige Thesen der Verständlichkeitskonzeption von Groeben/Christmann (1989, 1996) vorgestellt. Diese werden insbesondere kontrastiv zum Hamburger Modell in den Diskurs eingeführt. Das Karlsruher Verständlichkeitsmodell (Göpferich 2006, 2008, 2009) kann als Erweiterung und Verfeinerung des Hamburger Modells und des Ansatzes von Groeben/Christmann verstanden werden. Zunächst soll jedoch der Verständlichkeitsbegriff über einige 'vorangeschaltete' Konzepte hergeleitet werden.

Theoretische Voraussetzung für die Verständlichkeitsforschung ist, dass Menschen Texte überhaupt wahrnehmen können. Den physikalischen Voraussetzungen dazu wird unter dem Stichwort der Leserlichkeit nachgegangen, wobei überwiegend auf „mikrotypographische Textgestaltungsmerkmale“ (Groeben 1982, S. 174) abgehoben wird. Merkmale wie Form, Breite, Höhe und Stärke der Drucktype, Zeilenlänge und -abstand, Kontrast, Layout (Gruppierung von Wörtern, Spaltenabstand usw.) sind hier einige relevante Merkmale. Die Ergebnisse dieser Forschungsrichtung sind ohne Frage relevant für die grundlegendste Ebene des Textverständnis, der Textperzeption. Unter extremen Bedingungen, d.h. wenn viele leserlichkeitseinschränkende Merkmale zusammenkommen, leidet vermutlich auch die Verständlichkeit eines Textes unter den mikrotypografischen Voraussetzungen. In dieser Arbeit wird die Leserlichkeit von Texten als gegeben angenommen. Das Lesen am Bildschirm, wie es die VersuchsteilnehmerInnen taten, ist zwar sicherlich nicht die natürlichste Art, Texte zu rezipieren – zumindest sind aber die Effekte für alle Texte und zu jeder Zeit konstant.<sup>30</sup>

<sup>30</sup> Selbst der Fakt, dass das Lesen am Bildschirm in E-Book-Form immer beliebter wird (vgl. Börsenverein des Deutschen Buchhandels 2012), kann nicht uneingeschränkt als Argument herangezogen werden, da die Präsentation auf die Messung der Blickbewegungen ausgerichtet war. Schrift und Zeilenabstand sind hier größer als gewohnt.

Selbst wenn wir Leserlichkeit als gegeben annehmen, muss noch zum Konzept der Lesbarkeit Stellung bezogen werden. Die Lesbarkeitsforschung versucht(e), Oberflächenmerkmale von Texten, beispielsweise die durchschnittliche Satzlänge, Type-Token-Ratio usw., so miteinander zu verrechnen, dass am Ende ein Index für jeden Text steht, der die Lesbarkeit dieses Texts beschreibt. Lesbarkeitsformeln gibt es in vielen unterschiedlichen Ausführungen, insbesondere für das Englische. Klare (2000, S. 21) zählt etwa dreißig Stück. Er nennt zwei Charakteristika, anhand derer man die Auswahl einschränken könne: „The predictive accuracy of the formula and speed of application“ (ebd.). Hinzu kommt der Bereich, auf den die spezifische Formel normiert ist. So gibt es beispielsweise eine Formel, die speziell auf die Lesbarkeit von psychologischen Tests normiert ist (vgl. Forbes/Cottle 1953). Einige Formeln sind explizit auf eingeschränkte Gruppen von RezipientInnen ausgerichtet. So entwickelten beispielsweise Washburne/Morphett (1938) eine Lesbarkeitsformel, mit der Kinderbücher dahingehend eingeordnet werden können, für welche Klassenstufe (im speziellen Fall von der ersten bis zur neunten Klasse) sie geeignet sind. Eine umfassende Darstellung verschiedenster Lesbarkeitsformeln bietet Klare (1963). Ein Nachdruck des ersten Kapitels dieser Monografie (Klare 2000) vermittelt einen guten Überblick. Aus Platzgründen kann an dieser Stelle keine detaillierte Herleitung und Einführung in den Bereich der Lesbarkeitsformeln gegeben werden.

Lesbarkeitsformeln sind unter mehreren Gesichtspunkten kritisch zu betrachten. Einerseits werden beim Berechnungsvorgang nie Interaktionen zwischen Prädiktoren berücksichtigt. Das heißt, dass keine Einflüsse festgestellt werden können, die aus dem Zusammenwirken zweier oder mehrerer Faktoren entstehen. Der häufigste Kritikpunkt zielt darauf ab, dass nur Oberflächenmerkmale der Texte in die Formeln mit eingehen. Das ist einerseits natürlich der Reiz und genuines Merkmal der Lesbarkeitsformeln. Andererseits werden damit Satzkonstruktion, komplexe Kohärenz-Zusammenhänge oder Gliederungseigenschaften der Texte vollständig außer Acht gelassen. Groeben/Christmann (1989, S. 167) weisen darauf hin, dass zumindest „begrenzt Hinweise zur sprachlich-stilistischen Gestaltung von Texten“ abgeleitet werden können. Darunter fassen sie „kurze, geläufige Wörter“ und „grammatisch einfache Sätze“ (ebd.). Die prädiktive Validität – so Groeben/Christmann (1989) weiter – ist nur für die abhängige Variable Gesamtlesezeit gesichert. Versucht man das Leseverständnis für Texte vorherzusagen, die nicht in die Berechnung der Formeln eingingen, scheitert diese Vorhersage. „Dies ist darauf zurückzuführen, daß Aspekte der inhaltlichen Organisation und Strukturierung unberücksichtigt bleiben.“ (ebd.).

Aus den genannten Beschränkungen der Lesbarkeitsforschung heraus motiviert, stellen Groeben/Christmann (1989) an den Begriff der Verständlichkeit folgende Anforderungen: Das Konstrukt müsse „über die formal-stilistischen Beschreibungen der Textstruktur hinausgeh[en] und auf möglichst breiter Basis verständlichkeitsrelevante Dimensionen der Textstruktur einbezieh[en]“ (ebd., S. 168). Insbesondere heben sie „die semantische Struktur und Organisation von Textinhalten“ (ebd.) hervor. Außerdem müssten die Verarbeitungsprozesse des Rezipienten berücksichtigt werden. Bereits in Abschnitt 2 wurde darauf hingewiesen, dass der/dem RezipientIn eine „aktive Rolle“ (ebd.) beim Textverstehen zukommt. Groeben/Christmann (1989) greifen diesen Punkt ebenfalls auf: „Textverstehen ist [...] in Übereinstimmung mit der kognitiv-konstruktivistischen Erklärungsperspektive der Sprachverarbeitung [...] als Prozeß der Interaktion zwischen vorgegebenem Text und dem Kognitionssystem des Rezipienten zu sehen.“ (ebd.). Methodisch stehen dabei aber nicht unbedingt die inter-individuellen Unterschiede zwischen LeserInnen im Vordergrund. Durch eine möglichst große Anzahl von LeserInnen soll zu (meist gemittelten) Werten (in der linguistischen Verständlichkeitsforschung sind dies meist Bewertungen) für einzelne Texte gekommen werden, die losgelöst vom einzelnen Individuum etwas über die Verständlichkeit des Textes aussagen.

<b>Pol</b>	<b>Gegenpol</b>	<b>Dimension</b>
interessant	langweilig	Zusätzliche Stimulanz
zu kurz	zu lang	Kürze/Prägnanz
leicht verständlich	schwer verständlich	Einfachheit
folgerichtig	zusammenhanglos	Gliederung/Ordnung
für Schüler wenig ungeläufige Wörter	für Schüler viele ungeläufige Wörter	?
anregend	einschläfernd	Zusätzliche Stimulanz
stark gegliedert	wenig gegliedert	Gliederung/Ordnung
einfache Sätze	überladene Sätze	Einfachheit
konkret	abstrakt	Einfachheit
übersichtlich	unübersichtlich	Gliederung/Ordnung
flüssig	holprig	?
anschaulich	unanschaulich	Einfachheit
einfach	kompliziert	Einfachheit

Pol	Gegenpol	Dimension
einprägsam	schwer zu behalten	Einfachheit
weitschweifig	aufs Wesentliche beschränkt	Kürze/Prägnanz
kindgemäß	nicht kindgemäß	Einfachheit
gute Unterscheidung von Wesentlichem und Unwesentlichem	fehlende Unterscheidung von Wesentlichem und Unwesentlichem	Gliederung/Ordnung
abwechslungsreich	eintönig	Zusätzliche Stimulanz

Tab. 3.1: Die dem Hamburger Verständlichkeitsmodell zugrunde liegenden Gegensatzpaare, anhand derer die Polaritätsprofile für die Texte erhoben wurden (Langer et al. 2006, S. 190). In der dritten Spalte ist jeweils die Verständlichkeitsdimension angegeben, der die jeweilige Frage anhand der Faktorenanalyse „zugeschlagen“ wird (nach Langer et al. 1974, S. 51).

So gingen auch Langer et al. (1974) bei der Erstellung ihres Hamburger Verständlichkeitsmodells vor. Sie bedienten sich der Technik der semantischen Differentiale (vgl. Osgood/Luria 1954; Osgood/Suci/Tannenbaum 1957), um für jeden Text im ersten Schritt ein Polaritätsprofil zu erhalten. Die abgefragten Eigenschaften formten ein „kontextspezifisches Differential“ (Christmann 2004, S. 39), bei dem nicht die zunächst von Osgood postulierten „universellen semantischen Polaritäten“ (ebd., S. 37) eingesetzt wurden. Langer et al. (1974) erstellten das von ihnen eingesetzte Differential nach einer Sichtung der Rhetorik- und Stilistik-Forschung bis dato. Die 18 eingesetzten Gegensatzpaare sind in Tabelle 3.1 aufgeführt. Zehn ExpertInnen mussten ganze Texte (hier Schultexte) auf einer siebenstufigen Skala einschätzen. Aus den Bewertungen wurden anhand einer Faktorenanalyse vier relevante zugrunde liegende Dimensionen extrahiert: die in der Folge zu einiger Berühmtheit gelangten Dimensionen Einfachheit, Gliederung/Ordnung, Kürze/Prägnanz sowie anregende Zusätze<sup>31</sup> (vgl. Langer et al. 2006, S. 21). Diese vier Dimensionen der Verständlichkeit wurden dann selbst wiederum zur Bewertung von Texten durch ExpertInnen eingesetzt. In Tabelle 3.1 ist die Zuordnung von Fragen zu Verständlichkeitsdimensionen in der letzten Spalte vermerkt.<sup>32</sup> Die

<sup>31</sup> Diese vierte Dimension wird teilweise auch als ‘zusätzliche Stimulanz’ bezeichnet.

<sup>32</sup> Die Zuordnung der Fragen bzw. Eigenschaften zu Verständlichkeitsdimensionen bleibt in zwei Fällen unklar. In einer entsprechenden Aufstellung (Langer et al. 1974, S. 51) werden die Eigenschaften „für Schüler wenig/viele ungeläufige Wörter“ und „flüssig/holprig“ nicht aufgeführt. Später wird die Eigenschaft „geläufige Wörter“ unter dem Positivpol der Einfachheit geführt. „Flüssig/holprig“ wird nicht mehr genannt.

vier zugrunde liegenden Dimensionen sind somit einerseits auf methodischer Ebene ein Hilfskonstrukt für die Erhebung von Verständlichkeitsurteilen, andererseits haben sie auch theoretischen Status erlangt: Es sind laut dem Hamburger Verständlichkeitskonzept eben **die** Dimensionen, anhand derer die Verständlichkeit eines Textes eingeschätzt werden kann.

Im Folgenden wurden die Dimensionen mehrfach evaluiert. So konnte beispielsweise in einer groß angelegten Studie mit über 900 SchülerInnen gezeigt werden, dass Texte, die sich auf den vier Dimensionen unterschiedlich positionieren lassen, unterschiedliche Ergebnisse hinsichtlich der Verständnis- und Behaltensleistung hervorrufen. Die vier Dimensionen unterschieden sich dabei in ihrer Relevanz:

Gut verständliche Texte hatten hohe Werte in Einfachheit, Gliederung/Ordnung und Kürze/Prägnanz. Anregende Zusätze spielten keine besondere Rolle. Sie wirkten sich nur verständnisfördernd aus, wenn der Text gleichzeitig gut gegliedert war. Schlecht verständliche Texte hatten Mängel in Einfachheit. Diese konnten auch durch gute Gliederung/Ordnung oder Kürze/Prägnanz nicht ausgeglichen werden. Anregende Zusätze behinderten die Verständlichkeit auch bei großer Einfachheit, wenn der Text schlecht gegliedert und zu weit-schweifig war. (Langer et al. 2006, S. 192)

Die Verstehens- und Behaltensleistung der SchülerInnen wurde über die Beantwortung von Fragen zu den Texten operationalisiert. In einer weiteren Evaluationsstudie (vgl. Steinbach/Langer/Tausch 1972) wurden Texte mit anderen Themen und anderen Schülern evaluiert. Da hier Texte mit mathematischen Anwendungen (Multiplizieren mit dem Rechenschieber und Winkelhalbierung) zum Einsatz kamen, konnte die Performanz der SchülerInnen beim Lösen dieser Aufgaben direkt erhoben werden. Es zeigte sich wiederum, dass höhere Werte für das Merkmal 'Einfachheit' zu besseren Leistungen führen. Im Gegensatz zur ersten Evaluationsstudie waren Texte mit mittleren Werten für die Dimension 'Kürze/Prägnanz' besser für die Leistung der SchülerInnen als zu lange oder zu kurze Texte. „Gliederung/Ordnung und Anregende Zusätze hingen mit den Schülerleistungen *nicht* zusammen [Hervorh. i. Orig., S.W.]“ (Langer et al. 2006, S. 193) – vermutlich konnte dieser Zusammenhang nicht gezeigt werden, weil sich die verwendeten Texte auf diesen Dimensionen nicht unterschieden (vgl. ebd., S. 193f.). Das ist für die vorliegende Arbeit ebenfalls ein kritischer Punkt: Wählt man Texte, die auf relevanten Merkmalsdimensionen zu wenig variieren, läuft man Gefahr, dass keine relevanten Effekte zu Tage treten. Die Originalauszüge sowie die moderaten und starken Reformulierungsversionen dieser Textauszüge sind darauf ausgelegt, auf relevanten Dimensionen zu variieren. Die vollständigen, natürli-

chen Texte im Korpus sollten alleine aufgrund ihrer Zugehörigkeit zu verschiedenen Textgattungen (Zeitungsartikel, Pressemitteilungen, Urteile) schon untereinander so variieren, dass dieses Problem nicht bestehen dürfte.

Groeben (1982, S. 191) wirft einen interessanten Punkt bezüglich der ersten Dimension der Einfachheit auf. Die von Langer et al. (1974) ermittelte Dimension der Einfachheit kann demnach in der Tat ganz eng in Verbindung gebracht werden mit den von der Lesbarkeitsforschung untersuchten Merkmalen und den daraus resultierenden Empfehlungen. Das wird plausibel, wenn man sich die stichwortartige Definition der beiden Pole dieser Dimension vor Augen führt. Einfachheit – der positive Pol – ist demnach definiert durch eine einfache Darstellung, kurze, einfache Sätze (durchschnittliche Satzlänge), geläufige Wörter (Wortlisten), erklärte Fachwörter, konkrete (Anzahl abstrakter Wörter) und anschauliche Ausführungen. In Klammern hinter den von Langer et al. (1974) postulierten Eigenschaften des positiven Pols der Dimension 'Einfachheit' habe ich jeweils die damit verbundene Operationalisierung aus der Lesbarkeitsforschung aufgeführt. In keiner der anderen Dimensionsbeschreibungen lassen sich so viele Konzepte finden, die auch in der Lesbarkeitsforschung auftauchen. Lediglich die Eigenschaft „zu kurz“ in der Dimension 'Kürze/Prägnanz' sowie die Eigenschaft „abwechslungsreich“ in der Dimension 'zusätzliche Stimulanz' lässt sich auf Werte aus Lesbarkeitsformeln abbilden. Wie abwechslungsreich ein Text ist, lässt sich aber nur sehr grob in Form der *type token ratio* (lexikalische Vielfalt) in Lesbarkeitsformeln abbilden.

Langer et al. (1974, S. 14f.) unterteilen die Dimension 'Gliederung/Ordnung' in innere und äußere Ordnungsmerkmale. Innere Ordnung sei dann gegeben, wenn Sätze „nicht beziehungslos nebeneinanderstehen, sondern folgerichtig aufeinander bezogen sind“ und „die Information in einer sinnvollen Reihenfolge dargeboten wird“ (ebd., S. 14). Diese Formulierung zielt auf die Textkohärenz auf lokaler Ebene ab. Globale Kohärenz (zur Unterscheidung vgl. beispielsweise Schnotz 2006, S. 228f.) spielt in den Argumentationen zu den vier Verständlichkeitsdimensionen eher eine untergeordnete Rolle. Dieser Eindruck kann aber auch deshalb entstehen, weil nicht zwischen den beiden Konstrukten getrennt wird. Äußere Ordnungsmerkmale beschränken sich für Langer et al. (1974) auf die visuelle Gliederung eines Textes – der Aufbau des Textes soll sichtbar gemacht werden. Dazu zählen sie unter anderem Absatzüberschriften, gliedernde Vor- und Zwischenbemerkungen und die Visualisierung von Relevanz mit typografischen Mitteln. Beide Gesichtspunkte bewirken insbesondere bei längeren Texten, dass „der Leser sich zurechtfindet und die Zusammenhänge versteht“ (ebd., S. 14).

Auch die Dimension 'Kürze/Prägnanz' unterliegt einer gewissen Zweiteilung, denn „Sprachaufwand kann verschieden zustandekommen“ (ebd., S. 15). Einerseits können „Entbehrlichkeiten“ (ebd., S. 16) bezüglich des Informationszieles (im Falle von Langer et al. ist das hauptsächlich der an SchülerInnen zu vermittelnde Lehrinhalt) in Texten enthalten sein. Das wären alle nicht notwendigen Inhalte, wie beispielsweise Abschweifungen vom Thema, zu weites Ausholen oder überflüssige Zusatzinformationen. Auf der anderen Seite stehen „sprachliche Entbehrlichkeiten“ (ebd.). Diese umfassen überflüssige sprachliche Elemente wie beispielsweise Füllwörter, weitschweifige Formulierungen und Wiederholungen. Gerade diese Verständlichkeitsdimension dürfte sich sehr unterschiedlich für verschiedene Textgattungen auswirken. Vergleicht man die Textgattungen, die in der vorliegenden Arbeit verwendet werden, stellt sich die Frage, ob die Dimension der Kürze/Prägnanz für alle gleich relevant ist. Pressemitteilungen zu bestimmten Urteilen werden ohne Frage weite Teile der Argumentation in Urteilen außen vor lassen und die relevantesten Eckpunkte referieren. Noch weiter abstrahieren Zeitungsartikel, die wiederum den Inhalt der Pressemitteilungen so weit kondensieren, dass nur die relevantesten Beweggründe für eine Entscheidung noch explizit sprachlich kodiert werden. Andererseits erweitern Zeitungsartikel gegebenenfalls das Themenfeld, wenn die politisch-gesellschaftlichen Folgen von Urteilen abgeschätzt werden.

Anhand der vierten Dimension, zusätzliche Stimulanz, wollen Langer et al. all jene Maßnahmen des Autors fassen, „die beim Leser Interesse, Anteilnahme, Lust am Lesen hervorrufen sollen“ (ebd., S. 16). Einerseits werden hier Stilelemente genannt, die die/den LeserIn direkt ansprechen und in den Text „hineinziehen“ sollen. Dazu gehören beispielsweise Fragen zum Mitdenken oder direkte Ansprachen. Der andere Großbereich sprachlicher Mittel könnte unter dem Stichwort Personalisierung gefasst werden. Genannt werden lebensnahe Beispiele oder das Auftreten von Menschen (vgl. ebd.). Doch auch Reizworte, witzige Formulierungen oder Abwechslungsreichtum werden in dieser Dimension gefasst. Das lässt schon vermuten, was auch Langer et al. explizit ansprechen: „Nicht vollständig unabhängig voneinander sind die Dimensionen Kürze-Prägnanz und Zusätzliche Stimulanz. Denn zusätzliche Stimulierungen verlängern den Text.“ (Langer et al. 1974, S. 17). Außerdem ist ein mittleres bis mäßig hohes Ausmaß an zusätzlicher Stimulanz „in Verbindung mit hoher Ausprägung in Gliederung-Ordnung“ (ebd., S. 25) vorteilhaft. Ansonsten postulieren die Autoren die Unabhängigkeit der vier Dimensionen.



Die Vorgehensweise, die zur Herausbildung der vier Verständlichkeitsdimensionen nach dem Hamburger Verständlichkeitskonzept führte, war zum damaligen Zeitpunkt ohne Frage innovativ und ein Gewinn bringender Schritt auf dem „empirisch-induktiven Weg der Dimensionsbeschreibung“ (Groeben/Christmann 1989, S. 169). Es gibt aber auch einige Probleme und Kritikpunkte, die teilweise in die methodische Ausrichtung der vorliegenden Arbeit eingehen. Christmann (2004) unterstellt dem Ansatz „Theorielosigkeit“ (ebd., S. 43) und kritisiert außerdem die „relativ subjektiven und intuitiven Expertenratings“ (ebd.). Diese ließen eine Erfassung der Verständlichkeit eines Textes nur indirekt zu. Diese Expertenratings wurden nach der Reduktion der 18 Merkmalsdimensionen auf die vier referierten Verständlichkeitsdimensionen eingesetzt, um Texte zu bewerten. Langer et al. (1974, S. 14ff., 108ff.) fertigten sogenannte „Dimensionsbilder“ an, anhand derer speziell geschulte ExpertInnen Texte bewerten sollten. Dabei wird nur der Gesamteindruck der Dimension beurteilt. Die einzelnen Merkmale, die im Allgemeinen auf die der jeweiligen Dimension zugrunde liegenden Fragen des Polaritätsprofils (für die Zuordnung siehe Tab. 3.1) zurückgehen, sollten nicht beachtet werden. Nachdem trainierte ExpertInnen den Text „auf sich wirken“ ließen, sollten sie den Text für jede Dimension auf einer Skala von -2 bis +2 bewerten. Das sogenannte Wahrnehmungstraining der ExpertInnen bereitet gezielt auf den Umgang mit diesen Dimensionsbildern vor und macht explizite Vorgaben, welche Bewertungen für einige Beispieltexte die „richtigen“ sind (zum genauen Vorgehen während der ExpertInnen-Ausbildung vgl. Langer et al. 1974, S. 120ff.). Die Intuition der ExpertInnen beim Bewerten von Texten wird also gezielt in Richtung der Prinzipien des Hamburger Verständlichkeitsmodells gelenkt. Fairerweise muss man aber auch anmerken, dass die Bewertungen trainierter ExpertInnen evaluiert wurden, indem sie erfolgreich mit der Verstehensleistung von Probanden bei weiteren Textgattungen in Beziehung gesetzt wurden.

Groeben/Christmann greifen in ihrem Ansatz auf die „kognitions- und sprachpsychologische Grundlagenforschung“ (Groeben 1982, S. 4) zurück und wollen individuumszentrierte Grundlagenforschung mit anwendungsbezogenen Ansätzen aus der Instruktionspsychologie verbinden (vgl. Christmann/Groeben 1996; Groeben 1982; Groeben/Christmann 1989, 1996). Sie definieren sich im Gegensatz zur Hamburger Verständlichkeitskonzeption als „theoretisch-deduktive[n]“ (Groeben/Christmann 1996, S. 73) Ansatz, da nicht von „möglichen Textmerkmalen aus die relevanten Dimensionen auf[ge-]baut“ (Groeben 1982, S. 188), sondern „von der Theorie der Verarbeitungspro-

zesse die relevanten Dimensionen und ihnen zugeordnete Textmerkmale ab[ge]leitet“ (ebd.) werden. Überraschenderweise werden in den beiden Ansätzen trotz allem ganz ähnliche Dimensionen der Verständlichkeit postuliert (vgl. Groeben 1982; Groeben/Christmann 1996). Nach Groeben/Christmann (1996, S. 74) sind dies die folgenden Dimensionen. Nach dem Obertitel der jeweiligen Dimension werden zusätzlich „Handlungsanweisungen zur Erreichung optimaler Verständlichkeit bei der Textproduktion“ (ebd.) gegeben.

- 1) Sprachlich-stilistische Einfachheit: Geläufigkeit, Konkretheit, Anschaulichkeit der verwendeten Wörter, Kürze der Sätze, aktiv-positive Formulierungen, keine Nominalisierungen, keine Satzschachtelungen.
- 2) Semantische Kürze/Prägnanz bzw. Redundanz: Beschränkung aufs Wesentliche, keine wörtliche Wiederholung wichtiger Elemente versus synonym oder stilistisch variierende Wiederholungen, Ausdifferenzierungen, Elaborationen.
- 3) Kognitive Gliederung/Strukturierung: Gebrauch von Vorstrukturierungen,<sup>33</sup> sequenzielles Arrangieren, Hervorhebung wichtiger Konzepte, Unterschiede und Ähnlichkeiten von Konzepten, Beispielgebung, Zusammenfassungen.
- 4) Stimulierender kognitiver Konflikt: Neuheit und Überraschung von Konzepteneigenschaften, Inkongruenzen, konfliktevozierende Fragen.

Auch hier deckt sich die Dimension der sprachlich-stilistischen Einfachheit teilweise mit Kriterien aus der Lesbarkeitsforschung: Konzepte wie Geläufigkeit (frequente, familiäre Wörter), Konkretheit und kurze, nicht-verschachtelte Sätze tauchten dort ebenfalls auf.

Unter anderem um der im Hamburger Modell großen Abhängigkeit von Experten-Ratings entgegenzuwirken, plädiert Christmann (2004) für die Erhebung „textnaher Verstehensprodukte“ (ebd.). Unter einigen anderen (verschiedene Multiple-Choice-Verfahren, *cued recall* usw.) stellt sie als Alternativen das Verfahren der subjektiven Informationsmessung vor, das auf der *progressive cloze procedure* nach Shannon (1951) aufbaut. Textnahe Verstehensprodukte sollen das „Verstehen in enger Anlehnung an den Text überprüfen“ und so „das unmittelbare Textverständnis im Sinne einer Vorstellung über den Sinngehalt erfassen“ (Christmann 2004, S. 43). Hinzu treten textferne Methoden (freie Wiedergabe, Beantwortung von Textfragen usw.), bei denen „in hohem Maße auch Gedächtnis- und Schlussfolgerungsprozesse“ (ebd.) beteiligt sind.

<sup>33</sup> Diesen Ausdruck verwenden Groeben/Christmann (1996) als Übersetzung von *advance organizers* (vgl. ebd., S. 69). Dieser Ausdruck wurde durch Ausubel (1960) geprägt.

In der vorliegenden Arbeit wird als Hauptmethode die Messung von Blickbewegungen eingesetzt. Damit befinden wir uns zweifelsohne näher am Verstehens**prozess** denn am Verstehens**produkt**. Inwieweit dies erstrebenswert ist, wird klar, wenn wir uns den Weg des theoretischen Schlusses im Rahmen der Hamburger Verständlichkeitskonzeption vergegenwärtigen. Dieser Weg kann nur über mehrere Schritte erfolgen. Einerseits werden die Bewertungen von ExpertInnen auf den vier postulierten Dimensionen der Verständlichkeit zu Rate gezogen, um eine Einschätzung des Textes als Ganzes zu erlangen. Diese Einschätzungen können mit konkreten Maßen von VersuchsteilnehmerInnen korreliert werden. Diese Maße gehen, wie oben ansatzweise ausgeführt, aus diversen Tests **nach** dem eigentlichen Lesen des Textes hervor. Diese würden gemäß der Einteilung von Christmann (2004) in die Kategorie „textferne Verstehensprodukte“ fallen. Anhand der Ergebnisse dieser Tests kann überprüft werden, wie richtig die ExpertInnen mit ihren Einschätzungen der Textverständlichkeit lagen. Auf vielen Dimensionen schlecht bewertete Texte sollten dann Verstehensprodukte nach sich ziehen, die ebenfalls mangelhaft sind. Andererseits kann anhand der ExpertInnen-Ratings versucht werden, optimierte Fassungen der Texte zu erstellen, die dann wiederum Tests mit Laien unterzogen werden können. Direkte Folge dieses Vorgehens ist, dass (1) nur Rückschlüsse auf Textebene möglich sind und (2) nur die ExpertInnen-Urteile sowie die Testergebnisse der VersuchsteilnehmerInnen als abhängige Variablen zur Verfügung stehen. Ohne ein echtes Online-Maß (in der Terminologie von Christmann (2004) sogenannte textnahe Maße) als Ankerpunkt sind keine detaillierteren Rückschlüsse auf einzelne Abschnitte, Sätze, Phrasen oder Wörter innerhalb der Texte möglich. Das Leseverhalten der VersuchsteilnehmerInnen, operationalisiert über die aus den Blickbewegungen abgeleiteten Lesemaße, ist ein sehr viel direkteres Maß der Prozesse während der Textverarbeitung. Die Körnigkeitsstufe kann dabei zwischen dem Text als Ganzem bis hin zu einzelnen Wörtern oder Wortbestandteilen variiert werden.

Aus einer Kritik (siehe hierzu Göpferich 2006) an den Verständlichkeitskonzeptionen sowohl von Groeben/Christmann als auch des Hamburger Modells erwuchs das Karlsruher Verständlichkeitsmodell (vgl. Göpferich 2006, 2008, 2009). „Der gravierendste Mangel der beiden instruktionspsychologischen Konzepte“ sei demnach, „dass sie noch weitgehend textorientiert bleiben, also textexterner Bezugsgrößen entbehren“ (ebd., S. 296). Göpferich legt Wert darauf, dass die Verständlichkeit eines Textes „keine absolute Größe“ sei, sondern „in Relation zu verschiedenen textexternen und über den Text hinausweisenden Faktoren“ gesehen werden müsse, „die für Verständlichkeitsbetrachtungen einen Bezugsrahmen bilden“ (ebd.). Diesen Bezugsrahmen

versucht das Karlsruher Modell zu geben. Zu den bereits vorgestellten Verständlichkeitsdimensionen treten außerdem zwei weitere hinzu. Mit der Dimension 'Korrektheit' füllt das Karlsruher Modell die Lücke, die in den anderen Modellen dadurch entsteht, dass die Autoren „stillschweigend davon aus[gehen], dass die zu optimierenden Texte und auch die optimierten Versionen fehlerfrei sind“ (ebd., S. 304). Diese Dimension sei hauptsächlich auf das „mentale Denotatsmodell, das im Text vermittelt werden soll, und auf die Kodierung im Text“ (ebd.) zu beziehen. Die Einhaltung von Konventionen („mentales Konventionsmodell“) rechnet Göpferich ebenfalls zu dieser Dimension. Die Dimension der Perzipierbarkeit greift hauptsächlich die unter dem Stichpunkt Leserlichkeit bereits in die Diskussion eingeführten Konzepte und Merkmale auf. Außerdem soll die Leserlichkeit „nonverbaler Informationsträger“ (ebd., S. 309), wie beispielsweise Spiegelstriche oder Aufzählungen, beachtet werden. Gliederungshinweise wie Spiegelstriche werden in den anderen Modellen ebenfalls berücksichtigt. Dort werden solche Merkmale jedoch den Dimensionen 'Gliederung/Ordnung' (Hamburger Modell) bzw. 'Kognitive Gliederung' (Groeben/Christmann) zugeschlagen. Die Erweiterung auf textexterne Bezugsgrößen macht sich insbesondere dadurch fest, dass im Karlsruher Modell die für die Verständlichkeit eines Textes relevanten Faktoren in drei Hauptgruppen aufgeteilt werden: 1) Die Auftragsdaten, 2) die Textproduktions-Eckdaten und 3) die textgebundenen Verständlichkeitsfaktoren. Die Auftragsdaten umfassen die kommunikative Funktion eines Textes und setzen sich zusammen aus dem Zweck des Textes, den AdressatInnen und den SenderInnen des Textes (vgl. Göpferich 2008, S. 296ff.). Die Textproduktions-Eckdaten umfassen, so Göpferich (ebd., S. 298ff.), das mentale Denotatsmodell, das praktisch identisch ist mit dem von Kintsch (1988) definierten Situationsmodell, das während der Textrezeption „vor dem geistigen Auge des Rezipienten entstehen soll und im Idealfall, induziert durch die im Text verwendeten Zeichen und unter Ausnutzung der von ihnen ausgelösten *top-down*-Prozesse, auch entsteht [Hervorh. i. Orig., S.W.]“ (ebd., S. 298). Weiterhin gehört zu den Textproduktions-Eckdaten das mentale Konventionsmodell, das oben kurz angesprochen wurde. Dieses besteht aus den Textsortenkonventionen, d.h. dem Textsortenwissen des Senders und Empfängers. Diese sind im Gegensatz zu juristischen und redaktionellen Richtlinien, dem dritten Bestandteil innerhalb der Textproduktions-Eckdaten, nicht explizit verschriftlicht. Viertes Element ist das Vermittlungsmedium. Dessen Wahl wird von anderen Elementen der kommunikativen Funktion und der Textproduktions-Eckdaten beeinflusst (vgl. ebd., S. 301). Enthält beispielsweise das mentale Denotatsmodell komplexe Sachverhalte, „die sich nur schwer anschaulich verbalisieren lassen, so muss auf ein Medium zurückgegriffen werden, das auch nonverbale Darstellungen erlaubt“ (ebd.).

Göpferich leitet also für das Karlsruher Verständlichkeitsmodell sechs Verständlichkeitsdimensionen ab. Zu den an die anderen Verständlichkeitskonzeptionen angelehnten Dimensionen 'Struktur', 'Prägnanz', 'Motivation' und Simplizität fügt Göpferich die Dimensionen 'Korrektheit' und 'Perzipierbarkeit' hinzu. Mit diesen trägt sie

dem Umstand Rechnung, dass die Richtigkeit eines Textes auf allen Ebenen (Inhalt, Gestaltung, Typographie etc.) sowie die Leichtigkeit, mit der er über die Sinnesorgane aufgenommen und somit den kognitiven Systemen zur weiteren Verarbeitung zugeführt werden kann, zwei weitere wichtige Größen sind, von denen die Textverständlichkeit abhängt. (ebd.)

Göpferich (2008) nimmt für ihr Modell in Anspruch, dass die Dimensionen „präziser definiert [...]“ (ebd., S. 302) und daher bewusst anders benannt sind. Insgesamt scheinen die Überschneidungsbereiche der verschiedenen Verständlichkeitsmodelle trotzdem recht groß zu sein. Am ehesten sind unterschiedliche theoretische und/oder methodische Schwerpunkte zu konstatieren. Diametral entgegengesetzte Konzeptionen von Verständlichkeit sind nicht zu erkennen. Die Ableitung von Vorhersagen für die Verarbeitung konkreter Texte gestaltet sich trotzdem schwierig. Keines der vorgestellten Modelle hält konkrete Instrumentarien bereit, Vorhersagen auf der Mikroebene der Verarbeitung<sup>34</sup> zu machen. Ganz allgemein wird natürlich nach allen Modellen vorhergesagt, dass es zu Verständnisschwierigkeiten führen müsste, wenn relevante Prinzipien verletzt werden. Auf welche Ebene sich diese Verständnisschwierigkeiten auswirken müssten und wie genau diese zu operationalisieren wären, bleibt jedoch unklar. Gerade hinsichtlich der Blickbewegungen beim Lesen der Texte werden keine expliziten Prognosen gemacht. Diese sind aber eventuell nicht das Ziel solcher Modelle. Ihr Mehrwert, beziehungsweise der Zweck, zu dem sie hier referiert wurden, liegt darin, auf sprachinterne und -externe Phänomene aufmerksam zu machen, die ausschlaggebend für die Verständlichkeit eines Textes sind. In der vorliegenden Arbeit werden Hypothesen hauptsächlich in Kapitel 6 herausgebildet, in dem die entsprechenden Ergebnisse direkt angeschlossen werden können. Nun wollen wir zunächst auf der sprachlichen Ebene verweilen und sozusagen „hineinzoomen“ in die Textgattung der Rechtssprache. Im folgenden Abschnitt werden die wichtigsten linguistischen Merkmale genannt, die die Fachsprache des Rechts mit anderen Fachsprachen gemeinsam hat, aber insbesondere natürlich, welche sie zu diesen abgrenzt.

<sup>34</sup> In einer Klassifikation von Methoden zur Bestimmung der Verständlichkeit, in der Göpferich Schriver (1989) folgt, findet die Messung von Blickbewegungen denn auch keine Erwähnung. Im Originaltext ist diese interessanterweise genannt (Schriver 1989, S. 242).

### 3.3 Komplexitätsmerkmale der Rechtssprache

*Die Tarifierhöhung der Tabellenentgelte erfolgt unter dem Vorbehalt der Rückforderung und unter Ausschluss der Berufung auf den Wegfall der Bereicherung.*

Gehaltsabrechnung Uni Freiburg

Hansen-Schirra/Neumann (2004, S. 169) weisen darauf hin, dass sich Besonderheiten der Rechtssprache abseits „Fragen der Terminologie“ mit den Merkmalen decken, „die auch für die Fachsprachen anderer Disziplinen beschrieben wurden“:

So zielen grammatische Konstruktionen, die in Fachsprachen zu finden sind, auf eine sachliche Darstellung der Sachverhalte und auf Ausdrucksökonomie ab. Die Verwendung von Attribuierung [...] dient der komprimierten Beschreibung der Inhalte. Die Inhaltsbezogenheit dieser Texte kommt auch dadurch zum Ausdruck, dass viele Nominalisierungen in Kombinationen mit bedeutungsarmen Verben zu finden sind. (ebd.)

Auch im *Freiburg Legalese Reading Corpus* (FLRC), dem in dieser Arbeit erhobenen und ausgewerteten Lesekorpus, sind Beispiele für solche Konstruktionen mit bedeutungsarmen Verben zu finden. Als prototypisch kann Beispiel 6 aus der Entscheidung zum unerlaubten Aufenthalt in Deutschland<sup>35</sup> gesehen werden, in dem die verbale Konstruktion vermieden wird und durch das Muster [<Nominalisierung> <Vollzugsverb>] ersetzt wird.<sup>36</sup> In einem fortlaufenden – die Auslassungen im Beispiel enthalten im Original lediglich einen Einschub und Paragraphenverweise – Textabschnitt von zwei Sätzen tritt die entsprechende Konstruktion dreimal auf: Die Abschiebung wird durchgeführt oder geführt, die Duldung wird erteilt.

- (6) Dabei hat die Ausländerbehörde zu prüfen, ob und gegebenenfalls wann eine *Abschiebung* des Ausländers *durchgeführt* und zu welchem Zeitpunkt ein eventuelles Abschiebungshindernis behoben werden kann. Schon dann, wenn sich herausstellt, dass die *Abschiebung* nicht ohne Verzögerung *geführt* werden kann oder der Zeitpunkt

<sup>35</sup> Die im FLRC enthaltenen Texte werden in Abschnitt 5.1.1 noch im Einzelnen vorgestellt.

<sup>36</sup> 'Ersetzen' klingt hier nach einem sehr bewussten Vorgang, in dem eine schon vorhandene verbale Konstruktion mit diesem Muster ausgetauscht wird. Das ist in der Entstehung dieser Texte höchstwahrscheinlich nicht der Fall. Vielmehr sind die Muster der entsprechenden Textgattung vermutlich schon so bei den Produzierenden verankert, dass sie die primäre Wahl bei der Produktion der Texte sind.



der Abschiebung ungewiss bleibt, ist [...] als „gesetzlich vorgeschriebene förmliche Reaktion auf ein Vollstreckungshindernis“ eine *Duldung* zu erteilen [...].

Als weiteres Merkmal führen Hansen-Schirra/Neumann (2004, S. 169) „die häufige Verwendung passivischer Konstruktionen“ an, durch das „die Nennung des Agens, des Handlungsträgers, vermieden“ wird, „was eine objektive und sachbezogene Darstellung ermöglicht“. Auch dieses Charakteristikum kann in Beispiel 6 beobachtet werden: Das Agens der Abschiebe-Handlung wird nicht genannt (es ist auch nicht die Ausländerbehörde, die das Agens der Prüfhandlung ist), ebenso wenig das Agens, das die Duldung erteilt. Wer die vorliegende Dissertation bis zum aktuellen Punkt durchgehend gelesen hat, wird bemerkt haben, dass die eben genannten Punkte offenbar nicht nur auf die Rechtssprache zutreffen, sondern auch für die deutsche Wissenschaftssprache gelten. Ein weiteres Merkmal, das ebenfalls zu den „eher allgemeinen grammatischen Charakteristika von Fachsprachen“ (ebd., S. 170) gehört, aber auch in der Rechtssprache oft anzutreffen ist, betrifft das Satzbaumuster: Klassische Thema-Rhema-Strukturen, also die alternierende Abfolge bekannter und neuer Information, ist in Fachsprachen seltener anzutreffen als in allgemeinsprachlichen Textgattungen (vgl. ebd.). Beispiel 6 ist etwas zu kurz, um diese Aussage zu evaluieren, und die Einschätzung kann wohl nicht uneingeschränkt auf alle Teile bundesverfassungsgerichtlicher Entscheidungen ausgedehnt werden. Diese enthalten, sofern es der Gegenstandsbereich zulässt, oft längere Passagen, die geradezu narrativ sind. In solchen Textabschnitten folgt die Herleitung einer bestimmten Situation dann auch eher klassischen Thema-Rhema-Strukturen. Diese Passagen sind interessanterweise jene, die in den entsprechenden Pressemitteilungen wegfallen, weil sie Information enkodieren, die für Pressemitteilungen weniger relevant sind. Als „[k]ennzeichnend für die Fachsprache des Rechts und der Verwaltung“ identifizieren Hansen-Schirra/Neumann (2004, S. 170) in Anlehnung an Oksaar (1988) und Wagner (1981)

- lange Sätze,
- Personifizierungen unbelebter Gegenstände oder Sachverhalte,<sup>37</sup>
- nominale Umschreibungen statt einfacher Verben,<sup>38</sup>

<sup>37</sup> Auch dieses Merkmal kann in der Wissenschaftssprache beobachtet werden. So sind Konstruktionen wie ‘Diese Arbeit beschäftigt sich mit ...’ keine Seltenheit mehr. Hansen/Teich (1999, S. 319f.) besprechen dieses Phänomen in Zusammenhang mit Übersetzungen. Auch De Vogelaer (2007) geht im Rückgriff auf Hawkins (1986) auf sogenannte nicht-agentive Subjekte ein.

<sup>38</sup> Ein prototypisches Beispiel ist das Zitat, das dem aktuellen Abschnitt vorangestellt ist. Es entspringt dem „echten Leben“.



- Komplexität der Darstellungsweise durch Derivationen (also die Bildung neuer Wörter mit Affixen),
- Ketten einander untergeordneter Substantive,
- Genitivattribute sowie
- formelhafte Wendungen und archaische Formen.

Insbesondere über das Phänomen der langen Sätze und die zuweilen exzessive Attribuierung wird in dieser Arbeit noch zu reden sein. Beide sprachlichen Phänomene sind, ebenso wie die oben angesprochenen Nominalisierungen, über einen eigenen Komplexitätstypus im Reformulierungsteil des *Freiburg Legalese Reading Corpus* vertreten. In Abschnitt 5.2.1 werden die verwendeten Texte sowie deren Reformulierungen anhand korpuslinguistischer Maße untersucht. Hansen-Schirra/Neumann (2004) analysieren ebenfalls ausgewählte textgattungsspezifische Merkmale anhand eines Korpus aus Pressemitteilungen und Urteilen des Bundesverfassungsgerichts. Diese werden mit einem Referenzkorpus verglichen, das aus 15 (anderen) Textgattungen besteht. Die berichteten Ergebnisse bezüglich der Satzlängen (ebd., S. 172f.) in den Urteilen und Pressemitteilungen sind sehr ähnlich zu jenen, die in der vorliegenden Arbeit berichtet werden. So beträgt die durchschnittliche Satzlänge der von ihnen untersuchten Urteile 29,5 Wörter (im FLRC: 28,1 Wörter). Die durchschnittliche Satzlänge der Pressemitteilungen beträgt bei Hansen-Schirra/Neumann 20,4 Wörter (FLRC: 21,9 Wörter). Die Autorinnen weisen darauf hin, dass dies die durchschnittliche Satzlänge in ihrem Referenzkorpus noch knapp unterschreitet (dort: 21,6 Wörter). Sie schließen daraus, dass „die Autorinnen der Pressemitteilungen – sei es bewusst oder unbewusst – die Texte im Hinblick auf das Merkmal Satzlänge für die Rezipienten vereinfacht haben“ (ebd., S. 173). Im FLRC sind außerdem noch drei Zeitungsartikel enthalten. Diese weisen mit durchschnittlich 16,2 Wörtern noch einmal deutlich kürzere Sätze auf. Auch bei anderen Maßen – nicht alle sollen in diesem Abschnitt behandelt werden, da Abschnitt 5.2 eine eigene Korpusanalyse vorstellt – lassen sich Parallelen mit den Analysen von Hansen-Schirra/Neumann (2004) feststellen: Die lexikalische Vielfalt (operationalisiert über die Type-Token-Ratio) liegt für die beiden Textgattungen in ganz ähnlichen Regionen (Urteile Hansen-Schirra und Neumann: 0,39/FLRC: 0,30; Pressemitteilungen: 0,43/0,50). Hier sei allerdings auf den deutlichen Kontrast zwischen Type-Token-Verhältnis und die *mean segmental type-token ratio* verwiesen, auf den in Abschnitt 5.2.2 eingegangen wird. In Kürze: Die Type-Token-Ratio ist hochgradig von der Textlänge beeinflusst, was ihre Aussagekraft in Bezug auf die lexikalische Vielfalt von Texten deutlich einschränkt. Die *mean segmental type-token ratio* versucht, diesen Effekt abzufangen.

Weitere Ergebnisse der Studie von Hansen-Schirra/Neumann (2004) zeigen, dass für Urteile des Bundesverfassungsgerichts insbesondere formelhafte Ausdrücke („Antrag auf Erlass einer einstweiligen Anordnung“, „das Verfahren ist einzustellen / wird eingestellt“) als häufige Kollokationen auftreten. In der vorliegenden Arbeit kann am ehesten Abschnitt 7.2 zu dieser Diskussion beitragen, da dort die Einflüsse von Bi- und Trigrammfrequenz auf die Textrezeption untersucht werden. Bi- und Trigramme können nicht analog zu Kollokationsanalysen gesehen werden (zu verschiedenen Kollokationsmaßen vgl. beispielsweise Daudaravičius/Marcinkevičiene 2004; Gries 2013), bestenfalls als grobe Annäherung. Kollokationsmaße sind für das FLRC jedoch eher wenig aussagekräftig, weil es in seiner Natur als Lesekorpus weniger Wörter umfasst als die Korpora, anhand derer Kollokationsmaße üblicherweise berechnet werden. Die Bi- und Trigramm-Frequenzen wurden in der vorliegenden Arbeit aus einer lexikalischen Datenbank extrahiert. Ein letzter Punkt, der von Hansen-Schirra/Neumann (2004) in der kontrastiven Korpusuntersuchung als Merkmal der Rechtssprache bestätigt wurde, ist die spezielle Verwendung des Konjunktivs. Dieser liegt „auf den ersten Blick [...] nicht unbedingt als Verständnishindernis auf der Hand“ (ebd., S. 178). Laien sind sich jedoch der „fachspezifischen Funktion zur Kennzeichnung des streitigen Sachverhalts in Urteilen“ (ebd., S. 179) nicht bewusst, wie Neumann/Hansen-Schirra (2004) zeigen konnten. Für Laien kennzeichnet der Konjunktiv im allgemeinsprachlichen Verständnis hauptsächlich indirekte Rede. In den Urteilen verdreifacht sich der Anteil der Verben im Konjunktiv im Vergleich zum Referenzkorpus (18,8 zu 6,9%). Die Pressemitteilungen liegen mit 7,7% hingegen recht nah beim Wert für das Referenzkorpus. Hansen-Schirra/Neumann (2004) interpretieren dies als „eine Art Rephrasierung für die allgemeine Öffentlichkeit“ (S. 179). In der vorliegenden Arbeit spielt die Gegenüberstellung Konjunktiv – Indikativ keine Rolle. Grundsätzlich wäre diese Frage aber ebenfalls anhand des FLRC untersuchbar. Man müsste diese Annotations-ebene allerdings erst zum Korpus hinzufügen.

Über die Verwendung eines Referenzkorpus konnten Hansen-Schirra/Neumann (2004) zeigen, dass sich die Urteile des Bundesverfassungsgerichts in einigen fachsprachlichen Merkmalen, die zuvor bereits definiert wurden, von Allgemesprache abheben. Sie gehen außerdem auf die Unterschiede zwischen Urteilen und Pressemitteilungen ein – ein Kontrast, der auch in der vorliegenden Arbeit relevant ist. Sie unterstreichen dabei „den Charakter der Pressemitteilungen als intralinguale Übersetzungen“ (ebd., S. 182). Diese Auffassung vertreten wir ebenfalls (Wolfer/Hansen/Konieczny 2013), als wir Auszüge der Ergebnisse der vorliegenden Arbeit vorstellten. Auch in Wolfer et al. (2013) wählten wir einen ähnlichen Ansatz. Unsere Datenbasis bestand dort

aus einem Lesekorpus populärwissenschaftlicher Texte, die in gewisser Hinsicht als Übersetzungen von Wissenschaftstexten in eine allgemein verständlichere Form angesehen werden können.

Nachdem wir einen gewissen Überblick gewonnen haben, welche Komplexitätsmerkmale Rechtssprache aufweist, bleibt noch die Frage zu klären, inwiefern Rechtstexte überhaupt hinsichtlich ihrer Verständlichkeit optimiert werden müssen. Diese Frage wurde zwar bereits in der Einleitung gestreift, an dieser Stelle sollen nun aber widerstreitende Stimmen ausführlicher dargestellt werden.

### 3.4 Müssen Rechtstexte verständlich sein?

*Nie hat ein Dichter die Natur so frei ausgelegt  
wie ein Jurist die Wirklichkeit.*

Jean Giraudoux (1882-1944)

Towfigh (2009) bezieht von juristischer Seite Stellung zur Komplexität von Gesetzestexten. Auf linguistische Komplexität geht er dabei nur am Rande ein. Vielmehr definiert er die Komplexität von normativen Regelungen, also Gesetzen, anhand zweier Dimensionen: Dichte und Interdependenz. „*Dichte* meint die Anzahl der zu berücksichtigenden normativen Informationen, also die Zahl der Tatbestandsvoraussetzungen und Rechtsfolgeanordnungen bzw. Wirkungen innerhalb einer positiven Rechtsnorm. [Hervorhebung im Original, S.W.]“ (ebd., S. 31f.). Weiter führt er aus, dass „voraussetzungsvolle, vorgefragten- oder begriffsreiche Normen dicht im hiesigen Sinne“ (ebd.) seien. Die Komplexitätsdimension der Dichte bezieht sich also auf das einzelne Gesetz selbst. Zwar ist die Anzahl der in die Welt weisenden Verknüpfungen relevant, weil Voraussetzungen und Folgen der Norm mit einzubeziehen sind, die Verbindungen mit anderen Rechtsnormen ist aber für die Dimension der Dichte unwichtig. Diese Verbindungen zu anderen Rechtsnormen sind Teil der Komplexitätsdimension Interdependenz, genauer der externen Interdependenz (vgl. ebd., S. 32):

Durch externe Interdependenzen erhöht sich die Anzahl der bei der Beantwortung einer Rechtsfrage zu beachtenden Normen;[...] sie macht das, was als „Normenschungel“[...] oder „Hyperlexie“[...] bezeichnet wird, auch auf der Ebene der einzelnen Norm und bei der Beantwortung einer einzelnen Frage anhand dieser Norm spürbar.[...] Interdependenzen sind deswegen besonders heikel, weil sie die Zahl der zu berücksichtigenden Elemente durch mehrfach gestufte Wechsel- und Rückwirkungen dramatisch[...] zu erhöhen vermögen, und das über überschaubare Regelungskomplexe hinweg. [Fünf Fußnoten aus Originalzitat entfernt, S.W.]

Bestehen Verbindungen zwischen „Voraussetzungen und Wirkungen ein und derselben Norm“ (ebd.), spricht Towfigh (2009) von interner Interdependenz. Das Wechselspiel zwischen Dichte und Interdependenz charakterisiert er so, dass höhere Normendichte zu niedrigerer Interdependenz führen kann (und *vice versa*). Werden die Inhalte dependenter Regelungen im Gesetzestext selbst genannt, erhöht sich freilich die Dichte. Allerdings muss nur noch in Spezialfällen auf andere Normen zurückgegriffen werden – die Interdependenz ist also sehr gering. Auch der umgekehrte Fall ist leicht vorstellbar. Towfigh (2009, S. 33f.) gibt das Beispiel der landesrechtlichen Bauordnung, derzufolge eine Baugenehmigung zu erteilen ist, wenn dem Vorhaben öffentlich-rechtliche Vorschriften nicht entgegenstehen. Offensichtlich ist die Dichte dieser Norm sehr gering, „[e]s gibt lediglich eine geschriebene Tatbestandsvoraussetzung und eine klare Rechtsfolge“ (ebd.). Die Menge an Rechtsnormen, die über externe Interdependenzen mit dieser einfachen Norm verbunden sind, ist aber riesig. Towfigh (2009) bleibt unklar in seinen Ausführungen, ob er Dichte und Interdependenz als gegenläufige Konzepte definiert. Es scheint aber möglich, dass eine Norm sowohl eine hohe Dichte als auch eine hohe Interdependenz aufweist. Klar ist: „Komplexität ist [...] keine dichotome [...], sondern eine graduelle [...] Normeigenschaft.“ (ebd., S. 34). Darüber hinaus ist Komplexität, wie sie bei Towfigh (2009) beschrieben wird, weder positive, noch negative Eigenschaft einer Norm. Es können lediglich die Folgen „beobachtet und bewertet“ werden, so kann Komplexität „das Verständnis und die Anwendbarkeit der Normen erschwer[en]“. Mit höherer Normkomplexität werden die Adressaten unsicherer, welche Folgen das Gesetz für sie selbst haben könnte. Als Begründung argumentiert Towfigh (2009, S. 35) auf Basis kognitionswissenschaftlicher Erkenntnisse: „Je höher die Dichte und die Interdependenz, desto mehr kognitive Kosten (Aufmerksamkeit) fallen für das Entpacken von Inhalten an.“ Wie genau der Zusammenhang zwischen Kosten und Unsicherheit ist, bleibt unklar. Towfigh gesteht trotzdem Experten (und hier sind wohl juristische Experten, in seiner Terminologie „Intermediäre“, gemeint) optimierte „*Chunking*-Strategien [Hervorhebung im Original, S.W.]“ zu, anhand derer sie „mehr Informationen im Sinne größerer und komplexerer Informationseinheiten verarbeiten“ (ebd.). Dass dies wirklich die Kompetenz ist, die die/den juristische/n ExpertIn vom juristischen Laien unterscheidet, ist meines Erachtens zu bezweifeln – der Autor selbst kommt auf diese Aussage nicht mehr zurück. Schlüssiger wirken die Ausführungen zur perspektivischen Abhängigkeit der Komplexität. So kann Komplexität von zwei Blickwinkeln her eingeschätzt werden: Liegt eine bestimmte Fragestellung oder ein konkreter Sachverhalt vor, der anhand eines Gesetzestexts aufgelöst werden muss, spricht Towfigh (ebd., S. 35) von „konkreter Komplexität“. In diesem Fall könne das „abzuarbeitende Normprogramm verkürzt

sein, weil einzelne normative Anknüpfungspunkte ausgeblendet werden können“ (ebd.). Betrachtet man die Norm als solche, also ohne konkrete Fragestellung, lässt sich Komplexität auch abstrakt erfassen. Dann gehen die komplette Dichte sowie alle Interdependenzen in die Betrachtung mit ein. Die Komplexität einer Norm kann sich also ändern, je nachdem ob und mit welcher Fragestellung man an sie herantritt. Außerdem ist die Komplexität einer Norm abhängig von der/dem Betrachtenden. Je nachdem, ob die/der NormgeberIn, ein/e NormadressatIn oder eine/e Rechtskundige/r ein Gesetz vorliegen hat, „kann sich die Bewertung der Komplexität relativ zum Standpunkt des Betrachters verändern“ (ebd., S. 36). So ist es leicht vorstellbar, dass mit steigender Kenntnis des Rechtssystems mehr Interdependenzen zu anderen Normen sichtbar werden. Doch auch die andere Effektrichtung ist denkbar: Je mehr wir über das Rechtssystem als Ganzes wissen, desto eher werden wir einschätzen können, welche Teile einer Norm und welche Verbindungen zu anderen Gesetzen für die Beantwortung einer konkreten Sachfrage irrelevant sind. Rechtsexpertise kann also – anders ausgedrückt – sowohl zu einer Ausweitung als auch zu einer Einschränkung des Problemraums führen, gibt aber im Gegensatz zur Nichtkenntnis des Rechtssystems immer Orientierung im „Gesetzesdschungel“. Rechtsexpertise schlage sich dabei unter anderem in der Kenntnis juristischer Begriffsbildung nieder. JuristInnen können so „Rechtskonstruktionen [...] im Lebenssachverhalt wieder [erkennen]“ (ebd.). Towfigh charakterisiert dies als erworbene Gestalt- oder Schablonenerkennung (vgl. ebd.), anhand derer Rechtskonstruktionen – wie beispielsweise ein Verwaltungsakt – erkannt werden können. Durch diesen Vorgang können „größere, im konkreten Einzelfall unerhebliche Teile der Norm ‘ausgeblendet’“ (ebd., S. 37) werden – diese Schablonenerkennung ist also als ein Werkzeug zur Reduktion konkreter Komplexität anzusehen.

Durch das Gebot der Normenklarheit soll – so Towfigh weiter – Komplexität kontrolliert werden. Dabei können Regelungen außer Kraft gesetzt werden, die unverständlich und daher nicht justitiabel sind. Unverständlichkeit macht Towfigh dabei an der Komplexität der Regelung nach oben genannter Definition fest. Abgeleitet wird das Gebot der Normenklarheit meist aus dem Rechtsstaatsprinzip und der damit verbundenen Rechtssicherheit (vgl. hierzu auch Limbach 2008, S. 373). „In Einzelfällen schimmern Anleihen am Demokratieprinzip und dem Grundsatz der Gewaltenteilung durch“, so Towfigh (2009, S. 39). Auch Schröder/Würdemann (2008, S. 325) argumentieren in diese Richtung: „Die Forderung nach Erkennbarkeit des Rechts zielt darauf ab, dass den Bürgern, denen durch Gesetze Rechte und Pflichten entstehen, diese auch direkt aus dem Gesetz erkennen können.“ Es können an dieser Stelle nicht alle Urteilsprüche und Entscheidungen angeführt und besprochen

werden, die sich jemals mit dem Prinzip der Normenklarheit befasst haben. Dies leistet (Towfigh 2009, S. 39ff.) ohnehin juristisch kompetenter. Interessant ist dabei unter anderem, dass das Gebot der Normenklarheit dann strenger auszulegen ist, wenn der Eingriff in die Grundrechte des Adressaten stärker ist. Doch das bedeutet im Umkehrschluss nicht, dass bei Gesetzen, die nicht in die Grundrechte eingreifen, Normenklarheit irrelevant ist – wie das Bundesverfassungsgericht ausdrücklich feststellt: „Nicht nur bei Eingriffen in die Freiheitssphäre des Einzelnen [...], sondern auch bei der Gewährung von Leistungen und deren zivilrechtlicher Behandlung müssen die Normen in ihrem Inhalt entsprechend ihrer Zwecksetzung für die Betroffenen klar und nachvollziehbar sowie in ihrer Ausgestaltung widerspruchsfrei sein“ (BVerfGE 108, 52 <75>).

Ein Punkt, der für die Argumentation bezüglich der vorliegenden Arbeit interessant ist, ist der Bezug auf die/den sogenannten NormadressatIn. Die Bewertung, ob eine Norm klar genug ist, richtet sich am Verständnis der/des AdressatIn aus. Denn „[w]enn der Adressat den Normbefehl nicht versteht, kann er ihn auch nicht befolgen“ (Towfigh 2009, S. 42). Interessanterweise schließt das aus, dass Intermediäre als Bezugspunkt der Komplexitätsbewertung gelten. Zur Gruppe der Intermediäre zählen laut Towfigh (ebd., S. 43) „Rechtsanwälte, Notare, Steuerberater, Verbände, aber auch Gerichte und die Verwaltung“. In bestimmten Rechtsbereichen werden außerdem „Techniker“ (Patentanwälte, Ingenieure, Architekten usw.) subsumiert. Hier setzt Towfigh mit seiner Kritik unter anderem an: Denn die Rechtssprechung ist bezüglich des Einbezugs von Intermediären nicht so konsistent, wie es auf den ersten Blick scheinen mag. So scheint beispielsweise der Bundesfinanzhof zumindest in Ausnahmefällen davon auszugehen, dass sich potenzielle NormadressatInnen – im Grunde also alle Menschen, die in der Bundesrepublik Deutschland Steuern zahlen – von RechtsexpertInnen beraten lassen. In diesem Fall solle es dann genügen, wenn das fragliche Gesetz von den eingeschalteten ExpertInnen verstanden werden kann (vgl. ebd.). Aktueller Stand zur juristischen Debatte der Normenklarheit sei, so Towfigh (ebd.), dass „im Einzelfall plausible Entscheidungen zustande kommen“, die aber nicht von juristischer Dogmatik geleitet sind, sondern daher rühren, dass „schlaue Richter“ die „weite[n] Spielräume“ zu nutzen wissen, die diese bereitstellt. Weitere Kritik bezüglich der Ausrichtung am Normadressaten als Bezugspunkt der Komplexitätsbewertung von Gesetzen lässt sich in den folgenden Punkten wie folgt zusammenfassen:

Orientiert man sich am Normverständnis eines objektivierten Adressaten, ignoriere man dabei die „tatsächlich zu beobachtende Gesetzeskenntnis in der Gesellschaft“ (ebd., S. 46). Es sei eine „elementare Fiktion“ für das Funktionie-



ren des Rechts, dass die BürgerInnen Kenntnis vom Gesetz haben. Die Unkenntnis wird aber ebenfalls den BürgerInnen zugerechnet – was wiederum selbst Norm ist (vgl. ebd., S. 48). Der gemeine Normadressat habe zwar für gewöhnlich „eine diffuse Kenntnis des *Rechts*, nicht aber eine konkrete des *Gesetzes* [Hervorh. i. Orig., S.W.]“ (ebd., S. 46). Wenn es zu Verständnisschwierigkeiten komme, dann liege das nicht daran, dass es eine „*sprachliche* Herausforderung [Hervorh. i. Orig., S.W.]“ (ebd., S. 48) sei, sondern es scheitere in aller Regel am fehlenden rechtlichen Vorverständnis. Dies liege unter anderem darin begründet, dass die Rechtssprache eine Fachsprache sei.

Es sei fraglich, ob weniger komplexe, besser verständliche Gesetze tatsächlich häufiger befolgt werden als komplexere, unverständliche Normen (vgl. ebd., S. 50ff.). Towfigh hält es zwar für „plausibel“ (ebd., S. 53), dass Gesetze mit einer möglichst geringen Dichte und möglichst wenigen Interdependenzen gut funktionieren, weil ihr „Regelungsgehalt bzw. ihre Verhaltensanforderung verhältnismäßig leicht einsichtig und erlernbar“ (ebd.) sei. Er möchte den Umkehrschluss jedoch nicht zulassen, dass höhere Komplexität keine sachgerechten Ergebnisse ermögliche. An dieser Stelle sei auf einen Text von Enzensberger (2001) verwiesen, dessen Beitrag zur „Arena Verständnis des Rechts“ des Rechtshistorischen Journals geradezu als Wutrede bezeichnet werden kann. Enzensberger schließt:

Weite Zonen des Alltagslebens haben sich längst in der Illegalität eingerichtet. Schwarzarbeit gilt als selbstverständlich, Steuervermeidung ist zum Volkssport geworden, Schmuggel und Subventionsbetrug gehören zur Normalität der europäischen Einigung. Daß [sic] die Bevölkerung zur Notwehr greift, ist ebenso unvermeidlich wie die Ausbreitung unverständlicher Normen. (Enzensberger, S. 524)

Enzensberger etabliert hier einen Zusammenhang zwischen Rechtsverständnis und Rechtsbefolgung. Die BürgerInnen befolgten in vielen Fällen das Recht deshalb nicht, weil sie es nicht verstünden und es deshalb nicht akzeptierten. Den Grund für die Komplexität der Rechtssprache sieht er in einem elitären Herrschaftsanspruch von „Ministerialbürokratien, Parlamentsausschüsse[n], Richter[n], Staatsanwaltschaften und Advokatur“ (ebd., S. 523), die gemeinsam das Interesse teilten, dass „die Sphäre des Rechts ein Arkanaum bleibt“ (ebd.). Folgerichtig sieht er die Unverständlichkeit des Rechts nicht als Fehler, sondern als beabsichtigt.<sup>39</sup> Towfigh fordert ein „an Juristen

<sup>39</sup> Der bloße Vorwurf genügt natürlich weder, JuristInnen einen über die Sprache vermittelten Herrschaftsanspruch nachzuweisen, noch zu beweisen, dass verständlichere Gesetze häufiger befolgt würden. Die von Enzensberger (2001) niedergeschriebene Meinung kann aber sicherlich stellvertretend für Viele gelten und sollte deshalb hier Gehör finden.



geeichte[s] Gebot[.] der Normenklarheit“ (Towfigh 2009, S. 69). Dies sei aus demokratischen Gründen vertretbar (vgl. ebd., S. 61ff.). Ohnehin sei das Normenklarheitsgebot eher aus Sicht der Gewaltenteilung (vgl. ebd., S. 63ff.) und teilweise aus Grundrechtsperspektive (vgl. ebd., S. 65) zu motivieren – beidem stehe die Ausrichtung an juristischem Fachpersonal bzw. an Intermediären nicht entgegen. Einer der Gründe, den Towfigh außerdem mehrfach anführt, ist folgender: Juristisch nicht gebildete BürgerInnen könnten nicht zwischen der eingeführten rechtlichen Komplexität, die sich aus Dichte und Interdependenz zusammensetzt, und sprachlicher Komplexität unterscheiden (vgl. ebd., S. 58, 72). Folge von hoher Komplexität sei in beiden Fällen Unverständnis. Empirisch untermauert ist diese Aussage jedoch nicht. Meines Erachtens sind sehr wohl Fälle denkbar, in denen die von den Gesetzen betroffenen BürgerInnen zwischen den beiden Ebenen der Komplexität unterscheiden können. Überprüfbar wäre dies in einer Studie, in der die juristische Komplexität gleich bleibt, die Komplexität der sprachlichen Realisierung jedoch variiert. Im Ansatz ist das bei den Reformulierungsstudien der vorliegenden Arbeit der Fall. Allerdings mussten die VersuchsteilnehmerInnen hier weder explizite Einschätzungen zu juristischer noch zu sprachlicher Komplexität abgeben. Auch bleibt in der Argumentation von Towfigh (2009) unklar, warum es juristisch gebildeten Menschen soviel einfacher fallen soll, zwischen diesen Ebenen der Komplexität zu unterscheiden.

Towfigh (2009) eröffnet also die Perspektive einer Komplexitätseinschätzung auf juristischer Ebene und fordert in der Folge eine Neuausrichtung des Normenklarheitsgebots an juristisch ausgebildetem Fachpersonal. Dabei kommt der sprachlichen Komplexität meist eher eine untergeordnete Rolle zu und sie ist auch nicht Fokus seiner Argumentation. Daraus jedoch zu folgern, dass sprachliche Überarbeitungen von juristischem Textmaterial überflüssig sind, wäre zu kurz gegriffen. Wenn Towfigh (2009) abschließend darauf hinweist, dass „nicht jede sprachlich misslungene Fassung einer Norm [...] verfassungswidrig“ (S. 73) sei, mag das stimmen. Der Maßstab scheint bei einer solchen Aussage aber auch etwas verschoben. Wenn es nur noch darum geht, ob ein Gesetz wirklich in so starkem Maße sprachlich misslungen ist, dass es verfassungswidrig wird, klingt das doch sehr binär. Es sind Paare von Rechtstexten denkbar, bei denen beide Texte nicht verfassungswidrig wären, der eine aber trotzdem verständlicher ist. Für sprachliche Komplexität gilt das, was Towfigh auch für juristische Komplexität einfordert: Sie ist ein kontinuierliches Konzept.

Eine Voraussetzung dafür, dass JuristInnen als neuer Bezugspunkt für das Gebot der Normenklarheit gelten sollen, ist, dass eben diese mit der Komplexität von Gesetzen umgehen können. Doch was sagen die juristischen Exper-

tInnen zum Umgang mit sprachlicher Komplexität und ihrer eigenen Rolle im deutschen Rechtssystem? Eichhoff-Cyrus/Strobel (2009) konnten zeigen, dass sich die Mehrzahl der JuristInnen tatsächlich in einer Intermediär-Rolle sehen: Lediglich 11% der Befragten verneinten die Frage, ob sie sich selbst in einer Vermittler-Rolle sähen (vgl. Eichhoff-Cyrus/Strobel 2009, S. 141). Alle anderen bejahen diese Frage (63%) oder stimmen ihr zumindest situationsabhängig zu (27%). JuristInnen scheinen also größtenteils die Rolle als Intermediäre anzunehmen, die ihnen unter anderem auch Towfigh (2009) zuweisen möchte. Wäre die Rechtssprache als Fachsprache für JuristInnen völlig verständlich, könnte man hier die Argumentation beenden. Denn für ein reibungslos funktionierendes Rechtssystem wären viele Voraussetzungen erfüllt. Allerdings gibt die überwältigende Mehrheit (85%) der befragten JuristInnen an, dass sie im Zuge ihrer Arbeit in Situationen gelangen, in denen sie selbst Verständnisprobleme mit der Rechtssprache bekommen, davon 12% oft. Entsprechend stimmen auch 76% der Befragten der Aussage zu, dass eine Vereinfachung der Rechtssprache zu besseren Arbeitsbedingungen für juristische Fachleute führen würde, davon 28% auf jeden Fall. Gefragt danach, was ihnen zufolge in erster Linie verbessert werden sollte, ist ein recht eindeutig sprachliches Merkmal auf dem ersten Rang: 40% geben an, dass lange Sätze mit Einschüben in mehrere Sätze aufgeteilt werden sollten. 23% bemängeln die Definitionsgenauigkeit juristischer Fachtermini – ein Merkmal, das eher in Richtung der juristischen Komplexität geht. Die Ersetzung veralteter Wörter/Begriffe durch neue (19%) befindet sich im Schnittbereich, ebenso wie die verständlichere Gestaltung von Gesetzestexten durch eine übersichtliche Gliederung (17%).

Diese Befragung ist ein relativ deutlicher Hinweis darauf, dass es zu kurz gedacht ist, das Gebot der Normenklarheit einzuschränken bzw. auf juristische Fachleute neu auszurichten. Zumindest scheint es kein Allheilmittel zu sein, das Verständlichkeitsoptimierung juristischer Fachtexte überflüssig macht. Die ehemalige Verfassungsgerichtspräsidentin Jutta Limbach macht darauf aufmerksam, dass ganz besonders Verfassungsnormen „abstrakte[.] Sollenssätze“ (Limbach 2008, S. 374) seien, die sich durch einen „geringen Bestimmtheitsgrad“ (ebd.) auszeichneten. Der Weg von der sehr generellen Norm zur Entscheidung über den konkreten Einzelfall kann dabei weit und kompliziert sein. Die Kompetenz, „die Spannung zwischen dem Allgemeinen (dem Gesetz) und dem Besonderen (dem Fall) [zu] überbrücken“, gesteht Simon (2004, S. 410) denn auch nur „fachmännische[m] Verstand“ (ebd.) zu, also ausgebildeten JuristInnen. Trotzdem legt Limbach (2008) Wert darauf, dass das kein Grund sei, sich vom „Verständlichkeitspostulat“ (S. 376) zu verabschieden. Denn es ginge „hier nicht nur um ein Stil- und Informationsproblem, sondern

um einen eminent wichtigen politischen Tatbestand“ (ebd.). Sie präzisiert: „[I]n der Tat geht es um den Anspruch des Demokraten, sich Gesetzen und Richtersprüchen nicht blind unterwerfen zu müssen, sondern sich mit diesen kritisch auseinandersetzen zu können.“ (ebd.). Simon (2004) formuliert dies ähnlich und sieht das „Verlangen nach Allgemeinverständlichkeit als Verlangen nach Machtbegrenzung und Willkürschränken“ (S. 409).



## 4. Blickbewegungsmessung

Wo immer sich wissenschaftliche Hypothesen über das Blickbewegungsverhalten von Menschen formulieren lassen, ist Eyetracking eine attraktive Methode. Im Vergleich zu Verhaltensexperimenten ohne die Messung von Blickbewegungen ist der Ressourcenaufwand allerdings höher: Eyetracker sind etwas teurer als einfache Computer, auf denen beispielsweise „nur“ ein Reaktionszeitexperiment durchgeführt wird. Es braucht auch auf der technischen Seite etwas mehr Zeit, um sich mit den entsprechenden Abläufen (beispielsweise dem Kalibrieren der Messgeräte) vertraut zu machen. Im Repertoire der Experimentalparadigmen, die in der Kognitionswissenschaft eingesetzt werden, rangieren Eyetracking-Methoden bezüglich ihres Ressourcenaufwands wohl im Mittelfeld.<sup>40</sup>

Ein Vorteil bei der Messung von Blickbewegungen ist die Vielzahl an gemessenen Variablen. Häufig sind allerdings lediglich der Ort und die Dauer einzelner Fixationen relevant. Aus diesen Angaben werden auch die in dieser Arbeit verwendeten Lesemaße berechnet. Für andere Studien sind die Richtung oder die Geschwindigkeiten von Blicksprüngen (Sakkaden) interessant. Auch die augenblickliche Größe der Pupille, die ebenfalls von vielen Blickbewegungsmessgeräten gemessen wird, steht manchmal im Zentrum des Interesses. Die Veränderung der Pupillengröße (Mydriasis/Erweiterung oder Miosis/Verengung) wird mit verschiedenen Konstrukten in Verbindung gebracht: so beispielsweise Verarbeitungskomplexität (Hyönä/Tommola/Alaja 1995), emotionale Reaktion (Janisse 1974; Siegle et al. 2003) oder Arbeitsgedächtnisbelastung, wofür Kahnemann/Beatty (1966) zeigen konnten, dass die Stärke der Pupillenreaktion proportional zur Aufgabenschwierigkeit ist. Die zeitlich kleinste Ebene der Analyse ist die des einzelnen *Samples*. Ein Sample ist ein vom Eyetracker erzeugter Datenpunkt, der alle Informationen zum augenblicklich gemessenen Ort des Blicks und Zustand des Auges enthält. Die Menge der verfügbaren Samples hängt von der zeitlichen Auflösung des Eyetrackers ab, wobei die maximale Auflösung momentan bei 2000 Hertz liegt – ein solcher Eyetracker würde also pro Millisekunde zwei Samples bereitstellen. Fixationen und Sakkaden sind Konstrukte, die aus diesen einzelnen Datenpunkten errechnet werden. Einzelne Fixationen werden dabei nach bestimmten Heuristiken identifiziert, die Phasen relativer Ruhe des Blicks erfassen. Sakkaden sind die Blicksprünge zwischen diesen Ruhephasen.

---

<sup>40</sup> Insbesondere Imaging-Methoden (beispielsweise fMRI oder MEG) benötigen Gerätschaften, die mit einem Vielfachen an Wartungs- und Bedienungsaufwand verbunden sind.

Ein weiterer Vorteil der Messung von Blickbewegungen ist die Unmittelbarkeit, mit der kognitive Prozesse ihren Einfluss auf das Blickverhalten finden. Eyetracking ist in diesem Sinne eine Online-Messung:<sup>41</sup> Die Messung findet mit oder unmittelbar nach dem Phänomen statt, das gemessen werden soll. Die Zeit für verfälschende Einflüsse – sowohl von außen als für nachgeschaltete kognitive Prozesse – ist somit sehr kurz. Diese Verbindung von Blickbewegungen und kognitiven Prozessen im Allgemeinen war und ist nicht unumstritten. Die sogenannte *eye-mind hypothesis*, also die Annahme, dass Blickbewegungen tatsächlich sehr eng mit kognitiven Prozessen zusammenhängen, wurde prominent von Just/Carpenter (1976) vertreten. Sie zeigten für drei kognitive Domänen, mentale Rotation (vgl. Shepard/Metzler 1971), Satzverifikation (vgl. Carpenter/Just 1975) und den Vergleich von Größen und Mengen (vgl. Buckley/Gillman 1974), dass die menschlichen Blickbewegungen in der jeweiligen Aufgabe eng mit Vorhersagen eines entsprechenden kognitiven Modells zusammenhängen. Für die Leseforschung zeigten sie einige Jahre später (vgl. Just/Carpenter 1980), dass sich dieses Grundprinzip auf das Lesen von Texten übertragen lässt. Stellen, an denen laut Modell ein höherer Verarbeitungsaufwand (*processing load*) angenommen wird, werden länger gelesen. Ihr Modell (vgl. ebd., S. 331) nimmt noch eine relativ enge zeitliche Kopplung zwischen kognitiver Verarbeitung und Blickbewegungen an – diese Annahme wurde in den folgenden Jahrzehnten immer weiter hinterfragt. So präsentierten beispielsweise Anderson/Bothell/Douglass (2004) Hinweise im *fan effect*-Paradigma, die gegen eine sehr enge Verknüpfung von Blickbewegungen und Abrufvorgängen sprechen. Doch das ändert nichts an der Tatsache, dass die Arbeiten von Just und Carpenter im Sinne der Messbarkeit von kognitiven Abläufen über Blickbewegungen einen Meilenstein darstellten.

Diese Ausführungen weisen schon darauf hin, dass über Blickbewegungen auch die Auswirkungen von Zwischenschritten gemessen werden können – eben nicht nur das Produkt eines Verarbeitungsvorgangs. Nicht immer ist der Gegensatz zwischen Online- und Offline-Maßen relevant. Für jene zeitliche Bandbreite, in der sich die Effekte bewegen, die für uns in der vorliegenden Arbeit von Interesse sind, ist eine Online-Messung aber eindeutig von Vorteil.

In aller Kürze soll auf die technische Seite nicht-invasiver Blickbewegungsmessungen eingegangen werden. Eine ältere, aber noch immer eingesetzte

---

<sup>41</sup> Als Gegenbeispiel, also klassische Offline-Messung, gilt beispielsweise eine Fragebogenstudie. Hier wird mit deutlichem zeitlichen Versatz gemessen (Antwort auf dem Fragebogen) was eigentlich von Interesse ist (Verarbeitung des Fragebogens).

Technik zur Messung der Blickbewegungen beruht auf den sogenannten Purkinje-Bildern (auch Purkinje-Reflexionen oder Purkinje-Sanson-Bilder).<sup>42</sup> Dies sind Reflexionen an verschiedenen Stellen der Cornea und der Linse. Das Verhältnis zwischen erstem und viertem Purkinje-Bild (P1: äußere Reflexion an der Cornea, P4: Reflexion an der Innenseite der Linse) wird dann zur Berechnung der Blickposition verwendet. Bei einer neueren und meist einfacher zu bedienenden Messmethode wird das Auge von Infrarotemittern angestrahlt. Das von der Pupille reflektierte Licht wird von einer oder mehreren Hochgeschwindigkeits-Infrarotkamera(s) aufgefangen. Zusätzlich wird teilweise die Reflexion auf der Cornea gemessen. Die zweitgenannte Messmethode, die auch als *bright eye*-Methode bezeichnet wird, wurde in der vorliegenden Arbeit eingesetzt.

Die Messung von Blickbewegungen mittels Eyetracker ist in dieser Arbeit die zentrale empirische Methode. Das Blickverhalten der VersuchsteilnehmerInnen wird über mehrere abhängige Variablen operationalisiert, die im folgenden Abschnitt vorgestellt werden. Danach werden die Maße vorgestellt und diskutiert, die aus den erhobenen Blickbewegungssamples errechnet werden. Anhand von ihnen sollen Prozesse beobachtbar gemacht werden, die beim Lesen von Textmaterial wirken. Der genaue Ablauf sowie das technische Equipment, das für die Datenerhebung der vorliegenden Arbeit zum Einsatz kam, wird erst im nächsten Kapitel, insbesondere in Abschnitt 5.1.2, ausführlich dargestellt.

## 4.1 Blickbewegungsmaße beim Lesen

Zunächst müssen wir eine Frage aufgreifen, die absolut kritisch ist: Wie eng ist die Verbindung zwischen kognitiven Prozessen auf höheren Ebenen und Blickbewegungen? Sollte es der Fall sein, dass Blickbewegungen tatsächlich nur von oberflächlichen Eigenschaften des Perzepts (in unserem Fall dem geschriebenen Text) abhängen, würde das ein schwerwiegendes Problem darstellen, wenn wir diese Methodik einsetzen, um Aussagen zur Verständlichkeit abzuleiten. Starr/Rayner (2001) stellen diesen Konflikt als Widerstreit zweier Klassen von Modellen dar, die Blickbewegungssteuerung erklären wollen. In die Klasse der „oculomotor models“ (ebd., S. 157) fallen Modelle, die eine sehr lose Verbindung zwischen Blickbewegungen und höheren kognitiven Prozessen annehmen. Ausschlaggebend sind in solchen Erklärungsansätzen lediglich visuell-motorische Faktoren, die auf einer niedrigen (nicht-linguistischen) Ebene angesiedelt sind. Dazu gehören Eigenschaften des

---

<sup>42</sup> Benannt sind diese Bilder nach dem tschechischen Anatom Jan Evangelista Purkyně (1787-1869) und dem französischen Arzt Louis Joseph Sanson (1790-1841).



Perzepts, wie die Wortlänge oder der Abstand zwischen Wörtern. Auch Eigenschaften des Perzeptionsvorgangs selbst wird ein Einfluss eingeräumt. Dazu gehört die Landeposition im Wort (die optimale Landeposition ist dabei zwischen Wortanfang und -mitte, vgl. bspw. Kliegl/Nuthmann/Engbert 2006; Vitu/O'Regan/Mittau 1990) sowie die Amplitude der eingehenden Sakkade (vgl. z.B. Kliegl et al. 2006). Ein interessantes Argument für diese Modellklasse legten Vitu et al. (1995) vor. Sie präsentierten VersuchsteilnehmerInnen neben alltäglichen Texten auch Ketten von sich wiederholenden Buchstaben. Diese Buchstabenkette wurden erzeugt, indem in den Originaltexten jeder Buchstabe durch beispielsweise ein *x* ersetzt wurde. Damit waren zwar Wortlängen und -grenzen erhalten, sämtliche linguistische Information jedoch getilgt. Es zeigte sich, dass globale (Sakkadenlänge und Fixationsdauer) und lokale (Übersprungsrate, Landeposition sowie Refixierungswahrscheinlichkeit und -position) Blickbewegungsmuster sehr ähnlich zu den Originaltexten waren. Vitu et al. (1995, S. 352) argumentieren:

The finding that the eyes are capable of generating an autonomous oculomotor scanning strategy in the absence of any linguistic information to process argues in favor of the idea that such predetermined oculomotor strategies might be an important determinant of eye movements in reading.

Dieser Argumentation widersprachen Rayner/Fischer (1996) mit einem direkten Antwortartikel. Sie fanden für nicht-transformierten Text kürzere Fixationen, längere Sakkaden und weniger Übersprünge im Vergleich zu transformiertem Text. Außerdem hatte die Wortfrequenz für nicht-transformierten Text einen Einfluss auf die Fixationsdauern. Sie folgern: „Contrary to Vitu et al.'s (1995) findings, our results show that eye movements are not guided by a global strategy and local tactics, but by immediate processing demands.“ (ebd., S. 734).

Dieses Zitat ist gleichzeitig eine der Grundaussagen der konkurrierenden Modellklasse der „processing models“ (Starr/Rayner 2001, S. 157). Diese umschließt Modelle, die auch Prozessen auf höheren Ebenen Einfluss auf die Blickbewegungssteuerung einräumen. Wohin und wann der Blick bewegt wird, wird dabei unter anderem anhand von Eigenschaften wie Wortfrequenz oder der Vorhersagbarkeit eines Worts vorhergesagt.<sup>43</sup> Das „E-Z Reader“-Modell (vgl. Pollatsek/Reichle/Rayner 2006; Rayner/Reichle/Pollatsek 2000; Reichle et al. 1998; Reichle/Rayner/Pollatsek 1999; Reichle/Rayner 2000; Reich-

<sup>43</sup> Im Kontext der vorliegenden Arbeit ist auch dieser Einfluss noch eher auf einer niedrigen, hier lexikalischen, Ebene anzusiedeln. Im Kontext der widerstreitenden *oculomotor vs. processing models* ist die Aufnahme linguistischer Variablen aber hochrelevant und trennt die beiden Modellklassen.

le/Rayner/Pollatsek 2003; Reichle/Warren/McConnell 2009) als Instantiierung der *processing*-Modelle sagt die Blickbewegungen von Menschen beim Lesen von Textmaterial unter anderem auf Grundlage dieser linguistischen Variablen vorher. Starr/Rayner (2001, S. 158) fassen zusammen:

Given that frequency and predictability influence fixation time, researchers favoring a processing account believe that fixation times are primarily influenced by lexical factors and moment-to-moment comprehension processes. In general, processing theorists don't exclude the influence of low-level oculomotor factors, but they think that their influence on fixation time is small relative to the influence of cognitive factors. Thus, although they believe that the decision of when to move the eyes (how long to fixate) is strongly influenced by cognitive processing, the decision of where to move the eyes is primarily a function of oculomotor and visual factors (with perhaps some influence from linguistic factors).

Ohne Frage muss den Modellen der ersten (*oculomotor*) Klasse zugestanden werden, dass sie einflussreiche Faktoren der Blickbewegungssteuerung betonen. Das Problem liegt wohl eher in der Ausschließlichkeitsaussage, die teilweise damit einhergeht. Dass diese *low-level*-Einflüsse nicht die einzigen sind, die bei der Aufklärung von Fixationsdaten relevant sind, zeigt sich wissenschaftsgeschichtlich am Erfolg der verschiedenen Modelle der Blickbewegungssteuerung. Im Gesamtblick erschließt sich, dass Modelle, die auch eindeutig linguistische Faktoren einbeziehen, die erfolgreicher sind, wenn es um die Vorhersage von Lesedauer und/oder Übersprungswahrscheinlichkeiten geht. Die nicht-linguistischen Einflüsse auf niedrigeren Ebenen der Blickbewegungssteuerung sind ohne Frage notwendig für die Vorhersage von Fixationsdauern, sie sind aber offenbar nicht hinreichend. Inzwischen sind Maße wie die Wortfrequenz des augenblicklich verarbeiteten Wortes  $n$ , aber auch die Frequenzen der Wörter  $n-1$  und  $n-2$  (vgl. z.B. Kliegl et al. 2006), also der vorhergehenden Wörter, relevant für die Vorhersage von Fixationsdauern. Auch andere Maße, wie die Vorhersagbarkeit aus dem Kontext (vgl. Mielle/Sparrow/Sereno 2007) oder Wortfamiliarität (vgl. White 2008; Wolfer/Hansen/Konieczny 2012b), sind inzwischen Bestandteil solcher Vorhersagen. Diese eher auf der lexikalischen Ebene angesiedelten Maße werden in dieser Arbeit ebenfalls in die Analysen eingehen. Dabei werden Wortfrequenz und Wortfamiliarität einerseits oft als sogenannte Baseline-Prädiktoren<sup>44</sup> behandelt, andererseits in bestimmten Analysen zum zentralen Gegen-

---

<sup>44</sup> Baseline-Prädiktoren sind jene Prädiktoren, die zwar einen Einfluss auf die Verarbeitung haben, jedoch in der augenblicklich relevanten Analyse nicht von Interesse sind. Das typische Vorgehen ist häufig die Berechnung eines Baseline-Modells, das diese Prädiktoren enthält. Dieses Modell wird sukzessive um die in der Analyse relevanten Prädiktoren erweitert.

stand gemacht. In der vorliegenden Arbeit sind beide Sichtweisen durch entsprechende Analysen vertreten.

Wir wollen uns nun der Berechnung von Lesezeitmaßen widmen. Für die Berechnung werden nahezu ausschließlich diese Maße als abhängige Variablen verwendet – weshalb auch entschieden wurde, die Einflüsse auf der okulomotorischen Ebene (Amplitude der eingehenden Sakkade und relative Landeposition) zwar anzuerkennen, aber in den Analysen nicht zu beachten. Die Maße, die im FLRC zur Verfügung stehen, sind in Tabelle 4.1 aufgeführt. Es kann an dieser Stelle nicht auf jedes einzelne Maß eingegangen werden. Dies geschieht an anderem Ort ohnehin ausführlicher und besser (vgl. z.B. Rayner/Pollatsek 2006). Die Operationalisierung der einzelnen Maße ist ebenfalls Tabelle 4.1 zu entnehmen. Hier soll eine kurze Diskussion angeschlossen werden, welche Prozesse grob mit einzelnen Maßen assoziiert werden können. Es wird gezeigt, dass in der Literatur weitestgehend Einigkeit darüber herrscht, dass eine Eins-zu-eins-Zuordnung zwischen Maß und kognitivem Prozess nicht möglich ist. Sinnvoller erscheint eine Einteilung in frühe und späte Maße.

Es wäre zu einfach, jedem Maß einen ganz bestimmten kognitiven Prozess zuzuweisen. Aussagen wie *'first fixation durations'*<sup>45</sup> spiegeln den Aufwand beim lexikalischen Abruf wider' wären zu vereinfachend. Natürlich spielt der lexikalische Abruf eine Rolle für die *first fixation durations*, es ist aber nicht auszuschließen, dass andere Prozesse ebenfalls beteiligt sind. Die verschiedenen Maße können, Hyönä/Lorch/Rinck (2003) folgend, eingeteilt werden, indem sie eher mit unmittelbaren, frühen oder eher mit verzögerten, späten Effekten assoziiert werden. Zu ihrer Einteilung ist jedoch zu bemerken, dass die Zuteilung der *total reading time* zu verzögerten Effekten nicht ganz nachvollziehbar ist, da sich diese aus *first-pass reading times* (laut deren Einteilung unmittelbar) und *re-reading time* (verzögert) zusammensetzt. Liversedge/Paterson/Pickering (1998, S. 58) drücken sich, was frühe Effekte angeht, relativ vorsichtig aus. So schreiben sie im Zusammenhang mit experimentellen Studien:

If effects are found on measures of either first fixation duration or gaze duration in one condition relative to another, the experimenter will usually conclude that difficulty was experienced immediately on processing that region of text.

Kombiniert man nun diese Aussage mit *total reading times*, lässt sich eine schlüssigere Interpretation der Gesamtlesezeit auf einer Region ableiten (ebd.):

---

<sup>45</sup> Da der überwiegende Teil der psycholinguistischen Literatur in englischer Sprache verfasst ist, werde ich die englischen Namen für die verschiedenen Lesemaße verwenden.

If an effect is observed for total reading time on a region, *but not for earlier measures* such as first fixation duration or first pass reading time, then this is generally taken as an indication of the manipulation having a relatively late effect on processing [meine Hervorhebung, S.W.].

Bezeichnung	Operationalisierung	Typ
Single fixation duration (SFD)	Dauer der ersten Fixation auf der Region, wenn nur eine Fixation vorhanden	numerisch
First fixation duration (FFD)	Dauer der ersten Fixation auf der Region	numerisch
First-pass reading time (FPRT)	Dauer aller Fixationen auf der Region vom Eintreten in die Region bis zum ersten Verlassen	numerisch
Regression path duration (RPD)	Dauer aller Fixationen vom Eintreten in die Region bis zum ersten Verlassen nach rechts, inklusive aller Fixationen auf Regionen rechts von der aktuellen Region	numerisch
Regression contingent regression path duration (RCRPD)	Regression path duration nur im Falle einer regressiven Sakkade	numerisch
Re-reading time (RRT)	Total reading time minus first-pass reading time	numerisch
Total reading time (TRT)	Dauer aller Fixationen auf der Region	numerisch
Regressive Sakkade gestartet?	Wurde die Region beim ersten Lesen nach rechts verlassen?	T/F, p
Übersprungen? (skipped)	Wurde die Region beim ersten Lesen übersprungen?	T/F, p

Tab. 4.1: Lesemaße im *Freiburg Legalese Reading Corpus* (T = TRUE/wahr, F = FALSE/unwahr, p = Wahrscheinlichkeit). Binäre Maße werden in statistischen Modellen in Wahrscheinlichkeitsform überführt.

Diese Einteilung in frühe und späte Effekte ist nicht vollständig deckungsgleich mit einer anderen Einteilung, die sich eher auf die Eigenschaften der in das Maß eingehenden Fixationen bezieht. Dabei werden Maße, die sich nur aus Fixationen beim ersten Lesen einer Region zusammensetzen (*first-pass*-Maße) unterschieden von Maßen, in die auch Fixationen beim zweiten, dritten, *n*-ten Lesen einer Region mit eingehen. Dieser Einteilung folgend gehören *regression path durations* zu der ersten Kategorie, da bei deren Zählung noch keine Fixation rechts von der aktuellen Region aufgetreten ist. Während also in der erstgenannten Einteilung eher die assoziierten kognitiven Verarbeitungsprozesse betont werden (früh vs. spät), wird bei der Einteilung in *first*- und *second-pass*-Maße<sup>46</sup> streng nach der Eigenschaft der Zählung vorgegangen. Diese beiden Arten der Einteilung sind nicht deckungsgleich. Ein gutes Beispiel sind die eben schon erwähnten *regression path durations*, die laut der Einteilung von Hyönä et al. (2003) mit verzögerten Effekten assoziiert sind, jedoch formal zu den *first-pass*-Maßen gehören. Die Sinnhaftigkeit einiger Maße ergibt sich manchmal auch erst in Zusammenhang mit der zugrunde gelegten Region, für die sie berechnet werden. So weisen Rayner/Pollatsek (2006, S. 620) darauf hin, dass für längere (mehr als ein Wort umfassende) Regionen zwar manchmal *first fixation durations* berichtet werden, dies aber nur dann, wenn man sich für *spill-over*-Effekte<sup>47</sup> aus der vorherigen Region interessiert. Ansonsten sehen sie es eher als „inappropriate“ (ebd.) an, in solchen Fällen die Dauer der ersten Fixation heranzuziehen. Clifton et al. (2007) äußern sich zu den Termini der frühen und späten Maße insofern, dass sie einer direkten Abbildung auf erste und weitere Phasen der Satzverarbeitung, wie sie beispielsweise von Frazier/Fodor (1978) vorgeschlagen wurden, skeptisch gegenüberstehen. Sie schränken aber ein (Clifton et al. 2007, S. 349):

Nonetheless, careful examination of when effects appear may be able to shed some light on the underlying processes. Effects that appear only in the ‘late’ measures are in fact unlikely to directly reflect first-stage processes; effects that appear in the ‘early’ measures may reflect processes that occur in the initial stages of sentence processing, at least if the measures have enough temporal resolving power to discriminate among distinct, fast-acting, processes.

<sup>46</sup> Der Ausdruck *second-pass*-Maß wird meist im Sinne von *second-or-more* verwendet. Die in Tabelle 4.1 verzeichnete *re-reading time* wird daher manchmal auch als *second-pass reading time* bezeichnet (vgl. u.A. Calvo/Meseguer/Carreiras 2001; Clifton/Staub/Rayner 2007; Frenck-Mestre 2005; Rayner/Pollatsek 2006).

<sup>47</sup> *Spill-over*-Effekte sind Effekte, die sich erst in einer späteren Region zeigen als jener, die sie hervorgerufen hat.

In der vorliegenden Arbeit wird ebenfalls diese eher vorsichtige Herangehensweise vertreten. Eine Einteilung von Lesemaßen muss nicht völlig ohne die Einteilung in frühe und späte Prozesse auskommen. Eine starre Abbildung einzelner Maße auf Phasen der Sprachverarbeitung geht allerdings zu weit. Solche Aussagen können, wenn überhaupt, nur in der Zusammenschau mehrerer Maße und in Kombination mit der entsprechenden Einteilung der Regionen erreicht werden.

Nun, da die qualitative Kopplung von Lesezeitmaßen und kognitiven Prozessen etwas präzisiert wurde, möchte ich noch kurz auf die zeitliche Kopplung eingehen. Bei der Messung des menschlichen Leseprozesses werden Fixationen immer einem bestimmten Wort zugeordnet. Die Frage der Unmittelbarkeit betrifft nun die Zuordnung von Fixationsort und dem, was augenblicklich verarbeitet wird. Einfacher formuliert: Wird, während Wort  $n$  fixiert wird, auch nur Wort  $n$  verarbeitet? Gehen wir davon aus, dass Fixation  $n$  auf Wort  $n$  landet. Die vorherige Fixation  $n-1$  befand sich auf Wort  $n-2$ , das heißt Wort  $n-1$  wurde übersprungen. Dies ist beileibe keine Seltenheit. Im *Freiburg Legalese Reading Corpus* wurde fast die Hälfte aller Wörter übersprungen (43,3%). Doch warum wurde Wort  $n-1$  übersprungen? Mehrere Gründe sind denkbar: 1) Es trat ein okulomotorischer Fehler auf. Wort  $n-1$  sollte eigentlich fixiert werden, die Fixation landete aber 'aus Versehen' auf Wort  $n$ . 2) Vielleicht war Wort  $n-1$  aber auch hoch vorhersagbar. In diesem Fall musste die/der LeserIn das Wort gar nicht mehr fixieren, weil sie/er ohnehin schon wusste, welches Wort an dieser Stelle stehen würde. 3) Eine weitere Möglichkeit ist, dass Wort  $n-1$  gelesen wurde, während die Augen noch auf Wort  $n-2$  ruhten. Wort  $n-1$  war somit im Vorschaubereich, der perzeptuellen Spanne, während Wort  $n-2$  fixiert wurde. Diese Möglichkeit des *parafoveal preview* ist äußerst relevant, wenn man Aussagen über die zeitliche Unmittelbarkeit von Fixation und Wortverarbeitung treffen will. In Abschnitt 7.1 werden Effekte auf der lexikalischen Ebene im FLRC herausgearbeitet. Dort werden Effekte der parafovealen Vorschau in fast allen Modellen eine Rolle spielen. Kliegl et al. (2006) nennen Effekte, die der parafovealen Vorschau entspringen, *lag* und *successor effects*. Dabei gehen sie in der Effektbeschreibung von Fixation  $n$  aus, wie das in der sonstigen Literatur ebenfalls üblich ist. Als *lag effects* werden jene Effekte beschrieben, die von Eigenschaften bereits gelesener Wörter (meist Wort  $n-1$ ) ausgehen. Wort  $n-1$  kann noch nicht ganz verarbeitet sein. Es kann aber auch die perzeptuelle Spanne so modulieren, dass Wort  $n$  weniger gut vorverarbeitet werden konnte. Außerdem kann Wort  $n-1$  auch schlicht zu lang gewesen sein, als dass Wort  $n$  im Vorschaubereich hätte sein können. All diese Effekte, ebenso wie die *successor effects* – Einflüsse von Wörtern rechts von Wort  $n$  – werden in Abschnitt 7.1 nochmals mit einer Vorstel-

lung relevanter Literatur aufgegriffen. Im nun folgenden Abschnitt wird ein Eindruck davon vermittelt, welchen Zweck psycholinguistische Forschung mit Lesekorpora verfolgt, und wie versucht wird, mit potenziellen Störvariablen innerhalb der Korpora umzugehen.

## 4.2 Lesekorpora

Die klassische Herangehensweise der Psycholinguistik orientiert sich stark am Kreislauf empirischer Forschung. Hier werden anhand bestimmter theoretischer Fragestellungen Vorhersagen generiert, wie Menschen auf einen bestimmten linguistischen Stimulus reagieren. Anhand der aus den Theorien generierten Hypothesen wird ein Forschungsdesign erstellt, wobei ein oder mehrere Faktoren so manipuliert werden, dass eine Aussage über die Hypothesen möglich ist. Das linguistische Material wird dabei auf möglichst wenigen Dimensionen variiert, um Störeinflüsse möglichst gering zu halten. Dem Forschungsdesign inhärent sind bereits explizite statistische Hypothesen über das Verhalten der Versuchsteilnehmenden (beispielsweise „In Faktorausprägung (= Bedingung) 2 werden im Vergleich zu Bedingung 1 erhöhte *first-pass reading times* auf Wort *x* erwartet.“)<sup>48</sup> So ist die Zahl wissenschaftlicher Fragestellungen, die anhand eines Datensatzes bearbeitet werden können, naturgemäß auf den Raum eingeschränkt, den die zu Grunde liegende Hypothese aufspannt – zugunsten maximaler experimenteller Kontrolle.

Da im klassischen psycholinguistisch-experimentellen Ansatz alles außer den explizit variierten Variablen meist als potenzielle Störvariable gilt, stehen die Experimentalsätze nahezu immer isoliert. Auf begleitenden Kontext, der die Stimuli in einen natürlicheren Rahmen einbetten könnte, wird meist verzichtet, weil Kontext erzeugende Rahmensätze potenzielle Störeinflüsse darstellen (vgl. jedoch auch Fedorenko/Piantadosi/Gibson 2012; Gibson/Wu 2011, die zeigen, dass in bestimmten Fällen stärkere Effekte mit Kontext zu erwarten sind). Am ehesten lässt sich dieser Widerstreit zwischen experimenteller Kontrolle und Natürlichkeit der Stimuli im Trade-off zwischen interner und externer (oder ökologischer) Validität fassen: Je natürlicher die Stimuli, desto schwieriger ist der Einfluss potenzieller Störvariablen abzuschätzen. Je artifizierter die Experimentalmaterialien jedoch sind, desto eher läuft man Gefahr, artifizielles bzw. für das „wahre Leben“ untypisches Verhalten bei den Versuchsteilnehmenden zu evozieren.

<sup>48</sup> Selbstverständlich sind auch ausgefeiltere Hypothesen möglich. So lassen sich beispielsweise Interaktionshypothesen formulieren, die das Zusammenwirken von zwei oder mehr Faktoren annehmen.



#### 4.2.1 Der Zweck von Lesekorpora

Blickbewegungskorpora und im speziellen der *Freiburg Legalese Reading Corpus* sind darauf ausgelegt, diesen Trade-off zumindest ansatzweise auszuhebeln. Die Grundidee ist, große Mengen an schwach oder gar nicht experimentell kontrollierten sprachlichen Stimuli von vielen Menschen lesen zu lassen. Beim Zusammenstellen der Texte, die in das Lesekorpus eingehen, kommt ein kontrollierendes Element ins Spiel: Werden beispielsweise lediglich längere Zeitungsartikel oder Editorials in das Korpus aufgenommen (wie dies beim Dundee Corpus der Fall ist, Kennedy 2003; Kennedy/Hill/Pynte 2003), schränkt das die Menge an potenziellen Einflussvariablen (nämlich jene, die von der Textgattung abhängen) ein. Die grundsätzliche Idee von Blickbewegungskorpora ist – ebenso wie bei den zuvor beschriebenen klassisch-experimentellen Studien – ebenfalls die Überprüfung von Hypothesen. Der Unterschied besteht jedoch in der potenziellen Menge verschiedener Fragestellungen, die anhand eines Lesekorpus bearbeitet werden können: Während in klassisch-experimentellen Studien meist nur jene Hypothesen überprüft werden können, auf denen das Versuchsdesign beruht, können anhand von Blickbewegungskorpora verschiedenste Sachfragen auf den unterschiedlichsten Ebenen behandelt werden. So können Forschende anhand eines (zugegebenermaßen sehr großen) Datensatzes beispielsweise Fragen auf der physiologischen Ebene der Blickbewegungssteuerung, der Ebene lexikalischer Verarbeitung, der syntaktischen Ebene und/oder der Ebene des Textes untersuchen.

Das heißt im Umkehrschluss allerdings nicht, dass anhand eines Blickbewegungskorpus alle experimentellen Fragen beantwortet werden können. So ist es beispielsweise illusorisch, anzunehmen, dass alle Phänomene, die psycholinguistisch potenziell interessant sind, in einem Korpus vorkommen. Hier ist beispielsweise auf das Phänomen lokaler syntaktischer Kohärenzen (vgl. Tabor et al. 2004) zu verweisen: Es kann gezeigt werden, dass Menschen Sätze anders verarbeiten, wenn in ihnen Wortabfolgen vorkommen, die selbst kanonische Hauptsätze sind, von der Interpretation des Gesamtsatzes aber abweichen. Diese Effekte wurden im Visual-World-Paradigma (Tanenhaus et al. 1995) beispielsweise für lokale syntaktische Kohärenzen in Satzkomplementen (Müller-Feldmeth 2013; Konieczny et al. 2009) gezeigt. Auch die Referenzauflösung von Pronomina wird offenbar durch lokale syntaktische Kohärenzen moduliert (Weldle et al. 2010, siehe hierzu auch 2.2). Es ist jedoch überaus schwierig, in Blickbewegungskorpora, die sich aus natürlichen Texten zusammensetzen, jene sprachlichen Zielstrukturen zu finden, die zur Untersuchung solcher Fragen notwendig sind: Lokale syntaktische Kohärenzstrukturen, die einen kompletten Hauptsatz umfassen, sind so selten, dass

hier lediglich ein experimentell kontrolliertes Vorgehen sinnvoll erscheint. Es werden also gezielt solche Strukturen entworfen, die lokale syntaktische Kohärenzen enthalten. Diese Zielstrukturen füllen in einem experimentellen Design eine Zelle, die in der späteren Auswertung mit einer oder mehreren Kontrollbedingung(en) verglichen wird. Lokale syntaktische Kohärenzen sind hier nur als Stellvertreter für eine Vielzahl von psycholinguistischen Phänomenen, wie beispielsweise Holzwegsätze oder globale syntaktische Ambiguitäten (für einen Überblick siehe z.B. Hemforth/Konieczny 2008), zu verstehen, die kaum sinnvoll anhand von Lesekorpora untersucht werden können, weil die entsprechenden Strukturen in natürlicher Sprache schlicht nicht oder sehr selten vorkommen. Für viele Phänomene jedoch sind Blickbewegungskorpora eine attraktive Möglichkeit, die möglichst natürliche Verarbeitung von geschriebener Sprache (sozusagen „Lesen in freier Wildbahn“) zu untersuchen. Es wird im folgenden Abschnitt 5.1 für das *Freiburg Legalese Reading Corpus* und in Abschnitt 5.3 für andere Blickbewegungskorpora unter anderem dargestellt, wie man beim Zusammenstellen der Materialien für Blickbewegungskor gezielt darauf hinwirken kann, dass ganz bestimmte Phänomene untersuchbar werden.

Auch auf der Ebene der Datenauswertung ergeben sich grundsätzliche Unterschiede zwischen dem klassisch experimentellen Vorgehen und dem Einsatz von Lesekorpora. Experimentelle Stimulusätze sind – wie eben dargelegt – fast immer dazu entworfen, ein bestimmtes Phänomen zu untersuchen. Die experimentelle Manipulation umfasst daher nur kleine, sorgfältig ausgewählte Bereiche der Sätze. Die Sätze werden meist in Regionen unterteilt, die eine bestimmte Anzahl Wörter umfassen (zu diesem Vorgehen vgl. auch Liversedge et al. 1998, S. 57f.). Die gezielte Manipulation fällt in die kritische Region, und die Analyse von Lesezeiten beschränkt sich oft nur auf diese Region und meist noch die Region danach (*spill-over*-Region), um verzögerte Effekte ebenfalls abzugreifen. In Lesekorpora gibt es keine im Vorhinein definierten kritischen Regionen, weil zum Zeitpunkt der Datenerhebung noch nicht klar ist, welche Hypothesen später alle untersucht werden. Typischerweise bilden daher zunächst Wörter die Regionen in Lesekorpora. Einer späteren Einteilung in kritische Regionen steht dabei allerdings nichts im Wege. Im Falle des FLRC stehen auch Regionen zur Verfügung, die anhand der Phrasenstrukturknoten der syntaktischen Annotation definiert sind. Auf diese wird im Detail in den Abschnitten 7.3.2 und 7.3.3 eingegangen. Dort werden diese Regionendefinitionen zum ersten Mal eingesetzt. Für Lesekorpora ist diese Art der Regioneneinteilung vor der vorliegenden Arbeit meines Wissens noch nicht praktiziert worden.

#### 4.2.2 Kontrolle von Störvariablen

Möchte man Leseverhalten anhand von Lesekorpora untersuchen, bieten sich zwei Ansätze an, die Kliegl et al. (2004) als *experimental control approach* und *statistical control approach* vorstellen. Diese Ansätze können als alternative Herangehensweisen verstanden werden, wie 1) mit potenziellen Störfaktoren umgegangen wird und 2) bekanntermaßen hohen Korrelationen zwischen Prädiktorvariablen begegnet werden kann. In regressionanalytischen Termini gesprochen, liegt – wenn zwei oder mehrere Prädiktoren stark untereinander korreliert sind – Multikollinearität vor. Das heißt, große Anteile der Varianz in der vorherzusagenden Variablen werden sowohl von Prädiktorvariable  $x_1$  als auch von Prädiktorvariable(n)  $x_2$  bis  $x_n$  aufgeklärt.<sup>49</sup> Für Regressionsmodelle stellt dies ein Problem dar, da die statistischen Modelle bei hoher Multikollinearität schnell instabil werden (für Illustrationen, Erläuterungen sowie mögliche Lösungsansätze vgl. Baayen 2008, S. 181ff.).

Folgt man dem *experimental control approach*, werden Stimuli aufgrund ihrer Zugehörigkeit zu bestimmten Gruppen von Variablenkombinationen in die Analyse eingeschlossen. Das in Abschnitt 5.3 vorgestellte *Potsdam Sentence Corpus* (PSC, Kliegl et al. 2006) ist von Beginn an so konstruiert, dass in den Sätzen bestimmte Zielwörter vorkommen, die aus einer Zelle eines  $2 \times 2 \times 3$ -Designs stammen. Diese Zielwörter gehören immer einer Wortklasse (Nomen vs. Verb), Frequenzklasse (hoch vs. niedrig) sowie Wortlängenkategorie (lang, mittel, kurz) an. In Abschnitt 5.3 wird der Aufbau des PSC nochmals genauer dargestellt. Hier sollte lediglich darauf hingewiesen werden, wie anhand einer orthogonalen Korpuskonstruktion mit den hochkorrelierten Prädiktoren Wortlänge und Wortfrequenz umgegangen werden kann. So sind im PSC gezielt Wörter platziert, die auch untypische Kombinationen (lange häufige oder kurze seltene Wörter) repräsentieren. Gehen nun diese Wörter mit in die statistische Auswertung ein, umgeht man das Kollinearitätsproblem, dass die überwältigende Mehrheit der Wörter entweder kurz und hochfrequent oder lang und niedrigfrequent ist.

Eine etwas andere Ausprägung des *experimental control approach* liegt vor, wenn zwar nicht kontrolliert wird, welche Wörter in das Korpus aufgenommen werden, aber Selektionsbeschränkungen dahingehend eingeführt werden, welche Datenpunkte in die Analyse eingehen und welche nicht. Kliegl et al. (2004) beschreiben diese von Kliegl/Olson/Davidson (1983) erstmals ein-

<sup>49</sup> Dies ist nicht zu verwechseln mit der Varianz, die vom *Zusammenwirken* zweier Variablen aufgeklärt wird. Dieses Zusammenwirken drückt sich im Effekt der Interaktion zwischen zwei oder mehreren Variablen aus ( $x_1 \times x_2 [\dots \times x_n]$ ).

gesetzte Technik als „post-hoc selection of words from the reading material for an orthogonal design“ (Kliegl et al. 2004, S. 263); in Abgrenzung zur orthogonalen Korpuskonstruktion kann dieses Vorgehen orthogonales Sampling (vgl. Kennedy/Pynte 2005) genannt werden. Hier wird im Grunde ebenso ein ausgeglichenes Design angestrebt. Der Unterschied liegt jedoch darin, dass relativ unkontrolliert Material in das Korpus aufgenommen und von vielen Teilnehmenden gelesen wird, später aber nur ein Ausschnitt dessen in Analysen verwendet wird (dieses Vorgehen findet sich z.B. bei Kliegl et al. 1983; Radach/McConkie 1998; Radach/Heller 2000; Kennedy/Pynte 2005). Je nach Fragestellung können sich dann die Teilmengen des eingeschlossenen Materials verändern.

Grundsätzlich ist das Vorgehen nach dem *experimental control approach* recht unkompliziert und bietet die Möglichkeit, hochkorrelierte Konzepte zu entwirren. Zwei Nachteile sollen trotzdem kurz genannt werden:

- 1) Orthogonale Korpuskonstruktion ist nur dann sinnvoll möglich, wenn man das Blickbewegungskorpus komplett *from scratch* konstruiert. Jedes andere Vorgehen scheint entweder hochgradig kompliziert oder schlicht nicht praktikabel. Würde man ein Lesekorpus aus tatsächlich vorkommenden Sätzen konstruieren wollen, müsste man zusammenhängende Sätze entsprechend eines orthogonalen Designs aus einem natürlichen Ausgangskorpus sampeln. Beide Beschränkungen (zusammenhängend Sätze **und** orthogonal verteilte Variablen) gleichermaßen zu erfüllen, scheint mir nur äußerst schwer möglich.
- 2) Beim orthogonalen Sampling werden den VersuchsteilnehmerInnen natürliche Texte präsentiert. In die Analyse geht aber nur ein Bruchteil der Stimuli ein, nämlich eben jene, die dem Design entsprechen, das dem Sampling zu Grunde liegt. Dadurch gehen recht viele Datenpunkte verloren. Diese Daten wurden zuvor aber teuer erhoben. Es gibt auch keine Garantie, dass interessante Phänomene in der Teilmenge der Daten, die der Analyse zugeführt wird, tatsächlich beobachtet werden können. Der schlechteste Fall tritt dann ein, wenn sich interessante Effekte in dem Teil der Daten „verstecken“, der vor der Analyse aufgrund des Designs ausgeschlossen wurde.

Lorch/Myers (1990) diskutieren, inwieweit korrelierte Prädiktoren anhand von Messwiederholungsdesigns in den Griff zu bekommen sind. Inzwischen hat sich die statistische Modellierung anhand linearer gemischter Modelle (vgl. Pinheiro/Bates 2000) weitestgehend durchgesetzt und wurde mehrfach erprobt (so z.B. in Pynte/New/Kennedy 2008a). Die Grundidee ist, systema-

tisch auftretende Varianz zwischen Versuchspersonen zu kontrollieren. Gelingt dies, ist das gravierendere Problem, dass die von korrelierten Prädiktoren gemeinsam aufgeklärte Varianz noch immer nicht für statistische Tests auf Signifikanz benutzt werden kann, „leading to much more conservative tests compared to experimental designs with orthogonal factors, especially in the case of highly correlated predictors such as word length and word frequency“ (Kliegl et al. 2004, S. 264).

In der vorliegenden Arbeit wird diesem Problem meist mit einer Multiresidualisierungstechnik begegnet, wobei schrittweise hochkorrelierte Prädiktoren gegeneinander residualisiert werden. Wortlänge, Wortfrequenz und Wortfamiliarität sind naturgemäß hochkorreliert. Diese Prädiktoren werden dem Multiresidualisierungsverfahren ebenso zugeführt wie die Satzposition und die Einbettungstiefe von Wörtern. An entsprechender Stelle wird darauf eingegangen, inwieweit dies als konservatives Vorgehen aufgefasst werden kann. Es werden außerdem noch Hinweise darauf gezeigt, dass solche „Ketten“ von Residualisierungen nur wenige Schritte weit sein können. Der Vorteil eines statistischen Kontrollansatzes in Kombination mit Residualisierungstechniken liegt darin, dass ein Blickbewegungskorpus aus Texten kompiliert werden kann, die tatsächlich in der Welt vorkommen. Außerdem können alle Wörter in die entsprechenden Analysen mit einbezogen werden und nicht nur jene, die in bestimmte Zellen eines Sampling-Designs fallen.

Eye-tracking-korpora sind Sammlungen von Blickbewegungen auf natürlichem Textmaterial. Klassische linguistische Korpora werden dabei um eine Ebene erweitert, die sich aus Perzeptionsdaten von mehreren LeserInnen zusammensetzt. Lesekorpora sind im Wechselspiel zwischen interner und externer Validität eher in Richtung des Pols der externen Validität einzuordnen – zumindest im Vergleich zum klassischen psycholinguistischen Experiment. Dafür ist die potenzielle Menge an Sachfragen, die mit der Datenlage bearbeitet werden kann, höher als bei einer einzelnen experimentellen Studie. Kriterien experimenteller Forschung wie Replizierbarkeit der Ergebnisse und kontrollierte Erhebungsbedingungen im Labor sind natürlich trotzdem einzuhalten. Ein Problem bei der Verwendung von natürlichen Texten, die so tatsächlich schon „in freier Wildbahn“ aufgetreten sind, ist Multikollinearität von linguistischen Variablen. Ein gutes Beispiel auf lexikalischer Ebene ist das Konglomerat aus Wortlänge, Wortfrequenz und Wortfamiliarität, die untereinander alle hochkorreliert sind. Solche Kollinearitäten können jedoch auf quasi-experimentelle oder statistische Weise umgangen, vermieden oder kontrolliert werden. Im Folgenden wird detailliert auf das *Freiburg Legalese Reading Corpus* eingegangen. Dabei werden die den einzelnen Teilen zu Grun-

de liegenden Konstruktionsprinzipien beschrieben. Ferner werden technische Details zur Erhebung der Blickbewegungsdaten vorgestellt (5.1). Außerdem werden die Teile des Korpus untereinander (5.2) und das komplette Korpus mit etablierten und im vergangenen Abschnitt teilweise schon erwähnten Blickbewegungskorpora verglichen (5.3).

## 5. Das Freiburg Legalese Reading Corpus (FLRC)

### 5.1 Kompilierung des Freiburg Legalese Reading Corpus

Die Erstellung eines Blickbewegungskorpus erfolgt in zwei Schritten. Zunächst werden die Texte, die später gelesen werden sollen, ausgewählt bzw. erstellt. Ergebnis ist eine Textsammlung, die am ehesten mit dem traditionell-linguistischen Korpusbegriff verbunden ist. Während bei der klassisch-linguistischen Korpuskompilierung meist der Grundsatz „je größer, desto besser“ gilt, kann das nicht eins-zu-eins auf Blickbewegungskorpora übertragen werden. Bei Lesekorpora muss zusätzlich beachtet werden, dass die im Korpus enthaltenen Texte noch von Menschen gelesen werden müssen. Der zweite Schritt beim Erstellen eines Blickbewegungskorpus besteht darin, die Stimuli so aufzubereiten, dass sie der Blickbewegungserhebung zugeführt werden können. Hierbei müssen auch Fragen und Instruktionen erstellt werden, um die Blickbewegungserhebung in die Versuchssituation einzubetten. Um über einzelne linguistische Konstruktionen hinweg verallgemeinerbare Analysen durchführen zu können, werden die verwendeten Texte zusätzlich meist mit linguistischen Informationen annotiert. Diese Schritte werden in den folgenden Abschnitten für das *Freiburg Legalese Reading Corpus* dargestellt.

#### 5.1.1 Textauswahl und Gliederung des Freiburg Legalese Reading Corpus

Das FLRC (siehe Abb. 5.1) gliedert sich in zwei Teile: Dem Reformulierungsteil (FLRCRef), der auf der linken Seite der Abbildung dargestellt ist, sowie einem Teil mit ausschließlich natürlichen Texten (FLRCNat, rechte Seite der Abbildung).

##### **FLRCRef: Reformulierungsteil**

Der Reformulierungsteil besteht aus mittel- und stark reformulierten Auszügen aus Originalurteilen des Bundesverfassungsgerichts. Alle Materialien, die an den Reformulierungen beteiligt sind (Originale sowie moderat und stark reformulierte Texte) stammen aus einer Studie von Hansen et al. (2006) und wurden uns dankenswerterweise für die weitere Verwendung zur Verfügung gestellt. Als Ausgangsmaterial dienten 30 Sätze aus Urteilen des Bundesverfassungsgerichts, die sich auf unterschiedliche Weise als komplex bezeichnen lassen. Zehn Sätze wiesen viele Nominalisierungen auf (in Abb. 5.1 als *Nom.* gekennzeichnet), weitere zehn Sätze zeichneten sich insbesondere über komplexe Nominalphrasen aus (*NP*). Die letzte Komplexitätsgruppe



bilden Sätze, die eine komplexe syntaktische Struktur aufweisen (*Synt.*). Die zwei Reformulierungsstufen zu jedem Satz konzentrierten sich insbesondere auf die Auflösung der jeweiligen Komplexitätsart. Dabei wurden die Originalsätze in manchen Fällen auch geteilt. Welche **linguistisch** messbaren Folgen die Reformulierungen hatten, untersuche ich in Abschnitt 5.2.1. Dabei wird zu zeigen sein, ob die Reformulierungsgrundsätze tatsächlich eingehalten wurden und inwiefern sich die Reformulierungen auch auf andere linguistische Merkmale wie Satzlänge oder die Anteile bestimmter Parts-of-Speech auswirkten.

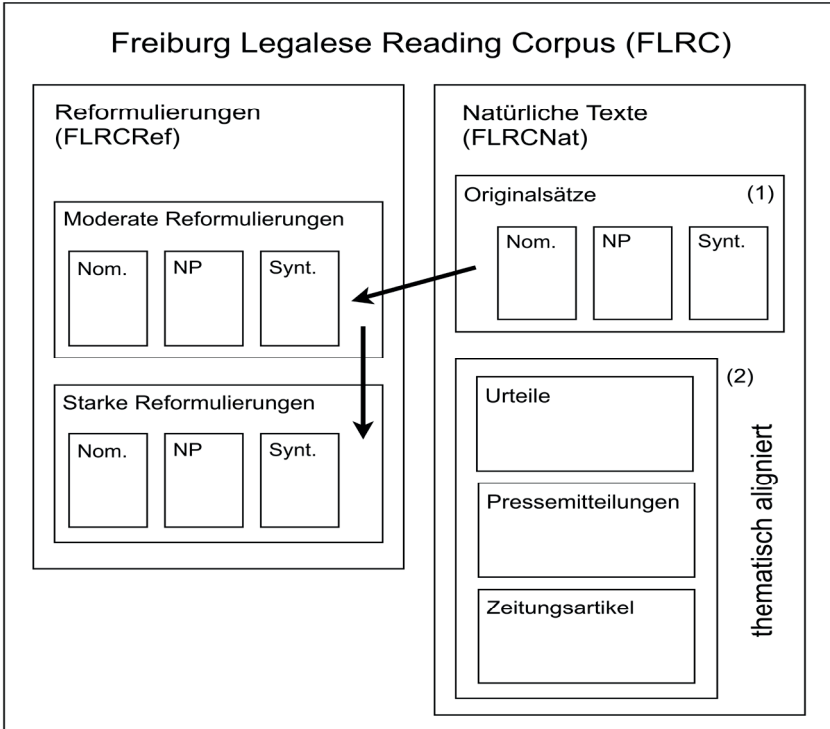


Abb. 5.1: Aufbau des Freiburg Legalese Reading Corpus. Auf der linken Seite der Abbildung ist der Reformulierungsteil (FLRCRef) abgebildet. Auf der rechten Seite (FLRCNat) sind sowohl die den Reformulierungen zu Grunde liegenden Originaltexte (1) als auch die natürlichen Texte (Urteile, Pressemitteilungen und Zeitungsartikel) (2) abgebildet.

		Komplexitätstyp		
		Nom.	NP	Synt.
Version	Original	A1 (200)	B1 (303)	C1 (434)
	moderat	A2 (214)	B2 (317)	C2 (439)
	stark	A3 (217)	B3 (334)	C3 (440)

Tab. 5.1: Das dem Reformulierungsteil des FLRC zu Grunde liegende Pseudodesign sowie die Gesamtanzahl der Wörter in der jeweiligen Zelle in Klammern.

Aus Gründen der einfacheren Darstellung werden die mit (1) bezeichneten Originalsätze, die Ausgangspunkt der Reformulierungen waren, zwar in Abbildung 5.1 dem FLRCNat-Teil zugerechnet. Bei den meisten Erklärungen, Darstellungen und späteren Analysen wird von diesen Sätzen jedoch als Teil des FLRCRef-Subkorpus gesprochen. Dies ist insofern sinnvoll, als dass sie in einem  $(3) \times 3$ -Design der ersten Ausprägung des Faktors Reformulierungsgrad (keine Reformulierung/Original) entsprechen. Dieses Design, das der Reformulierungsstudie zu Grunde liegt, ist in Tabelle 5.1 dargestellt. Das Design ist kein vollständig gekreuztes Design, da nicht jeder Text in jedem Komplexitätstyp vorliegt. Der „Pseudofaktor“ Komplexitätstyp ist daher sowohl im Fließtext als auch in Abbildung 5.1 geklammert. Die Bedingungen sind entsprechend in der Horizontalen nicht durchnummeriert, sondern über Buchstaben repräsentiert. Jede Zelle des Designs ist mit zehn Texten belegt, wobei zwischen den Texten in einer Zeile (also in der Variation des Faktors Komplexitätstyp) keine Beziehung besteht. Von Zeile zu Zeile steigt für jeden der Texte der Grad der Reformulierung. Über das gesamte Design hinweg sind 90 Texte enthalten, wobei 60 davon (Z. 2 und 3) Reformulierungen sind.<sup>50</sup> Für sechs der zehn Texte mit vielen Nominalisierungen (Komplexitätstyp *Nom.*) gilt, dass sich Bedingungen A2 und A3 nicht unterscheiden. Dies war bereits bei der Studie von Hansen et al. (2006) der Fall. Der Grund hierfür ist, dass einige der Originaltexte aus der Komplexitätskategorie *Nom.* zu kurz waren, um noch mehr Nominalisierungen in verbale Strukturen umzuwandeln. Dies ist insofern verständlich, als ohnehin nur Nominalisierungen mit der Endung *-ung* reformuliert wurden, was die Menge an potenziellen Reformulierungen weiter einschränkt. Wir haben diese Besonderheit übernommen, um die Vergleichbarkeit der Studien zu gewährleisten. Die Wortanzahlen in jeder der

<sup>50</sup> Eventuell ist dieses Design einfacher zu verstehen, wenn man es anhand eines hypothetischen Textes darstellt: Gegeben sei ein Originaltext  $x_1$  mit vielen komplexen Nominalphrasen. Somit fällt er in Zeile 1, Spalte 2 (Zelle B1) des in Tabelle 5.1 dargestellten Designs. Die moderate Reformulierung  $x_2$  dieses Textes fällt in Zelle B2, die starke Reformulierung  $x_3$  in Zelle B3. Da keine Originalversion des Texts  $x_1$  mit vielen Nominalisierungen vorliegen **kann** (da es dann kein Originaltext mehr wäre), geht Text  $x_1$  nicht in die Spalten 1 und 3 des Designs ein.

neun Zellen (in Tab. 5.1 in Klammern abgetragen) zeigen, dass alle Originaltexte durch die Reformulierungen etwas länger werden. Bei den von Hansen et al. (2006) angewandten Reformulierungsstrategien ist dies zu erwarten. Das wird in Abschnitt 5.2 klar, wenn die Folgen der Reformulierungen herausgearbeitet werden.

Alle Reformulierungen und Originalauszüge, aus denen sie erstellt wurden, wurden mit der halbautomatischen Annotationssoftware *Annotate*<sup>51</sup> (Plaehn 1998) mit einer Phrasenstrukturannotation versehen. Diese umfasst außerdem Parts-of-Speech gemäß dem Stuttgart-Tübingen-Tagset.<sup>52</sup> Die Phrasenstrukturannotation enthält außerdem Kantenlabels, die jeder Phrase eine grammatische Funktion (als Beziehungen zwischen Phrasen) zuweisen. Alle Texte wurden von zwei Personen unabhängig voneinander annotiert. Nicht-Übereinstimmungen wurden in Diskussionen geklärt.

### FLRCNat: Natürliche Texte

Die natürlichen Texte sind in Abbildung 5.1 im Kasten unten rechts anzusiedeln (bezeichnet mit (2)). Dieser Korpusenteil enthält insgesamt neun Texte, die in diesem Wortlaut in den jeweiligen Veröffentlichungsmedien erschienen sind. Dabei sind drei Themenbereiche sowohl über das Originalurteil als auch über die dazugehörige Pressemitteilung und einen Zeitungsartikel, der von diesem Urteil handelt, abgedeckt. Diese Themenbereiche umfassen im Einzelnen

- 1) die Verfassungsmäßigkeit der Strafvorschriften über den Umgang mit Cannabis (konkretes Normenkontrollverfahren),
- 2) die Beteiligung deutscher Soldaten am AWACS-Einsatz der NATO in der Türkei (Verfahren des vorläufigen Rechtsschutzes),<sup>53</sup>
- 3) sowie die Strafbarkeit des unerlaubten Aufenthalts in Deutschland (Verfassungsbeschwerde).

Mit Entscheidung Nummer 1 (AZ 2 BvL 8/02, 29. Juni 2004) bestätigte das Bundesverfassungsgericht eine frühere Entscheidung zur Strafbarkeit des Umgangs mit Cannabis. Eine Vorlage des Amtsgerichts Bernau wurde abgelehnt, da das höchste deutsche Gericht keine „entscheidungserheblichen neuen Erkenntnisse“ (Bundesverfassungsgericht 2004) zur (Un-)Gefährlichkeit des Cannabiskonsums sah, die eine Neuregelung nahegelegt hätten.

<sup>51</sup> An dieser Stelle sei unserer damaligen Hilfskraft Peter Baumann herzlich gedankt, der das nicht ganz unerhebliche Problem einer Installation von *Annotate* löste.

<sup>52</sup> Das Tagset ist unter anderem abzurufen unter [www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/TagSets/stts-table.html](http://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/TagSets/stts-table.html) (Stand: 15.3.2015).

<sup>53</sup> Beim Hauptsacheverfahren handelte es sich um einen Organstreit.

Mit Entscheidung Nummer 2 (AZ 2 BvQ 18/03, 5. März 2003) lehnte das Bundesverfassungsgericht einen Antrag der FDP-Bundestagsfraktion auf Erlass einer einstweiligen Anordnung gegen die Bundesregierung ab. Die FDP-Fraktion wollte erreichen, dass der Bundestag unverzüglich über die Teilnahme deutscher Soldaten an einem AWACS-Einsatz in der Türkei abstimmt. Dies begründete die FDP-Fraktion damit, dass die Teilnahme deutscher Soldaten an dem NATO-Einsatz keine Routinemaßnahme sei. Es handele sich vielmehr um einen militärischen Einsatz, der der Zustimmung des Bundestags bedürfe, und es sei unrealistisch anzunehmen, dass der Einsatz der AWACS-Flugzeuge strikt zu trennen sei vom Einsatz anderer Flugzeuge über dem Irak. Das Bundesverfassungsgericht lehnte diesen Antrag ab: „Eine Folgenabwägung fällt zu Lasten der Antragsstellerin [= der FDP-Bundestagsfraktion, S.W.] aus.“ (Bundesverfassungsgericht 2003b). In der angesprochenen Folgenabwägung wurden die Rechte des Parlaments gegen die außenpolitische Verantwortung der Exekutive abgewogen. Durch eine einstweilige Verfügung wäre nach Einschätzung des Bundesverfassungsgerichts die Handlungsfähigkeit der Bundesregierung zu stark eingeschränkt worden – diese stehe der gefährdeten Rechtsposition des Bundestags zumindest gleichwertig gegenüber.

Entscheidung Nummer 3 (AZ 2 BvR 397/02) erklärte die Verfassungsbeschwerde eines Syrers für erfolgreich, der mit gefälschtem Reisepass nach Deutschland eingereist war und seine eigenen Identitätspapiere bewusst im Heimatland zurückgelassen hatte. Sein Asylantrag wurde abgelehnt, und

[o]bwohl er vollziehbar ausreisepflichtig war, wurde seine Abschiebung seitens der Ausländerbehörde nicht in die Wege geleitet. Die Beschaffung eines Heimreisedokuments verzögerte sich mangels Vorliegens der notwendigen Identitätsnachweise. (Bundesverfassungsgericht 2003a)

Die Aussetzung der Abschiebung (d.h. die Duldung) wurde dem Syrer aufgrund der Verzögerungen erst neun Monate später bewilligt. Da ein Ausländer, der eine nach dem Ausländergesetz erforderliche Aufenthaltsgenehmigung nicht besitzt, zur Ausreise verpflichtet ist, der Syrer sich jedoch in der Zwischenzeit bis zur Erteilung der Duldung im Bundesgebiet befand, wurde er zu einer Freiheitsstrafe von vier Monaten (ohne Bewährung) verurteilt. Das Bundesverfassungsgericht hielt dieses Vorgehen für verfassungswidrig, da unabhängig vom Verschulden des Ausländers eine Duldung zu erteilen sei. „Die Systematik des Ausländergesetzes lässt dabei grundsätzlich keinen Raum für einen unregelmäßigen Aufenthalt, der den Zeitpunkt der Duldungserteilung ins Belieben der Behörden stellt.“ (Bundesverfassungsgericht 2003a).

Dem Ausländer muss auch dann eine Duldung erteilt werden (und zwar sofort), wenn er selbst für die Entstehung eventueller Hindernisse verantwortlich ist.

Die Texte wurden einerseits hinsichtlich ihrer Länge ausgewählt (je ein kürzeres, mittellanges und etwas längeres Urteil), andererseits mussten Pressemitteilungen und Zeitungsartikel zu den Urteilen verfügbar sein. Als drittes Auswahlkriterium sollten die Themen möglichst breit gefächert sein. Außerdem sind – sofern man das Verfahren des vorläufigen Rechtsschutzes entsprechend des Hauptsacheverfahrens einordnet – alle wichtigen Verfahrensarten des Bundesverfassungsgerichts abgedeckt.

Sicherlich wäre es wünschenswert, eine breitere Auswahl an Texten und somit auch Themen zur Verfügung zu haben. Der (pragmatische) Flaschenhals sind jedoch die Urteile, die im Allgemeinen sehr lang sind und mit teilweise über 80 Bildschirmseiten schon jetzt für einige VersuchsteilnehmerInnen eine große Herausforderung darstellten. Es ist daher illusorisch, mehr als ein Urteil pro Versuchsdurchlauf zu präsentieren. Wir haben uns deshalb dazu entschieden, lediglich jeweils drei Urteile, Pressemitteilungen und Zeitungsartikel aufzunehmen. Diese wurden dafür jeweils von mehr TeilnehmerInnen gelesen. Für die erhobenen Daten bedeutet das, dass sie eher über die Population der Leser als über die Menge aller möglichen Texte verallgemeinerbar sind. Eine gewisse Verallgemeinerbarkeit ist trotzdem über die breite Themenauswahl gewährleistet. Außerdem sind insbesondere die Urteile sehr lang. Sie sollten also auch viele Instanzen der Merkmale enthalten, die für die vorliegende Arbeit von Interesse sind.

Die kompletten Zeitungsartikel, Pressemitteilungen und Urteile wurden vollautomatisch mit dem Berkeley Parser (Petrovet et al. 2006; Petrov/Klein 2007) annotiert. Dieser Parser erwies sich als geeigneter als vergleichbare Software. Einerseits, weil er in Tests genauere Annotationen anfertigte, andererseits weil das Ausgabeformat sowie das verwendete Tagset besser in das Annotate-Ausgabeformat überführbar war. Diese Überführung geschah anhand eines eigens dafür angefertigten R-Scripts. Die vollautomatischen Annotationen wurden stichprobenhaft überprüft und wo nötig angepasst. Es wurde eine Parametrisierung des Parsing-Vorgangs gewählt, die maximale Genauigkeit bei geringerer Geschwindigkeit erlaubte.

### 5.1.2 Methode und Versuchsablauf

Die Blickbewegungsdaten der Versuchsteilnehmenden wurden mit einem *SR Research EyeLink 1000* gemessen, der mit einer zeitlichen Auflösung von

1.000 Hz Daten über die Blickposition zur Verfügung stellt. Dies entspricht einem Datenpunkt (Sample) pro Millisekunde. Interne Algorithmen des *SR Research Data Viewer* – es wurden ihre Standardeinstellungen benutzt – rechnen diese Samples in Fixationen und Sakkaden um, die als Grundlage weiterer Auswertungen dienen. Um die größtmögliche Messgenauigkeit zu gewährleisten, wurde eine Kinnstütze eingesetzt, die in einem festen Abstand (60 cm) zum Monitor montiert wird und dabei Kinn und Stirn der Teilnehmenden abstützt. Vor der Versuchsdurchführung wurde mittels eines standardisierten Tests das dominante Auge<sup>54</sup> der Versuchsteilnehmenden ermittelt. Die Blickbewegungen wurden monokular anhand des dominanten Auges erfasst.

Die Daten für den FLRCRef-Teil mit den Originalsätzen wurden getrennt von jenen der natürlichen, ganzen Texte (Urteile, Pressemitteilungen und Zeitungsartikel) erhoben. Alle VersuchsteilnehmerInnen rekrutierten sich aus der Studierendenschaft der Universität Freiburg. Für die Teilnahme erhielten sie entweder eine Vergütung von 7,50 € pro Stunde oder VersuchsteilnehmerInnen-Stunden, die sie bis zu einem gewissen Punkt im Studium sammeln mussten. Das Alter der VersuchsteilnehmerInnen wurde nicht erhoben. Keine/r der TeilnehmerInnen war juristisch vorgebildet, was über die Abfrage von ehemaligen und aktuellen Studienfächern sichergestellt wurde. 40 LeserInnen nahmen an der Datenerhebung für den FLRCRef-Teil (mit Originalsätzen) teil. Davon waren 26 (65%) weiblich. Bei 10 LeserInnen (25%) musste das Sehen mit Kontaktlinsen oder einer Brille korrigiert werden. Die TeilnehmerInnen wurden im Vorfeld gebeten – sofern vorhanden – sowohl Brille als auch Kontaktlinsen mitzubringen, damit die für das Blickbewegungsmessgerät am wenigsten problematische Sehhilfe (meist die Brille) benutzt werden konnte. 41 LeserInnen nahmen an der Datenerhebung für die natürlichen Texte teil, die Daten einer/s Lesenden mussten verworfen werden. 25 (62,5%) der verbleibenden 40 TeilnehmerInnen waren weiblich. 14 TeilnehmerInnen (35%) trugen eine Brille oder Kontaktlinsen bei der Datenerhebung.

Die Texte wurden auf einem 20-Zoll-Flachbildschirm mit einer Auflösung von 1.600 × 1.200 Pixeln präsentiert. Für eine möglichst natürliche Leseerfahrung wurde eine proportionale Serifenschriftart (Times New Roman) für die Auswertung verwendet. Eine Bildschirmseite enthielt maximal elf Zeilen Text im 1,5-Zeilenabstand. Auf diese Weise wurden maximal ca. 540 Zeichen pro

---

<sup>54</sup> Das dominante Auge wird im Allgemeinen als das Auge definiert, dessen Input in Aufgaben, in denen nur ein Auge benutzt werden kann, bevorzugt wird. Beim Lesen „führt“ das dominante Auge meist den Leseprozess, das nicht-dominante Auge folgt wenige Millisekunden später. Für eine Diskussion verschiedener Befunde zur Augendominanz, insbesondere unter dem Blickwinkel von Beeinträchtigungen, vgl. Porac/Coren (1976).

Bildschirm präsentiert. Die Texte wurden so auf den Bildschirmseiten angeordnet, dass – wo immer möglich – Sätze nicht über mehrere Bildschirmseiten verteilt waren. Um die Leseerfahrung möglichst natürlich zu halten, wurde den VersuchsteilnehmerInnen bei der Erhebung der kompletten natürlichen Texte (in Abb. 5.1 mit (2) markiert) die Möglichkeit eingeräumt, im Text vor und zurück zu blättern. Für die Texte aus dem FLRCRef-Teil war dies nicht nötig, da diese Texte immer auf eine Bildschirmseite passten.

### **Versuchsablauf FLRCRef: Reformulierungsteil**

Die Versuchsabläufe waren leicht unterschiedlich für den FLRCRef-Teil (und Originaltexte) und die kompletten Texte des FLRCNat-Teils. Im ersten Fall wurde den Versuchsteilnehmenden eine Liste von Stimuli präsentiert, die gemäß eines Lateinischen-Quadrat-Versuchsdesigns rotiert waren. Das heißt, dass jeder Text von einer/einem VersuchsteilnehmerIn nur in einer der drei Reformulierungsstufen (Original/moderat/stark) gelesen wurde. Bei drei Versuchsbedingungen (gegeben durch die verschiedenen Reformulierungsebenen, siehe Tab. 5.1) ergeben sich drei Listen mit je 30 Texten, die in je acht verschiedenen, randomisierten Reihenfolgen vorlagen. Nach der Ermittlung des dominanten Auges, dem Ausrichten der Kameras sowie der anschließenden Kalibrierung und Validierung der Messung wurde den TeilnehmerInnen eine Versuchsinstruktion präsentiert. Die TeilnehmerInnen wurden in der Instruktion darauf hingewiesen, „gründlich und aufmerksam“ zu lesen. Ihnen wurde außerdem der Ablauf jedes Versuchsdurchlaufs (inkl. der Annahme des *Drift Checks*) erklärt und erläutert, wie sie auf die Fragen antworten können. Danach folgten sechs Trainingsdurchläufe. Für jeden Trial folgte der Ablauf immer dem folgenden Schema:

- 1) *Drift Check*: Auf dem Bildschirm erschien ein Punkt, den die Versuchsteilnehmenden anschauen sollten. Währenddessen drückten sie eine Taste auf einer Antwortbox,<sup>55</sup> die sie über die komplette Dauer des Experiments in der Hand hielten. Entstanden hier zu hohe Abweichungen zwischen der tatsächlichen Position des Punkts und der gemessenen Blickposition, musste das Blickbewegungsgerät neu kalibriert werden. Der Punkt zum *Drift Check* erschien auf der Position des ersten Buchstabens des darauf folgenden Texts.
- 2) Vor dem eigentlichen Stimulustext erschien in einigen Fällen ein Text, der Kontext zur Interpretation des eigentlichen Stimulustexts bereitstellte.

<sup>55</sup> Unter [www.wolferonline.de/diss](http://www.wolferonline.de/diss) kann ein Bild abgerufen werden, das die Antwortbox und die Tastenbelegung illustriert (Stand: 6.3.2017).



Nach dem Lesen dieses Texts drückten die Versuchsteilnehmenden die Antwort-Box, damit ihnen der eigentliche Stimulustext präsentiert wurde. Benötigten die Versuchsteilnehmenden länger als 30 s um den Knopf zu drücken, erschien der Stimulustext automatisch.

- 3) Die Versuchsteilnehmenden hatten 60 s Zeit, um den Stimulustext zu lesen. Im Normalfall betätigten sie jedoch vorher die Antwortbox, um mit der Frage fortzufahren. Benötigten sie mehr als eine Minute, um den Text zu lesen, wurde automatisch die Frage präsentiert.
- 4) Die Versuchsteilnehmenden hatten nun 15 s Zeit, um die Frage zu lesen und zu beantworten. Um die Frage mit 'ja' zu beantworten, mussten sie die Taste an ihrem linken Zeigefinger (an der Antwortbox hinten links) betätigen. Um die Frage mit 'nein' zu beantworten, mussten sie die Taste an ihrem rechten Zeigefinger (an der Antwortbox hinten rechts) betätigen. Zur Unterstützung der Antwort wurden die Antwortmöglichkeiten unten links und unten rechts auf dem Bildschirm eingeblendet. Die VersuchsteilnehmerInnen bekamen **keine** Rückmeldung, ob die Antwort, die sie auf die Frage gegeben hatten, korrekt war oder nicht. Dadurch wird vermieden, dass die LeserInnen sich noch während des folgenden Versuchsdurchgangs über die zuvor beantwortete Frage Gedanken machen. Diese Gefahr besteht insbesondere bei falsch beantworteten Fragen.
- 5) Der nächste Versuchsdurchlauf begann bei Punkt 1 dieser Liste.

Außer den Lesezeitdaten wurden zwei Reaktionszeiten aufgezeichnet: die Zeit von der Darbietung des Stimulustexts bis zum Fortschreiten zur Frage (Punkt 3) sowie die Zeit von der Darbietung der Frage bis zur Antwort (Punkt 4).

### **Versuchsablauf FLRCNat: Natürliche Texte**

Der Ablauf vor dem eigentlichen Versuch war für die Erhebung der Blickbewegungen auf natürlichen Texten identisch mit der Reformulierungsstudie. Auch hier waren die neun Texte anhand eines lateinischen Quadrats auf Listen verteilt. Auf diese Weise las jede/r TeilnehmerIn je ein Urteil, eine Pressemitteilung und einen Zeitungsartikel – allerdings niemals zwei Texte zu demselben Thema. Es ergeben sich durch die Kombination aller Texte gemäß des Lateinischen-Quadrat-Designs 18 Listen. Nach einer Instruktion, die in relevanten Teilen identisch war mit der bereits vorgestellten, folgte die Erhebung dem folgenden Ablauf:

- 1) Vor der ersten und danach jeder siebten Bildschirmseite wurde ein *Drift Check* (siehe Punkt 1 im Ablauf der Reformulierungsstudie) durchgeführt. Hierdurch sollte eine natürlichere Leseerfahrung gewährleistet werden, die eventuell durch einen *Drift Check* vor jeder Seite unterbrochen worden wäre. Der *Drift Check* musste mit einer Taste auf der Antwortbox bestätigt werden.
- 2) Eine Bildschirmseite wurde präsentiert, die entweder zur nächsten Seite (rechte Taste an der Antwortbox) oder zur vorherigen Seite (linke Taste an der Antwortbox) verlassen werden konnte. Unten rechts auf der Bildschirmseite wurde die aktuelle und die Gesamtseitenzahl präsentiert. Diese Einblendung folgte dem Muster Seite<sub>aktuell</sub>/Seite<sub>max</sub> und war während der gesamten Zeit präsent.
- 3) Nach jedem Text wurden fünf Fragen präsentiert.<sup>56</sup> Es war den Versuchsteilnehmenden nicht möglich, zurück in den Text zu blättern, sobald sie zu den Fragen fortgeschritten waren. Diese Fragen waren nach dem Multiple-Choice-Verfahren zu beantworten. Von vier Antwortmöglichkeiten war immer eine korrekt. Die Fragen waren mit den Zifferntasten 1 bis 4 auf der Tastatur zu beantworten. Über ein akustisches Signal wurde den VersuchsteilnehmerInnen rückgemeldet, ob sie die Frage korrekt beantwortet haben. Es bestand kein Zeitdruck beim Beantworten der Fragen, die nächste Frage bzw. die erste Seite des nächsten Texts erschien immer erst nach einem Tastendruck.
- 4) Ein Punkt zum *Drift Check* wurde zu Beginn des folgenden Texts präsentiert. Der Ablauf beginnt somit erneut bei Punkt 1.

### 5.1.3 Fragen

Um die VersuchsteilnehmerInnen dazu anzuhalten, die ihnen präsentierten Texte aufmerksam zu lesen, wurde(n) ihnen nach jedem Text eine oder mehrere Fragen gestellt. Nach den Texten aus der Reformulierungsstudie war dies je eine Entscheidungsfrage (ebenfalls übernommen aus Hansen et al. 2006), die sich meist auf eine propositionale Einheit des Texts bezog. Bei wenigen Fragen waren einfache Inferenzen nötig. Über einen kompletten Versuch hinweg waren die Fragen gleich häufig positiv und negativ zu beantworten.

Beispiel 7 zeigt Kontext (Beispiel 7a), Stimulustext (Beispiel 7b) sowie Frage (Beispiel 7c) für einen Versuchsdurchgang (ein Stimulustext, der nach Tabelle 5.1 in Zelle B3, Komplexitätstyp Nominalphrasen, starke Reformulierung, fällt).

<sup>56</sup> Die Alternative, längere Texte durch mehrere Fragenblöcke zu unterbrechen, haben wir deshalb nicht gewählt, weil wir den natürlichen Lesefluss nicht unterbrechen wollten.

- (7) a. Der Verfassungsbeschwerde kommt keine grundsätzliche verfassungsrechtliche Bedeutung zu.
- b. Die Frage, ob dem Land Schleswig-Holstein die Gesetzgebungskompetenz zur Erhebung der Abgabe zusteht, lässt sich mit Hilfe der vorhandenen Maßstäbe entscheiden. Diese sind im Beschluss des Zweiten Senats des Bundesverfassung vom 7. November 1995 entwickelt worden.
- c. Lässt sich entscheiden, wem die Gesetzgebungskompetenz zusteht?

In diesem Fall wäre Frage 7c korrekt mit 'ja' zu beantworten, denn der Matrixsatz in Beispiel 7b ('Die Frage [...] lässt sich [...] entscheiden') ist nicht negiert. Die Fragen für die Reformulierungsstudie wurden wie die Texte von Hansen et al. (2006) übernommen.

Die Fragen für die natürlichen Texte wurden gemeinsam mit einer juristisch vorgebildeten Hilfskraft erstellt. Dabei wurde darauf geachtet, dass einige Fragen (ähnlich wie in der Reformulierungsstudie) auf die propositionale Ebene abzielen. Diese Fragen konnten meist über einen einfachen Rückbezug auf eine Proposition im Text beantwortet werden. Andere Fragen – und dies stellt eine Abweichung zur Reformulierungsstudie dar – wurden so konzipiert, dass sie dann korrekt beantwortet werden konnten, wenn die/der LeserIn zuvor ein kohärentes Situationsmodell (Kintsch/van Dijk 1978; Kintsch 1988, 1998; Zwaan/Radvansky 1998) des Texts aufgebaut hatte. Beispiel 8 zeigt eine solche Frage exemplarisch, sie wurde den TeilnehmerInnen nach Entscheidung Nummer 2 (AWACS, AZ 2 BvQ 18/03, siehe 5.1.1) präsentiert. Die Sätze 8a bis 8d wurden den VersuchsteilnehmerInnen als Antwortmöglichkeiten präsentiert.

- (8) Welcher mögliche Nachteil könnte sich nach Ansicht der Antragsgegnerin daraus ergeben, wenn die einstweilige Anordnung erginge?
- a. Deutschland würde sich ohne Zustimmung des Bundestags an einem militärischen Auslandseinsatz beteiligen.
- b. Deutschland würde wirtschaftlich Schaden nehmen, da es an Projekten zum Wiederaufbau der irakischen Infrastruktur nicht entsprechend beteiligt würde.
- c. Die Zahl der deutschen Besatzungsmitglieder in den AWACS-Flugzeugen der NATO würde stark zunehmen.
- d. Deutschland würde die gemeinsam mit anderen NATO-Mitgliedern vereinbarte Hilfe für die Sicherung der territorialen Integrität der Türkei zum Scheitern bringen.

Korrekt ist Aussage 8d. Allerdings kann dies nicht direkt aus dem Urteilstext geschlossen werden, sondern muss über mehrere Zusammenhänge rekonstruiert werden. Insbesondere muss im mentalen Situationsmodell repräsentiert sein, wer Antragsgegnerin war (in diesem Fall die Bundesregierung). Warum Informationen aus der gesamten Textrepräsentation zur Lösung der Frage genutzt werden müssen, wird noch deutlicher, wenn man die Antworten berücksichtigt, die als Disktraktoren dienen. Antwort 8a ist zwar rein oberflächlich sinnvoll, aber als Antwort auf genau diese Frage falsch, weil das nicht die Befürchtung der Antragsgegnerin ist, sondern jene der **Antragsstellerin** (der FDP-Fraktion des Bundestags) – diese Befürchtung ist ja eben der Grund für den Antrag auf einstweilige Verfügung, über den das Bundesverfassungsgericht entschied. Antwort 8b mag plausibel klingen, dieser Themenkomplex wird aber im kompletten Urteil nicht einmal angeschnitten.

An dieser Frage mag es relativ einfach sein, zu exemplifizieren, warum ein Zugriff auf das Situationsmodell des Textes eher nötig ist als der Abruf einer einzelnen Proposition. Dies ist aber nicht immer so einfach trennbar. Selbst wenn die relevante Information in Form einer propositionalen Einheit im Text vorkommt, kann es unter Umständen für LeserInnen einfacher sein, sich die Information über das Situationsmodell zu erschließen. Das ist insbesondere dann der Fall, wenn die Proposition, auf die sich eine Frage direkt bezieht, am Anfang eines langen Texts kodiert ist. Beispiel 9 soll das verdeutlichen. Diese Frage wurde ebenfalls nach Entscheidung Nummer 2 präsentiert und fragt die grundlegende Rollenverteilung der Konfliktparteien ab. Korrekt ist Antwort 9a.

- (9) Wer ist Antragsstellerin, wer Antragsgegnerin?
- a. Antragsstellerin ist die FDP-Fraktion im Deutschen Bundestag, Antragsgegnerin die Bundesregierung.
  - b. Antragsstellerin ist die Bundesregierung, Antragsgegnerin die FDP-Fraktion im Deutschen Bundestag.
  - c. Antragsstellerin ist die NATO, Antragsgegnerin die türkische Regierung.
  - d. Antragssteller ist der Bundestag, Antragsgegnerin die Bundeswehr.

Antragsstellerin und Antragsgegnerin werden zu Beginn der Entscheidung explizit benannt. Im weiteren Text werden die beiden Akteure immer wieder aufgegriffen und die anderen, in den Distraktoren genannten Akteure (die NATO, die türkische Regierung, der Bundestag und die Bundeswehr) sind im Laufe der Entscheidung von Bedeutung. Es ist davon auszugehen, dass die

Rollenverteilung zentral für den Aufbau des Situationsmodells ist, denn eine der relevantesten Informationen in Entscheidungen ist, welche Parteien sich gegenüberstehen und welche Positionen gegeneinander abgewogen werden. Wer den Text also verstanden (= ein korrektes Situationsmodell zum Text aufgebaut) hat, sollte diese Rollenverteilung schnell daraus ableiten können. Dabei sollte es nicht mehr nötig sein, auf die eine ganz zu Beginn des Leseprozesses aufgebaute Proposition zuzugreifen. Anders formuliert: Diese eine Proposition nimmt in der Gesamtrepräsentation des Texts eine solch zentrale Stellung ein,<sup>57</sup> dass die Antwortperformanz nicht mehr vom Abstand zwischen erster Enkodierung und Abfrage beeinflusst sein sollte.

Die Aufteilung der Fragen in 'auf eine einzelne Proposition abzielend' vs. 'auf das Situationsmodell abzielend' ist also nicht so eindeutig, wie sie zunächst scheinen mag. In den entsprechenden Analysekapiteln wird die Aufteilung der Fragen in diese Kategorien trotzdem aufrechterhalten.

Noch ein Punkt ist bezüglich der verwendeten Fragen als kritisch einzuschätzen. Die Fragen (insbesondere im FLRCRef-Teil) wurden hauptsächlich dazu eingesetzt, um gewährleisten zu können, dass die VersuchsteilnehmerInnen die Texte aufmerksam lesen. Damit sind die Fragen für Analysen von textfernen Verstehensprodukten (vgl. Christmann 2004) wie erworbenes Faktenwissen, Überblicks- und Zusammenhangswissen eher ungeeignet. Um solche Konstrukte genauer erheben zu können, sind elaboriertere Tests notwendig (vgl. z.B. Wolfer et al. 2013). Die während der Erhebung des FLRC verwendeten Fragen sind hierzu nicht optimal geeignet. Auf diese Einschränkung wird an den relevanten Stellen erneut eingegangen.

## 5.2 Linguistische Oberflächenanalysen der verwendeten Texte

Bevor in den Kapiteln 6 und 7 die Ergebnisse auf der rezeptiven Ebene vorgestellt werden, sollen in den folgenden Abschnitten zunächst die linguistischen Eigenschaften der präsentierten Texte beleuchtet werden. Da dies keine korpuslinguistische Arbeit ist, werden dies keine erschöpfenden Analysen sein – sie werden aber dabei helfen, einen Eindruck der verwendeten Texte zu erhalten und die folgenden Blickbewegungsanalysen besser einzuschätzen.

Zunächst werden in Abschnitt 5.2.1 die in der Reformulierungsstudie verwendeten Texte analysiert. Hier werden wir insbesondere die verschiedenen Reformulierungsstufen untereinander vergleichen, um einen Eindruck davon zu bekommen, wie die reformulierten Texte aus der Studie von Hansen et al.

---

<sup>57</sup> Im Sinne von Kintsch/van Dijk (1978) ist diese Proposition als Makroproposition einzustufen.

(2006) entstanden sind. Die Komplexität eines durchschnittlichen Zeitungstexts diente ihnen als Vorbild während des Reformulierungsprozesses, weil Hansen et al. annehmen, dass „newspapers use a range of language well adapted to the reading habits of the lay reader“ (ebd., S. 25). Die Reformulierungen wurden von LinguistInnen vorgenommen. Die reformulierten Texte wurden einer juristisch vorgebildeten Gewährsperson vorgelegt, die sicherstellte, dass die Reformulierungen nicht den juristischen Gehalt der Texte verfälschte. Es lagen Richtlinien zur Reformulierung der Texte vor, die bei der konkreten Ausgestaltung der Reformulierungen Raum für intuitive Zugänge lassen. Die Richtlinie für Texte mit vielen Nominalisierungen (Zelle A1 in Tab. 5.1) war: „Ersetze \*ung-Nominalisierungen mit der verbalen Form des Lemmas, wodurch eine Nebensatzkonstruktion entsteht.“ In den Fällen, in denen die Reformulierung einen Infinitiv im Nebensatz zur Folge hat, wird dieser Nebensatz nochmals umformuliert, um eine finite Nebensatzstruktur zu erhalten. Beispiel 10 zeigt diesen Ablauf für die in Satz 10a hervorgehobene Konstruktion. Ziel der Reformulierung war die Nominalisierung ‚Veröffentlichung‘, die zunächst in eine erweiterte Infinitivkonstruktion umformuliert wurde (Satz 10b). Im zweiten Reformulierungsschritt (Satz 10c) wurde diese in eine Nebensatzkonstruktion mit finitem Verb umgewandelt. Für die Texte mit vielen Nominalisierungen bestand die Freiheit bei der Reformulierung insbesondere darin, welche konkrete Nominalisierung umgewandelt wurde. „The most plausible candidates [...] are accumulations of ‘ung’-nominalisations within the same noun phrase construction.“ (Hansen et al. 2006, S. 27).

(10) a. (Originalversion)

Bei der Abwägung zwischen den Belangen der Beschwerdeführerin einerseits und dem Persönlichkeitsrecht der Kläger andererseits gebühre dem Interesse der Kläger an einer **Veröffentlichung der Richtigstellung** auf der Titelseite der Vorrang.

b. (moderate Reformulierung)

Bei der Abwägung zwischen den Belangen der Beschwerdeführerin einerseits und dem Persönlichkeitsrecht der Kläger andererseits gebühre dem Interesse der Kläger daran, **eine Richtigstellung auf der Titelseite zu veröffentlichen**, der Vorrang.

c. (starke Reformulierung)

Bei der Abwägung zwischen den Belangen der Beschwerdeführerin einerseits und dem Persönlichkeitsrecht der Kläger andererseits gebühre dem Interesse der Kläger daran, **dass eine Richtigstellung auf der Titelseite veröffentlicht wird**, der Vorrang.

Bei Texten mit komplexen Nominalphrasen (Zelle B1) sollte Komplexität zunächst in untergeordnete Nebensätze ausgelagert werden (was die moderate Reformulierungsversion ergab, Zelle B2). Für die starke Reformulierungsversion (Zelle B3) sollten die Sätze geteilt werden, um subordinierende Strukturen zu vermeiden. In syntaktisch komplexen Texten (Zelle C1) sollten die Sätze geteilt werden. Die AutorInnen weisen darauf hin, dass es dabei insbesondere eine Herausforderung darstellte, die logischen Beziehungen zwischen Sätzen aufrecht zu erhalten:

Transferring the intrasentential logical relation to a sequence of logically related sentences may involve major restructuring, because sentence splitting can lead to ambiguous intersentential reference and to a loss of textual coherence. (Hansen et al. 2006, S. 26)

Dieser Sachverhalt und die Konsequenzen, die dadurch für die Verarbeitung entstehen, wird in der vorliegenden Arbeit noch häufiger aufgegriffen werden.

Welche linguistischen Folgen die Reformulierungen tatsächlich auf der sprachlichen Oberfläche hatten, werden wir für ausgewählte Phänomene zeigen. Man könnte diesen Abschnitt auch als eigene Ministudie lesen, in der untersucht wird, wie ausgebildete LinguistInnen (im Fall von Hansen et al. (2006) in Begleitung von juristischen Beratern) mit Texten umgehen, die sie anhand von mehr oder weniger genau gefassten Richtlinien reformulieren sollen. Wie bereits erwähnt: Die Richtlinien waren nicht so hart formuliert, dass für jeden Text nur eine mögliche reformulierte Version möglich war – wie diese Freiheiten genutzt wurden, werden wir analysieren.

Die natürlichen Texte werden in Abschnitt 5.2.2 untersucht. Wir werden dabei die verschiedenen Textgattungen untereinander vergleichen, um systematische Unterschiede zwischen ihnen herauszuarbeiten.

### 5.2.1 FLRCRef: Reformulierungskorpus

Aus den Reformulierungsrichtlinien ergeben sich Vorhersagen, auf welchen Dimensionen sich die Texte verändern sollten. So war es das Ziel, in den Texten mit vielen Nominalisierungen die Anzahl an \*ung-Nominalisierungen zu verringern. Ersetzt werden sollten diese durch die Bildung verbaler Strukturen. Der Anteil von Nominalisierungen am Gesamtvolumen lässt sich für jeden Text über  $n(\text{Nominalisierungen})/n(\text{Wörter})$  messen. Dieser Anteil ist im linken Teil von Abbildung 5.2 abgetragen.



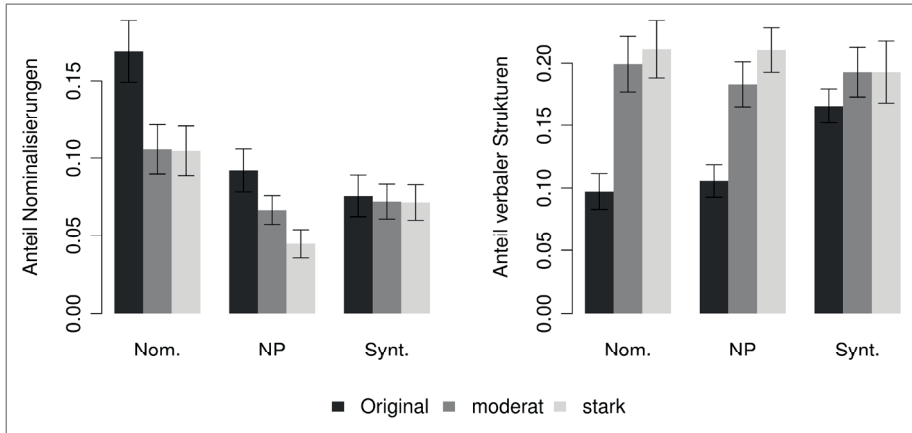


Abb. 5.2: Anteil Nominalisierungen (links) und verbaler Strukturen (rechts) in Abhängigkeit des Komplexitätstyps und der Reformulierungsversion. Fehlerbalken symbolisieren hier und in allen folgenden Schaubildern  $\pm 1$  Standardfehler. Weicht die Beschriftung von dieser Konvention ab, ist dies vermerkt.

Ob es sich bei einem Wort um eine Nominalisierung handelt, wurde manuell annotiert. Der Anteil verbaler Strukturen wurde in diesem Fall folgendermaßen gefasst:  $(n(\text{Verben})+n(\text{Partizipien}))/n(\text{Wörter})$ . Der rechte Teil von Abbildung 5.2 zeigt diesen Anteil in Abhängigkeit des Komplexitätstyps ( $x$ -Achse) und der Reformulierungsversion (Balken). Ein Fall ist in den folgenden Abbildungen ein Text. Pro Balken ist somit  $n = 10$ .

Offenbar waren die Reformulierungen in diesem Punkt erfolgreich. Der Anteil an Nominalisierungen sinkt für den Komplexitätstyp Nominalisierungen (jeweils die linke Balkengruppe) von Originalversion zu moderater Reformulierung um 6,3% von durchschnittlich 16,9% auf 10,6% ( $t(17,2) = -2,49$ ;  $p < 0,05$ ).<sup>58</sup> Der Anteil an Nominalisierungen bleibt dann konstant (stark reformulierte Version: 10,5%). Das ist einerseits darauf zurückzuführen, dass einige stark reformulierte Texte identisch sind zu den moderat reformulierten Texten.<sup>59</sup> Andererseits geht dies auch aus der von Hansen et al. (2006) verwendeten Reformulierungsrichtlinie hervor. Für den Übergang von Originaltext zu moderater Reformulierung wurde mindestens eine \*ung-Nominalisierung zur Umformulierung in erweiterte Infinitivkonstruktionen ausgewählt (siehe

<sup>58</sup> Wann immer in der vorliegenden Arbeit statistische Tests mit Anteilen berechnet wurden, wurden diese Tests mit Logit-Werten ( $L(x) = \ln(p(x)/1-p(x))$ ) parallel durchgeführt. Wo dies zu einem anderen Ergebnis führte, ist dies vermerkt.

<sup>59</sup> Dies gilt, wie in Abschnitt 5.1.1 dargelegt, nur für die Texte mit vielen Nominalisierungen (Komplexitätstyp *Nom.*).

Beispiel 10). Im weiteren Schritt (von moderater zu starker Reformulierung) wurden dann nicht weitere Nominalisierungen umformuliert, sondern – sofern überhaupt eine Veränderung von moderater zu starker Reformulierung vorlag – diese erweiterten Infinitivkonstruktionen umgewandelt, meist in Nebensätze mit finitem Verb. Dieses Vorgehen spiegelt sich unter anderem im signifikanten Anstieg (Original vs. moderate Reformulierung) von 8,4% auf 20,0% ( $t(17,7) = 3,93$ ;  $p < 0,01$ ) der verbalen Strukturen (Abb. 5.2, rechte Seite) wider, der gegenläufig zum Abfall der Nominalisierungen erfolgt. Auch hier ist kein signifikanter Unterschied mehr zwischen moderater und starker Reformulierungsversion festzustellen.

Beispiel 11 illustriert dieses Vorgehen und die linguistischen Auswirkungen. Einige Part-of-Speech-Tags (POS-Tags) sind zu Illustrationszwecken über Indizes vermerkt; verwendet wird, wie bereits a.a.O. erwähnt, das Stuttgart-Tübingen-Tagset.<sup>60</sup> Von Satz 11a (Original) zu Satz 11b (moderate Reformulierung) wurden vier Nominalisierungen ('Herstellung', 'Reduzierung', 2 x 'Aufhebung') in verbale Strukturen überführt ('hergestellt wurde', 'zu reduzieren oder aufzuheben', 'aufzuheben'). Dabei wurden zwei Infinitivgruppen eingeführt, die 'Ermächtigung' und 'Rechtspflicht' (beide kursiv hervorgehoben) modifizieren. Die erste Infinitivgruppe wurde dann im zweiten Schritt (Satz 11b zu Satz 11c) in den mit 'dass' eingeleiteten, untergeordneten Nebensatz transformiert. Dabei werden zwei Infinitivformen ('reduzieren', 'aufzuheben') in drei finite Formen ('reduziert', 'aufgehoben', 'wird') umgewandelt. Die zweite Infinitivgruppe bleibt gleich wie in Satz 11b.

(11) a. (Originalversion)

Mehr als zwölf Jahre nach<sub>APPR</sub> der Herstellung<sub>NN</sub> der deutschen Einheit habe sich die *Ermächtigung* zur Reduzierung<sub>NN</sub> oder Aufhebung<sub>NN</sub> der Gebühreneräßigung durch Rechtsverordnung zu einer *Rechtspflicht* zur Aufhebung<sub>NN</sub> des Gebührenabschlags verdichtet.

b. (moderate Reformulierung)

Mehr als zwölf Jahre nachdem<sub>KOUS</sub> die deutsche Einheit herstellt<sub>VPPP</sub> wurde<sub>VAFIN'</sub> habe sich die *Ermächtigung*, die Gebührenermäßigung durch Rechtsverordnung zu reduzieren<sub>VVINFINF</sub> oder aufzuheben<sub>VVIZU'</sub> zu der *Rechtspflicht* verdichtet, den Gebührenabschlag aufzuheben<sub>VVIZU'</sub>.

<sup>60</sup> Das Tagset ist unter anderem abzurufen unter [www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/TagSets/stts-table.html](http://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/TagSets/stts-table.html) (Stand: 15.3.2015).

c. (starke Reformulierung)

Mehr als zwölf Jahre nachdem<sub>KOUS</sub> die deutsche Einheit hergestellt<sub>VVPP</sub> wurde<sub>VAFIN'</sub> habe sich die *Ermächtigung*, dass<sub>KOUS</sub> die Gebührenermäßigung durch Rechtsverordnung reduziert<sub>VVPP</sub> oder aufgehoben<sub>VVPP</sub> wird<sub>VAFIN'</sub> zu der *Rechtspflicht* verdichtet, den Gebührenabschlag aufzuheben<sub>VVIZU'</sub>.

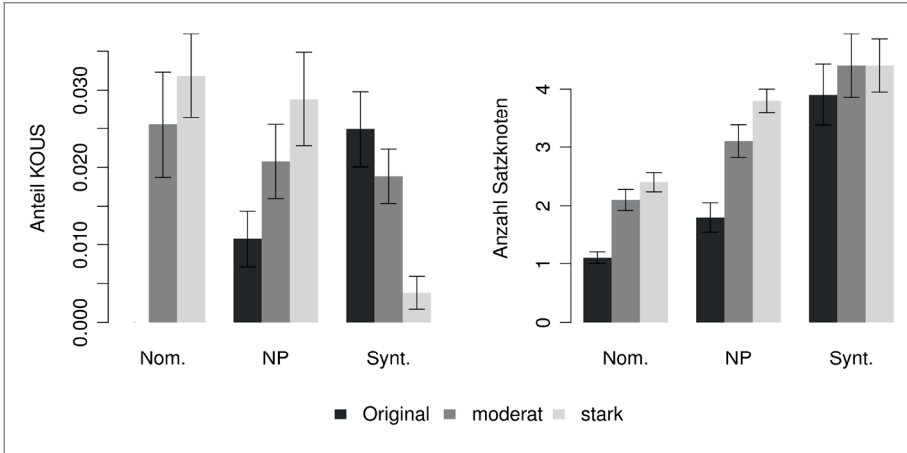


Abb. 5.3: Anteil des Part-of-Speech-KOUS, unterordnende Konjunktion mit Satz (links) und Anzahl der Satzknotten (rechts) in Abhängigkeit des Komplexitätstyps und der Reformulierungsversion.

Dass Beispiel 11 stellvertretend für ein häufiges Vorgehen bei der Reformulierung von Sätzen mit vielen Nominalisierungen in Hansen et al. (2006) ist, zeigt sich in den Abbildungen 5.2 und 5.3. Während die Anzahl der Nominalisierungen erwartungsgemäß sinkt, steigt die Anzahl verbaler Strukturen. Gleichzeitig werden in Texten des Komplexitätstyps *Nom.* überhaupt erst durch die Reformulierungen unterordnende Konjunktionen eingeführt (siehe linker Teil der Abb. 5.3). Beispiel 11 trägt zu dieser Verteilung durch das neu eingeführte ‘nachdem’ in Satz 11b sowie durch das den untergeordneten Nebensatz einleitende ‘dass’ (Satz 11c) bei.

Steigt die Anzahl unterordnender Konjunktionen, sollte die Anzahl der Satzknotten steigen. Wie Abbildung 5.3 (rechte Seite) zeigt, ist dies der Fall. Die durchschnittliche Anzahl Satzknotten in den Originalsätzen beträgt: 1,1 ( $s = 0,32$ ); moderate Reformulierungen: 2,1 ( $s = 0,57$ ); starke Reformulierungen: 2,4 ( $s = 0,52$ ). Ein signifikanter Anstieg über alle Texte des Komplexitätstyps *Nom.* hinweg ist wiederum nur für den Übergang von der Originalversion zu mo-

derater Reformulierung ( $t(14,8) = 4,68; p < 0,001$ ) festzustellen. Beispiel 11 illustriert auch diesen Sachverhalt: Während in der ersten Reformulierungsstufe ein Satzknoten (mit dem finiten Verb 'wurde') hinzukommt, ist es im Übergang von moderater zu starker Reformulierung ein weiterer (finites Verb: 'wird'). In bestimmter Hinsicht kann das vorgestellte Muster als Verlagerung der Komplexität von der lexikalischen auf die syntaktische Ebene verstanden werden: Wo vorher komplexe Nominalisierungen verwendet wurden, stehen jetzt untergeordnete Nebensätze oder erweiterte Infinitivkonstruktionen.

Besonders interessant wird diese Beobachtung, wenn wir sie mit syntaktisch komplexen Texten (Komplexitätstyp *Synt.*) vergleichen. Die Menge aller bisher eingeführten Parts-of-Speech bzw. Knotentypen bleibt konstant, bis auf jene der unterordnenden Konjunktionen mit Satz. Während zwischen den Originaltexten ( $\bar{x} = 2,5\%; s = 1,5$ ) und den moderaten Reformulierungen ( $\bar{x} = 1,9\%; s = 1,1$ ) kein signifikanter Unterschied besteht, fällt der Anteil unterordnender Konjunktionen mit Satz im Übergang zur stark reformulierten Version ( $\bar{x} = 0,38\%; s = 0,67$ ) deutlich ab ( $t(14,8) = -3,69; p < 0,01$ ). Im Lichte der vorherigen Ausführungen zu Texten des Komplexitätstyps *Nom.*, in dem ein Anstieg dieser Konjunktionen mit einem Anstieg der Menge an Satzknoten einherging, wäre zu erwarten, dass sich dieser Zusammenhang auch für die Texte mit komplexer Syntax zeigt. Dies ist nicht der Fall, wie Abbildung 5.3 (rechts) zeigt. Die Anzahl der Satzknoten verändert sich bei syntaktisch komplexen Texten nicht signifikant über die verschiedenen Reformulierungsstufen hinweg. Die Erklärung hierfür kann teilweise aus Abbildung 5.4 abgelesen werden. In syntaktisch komplexen Texten wurden Sätze geteilt.

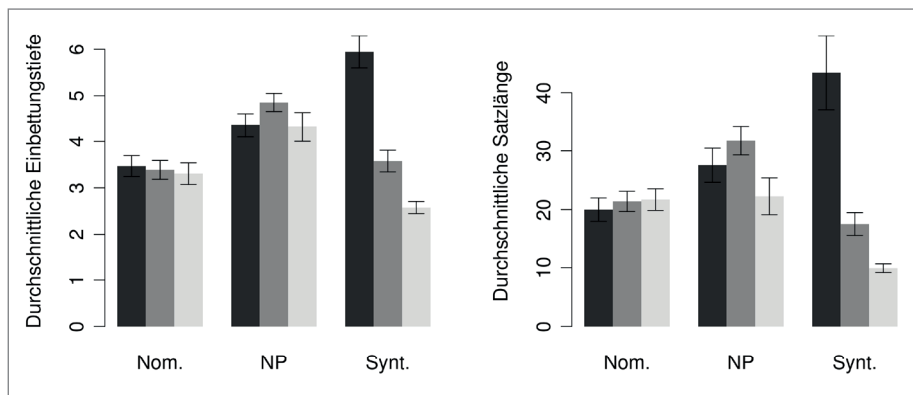


Abb. 5.4: Durchschnittliche Einbettungstiefe der Wörter (links) und durchschnittliche Länge der Sätze (rechts) in Abhängigkeit des Komplexitätstyps und der Reformulierungsversion.

Da nicht mehr semantischer Gehalt hinzukommt, hat dies deutliche Auswirkungen auf die durchschnittliche Satzlänge (Abb. 5.4 rechts), die in allen Reformulierungsschritten signifikant fällt (Vergleich Original ( $\bar{x} = 43,4$ ;  $s = 19,9$ ) – moderat ( $\bar{x} = 17,6$ ;  $s = 9,70$ ):  $t(10,8) = -3,93$ ;  $p < 0,01$ ; Vergleich moderat – stark ( $\bar{x} = 9,9$ ;  $s = 4,67$ ):  $t(31,1) = -3,69$ ;  $p < 0,001$ ).<sup>61</sup> Durch unterordnende Konjunktionen eingeleitete Nebensätze wurden also offenbar besonders oft vom Ursprungssatz abgetrennt und in einen eigenständigen Satz überführt. Dadurch sinkt die Anzahl der entsprechenden Konjunktionen und die durchschnittliche Satzlänge, jedoch nicht die Anzahl der Satzknoten.

Eine weitere Beobachtung (ohne Abbildung) bestätigt dieses Muster: Die Anzahl der Sätze pro Text der Komplexitätskategorie *Synt.* steigt deutlich an. Von 1 (Original, konstant) auf 2,5 (moderat,  $s = 0,71$ ) auf 4,1 (starke Reformulierung,  $s = 1,20$ ).<sup>62</sup> Wenn ehemals lange Sätze in mehrere kurze geteilt werden, sollten diese kürzeren Sätze auch geringere Einbettungstiefen aufweisen als die Originalsätze – nicht zuletzt sind tiefe Einbettungen als Komplexitätsquelle bekannt und sollten durch Reformulierungen aufgelöst werden. Genau dies ist für die Texte der Komplexitätsklasse *Synt.* der Fall (siehe Abb. 5.4 links, Rohwerte siehe Tab. 5.2). Abgetragen ist die durchschnittliche Einbettungstiefe der Wörter in den Texten. Ein alternatives Maß ist die maximale Einbettungstiefe pro Satz, die dasselbe Muster zeigt (ohne Abb., siehe aber Tab. 5.2).

	Nom.	NP	Synt.
Durchschnittliche Einbettungstiefe			
Original	3,47 (0,71)	4,35 (0,81)	5,94 (1,09)
moderat	3,39 (0,64)	4,86 (0,62)	3,58 (0,74)
stark	3,31 (0,74)	4,32 (1,00)	2,58 (0,39)
Durchschnittliche maximale Einbettungstiefe			
Original	5,1 (1,37)	7,2 (1,55)	9,7 (1,42)
moderat	5,3 (1,16)	7,9 (1,29)	6,3 (1,83)
stark	5,2 (1,40)	7,3 (1,64)	4,33 (1,22)

Tab. 5.2: Durchschnittliche Einbettungstiefe der Texte (mit Standardabweichung); Rohwerte zu Abbildung 5.4 links, sowie maximale Einbettungstiefe (ohne Abbildung).

<sup>61</sup> Die nach Holm korrigierten Signifikanzschwellen entsprechen den hier angegebenen.

<sup>62</sup> Beide Schritte stellen eine signifikante Erhöhung der Satzanzahl pro Text dar (Vergleich Original – moderat:  $t(9) = 6,71$ ;  $p < 0,001$ ; Vergleich moderat – stark:  $t(14,6) = 3,64$ ;  $p < 0,01$ , die nach Holm korrigierten Signifikanzschwellen entsprechen den hier angegebenen).

Es zeigt sich, dass die durchschnittliche Einbettungstiefe, wie vorhergesagt, stark abfällt. Dies ist für den Übergang von den Originaltexten zu den moderaten Reformulierungen ( $t(15,8) = -5,67; p < 0,001$ ) ebenso der Fall wie für den Vergleich zwischen moderaten und starken Reformulierungen ( $t(14,0) = -3,73; p < 0,01$ ). Es ergibt sich also bisher folgendes recht klares Bild für die syntaktisch komplexen Texte: Die Reformulierungen von Hansen et al. (2006) ergeben Texte mit mehr Sätzen, die kürzer sind und geringere Einbettungstiefen aufweisen als die Originalversionen. Es muss jedoch der Inhalt und die Kohärenz zwischen den Sätzen und den darin beschriebenen Sachverhalten aufrechterhalten werden. Dementsprechend müssen sprachliche Kohäsionsmittel eingesetzt werden, die über Satzgrenzen hinweg Inhalte und Personen aufeinander beziehen. Unter anderem bieten sich hierzu Demonstrativa an, denen im Stuttgart-Tübingen-Tagset (STTS) die Parts-of-Speech PDS (substituierendes Demonstrativpronomen) und PDAT (attribuierendes Demonstrativum) zugewiesen sind. Weitere Elemente, die zum Einsatz kommen können, um Beziehungen zwischen Sätzen herzustellen, sind beispielsweise Indefinitpronomen – im STTS mit den Tags PIS (substituierendes Indefinitpronomen), PIAT und PIDAT (attribuierende Indefinitpronomen ohne und mit Determiner) verzeichnet.

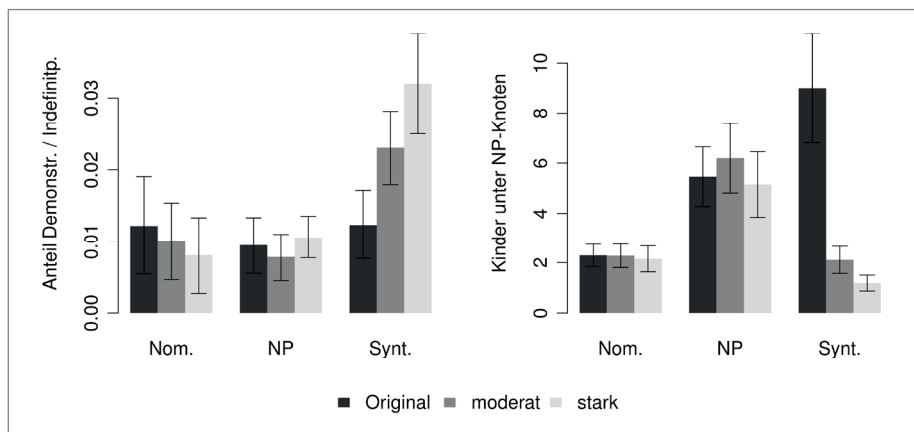


Abb. 5.5: Anteil der Demonstrativa und Indefinitpronomen (links) sowie durchschnittliche Anzahl Kinderknoten unter NPs (rechts) in Abhängigkeit des Komplexitätstyps und der Reformulierungsversion.

Abbildung 5.5 (linkes Schaubild) zeigt, dass über die Reformulierungsstufen syntaktisch komplexer Texte hinweg ein Anstieg eben jener Elemente zu verzeichnen ist – ein signifikanter Unterschied ist zwischen der Originalversion ( $\bar{x} = 1,24\%; s = 1,51$ ) und den starken Reformulierungen ( $\bar{x} = 3,20\%; s = 2,22$ )

festzustellen ( $t(15,9) = 2,31; p < 0,05$ ). Dies ist für die anderen Komplexitätstypen nicht der Fall. Beispiel 12c zeigt für einige Stimulustexte, wie die inner-textliche Kohärenz aufrechterhalten wird, obwohl die Sätze geteilt werden.

(12) a. (Originalversion)

Im Übrigen sei der Wille des Gesetzgebers zu respektieren, [das Mutter-Kind-Verhältnis vor Beeinträchtigungen zu bewahren]<sub>1</sub>, [die aus einem jederzeit möglichen Sorgerechtsübergang auf den Vater herrühren könnten]<sub>2</sub>.

b. (moderate Reformulierung)

Im Übrigen sei der Wille des Gesetzgebers zu respektieren, [das Mutter-Kind-Verhältnis vor Beeinträchtigungen zu bewahren]<sub>1</sub>. [Solche<sub>PIS</sub> könnten aus einem jederzeit möglichen Sorgerechtsübergang auf den Vater herrühren]<sub>2</sub>.

c. (starke Reformulierung)

Im Übrigen sei der Wille des Gesetzgebers zu respektieren. Dieser<sub>PDS</sub> wolle das Mutter-Kind-Verhältnis vor Beeinträchtigungen bewahren. [Solche<sub>PIS</sub> könnten aus einem jederzeit möglichen Sorgerechtsübergang auf den Vater herrühren]<sub>2</sub>.

Möchte man die erweiterte Infinitivgruppe (in Satz 12a und 12b durch 1 indiziert) in einen eigenständigen Satz umwandeln, muss über das Demonstrativpronomen der Bezug zum vorherigen Satz hergestellt werden. Ob das in diesem Fall gut gelungen ist, ist fraglich. Meist bezieht sich ein Demonstrativum – sofern keine NP in Objektposition vorhanden ist – auf das Subjekt des vorhergehenden Satzes (vgl. Diessel 1999).<sup>63</sup> Hier wäre dies 'der Wille des Gesetzgebers'. Dieser will jedoch **nicht** das Mutter-Kind-Verhältnis vor Beeinträchtigung bewahren – der Gesetzgeber möchte das. Die Bedeutung des Gesamtsatzes verändert sich hierdurch jedoch nicht bzw. minimal. Wir werden in den Abschnitten 7.4 und 7.5 unter anderem auf die Auswirkungen von Referenzen auf das Leseverhalten der VersuchsteilnehmerInnen zurückkommen. Der Relativsatz, der in Satz 12a mit 2 indiziert ist, wird ebenfalls aufgelöst und in einen eigenen Satz transformiert (in Satz 12b und 12c ebenfalls durch 2 indiziert). Dieser neue Satz beginnt mit dem substituierenden Indefinit-

<sup>63</sup> 'Subjekt- und Objektposition' trifft nicht exakt den Kern der Sache, für den augenblicklichen Kontext ist die Terminologie jedoch ausreichend. Diessel (1999) fasst die Ergebnisse zahlreicher Studien zusammen: „[A]naphoric demonstratives are often used to indicate a referent that is somewhat unexpected and not currently in the focus of attention.“ (S. 96). Oft fallen Topikalität und grammatische Funktion (Subjekt vs. Objekt) zusammen.



pronomen 'solche'. Der Relativsatz mit Verbendstellung wird entsprechend in einen eigenständigen Satz mit Verbzweitstellung umgewandelt (womit eine Auflösung des Verbclusters 'herrühren könnten' einhergeht).

Vereinfachend ausgedrückt könnte man also konstatieren, dass die syntaktisch komplexen Texte auf der Ebene des Satzes (Satzlänge und -einbettung) einfacher werden. Diese Vereinfachung geht aber zu Lasten einer höheren referenziellen Komplexität. Während also bei Texten mit vielen Nominalisierungen die Komplexität in gewisser Weise von der lexikalischen auf die syntaktische Ebene verlagert wird, findet diese Verlagerung bei den syntaktisch komplexen Texten eine Ebene höher, nämlich von der Satz- auf die Textebene, statt.

Im Lichte der eben vorgestellten Analyse soll kurz nochmals auf die Texte der Komplexitätsklasse *Nom.* eingegangen werden. Es ist etwas zu einfach, zu behaupten, dass die Komplexität von der lexikalischen auf die syntaktische Ebene verlagert wird. Zwar steigt – wie oben dargelegt – die Anzahl der Satzknotten, interessanterweise werden die Sätze im Schnitt aber weder signifikant länger, noch entstehen signifikant tiefere Einbettungen (siehe Abb. 5.4 sowie Tab. 5.2). Das liegt nicht daran, dass die vorgenommenen Umformungen keinerlei Einfluss auf diese Variablen haben. Der Umfang der Umformungen verglichen mit der Gesamtwortanzahl im Satz ist nur recht gering. Außerdem wird meist auch nicht mehr als eine zusätzliche Einbettungsebene eingeführt. Das ist eine schwache Veränderung, die sich meist nur auf wenige Wörter beschränkt – im Kontext des Gesamtsatzes tritt diese Veränderung daher statistisch nicht zu Tage, zumindest nicht bei den genannten Variablen Satzlänge und durchschnittliche Einbettungstiefe.

Texte mit komplexen Nominalphrasen (Komplexitätstyp *NP*) wurden ebenfalls in zwei Stufen reformuliert. Hier sollten die komplexen Nominalphrasen zunächst in Nebensätze ausgelagert werden. Die dadurch entstehenden subordinierten Strukturen sollten im zweiten Schritt in eigenständige Sätze ausgelagert werden. Interessanterweise sinkt bei diesem Vorgehen auch der Anteil der Nominalisierungen, obwohl dies nicht das primäre Ziel der Reformulierungen war. Während in den Originaltexten noch 9,20% ( $s = 4,36$ ) der Wörter in den Texten Nominalisierungen waren, fällt der Anteil in den moderaten Reformulierungen auf 6,65% ( $s = 2,94$ ) und schließlich auf 4,46% ( $s = 2,90$ ) für die stark reformulierten Texte (siehe Abb. 5.2). Lediglich der Unterschied zwischen den Originalen und den stark reformulierten Texten ist hierbei statistisch signifikant ( $t(15,6) = -2,87; p < 0,05$ ).

## (13) a. (Originalversion)

Die Änderungskompetenz des Bundes ist an [die *Beibehaltung* der wesentlichen Elemente der in dem fortbestehenden Bundesgesetz enthaltenen Regelung]<sub>1</sub> geknüpft.

## b. (moderate Reformulierung)

Die Änderungskompetenz des Bundes ist daran geknüpft, [die wesentlichen Elemente der Regelung *beizubehalten*, die in dem fortbestehenden Bundesgesetz enthalten sind]<sub>1</sub>.

## c. (starke Reformulierung)

[Die wesentlichen Elemente der Regelung, die in dem fortbestehenden Bundesgesetz enthalten sind, *müssen beibehalten werden*.]<sub>1</sub>  
Daran<sub>PROAV</sub> ist die Änderungskompetenz des Bundes geknüpft.

Beispiel 13 gibt einen guten Eindruck dieser Reformulierungspraxis. Die komplexe Nominalphrase (in Beispiel 13a mit 1 indiziert) wird zunächst in einen subordinierten Satz ausgelagert, dem ein Relativsatz angefügt wird, der aus der ehemaligen Prämodifikation der Regelung ‘in dem fortbestehenden Bundesgesetz enthaltenen’ besteht. Außerdem wird die Nominalisierung ‘Beibehaltung’ (kursiv hervorgehoben) umgewandelt in eine Infinitivkonstruktion. Beide Maßnahmen beschränken sich nicht nur auf dieses Beispiel, was sich in einem signifikanten Anstieg unterordnender Konjunktionen sowie an Satzknotten widerspiegelt (beides siehe Abb. 5.3 sowie Tab. 5.3).

	<b>Original</b>	<b>moderat</b>	<b>stark</b>
KOUS in %	1,07 (1,15)	2,07 (1,51)	2,89 (1,92)
Satzknotten	1,8 (0,79)	3,1 (0,88)	3,8 (0,63)

Tab. 5.3: Durchschnittlicher Anteil (und Standardabweichung in Klammern) unterordnender Konjunktionen (KOUS) sowie durchschnittliche Anzahl Satzknotten für Texte mit komplexen Nominalphrasen (Komplexitätstyp NP).

Die stark reformulierten Texte enthalten anteilig signifikant mehr unterordnende Konjunktionen als die Originalversionen ( $t(13,1) = 2,91; p < 0,05$ ), dasselbe gilt für die Anzahl der Satzknotten ( $t(17,2) = 6,26; p < 0,001$ ), bei der auch die Zunahme von Originalversion zu moderater Reformulierung signifikant ist ( $t(17,8) = 3,49; p < 0,01$ ).<sup>64</sup>

<sup>64</sup> Die nach Holm korrigierten Signifikanzschwellen entsprechen sowohl für unterordnende Konjunktionen als auch für Satzknotten den hier angegebenen.

Im Übergang von Beispiel 13b zu 13c wird der Satz geteilt und die Änderungskompetenz (im ursprünglichen Satz noch Subjekt) wird nun erst nach der Bedingung genannt, an die sie geknüpft ist. Mit der Teilung des Satzes wird auch das Pronominaladverb 'daran' eingeführt, um die Kohärenz zwischen den Sätzen aufrecht zu erhalten. Dieses Vorgehen führt zwar nicht zu solch auffälligen Ergebnissen wie bei den syntaktisch komplexen Texten, es schlägt sich jedoch statistisch messbar nieder (siehe Abb. 5.4). Während die Originalsätze noch durchschnittlich 27,5 Wörter ( $s = 9,61$ ) enthalten, steigt die Satzlänge zunächst (nicht signifikant) auf durchschnittlich 31,7 Wörter ( $s = 7,60$ ) für die moderaten Reformulierungen. Das Teilen der Sätze macht sich dann jedoch in der starken Reformulierungsversion bemerkbar: Die durchschnittliche Satzlänge sinkt deutlich (und signifikant)<sup>65</sup> auf 22,3 Wörter ( $s = 12,1$ ).

Zusammenfassend gilt für Texte des Komplexitätstyps *NP*, dass sowohl die Reformulierungsrichtlinien als auch die tatsächlichen Konsequenzen eine Mischung aus den beiden anderen Herangehensweisen (für *Nom.* und *Synt.*) darstellen. Die Anzahl der Nominalisierungen sinkt, damit einhergehend steigt der Anteil verbaler Strukturen. Außerdem werden weitere Satzknöten eingeführt, die meist zu untergeordneten Sätzen gehören (gemessen am steigenden Anteil unterordnender Konjunktionen). Weiterhin werden Sätze geteilt und damit im Schnitt kürzer – jedoch sinkt die Einbettungstiefe nicht (siehe Tab. 5.2 sowie Abb. 5.4).

NP-Komplexität lässt sich unter anderem dadurch operationalisieren, dass man die Anzahl der Kinderknöten (nicht-terminale Knöten und Wörter) unter Nominalphrasen zählt. Dieses in Abbildung 5.5 (rechts) abgetragene Maß misst relativ direkt, wie komplex eine gegebene Nominalphrase ist. Da es das primäre Ziel der Reformulierungen für die Texte mit komplexen NPs war, NP-Komplexität zu verringern, sollte zu erwarten sein, dass die Anzahl der Kinderknöten unter NP-Knöten sinkt. Erstaunlicherweise zeigt sich dieses Muster nicht für die Texte der Komplexitätsklasse *NP*, sondern nur für syntaktisch komplexe Texte. Das Argument, dass bei diesen Texten einfach mehr Potenzial zur Vereinfachung gegeben war, zählt nur teilweise. Zwar liegt die mittlere Anzahl an Kinderknöten unter NPs für die Originaltexte hier mit 9,00 ( $s = 16,5$ ) recht hoch, andererseits sind im Mittel für die Original-*NP*-Texte immerhin auch 5,46 ( $s = 8,20$ ) Knöten unter NP-Knöten eingebettet. Hier wäre also noch „Raum nach unten“ gewesen, der über die Reformulierungen jedoch offenbar nicht abgedeckt wurde.

<sup>65</sup> Für den Unterschied zwischen moderater und starker Reformulierungsversion gilt:  $t(23,0) = -2,39$ ;  $p < 0,05$ . Korrigiert man die Signifikanzschwellen aufgrund der mehrfachen Vergleiche (Original vs. moderat, Original vs. stark, moderat vs. stark) nach Holm, ist dieser Unterschied nur noch bei einem einseitigen Test signifikant.

Abschließend sollen Fernabhängigkeiten im FLRCRef-Teil beleuchtet werden. Genauer: Die Distanzen zwischen abhängigen Elementen, gemessen über die Anzahl intervenierender Wörter. Die Abhängigkeit zwischen einem Verb und seinen Argumenten ist ein wichtiger Faktor bei der Sprachverarbeitung (vgl. z.B. Baumann 2012; Gibson 2000; Konieczny 2000; Kaan 2002). In der vorliegenden Arbeit werden wir die Abhängigkeit zwischen Subjekt und Verb an mehreren Stellen aufgreifen. In Abschnitt 7.3 wird eine Analyse zu ähnlichkeitsbasierter Interferenz (Gordon/Hendrick/Johnson 2001; Gordon et al. 2006) und hinweisbasierter Satzverarbeitung (Lewis/Vasishth 2005; Van Dyke/McElree 2006) präsentiert. An entsprechender Stelle wird detailliert auf diese Theorien eingegangen.

An aktueller Stelle soll auf linguistischer Ebene die lineare Distanz zwischen einem Verb und dessen Subjekt im Zentrum des Interesses stehen. Es ist zu erwarten, dass die Reformulierungen darauf einen Einfluss haben. Für das Englische zeigte unter anderem Gibson (1998, 2000), dass höhere Distanzen mit einem höheren Verarbeitungsaufwand in Beziehung stehen (*locality effect*). Für das Deutsche jedoch zieht Konieczny (2000) diesen Zusammenhang in Zweifel. Er zeigt, dass unter bestimmten Umständen mehr intervenierendes Material zwischen Subjekt und Verb dazu führen kann, dass das Verb besser vorhersagbar wird und somit schneller verarbeitet wird (*anti-locality effect*). An entsprechender Stelle wird auf diese widersprüchlichen Ergebnisse nochmals genauer eingegangen. Wie verhalten sich die Texte im FLRCRef-Teil bezüglich dieser Variable? Um dies beantworten zu können, wurden anhand einer einfachen Heuristik Verb-Subjekt-Paare im FLRC identifiziert. Die verwendete Heuristik ist folgende: 1) Identifiziere ein finites Verb anhand der POS-Tags. 2) Identifiziere den nächsten Satzknotten, an dem dieses finite Verb hängt. 3) Identifiziere alle Knoten, die an diesem Satzknotten hängen. 4) Selektiere jenen Knoten, der mit dem Kantenlabel 'SB' für Subjekt annotiert ist. 5) Identifiziere alle Wörter, die dieses Subjekt enkodieren. Wo diese Heuristik fehlschlug, wurde die Abhängigkeit manuell nachannotiert. Das Subjekt umfasst in den allermeisten Fällen mehrere Wörter (das ist beispielsweise bei Pronomen nicht der Fall). Hier wurde entschieden, dass immer die Distanz zwischen Verb und jenem Wort zählt, das am nächsten zum Verb steht. Steht also das Verb nach dem Subjekt (was für 196 von 253 (77,5%) Verben der Fall ist), wird die Distanz bis zum letzten Wort des Subjekts berechnet. Dies ist in Beispiel 14 exemplifiziert. Die Indizes geben die Subjekt-Verb-Abhängigkeiten an (das Verb selbst zählt ebenfalls, damit keine Distanzen von 0 auftreten), hochgestellt an den Verben ist die Dependenzdistanz notiert. Abgetrennte Verbzusätze wie das 'voraus' in Beispiel 14 sind nicht beachtet. Als Ausgangspunkte wurden alle finiten Verben gewählt. Das heißt, dass im Falle einer

Konstruktion mit Hilfs- oder Modalverben das entsprechende Hilfs- oder Modalverb als Ausgangspunkt der Zählung diente. Das ist auch anhand der Passivhilfsverben in Beispiel 16 zu sehen. Außerdem wurde Beispiel 16 aufgenommen, weil es die längste Fernabhängigkeit zwischen Verb und Subjekt enthält, die im FLRC auftritt (das mit *i* indizierte 'zwingen' kurz vor Ende des Satzes).

- (14) [Die gemeinsame Ausübung der Elternverantwortung]<sub>i</sub> setzt<sub>i</sub><sup>1</sup> eine tragfähige, soziale Beziehung zwischen den Eltern voraus und erfordert<sub>i</sub><sup>1</sup> ein Mindestmaß an Übereinstimmung zwischen ihnen.
- (15) Bei der Abwägung zwischen den Belangen der Beschwerdeführerin einerseits und dem Persönlichkeitsrecht der Kläger andererseits gebühre<sub>i</sub><sup>13</sup> dem Interesse der Kläger an einer Veröffentlichung der Richtigstellung auf der Titelseite [der Vorrang]<sub>i</sub>.
- (16) [Die Ausführungen des Bayerischen Staatsministeriums der Justiz]<sub>r</sub> [der Vertrag]<sub>j</sub> gelte<sub>j</sub><sup>1</sup> als auf unbestimmte Zeit abgeschlossen, weil [das Schriftformerfordernis des Paragraphen 566 Satz 1]<sub>k</sub> in Verbindung mit Paragraph 580 BGB dadurch nicht eingehalten wurde<sub>k</sub><sup>10</sup>, dass, wie vom Kläger vorgetragen, [Renovierungs- und Instandsetzungspflichten]<sub>l</sub> vereinbart wurden<sub>i</sub><sup>2</sup>, zwingen<sub>i</sub><sup>37</sup> zu keiner anderen Beurteilung.

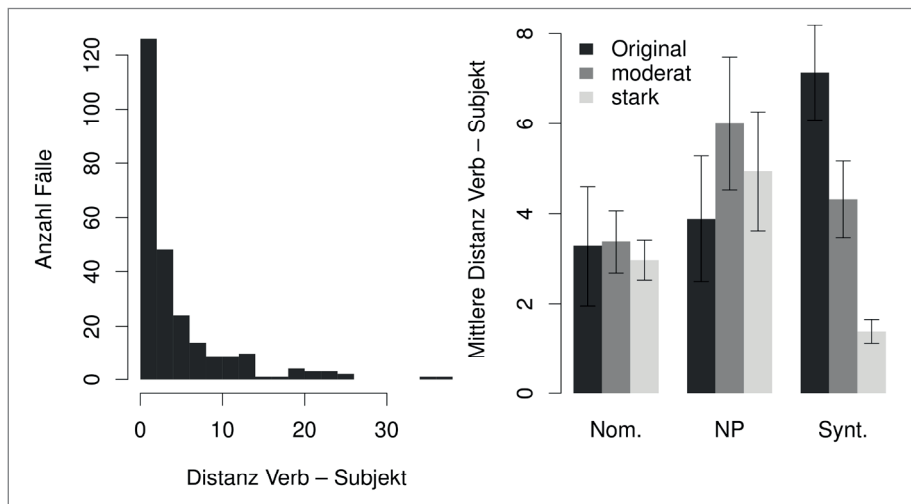


Abb. 5.6: Histogramm für die Verb-Subjekt-Distanzen (links) sowie in Abhängigkeit des Komplexitätstyps und der Reformulierungsversion (rechts).

Abbildung 5.6 gibt einen Eindruck der Verteilung der Verb-Subjekt-Distanzen. Das Histogramm auf der linken Seite zeigt, dass die überwiegende Mehrzahl der Verben nah am Subjekt steht. Fast die Hälfte aller Verben (48,6% stehen direkt neben dem zugehörigen Subjekt (Distanz = 1). Ein differenzierteres Bild ergibt sich, wenn man über die Reformulierungsversionen innerhalb der verschiedenen Komplexitätstypen aggregiert (rechte Seite von Abb. 5.6). Es zeigt sich, dass es insbesondere die syntaktisch komplexen Texte sind, innerhalb derer die Verb-Subjekt-Distanz über die Reformulierungsversionen hinweg deutlich absinkt (alle Kontraste innerhalb der rechten Balkengruppe sind hier laut *t*-Test mit Signifikanz-Korrektur nach Holm signifikant). Für die Texte mit vielen Nominalisierungen (linke Balkengruppe) ist nahezu keine Entwicklung zu beobachten. Innerhalb der Texte mit komplexen Nominalphrasen (mittlere Balkengruppe) scheint es leichte Schwankungen zu geben, die aber nicht so stark sind, als dass die Unterschiede signifikant würden. Die Teilung der Sätze in den syntaktisch komplexen Texten hat also offenbar auch dazu geführt, dass die Distanzen zwischen Verben und deren Subjekten deutlich fallen. Das ist nicht unbedingt eine Überraschung, denn wenn mehr Textmaterial in einem Satz ist, kann mehr davon zwischen Verb und Subjekt intervenieren. Das lässt sich sehr gut im Vergleich der Abbildungen 5.6 (rechts) und 5.4 (rechts) sehen: Die Satzlängen verhalten sich in einem sehr ähnlichen Muster wie die Subjekt-Verb-Distanzen. Dieser Zusammenhang wird auch durch eine signifikante Korrelation ( $r = 0,64$ ,  $p = \min p$ )<sup>66</sup> zwischen Satzlänge und durchschnittlicher Subjekt-Verb-Distanz in diesem Satz bestätigt.

Nimmt man alle Komplexitätstypen zusammen, zeigt sich keine signifikante Verschiebung der Verteilung bezüglich der relativen Position des Subjekts zum Verb. Es werden zwar über die Reformulierungsversionen hinweg immer mehr finite Verben (Original: 62, moderat: 91, stark: 100), weil es immer mehr Sätze werden. In der Gesamtheit betrachtet verändert sich aber deren relative Position zum Subjekt nicht signifikant (Original: 83,1% der Verben hinter dem Subjekt, moderat: 81,3%, original: 71%). Betrachtet man aber wiederum nur die syntaktisch komplexen Texte (also jenen Komplexitätstyp, in dem die Abhängigkeitsdistanz deutlich sinkt), zeigt sich, dass eine recht deutliche (und signifikante) Verschiebung eintritt. Während in den Originaltexten fast alle (94,6%) der Subjekte vor dem Verb stehen, sind es für die moderat reformulierten Texte noch 79,5% und für die starken Reformulierungen noch etwas mehr als die Hälfte (55,8%). Diese Verschiebung von der kanonischen Reihen-

<sup>66</sup> Ich verwende  $p = \min p$ , wenn die Irrtumswahrscheinlichkeit den kleinsten von  $R$  ausgegebenen Wert,  $p = 2,2 \times 10^{-16}$ , annimmt.

folge Subjekt vor Verb hin zur Reihenfolge Verb vor Subjekt hat nach manueller Durchsicht der Stimuli zwei Gründe: 1) In einigen Stimulustexten wurden Fragen eingeführt, um die Komplexität zu verringern. Formt man einen Deklarativsatz in eine Frage um, rückt dabei das Subjekt hinter das Verb, das die erste Stelle im Fragesatz einnimmt. 2) Außerdem wurden – wie im aktuellen Abschnitt bereits gezeigt (siehe Abb. 5.5, linke Seite) – Demonstrativa und Pronominaladverbien beim Trennen der Sätze eingeführt. Das Demonstrativum oder Pronominaladverb referiert dabei oft auf jene Entität aus dem Satz zuvor, die entweder im Akkusativ oder Dativ steht. Nur sehr selten wird auf das Subjekt des vorhergehenden Satzes referiert. Wird der neu eingeführte Satz mit dem referierenden Element eingeleitet, und steht das Verb wie üblich an zweiter Stelle, kommt meist als nächstes das Subjekt im neu eingeführten Satz. Somit steht es hinter dem Verb. Beispiel 17 illustriert beide eben genannten Effekte der Reformulierungen.

(17) a. (Originalversion)

Sie<sub>i</sub> erfüllen<sub>i</sub> diese Pflicht unter anderem dadurch, dass sie bei gegebenem Anlass Gruppen und auch politische Parteien beobachten, um feststellen zu können, ob von ihnen [eine Gefahr für die freiheitliche demokratische Grundordnung]<sub>j</sub> ausgeht<sub>i</sub>.

b. (moderate Reformulierung)

Sie<sub>i</sub> erfüllen<sub>i</sub> diese Pflicht unter anderem durch folgendes: Sie<sub>j</sub> beobachten<sub>j</sub> bei gegebenem Anlass Gruppen und auch politische Parteien. **Damit wollen<sub>k</sub> sie<sub>k</sub>** feststellen, ob von ihnen [eine Gefahr für die freiheitliche demokratische Grundordnung]<sub>l</sub> ausgeht<sub>i</sub>.

c. (starke Reformulierung)

Sie<sub>i</sub> erfüllen<sub>i</sub> diese Pflicht unter anderem durch folgendes: Sie<sub>j</sub> beobachten<sub>j</sub> bei gegebenem Anlass Gruppen und auch politische Parteien. **Damit wollen<sub>k</sub> sie<sub>k</sub>** feststellen: **Geht<sub>i</sub> von diesen** [eine Gefahr für die freiheitliche demokratische Grundordnung]<sub>l</sub> aus?

Die Bedeutung der Indizes ist parallel gehalten zu jener aus den Beispielen 14 bis 16. Die zu den eben genannten Effekten korrespondierenden Strukturen sind fett hervorgehoben. In der starken Reformulierungsversion (Satz 17c) wird eine Frage eingeführt. Bereits in der moderaten Reformulierungsversion, Satz 17b, wird ein Satz abgetrennt und über das Pronominaladverb 'damit' eingeführt. Beide Reformulierungsstrategien führen zu einer Umkehrung der Subjekt-Verb-Reihenfolge.



Dieser Stimulustext kann bezüglich der Effekte auf die Umkehrung der Verb-Subjekt-Reihenfolge als prototypisch gelten. Das einzige abweichende Muster, das dieselbe Konsequenz (nämlich eine Umkehrung der Verb-Subjekt-Reihenfolge) hat, ist die Einführung einer Aufzählung, in der Sätze mit ‘erstens’ und ‘zweitens’ eingeführt werden. Der Inhalt dieser Sätze war zuvor in subordinierten Strukturen enthalten. Durch die Kombination der Aufzählungswörter mit Hilfsverben (in dem konkreten Fall ‘müssten’) wird ebenfalls die Reihenfolge Verb – Subjekt eingeführt.

Diese letzte Analyse des sprachlichen Materials aus dem FLRCRef-Teil bezüglich Dependenzdistanzen zeigt insbesondere, welche vielfältigen Konsequenzen das Teilen von Sätzen haben kann. Während bei der Reformulierung von Nominalisierungen und Nominalphrasen keine eindeutigen Effekte zu beobachten waren, waren die Konsequenzen des Teilens von Sätzen recht ausgeprägt. Es konnte außerdem gezeigt werden, dass sich nicht nur die Distanz zwischen Subjekt und Verb ändert, sondern auch die Abfolge der beiden Elemente im Satz.

### 5.2.2 FLRCNat: Zeitungsartikel, Pressemitteilungen, Urteile

Während sich die Vergleiche im Reformulierungskorpus direkt aus dem zugrunde liegenden Design ergeben, werden die Texte aus dem Korpusteil mit kompletten zusammenhängenden Texten aufgrund ihrer Zugehörigkeit zu einer Textgattung (Zeitungsartikel vs. Pressemitteilungen vs. Urteile) verglichen. Gegebenenfalls tritt der Vergleich der verschiedenen Themengebiete hinzu.

Tabelle 5.4 gibt einen Überblick über einfache Oberflächenmaße der verwendeten Texte. Die Anzahl der Bildschirmseiten (letzte Kategorie) bezieht sich auf die Präsentation der Texte für die VersuchsteilnehmerInnen. Wo immer möglich, wurden komplette Sätze auf einer Bildschirmseite dargeboten.<sup>67</sup> Für alle Längenmaße (Sätze, Token, Bildschirmseiten) zeigt sich ein klares Kontinuum von Zeitungsartikeln über Pressemitteilungen hin zu Urteilen, die mit Abstand die längsten Texte sind. Selbstverständlich ist hier nicht mehr im Ansatz davon zu sprechen (wie es bei den verschiedenen Reformulierungsversionen noch möglich war), dass die Texte dieselbe Menge an semantischem Gehalt haben.

Die Anzahl der Types bezieht sich nicht auf einzelne Wortformen, sondern auf Wortstämme.<sup>68</sup> Für die Type-Token-Ratio (auch als lexikalische Vielfalt

<sup>67</sup> Nur bei sehr langen Sätzen war dies nicht möglich. Von insgesamt 376 dargebotenen Sätzen war dies bei 11 (2,93%) der Fall. All diese Sätze stammen aus Urteilen.

<sup>68</sup> Für das Stemming wurde das R-Paket `Rstem` (Lang 2012) verwendet.

bezeichnet) zeigt sich, dass diese mit steigender Textlänge immer geringer wird. Das ist ein bekanntes Phänomen, das unter anderem auch in Perkuhn/Keibel/Kupietz (2012, S. 88) beschrieben ist. Dort wird ein Vorgehen vorgeschlagen, das in den Grundzügen Johnson (1944) folgt: Die Type-Token-Ratio wird nicht mehr über einen kompletten Text mit  $t$  Token ermittelt, sondern für einzelne Abschnitte des Texts, die immer denselben Token-Umfang  $n$  haben (als kleinsten Umfang schlagen Perkuhn et al. (2012)  $n = 100$  Token vor). Diese Teil-Ratios werden dann gemittelt (daher auch die Bezeichnung *mean segmental type-token ratio*, MSTTR). „Übrige“ Token (falls der Rest der Division  $t/n$  ungleich 0 ist) werden üblicherweise ignoriert. Tabelle 5.4 führt auch dieses Maß auf. Hier wurde  $n = 100$  gewählt, da der kürzeste Text 111 Token zählt – wie üblich wurden Token im letzten Segment, das nicht mehr die Länge 100 erreicht, ignoriert.

	Zeitungsartikel	Pressemitteilung	Urteil
	Anzahl Sätze		
Aufenthalt	10	28	125
AWACS	14	29	124
Cannabis	7	24	125
Anzahl Token (Types)			
Aufenthalt	156 (124)	600 (293)	3.634 (970)
AWACS	234 (136)	602 (295)	2.979 (987)
Cannabis	111 (89)	570 (288)	3.883 (1.184)
Type-Token-Ratio (MSTTR)			
Aufenthalt	0,80 (0,81)	0,49 (0,77)	0,27 (0,76)
AWACS	0,58 (0,72)	0,49 (0,76)	0,33 (0,76)
Cannabis	0,80 (0,83)	0,51 (0,75)	0,31 (0,77)
Anzahl Bildschirmseiten			
Aufenthalt	4	14	81
AWACS	5	15	75
Cannabis	3	11	67

Tab. 5.4: Übersicht über einige Oberflächenmaße der neun Texte im Korpusteil mit den kompletten Texten (siehe Abb. 5.1, Markierung (2)).

Es ergibt sich sehr deutlich, dass die recht niedrige Type-Token-Ratio der Urteile (und in geringerem Maße auch die der Pressemitteilungen) ein Artefakt der Textlänge ist. Die MSTTR-Werte unterscheiden sich minimal untereinander und auch wenig von den Werten der Zeitungsartikel. Die lexikalische Vielfalt der verschiedenen Textsorten scheint also – verwendet man das geeignetere Maß für deren Bestimmung – ungefähr gleich hoch in den verschiedenen Texten zu sein.

Abbildung 5.7 zeigt, wie sich die verwendeten Texte bezüglich der Textlänge im Gesamtkontext der Pressemitteilungen und Urteile des Bundesverfassungsgerichts einordnen lassen.<sup>69</sup> Die Pressemitteilungen gehören alle drei eher den kürzeren an. Die Pressemitteilung zum Thema „unerlaubter Aufenthalt“ befindet sich auf dem 43,6. Perzentil aller Textlängen aller Pressemitteilungen, jene zum AWACS-Urteil auf dem 43,8. Perzentil. Die Pressemitteilung zum Thema „Cannabis“ ist noch etwas kürzer und liegt auf dem 40,1. Perzentil. Die Urteile befinden sich noch mittiger in der Verteilung, die von allen Urteilen aufgespannt wird (unerlaubter Aufenthalt: 53,2. Perzentil; AWACS: 44,1. Perzentil; 56,6. Perzentil). Insgesamt jedoch zeigt sich, dass keine der verwendeten Pressemitteilungen und keines der verwendeten Urteile völlig aus dem Rahmen fällt, gegeben die gängige Länge der jeweiligen Texte beim Bundesverfassungsgericht.

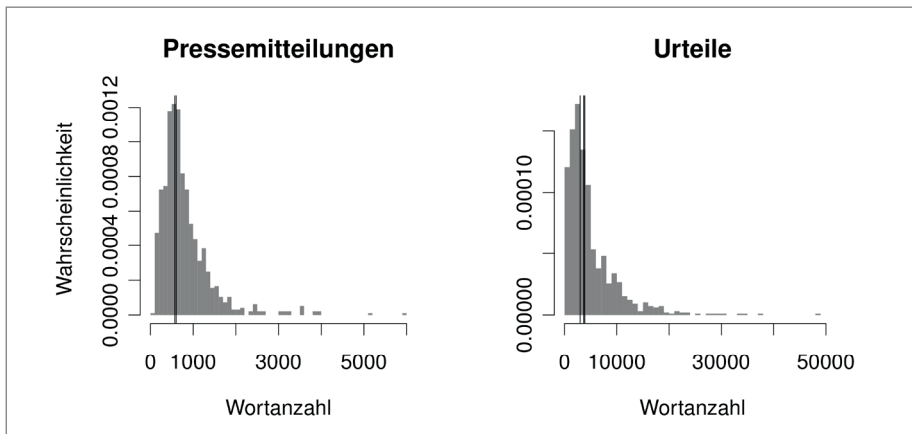


Abb. 5.7: Histogramme für die Textlängen der Pressemitteilungen und Urteile des Bundesverfassungsgerichts. Senkrechte Striche markieren die Texte im FLRC.

<sup>69</sup> Die Datengrundlage besteht aus 1.945 Pressemitteilungen und Urteilen, die bis zum 2. November 2010 unter [www.bverfg.de](http://www.bverfg.de) verfügbar waren (insgesamt ca. 5,6 Millionen Tokens). Die ersten Texte, die online verfügbar waren, sind aus dem Jahr 1998.

Die weiteren Analysen werden sich an jene des vorherigen Abschnitts anlehnen. Juristische Fachsprache ist u.a. für einen ausgeprägten Nominalstil bekannt (vgl. u.a. Fluck 1997; Sander 2004). Es wäre zu erwarten, dass sich dieser Effekt auch im Vergleich zwischen den verschiedenen Textgattungen im FLRCNat-Korpus niederschlägt. Für den FLRCRef-Teil wurden alle Nominalisierungen manuell annotiert. Aufgrund der höheren Textmenge wurden in den hier analysierten Texten lediglich \*ung-Nominalisierungen identifiziert.<sup>70</sup> Abbildung 5.8 zeigt zwei Maße, die beschreiben, wie viele Nominalisierungen in den jeweiligen Textgattungen vorkommen.

Im linken Schaubild der Abbildung 5.8 zeigt sich – wie vorhergesagt – ein Anstieg von Zeitungsartikeln zu Pressemitteilungen zu Urteilen hinsichtlich der **Anzahl** der \*ung-Nominalisierungen pro Satz. Die Sätze in den Zeitungsartikeln enthalten demnach durchschnittlich 0,55 ( $s = 0,72$ ) \*ung-Nominalisierungen. Pressemitteilungen bewegen sich mit 0,77 ( $s = 0,97$ ) \*ung-Nominalisierungen im Mittelfeld und die Urteile enthalten im Schnitt in fast jedem Satz eine \*ung-Nominalisierung ( $\bar{x}_{\text{Urteile}} = 0,96$ ;  $s = 1,13$ ).<sup>71</sup> Dieses Ergebnis entspricht zunächst den Erwartungen. Es ist jedoch weniger eindeutig als es scheint: Das eigentlich sauberere Maß zieht die Länge der Sätze in Betracht. Die Anzahl der \*ung-Nominalisierungen sollte anhand der Wortanzahl normiert werden, womit das Maß den durchschnittlichen **Anteil** der \*ung-Nominalisierungen pro Satz widerspiegelt. Das ist insofern sauberer, als dass der Tatsache Rechnung getragen wird, dass mit steigender Satzlänge die Wahrscheinlichkeit des Vorkommens einer Nominalisierung steigt. Dieser Effekt ist nicht eingeschränkt auf Nominalisierungen, sondern gilt für jedes Wort oder jede Wortart: Je länger ein Satz wird, desto größer ist die Wahrscheinlichkeit, dass ein bestimmtes Wort in diesem Satz auftaucht. Normiert man also die Anzahl der Nominalisierungen an der Satzlänge, zeigt sich keinerlei Unterschied mehr zwischen den verschiedenen Textgattungen.

Die durchschnittlichen Werte für den Anteil der \*ung-Nominalisierungen pro Satz bewegen sich hier in einem sehr kleinen Bereich um 3,5%. Hansen et al. (2006) analysieren die Verteilung der Nominalisierungen auf eine leicht andere Weise und berechnen den Anteil der \*ung-Nominalisierungen im Vergleich zur Gesamtzahl der Nomen in den Sätzen (vgl. Hansen et al. 2006, S. 23). Auch

<sup>70</sup> Die Abfrage in Form eines regulären Ausdrucks lautete  $\text{^[[:upper:]]+[[:alpha:]]*ung\$}$ . Alle Treffer wurden überprüft. Eine Analyse, die zusätzlich die Suffixe „-ion“, „-ismus“, „-heit“, „-keit“, „-ität“ und „-schaft“ umfasste (vgl. Hansen et al. 2006, S. 20) bringt keine anderen Effektmuster hervor.

<sup>71</sup> Der Unterschied zwischen Zeitungsartikeln und Urteilen ist statistisch signifikant ( $t(43,3) = 2,89$ ;  $p < 0,01$ ). Weder Zeitungsartikel noch Urteile unterscheiden sich signifikant von Pressemitteilungen.

diese Art der Analyse ergibt für die verwendeten Texte keine signifikanten Unterschiede (ohne Abbildung). Es ist zwar ein minimaler Anstieg von Zeitungsartikeln zu Pressemitteilungen festzustellen, die Varianz innerhalb der Sätze ist hier aber einfach zu groß, als dass die kleine Anzahl der Sätze (gerade in den Zeitungsartikeln) hier den Effekt vor dem Rauschen zu Tage treten lassen könnte. Ob die Urteile als prototypischste Vertreter juristischer Fachsprache tatsächlich mehr Nominalisierungen enthalten, oder ob dieser Effekt einfach daran liegt, dass diese Textgattung auch im Schnitt die längsten Sätze aufweist (siehe Abb. 5.9, rechts), lässt sich daher nicht abschließend beantworten – eine breit angelegte Korpusstudie könnte sicherlich dazu dienen, die Ergebnisse von Hansen et al. (2006) zu überprüfen. In der vorliegenden Arbeit liegt der Fokus jedoch auf anderen Fragen.

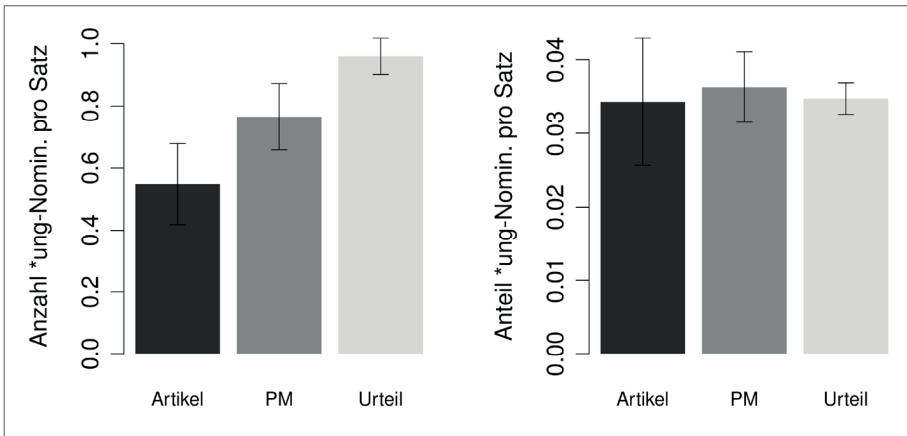


Abb. 5.8: Anzahl der \*ung-Nominalisierungen pro Satz (links) und der Anteil der \*ung-Nominalisierungen pro Satz (rechts) in Abhängigkeit der Textgattung (PM = Pressemitteilung).

Ein etwas klareres Bild ergibt sich für die durchschnittliche Einbettungstiefe sowie die eben schon angesprochene durchschnittliche Satzlänge (siehe Abbildung 5.9) innerhalb der neun Texte. Wörter in Urteilen sind im Schnitt signifikant tiefer eingebettet als Wörter in Zeitungsartikeln ( $\bar{x}_{ZA} = 3,70$ ;  $\bar{x}_U = 4,24$ ;  $t(37,3) = 3,19$ ;  $p_{\text{Holm}} < 0,05$ ), alle anderen Vergleiche ergeben keine signifikanten Unterschiede bei einer Mehrfachvergleich-Korrektur nach Holm. Wiederum befinden sich die Pressemitteilungen in Sachen Einbettungstiefe zwischen den Zeitungsartikeln und den Urteilen. Interessanterweise gilt dies nur für zwei der drei Themengebiete. Der signifikante Unterschied hinsichtlich der mittleren Einbettungstiefe ist lediglich auf die Texten zum Thema Aufenthalt und Cannabis zurückzuführen. Für das Thema AWACS hingegen ist dieser

Unterschied nicht beobachtbar. Das alternative Maß der durchschnittlichen maximalen Einbettungstiefe zeigt bei einer Korrektur nach Holm das gleiche Effektmuster, lediglich der Unterschied zwischen Zeitungsartikeln und Urteilen zeigt sich noch etwas ausgeprägter:  $\bar{x}_{ZA} = 5,16$ ;  $\bar{x}_U = 6,32$ ;  $t(39,8) = 4,03$ ;  $p_{Holm} < 0,01$ .

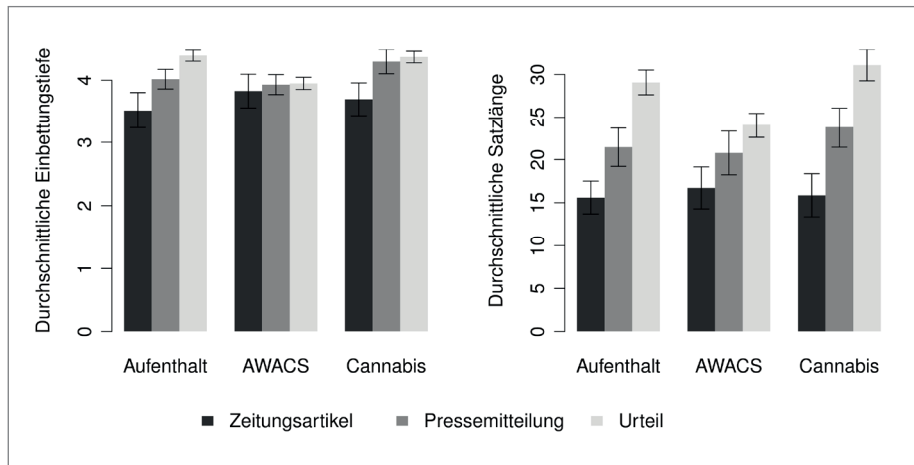


Abb. 5.9: Durchschnittliche Einbettungstiefe der Wörter (links) und durchschnittliche Länge der Sätze (rechts) in Abhängigkeit des Themas und der Textgattung.

„Immer wieder wird [...] deutschen Juristen vorgeworfen, dass Ihre Sätze zu lang und zu verschachtelt seien.“ (Heutger 2006, S. 61). Dass die verwendeten Urteile des Bundesverfassungsgerichts in der Tat verschachtelter sind als die verwendeten Zeitungsartikel, wurde soeben gezeigt. Dass die Sätze tatsächlich auch deutlich länger sind als jene in Zeitungsartikeln und Pressemitteilungen, zeigt Abbildung 5.9 (rechts). Während der durchschnittliche Zeitungsartikelsatz gerade einmal 16,2 Wörter ( $s = 7,45$ ) umfasst, enthalten Pressemitteilungen im Schnitt 21,9 ( $s = 12,3$ ) und die Urteilssätze gar 28,1 Wörter ( $s = 17,8$ ). In den verwendeten Texten fallen die Unterschiede für das AWACS-Thema wiederum am geringsten aus. Über alle Texte hinweg ist sowohl der Unterschied Zeitungsartikel vs. Urteil ( $t(64,1) = 7,32$ ;  $p < 0,001$ ) als auch Pressemitteilung vs. Urteil ( $t(163) = 3,76$ ;  $p_{Holm} < 0,01$ ) signifikant.<sup>72</sup> Hinsichtlich Satzlänge und Einbettung scheint also das erwartete Muster für die verwendeten Texte vorzuliegen: Zeitungsartikel sind weniger tief eingebettet und haben kürzere Sätze als die Pressemitteilungen des Bundesverfassungs-

<sup>72</sup> Aufgrund der Korrektur für Mehrfachvergleiche nach Holm ist der Unterschied zwischen Zeitungsartikeln und Pressemitteilungen hier nicht signifikant.

gerichts. Für diese wiederum gilt das gleiche im Vergleich zu dessen Urteilen. Abbildung 5.10 hilft uns, einen besseren Eindruck der Satzlängen innerhalb der drei Textgattungen zu erhalten. Die durchschnittliche Satzlänge in allen Pressemitteilungen und Urteilen des Bundesverfassungsgerichts zwischen 1998 und 2010 ist dort vermerkt (gestrichelte Linien in Histogrammen, Kreuze in Boxplots).

Für das Reformulierungskorpus (siehe vorheriger Abschnitt) zeigt sich recht deutlich ein Anstieg des Anteils verbaler Strukturen (operationalisiert über die Anzahl der Verben und Partizipien in Relation zur Textlänge) mit zunehmender Reformulierungsstufe der *Nom.*- und *NP*-Texte (siehe Abb. 5.2). Die analoge Hypothese für die natürlichen Texte würde höhere Anteile verbaler Strukturen für die Zeitungsartikel im Vergleich zu Pressemitteilungen und Urteilen annehmen. Dieses Bild lässt sich zumindest für die hier verwendeten Texte nicht bestätigen (siehe Abb. 5.11, links). Lediglich für das Thema Cannabis zeichnet sich die angenommene Verteilung ansatzweise ab – hier ist jedoch die Varianz für die Zeitungsartikel recht hoch. Für das Thema AWACS zeigt sich numerisch das entsprechende Muster, dies ist jedoch nicht signifikant. Auch hier gilt also, dass nur eine größer angelegte Korpusstudie eventuell in der Lage wäre, diesen Sachverhalt nachzuweisen – für die verwendeten Texte kann aber nicht postuliert werden, dass in Zeitungsartikeln anteilig signifikant mehr verbale Strukturen zu finden sind als in Pressemitteilungen oder Urteilen.

Die Anzahl Satzknoten pro Satz kann Aufschluss darüber geben, wie viele vollständige Satzkonstrukte innerhalb eines durch ein Satzendezeichen abgetrennten Satzes enthalten sind. In Abschnitt 5.2.2 (unter anderem an Beispiel 11) wurde gezeigt, wie zusätzliche Satzknoten eingeführt werden. Dies können unter- und nebengeordnete Sätze sein, die auch komplexe Sachverhalte (wie beispielsweise der Inhalt einer Überprüfung) ausdrücken können. Beispiel 18 ist der erste Satz der Pressemitteilung zum Thema „Aufenthalt“ und zeigt eine solche Konstruktion.

- (18) Wird einem Ausländer die Straftat des unerlaubten Aufenthalts im Bundesgebiet nach § 92 Abs. 1 Nr. 1 Ausländergesetz (AuslG) vorgeworfen, müssen die Strafgerichte von Verfassungs wegen selbstständig prüfen, [ob<sub>KOUS</sub> die gesetzlichen Voraussetzungen für die Erteilung einer ausländerrechtlichen Duldung im Tatzeitraum gegeben waren].



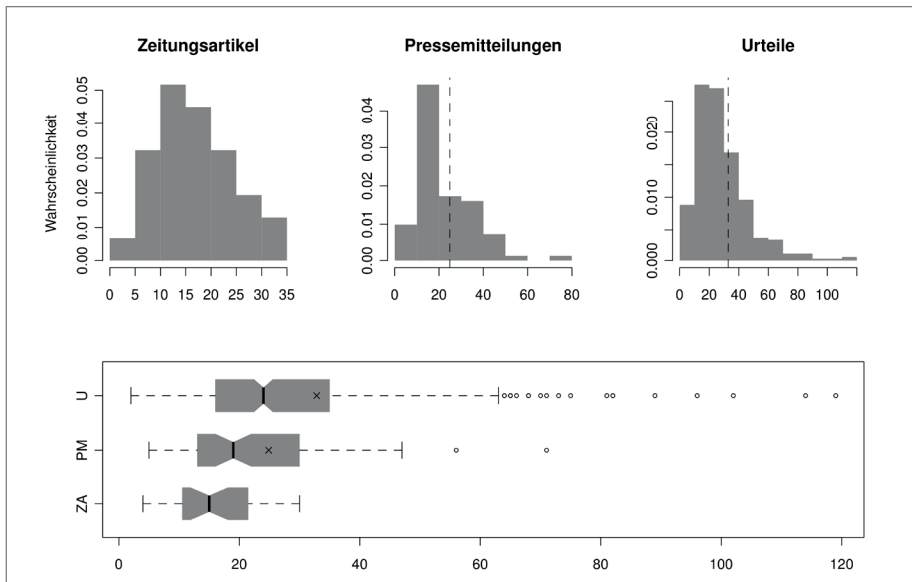


Abb. 5.10: Verteilung der Satzlängen in Wörtern in Abhängigkeit der Textgattung (Histogramme und Boxplots). Die Balkenhöhen pro Histogramm sind standardisiert und ergeben in der Summe 1. ZA = Zeitungsartikel, PM = Pressemitteilung, U = Urteil.

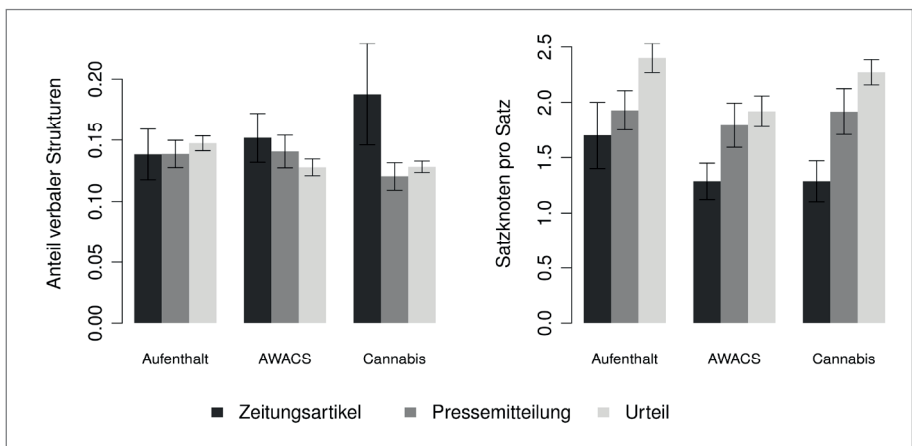


Abb. 5.11: Anteil der verbalen Strukturen pro Satz (links) sowie Anzahl Satzknöten pro Satz (rechts) in Abhängigkeit des Themas und der Textgattung.

Die unterordnende Konjunktion (KOUS) 'ob' leitet hier den Nebensatz ein, der näher spezifiziert, was die Strafgerichte selbstständig prüfen müssen. Diesem Nebensatz kommt ein eigener Satzknoten zu. Glaubt man der Einschätzung, dass Sätze in Zeitungen grundsätzlich von einfacherer Struktur als jene in Gerichtsurteilen seien, sollte sich dies auch in der Menge der Satzknöten pro Satz niederschlagen. Es zeigte sich außerdem für die Sätze aus dem Reformulierungskorpus, dass eben dieses Maß sensitiv für die Reformulierungen der syntaktisch komplexen Texte war. Für alle Themenbereiche der natürlichen Texte zeigt sich in der Tat, dass die Urteile signifikant mehr Satzknöten pro Satz enthalten als die Zeitungsartikel – wiederum bewegen sich die Pressemitteilungen dazwischen. Gemittelt über alle Themengebiete ergibt sich nach der Korrektur für Mehrfachvergleiche nach Holm ein signifikanter Unterschied ( $t(159) = 2,40$ ;  $p_{\text{Holm}} < 0,01$ ) zwischen Zeitungsartikeln ( $\bar{x} = 1,42$ ;  $s = 0,72$ ) und Urteilen ( $\bar{x} = 2,20$ ;  $s = 1,44$ ).<sup>73</sup> Offenbar handelt es sich bei den zusätzlichen Satzknöten in den Pressemitteilungen und Urteilen in der Tat mehrheitlich um eingebettete Satzknöten. Die durchschnittliche Einbettungstiefe der Satzknöten (ohne Abbildung) in Zeitungsartikeln ( $\bar{x} = 0,36$ ;  $s = 0,55$ , wobei 0 der oberste Satzknöten ist) ist geringer als jene in Pressemitteilungen ( $\bar{x} = 0,67$ ;  $s = 0,80$ ) und Urteilen ( $\bar{x} = 0,81$ ;  $s = 0,86$ ). Der Unterschied zwischen Zeitungsartikeln und den anderen Textgattungen ist dabei höchst signifikant ( $t(39,2) = 3,98$ ;  $p < 0,001$ ). Das greift einerseits das bereits erwähnte Ergebnis auf, dass Wörter in Urteilen tendenziell tiefer eingebettet sind. Es ist daher nicht verwunderlich, dass es auch tiefer eingebettete Satzknöten gibt. Andererseits zeigt es aber auch, dass es nicht bloße Koordinationen von Hauptsätzen auf der obersten Ebene sind, die bei Pressemitteilungen und Urteilen hinzukommen. Wie tief Satzknöten tatsächlich eingebettet sein können, wird an einem Satz deutlich, der aus dem Urteil zum AWACS-Einsatz stammt und 96 Wörter lang ist:

- (19) [Erginge die einstweilige Anordnung]<sub>s1</sub> und [stellte sich im Hauptsacheverfahren heraus]<sub>s2'</sub> dass die Zustimmung des Bundestages nicht erforderlich gewesen sei, müsste Deutschland entweder einem Verbündeten die im Rahmen der NATO eingeforderte Unterstützung versagen und die gemeinsam mit anderen NATO-Mitgliedern vereinbarte Hilfe für die Sicherung der territorialen Integrität der Türkei zum Scheitern bringen, [weil die AWACS-Flugzeuge nach Abzug der relativ großen Zahl der deutschen Besatzungsmitglieder ihren defensiven Auftrag nicht mehr erfüllen könnten]<sub>VVF<sub>FIN</sub>s3'</sub> oder

<sup>73</sup> Die Vergleiche Zeitungsartikel vs. Pressemitteilungen und Pressemitteilungen vs. Urteile sind bei einseitiger Testung marginal signifikant mit  $p_{\text{Holm}} < 0,06$ .

[die Bundesregierung müsste<sub>VAFIN</sub> vom Bundestag die Zustimmung zu einem militärischen Einsatz einholen]<sub>S4</sub> und [wäre<sub>VVFIN</sub> damit in dem ihr zustehenden exekutiven Handlungsspielraum unbegründet erheblich eingeschränkt worden]<sub>S5</sub>.

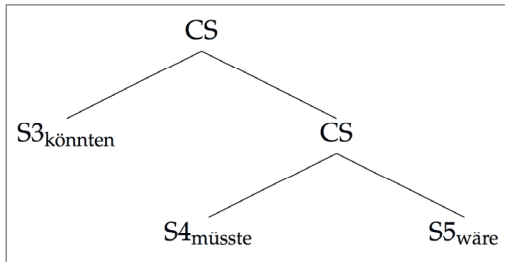


Abb. 5.12: Baumstruktur der Satzknöten S3 bis S5 aus Beispiel 19. Die Satzknöten sind mit den finiten Verben der Teilsätze indiziert.

Da der verwendete Berkeley-Parser für die Koordination zweier Sätze (wie die in Beispiel 19 mit S1 und S2 indizierten Teilsätze) CS-Knoten (*Coordinated Sentence*) immer eine Ebene über den beiden koordinierten Satzknöten annimmt, und nicht mehr als zwei Satzknöten (S bzw. CS) an einem CS-Knoten hängen dürfen, steigt hier die annotierte Einbettung recht schnell. Die Satzknöten für die Teilsätze S3 bis S5 (im Beispiel sind nicht alle Teilsätze indiziert) sind daher wie in Abbildung 5.12 verschachtelt. Zu bedenken ist, dass der in Abbildung 5.12 zuoberst abgebildete CS-Knoten selbst bereits mehrfach eingebettet ist.

Koordinierte Sätze selbst (repräsentiert über CS-Knoten) kommen in den Urteilen deutlich häufiger vor als in den beiden anderen Textgattungen ( $t(285) = 3,75$ ;  $p < 0,001$ ). Im Lichte von Beispiel 19 betrachtet ist das wenig überraschend. Es ist schlichtweg unvorstellbar, Sätze mit nahezu hundert Wörtern zu formulieren, ohne einige Satzknöten zu koordinieren.

Wie in Abschnitt 5.2.1 gezeigt wurde, variierte die Verwendung unterordnender Konjunktionen (KOUS) in Texten aller Komplexitätstypen, wenn auch teilweise in entgegengesetzte Richtungen (siehe Abb. 5.3 sowie die Ausführungen im vorherigen Kapitel). In den hier verwendeten natürlichen Texten zeigt sich zwar ein numerischer Unterschied für die Anteile unterordnender Konjunktionen ( $\bar{x}_{ZA} = 1,15\%$ ;  $s_{ZA} = 2,03$ ;  $\bar{x}_U = 1,70\%$ ;  $s_U = 2,36\%$ ). Abbildung 5.13 (links) zeigt jedoch gut auf, dass dies extrem zwischen den einzelnen Themenbereichen schwankt – keines der drei Themen zeigt die gleiche Rangordnung unter den Textgattungen. Wie bereits für die erste Analyse in diesem

Abschnitt (\*ung-Nominalisierungen) gilt auch hier, dass die **Anzahl** der unterordnenden Konjunktionen durchaus von Zeitungsartikeln zu Urteilen hin steigt (ohne Abb.). Nur wird dieses Muster stark abgeschwächt, wenn wir die **Anteile** zur Grundlage der Analyse machen.

Die hier vorgestellten Analysen der verwendeten natürlichen Texte zeigen in weiten Teilen nicht so klare Resultate wie für die Texte aus dem Reformulierungsteil. Das ist aus zwei Gründen nicht verwunderlich: Erstens wurden die Analysen größtenteils parallel gehalten. Das bedeutet, dass die Analysen, die eigentlich auf das Reformulierungskorpus zugeschnitten sind und dort auch größtenteils trennscharfe Ergebnisse liefern, auf ein anderes Korpus übertragen werden, das offenbar nicht in gleicher Maße Weise zugänglich ist. Zweitens sind die natürlichen Texte um ein Vielfaches länger als die Texte aus dem Reformulierungsteil. Dies führt offenbar zu stärkerem Rauschen im Datensatz als wenn man gezielt einzelne Sätze aus einem breiten Kontext herauslöst, um spezifische Phänomene zu untersuchen (wie dies beim Reformulierungsteil der Fall ist). Mit den Einbußen, die man hier bezüglich der internen Validität in Kauf nimmt, geht dafür ein natürlicherer Kontext einher. Die VersuchsteilnehmerInnen lesen nicht isolierte Teile von Texten, sondern komplette Texte, die in einem inhaltlichen Gesamtzusammenhang stehen.

Ein Eindruck zu den natürlichen Texten konnte dennoch gewonnen werden. Zieht man die vorgestellten Maße gemeinsam in Betracht, zeigt sich ein Komplexitätskontinuum von den Zeitungsartikeln über die Pressemitteilungen hin zu den Urteilen. Dies bestätigt allein anhand der neun hier verwendeten Texte eine breitere Untersuchung von Hansen et al. (2006), die zu dem Schluss kommen, dass „[t]he press releases’ complexity lies between that of the court decisions and the newspaper reports.“ (S. 24). Die in der vorliegenden Arbeit gezeigten Analysen gehen teilweise über jene aus Hansen et al. (2006) hinaus (dies ist umgekehrt auch der Fall). Einige Variablen zeigten das angesprochene Kontinuitätsmuster von Artikeln über Pressemitteilungen zu Urteilen. Wobei meist gilt, dass der Unterschied zwischen Artikeln und Urteilen signifikant ist und Pressemitteilungen sich ohne signifikante Unterschiede dazwischen bewegen. Diese sind im Einzelnen

- die durchschnittliche Anzahl der \*ung-Nominalisierungen pro Satz (hier ist zu bedenken, dass dies nicht für den absoluten Anteil der \*ung-Nominalisierungen gilt),
- die durchschnittliche Satzlänge in Wörtern,
- die durchschnittliche Einbettungstiefe der Wörter,
- die durchschnittliche Anzahl Satzknotten (S) pro Satz,

- die durchschnittliche Einbettungstiefe dieser Satzknotten (Vergleich Artikel vs. Pressemitteilungen/Urteile) sowie
- die durchschnittliche Anzahl koordinierender Satzknotten (CS, Vergleich Urteile vs. Zeitungsartikel/Pressemitteilungen).

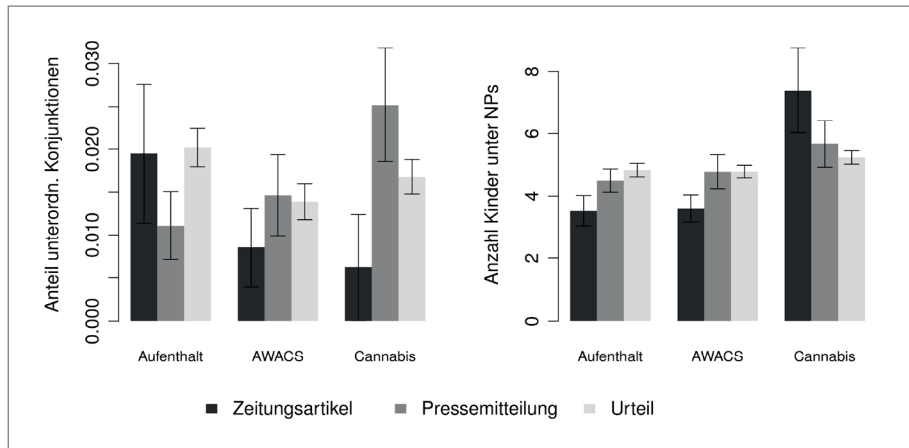


Abb. 5.13: Anzahl der unterordnenden Konjunktionen (KOUS, links) sowie Anzahl der Kinderknoten unter NPs (rechts) in Abhängigkeit des Themas und der Textgattung.

Bei einigen Variablen gruppieren sich die Pressemitteilungen eher zu den Artikeln, bei anderen eher zu den Urteilen. Auch dies bestätigt Ergebnisse von Hansen et al. (2006). Andere Variablen (wie beispielsweise der Anteil unterordnender Konjunktionen oder die Anzahl an Kinderknoten unter NP-Knoten), die sich noch sensitiv für den Übergang der Reformulierungsstufen zeigten, sind nicht trennscharf in der Unterscheidung der verwendeten Textgattungen – zumindest nicht bei der kleinen Stichprobe, die hier vorliegt. Oft werden hier mögliche Effekte überlagert von Unterschieden zwischen den verschiedenen Themenbereichen.

### 5.2.3 Referenzielle Ausdrücke

Zum Abschluss dieses Kapitels soll die im FLRC verwendete Referenzannota-tion vorgestellt werden. Dies weist zurück auf Kapitel 2.2, in dem Modelle für die Anaphernverarbeitung vorgestellt wurden. Außerdem ist ein großer Teil des Ergebniskapitels (7.4) der Analyse referenzieller Ausdrücke und deren Auswirkungen auf die Verarbeitung gewidmet. Alle referenziellen Ausdrücke im *Freiburg Legalese Reading Corpus* sind manuell annotiert. Dabei wurden die folgenden Eigenschaften festgehalten:

**Referenztyp:** Pro-Form, NP-Referenz, Relativpronomen oder Possessivpronomen.

**Referenzrichtung:** Ist der referenzielle Ausdruck eine Katapher (Ausprägung *vorw*) oder eine Anapher (*rueckw*)? In der vorliegenden Arbeit wird ein Hauptaugenmerk auf Anaphern gelegt.

**Referenzdistanz:** Anzahl der Wörter zwischen Referenz und Koreferent, numerisch angegeben. Wird mehrfach auf denselben Referenten verwiesen, wird die Distanz bis zur vorherigen Referenz in der Kette angegeben. Ausschlaggebend ist der Anfang des jeweiligen Koreferenten. Das bedeutet, dass bei sehr langen Koreferenten bei Anaphern die Distanz überschätzt werden kann. Bei Kataphern kann sie unterschätzt werden.

**Referenz über Satzgrenze:** Liegt zwischen referenziellem Ausdruck und Koreferent eine Satzgrenze (TRUE) oder nicht (FALSE)? Es wird nicht zwischen einer und mehreren Satzgrenzen unterschieden.

**Exakter Wortlaut:** Ist der Wortlaut von Referenz und Koreferent identisch (TRUE) oder nicht (FALSE)? Diese Kategorie ist nur für NP-Referenzen sinnvoll.

**Spezifitätsrichtung:** Ist die Referenz im taxonomischen Sinne spezifischer als das Referierte (*spez*), abstrakter (*abst*) oder auf derselben Ebene (*gleich*)? Diese Kategorie ist nur für NP-Referenzen sinnvoll. Ist die Spezifitätsrichtung nicht eindeutig oder kann sie nicht beurteilt werden, wird *unbek* vergeben. Wenn die Kategorie „Exakter Wortlaut“ TRUE ist, muss hier *same* vergeben werden.

**Neue Information:** Fügt die Referenz Informationen zum Referierten hinzu (TRUE) oder nicht (FALSE)? Diese Kategorie ist nur für NP-Referenzen sinnvoll. Ist die Frage nicht eindeutig beantwortbar, wird wiederum *unbek* vergeben. Wenn die Kategorie „Exakter Wortlaut“ TRUE ist, muss hier zwingend FALSE vergeben werden.

Einige Beispiele sollen die verwendeten Annotationen verdeutlichen. Beispiel 20 enthält zwei Referenzen zu Beginn (*x* und *y*), die auf vorher eingeführte Referenten referieren. Der referenzielle Ausdruck ‘Dieses Beteiligungsrecht’ bezieht sich auf den im vorherigen Satz erwähnten ‘konstitutiven Parlamentsvorbehalt’. Der Referenztyp ist hier NP, weil die komplette Nominalphrase nötig ist, um die Referenz herzustellen. Die Spezifitätsrichtung ist *abst*, da der konstitutive Parlamentsvorbehalt eines von mehreren Beteiligungsrechten und damit abstrakter ist. Es wird neue Information zum Konzept des Koreferenten hinzugefügt, weil im Text bisher nicht die Information verfügbar war, dass es sich beim konstitutiven Parlamentsvorbehalt um ein Beteiligungsrecht

handelt. Außerdem geht die Referenz über eine Satzgrenze. Die Bundeswehr wird als Parlamentsheer bezeichnet, was ebenfalls eine neue Information darstellt.

- (20) [Die Antragsstellerin]<sub>x</sub> beruft sich für [den Bundestag]<sub>y</sub> auf den [konstitutiven Parlamentsvorbehalt]<sub>i</sub>. [Dieses Beteiligungsrecht]<sub>i</sub> hat ein hohes Gewicht, weil [die Bundeswehr]<sub>j</sub> [ein Parlamentsheer]<sub>j</sub> ist.

Beispiel 21 stammt aus dem Zeitungsartikel zum Einsatz von Bundeswehrosoldaten in AWACS-Flugzeugen im Luftraum der Türkei. Es sind alle annotierten referenziellen Ausdrücke im Beispieltext markiert. Koreferenz wird über identische Indizes angezeigt. Das Verhältnis zwischen Verteidigungsminister Peter Struck, Referent *i* und damaliges Mitglied der Bundesregierung, und der Bundesregierung selbst, Referent *j*, ist komplex. Bei der Annotation mussten hier Kompromisse gemacht werden. So wird der Ausdruck 'der Bundesregierung' als abstrakte NP-Referenz auf Verteidigungsminister Peter Struck bezogen, weil Peter Struck zum damaligen Zeitpunkt eine Teilmenge der Bundesregierung bildete. Die nächste Nennung der Regierung wird dann auf die letzte Nennung bezogen, was in Beispiel 21 über den Index *j* gefasst wird. Die pronominale Referenz 'er' kurz vor Ende des Beispiels wird dann jedoch nicht auf die Bundesregierung (Referent *j*) bezogen, sondern ganz zurück auf Verteidigungsminister Peter Struck. Durch diese Art der Annotation ist sowohl dem Fakt Genüge getan, dass Peter Struck Mitglied der Bundesregierung ist, als auch dass es sonderbar anmuten würde, das 'er' auf die Bundesregierung zu beziehen. Dazwischen liegt noch eine NP-Referenz über eine Satzgrenze ('Das Gericht' → 'des Bundesverfassungsgerichts'), die als Spezifitätsrichtung *abstr* (abstrakt) annotiert wird, da das Bundesverfassungsgericht ein bestimmtes Gericht ist. Eine der Referenzen führt neue Information in den Diskurs ein. Das ist die erste Referenz 'Verteidigungsminister Peter Struck', da zu keinem Zeitpunkt vorher im Text Peter Struck als Verteidigungsminister titulierte wurde.

- (21) [Verteidigungsminister Peter Struck]<sub>i</sub> hat im Namen [der Bundesregierung]<sub>j</sub> die Entscheidung [des Bundesverfassungsgerichts]<sub>k</sub> begrüßt. [Das Gericht]<sub>k</sub> habe mit der Abweisung der Klage bestätigt, dass [die Regierung]<sub>j</sub> klar im Rahmen [ihrer]<sub>j</sub> Regierungskompetenz und [ihrer]<sub>j</sub> Bündnisverpflichtungen über den Einsatz entscheiden könne, sagte [er]<sub>i</sub> in Berlin.

Ich habe mich bewusst dagegen entschieden, die Verteilung der referenziellen Ausdrücke an dieser Stelle zu analysieren. Dies wird in Abschnitt 7.4 gesche-



hen. Durch die unmittelbare Nähe zu den entsprechenden Ergebnissen der Textverarbeitung soll es einfacher gemacht werden, die linguistische und die Verarbeitungsebene aufeinander zu beziehen.

Im folgenden Abschnitt soll das FLRC mit zwei seit Längerem bekannten Lesekorpora verglichen werden. Hinzu tritt ein recht neues Korpus, das ebenfalls zur Analyse von Blickbewegungsverhalten während des Lesens herangezogen werden kann. Bei der Erstellung der verschiedenen Korpora wurden jeweils unterschiedliche Schwerpunkte gesetzt. Dies macht es interessant, sie untereinander zu vergleichen und zu sehen, wie sich das FLRC positioniert.

### 5.3 Vergleich des FLRC mit vorhandenen Lesekorpora

Die Blickbewegungskorpora, auf denen die meisten Veröffentlichungen basieren, sind das *Dundee Corpus* (Kennedy 2003; Kennedy et al. 2003) und das *Potsdam Sentence Corpus* (PSC, Kliegl et al. 2006). Hinzu tritt ein neueres Korpus von Frank et al. (2013), das *self-paced reading*- und Blickbewegungsdaten enthält. Der Vergleich dieser vier Korpora ist auch deshalb ergiebig, weil das *Dundee Corpus* und das PSC recht unterschiedlich ausgerichtet sind: Während das *Dundee Corpus* sich in der Textauswahl bei Editorials und längeren Artikeln aus *Le Monde* und *The Independent* bedient und somit Blickbewegungen auf französischen und englischen Texten enthält, besteht das PSC aus nach bestimmten Kriterien konstruierten deutschen Einzelsätzen, die keinerlei Verbindung untereinander aufweisen.

Frank et al. (2013) orientierten sich bei der Erstellung ihres Korpus am PSC. Da ihr Korpus hauptsächlich als „gold standard“ (ebd., o.S.) für die Evaluation von psycholinguistisch-komputationalen Modellen gedacht ist, möchten Frank et al. gerade **nicht** mit zusammenhängenden Sätzen arbeiten:

Two potential problems arise when using newspaper or narrative texts for model evaluation. First, the sentence-processing models invariably treat sentences as independent entities, whereas the interpretation of a sentence in text depends on the previous sentences. Second, and perhaps more important, understanding newspaper or narrative texts requires vast amounts of extra-linguistic knowledge to which the models have no access. (Frank et al. 2013, o.S.)

Bei der Erstellung des Korpus waren sich Frank et al. (2013) außerdem darüber im Klaren, dass „the Potsdam Corpus sentences are unlikely to form a representative sample of the language, because they were manually constructed as experimental stimuli“ (o.S.). Frank et al. verwendeten daher einzelne Sätze aus frei verfügbaren Romanen. Es wurden alle Sätze selektiert, die nur Wörter enthielten, die entweder zu den 200 häufigsten englischen Wörtern

(hauptsächlich Funktionswörter) gehören oder auf der Liste der 7.754 hochfrequenten Inhaltswörter von Andrews/Vigliocco/Vinson (2009) auftauchen. Die Sätze mussten außerdem mindestens fünf Wörter lang sein und mindestens zwei Inhaltswörter enthalten. Aus diesen Sätzen wurden manuell 361 Sätze ausgesucht, die ohne Kontext „with minimal involvement of extra-linguistic knowledge“ (Frank et al. 2013, o.S.) interpretierbar sind. Sich wiederholende Eigennamen in den Sätzen wurden gegen Eigennamen desselben Geschlechts ausgetauscht, so dass jeder Name maximal zweimal auftauchte. Frank et al. wollten damit vermeiden, dass die VersuchsteilnehmerInnen versuchen, Verbindungen unter den Sätzen herzustellen. 205 dieser Sätze (alle, die in eine Bildschirmzeile passten) wurden von 43 VersuchsteilnehmerInnen gelesen, während ihre Blickbewegungen aufgezeichnet wurden. Das Lesekorpus von Frank et al. (2013) enthält für alle 361 selektierten Sätze *self paced reading center presentation*-Daten, für die Teilmenge (205 Sätze) wurden zusätzlich Blickbewegungsdaten erhoben.

Bei den Erhebungen des *Dundee Corpus* und des *Potsdam Sentence Corpus* (PSC) wurden hinsichtlich Teilnehmerzahl und Korpusumfang entgegengesetzte Ansätze verfolgt. Beim *Dundee Corpus* wurden viele Wörter von wenigen Menschen gelesen. Beim PSC wurden wenige Wörter von vielen Menschen gelesen. Bei der Erhebung des FLRC wurde ein Kompromiss zwischen diesen beiden Extremen gesucht. Ich werde die in den Texten enthaltenen Texte nicht inhaltlich vergleichen – im Falle des *Dundee Corpus* ist dies nicht möglich, da es sich um urheberrechtlich geschützte Texte handelt. Hier werden wir uns mit der Feststellung begnügen müssen, dass es sich um Zeitungstexte (so ungenau diese Bezeichnung auch sein mag) handelt.

Eine detaillierte Textgattungsanalyse für den PSC ist hinfällig, da es sich um Kunstsätze handelt, die so nie in „freier Wildbahn“ aufgetaucht sind. In Abschnitt 4.2.1 wurde im Kontext des experimentellen Kontrollansatzes kurz angesprochen, dass der Aufbau jedes Satzes im PSC einem experimentellen Design folgt. Jeder Satz enthält ein Zielwort, das aus der CELEX-Datenbank (Baayen/Piepenbrock/van Rijn 1993) aufgrund eines  $2 \times 2 \times 3$ -Designs ausgewählt wurde. Dieses Design setzt sich aus den Faktoren Wortklasse (Nomen vs. Verb), gedruckte Frequenz (hoch vs. niedrig)<sup>74</sup> und Wortlänge (kurz/mittel/lang)<sup>75</sup> zusammen. Jede Zelle des Designs ist mit zwölf Sätzen belegt, wobei die Position des Zielworts zwischen zweitem und vorletztem Wort im Satz vari-

<sup>74</sup> Hochfrequente Wörter haben eine höhere Frequenz als 50 Vorkommen pro einer Million Wörter, niedrigfrequente Wörter befinden sich im Frequenzband von 1 bis 4 Vorkommen pro einer Million Wörter.

<sup>75</sup> Als kurze Wörter wurden 3- und 4-buchstabile Wörter ausgewählt, mittellange Wörter sind 5 bis 7 Buchstaben lang, lange Wörter 8 oder 9 Buchstaben.

iert. Für eine Teilmenge von 32 Sätzen werden zwei direkt aneinander grenzende Zielwörter verwendet, wobei es sich immer um Nomen-Verb- oder Verb-Nomen-Kombinationen handelt. Dieser Teilmenge liegt ein  $2 \times 2 \times 2 \times 2$ -Design zu Grunde. Zusätzlich zu Wortklasse, Frequenz und Länge des ersten Zielwortes kommt hier der Faktor Frequenz des zweiten Zielworts hinzu. Das zweite Zielwort ist immer genauso lang wie das erste (vgl. Kliegl et al. 2004, S. 268). Um die Zielwörter herum wurden Sätze konstruiert, die diese in einen (mehr oder weniger sinnvollen) Zusammenhang stellen.

### 5.3.1 Einfache Kennzahlen

Klassische linguistische Korpora sind Textsammlungen, bei denen neben der Reichhaltigkeit der Annotation und gegebenenfalls der Ausgewogenheit der aufgenommenen Textgattungen hauptsächlich die Anzahl der Wörter (Token) als Qualitätsmerkmal herangezogen werden.

	FLRC	PSC	DunE	DunF	Frank et al.
Tokens	16.201	1.138	56.212	52.173	1.931
Types	2.755	n.v.	7.367	11.321	676
Vtn	81	222	10	10	43
Wörter/Satz	23,8	7,9	22,7	n.v.	9,4

Tab. 5.5: Anzahl Wörter (Tokens und Types), VersuchsteilnehmerInnen (Vtn) sowie die durchschnittliche Satzlänge für das FLRC, das PSC, den englischen (DunE) und französischen (DunF) Teil des *Dundee Corpus* sowie das Korpus von Frank et al. (2013).

Bei Lesekorpora kommt ein weiterer Faktor hinzu. Je mehr VersuchsteilnehmerInnen die im Lesekorpus enthaltenen Texte lesen, desto eher sind die gewonnenen Erkenntnisse auf die Grundgesamtheit der LeserInnen übertragbar. Je mehr Datenpunkte (Fixationen, Lesezeiten) vorliegen, desto eher treten auch subtile Effekte aus dem – insbesondere bei natürlichen Texten nicht unbeträchtlichen – Rauschen hervor. Tabelle 5.5<sup>76</sup> sowie Abbildung 5.14 zeigen,

<sup>76</sup> Die Typeangaben in Tabelle 5.5 sind für das FLRC, den englischen Teil des *Dundee Corpus* und das Korpus von Frank et al. (2013) selbst berechnet und basieren auf Wortstämmen (Lang 2012). Für den französischen Teil des *Dundee Corpus* wurde die Angabe aus Kennedy/Pynte (2005) übernommen. Ob es sich hierbei um stamm- oder wortformbasierte Types handelt, ist leider unklar. Es ist zu beachten, dass Abweichungen zwischen den Angaben in Frank et al. (2013) und den hier referierten Kennwerten höchstwahrscheinlich dadurch zustande kommen, dass das Eyetracking-Korpus von Frank et al. (2013) eine Teilmenge des Korpus darstellt, für das im entsprechenden Quelltext Angaben vorhanden sind.

dass das PSC und das *Dundee Corpus* als Extrempole eines Kontinuums von „wenig Wörter, viele VersuchsteilnehmerInnen“ vs. „viele Wörter, wenig VersuchsteilnehmerInnen“ gesehen werden können. Die optimale Kombinationsmöglichkeit, bei der viele Wörter von vielen VersuchsteilnehmerInnen gelesen werden, ist natürlich wünschenswert und auch theoretisch möglich. Hier hat man es jedoch mit einem pragmatischen Flaschenhals zu tun, bei dem wir für den FLRC einen Kompromiss anstreben. Der Umfang des *Dundee Corpus* mit insgesamt über 100.000 Tokens ist zweifellos imposant. Der Fakt, dass die Daten zu einer Sprache auf 10 VersuchsteilnehmerInnen beruhen, legt jedoch die Frage nahe, inwieweit persönliche Voraussetzungen eines Individuums die Daten als Ganzes beeinflussen können. Beim PSC gehen individuelle Abweichungen einzelner TeilnehmerInnen eher in der „ausgleichenden Masse“ der 221 anderen TeilnehmerInnen unter. Im PSC sind lexikalische Variablen (Wortklasse, Frequenz, Länge) im Gegensatz zum *Dundee Corpus* und dem FLRC maximal kontrolliert. Das FLRC stellt einen Kompromiss dar. Mit 80 Menschen, die die Texte gelesen haben und etwa 16.000 Wörtern befindet es sich nahezu in der Mitte der Diagonalen zwischen dem PSC und den beiden Teilen des *Dundee Corpus*.

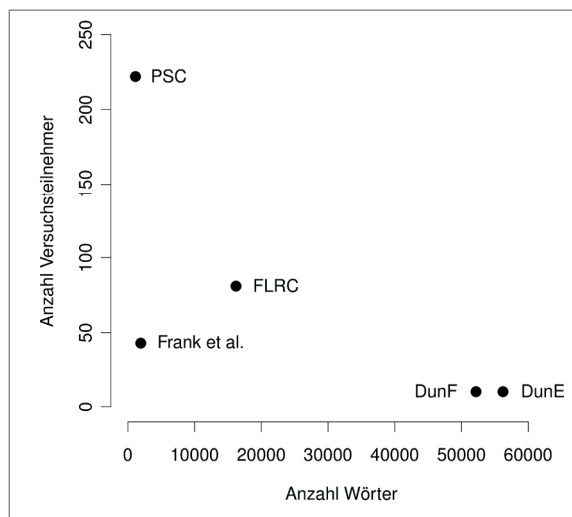


Abb. 5.14: Visualisierung der Tabelle reftab:kennzahlen. Im pragmatischen Trade-off zwischen hoher Wortanzahl und vielen VersuchsteilnehmerInnen bildet das FLRC einen Kompromiss.

Die durchschnittlichen Satz­längen des englischen Teils des *Dundee Corpus* sind vergleichbar mit jenen des FLRC, obwohl es sich um eine grundlegend andere Textgattung handelt. Man sollte annehmen, dass Zeitungstexte aus

*The Independent* unter anderem auch auf eine gute Verständlichkeit hin geschrieben werden. Dies scheint sich jedoch nicht auf die durchschnittliche Satzlänge auszuwirken, die nur circa ein Wort weniger beträgt als im gesamten FLRC. Die detaillierte Aufschlüsselung der Satzlängen im FLRC ist den Abschnitten 5.2.1 und 5.2.2 zu entnehmen. Dort wird sowohl zwischen den verschiedenen Korpusteilen des FLRC als auch zwischen verschiedenen Komplexitätstypen bzw. Themengebieten unterschieden.

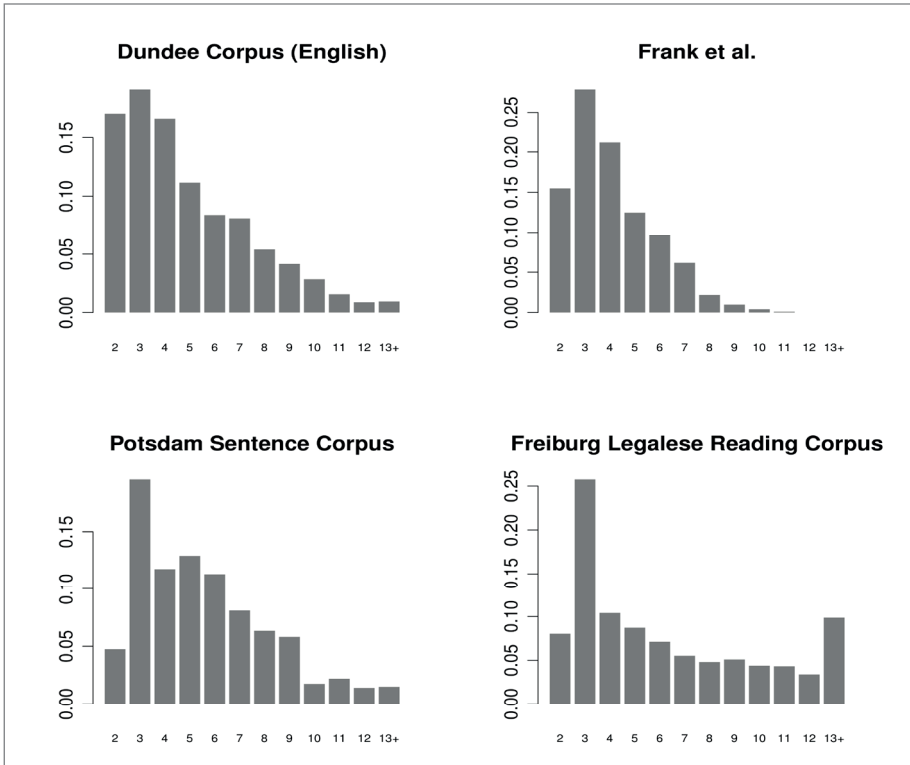


Abb. 5.15: Häufigkeitsverteilungen der Wortlängen in den Lesekorpora. Die Werte wurden an der Gesamtkorpusgröße standardisiert.

Abbildung 5.15 gibt einen Überblick über die Verteilung der verschiedenen Wortlängen in den Korpora. Für das *Dundee Corpus* konnte – aufgrund unzureichender Quellenlage – lediglich die Verteilung für den englischen Teil berechnet werden, da uns nur dieser Teil vollständig vorlag. Die Verteilung von Wortlängen ist für das PSC und beide Teile des FLRCs ähnlich. Ein bemerkenswerter Unterschied besteht im relativ hohen Anteil von Wörtern mit 13 oder mehr Buchstaben im FLRC. Dies kann wohl auf die Textgattung Rechts-

sprache zurückgeführt werden, die den Großteil des FLRCs ausmacht. Lange Komposita werden dort besonders häufig verwendet. Dies wird durch einen Blick auf die fünf häufigsten Wörter mit mehr als zwölf Buchstaben bestätigt: ‘Bundesverfassungsgericht(s)’, ‘Beschwerdeführer’, ‘Bundesregierung’, ‘Ausländerbehörde’, ‘einstweiligen’. 14 der 20 häufigsten Wörter mit mehr als zwölf Buchstaben sind Komposita. Sowohl das PSC als auch das FLRC unterscheiden sich vom englischen Teil des *Dundee Corpus* dadurch, dass eindeutig weniger zweibuchstabile Wörter im Korpus enthalten sind. Dies wird insbesondere im jeweiligen Vergleich zu den dreibuchstabigen Wörtern deutlich. Hier handelt es sich wohl um einen Spracheffekt. Im Englischen sind zweibuchstabile Wörter generell häufiger als im Deutschen, was bis ins Lesekorpus-Sampling durchschlägt. Eine Auszählung mit dem deutschen Teil des Google-Books-Korpus<sup>77</sup> (Michel et al. 2011) spiegelt die Verhältnisse zwischen zwei- und dreibuchstabigen Wörtern wider. Im deutschen Google-Books-Unigram-Korpus gilt:  $n(\text{zweibuchstabile Wörter})/n(\text{dreibuchstabile Wörter}) = 0,288$ . Für das FLRC liegt dieser Wert bei 0,31. Die analogen Statistiken für das englische Google-Books-Korpus bilden das Verhältnis von zwei- zu dreibuchstabigen Wörtern im englischen Teil des *Dundee Corpus* ab (Google Books: 0,883; *Dundee Corpus*: 0,892). Das Korpus von Frank et al. (2013) zeigt diesen Spracheffekt nur teilweise. Das entsprechende Verhältnis liegt hier mit 0,56 zwischen dem FLRC und dem englischen Teil des *Dundee Corpus*.

Betrachtet man die hier eingeführten Kennzahlen, ergibt sich folgendes Bild: Das FLRC ist dem englischen Teil des *Dundee Corpus* recht ähnlich. Unterschiede rühren entweder vom Sprachunterschied (weniger zweibuchstabile Wörter im Deutschen) oder von einem Textgattungseffekt (etwas längere Sätze und mehr sehr lange Wörter im FLRC). Im *Potsdam Sentence Corpus* und dem Korpus von Frank et al. sind die Sätze generell viel kürzer, beim PSC weniger als halb so lang wie jene im FLRC oder dem *Dundee Corpus*. Dies ist vermutlich auf die Künstlichkeit der Sätze zurückzuführen. Die Zielwörter wurden offenbar mit so wenig linguistischem Material wie möglich „verpackt“, was insgesamt zu viel kürzeren Sätzen führte, als es in natürlicher Sprache zu erwarten wäre. Das Korpus von Frank et al. (2013) enthält keine Wörter, die länger sind als elf Buchstaben – das erscheint im Lichte der anderen Korpora untypisch.

<sup>77</sup> Unter <http://storage.googleapis.com/books/ngrams/books/datasetsv2.html> (Stand: 3.6.2016) können die dem *Google Ngram Viewer* zugrunde liegenden Datensätze abgerufen werden.

### 5.3.2 Verfügbare Variablen

	FLRC	PSC	Dundee
Tokenfrequenz	Wolfer et al. (2012b)	Kliegl et al. (2004)	Kennedy/Pynte (2005)
Bigram-Frequenz	n.n.v.	Boston et al. (2008)	Baumann (2012)
Familiarität	Wolfer et al. (2012b)	–	Kennedy und Pynte (2005)
Predictability	–	Kliegl et al. (2004)	–
Einbettungstiefe	n.n.v.	n.n.v.	Pynte et al. (2008a)
Surprisal	–	Patil et al. (2009)	Baumann (2012); Demberg/Keller (2008)
Cue-Interferenz	–	Patil et al. (2009)	–
LSA-Constraint	–	–	Pynte et al. (2008a)
Referenz	Wolfer et al. (2012a)	–	–

Tab. 5.6: Die verfügbaren Prädiktoren für die Lesekorpora sowie eine stellvertretende Veröffentlichung (n.n.v. = noch nicht veröffentlicht, Prädiktor aber verfügbar).

Lesekorpora unterscheiden sich – wie klassische Korpora ebenfalls – in der Menge der sprachlichen Variablen, mit denen das Korpus annotiert ist. Bei Lesekorpora werden diese Variablen hauptsächlich zur Vorhersage des Blickbewegungsverhaltens der Versuchsteilnehme eingesetzt, weshalb in regressionsanalytischer Terminologie auch häufig von Prädiktoren anstatt von unabhängigen Variablen gesprochen wird. Tabelle 5.6 gibt einen Überblick über einige der verfügbaren Prädiktoren für die eingeführten Lesekorpora. Für das PSC und das *Dundee Corpus* folgen wir dabei der bisherigen Veröffentlichungslage. Ein Prädiktor wird dann als verfügbar angenommen, wenn mir eine Veröffentlichung bekannt ist, in dem dessen Einfluss überprüft wird. Einige Prädiktoren sind recht einfach nachträglich zu erheben (beispielsweise die Familiarität eines Worts), bei anderen ist dies schwieriger (beispielsweise bei der Vorhersagbarkeit/Predictability oder Referenzauflösung).

Einige Prädiktoren bedürfen einer kurzen Einführung. Die **Familiarität** ( $\approx$  Vertrautheit) wird definiert als „[d]ie kumulierte Frequenz aller Wörter einer



Länge, die mit dem gleichen Trigramm beginnen“ (Heister et al. 2011).<sup>78</sup> Die Längenrestriktion wird dabei manchmal aufgeweicht – so verwenden beispielsweise Kennedy/Pynte (2005) die kumulierte Frequenz aller Wörter mit demselben initialen Zeichentrigramm, die dieselbe Länge haben wie Wort  $n$  +/- 2 Buchstaben. Egal, wie Familiarität im Detail gefasst wird, sie ist ein Maß dafür, wie oft die Form eines Wortes (erfasst durch die Länge und den Anfang des Wortes) vorkommt. Im Allgemeinen wird angenommen, dass höhere Werte für Familiarität bedeuten, dass ein Wort schneller gelesen wird. In Abschnitt 7.1 stelle ich jedoch eine Analyse vor, in der dieser einfache Zusammenhang teilweise in Frage gestellt wird. Es ist darauf hinzuweisen, dass das hier verwendete Konzept der Familiarität lediglich auf Frequenzmaßen beruht. Für ein Familiaritätskonzept, das auf der subjektiven Erfahrung von Menschen beruht und mit Rating-Studien erhoben wurde, siehe „experiential familiarity“ (Gernsbacher 1984).

**Predictability** erfasst die Vorhersagbarkeit eines Wortes gegeben den vorherigen Satzkontext. Predictability wird über eine Lückentextaufgabe erhoben, in der der Satzkontext bis zu Wort  $n$  gegeben ist, und die/der VersuchsteilnehmerIn das nächste Wort einsetzen muss (vgl. Taylor 1953; Kliegl et al. 2004). Es wurde gezeigt, dass Wörter, die in diesem Sinne vorhersagbarer sind, auch schneller gelesen werden (vgl. u.a. Boston et al. 2008).

**Surprisal** ist ein Maß für die Reallokation von Ressourcen während der parallelen, probabilistischen Ambiguitätsauflösung. Dabei werden alle möglichen strukturellen Satzanalysen gemäß ihrer Konsistenz mit dem bisherigen Input in eine Rangfolge gebracht. Die Schwierigkeit zum Zeitpunkt des Wortes  $n$  wird dabei dadurch operationalisiert, wie stark diese Rangfolge von Wort  $n-1$  zu Wort  $n$  verändert werden muss, sobald Wort  $n$  gelesen wurde (vgl. Levy 2008). Hale (2001) schlug dieses Maß ebenfalls als Operationalisierung der Schwierigkeit von syntaktischer Integration vor. Ein ähnliches, aber – wie Levy (2008, S. 1128) klarstellt – nicht unbedingt übereinstimmendes Maß ist *Entropy Reduction* (Hale 2003). Hier wird die von einem Wort hervorgerufene Veränderung der Unsicherheit bezüglich der kompletten Satzanalyse erfasst. Werden durch ein Wort viele mögliche Satzfortsetzungen ausgeschlossen, ist die Reduzierung der Unsicherheit größer.

**Cue-Interferenz** ist ein Maß, das an die interferenz-basierte Theorie des *cue-based retrieval* während des Satzverstehens angelehnt ist (vgl. Lewis/Vasishth 2005; Lewis/Vasishth/Van Dyke 2006; Van Dyke 2007). Für das PSC wurde dieses Maß direkt aus den Verarbeitungsvorhersagen des ACT-R-Modells

<sup>78</sup> Die Frequenzen der Wörter „Strafe“ und „Ströme“ tragen beispielsweise zur Familiarität des Worts „Straße“ bei.

von Lewis/Vasishth (2005) abgeleitet (vgl. Patil et al. 2009). *Cue-based retrieval* wurde bereits mehrfach in Abschnitt 2 angesprochen. Die Grundidee ist, dass bei der Sprachverarbeitung ähnlichkeitsbasierte Interferenzeffekte während des Abrufs wirken. Grammatikalische Beziehungen (beispielsweise zwischen einem Verb und einem dazugehörigen Nomen) können in diesem Sinne erst hergestellt werden, wenn erforderliche Konstituenten gemäß der Abrufhinweise des grammatikalischen Kopfs abgerufen wurden. In die Abrufhinweise können Eigenschaften wie Kasus, die thematische Rolle oder semantische Eigenschaften eingehen. Je mehr Konstituenten vorliegen, die auf diese Abrufhinweise (mehr oder weniger) passen, desto länger dauert der Abruf des korrekten Konstituenten und desto länger dauert die Verarbeitung des grammatikalischen Kopfs (zu empirischen Hinweisen siehe auch Van Dyke 2007; Wolfer et al. 2013). Implementiert ist das Modell (vgl. Lewis/Vasishth 2005) bisher für grammatikalisch-syntaktische (Kasus, Genus), nicht aber für semantische Passung (beispielsweise Animiertheit).

**Der LSA-Constraint** (LSA: Latent Semantic Analysis, vgl. Landauer/Dumais 1994) ist ein Maß, das in Pynte et al. (2008a) als eine Art semantisches Pendant zu Surprisal eingeführt wird. Hier wird anhand eines zuvor berechneten LSA-Raums<sup>79</sup> bestimmt, wie ähnlich Wort  $n$  dem bisherigen Satz ist (erfasst über  $\cos(x)$ , wobei  $x$  der Winkel zwischen dem mit Wort  $n$  assoziierten Vektor und dem Vektor des Satzes bis zu Wort  $n$  ist). Pynte et al. (2008a) zeigen, dass ein höherer LSA-Constraint (Pynte nennt ihn „semantischen Constraint“) mit niedrigeren Lesezeiten assoziiert ist.

Noch sind im FLRC eher wenig Maße implementiert. Die syntaktische Annotation des Korpus bietet jedoch relativ viele Möglichkeiten, diese Informationen nach und nach zu erweitern. Was ebenfalls nicht im FLRC enthalten ist, sind Prädiktoren, die sich selbst aus dem Blickverhalten der VersuchsteilnehmerInnen ergeben. So sind sowohl im PSC als auch im *Dundee Corpus* die Amplitude der eingehenden Sakkade sowie die initiale Landeposition im Wort als Prädiktoren enthalten. Diese Maße sind insbesondere für Modelle der Blickbewegungssteuerung interessant. In Modellen wie *E-Z Reader* (Reichle et al. 1998) und SWIFT (Engbert et al. 2005) werden insbesondere kognitive Operationen der Blickbewegungssteuerung simuliert. Variablen,

<sup>79</sup> Das Konzept eines semantischen Raums basierend auf der *Latent Semantic Analysis* und wie dieser erstellt wird, kann an dieser Stelle nicht ausführlich beschrieben werden. Die Grundidee ist, dass sprachliche Strukturen wie Wörter, Teilsätze oder Sätze, die oft gemeinsam in einem Textabschnitt miteinander auftreten, eine ähnliche Bedeutung haben. Diese Bedeutung kann über eine Dimensionsreduktion statistisch approximiert werden. Ist diese approximierte Bedeutung für möglichst viele sprachliche Strukturen bekannt, können Bedeutungsunterschiede zwischen einzelnen Strukturen berechnet werden.

die die Blickbewegungssteuerung operationalisieren, sind notwendig, um diese Modelle zu trainieren und zu testen. Mit dem FLRC verfolgen wir nicht das Ziel, Modelle der Blickbewegungssteuerung zu evaluieren. Deshalb haben wir uns entschlossen, weder die ein- oder ausgehende Sakkadenamplitude, noch die initiale Landeposition innerhalb eines Wortes als Prädiktor in das Korpus aufzunehmen. Wir verlieren diese Prädiktoren somit für unsere Baseline-Modellierungen, weshalb wir bei den statistischen Analysen mit mehr Rauschen rechnen müssen.

## 5.4 Zusammenfassung

In Kapitel 5 wurde zunächst gezeigt, welche Texte in das *Freiburg Legalese Reading Corpus* eingingen und wie diese Texte beschaffen sind. Ich habe einige systematische Unterschiede zwischen den verschiedenen Komplexitätstypen und Reformulierungsstufen herausgearbeitet. Auch im FLRCNat-Teil des Korpus (natürliche, lange Texte) konnten allgemeine Unterschiede gezeigt werden – insbesondere zwischen den verschiedenen Textsorten (Zeitungartikel vs. Pressemitteilungen vs. Urteile). Im dritten Abschnitt von Kapitel 5 wurde gezeigt, dass das FLRC bezüglich der Natürlichkeit der enthaltenen Texte eher auf dem Niveau des *Dundee Corpus* liegt. Im Vergleich dazu wurde bei der Erhebung des FLRC darauf geachtet, dass jeder Text von mehr als zehn VersuchsteilnehmerInnen gelesen wurde, um eine bessere Verallgemeinerbarkeit über VersuchsteilnehmerInnen hinweg zu gewährleisten. Das relativ neue Lesekorpus von Frank et al. (2013) sowie das *Potsdam Sentence Corpus* enthält lediglich Sätze, die keine Beziehung untereinander haben. Im Falle des PSC sind diese um Zielwörter herum konstruiert, im Falle des Korpus von Frank et al. (2013) wurden an sich natürliche Sätze so ausgewählt und gegebenenfalls umkonstruiert, dass sie die Beziehung untereinander verlieren. Im FLRC wurde gezielt darauf Wert gelegt, menschliche Diskursverarbeitung zu elizieren. Insbesondere die langen, natürlichen Texte, aber auch einige Texte aus dem Reformulierungsteil sind dazu geeignet.



## 6. Verhaltensdaten

Dieses Kapitel ist das erste der beiden Kernkapitel dieser Dissertation. In diesen beiden Kapiteln werden die empirischen Ergebnisse der Lesestudien präsentiert. Dabei werden im aktuellen Kapitel die Verhaltensdaten vorgestellt. Damit sind nicht-blickbewegungs-basierte Maße wie Gesamtlesedauer (Abschnitt 6.2.1) und Antwortperformanz (6.2.2) gemeint. Eine Überleitung hin zur Analyse der Blickbewegungsdaten ist Abschnitt 6.2.3. Dort wird der Zusammenhang zwischen der Lesezeit von fragenrelevanten Abschnitten und der Antwortperformanz analysiert. Im folgenden Kapitel 7 werden dann die Blickbewegungsdaten im Detail ausgewertet. Zunächst wird im aktuellen Kapitel eine kurze Einführung in die vorwiegend verwendeten statistischen Methoden gegeben (6.1). Diese statistischen Methoden gelten gleichsam auch für Kapitel 7.

### 6.1 Statistisches Vorgehen und Datenaufbereitung

In der vorliegenden Arbeit werden an mehreren Stellen lineare gemischte Modelle berichtet (Pinheiro/Bates 2000), die mit dem R-Paket `lme4` (Bates, Maechler/Bolker 2012) berechnet wurden. Diese erlauben neben der üblichen Analyse fester Faktoren (*fixed effects*) auch die Aufnahme von Zufallsfaktoren (*random effect structure*). Diese werden hauptsächlich deshalb ins Modell aufgenommen, weil sie bestimmte Anteile der Gesamtvarianz aufklären, beispielsweise die Varianz, die generell zwischen Versuchspersonen herrscht (langsame vs. schnelle Leser). Das kann dazu führen, dass die Effekte der festen Faktoren sauberer hervortreten. Die Stufen von Faktoren, die als Zufallseffekte in gemischte Modelle eingehen, müssen immer als zufällige Ziehung aus einer größeren Menge an möglichen Stufen gesehen werden (vgl. Tagliamonte/Baayen 2012). Das eingängigste Beispiel ist hier die Aufnahme der VersuchsteilnehmerInnen als Zufallseffekt: Jene Menschen, die an einer Datenerhebung teilnehmen, sind eine Ziehung aus einer weitaus größeren Population an Menschen. Somit ist die Behandlung der VersuchsteilnehmerInnen als Zufallseffekt gerechtfertigt. Mit der Verwendung gemischter Modelle werden außerdem Probleme, die mit klassischen F1-F2-Analysen im Paradigma der Varianzanalyse (ANOVA) zusammenhängen, minimiert (vgl. Clark 1973).

Für die Modellanpassungen werden Ausschlussverfahren angewandt, bei denen zunächst alle Prädiktoren in das Modell aufgenommen werden. Im Sinne von Hüttner (1986, S. 88ff.) handelt es sich also um eine rückwärtsschreitende Elimination. Eine alternative Vorgehensweise wäre die vorwärtsschreitende In-

klusion von Prädiktoren bzw. Effekten. Bei beiden Arten der Modellanpassung kann es in bestimmten Fällen vorkommen, dass die Reihenfolge der Eliminierung bzw. Inklusion von Prädiktoren einen Einfluss auf das Ergebnismodell hat. Dies kann auch in der vorliegenden Arbeit nicht vollständig ausgeschlossen werden. Es wurde aber versucht, mehrere Reihenfolgen abzu prüfen. Es liegen keine expliziten Hypothesen zu dreifachen Interaktionen vor. Für die Anpassung der gemischten Modelle wurden daher lediglich zweifache Interaktionen aufgenommen – wo das Vorgehen abweicht, ist es explizit im Text erwähnt. Schrittweise werden dann Prädiktoren, die die Vorhersage der Korrektheit der Antwort nicht signifikant verbessern, aus dem Modell ausgeschlossen. Dabei wird folgendes Verfahren iterativ angewandt.

- 1) Berechnung des Referenzmodells  $i$  mit  $d_i$  Effekt-Freiheitsgraden<sup>80</sup> und der Faktoren-Menge  $m_i$ , die sich aus Zufallsfaktoren und festen Faktoren zusammensetzt.
- 2) Berechnung des Modells  $i+1$  mit  $d_{i+1}$  Effekt-Freiheitsgraden wobei  $d_{i+1} < d_i$  und das Modell  $i+1$  in Modell  $i$  geschachtelt ist,<sup>81</sup> das bedeutet, dass  $m_{i+1}$  eine Teilmenge von  $m_i$  ist.
- 3) Modellvergleich über *likelihood ratio test*.<sup>82</sup>
- 4) Wenn dieser Test signifikant mit  $p < 0,05$  ist, wird Modell  $i$  beibehalten, da die höhere Anzahl an Effekt-Freiheitsgraden durch eine Verbesserung des Modells gerechtfertigt ist.
- 5) Wenn der Test nicht signifikant ist, wird Modell  $i+1$  übernommen und dient im nächsten Schritt als Referenzmodell.

Beim Ausschluss wird bei der Struktur der Zufallseffekte begonnen, wobei zuerst überflüssige Korrelationsparameter<sup>83</sup> und *random slopes*<sup>84</sup> identifiziert werden, bevor Zufallseffekte ganz aus dem Modell genommen werden.

<sup>80</sup> Die Anzahl der Effekt-Freiheitsgrade ergibt sich direkt aus der Anzahl und Skalierung der aufgenommenen Prädiktoren.

<sup>81</sup> „Geschachteltes Model“ wird im Deutschen als Übersetzung für den englischen Fachterminus „nested model“ verwendet.

<sup>82</sup> In R ist dieser *likelihood ratio test* über `anova(<Modell 1>, <Modell 2>)` anzufordern. Hierbei wird die Funktion `anova()` aus dem Paket `nlme` (Pinheiro et al. 2010) aufgerufen. Dieses Vorgehen wird unter anderem auch in Pinheiro/Bates (2000) für den Vergleich zweier Modelle exemplifiziert.

<sup>83</sup> Korrelationsparameter werden über das 1+ vor dem Parameter für die *random slope* notiert. Kein Korrelationsparameter wird angenommen, wenn 0+ notiert wird.

<sup>84</sup> Eine *random slope* weist das Modell an, für die verschiedenen Stufen eines Zufallsfaktors unterschiedliche Steigungen in Abhängigkeit eines anderen Faktors zuzulassen. Ein gängiges Beispiel ist der Zufallseffekterterm `(0+trial.pos | vtn)` für die Vorhersage beispielsweise

Nachdem die Struktur der Zufallseffekte optimiert wurde, werden schrittweise feste Faktoren ausgeschlossen. Gemischte Modelle werden in der vorliegenden Arbeit immer in der folgenden Form berichtet: Der erste Block gibt die Zufallsfaktoren an, der zweite Block die festen Faktoren. Die Signifikanzindikatoren sind der Konvention entsprechend folgendermaßen abgestuft:  $p < 0,1$ ; \*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ . Die Signifikanzwerte linearer gemischter Modelle werden – wo dies möglich und nötig ist – mit MCMC-Simulationen berechnet.<sup>85</sup> Bei Modellen mit Korrelationsparameter(n) ist dies leider nicht möglich.

Wie bereits in Abschnitt 5.1.2 erwähnt, konnte beim Lesen der natürlichen Texte vor- und zurückgeblättert werden. Es wurden Bereinigungsheuristiken angewandt, um dabei Lesedurchgänge von bloßen „Blätterdurchgängen“ zu unterscheiden. Diesen Heuristiken liegt die Intuition zugrunde, dass eine Seite nur dann als gelesen gelten sollte, wenn ein gewisser „Leseaufwand“ betrieben wurde. Dieser Leseaufwand wurde über zwei nacheinander geschaltete Heuristiken operationalisiert:

**Heuristik 1:**  $x_1$  ist definiert als  $x_1 = n_{\text{fix}} / n_{\text{words}}$ , wobei  $n_{\text{fix}}$  als die Anzahl der Fixationen auf  $n_{\text{words}}$  einer Seite bei einem Durchgang und  $n_{\text{words}}$  als die Anzahl der Wörter auf dieser Seite definiert ist. Es muss gelten:  $x_1 > 0,3$ . Wenn bei einem Lesedurchgang jedes Wort ein Mal fixiert würde, würde  $x_1$  gleich 1 sein. Ebenso würde  $x_1$  gleich 1 sein, wenn die ersten Wörter auf einer Seite bei einem Lesedurchgang so viele Fixationen erhalten würden, wie Wörter auf einer Seite sind.  $x_1$  geht gegen 0 je weniger Fixationen bei einem Lesedurchgang einer Seite auftreten.  $x_1$  ist größer 1, wenn mehr Fixationen bei einem Lesedurchgang auftreten als Wörter auf der Seite sind.

**Heuristik 2:**  $x_2$  ist definiert als  $x_2 = n_{\text{fixwords}} / n_{\text{words}}$ , wobei  $n_{\text{fixwords}}$  als die Anzahl der fixierten Wörter bei einem Lesedurchgang einer Seite definiert ist. Es muss gelten:  $x_2 > 0,3$ . Das bedeutet, dass mindestens 30% der Wörter auf einer Seite fixiert werden mussten.  $x_2$  kann nicht größer 1 sein. Wenn jedes Wort auf einer Seite fixiert wird (egal wie oft), ist  $x_2$  gleich 1.  $x_2$  ist typischerweise kleiner 1, da einige Wörter beim Lesen übersprungen werden.

40 VersuchsteilnehmerInnen, die an der Erhebung des FRLCNat-Korpus teilgenommen haben, lasen insgesamt 3.895 Bildschirmseiten. Von diesen Sei-

---

weise der Reaktionszeit, durch den die Aufnahme einer *random slope* für VersuchsteilnehmerInnen in Abhängigkeit der Trial-Position im Experiment sowohl Ermüdungs- als auch Übungseffekte aufgefangen werden können.

<sup>85</sup> Hierzu wurde die Funktion `pvals.fnc()` aus dem Paket `languageR` (Baayen 2011) verwendet, die wiederum auf die Sampling-Funktion `mcmcSamp()` aus dem Paket `lme4` (Bates et al. 2012) zugreift.



tendurchgängen wurden 59 (1,5%) durch Heuristik 1 ausgeschlossen. Von den übrigen Seitendurchgängen wurden 23 (0,4%) aufgrund von Heuristik 2 ausgeschlossen. Insgesamt wurden 2,1% (82 von 3.895) der Seitendurchgänge ausgeschlossen.

In Abschnitt 7 werden statistische Modelle berichtet, anhand derer die Einflussfaktoren auf das Blickbewegungsverhalten der VersuchsteilnehmerInnen identifiziert werden (für eine genauere Besprechung einiger dieser Prädiktoren siehe 5.3.2). Die Faktoren, die dort eingehen, umfassen unter anderem Wortlänge, Wortfrequenz, Wortfamiliarität (hier operationalisiert durch kumulierte Anzahl aller Wörter gleicher Länge mit demselben initialen Zeichentrigramm) sowie die Anzahl und kumulierte Frequenz aller Levenshtein-Nachbarn. Diese Maße sind untereinander hoch korreliert.

	Wortlänge	Frequenz	Familiarität	kumFreq. Lv
Frequenz	-0,69	–		
Familiarität	-0,62	0,90	–	
kumFreq. Lv	-0,79	0,80	0,76	–
Anzahl Lv	-0,85	0,82	0,76	0,92

Tab. 6.1: Korrelationsmatrix von Wortlänge, -frequenz, -familiarität sowie der kumulierten Frequenz (kumFreq. Lv) und Anzahl aller Wörter (Anzahl Lv) mit einer Levenshtein-Distanz von 1. In der Tabelle ist der Pearson'sche Korrelationskoeffizient  $r$  abgetragen. Alle Korrelationen sind höchst signifikant.

In Tabelle 6.1 sind die entsprechenden Korrelationskoeffizienten für den kompletten FLRC abgetragen. Diese sogenannte Multikollinearität ist ein bekanntes Problem in Regressionsmodellen (vgl. Baayen 2008, S. 40). Multikollinearität bedeutet, dass zwei oder mehr Prädiktorvariablen so stark miteinander korreliert sind, dass nicht bestimmt werden kann, auf welchen dieser Prädiktoren ein Einfluss auf die vorherzusagende Variable zurückzuführen ist. Rechnerisch kann Multikollinearität dazu führen, dass die berechnete Signifikanz der Effekte nicht korrekt ist. Aus diesem Grund wird hier eine Residualisierungstechnik angewandt, mit der die Multikollinearität der Prädiktorvariablen rechnerisch drastisch reduziert wird. Diese Technik besteht aus vier Schritten (siehe auch Wolfer et al. 2012b):

- 1) Zunächst wird die aus *dlexDB* (Heister et al. 2011) extrahierte logarithmierte normalisierte Wortfrequenz anhand eines einfachen linearen Modells aus der Wortlänge vorhergesagt. Aus dieser Vorhersage werden die Residuen extrahiert und zurück in die Datendatei geschrieben.

- 2) Die ebenfalls aus *dlexDB* extrahierte Wortfamiliarität wird aus der Wortlänge und der residualisierten Wortfrequenz aus Schritt 1 vorhergesagt.<sup>86</sup>
- 3) Dasselbe Verfahren wird für die kumulierte Frequenz aller Levenshtein-Nachbarn angewandt. Diese wird aus Wortlänge, residualer Frequenz aus Schritt 1 und residualer Familiarität aus Schritt 2 vorhergesagt.
- 4) Die Residuen aus den Schritten 1, 2 und 3 werden verwendet, um die Anzahl der Levenshtein-Nachbarn vorherzusagen. Die resultierenden Residualdatenreihen werden als Prädiktoren verwendet. Die residualisierten Maße werden in den Modelltabellen über das Präfix *res* markiert (so steht beispielsweise der Variablenname *res.freq.dlex.lm* für die mit einem linearen Modell (*lm*) residualisierte (*res*) Frequenz nach *dlexDB*.

Wie oben bereits beschrieben, wird die Korrelation, die zwischen den beteiligten Variablen besteht, durch diese Technik weitestgehend aufgelöst. Eine andere Folge der Residualisierungstechnik ist, dass die Residuen der entsprechenden Datenreihen nicht mehr auf der ursprünglichen Skala interpretierbar sind. Machen wir uns nochmals bewusst, aus welchen Teilkonstrukten das Konstrukt der Familiarität eigentlich zusammengesetzt ist. Einerseits geht die Frequenz aller Wörter mit demselben Wortanfang und derselben Länge mit ein. Aber auch die Frequenz des Wortes selbst wird dessen Familiarität zugerechnet – was der Hauptgrund für die hohe Korrelation zwischen Wortfrequenz und Wortfamiliarität ist. Bereinigt man nun die Frequenz des Wortes selbst, bleibt nurmehr der Einfluss der Wortfrequenzen aller gleich langen Wörter mit demselben Anfang zurück. Die residuale Familiarität fasst also nicht mehr die **Vertrautheit** mit einer bestimmten Wortform, sondern eher die **Verwechselbarkeit** des Aussehens. Die damit verbundene Verarbeitungshypothese ändert also ihre Richtung: Die Hypothese für die rohe Familiarität besagt, dass vertrautere Wörter schneller gelesen werden sollten, weil man öfter auf dieses „Wortaussehen“ trifft. Ein höherer Wert für die **residuale** Familiarität sollte aber eher mit längeren Lesezeiten assoziiert sein, weil mehr Abrufkandidaten zur Verfügung stehen, die mit dem korrekten Wort um den Abruf konkurrieren. Dieser Hypothese wird in Abschnitt 7.1 nachgegangen.

Die beschriebene Residualisierungstechnik ist in gewisser Hinsicht ein konservatives Vorgehen. Der Grund dafür liegt in der Reihenfolge der Residualisierungsvorgänge. Wenn zwei Prädiktoren hoch korreliert sind, klären diese im Allgemeinen auch gemeinsame Varianz der Kriteriumsvariable auf. Das

---

<sup>86</sup> Man könnte die Familiarität auch direkt aus der nicht-residualisierten Wortfrequenz vorherzusagen. Die Kollinearität zwischen Länge und Frequenz ist irrelevant, da der Signifikanzwert nicht interpretiert wird. Auf die Residuen schlägt sich die Kollinearität nicht nieder.

wird deutlich, wenn wir nochmals das Beispiel Wortlänge und Frequenz heranziehen. Ein hypothetisches Verhältnis der unterschiedlichen Varianzaufklärungen ist in Abbildung 6.1 dargestellt. Wird nun ein Wort langsamer gelesen als andere, liegt das sowohl daran, dass es länger ist, als auch daran, dass es seltener ist.

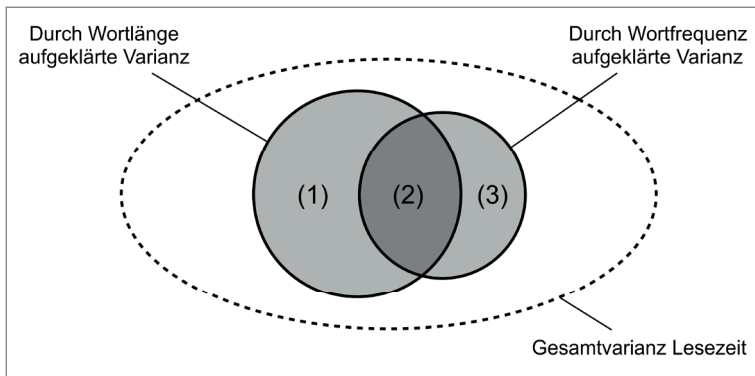


Abb. 6.1: Schematisierung der Varianzaufklärung von Wortlänge (1), Wortfrequenz (3) sowie gemeinsam aufgeklärte Varianz (2) der Lesezeit.

Diese gemeinsam aufgeklärte Varianz ist in Abbildung 6.1 durch (2) gekennzeichnet. Residualisiert man nun aber die Frequenz nach der Wortlänge, wird die Varianz in der Lesezeit, die von beiden Faktoren gemeinsam aufgeklärt wird, dem Faktor Wortlänge zugeschlagen. Folgt man Abbildung 6.1, werden die mit (1) und (2) bezeichneten Bereiche der Wortlänge zugeschlagen. Natürlich kann noch immer ein Einfluss der Frequenz nachgewiesen werden. Aber eben nur dann, wenn die Wortfrequenz über die gemeinsam mit der Länge aufgeklärte Varianz hinweg einen eigenständigen Einfluss auf die Lesezeit hat (Bereich (3) in Abb. 6.1). Bildhaft gesprochen beraubt man den Faktor Wortfrequenz eines Teils der Möglichkeiten, Varianz der Lesezeit aufzuklären, weil man die gemeinsam mit der Wortlänge aufgeklärte Varianz (2) „herausschneidet“ und vollständig dem Faktor Wortlänge zubilligt.

Um die Residualisierungstechnik noch einmal zusammenzufassen: Sie wird notwendig, weil einige Prädiktoren der Lesezeit stark untereinander korreliert sind. Dies führt dazu, dass die Einflüsse auf die abhängigen Variablen rechnerisch nicht klar getrennt werden können. Die Residualisierungstechnik separiert in oben beschriebener Weise die Beiträge zur Varianzaufklärung durch eine Dekorrelation der Prädiktorvariablen. Dadurch werden die Einflüsse der beteiligten Prädiktoren leichter unterscheidbar.

## 6.2 Analyse der Verhaltensdaten

Im ersten Analysekapitel werden Ergebnisse berichtet, die nicht direkt mit der Messung der Blickbewegungen in Beziehung stehen. Dies ist unter anderem deshalb interessant, weil diese Daten direkt mit jenen von Hansen et al. (2006) verglichen werden können. Hansen et al. (2006) erhoben für die Texte aus dem FLRCRef-Korpus (inklusive der zugrunde liegenden Originaltexte) Lesezeiten im kumulativen *self-paced reading*-Paradigma.<sup>87</sup>

Ein Problem dieses Vorgehens ist, dass die VersuchsteilnehmerInnen die Möglichkeit haben, einfach so oft das nächste Wort einzublenden (hier über einen Mausklick) bis der komplette Text auf dem Bildschirm ist und dann erst anfangen, den Text zu lesen. Die Gesamtlesezeit ist dann noch immer einigermassen aussagekräftig, nicht aber die Lesezeit der einzelnen Wörter. Auch Liversedge et al. (1998) weisen auf die Vorteile hin, die die Messung von Blickbewegungen gegenüber dem *self-paced reading*-Paradigma hat:

One of the major advantages of eye movement methodology over alternative reading time measures (e.g. self-paced reading) is that the experimenter can separate those fixations that were made when a region was first read from fixations that were made later in the eye movement record. This is important as it allows experimenters to determine when a characteristic of the text first influenced processing and therefore permits them to make inferences about the time course of processing and during text comprehension.

Die Antwortperformanz der VersuchsteilnehmerInnen bezüglich der Ja/Nein-Verständnisfragen für den FLRCRef-Teil ist direkt vergleichbar mit der berichteten Antwortperformanz aus Hansen et al. (2006).

### 6.2.1 Lesedauer

Die rohe Lesedauer der Texte hängt stark von der Menge des präsentierten Materials ab. Da dieser Zusammenhang typischerweise nicht interessiert, können Residuen berechnet werden. Hierzu wird anhand einer linearen Regression die Lesezeit aus der Anzahl Wörter oder Zeichen vorhersagt. Die Residuen dienen dann als eigentliche abhängige Variable für die weiteren Analysen. Dieses zweistufige Vorgehen kann in ein einstufiges umgewandelt

---

<sup>87</sup> Das *self-paced reading*-Paradigma kann in verschiedenen Varianten eingesetzt werden. Die typischste Variante ist ein *self-paced reading moving-window*-Paradigma. Hierbei verschwindet ein Wort (oder eine Phrase) sobald das/die nächste eingeblendet wird, es bewegt sich also ein Fenster über den Text, das immer nur das aktuelle Wort (oder die aktuelle Phrase) zeigt. Im kumulativen *self-paced reading*-Paradigma erscheinen die Wörter bei Tastendruck auf dem Bildschirm, die vorherigen werden aber nicht ausgeblendet.

werden, indem man die Anzahl der Wörter oder Zeichen als Prädiktor in das eigentliche Modell aufnimmt. Die Herauspriorisierung des Effekts der Materialmenge geschieht in diesem Fall modellintern. Der Effekt der Materialmenge ist einfach einer von mehreren, die im Modell berücksichtigt werden.

### Originalauszüge und Reformulierungen

Hansen et al. (2006) berichten für die Komplettlesezeit pro Seite einen marginal signifikanten Effekt der Reformulierungsversion.<sup>88</sup> Der entsprechende Plot (vgl. Hansen et al. 2006, S. 34) legt nahe, dass dieser Effekt lediglich von der starken Reformulierungsversion stammt. Dies ist im Text so beschrieben, jedoch ohne Angabe von Signifikanzschwellen. Die vorliegende Arbeit kann dieses Ergebnis nicht bestätigen. Tabelle 6.2 gibt einen ersten Überblick über die untransformierten, ausreißerkorrigierten Lesezeiten. Das in Tabelle 6.3 gezeigte Modell ist das Ergebnis der in Abschnitt 6.1 beschriebenen Anpassungsprozedur für logarithmierte Gesamtlesedauern auf den Texten. Die logarithmierten Lesezeiten wurden einer Outlier-Korrektur unterzogen.<sup>89</sup> Zwölf Datenpunkte (1% der Fälle) wurden ausgeschlossen. Insgesamt gehen 1.187 Datenpunkte in die Analyse ein.

		Komplexitätstyp		
		Nom.	NP	Synt.
Version	Original	9,5 s (7,6)	12,6 s (7,9)	17,1 s (11,5)
	moderat	8,6 s (7,7)	13,4 s (7,2)	14,1 s (9,4)
	stark	8,3 s (6,1)	14,8 s (9,3)	17,4 s (10,5)

Tab. 6.2: Mittelwerte (Standardabweichungen) für die Gesamtlesedauer der Texte nach Version und Komplexitätstyp.

Zunächst ist anzumerken, dass der Faktor *version* und die Interaktion *version* × *typ* im Modell bleiben, obwohl keiner der beteiligten Kontraste signifikant ist. Trotzdem wird die Passung des Modells signifikant schlechter, wenn man die Faktoren aus dem statistischen Modell entfernt. Der Komplexitätstyp (*typ*) schlägt sich als signifikanter Haupteffekt mit einer Irrtumswahrscheinlichkeit von  $p_{\text{MCMC}} = 0,003$  im Modell nieder.

<sup>88</sup> Es wird dabei lediglich der Signifikanzwert für den Omnibustest über alle drei Stufen hinweg berichtet, dieser liegt bei  $p = 0,053$ .

<sup>89</sup> Outlier wurden über Boxplots, gruppiert nach VersuchsteilnehmerInnen, identifiziert. Alle Datenpunkte, die außerhalb des doppelten Interquartilabstands, angelegt am oberen bzw. unteren Rand der Box, lagen, wurden ausgeschlossen.

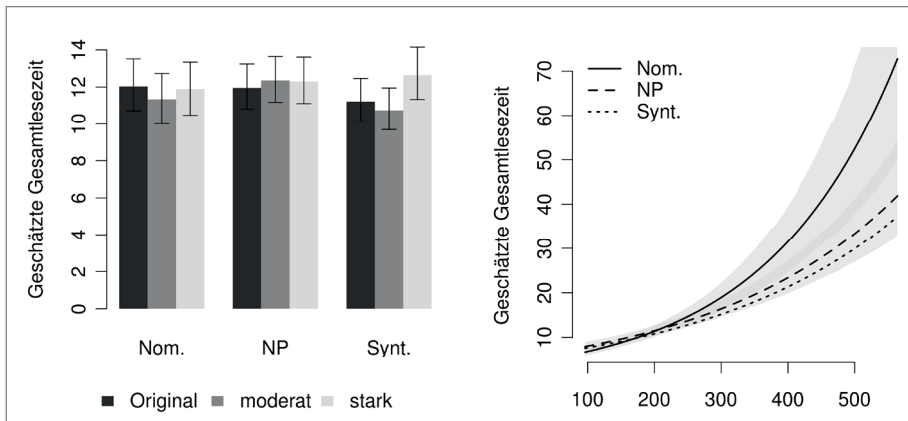


Abb. 6.2: Geschätzte mittlere Textlesezeiten in Sekunden sowie 95%-Konfidenzintervalle für die Texte aus dem Reformulierungskorpus. Die Effektschätzer aus dem entsprechenden gemischten Modell (Tabelle 6.3) sind zurücktransformiert auf eine Sekundenskala. Die Interaktion *version* × *typ* ist links abgetragen, die Interaktion *n.chars* × *typ* rechts. Die Konfidenzintervalle für *typ* gleich *Synt.* sind nicht abgetragen.

Sowohl Texte mit komplexen Nominalphrasen als auch Texte mit komplexer Syntax werden langsamer gelesen als jene mit vielen Nominalisierungen (dieser Komplexitätstyp ist die Referenzstufe im Modell und taucht daher nicht in Tabelle 6.3 auf). Bei der Interpretation dieser Effekte ist Vorsicht geboten, da die Texte der Komplexitätstypen *NP* und *Synt.* gleichzeitig die längsten Texte sind. Es ist also davon auszugehen, dass der große Anteil dieser Kontraste durch den Effekt von *n.chars* (Anzahl der Buchstaben) erklärt wird.<sup>90</sup> Die Anzahl der Wörter im Stimulustext (*res.n.words*) hat über die Zeichenanzahl hinaus<sup>91</sup> einen Effekt auf die Gesamtlesezeit der Texte. Auch der aus der Lesbarkeitsformel von Dickes/Steiwer (1977) abgeleitete Index (Prädiktor *res.DS.index*) hat den vorhergesagten Effekt auf die Gesamtlesezeit des Texts.<sup>92</sup> Je niedriger der residuale Dickes-Steiwer-Index, desto einfacher der Text (so die Effektrichtung nach Dickes und Steiwer) und desto niedriger die Lesezeit. Die Position des Versuchsdurchgangs im Gesamtexperiment *trial.pos* hat einen beschleunig-

<sup>90</sup> Im Laufe der Arbeit werden an mehreren Stellen Residualisierungstechniken angewandt, die solche Kollinearitäten auflösen können. Im vorliegenden Fall gestaltet sich das leider schwierig, da die vorherzusagende Variable *typ* dreistufig nominalskaliert ist.

<sup>91</sup> Die Anzahl der Wörter wurde anhand eines einfachen linearen Modells aus der Anzahl der Buchstaben vorhergesagt, um der natürlichen Kollinearität der Variablen entgegen zu wirken. Die Residuen dieser Vorhersage gehen als *res.n.words* ins Modell ein.

<sup>92</sup> Der Dickes-Steiwer-Index wurde anhand eines einfachen linearen Modells aus der Anzahl der Wort- und Zeichenanzahl vorhergesagt, da diese Variablen mittelbar in die Berechnung des Indizes eingehen und die Variablen daher kollinear sind.

genden Effekt auf die Lesezeit. Es kann nicht gezeigt werden, dass dieser Effekt über verschiedene VersuchsteilnehmerInnen variiert, denn die entsprechende *random slope* wurde aus dem Modell ausgeschlossen. Auch die Anwesenheit eines Kontext stiftenden Satzes vor dem eigentlichen Stimulussatz (*kontext*) führt zu einer signifikanten Beschleunigung beim Lesen.

Groups	Name	Variance	Std.Dev.			
vtn	(Intercept)	0.12269	0.35027			
Residual		0.10929	0.33059			
		Estimate	Std. Error	t value	$P_{\text{MCMC}}$	
(Intercept)		1.5347101	0.1002072	15.315	0.0001	***
version moderat		-0.0588167	0.0411791	-1.428	0.1488	
version stark		-0.0119029	0.0415442	-0.287	0.7602	
typ NP		0.3200774	0.1006356	3.181	0.0012	**
typ Synt		0.2885664	0.0930220	3.102	0.0020	**
n.chars		0.0050972	0.0004453	11.447	0.0001	***
res.n.words		0.0075169	0.0020844	3.606	0.0004	***
res.DS.index		-0.0089350	0.0022813	-3.917	0.0001	***
trial.pos		-0.0082591	0.0011255	-7.338	0.0001	***
kontext ja		-0.1433588	0.0244378	-5.866	0.0001	***
version moderat:typ np		0.0922733	0.0577945	1.597	0.1094	
version stark:typ np		0.0406402	0.0588730	0.690	0.4876	
version moderat:typ synt		0.0159634	0.0591870	0.270	0.7918	
version stark:typ synt		0.1311218	0.0622833	2.105	0.0376	*
typ np:n.chars		-0.0015477	0.0005141	-3.010	0.0048	**
typ synt:n.chars		-0.0017001	0.0004658	-3.650	0.0002	***

Tab. 6.3: Angepasstes gemischtes Modell für die logarithmierte Lesedauer der Texte im Reformulierungsteil des FLRC.

Die beiden signifikanten Interaktionen des angepassten Modells aus Tabelle 6.3 sind in Schaubild 6.2 abgetragen. Die Interaktion zwischen Reformulierungsversion (*version*) und Komplexitätstyp (*typ*) wird mit  $p_{\text{MCMC}} = 0,032$  signifikant. Dieser Effekt wird offenbar hauptsächlich von der starken Reformulierungsversion für syntaktisch komplexe Texte (in Abb. 6.2 der Balken ganz rechts) gespeist. Diese Texte werden im Gegensatz zu den anderen Reformulierungsstufen länger gelesen als die Originaltexte und die moderaten Reformulierungen. Dieses Effektmuster gilt aber nur für die syntaktisch komplexen Texte. Die Effektrichtung ist interessant, deutet sie doch darauf hin, dass die stark reformulierten Texte besonders aufwändig zu lesen sind. Der Komplexitätstyp der Texte ist an einer zweiten Interaktion beteiligt (*n.chars*  $\times$  *typ*:  $p_{\text{MCMC}} = 0,002$ ). Diese ist in Abbildung 6.2 (rechts) visualisiert. Die Länge der Texte hat auf den Komplexitätstyp *Nom.* eine stärkere Auswirkung als auf



Texte der anderen beiden Komplexitätstypen, bei denen sich der Effekt der Textlänge auf vergleichbarem Niveau bewegt. Diese Interaktion könnte daher stammen, dass sich der Effekt der Textlänge nicht über alle Zeichenanzahlen hinweg gleich verteilt. Sollte es der Fall sein, dass die Textlänge bei kurzen Texten einen größeren Einfluss hat als bei längeren Texten, könnte eine solche Interaktion zustande kommen, da nur für Texte der Komplexitätstypen *NP* und *Synt.* längere Texte vorhanden sind, die den (linearen) Gesamtschätzer „nach unten ziehen“. Eine Folgeanalyse legt in der Tat eine solche Interpretation nahe. Hierfür werden die Texte zunächst so aufgeteilt, dass sich gemäß der Textlänge drei gleich große Gruppen ergeben (Aufspaltung der Datenmenge am 33. und 66. Perzentil). Für jede der drei Textgruppen wird ein gemischtes Modell mit der Textlänge als Prädiktor der Gesamtlesezeit angepasst. Die Effektschätzer für die drei Modelle zeigen recht deutlich, dass der Effekt der Textlänge im Drittel der längsten Texte (227 bis 564 Zeichen) deutlich abfällt (erstes Drittel:  $\beta = 0,0073$ ; zweites Drittel:  $\beta = 0,0075$ ; drittes Drittel:  $\beta = 0,0027$ ; die Standardfehler der Schätzer sind sehr gering und bewegen sich ungefähr in der Größenordnung eines Zehntels der jeweiligen Effektschätzer). Der Längeneffekt ist also für jede der drei Gruppen signifikant, die Effektgröße sinkt aber deutlich für die längsten Texte.

### Natürliche, komplette Texte

Die untransformierten Rohwerte der Gesamtlesezeit der Texte aus dem FLRCNat-Korpus sind in Tabelle 6.4 abgetragen. Die Datenbasis enthält 122 „Fälle“ (40 VersuchsteilnehmerInnen mal 3 plus 2 Lesedurchgänge, die ein/e VersuchsteilnehmerIn ausnahmsweise zusätzlich absolviert hat). Jede Gesamtlesedauer wird logarithmiert und aus Buchstaben- und Wortanzahl anhand eines einfachen linearen Modells vorhergesagt.

Die Residuen dieser Vorhersage, also die von der Textlänge bereinigte logarithmierte Gesamtlesezeit (in Abb. 6.3 auf der rechten Seite), werden im Weiteren als Kriteriumsvariable verwendet. Mir ist bewusst, dass ein solches Vorgehen dazu führt, dass die Interaktionseffekte Textlänge  $\times$  Textgattung und Textlänge  $\times$  Thema nicht mehr in das endgültige Modell aufgenommen werden können (das wäre ein Interaktionseffekt, der dem im vorherigen Abschnitt diskutierten Effekt Textlänge  $\times$  Komplexitätstyp am nächsten kommen würde). Residualisiert man die Gesamtlesezeit aber nicht, muss man mit einer trimodalen Verteilung der abhängigen Variable umgehen, da die Textlängen von Zeitungsartikeln, Pressemitteilungen und Urteilen sehr weit auseinander liegen (siehe hierzu auch Tab. 6.4) und somit zu drei getrennten Gipfeln in der Verteilung der Gesamtleszeiten führen. Dies ist sehr gut im linken

Schaubild von Abbildung 6.3 abzulesen. Auf dem rechten Schaubild, in dem die residualisierten Lesezeiten abgetragen sind, überlappen sich die residualen Lesezeiten für Artikel, Pressemitteilungen und Urteilen zu großen Teilen (signalisiert durch die dunklere Einfärbung der Balken) und die Trimodalität der Verteilung ist verschwunden. Für die rohen logarithmierten Lesezeiten (linkes Schaubild in Abb. 6.3) sind ganz eindeutig drei Verteilungen mit minimalen Überlappungsbereichen zu erkennen.

		Textgattung		
		Zeitung- artikel	PMen	Urteile
Thema	Aufenthalt	56,5 s (14,3)	203 s (82,0)	1034 s (207)
	AWACS	58,9 s (15,4)	241 s (67,0)	865 s (217)
	Cannabis	45,5 s (14,3)	209 s (41,2)	1005 s (239)

Tab. 6.4: Mittelwerte (Standardabweichungen) für die Gesamtlesedauer der natürlichen, kompletten Texte (FLRCNat).

Als Vorhersagevariablen auf dieser Granularitätsstufe der Analyse bleiben nun noch das Thema des Texts sowie die Textgattung. Außerdem geht die Position des Texts für die/den jeweiligen VersuchsteilnehmerIn in die Vorhersage ein. Eine Ausreißerkorrektur ist nicht notwendig, da keiner der 122 Datenpunkte die Kriterien eines Ausreißers erfüllt.

Das angepasste gemischte Modell (Tab. 6.5) enthält noch alle Prädiktoren bis auf die *random slope* für VersuchsteilnehmerInnen in Abhängigkeit der Textposition (diese hat zwar einen beschleunigenden Einfluss, der sich aber offenbar nicht signifikant über die TeilnehmerInnen hinweg verändert). Der Textgattungseffekt (*textgatt*) ist sehr ausgeprägt. Pressemitteilungen werden langsamer gelesen (gegeben die residualisierten Gesamtlesezeiten) als Zeitungsartikel, ebenso die Urteile, die sich zwischen den beiden Stufen befinden (das deuten die Abbildungen 6.3 (rechtes Schaubild) und 6.4 schon an und es wird durch ein gemischtes Modell mit ungeordneten Faktorenstufen für die Textgattung bestätigt:  $t_{ZAvsU} = -8,63$ ;  $t_{PMvsU} = 9,23$ ; beide  $p_{MCMC} = 0,0001$ ). Auch das Textthema (*thema*) wirkt sich signifikant auf die residuale Lesezeit der natürlichen Texte aus ( $p_{MCMC} = 0,003$ ). Insbesondere wird dieser Effekt vom Kontrast zwischen dem Thema 'unerlaubter Aufenthalt' und 'Cannabis' hervorgerufen. Wie ein gemischtes Modell mit ungeordneten Faktorstufen zeigt, liegt das Thema 'AWACS' zwischen den beiden anderen Themengebieten, ohne einen verlässlichen, signifikanten Unterschied zu einem der beiden ( $t_{AUvsAW} =$

2,90;  $p_{\text{MCMC}} = 0,083$ ;  $t_{\text{CAvsAW}} = -1,17$ ;  $p_{\text{MCMC}} = 0,27$ ).<sup>93</sup> Die Interaktion Textgattung  $\times$  Thema ist insgesamt nicht signifikant ( $p_{\text{MCMC}} = 0,199$ ), und lediglich ein Kontrast wird marginal signifikant (siehe Tab. 6.5). Trotzdem verschlechtert sich das Modell signifikant, wenn man die Interaktion ausschließt.

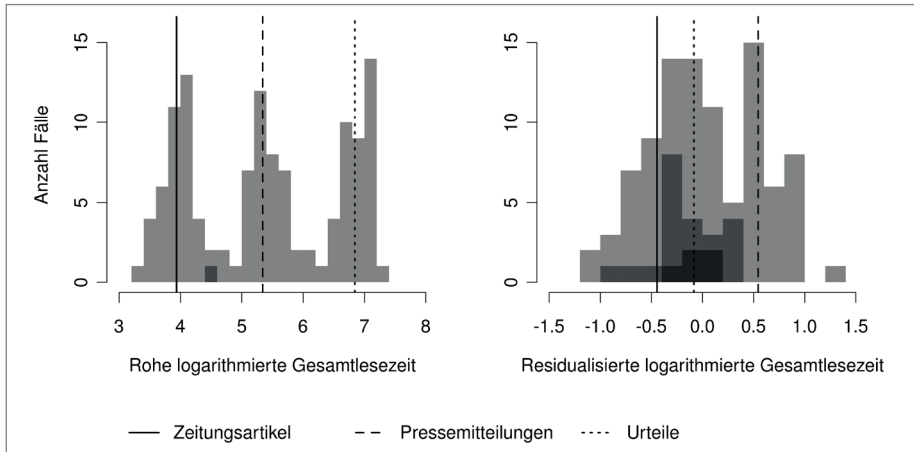


Abb. 6.3: Histogramme der logarithmierten Lesezeiten (links) und der residualisierten logarithmierten Lesezeiten (rechts) sowie die Mittelwerte der jeweiligen Maße (vertikale Linien) für die Zeitungsartikel, Pressemitteilungen und Urteile. Überlappen sich die Histogramme, werden die entsprechenden Bereiche dunkler eingefärbt.

Groups	Name	Variance	Variance	Variance		
vtn	(Intercept)	0.062276	0.062276	0.062276		
Residual		0.012750	0.012750	0.012750		
		Estimate	Std. Error	t value	$P_{\text{MCMC}}$	
(Intercept)		-0.31516	0.05607	-5.621	0.0001	***
textgatt PM		0.99243	0.05213	19.037	0.0001	***
textgatt U		0.46961	0.05440	8.632	0.0001	***
thema awacs		0.15668	0.05619	2.789	0.3664	
thema cannabis		-0.21576	0.05128	-4.207	0.0074	**
text.pos		-0.05031	0.01292	-3.893	0.0358	*
textgatt PM:thema awacs		-0.14714	0.08405	-1.751	0.8628	
textgatt U:thema awacs		-0.32233	0.08547	-3.771	0.0724	.
textgatt PM:thema cannabis		0.12092	0.08080	1.497	0.2592	
textgatt U:thema cannabis		-0.01039	0.08248	-0.126	0.7854	

Tab. 6.5: Angepasstes gemischtes Modell für die residualisierte logarithmierte Lesedauer der Zeitungsartikel, Pressemitteilungen und Urteile im FLRCNat-Teil.

<sup>93</sup> t-Wert-Indizierung: AU = Aufenthalt, AW = AWACS, CA = Cannabis.

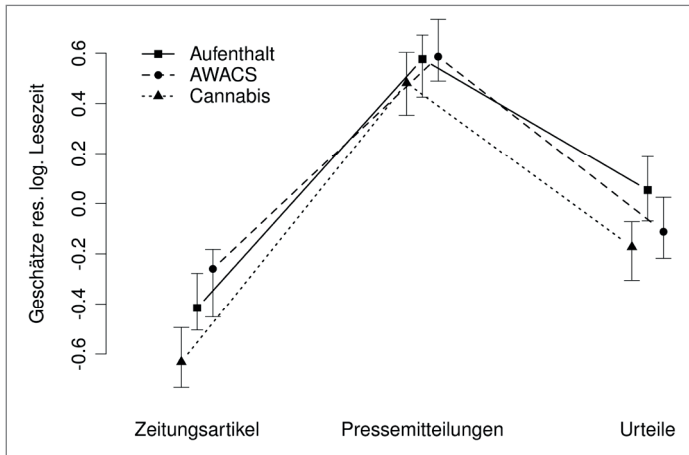


Abb. 6.4: Geschätzte residualisierte logarithmierte Gesamtlesezeit für Zeitungsartikel, Pressemitteilungen und Urteile in den unterschiedlichen Themengebieten sowie 95%-Konfidenzintervalle.

Abbildung 6.4<sup>94</sup> visualisiert den Interaktionseffekt. Zunächst wird deutlich, dass sowohl der Zeitungsartikel als auch das Urteil zum Thema Cannabis schneller gelesen werden und auch die Pressemitteilung ist zumindest numerisch jene, die am schnellsten gelesen wird. Das Urteil zum Thema 'AWACS' ist sowohl unter den Zeitungsartikeln als auch unter den Pressemitteilungen das am längsten gelesene – dieser Effekt ist jedoch nicht signifikant. Im Grunde liegen diese Texte gleichauf mit jenen zum Thema 'unerlaubter Aufenthalt'. Trotzdem reicht offenbar die Umkehrung dieser Reihenfolge bei den Urteilen für einen marginal signifikanten Kontrast. Am auffälligsten in dieser Analyse ist sicherlich der große Unterschied zwischen den verschiedenen Textgattungen, der auch noch besteht, **nachdem** die Textlänge aus den Gesamtlesezeiten herausgerechnet wurde. Dass Zeitungsartikel die am schnellsten gelesenen Texte sind, ist nicht verwunderlich – sind es doch die einzigen Texte, deren primäre Funktion es ist, von juristischen Laien rezipiert zu werden. Das könnte man auch von Pressemitteilungen behaupten, doch bereits in Abschnitt 5.2.2 und auch in der Arbeit von Hansen et al. (2006) zeigte sich, dass die Pressemitteilungen zum Teil sehr anspruchsvoll sind.

Eine andere Beobachtung ist meines Erachtens überaus interessant: Auch wenn Textlänge aus den Gesamtlesezeiten für einen Text herauspartialisiert ist, gruppieren sich die Texte einer Textgattung stark zusammen. Abbildung 6.4

<sup>94</sup> Die abgetragenen Konfidenzintervalle sind über MCMC-Simulationen ermittelt. Wann immer ein gemischtes Modell visualisiert wird, ist dies der Fall. In den Bildunterschriften wird dies nicht mehr einzeln vermerkt.

macht das deutlich. So scheint bereits ein sehr grobes Maß wie die residualisierte Lesezeit genug Sensitivität zu besitzen, zwischen Texten unterschiedlicher Gattung unterscheiden zu können. Solch deutliche Effekte waren nicht zwingendermaßen zu erwarten, auch wenn in Abschnitt 5.2.2 einige Unterschiede auf der linguistischen Ebene zwischen den Textgattungen herausgearbeitet wurden. Die gezeigten Unterschiede waren jedoch nie dergestalt, dass die Pressemitteilungen linguistisch am komplexesten waren. Sie gruppieren sich dagegen meist zwischen Zeitungsartikeln und Pressemitteilungen ein. Ein ganz ähnliches Muster zeigte sich für einige Dimensionen in den kontrastiven Korpusanalysen von Hansen-Schirra/Neumann (2004).

Analysen zur Antwortperformanz der VersuchsteilnehmerInnen stellen ein im Vergleich zur Lesedauer textferneres Maß dar (zur Herleitung dieses Begriffs siehe Abschnitt 3.2 der vorliegenden Arbeit). Dafür kann man aber argumentieren, dass sie eher auf die mentale Repräsentation abzielen, die nach dem Lesen des Textes besteht.

## 6.2.2 Antwortperformanz

### Originalauszüge und Reformulierungen

Nachdem sie die Texte gelesen hatten, mussten die VersuchsteilnehmerInnen nach jedem Text eine ja/nein-Entscheidungsfrage beantworten. Zum genaueren Aufbau der Fragen siehe Abschnitt 5.1.3. Wie in Tabelle 6.6 ersichtlich, liegt die Antwortperformanz der VersuchsteilnehmerInnen generell über Zufallsniveau.<sup>95</sup> Für die Texte mit vielen Nominalisierungen und komplexen Nominalphrasen ist dies recht deutlich. Die Texte mit komplexer Syntax fallen etwas ab – was vermutlich auch daran liegt, dass dies gleichzeitig die längsten Texte sind. Um diesen und andere Effekte kontrollieren zu können, ist hier ein gemischtes Modell nötig.

		Komplexitätstyp		
		Nom.	NP	Synt.
Version	Original	84,1%	84,0%	70,2%
	moderat	87,8%	81,9%	61,1%
	stark	87,0%	84,0%	63,0%

Tab. 6.6: Anteil korrekter Antworten für Originalauszüge und Reformulierungen, aufgeteilt nach Version und Komplexitätstyp.

<sup>95</sup> Es antwortete kein/e einzelner VersuchsteilnehmerIn unter Zufallsniveau. Dasselbe gilt auch für die natürlichen, vollständigen Texte.

Groups	Name	Variance	Std.Dev.			
vtn	(Intercept)	0.024303	0.1559			
		Estimate	Std. Error	z value	p value	
(Intercept)		4.368252	0.613557	7.120	1.08e-12	***
typ NP		-5.387833	0.893157	-6.032	1.62e-09	***
typ Synt		-2.614603	0.662216	-3.948	7.87e-05	***
n.chars		-0.011042	0.003440	-3.210	0.00133	**
korr.antwort ja		0.507033	0.211919	2.393	0.01673	*
kontext ja		-0.996261	0.224251	-4.443	8.89e-06	***
typ np:n.chars		0.025556	0.004726	5.408	6.39e-08	***
typ synt:n.chars		0.008392	0.003497	2.399	0.01642	*

Tab. 6.7: Angepasstes gemischtes logistisches Regressionsmodell für die Wahrscheinlichkeit einer korrekten Antwort für die Originaltexte sowie die moderaten und starken Reformulierungen.

Das angepasste Modell für die Wahrscheinlichkeit einer korrekten Antwort auf die Frage ist in Tabelle 6.7 abgedruckt. Es handelt sich um ein gemischtes logistisches Regressionsmodell.<sup>96</sup> In diesem Fall werden die mit einem Kontrast assoziierten *p*-Werte direkt ausgegeben.

Es ist kein signifikanter Effekt der Reformulierungsversion festzustellen – weder als Haupteffekt, noch als beteiligter Faktor in einer Interaktion. Die Position im Experiment, für die Gesamtlesedauer noch ausschlaggebend, hat ebenfalls keinen signifikanten Einfluss. Die Kontraste zwischen den verschiedenen Komplexitätstypen sind ausgeprägt. Auch die nicht überprüften Kontraste stellen sich nach einer Umordnung der Faktorstufen als höchst signifikant heraus:  $z_{\text{NOMvsSYNT}} = 3,95$ ;  $p < 0,0001$ ;  $z_{\text{NPvsSYNT}} = -3,83$ ;  $p = 0,0001$ . Fragen nach Texten mit vielen Nominalisierungen werden demnach am ehesten korrekt beantwortet. Fragen nach Texten mit komplexer Syntax liegen im Mittelfeld mit signifikanten Abweichungen sowohl zum Komplexitätstyp *Nom.* als auch zu Texten des Komplexitätstyps *NP*, für die es am unwahrscheinlichsten ist, dass sie korrekt beantwortet werden. Je länger die Texte sind (*n.chars*), desto unwahrscheinlicher wird eine korrekte Antwort. Müssen die VersuchsteilnehmerInnen mit ‘ja’ auf die Frage antworten (*korr.antwort ja*), wird eine korrekte Antwort wahrscheinlicher. Dieser Effekt ging als ‘inhaltsunabhängige Zustimmungstendenz’ in die Literatur ein (vgl. Schnell/Hill/Esser 2008, S. 154) und besagt, dass es einer/einem VersuchsteilnehmerIn leichter fällt, positiv auf eine Entscheidungsfrage zu antworten als negativ.

<sup>96</sup> Ein logistisches Regressionsmodell kann über den Parameter `glmer(..., family = "binomial")` angefordert werden. Es werden dabei keine *t*-Werte, sondern *z*-Werte als Prüfgröße angegeben (vorletzte Spalte der Tabelle).

Ein Kontext stiftender Text vor dem eigentlichen Stimulustext wirkt sich negativ auf die Wahrscheinlichkeit einer korrekten Antwort aus.

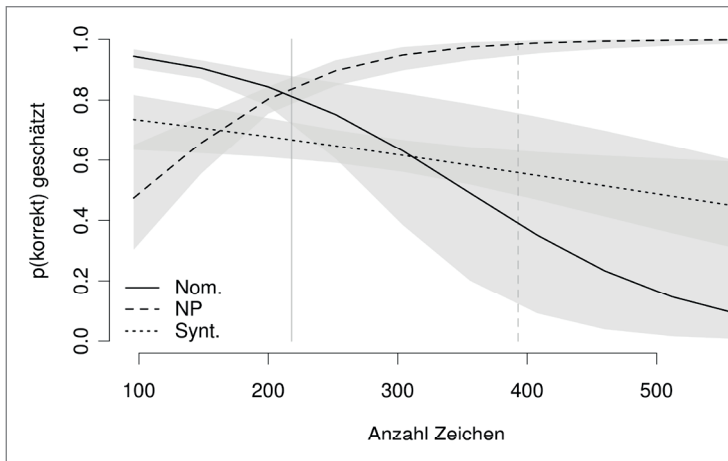


Abb. 6.5: Geschätzte Wahrscheinlichkeit für eine korrekte Antwort in Abhängigkeit von Textlänge (Anzahl Zeichen) und Komplexitätstyp (inkl. 95%-Konfidenzintervalle). Die senkrechten Linien geben die jeweiligen Maximalwortanzahlen für die entsprechenden Komplexitätstypen an.

Außerdem interagieren die Faktoren Komplexitätstyp und Textlänge. Diese Interaktion ist in Abbildung 6.5 abgetragen.<sup>97</sup> Es wird deutlich, dass die Texte mit vielen Nominalisierungen und jene mit komplexer Syntax die erwartete Effektrichtung zeigen: Je länger ein Text ist, desto unwahrscheinlicher wird es, dass korrekt darauf geantwortet wird. Für Texte mit komplexen Nominalphrasen jedoch gilt, dass die Fragen nach längeren Texten eher korrekt beantwortet werden – der Effekt der Textlänge dreht sich hier also um. In Abbildung 6.5 sind außerdem die Maximalbuchstabenanzahlen für die Texte unterschiedlichen Komplexitätstyps mit senkrechten Linien eingetragen. Das gibt einen Eindruck davon, in welchem Bereich die Schätzer ermittelt wurden. So ist insbesondere für die Texte mit vielen Nominalisierungen ersichtlich, dass die Effektschätzer eigentlich nur bis zu einer Textlänge von 218 Zeichen angepasst sind. Alles darüber hinaus stellt eine Extrapolation des Effekts dar, der nicht unbedingt zu vertrauen ist. Das ist auch durch große Konfidenzintervalle im entsprechenden Bereich gekennzeichnet. Im Lichte dessen muss der Effektkontrast, der in Tabelle 6.7 in der letzten Zeile abgetragen ist, sehr vorsichtig interpretiert werden. Dieser besagt, dass der Effekt der Länge auf die Korrektheit der Antwort in Komplexitätstyp *Synt.* im Vergleich zu

<sup>97</sup> Die Effektschätzer sind anhand des R-Pakets *effects* (Fox 2003) extrahiert.



*Nom.* signifikant nach oben korrigiert wird. Anders ausgedrückt: Die Steigung der Funktion, die den Schätzer für die Korrektheit der Antwort in Abhängigkeit der Textlänge beschreibt, wird von *Nom.* zu *Synt.* positiv verändert. Betrachtet man jedoch die Linienverläufe in Abbildung 6.5 bis zur ersten senkrechten Markierung, also der maximalen tatsächlich vorkommenden Buchstabenanzahl für Texte des Komplexitätstyps *Nom.*, kann nur ein minimaler Unterschied der Steigungen der beiden Effektschätzer festgestellt werden.

Der Interaktionseffekt bezüglich der Texte mit komplexen Nominalphrasen ist in Abbildung 6.5 ersichtlich. Die Fragen nach Texten des Komplexitätstyps *NP* werden mit steigender Länge des Textes offenbar eher korrekt beantwortet. Das ist etwas überraschend, geht dieser Effekt doch in die entgegengesetzte Richtung wie bei den beiden anderen Komplexitätstypen. Ein Grund hierfür könnte sein, dass eine oder mehrere Fragen nach sehr langen Texten recht einfach sind. Und in der Tat ist die Frage nach den beiden längsten Texten des Komplexitätstyps *NP* (Item 12, moderate und starke Reformulierungen) sehr einfach. In Beispiel 22 ist die moderate Reformulierungsversion (der zweitlängste Text im Komplexitätstyp *NP*) dieses Items abgedruckt. Die Frage 22b bezieht sich auf den in 22a hervorgehobenen Teil. Tatsächlich hat kein/e VersuchsteilnehmerIn diese Frage falsch beantwortet. Für die starke Reformulierungsversion gilt dasselbe. Das ist kaum verwunderlich, ist die Frage doch fast eine Kopie des letzten Teilsatzes im Stimulustext. In diesem Fall ist die Länge des Textes natürlich nicht mehr ausschlaggebend. Nicht alle Fragen nach den längsten Texten sind tatsächlich so einfach, aber Stimulustext Nummer 12 trägt auf jeden Fall zum beobachteten Längeneffekt bei Texten des Komplexitätstyps *NP* bei.

(22) a. (Stimulustext)

Die Ausführungen des Bayerischen Staatsministeriums der Justiz, der Vertrag gelte als auf unbestimmte Zeit abgeschlossen, weil das Schriftformerfordernis des Paragraphen 566 Satz 1 in Verbindung mit Paragraph 580 BGB hinsichtlich vom Beschwerdeführer vorgetragener Vereinbarungen zu Renovierungs- und Instandsetzungspflichten nicht eingehalten sei, **zwingen zu keiner anderen Beurteilung.**

b. (Frage)

Zwingen die Ausführungen des Bayerischen Staatsministeriums zu einer anderen Beurteilung?

Die eben vorgestellten Ergebnisse können direkt mit den Erkenntnissen von Hansen et al. (2006) verglichen werden. Sie berichten (vgl. Hansen et al. 2006, S. 35f.) einen Effekt der Reformulierungsversion (für den Omnibustest:

$F(2,58) = 5,88; p < 0,0047$ ), wobei die Wahrscheinlichkeit einer korrekten Antwort nach den moderat reformulierten Texten signifikant größer ist als für die Originaltexte und die starken Reformulierungen (zwischen diesen beiden Stufen besteht kein signifikanter Unterschied). Dieser Effekt kann anhand der vorliegenden Daten in keinem Modell bestätigt werden. Selbst eine einfache Varianzanalyse, die im Vergleich zu einem gemischten logistischen Regressionsmodell für diese Auswertung ohnehin schlechter geeignet ist (vgl. Jaeger 2008), zeigt keinen Effekt in diese Richtung. Auch ein gemischtes logistisches Regressionsmodell, das **nur** den Faktor Reformulierungsversion und *random intercepts* für VersuchsteilnehmerInnen enthält, gibt keine Hinweise darauf, dass der von Hansen et al. gefundene Effekt repliziert werden kann.

		Textgattung		
		Zeitungsartikel	PMen	Urteile
Thema	Aufenthalt	78,8 %	60 %	76,7 %
	AWACS	74,5 %	68,6 %	70 %
	Cannabis	80 %	61,5 %	56 %

Tab. 6.8: Anteil korrekter Antworten für die kompletten Zeitungsartikel, Pressemitteilungen und Urteile, aufgeteilt nach Thema und Textgattung.

Für die kompletten Texte aus dem FLRCNat-Teil wurden pro Text fünf Multiple-Choice-Fragen mit je vier Antwortmöglichkeiten erstellt, das Zufallsniveau liegt somit bei 25% korrekt beantworteter Fragen (für eine genauere Erläuterung der Fragen siehe Abschnitt 5.1.2).

Wie in Tabelle 6.8 ersichtlich, liegt die Antwortperformanz der VersuchsteilnehmerInnen konstant deutlich über Zufallsniveau.<sup>98</sup> Alle Fragen wurden erst nach dem jeweiligen Text gestellt. Daher wäre es zu erwarten gewesen, dass die Performanz bei den Urteilen deutlich abfällt, da vor den Fragen sehr viel Textmaterial steht. Das ist nicht der Fall, wie ebenfalls in Tabelle 6.8 abgelesen werden kann.

Wie in Abschnitt 5.1.2 erläutert, wurden die Fragen so konzipiert, dass sie entweder ganz konkret propositionale Einheiten in den Texten abfragen oder sich eher auf generelle Sachverhalte oder globale Zusammenhänge beziehen. Grob in die Terminologie von Kintsch (1988) übersetzt, würde sich diese zweite Art

<sup>98</sup> Einschränkung muss angemerkt werden, dass zwei der 45 Fragen tatsächlich unter Zufallsniveau (6,7% und 15,4% korrekt) beantwortet wurden. Es handelt sich dabei um eine Frage nach dem 'Cannabis'- Urteil und eine Frage nach der Pressemitteilung zum Thema 'Aufenthalt'.

von Fragen eher auf das Situationsmodell beziehen, das beim Lesen der Texte aufgebaut wird. Um diese und weitere Einflussfaktoren auf die Antwortperformanz der VersuchsteilnehmerInnen zu analysieren, wird auch hier ein gemischtes logistisches Regressionsmodell herangezogen (siehe Tab. 6.9).

Groups vtn	Name (Intercept)	Variance 0.41763	Std.Dev. 0.64624			
		Estimate	Std. Error	z value	p value	
(Intercept)		1.23525	0.32410	3.811	0.000138	***
textgatt PM		-1.13160	0.42753	-2.647	0.008125	**
textgatt U		0.14279	0.47153	0.303	0.762021	
thema awacs		-0.29475	0.46809	-0.630	0.528900	
thema cannabis		0.21622	0.45577	0.474	0.635216	
fragentyp sitmod		0.51288	0.21129	2.427	0.015209	*
textgatt PM:thema awacs		0.95821	0.65524	1.462	0.143637	
textgatt U:thema awacs		-0.38330	0.69183	-0.554	0.579548	
textgatt PM:thema cannabis		-0.03476	0.64328	-0.054	0.956906	
textgatt U:thema cannabis		-1.46737	0.66847	-2.195	0.028155	*

Tab. 6.9: Angepasstes gemischtes logistisches Regressionsmodell für die Wahrscheinlichkeit einer korrekten Antwort für die Zeitungsartikel, Pressemitteilungen und Urteile.

Die Kontraste zwischen den verschiedenen Textgattungen (Faktor *textgatt*) werden im Kontext der signifikanten Interaktion Textgattung × Thema interpretiert. Die Unterschiede zwischen den Themen (*thema*) sind – solange die Interaktion außen vor gelassen wird – nicht signifikant. Der Effekt verbleibt jedoch im endgültigen Modell, da er an der signifikanten Interaktion beteiligt ist. Fragen, die eher auf die globale Interpretation des Textes abzielen, werden generell eher korrekt beantwortet als Fragen, die sich auf eine einzelne Proposition im Text beziehen (*fragentyp*). Auch die Rohwerte legen das nahe: 74,4% aller Fragen nach globalen Zusammenhängen werden korrekt beantwortet, aber nur 67,0% der Fragen nach einzelnen Propositionen.

Die Interaktion Textgattung × Thema ist in Abbildung 6.6 abgetragen. Es zeigt sich, dass sich die Antwortperformanz nach Zeitungsartikeln kaum unterscheidet. Nach Pressemitteilungen werden Fragen eher schlechter beantwortet, was aber hauptsächlich auf die Themen ‘Cannabis’ und ‘Aufenthalt’ zurückzuführen ist. Die größte Ausdifferenzierung unter den Themengebieten findet sich für die Urteile. Die beste Antwortperformanz zeigen die VersuchsteilnehmerInnen hier nach dem Thema ‘Aufenthalt’. Der Unterschied zum ‘AWACS’-Urteil ist marginal signifikant.<sup>99</sup> Am schlechtesten werden Fragen

<sup>99</sup> In einem Modell, das nur die Urteile einschließt, gilt für den Unterschied ‘AWACS’ vs. ‘Aufenthalt’:  $\beta = -0,76$ ;  $SE = 0,45$ ;  $z = -1,69$ ;  $p = 0,09$ .

nach dem 'Cannabis'-Urteil beantwortet. Der Unterschied zum 'AWACS'-Urteil ist dabei nicht signifikant ( $p = 0,26$ ), der Unterschied zum 'Aufenthalts'-Urteil hingegen deutlich ausgeprägt.<sup>100</sup> Betrachtet man die verschiedenen Themengebiete, fallen drei grundlegend unterschiedliche Linienverläufe auf: Für das Thema 'unerlaubter Aufenthalt in Deutschland' ist ein klarer Einbruch bei den Pressemitteilungen zu sehen. Die Antwortperformanz für Zeitungsartikel und Urteile befindet sich auf einem (recht hohen) Niveau. Überhaupt keine Entwicklungen sind für die 'AWACS'-Texte festzustellen. Hier scheinen die Fragen nach allen drei Texten ähnlich gut beantwortet worden zu sein. Beim Thema 'Cannabis' können die VersuchsteilnehmerInnen lediglich die Fragen nach Zeitungsartikeln gut beantworten. Die Performanz für Pressemitteilungen und Urteile fällt demgegenüber deutlich ab.

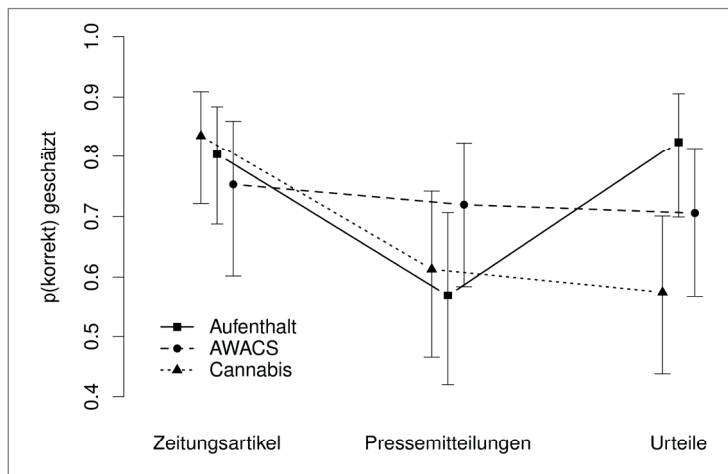


Abb. 6.6: Geschätzte Wahrscheinlichkeit für eine korrekte Antwort in Abhängigkeit von Textgattung und Thema (inkl. 95%-Konfidenzintervalle).

Es ergibt sich somit kein solch einheitliches Bild, wie das noch bei der residualen Lesedauer (siehe Abb. 6.4) der Fall war. Dort sahen wir, dass die verschiedenen Textgattungen intern sehr konsistent und im Vergleich zu den anderen Textgattungen deutlich abgegrenzt waren. Diese Gruppierung der verschiedenen Themen innerhalb der Textgattung ist in Abbildung 6.6 am ehesten bei den Zeitungsartikeln zu beobachten, aber insgesamt sicher nicht so ausgeprägt wie beim Maß der residualen Lesedauer. Die Gründe hierfür können vielfältig sein, am plausibelsten erscheinen meines Erachtens aber zwei eher methodische Gründe. 1) Das nicht-distinktive Effektmuster be-

<sup>100</sup> In einem Modell, das nur die Urteile einschließt, gilt für den Unterschied 'Aufenthalt' vs. 'Cannabis':  $\beta = 1,20$ ;  $SE = 0,42$ ;  $z = 2,86$ ;  $p = 0,004$ .

züglich der Textgattungen liegt am Maß Antwortperformanz selbst. Während sich die residuale Lesedauer noch direkt auf den Text selbst bezieht, ist die Antwortperformanz bei Fragen **nach** dem eigentlichen Text ein viel indirekteres Maß. Effekte der Textgattung können durch Effekte überschattet werden, die zwischen dem Lesen selbst und den für die Antwort nötigen Verarbeitungen stattfinden. 2) Der zweite Grund könnten die Fragen selbst sein. Diese waren für die Zeitungsartikel, Pressemitteilungen und Urteile immer unterschiedlich. Vergleichen wir beispielsweise ein Urteil und einen Zeitungsartikel, wird recht schnell deutlich, warum das nötig war. Im Urteil sind die Informationen viel detaillierter und meist auch mit mehr Bezug zum verfassungsgerichtlichen und/oder staatsrechtlichen Kontext dargestellt. In Zeitungsartikeln sind die wichtigsten Informationen knapp zusammengefasst, und das meist auf einer ganz anderen Abstraktionsebene als in den Urteilen selbst. Längere Passagen werden praktisch nie formulierungsgetreu aus dem Urteil übernommen. Somit überschneiden sich propositionaler Gehalt von Urteil und Zeitungsartikel meist nur sehr wenig. Da die Fragen auf den propositionalen Gehalt eines Textes abgestimmt sind, werden nach Texten, die zwar zum selben Thema verfasst wurden, aber unterschiedlichen Textgattungen angehören, unterschiedliche Fragen gestellt. Daraus wiederum kann folgen, dass die Textgattungen nicht effektiv anhand der Antwortperformanz der LeserInnen unterschieden werden können und ein Muster entsteht, wie es in Abbildung 6.6 zu sehen ist. Dies muss hier allerdings als Verdacht stehen bleiben und kann anhand der vorliegenden Daten nicht untermauert werden.

### 6.2.3 Antwortperformanz und Lesezeit

Obwohl wir uns erst im folgenden Abschnitt 7 der detaillierten Auswertung der Blickbewegungsdaten zuwenden, soll hier auf einen Teilaspekt der Blickbewegungsdaten vorgegriffen werden. Es geht dabei gleichsam um einen Kernbereich dieser Dissertation, nämlich das Zusammenspiel von Blickbewegungen und der Verständlichkeit von Texten. Wie in Kapitel 3 dargelegt wurde, kann Textverständlichkeit unter mehreren Blickwinkeln angegangen werden. Blickbewegungen sagen dabei eher etwas über den Verstehens**prozess** aus – die Messung von Blickbewegungen kann somit als eine textnahe Methode beschrieben werden (vgl. Christmann 2004). Wie gut auf Fragen nach einem Text geantwortet wird, würde gemäß dieser Logik den textfernen Maßen zugerechnet. Im aktuellen Abschnitt wollen wir nun versuchen, diese beiden Pole miteinander in Verbindung zu bringen. Unter anderem hatten Reynolds/Anderson (1982) eine ähnliche Idee im Rahmen einer Studie mit experimentell kontrollierten Texten. Sie stellten ihren VersuchsteilnehmerInnen Fragen, die lediglich auf ganz bestimmte Kategorien von Textinformationen abzielten.

Diese *between participants*-Manipulation variierte technische Informationen vs. Namensinformationen. Reynolds und Anderson konnten zeigen, dass die VersuchsteilnehmerInnen Textabschnitten mehr Aufmerksamkeit widmeten, die zur Kategorie der Frage passten (also Textabschnitte mit Namensinformationen oder Textabschnitte mit technischen Informationen).<sup>101</sup> Textabschnitte, die zur jeweiligen Fragenkategorie passten, wurden länger gelesen. Die Autoren schließen daraus, dass die Fragen dazu führten, dass die LeserInnen den für sie relevanten Textabschnitten mehr Aufmerksamkeit widmeten. Die von Reynolds/Anderson (1982) untersuchte Wirkungsrichtung kann also so beschrieben werden, dass die zuvor gestellten Fragen ursächlich das Lesen beeinflussen.

Im aktuellen Abschnitt soll die entgegengesetzte Wirkungsrichtung untersucht werden, also wie der vorherige Leseprozess ursächlich auf die Antwortperformanz bei Fragen wirkt. Im Gegensatz zur eben vorgestellten Studie wurden die Fragen im Falle des FLRC immer erst **nach** dem Text gestellt, auf den sie Bezug nahmen. Mit diesem Verfahren sollte hauptsächlich sichergestellt werden, dass der Lesefluss nicht von Fragen mitten im Text unterbrochen wird, die VersuchsteilnehmerInnen aber trotzdem aufmerksam die Texte lesen, weil sie wissen, dass danach Fragen gestellt werden und damit ihr Leseverständnis überprüft wird. Der inhaltliche Typ der erfragten Information variierte und wurde nicht streng experimentell kontrolliert. Durch eine nachträglich eingeführte Annotationsebene kann aber kontrolliert werden, wie lange die VersuchsteilnehmerInnen die für die Beantwortung der Fragen relevante Information im Text gelesen haben. Von Interesse ist nun, ob die Zeit, die die VersuchsteilnehmerInnen mit dem Lesen der relevanten Information verbracht haben, einen Einfluss auf die korrekte Beantwortung der Frage hat. Der Datensatz für diese Analyse ist so aufgebaut, dass ein Fall immer aus einer Beantwortung einer Frage durch eine/n VersuchsteilnehmerIn besteht. Abhängige Variable ist die Antwortperformanz. Als Prädiktoren stehen bereit: die Gesamtlesezeit (*total reading time*) auf relevanten Textabschnitten, die Gesamtlesezeit auf nicht-fragenrelevanten Textabschnitten, die Reaktionszeit auf die Frage sowie (im Falle der natürlichen Texte) der Fragentyp (Einzelproposition vs. Situationsmodell). Alle Zeitspannen gehen als logarithmierte Variablen in die Modelle ein.<sup>102</sup> Als Zufallseffekte (*random intercepts*) gehen die Identität der VersuchsteilnehmerInnen sowie die Text-ID in die

---

<sup>101</sup> Die Textabschnitte wurden gezielt so formuliert, dass tatsächlich nur eine Art von Information in einem Textabschnitt enthalten war.

<sup>102</sup> Durch diese Transformation werden die Variablen in eine Normalverteilung überführt. Für Prädiktoren ist das nicht zwingend notwendig. Allerdings zeigten sich die Modelle mit logarithmierten Prädiktoren robuster als jene mit nicht-logarithmierten Prädiktoren.

Modelle ein. Da die Art der Fragen für die reformulierten Texte und die natürlichen Texte unterschiedlich waren (Reformulierungen und Originalauszüge: Entscheidungsfragen; natürliche, komplette Texte: Multiple-Choice-Fragen), werden unterschiedliche Modelle für die beiden Fragentypen berechnet. Der Datensatz für die Reformulierungen und Originalauszüge umfasst 1.178 Fälle, jener für die natürlichen Texte 530 Fälle.

Gegeben die Ergebnisse von Reynolds/Anderson (1982) ist zu erwarten, dass eine Frage dann einfacher beantwortet werden kann, wenn man jenem Textmaterial mehr Aufmerksamkeit zukommen ließ, das relevant für die Beantwortung der Frage war. Außerdem schließen wir die Gesamtzeit auf Textabschnitten mit ein, die nicht fragenrelevant waren. Dadurch soll kontrolliert werden, dass etwaige Effekte nicht einfach darauf zurückzuführen sind, dass der Text insgesamt länger gelesen wurde. In diesem Fall würde sowohl die Lesezeit auf fragenrelevanten als auch auf nicht-fragenrelevanten Textabschnitten steigen.

Groups	Name	Variance	Std.Dev.			
text.id	(Intercept)	0.8983052	0.947790			
vtn	(Intercept)	0.0020923	0.045742			
		<b>Estimate</b>	<b>Std. Error</b>	<b>z value</b>	<b>p value</b>	
(Intercept)		8.2852	1.6885	4.907	9.25e-07	***
log.q.trt		0.4151	0.1710	2.428	0.0152	*
log.q.rt		-0.8529	0.1724	-4.946	7.56e-07	***
log.nq.trt		-0.3405	0.1718	-1.981	0.0476	*

Tab. 6.10: Gemischtes logistisches Regressionsmodell für die Antwortperformanz in Abhängigkeit der Gesamtlesezeit auf fragenrelevanten Abschnitten (*log.q.trt*), der Gesamtlesezeit auf nicht-fragenrelevanten Abschnitten (*log.nq.trt*) sowie der Reaktionszeit auf die Frage (*log.q.rt*), Originalauszüge und Reformulierungen.

Das in Tabelle 6.10 einsehbare Modell zeigt, dass die Vorhersage basierend auf den Ergebnissen von Reynolds/Anderson (1982) zutrifft. Auch der Effekt der Reaktionszeit zeigt in die erwartete Richtung: Je länger der/die LeserIn für die Antwort benötigt, desto unwahrscheinlicher wird eine korrekte Antwort. Interessanterweise geht der Effekt für die Lesezeit auf nicht-fragenrelevanten Textabschnitten (*log.nq.trt*) in die entgegengesetzte Richtung. Die Effekte scheinen sich also nicht einfach darauf reduzieren zu lassen, dass die Fragen besser beantwortet werden, wenn der Text als Ganzes länger gelesen wird. Es scheint in der Tat wichtig zu sein, dass die **relevanten** Textstellen mit genug Aufmerksamkeit gelesen werden.

Dies legt zumindest für einfache Entscheidungsfragen und sehr kurze Texte, wie sie im FLRCRef-Teil verwendet wurden, nahe, dass die Zeit, die man auf



fragenrelevanten Abschnitten verbringt, signifikant dazu beiträgt, ob die Frage später korrekt beantwortet wird. Würde man nach dem Lesen der Frage die Möglichkeit haben, wieder zurück in den Text zu springen, könnte man eventuell noch nach relevanter Information suchen. Das war während der Erhebungen für das FLRC aber nicht möglich. Ob sich diese Effekte auf die Multiple-Choice-Fragen nach den kompletten natürlichen Texten ausdehnen lassen, soll als nächstes analysiert werden. Als zusätzlicher Prädiktor steht uns der Typ der Frage (*q.type*) zur Verfügung. Die Fragen wurden in die Typen *prop* (für Proposition) und *sitmod* (für Situationsmodell) eingeteilt. Um unterschiedlich starke Einflüsse der Gesamtlesezeit fragenrelevanter Textregionen (*log.q.trt*) für diese beiden Fragentypen zuzulassen, geht die Variable Fragentyp (*q.type*) als einfacher Effekt und als interagierender Faktor in das Modell ein. Dabei wird außerdem die Interaktion mit der Variable *log.nq.trt* (Gesamtlesezeit nicht-fragenrelevanter Textregionen) beachtet. Die Zufallseffektstruktur in Vergleich zum vorherigen Modell bleibt unverändert.

Tabelle 6.11 zeigt die Ergebnisse dieser Analyse. Es ist kein Effekt der Lesezeit für nicht-fragenrelevante Textregionen festzustellen, weder als einfacher Haupteffekt noch als Interaktion mit dem Fragentyp. Numerisch gehen die Tendenzen in die gleiche Richtung wie der für den FLRCRef-Teil signifikante Effekt. Die Einflüsse sind aber weit davon entfernt, signifikant zu sein. Der Faktor *q.type* ist an einer signifikanten Interaktion mit der Lesezeit fragenrelevanter Textteile beteiligt. Die Reaktionszeit hat den bereits zuvor gezeigten klaren Effekt: Längere Reaktionszeiten sind eindeutig mit schlechterer Antwortperformanz assoziiert. Der Einfluss der Lesezeit auf fragenrelevante Textregionen ist marginal signifikant, wir wollen uns aber direkt der signifikanten Interaktion widmen, an der dieser Effekt ebenfalls beteiligt ist.

Anhand von Abbildung 6.7 kann diese Interaktion leicht interpretiert werden. Offenbar ist der Effekt für Fragen nach dem Situationsmodell sehr viel stärker als für Fragen nach einzelnen Propositionen. Das bestätigt sich, wenn man die Analysen getrennt für die beiden Fragentypen durchführt. Hier hat die Gesamtlesezeit relevanter Textabschnitte beim Fragentyp *prop* nurmehr einen marginal signifikanten Einfluss ( $\beta = 0,23$ ;  $SE = 0,12$ ;  $z = 1,86$ ;  $p = 0,06$ ), für Fragen nach dem Situationsmodell aber einen recht deutlichen ( $\beta = 1,09$ ;  $SE = 0,30$ ;  $z = 3,70$ ;  $p = 0,0002$ ). Fragen, die eher auf das Situationsmodell abzielten, decken dabei über die Lesezeit relevanter Textabschnitte hinweg eine recht große Spannweite der Wahrscheinlichkeiten einer korrekten Antwort ab: von 20% bis nahezu 100% Wahrscheinlichkeit für eine korrekte Antwort. Auch ein anderer Effekt ändert sich in dieser getrennten Analyse der beiden Fragentypen. Während bei Fragen nach einzelnen Propositionen die Reaktionszeit auf die Frage (*log.q.rt*) noch die übliche Effektrichtung zeigt ( $\beta = -0,76$ ;  $SE = 0,22$ ;

$z = -3,48$ ;  $p = 0,0005$ ), ist dieser Einfluss für die Situationsmodell-Fragen nicht mehr nachzuweisen ( $z < 1$ ). Das heißt: Wurden die LeserInnen nach einzelnen Propositionen gefragt, scheint es nicht so wichtig zu sein, wie lange sie die Stelle im Text gelesen haben, die die Information enthält. Je kürzer sie aber überlegen, bevor sie eine Antwort geben, desto eher ist diese Antwort auch richtig. Fragt man allerdings nach dem Situationsmodell, das die LeserInnen herausgebildet haben, also nach eher globalen Zusammenhängen, ist es nicht mehr ausschlaggebend, wie lange die LeserInnen benötigen, bis sie eine Antwort geben. Hier allerdings ist es sehr relevant, wie lange sie Textstellen gelesen haben, die diese globalen Zusammenhänge beschreiben.

Groups	Name	Variance	Std.Dev.			
vtn	(Intercept)	0.307742	0.55475			
text.id	(Intercept)	0.098013	0.31307			
		<b>Estimate</b>	<b>Std. Error</b>	<b>z value</b>	<b>p value</b>	
(Intercept)		6.23950	2.15558	2.895	0.00380	**
log.q.trt		0.23027	0.12370	1.861	0.06268	.
q.type sitmod		-4.48020	3.18766	-1.405	0.15988	
log.q.rt		-0.63658	0.19687	-3.234	0.00122	**
log.nq.trt		-0.10502	0.12812	-0.820	0.41239	
log.q.trt:q.type sitmod		0.66041	0.28921	2.283	0.02240	*
q.type sitmod:log.nq.trt		-0.07603	0.26798	-0.284	0.77662	

Tab. 6.11: Gemischtes logistisches Regressionsmodell für die Antwortperformanz in Abhängigkeit der Gesamtlesezeit auf fragenrelevanten Abschnitten (*log.q.trt*), der Gesamtlesezeit auf nicht-fragenrelevanten Abschnitten (*log.nq.trt*), der Reaktionszeit auf die Frage (*log.q.rt*) und den Fragentyp (*q.type*), komplette Texte.

Dieses Muster kann auf mehrere Weisen interpretiert werden: Eine Möglichkeit ist, dass jene Textregionen, die für Fragen nach dem Situationsmodell relevant sind, eben für den ganzen Text sehr zentral sind. Liest man diese wichtigen Stellen „zu schnell“, hat dies nachhaltigere Auswirkungen als bei den Fragen nach für das Verständnis des Gesamtzusammenhangs tendenziell unwichtigeren einzelnen Propositionen. Eine zweite Erklärung für die gefundene Effektstruktur könnte sein, dass Antworten auf Situationsmodell-Fragen einem „Alles-oder-nichts-Muster“ folgen: Wenn die LeserInnen die Antwort nicht wissen, versuchen sie auch nicht lange, die Antwort zu rekonstruieren. Je länger sie aber beim Lesen relevanter Abschnitte (und eventuell damit verbundenen Prozessen wie der Integration der beinhalteten Propositionen in die Textrepräsentation) zubringen, desto eher wissen sie die Antwort und beantworten die Frage richtig.

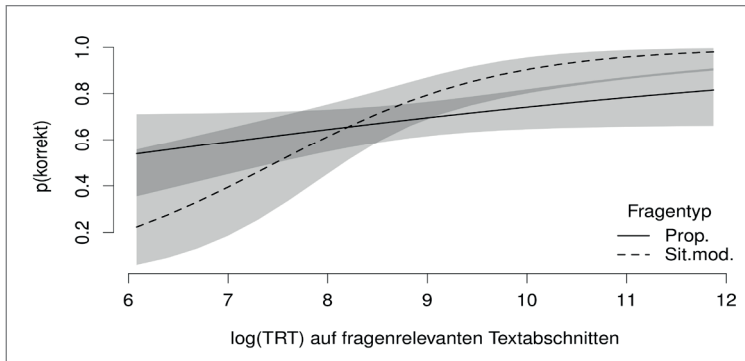


Abb. 6.7: Interaktion der logarithmierten Gesamtlesezeit auf fragenrelevanten Abschnitten im FLRCNat-Teil und dem Fragentyp. Grundlage sind die Schätzwerte aus dem gemischten logistischen Regressionsmodell.

Zwischen diesen beiden Erklärungsversuchen kann an dieser Stelle und mit der gegebenen Datenbasis nicht weiter unterschieden werden. Ein allgemeinerer Punkt ist ohnehin relevanter im Rahmen dieser Arbeit. Die Analysen zeigen nämlich, dass der Zusammenhang zwischen Textverständnis und Lesezeiten nicht trivial ist. Diese Diskussion greift auch zurück auf Abschnitt 3.2 und die von Christmann (2004) eingeführten Konzepte der textnahen und -fernen Verstehensprodukte. Im selben Atemzug wurde auch zwischen Verstehensprozess und -produkt unterschieden. In der präsentierten Analyse wurden beide Ebenen verbunden. Die abhängige Variable gehört dabei der Produktseite an: Die Korrektheit der Antwort auf die Frage(n) nach dem Text wurde vorhergesagt aus der „Vertretervariable“ der Prozess-Seite: der Gesamtlesezeit auf Textregionen, die für die Beantwortung der Frage relevant waren.<sup>103</sup> Offenbar besteht eine gewisse Verbindung zwischen den beiden Variablen. Es gibt relativ deutliche Hinweise darauf, dass längere Lesezeiten auf fragenrelevanten Bereichen zu besseren Ergebnissen bei der jeweiligen Frage führen. Die einfache Gleichung, dass schnellere Lesezeiten immer positiv für das Verstehen eines Textes sind, scheint also die Realität unzulässig zu vereinfachen: Wären schnellere Lesezeiten stets mit flüssigerem und damit besserem Verstehen assoziiert, würde sich dieses Muster so nicht zeigen. Die Antwortperformanz müsste sich dann mit schnelleren Lesezeiten auf fragenrelevanten Regionen erhöhen. Es sei hier allerdings nochmals darauf hin-

<sup>103</sup> Fügt man zu den obigen Modellen die Gesamtlänge des fragenrelevanten Abschnitts hinzu, was eigentlich ein guter Prädiktor für die Gesamtlesezeit sein sollte, zeigt sich, dass die Buchstabenanzahl keinerlei Einfluss mehr hat. Auch die Effektstruktur der Modelle wird nur auf numerischer Ebene verändert – alle signifikanten Effekte bleiben signifikant, alle nicht-signifikanten Effekte bleiben ebenfalls unverändert.

gewiesen, dass die Fragen für den FLRCRef-Teil von Hansen et al. (2006) übernommen wurden. Diese Fragen waren insbesondere dafür gedacht, die Aufmerksamkeit der VersuchsteilnehmerInnen auf einem konstanten Niveau zu halten. Auch die Fragen im FLRCNat-Teil waren nicht explizit für diese Art von Analyse entworfen – trotzdem: Die Effekte gehen sowohl für Entscheidungs- als auch für Multiple-Choice-Fragen in vergleichbare Richtungen. Für die weitere Arbeit bedeutet dies, dass argumentativ getrennt werden muss in Prozess- und Produktebene der Textverständlichkeit. Es werden im Weiteren hauptsächlich Maße als abhängige Variablen dienen, die der Prozessseite zuzurechnen sind. Mit Übertragungen der Schlussfolgerungen auf textferne Verstehensprodukte muss vorsichtig umgegangen werden.

Es muss zum Ende dieses Abschnitts noch ein Nicht-Effekt Erwähnung finden. Neben der Gesamtlesezeit auf fragenrelevanten Abschnitten wurde das Rücksprungverhalten der VersuchsteilnehmerInnen und dessen Korrelation mit dem Antwortverhalten untersucht. Hier können keine Effekte festgestellt werden. Mehr regressive Sakkaden von Wörtern in fragenrelevanten Abschnitten führen weder zu besseren noch zu schlechteren Ergebnissen bei der Beantwortung der jeweiligen Frage.

#### 6.2.4 Zusammenfassung

Die vorgestellten Analysen der Gesamtlesezeiten der Texte sowie der Antwortperformanzen der VersuchsteilnehmerInnen konnten die von Hansen et al. (2006) berichteten Effekte für die Originalauszüge und Reformulierungen nicht bestätigen. Für die Reformulierungen konnte kein Effekt festgestellt werden, der über alle Komplexitätstypen hinweg gleich wirkt. Das gilt sowohl für die Gesamtlesezeit als auch für die Antwortperformanz. Bezüglich der Gesamtlesezeit fallen insbesondere die stark reformulierten syntaktisch komplexen Texte aus dem Rahmen, sie werden von allen Texten am längsten gelesen. Außerdem zeigte sich in einer Nachanalyse, dass der Effekt der Textlänge auf die Lesedauer in langen Texten schwächer wird (im Vergleich zu kürzeren Texten). Anders formuliert: Wenn ein Text schon relativ lang ist, scheint es weniger ausschlaggebend für die Gesamtlesezeit zu sein, ob er noch etwas länger wird. Kommen bei einem kurzen Text noch einige Wörter hinzu, scheint dies größere Auswirkungen zu haben. In Sachen Antwortperformanz sind keine systematischen Unterschiede zwischen den Reformulierungsversionen festzustellen. Dafür zeigten sich Unterschiede für die verschiedenen Komplexitätstypen. Insgesamt waren die Fragen nach syntaktisch komplexen Texten am schwierigsten zu beantworten, was in großen Teilen – aber nicht ausschließlich – an der Länge der Texte liegt.

Während sich die residuale Gesamtlesedauer der kompletten Texte als überaus trennscharfes Maß bezüglich der Textgattung erweist (Zeitungsartikel < Urteile < Pressemitteilungen), lässt sich kein solch klares Muster für die Antwortperformanz feststellen. Die verschiedenen Themenbereiche unterscheiden sich bei Pressemitteilungen und Urteilen relativ stark. Am ehesten kann noch für Zeitungsartikel eine interne Konsistenz der Textgattung beobachtet werden. Nach allen Texten lag die Antwortperformanz über dem Zufallsniveau. Die primäre Funktion der Fragen, dafür zu sorgen, dass die VersuchsteilnehmerInnen die Texte aufmerksam lesen, wurde also offenbar erfüllt. Außerdem zeigte sich, dass Fragen, die nicht auf einzelne Propositionen, sondern auf Inhalte des Situationsmodells abzielten, häufiger korrekt beantwortet wurden. Laut dem in Tabelle 6.9 gezeigten Modell steigt die Wahrscheinlichkeit um 67%, wenn die Frage auf das Situationsmodell abzielte (im Vergleich dazu, wenn sie auf einzelne Propositionen abzielte). Für die Abschätzung dieser *odds* siehe Pampel (2000, S. 21ff.). In Abschnitt 5.1.3 wurden zwar einige Bedenken bezüglich dieser Einteilung der Fragen formuliert, die hier gefundene Effektrichtung ist jedoch zu erwarten: Inhalte des Situationsmodells sollten in der mentalen Repräsentation der VersuchsteilnehmerInnen eine prominentere Stellung einnehmen als Fragen zu einzelnen Propositionen, die gegebenenfalls gar nicht mehr Teil der Textrepräsentation sind. Als weiterer Einflussfaktor könnte auch noch hinzukommen, wie gut die Distraktoren, also die anderen Antwortmöglichkeiten zum Frageninhalt passen, bzw. wie verwechselbar sie im Vergleich zur korrekten Antwort sind. Dies wurde in der vorliegenden Analyse nicht berücksichtigt. In einem letzten Schritt wurde bereits eine erste Verbindung zu den erhobenen Blickbewegungsdaten gezogen. Es wurde gezeigt, dass Fragen eher korrekt beantwortet wurden, wenn relevanten Textstellen mehr Aufmerksamkeit zuteil wurde.



## 7. Blickbewegungsdaten

In Kapitel 4 haben wir uns bereits mit der Erhebung von Blickbewegungsdaten beim Lesen beschäftigt. Zumindest implizit wurde dort auch vermittelt, dass die interessantesten Daten, die Lesekorpora mitbringen, die Blickbewegungen der VersuchsteilnehmerInnen auf dem Textmaterial sind. Mit diesen Daten können wir den Verarbeitungsprozess nachvollziehen, der während des Lesens abläuft. Dies wollen wir im aktuellen Kapitel tun. Hierzu werden zunächst beide Korpusteile zusammengefasst, um eine möglichst umfassende Datenbasis zur Verfügung zu haben. Das Reformulierungskorpus und die vollständigen natürlichen Texte werden dazu zu einem Datensatz unifiziert. Anhand dieses unifizierten Datensatzes werden in Abschnitt 7.1 zunächst die Einflüsse der (sub-)lexikalischen Ebene erfasst. In diesem Abschnitt werden wir einige auf dieser Ebene bereits bekannte Effekte aufgreifen und teilweise zum ersten Mal in Lesekorpora nachweisen. Es wird gezeigt werden, dass manche Effekte, die bisher nur in experimentellen Untersuchungen nachweisbar waren, auch während des Lesens natürlicher Texte gefunden werden können. In Abschnitt 7.3 werden einige Prädiktoren auf syntaktischer Ebene hinzutreten. Außerdem wird auch ein neues Verfahren der Regioneneinteilung in Blickbewegungskorpora eingeführt. Mit dieser Regioneneinteilung anhand der annotierten Phrasenstruktur wird analysiert, ob schnellere Blickbewegungen in syntaktisch tiefer eingebetteten Phrasen als ein Relevanzeffekt gefasst werden kann. Im weiteren Verlauf des Abschnitts werden dann auch ähnlichkeitsbasierte Interferenzeffekte im FLRC untersucht. Abschnitt 7.4 wird sich schließlich der Verarbeitung auf der Textebene widmen. Dort werden wir insbesondere die Auflösung von Anaphern beleuchten und auf einige der in Abschnitt 2.2 vorgestellten Theorien und Modelle zur Referenzauflösung eingehen.

In Abschnitt 7.5 werden Unterschiede in der Verarbeitung der verschiedenen Reformulierungsversionen herausgearbeitet. Ausgangspunkte werden die verschiedenen Komplexitätstypen sein, die für die Rechtssprache allgemein bereits in Abschnitt 3.3 und für das FLRC in Abschnitt 5.2.1 herausgearbeitet wurden. Außerdem werden anhand einer manuellen Annotation von Fernabhängigkeiten zwischen Verben und deren Subjekten Effekte der Dependenzdistanz analysiert.

Das in Abschnitt 6.1 vorgestellte Modellanpassungsverfahren muss für die nun folgenden Analysen leicht modifiziert werden, da die potenzielle Zahl an Effektermen, die in die Modelle aufgenommen werden können, sehr groß ist.



Für Analysen auf der lexikalischen Ebene stehen beispielsweise schon über zwanzig Prädiktoren zur Verfügung. Dies liegt hauptsächlich daran, dass die Wortlänge, residuale Wortfrequenz und residuale Wortfamiliarität für die Wörter  $n+/-1$  und  $n+/-2$  mit aufgenommen werden können und sollten, da gezeigt wurde, dass auch deren Eigenschaften Einfluss auf die Lesezeit von Wort  $n$  haben können (vgl. z.B. Kliegl et al. 2006). Würde man nun ein Modell als Ausgangspunkt verwenden, in dem alle möglichen Interaktionen enthalten sind, würde dies zu extrem unübersichtlichen Modellen mit faktisch uninterpretierbaren  $n$ -fachen Interaktionen führen. Als Lösungsansatz wird von diesem Punkt an folgendes Modellanpassungsverfahren angewendet:

- Es wird ein Referenzmodell berechnet, in dem alle Haupteffekte dieser Ebene (also für Analysen auf der lexikalischen Ebene alle lexikalischen Prädiktoren usw.) sowie *random intercepts* für die Textidentität und die VersuchsteilnehmerInnen enthalten sind. Außerdem enthält das Referenzmodell *random slopes* für VersuchsteilnehmerIn in Abhängigkeit der Position im Experiment und einen Korrelationsparameter.
- Die Zufallseffektstruktur wird gemäß des Vorgehens, das in Abschnitt 6.1 beschrieben wurde, angepasst. Zunächst werden überflüssige Korrelationsparameter, dann *random slopes*, dann *random intercepts* identifiziert.
- Die Effektstruktur der Haupteffekte wird gemäß des in Abschnitt 6.1 beschriebenen Vorgehens angepasst. Haupteffekte, die nicht signifikant zur Verbesserung des Modells beitragen, werden schrittweise ausgeschlossen.
- Es werden dann schrittweise Interaktionen zum Modell hinzugefügt. Interaktionen werden hinzugefügt sofern die relevante Forschungsliteratur Hypothesen zu solchen Interaktionen nahelegen. Dabei werden maximal dreifache Interaktionen hinzugefügt. Geht aus einem *likelihood ratio test* hervor, dass die hinzugefügte Interaktion das Modell nicht signifikant verbessert, wird sie wieder aus dem Modell ausgeschlossen.
- Es gilt das Prinzip, dass ein Haupteffekt im Modell belassen wird, sofern er an einer Interaktion beteiligt ist. Dies gilt auch für Haupteffekte, die nicht signifikant zur Verbesserung des Modells beitragen. Dieses Vorgehen ist aus statistischer Perspektive notwendig (vgl. Field/Miles/Field 2012, S. 350).

Durch dieses hybride Vorgehen mit ein- und ausschließenden Elementen soll erreicht werden, dass zunächst die Varianz in der Kriteriumsvariable, die von einfachen Effekten aufgeklärt wird, auch diesen zugerechnet wird. Erst danach werden gegebenenfalls Interaktionen postuliert, die zusätzliche Varianz aufklären.

## 7.1 Analysen auf lexikalischer Ebene

Als Prädiktoren der lexikalischen Ebene gehen die Wortlänge, die Frequenz und die Familiarität von Wort  $n$ , den Wörtern  $n+/-1$  und den Wörtern  $n+/-2$  ein. Es treten weitere Eigenschaften von Wort  $n$  als Prädiktoren hinzu. Dazu gehört die kumulierte Frequenz aller Wörter, die eine Levenshtein-Distanz von 1 zum Wort haben (ein Maß für die Dichte der lexikalischen Nachbarschaft, in der sich Wort  $n$  befindet) sowie die Position der jeweiligen Bildschirmseite im Experiment. Es werden ferner zwei binäre Prädiktoren eingeführt, die angeben, ob Wort  $n$  das erste oder letzte in seiner jeweiligen Zeile ist. Typischerweise werden die ersten und letzten Wörter einer Zeile aus dem Datensatz entfernt. Meines Erachtens ist das unnötig, weil dies durch die Hinzunahme dieser festen Faktoren kontrolliert werden kann. Als *random intercepts* fungieren VersuchsteilnehmerInnen sowie die Identität des präsentierten Textes. Zusätzlich wird eine *random slope* für VersuchsteilnehmerInnen in Abhängigkeit der Position des Textes im Experiment und ein Korrelationsparameter eingeführt.

Vor der eigentlichen Modellanpassung wird im Vergleich zum Vorgehen in den vorhergehenden Abschnitten eine etwas ausgefeiltere Methode der Ausreißerkorrektur angewendet. Anhand des kompletten Modells wurde für die Residuen dieser Vorhersage ein Boxplot erstellt. Alle Datenpunkte, deren Residuen mehr als einen doppelten Interquartilabstand vom unteren oder oberen Ende der Box entfernt waren, wurden entfernt. Das heißt, dass der Fakt, ob ein Datenpunkt als Ausreißer zählt oder nicht, davon abhängt, ob das assoziierte Residuum ein Ausreißer im Kontext aller Residuen ist. Der Vorteil dieses Vorgehens ist, dass das Gruppierungsproblem umgangen wird. Dieses Gruppierungsproblem besteht darin, dass man Ausreißer eigentlich für die verschiedenen VersuchsteilnehmerInnen getrennt identifizieren sollte, denn man möchte der Tatsache Rechnung tragen, dass manche Menschen schneller lesen als andere. Andererseits müsste man aber beispielsweise auch für verschiedene Wortlängen getrennt Ausreißer korrigieren, denn dass lange Wörter deutlich länger gelesen werden als kurze Wörter ist keine große Überraschung. Im Grunde könnte also nach jedem Faktor, der in ein Modell eingeht, gruppiert werden. Würde man dies tun, würde die Ausreißerkorrektur jedoch unbrauchbar, da die Gruppen, innerhalb derer Ausreißer (beispielsweise über Boxplots) gesucht werden, mit jedem Gruppierungsfaktor kleiner werden. Enthält eine Gruppe nur noch wenige Datenpunkte, wird die Wahrscheinlichkeit immer geringer, dass überhaupt Ausreißer gefunden werden. Die Gruppierungsmethode fällt in diesem Kontext also als Ansatz zur Ausreißeridentifizierung weg.

## 7.1.1 First fixation duration (FFD)

vtn	trial	2.3042e-07	0.00048002			
Residual		1.3273e-01	0.36431564			
		<b>Estimate</b>	<b>Std. Error</b>	<b>t value</b>	<b>p<sub>MCMC</sub></b>	
(Intercept)		5.165e+00	1.731e-02	298.29	0.0001	***
w.len		2.686e-03	5.112e-04	5.25	0.0001	***
res.freq.dlex.lm		-4.380e-03	1.728e-03	-2.53	0.0118	*
res.famil.lm		1.376e-02	2.130e-03	6.46	0.0001	***
res.all.levn.n		-2.374e-02	5.876e-03	-4.04	0.0001	***
trial		3.714e-04	8.520e-05	4.36	0.0002	***
w.len.1l		1.727e-03	5.616e-04	3.08	0.0018	**
res.freq.dlex.lm.1l		-4.690e-03	2.376e-03	-1.97	0.0488	*
w.len.2l		2.122e-03	2.562e-04	8.28	0.0001	***
first.word.row yes		-5.966e-02	3.945e-03	-15.12	0.0001	***
last.word.row yes		-1.059e-01	3.166e-03	-33.45	0.0001	***
w.len:w.len.1l		6.744e-04	6.855e-05	9.84	0.0001	***
res.freq.dlex.lm:w.len.1l		-1.413e-03	2.561e-04	-5.52	0.0001	***
w.len:res.freq.dlex.lm.1l		-7.686e-04	2.993e-04	-2.57	0.0084	**

Tab. 7.1: Angepasstes gemischtes Modell für *first fixation durations*, lexikalische Ebene.

Für logarithmierte *first fixation durations* (FFDs) gelten nach dem oben vorgestellten Identifikationsverfahren 2.420 Datenpunkte (1,1%) als Ausreißer und werden aus der Analyse ausgeschlossen. Das angepasste Modell für FFDs ist in Tabelle 7.1 abgedruckt. Die Effekte der Wortlänge (*w.len*) und der residuellen Wortfrequenz (*res.freq.dlex.lm*) gehen in die jeweils erwartete Richtung. Auch die residuale Familiarität (*res.famil.lm*) scheint bereits auf die FFDs einen Effekt zu haben: Die ersten Fixationen auf Wörtern, zu denen mehr andere Wörter gleicher Länge mit demselben Anfang in Konkurrenz stehen, sind deutlich länger. Ein unerwarteter Effekt ergibt sich für die residuale Anzahl aller Levenshtein-Nachbarn (*res.all.levn.n*). Diese hat einen beschleunigenden Effekt auf die FFDs. Generell werden die VersuchsteilnehmerInnen langsamer, was am positiven Effekt der Position im Experiment *trial* abzulesen ist. Die angepasste Zufallseffekt-Struktur legt jedoch mit der eingeschlossenen *random slope* nahe, dass dieser Effekt über VersuchsteilnehmerIn hinweg variiert. Sowohl die Wortlänge (*w.len.1l*) als auch die residuale Frequenz (*res.freq.dlex.lm.1l*) des Worts *n-1* haben einen Effekt auf die FFDs des Worts *n*. Diese Effekte sind etwas schwächer als jene von Wortlänge und -frequenz für das Wort selbst, gehen aber in dieselbe Richtung. Für das Wort *n-2* hat lediglich

die Wortlänge (*w.len.2l*) noch einen signifikanten Effekt auf die FFDs auf Wort *n*. Für Wörter, die an erster Stelle in der Zeile (*first.word.row*) stehen als auch für jene, die am Zeilenende (*last.word.row*) stehen, sind die FFDs deutlich kürzer. Das ist insbesondere in Kombination mit den Ergebnissen zu *first-pass reading times* interessant und wird an entsprechender Stelle diskutiert. Drei ins Modell aufgenommene Interaktionen zeigen signifikante Effekte.

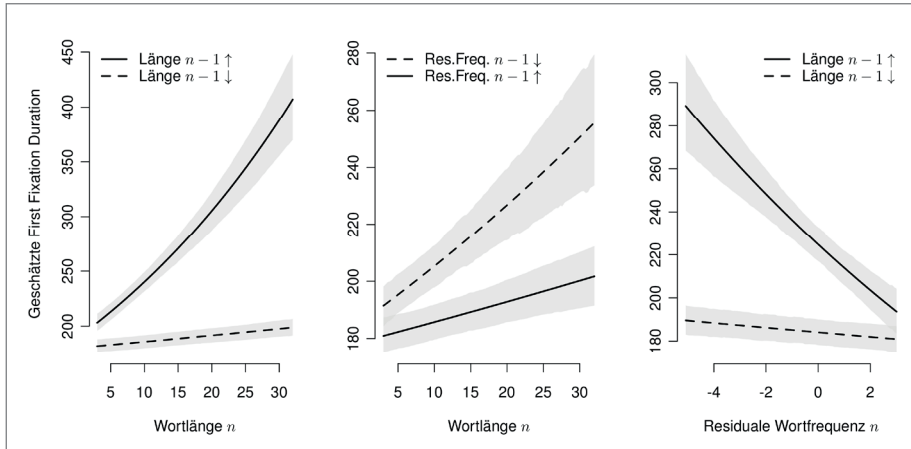


Abb. 7.1: Interaktionseffekte auf die geschätzte *first fixation duration* laut des Modells aus der Tabelle *reftab:mod-ffd-lex* inkl. 95%-Konfidenzintervalle: Wortlänge *n* × Wortlänge *n-1* (links), Wortlänge *n* × res. Wortfrequenz *n-1* (Mitte), res. Wortfrequenz *n* × Wortlänge *n-1*.

Abbildung 7.1 visualisiert diese Interaktionen. Im linken Schaubild ist die Interaktion Wortlänge *n* × Wortlänge *n-1* abgetragen.<sup>104</sup> Es zeigt sich, dass der Effekt der Wortlänge von Wort *n* selbst deutlich stärker ist, wenn das vorherige Wort länger ist. Dasselbe gilt, wenn die residuale Frequenz für Wort *n-1* niedrig ist. Auch der Effekt der residualen Frequenz von Wort *n* selbst wird durch die Wortlänge des vorherigen Worts moduliert: Ist Wort *n-1* länger, ist der Effekt der residualen Frequenz von Wort *n* stärker ausgeprägt.

Die Haupteffekte der Wortlänge und residualen Wortfrequenz sind erwartbar. Dass die residuale Familiarität von Wort *n* bereits auf die FFDs einen verlangsamenden Einfluss hat, bestätigt die These, dass die residuale Familiarität nicht mehr als Vertrautheitsmaß interpretiert werden sollte, sondern als Maß dafür,

<sup>104</sup> Die Visualisierung der Effektschätzer für zwei interagierende Prädiktoren, die beide intervallskaliert sind, ist nicht ganz trivial. Hier wurde der jeweilige Maximal- und Minimalwert für den interagierenden Faktor abgetragen. Da es sich um lineare Modelle handelt, befinden sich die Schätzer für alle anderen Werte zwischen diesen beiden Linien.

wie viele konkurrierende Abrufkandidaten ein Wort hat. Wenn es sich dabei tatsächlich um einen Effekt handelt, der während des Wortabrufs von Bedeutung ist, sollte dieser bereits in frühen Phasen der Verarbeitung wirken.

Die Interaktionen, an denen die Wortlänge des vorherigen Wortes beteiligt ist (in Abb. 7.1 sind diese im linken und rechten Schaubild visualisiert) sind über die rein physikalische Ausdehnung des vorherigen Wortes (Wort  $n-1$ ) zu erklären. Je länger dieses ist, desto weniger von Wort  $n$  befindet sich in der perceptuellen Spanne, die typischerweise 15 Buchstaben rechts von der aktuellen Fixationsposition umfasst (für eine Herleitung und weitere Hinweise hierzu vgl. Rayner 1998, S. 378ff.). McConkie/Rayner (1975) zufolge können nur bis zu einer Grenze von ungefähr zehn Buchstaben einzelne Buchstaben und die Form eines Wortes identifiziert werden. Bis zu 15 Buchstaben rechts von der aktuellen Fixationsposition kann Wortlängeninformation extrahiert werden. Wenn Wort  $n-1$  sehr lang ist, kann Wort  $n$  also nicht vorverarbeitet werden und die Effekte der Eigenschaften von Wort  $n$  schlagen sich erst dann auf die Lesezeit nieder, wenn der Blick auch bei Wort  $n$  ankommt. Die (residuale) Wortfrequenz von Wort  $n$  hat insbesondere dann einen starken Einfluss auf die *first fixation duration*, wenn das Wort nicht bereits während des Lesens von Wort  $n-1$  vorverarbeitet werden konnte (Abb. 7.1, rechts). Die Modulation des Wortlängeneffekts durch die residuale Frequenz des vorherigen Wortes (Abb. 7.1, Mitte) kann auf eine ähnliche Art erklärt werden. Hier ist jedoch davon auszugehen, dass eine Modulation der perceptuellen Spanne im Sinne von Henderson/Ferreira (1990) stattfindet: Ist Wort  $n-1$  seltener, ist es schwieriger zu verarbeiten. Dadurch wird die perceptuelle Spanne kleiner (der Fokus verdichtet sich sozusagen auf die Verarbeitung schwieriger Wörter). Deshalb kann während des Lesens von Wort  $n-1$  weniger von Wort  $n$  vorverarbeitet werden, und der Effekt der Wortlänge verstärkt sich, sobald man im Leseprozess bei Wort  $n$  ankommt. In der Literatur werden diese Art von Effekten oft unter dem Stichwort *parafoveal preview benefit* (vgl. Kennedy 2000; Kennison/Clifton 1995; Rayner et al. 1982) zusammengefasst.

### 7.1.2 *First-pass reading time* (FPRT)

Für logarithmierte FPRTs, also die Zeit, die die LeserInnen beim kompletten ersten Lesen des Wortes benötigen, werden nach der oben beschriebenen Prozedur für die Bereinigung von Ausreißern 1.902 Fälle (0,88%) ausgeschlossen. Das angepasste Modell ist in Tabelle 7.2 einzusehen.

Wieder sind Effekte in die erwarteten Richtungen für die Wortlänge, die residuale Frequenz und Familiarität festzustellen. Die residuale Gesamtfrequenz der Levenshtein-Nachbarn (*res.all.levn.cumfreq*) hat einen Effekt, der in die er-

wartete Richtung zeigt: Je höher die residuale Frequenz aller Levenshtein-Nachbarn ist, desto höher sind die *first pass reading times* auf Wort  $n$ . Die Position im Experiment und die Wortlänge des vorherigen Wortes haben beide keinen signifikanten Effekt, verbleiben aber im Modell, da sie an anderen Effekten beteiligt sind (*trial* als *random slope* in der Zufallseffektstruktur und *w.len.1l* an den Interaktionen mit *w.len* und *res.freq.dlex.lm*). Die residuale Frequenz des vorherigen Wortes (*res.freq.dlex.lm.1l*) hat einen beschleunigenden Effekt auf die FPRTs. Auch die Wortlänge sowie die residuale Frequenz von Wort  $n-2$  (*w.len.2l* und *res.freq.dlex.lm.2l*) haben noch Einfluss auf die Lesezeit auf Wort  $n$ . Wenden wir uns der rechten Seite des aktuellen Wortes zu, können wir außerdem einen beschleunigenden Effekt der Wortlänge (*w.len.1r*), der residualen Wortfrequenz (*res.freq.dlex.lm.1r*) und der residualen Wortfamiliarität (*res.famil.lm.1r*) des Wortes  $n+1$  feststellen. Befindet sich ein Wort am Anfang einer Zeile, wird es länger gelesen, für das Ende der Zeile gilt das Gegenteil. Auch die Textzeile, in der ein Wort steht, hat einen signifikanten beschleunigenden Einfluss auf die *first-pass reading times* eines Wortes.

Die vier signifikanten Interaktionen sind in Abbildung 7.2 dargestellt. Wie für FFDs gilt auch für FPRTs: Ist Wort  $n-1$  länger, wirkt sich der Effekt der Wortlänge von Wort  $n$  stärker aus (oben links). Dasselbe gilt für den Effekt der residualen Frequenz von Wort  $n$  (unten links), der ebenfalls durch die Wortlänge  $n-1$  moduliert wird. Der Wortlängeneffekt wird außerdem moduliert durch die residuale Frequenz von Wort  $n-1$  (oben rechts): Dieser Effekt ist nicht stark ausgeprägt, aber signifikant. Wenn das vorherige Wort eine hohe residuale Frequenz aufweist, ist der Effekt der Wortlänge  $n$  minimal stärker, außerdem überkreuzen sich die beiden Effektschätzer.<sup>105</sup> Der Wortlängeneffekt auf Wort  $n$  wird außerdem noch von der residualen Familiarität des folgenden Wortes moduliert (unten rechts). Weist das folgende Wort einen höheren residualen Familiaritätswert auf, wirkt sich die Wortlänge von Wort  $n$  stärker auf die Verarbeitung aus.

Die einfachen Haupteffekte der Wortlänge sowie residualer Frequenz und Familiarität replizieren die Haupteffekte, die bereits für *first fixation durations* gefunden wurden. Sind die residualen Frequenzen für die vorherigen beiden Wörter niedrig, wird auch Wort  $n$  langsamer gelesen. Dies deutet auf einen ausgeprägten *spill-over*-Effekt hin, der besagt, dass Effekte von Eigenschaften vergangener Wörter auch dann noch ihren Einfluss ausüben können, wenn der Blick schon auf einem Wort weiter rechts im Text ruht (vgl. Henderson/

<sup>105</sup> In den Plots sind rücktransformierte, geschätzte Lesezeiten abgetragen. Die Modelle selbst sind an logarithmierte Lesezeiten angepasst. In den Modellen sind also tatsächlich nur lineare Schätzer enthalten. Durch die Rücktransformation auf die einfacher lesbare Skala entstehen dadurch Kurven in den Schaubildern.

Ferreira 1990; Kennison/Clifton 1995). Es finden sich zudem Effekte, die aus der anderen Richtung wirken, nämlich vom folgenden Wort aus. Je länger das folgende Wort ist, desto **kürzer** wird Wort  $n$  gelesen. Dasselbe gilt für die residuale Frequenz von Wort  $n+1$ . Ist diese höher, wird Wort  $n$  ebenfalls schneller gelesen. Das ist zunächst etwas rätselhaft, könnte aber dadurch erklärt werden, dass Wort  $n+1$  bereits vorverarbeitet wird, während noch Wort  $n$  fixiert wird. Ist die residuale Frequenz  $n+1$  höher, ist diese Vorverarbeitung einfacher, also sinken die FPRTs auf Wort  $n$ . Dieselbe Erklärung kann jedoch nicht für die Wortlänge  $n+1$  gelten, denn dieser Effekt zeigt – wenn man eben dieser Erklärung folgt – in die falsche Richtung. Die Erklärung, die hier heranzuziehen wäre, wäre folgende: Ist Wort  $n+1$  länger, liegen größere Teile außerhalb der parafovealen Vorschau, während man sich noch bei Wort  $n$  befindet. Das Wort ist also unter Umständen gar nicht voll identifizierbar. Ist dies der Fall, kann Wort  $n+1$  also nicht vorverarbeitet werden, was wiederum zu einer

Groups	Name	Variance	Std.Dev.			
text.id	(Intercept)	1.1988e-03	0.03462406			
vtn	(Intercept)	2.5056e-02	0.15829206			
vtn	trial	3.7232e-07	0.00061018			
Residual		1.8780e-01	0.43335737			
		<b>Estimate</b>	<b>Std. Error</b>	<b>t value</b>	<b><math>p_{MCMC}</math></b>	
(Intercept)		5.104e+00	2.022e-02	252.38	0.0001	***
w.len		3.692e-02	6.661e-04	55.42	0.0001	***
res.freq.dlex.lm		-5.018e-03	2.396e-03	-2.09	0.0328	*
res.famil.lm		3.722e-02	2.797e-03	13.31	0.0001	***
res.all.levn.cumfreq		7.795e-03	1.696e-03	4.60	0.0001	***
trial		1.893e-04	1.090e-04	1.74	0.0890	.
w.len.1l		8.708e-05	7.388e-04	0.12	0.9170	
res.freq.dlex.lm.1l		-1.520e-02	3.113e-03	-4.88	0.0001	***
w.len.2l		2.927e-03	3.455e-04	8.47	0.0001	***
res.freq.dlex.lm.2l		-5.508e-03	1.412e-03	-3.90	0.0001	***
w.len.1r		-1.721e-03	3.745e-04	-4.59	0.0001	***
res.freq.dlex.lm.1r		-3.704e-03	1.497e-03	-2.47	0.0138	*
res.famil.lm.1r		-3.098e-02	6.011e-03	-5.15	0.0001	***
first.word.row yes		1.501e-01	5.185e-03	28.95	0.0001	***
last.word.row yes		-1.537e-01	4.412e-03	-34.84	0.0001	***
text.row		-7.306e-03	7.309e-04	-10.00	0.0001	***
w.len:w.len.1l		7.135e-04	8.853e-05	8.06	0.0001	***
w.len:res.freq.dlex.lm.1l		9.317e-04	3.865e-04	2.41	0.0142	*
res.freq.dlex.lm:w.len.1l		-1.594e-03	3.448e-04	-4.62	0.0001	***
w.len:res.famil.lm.1r		3.062e-03	6.899e-04	4.44	0.0001	***

Tab. 7.2: Angepasstes gemischtes Modell für *first-pass reading times*, lexikalische Ebene.



kürzeren FPRT auf Wort  $n$  führt. Führt man diese Erklärung an, macht man gleichzeitig eine Vorhersage: Sollte es so sein, dass keine Vorverarbeitung von längeren Wörtern  $n+1$  stattfindet, sollte die Wortfrequenz dieser Wörter auch keinen Effekt auf die Lesezeit von Wort  $n$  haben. In der Tat finden sich in einer Nachanalyse Hinweise, die diese Hypothese zu bestätigen scheinen. Berechnet man dasselbe Modell wie in Tabelle 7.2 nur für Wörter  $n+1$  mit einer Länge mit mehr als fünf Buchstaben, verschwindet der Effekt der residualen Wortfrequenz  $n+1$  ( $t = 0,90$ ).<sup>106</sup>

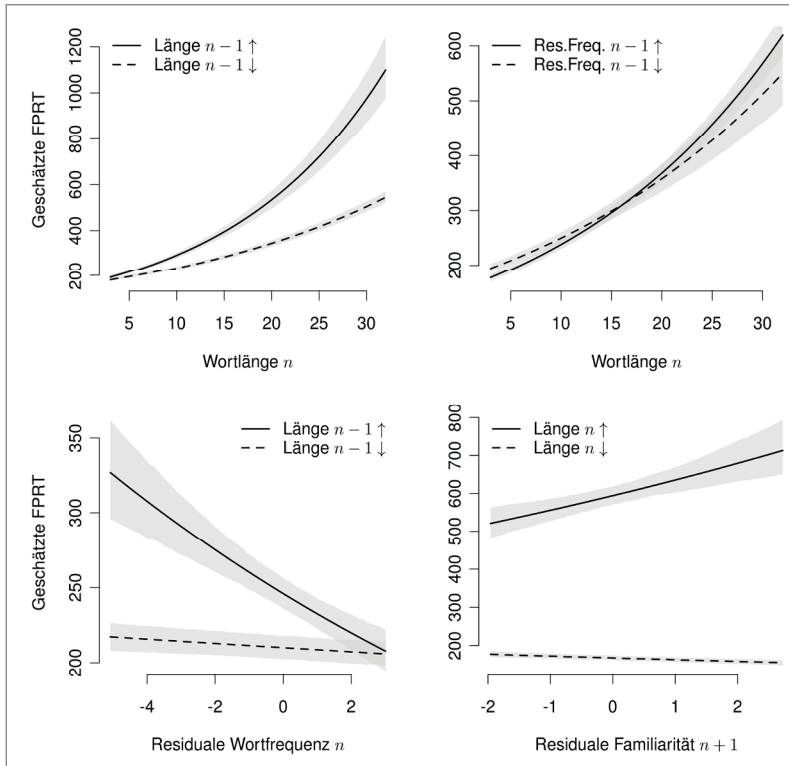


Abb. 7.2: Interaktionseffekte auf die geschätzte *first-pass reading time* laut des Modells aus der Tabelle reftab:mod-fprt-lex inkl. 95%-Konfidenzintervalle: Wortlänge  $n \times$  Wortlänge  $n-1$  (oben links), Wortlänge  $n \times$  res. Frequenz Wort  $n-1$  (oben rechts), res. Frequenz Wort  $n \times$  Wortlänge  $n-1$  (unten links), res. Familiarität Wort  $n+1 \times$  Wortlänge  $n$  (unten rechts). FPRTs sind zurücktransformiert aus logarithmierten Zeiten.

<sup>106</sup> Dies unterstützt die vorstehende Erklärung nicht bedingungslos. Erstens kann man nicht davon ausgehen, dass ein Effekt tatsächlich nicht vorhanden ist, wenn er statistisch nicht signifikant ist – im Grunde dürfen Nichteffekte nicht interpretiert werden. Zweitens ist der *cut-off* bei fünf Buchstaben relativ willkürlich. Es zeigt sich jedoch, dass für *cut-offs* bei sechs und sieben Buchstaben ebenfalls kein Effekt mehr festzustellen ist.

Interessant ist weiterhin, dass im Kontrast zu *first fixation durations*, wo erste Wörter in der Zeile noch mit niedrigeren Lesezeiten assoziiert waren, sich dieser Effekt nun umkehrt. Dies könnte an folgendem Muster liegen: Die meisten Blicksprünge auf Wörter am Anfang einer Zeile kommen aus der vorherigen Zeile. Da dies relativ lange Sakkaden sind, ist die Wahrscheinlichkeit größer, dass es zu Fehlfixationen kommt. Das würde bedeuten, dass die Leserin/der Leser zwar das erste Wort der Zeile trifft, aber nicht an der sogenannten *preferred viewing location* (vgl. McConkie et al. 1988; Rayner 1979; Vitu et al. 1990). In solchen Fällen muss dann gegebenenfalls sogleich eine Korrektursakkade „nachgeschoben“ werden. Diese initiale Fehlfixation scheint im Vergleich zu den anderen ersten Fixationen auf Wörtern sehr kurz – daher die kürzeren *first fixation durations* für Wörter, die am Zeilenanfang stehen. Dass die *first-pass reading times* für diese Wörter länger sind, ist dann nicht mehr überraschend. Denn sowohl die initiale Fehlfixation als auch alle nachfolgenden Fixationen auf dem Wort, bis es zum ersten Mal verlassen wird, gehen in die *first-pass reading times* mit ein. Zusätzlich könnte noch ein anderer Effekt wirken, der wiederum die perzeptuelle Spanne als Erklärung mit einbezieht. Steht ein Wort am Zeilenanfang, kann es nicht vorverarbeitet werden, während der Blick noch auf dem vorherigen Wort ruht – denn das vorherige Wort liegt eine Zeile weiter oben am rechten Rand der Zeile.

Die für FPRTs gefundenen Interaktionen erinnern teilweise an jene für FFDs. Wiederum hat die Wortlänge des aktuellen Worts einen stärkeren Einfluss auf die Lesezeit, wenn das vorherige Wort länger ist (Abb. 7.2, oben links). Auch die Interaktion zwischen residualer Frequenz  $n$  und Länge  $n-1$  kann für FPRTs repliziert werden (Abb. 7.2, unten links). Etwas rätselhaft erscheint der relativ schwache Interaktionseffekt zwischen der Wortlänge  $n$  und der residualen Frequenz  $n-1$  (Abb. 7.2, oben rechts). Die Wortlänge  $n$  hat einen stärkeren Einfluss, wenn die residuale Frequenz  $n-1$  hoch ist (Wort  $n-1$  also einfacher zu verarbeiten sein sollte). Im Sinne einer dynamischen perzeptuellen Spanne wäre ein Effekt in die andere Richtung zu erwarten. Woher dieser Interaktionseffekt stammt, ist momentan nicht zu klären. Einschränkend muss angemerkt werden, dass der fragliche Effekt im Vergleich zu den anderen relativ schwach ausgeprägt ist. In die vierte Interaktion sind die residuale Familiarität des jeweils folgenden Worts und die Wortlänge des aktuellen Worts involviert (unten rechts). Offenbar hat die residuale Familiarität des folgenden Wortes erst dann den erwarteten Einfluss, wenn Wort  $n$  lang ist. Die Effektgröße für kurze Wörter  $n$  liegt lediglich bei 22 ms, während sie für längere Wörter  $n$  192 ms beträgt. Eine Erklärung hierfür könnte der folgende Prozess sein: Wenn Wort  $n$  lang ist, ist davon auszugehen, dass LeserInnen mehrere Fixationen benötigen, um das komplette Wort zu lesen. Bei Fixa-

tionen im hinteren Bereich des Wortes  $n$  ist es dann wahrscheinlicher, dass Wort  $n+1$  parafoveal vorverarbeitet wird und die residuale Familiarität dieses Worts ihren erwarteten Einfluss nimmt. Das würde erklären, warum dieser Effekt für *first fixation durations* nicht zu beobachten ist.

### 7.1.3 Regression path duration (RPD)

vtn	(Intercept)	4.6447e-02	0.2155146			
vtn	trial	2.4729e-06	0.0015725			
Residual		4.4053e-01	0.6637233			
		<b>Estimate</b>	<b>Std. Error</b>	<b>t value</b>	<b><math>p_{\text{MCMC}}</math></b>	
(Intercept)		5.5141742	0.0303354	181.77	0.0001	***
w.len		0.0486240	0.0006175	78.74	0.0001	***
res.freq.dlex.lm		-0.0184989	0.0026471	-6.99	0.0001	***
res.famil.lm		0.0270060	0.0086356	3.13	0.0016	**
res.all.levn.cumfreq		0.0067263	0.0028483	2.36	0.0192	*
trial		-0.0003176	0.0002537	-1.25	0.2082	
w.len.1l		-0.0015714	0.0005990	-2.62	0.0088	**
res.freq.dlex.lm.1l		-0.0152880	0.0025947	-5.89	0.0001	***
res.famil.lm.1l		0.0103676	0.0051934	2.00	0.0494	*
w.len.2l		0.0023080	0.0005966	3.87	0.0001	***
res.freq.dlex.lm.2l		-0.0081834	0.0024984	-3.28	0.0008	***
res.famil.lm.2l		0.0150336	0.0050672	2.97	0.0032	**
w.len.1r		-0.0054360	0.0006354	-8.55	0.0001	***
res.freq.dlex.lm.1r		-0.0117174	0.0026217	-4.47	0.0001	***
res.famil.lm.1r		-0.0438286	0.0101776	-4.31	0.0001	***
first.word.row yes		0.0497042	0.0086702	5.73	0.0001	***
last.word.row yes		-0.1725892	0.0073958	-23.34	0.0001	***
text.row		-0.0121491	0.0012510	-9.71	0.0001	***
w.len.res.famil.lm.1r		0.0026592	0.0011547	2.30	0.0230	*
res.famil.lm:w.len.1l		0.0024242	0.0010815	2.24	0.0260	*

Tab. 7.3: Angepasstes gemischtes Modell für *regression path durations*, lexikalische Ebene.

7.151 Fälle (3,32%) wurden bei den logarithmierten RPDs als Ausreißer identifiziert und aus der Analyse ausgeschlossen. Tabelle 7.3 zeigt das angepasste Modell für logarithmierte RPDs.

Wortlänge, residuale Frequenz und residuale Familiarität zeigen die erwarteten Effekte. Die kumulierte Anzahl aller Levenshtein-Nachbarn zeigt wie schon für FPRTs einen verlangsamenden Effekt. Die Position im Experiment verbleibt wiederum nur deshalb im Modell, weil sie an der Zufallseffektstruktur beteiligt ist. Während die Länge des vorherigen Wortes bei FPRTs noch keinen Effekt zeigte, ergibt sich für RPDs ein Effekt in eine unerwartete Rich-

tung: Je länger Wort  $n-1$  ist, desto **kürzer** ist der Regressionspfad für Wort  $n$ . Dieser Prädiktor ist zudem an einer Interaktion beteiligt, weshalb er später interpretiert wird. Die residuale Frequenz und Familiarität von Wort  $n-1$  zeigen erwartbare Effekte auf die RPDs. Der Effekt der residualen Frequenz  $n-1$  war bereits für FPRTs zu beobachten. Alle Effekte, die mit dem Wort  $n-2$  assoziiert sind (*w.len.2l*, *res.freq.dlex.lm.2l* und *res.famil.lm.2l*), gehen in die jeweils erwarteten Richtungen. Alle Effekte für das folgende Wort (*w.len.1r*, *res.freq.dlex.lm.1r* und *res.famil.lm.1r*) gehen in die jeweils gleichen Richtungen wie für FPRTs. Dasselbe gilt für die Prädiktoren, die mit der Position des Wortes in der Zeile (*first.word.row*, *last.word.row*) bzw. auf dem Bildschirm (*text.row*) zusammenhängen. Zwei Interaktionen zeigen einen signifikanten Effekt auf die RPDs (siehe Abb. 7.3).

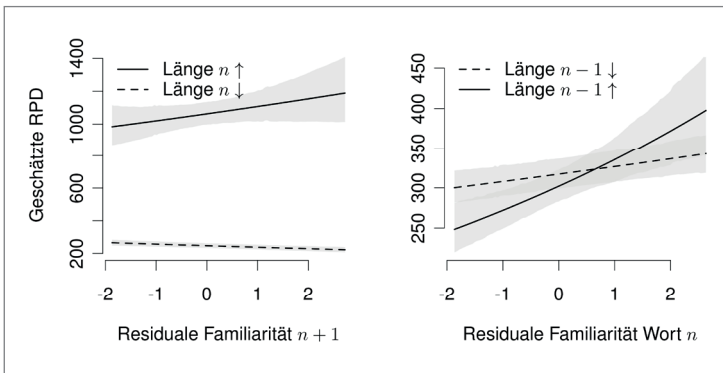


Abb. 7.3: Interaktionseffekte auf die geschätzte *regression path duration* laut des Modells aus Tabelle 7.4 inkl. 95%-Konfidenzintervalle: Wortlänge  $n \times$  res. Familiarität  $n+1$  (links) und res. Familiarität  $n \times$  Wortlänge  $n-1$  (rechts). RPDs sind zurücktransformiert aus logarithmierten Zeiten.

Die Interaktion zwischen residualer Familiarität des jeweils folgenden Wortes mit der Wortlänge des aktuellen Wortes (Abb. 7.3, links) ist ebenfalls sehr ähnlich zur Interaktion, die für FPRTs beobachtet werden konnte. Außerdem wird der Effekt der residualen Familiarität durch die Länge des vorherigen Wortes moduliert (rechts): Es kann gezeigt werden, dass die Familiarität des aktuellen Wortes dann einen höheren Einfluss auf die RPDs hat, wenn das vorherige Wort länger ist.

Im Lichte der vorangegangenen Abschnitte sind die für RPDs gefundenen Haupteffekte nicht sehr überraschend. Lediglich der beschleunigende Effekt für lange Wörter  $n-1$  stört das Gesamtbild etwas. Warum längere Wörter  $n-1$  zu kürzeren *regression path durations* auf Wort  $n$  führen sollten, ist nur schwer

zu klären. Insbesondere die Umkehrung der Effektrichtung im Vergleich zu FFDs und FPRTs ist sonderbar. Nachanalysen legen nahe, dass der Effekt weder auf bestimmte Wortklassen (Inhaltswörter vs. Funktionswörter) noch auf eine systematische Ungleichverteilung der Wortlängen auf den Satz- oder Seitenanfang zurückzuführen ist.

Die Interaktionen (siehe Abb. 7.3) sind einfacher erklärbar. Bei jener zwischen residualer Familiarität  $n+1$  und Länge  $n$  (links) sei auf die nicht abschließend zu bestätigende Erklärung aus dem letzten Abschnitt verwiesen: Die Familiarität des folgenden Wortes kann eventuell erst dann ihren Einfluss bereits bei der Verarbeitung des aktuellen Wortes „ausspielen“, wenn das aktuelle Wort länger ist und daher mehrere Fixationen im Wort nötig sind. Werden dann hintere Regionen von Wort  $n$  fixiert, hat die Familiarität von Wort  $n+1$  Einfluss. Die zweite Interaktion (Abb. 7.3, rechts) deutet auf den schon mehrfach gezeigten Effekt hin, dass bei längeren Wörtern  $n-1$  das Wort  $n$  nicht mehr in der perzeptuellen Spanne liegt und daher die Familiarität erst dann einen starken Einfluss hat, wenn der Leser bei Wort  $n$  anlangt.

#### 7.1.4 Total reading time (TRT)

418 Fälle (0,2%) der logarithmierten *total reading times* wurden als Ausreißer identifiziert und von der Analyse ausgeschlossen.

Das angepasste Modell (siehe Tab. 7.4) zeigt deutliche Effekte für Wortlänge, Wortfrequenz und Wortfamiliarität. Diese Effekte zeigen – gegeben die vorhergehenden Analysen – in die erwartete Richtung. Die Effekte der Wortlänge und der Wortfrequenz werden durch Eigenschaften des vorherigen Wortes moduliert, was im Modell über die signifikanten Interaktionen zum Ausdruck kommt. Die Wortlänge des vorhergehenden Wortes hat – wie schon für *regression path durations* – einen beschleunigenden Effekt, ebenso die residuale Frequenz von Wort  $n-1$ . Je länger Wort  $n-2$  ist, desto länger wird Wort  $n$  gelesen. Ist die residuale Frequenz von Wort  $n-2$  höher, fällt die TRT für Wort  $n$ . Für das jeweils folgende Wort  $n+1$  gilt: Mit steigender Länge und steigender residualer Familiarität  $n+1$  fällt die TRT für Wort  $n$ . Diese Effekte konnten sowohl für FPRTs als auch für RPDs bereits beobachtet werden. Die physikalische Position des Wortes auf dem Bildschirm hat ebenfalls einen Einfluss. Mit fortschreitender Zeilenzahl sinkt die TRT, ebenso wenn ein Wort am Zeilenende steht. Drei Interaktionen haben signifikante Effekte auf die Gesamtlesezeit auf einem Wort (siehe Abb. 7.4). Die Wortlänge des Wortes  $n$  hat einen stärkeren Einfluss auf die TRT, wenn das Wort davor länger ist (Abb. 7.4, links). Auch die residuale Frequenz des Wortes  $n$  hat einen starken

Einfluss in die erwartete Richtung, wenn das vorherige Wort länger ist (Mitte). Der dritte Interaktionseffekt erfasst die Modulation des Frequenzeffekts für Wort  $n$ : Die residuale Frequenz von Wort  $n$  hat einen stärkeren Einfluss auf die Gesamtlesezeit, wenn das vorherige Wort hochfrequent war.

Groups	Name	Variance	Std.Dev.			
text.id	(Intercept)	2.4131e-02	0.1553420			
vtn	(Intercept)	4.9521e-02	0.2225326			
vtn	trial	2.0193e-05	0.0044937			
Residual		3.2665e-01	0.5715346			
		<b>Estimate</b>	<b>Std. Error</b>	<b>t value</b>	<b><math>p_{\text{MCMC}}</math></b>	
(Intercept)		5.5647678	0.0344428	161.57	0.0001	***
w.len		0.0601879	0.0007032	85.58	0.0001	***
res.freq.dlex.lm		-0.0206137	0.0027470	-7.50	0.0001	***
res.famil.lm		0.0420184	0.0033130	12.68	0.0001	***
res.all.leavn.n		-0.0612615	0.0091298	-6.71	0.0001	***
trial		-0.0028336	0.0005475	-5.18	0.0001	***
w.len.l1		-0.0035520	0.0008096	-4.39	0.0001	***
res.freq.dlex.lm.l1		-0.0104279	0.0016696	-6.25	0.0001	***
w.len.l2		0.0020689	0.0003953	5.23	0.0001	***
res.freq.dlex.lm.l2		-0.0073493	0.0016268	-4.52	0.0001	***
w.len.lr		-0.0010811	0.0004090	-2.64	0.0070	**
res.famil.lm.lr		-0.0072320	0.0034208	-2.11	0.0368	*
last.word.row.yes		-0.2448102	0.0051984	-47.09	0.0001	***
text.row		-0.0237826	0.0008467	-28.09	0.0001	***
w.len:w.len.l1		0.0005016	0.0000997	5.03	0.0001	***
res.freq.dlex.lm:w.len.l1		-0.0025330	0.0003949	-6.41	0.0001	***
res.freq.dlex.lm:res.freq.dlex.lm.l1		-0.0047763	0.0015388	-3.10	0.0022	**

Tab. 7.4: Angepasstes gemischtes Modell für *total reading times*, lexikalische Ebene.

Die Haupteffekte für TRTs sind im Allgemeinen wenig überraschend. Einige Eigenschaften des Wortes  $n-2$  haben einen Einfluss auf die Verarbeitung von Wort  $n$ . Effekte in die entsprechenden Richtungen konnten bereits für FPRTs, RPDs und teilweise für FFDs (dort aber nur die Wortlänge  $n-2$ , siehe Tab. 7.1) nachgewiesen werden. Der beschleunigende Effekt für längere Wörter  $n+1$  wurde bereits für *first-pass reading times* diskutiert. Die Interaktion zwischen der Länge des aktuellen Wortes und jener des vorherigen Wortes (siehe Abb. 7.4, links) war bereits für FFDs und FPRTs signifikant und zeigte dasselbe Muster. Auch die Interaktion zwischen der residualen Frequenz von Wort  $n$  und der Länge des vorhergehenden Wortes konnte bereits für FPRTs beobachtet werden. Das gilt nicht für die Interaktion, die in Abbildung 7.4 rechts abgetragen ist. Hier zeigt sich, dass die residuale Frequenz von Wort  $n$  dann einen stärkeren

ren Einfluss hat, wenn Wort  $n-1$  hochfrequent ist. Das steht zunächst im Gegensatz zur Auffassung einer dynamischen Anpassung der perzeptuellen Spanne. Folgt man dieser Auffassung, müsste sich die perzeptuelle Spanne bei der Verarbeitung eines hochfrequenten Wortes  $n-1$  ausdehnen, denn hochfrequente Wörter sind einfacher zu verarbeiten. Ist die perzeptuelle Spanne größer, müsste Wort  $n$  bereits vorverarbeitet werden können. Der Effekt der Frequenz  $n$  sollte dann, sobald man bei Wort  $n$  angekommen ist, **schwächer** sein. Das ermittelte Interaktionsmuster ist aber umgekehrt. Unter einer Bedingung ergibt diese Interaktion einen Sinn: Wenn Wort  $n-1$  hochfrequent ist, steigt die Häufigkeit, dass es beim Lesen komplett übersprungen wird. Wenn dies der Fall ist, kann Wort  $n$  nicht vorverarbeitet werden, weil die nächste Sakkade direkt in Wort  $n$  selbst fällt – genau dann hat aber die residuale Wortfrequenz von  $n$  auch einen stärkeren Einfluss auf die Verarbeitungszeit. Eine Nachanalyse bestätigt dieses Muster. Schränkt man das in Tabelle 7.4 gezeigte Modell auf jene Fälle ein, wo Wort  $n-1$  **nicht** übersprungen wird, verschwindet als einzige Interaktion die kritische ( $res.freq.dlex.lm \times res.freq.dlex.lm.1l$ ,  $t = -0,39$ ). Schränkt man das Modell auf die Fälle ein, wo Wort  $n-1$  tatsächlich übersprungen wird, bleibt die Interaktion bestehen ( $t = -5,51$ ). Das spricht stark dafür, dass es jene Fälle sind, in denen Wort  $n-1$  übersprungen wird, die für den kritischen Interaktionseffekt hauptsächlich verantwortlich sind.

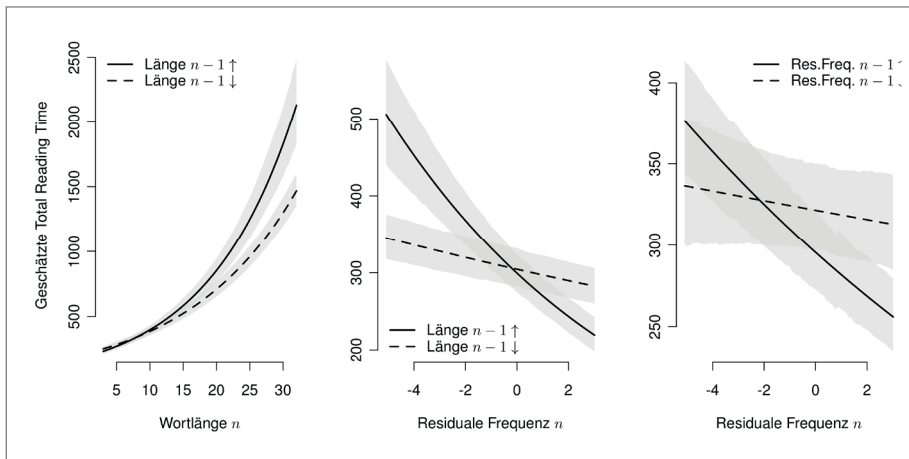


Abb. 7.4: Interaktionseffekte auf die geschätzte *total reading time* laut des Modells aus der Tabelle *reftab:mod-trt-lex* inkl. 95%-Konfidenzintervalle: Wortlänge  $n \times$  Wortlänge  $n-1$  (links), res. Frequenz  $n \times$  Wortlänge  $n-1$  (Mitte) sowie res. Frequenz  $n \times$  res. Frequenz  $n-1$  (rechts). TRTs sind zurücktransformiert aus logarithmierten Zeiten.



## 7.1.5 Zusammenfassung und Evaluation

	FFD	FPRT	RPD	TRT
Länge	+	+	+	+
Res.Freq.	-	-	-	-
Res.Famil.	+	+	+	+
kumFreq. Lv	0	+	+	0
Anzahl Lv	-	0	0	-
Länge n - 1	+	+	-	-
Res.Freq.n - 1	-	-	-	-
Res.Famil.n - 1	+	0	+	0
Länge n - 2	+	+	+	+
Res.Freq. n - 2	0	-	-	0
Res.Famil. n - 2	0	0	+	0
Länge n + 1	0	-	-	-
Res.Freq. n + 1	0	-	-	0
Res.Famil. n + 1	0	-	-	-
Länge n + 2	0	0	0	0
Res.Freq. n + 2	0	0	0	0
Res.Famil. n + 2	0	0	0	0
Erstes Wort in Zeile	-	+	+	0
Letztes Wort in Zeile	-	-	-	-
Textzeile	0	-	-	-

Tab. 7.5: Übersicht der einfachen Effekte auf die verschiedenen Lesezeitmaße auf lexikalischer und Darstellungsebene.

Tabelle 7.5 gibt einen Überblick über die durchgeführten Analysen auf lexikalischer Ebene. So wird auf einen Blick sichtbar, wo Effekte konsistent über verschiedene Lesezeitmaße hinweg auftreten und wo widersprüchliche Effekte gefunden wurden. In die Tabelle geht lediglich das Vorzeichen des jeweiligen Effektschätzers für den in der ersten Spalte angegebenen Haupteffekt ein. Dabei sind mehrere Dinge zu beachten. 1) Eine 0 in Tabelle 7.5 bedeutet nicht, dass ein solcher Effekt nicht besteht. Eine 0 bedeutet hier lediglich, dass kein solcher Haupteffekt innerhalb des unifizierten FLRC-Datensatzes nachgewiesen werden konnte. 2) Sollten die abgetragenen Hauptef-

fekte in einer oder mehreren Interaktionen beteiligt sein, ist dies in der Tabelle nicht beachtet. Es sei diesbezüglich auf den entsprechenden Textabschnitt verwiesen. Eine Interaktion kann unter Umständen dazu führen, dass der Haupteffekt selbst nicht mehr interpretierbar ist (vgl. Field et al. 2012, S. 531).

Tabelle 7.5 kann auf mehrere Arten gelesen werden. Betrachtet man nur eine Spalte von oben nach unten, lässt sich erkennen, für welche Effekte sich ein bestimmtes Lesezeitmaß sensitiv zeigt. So erkennt man beispielsweise auf einen Blick, dass für FFDs kein einziger Effekt folgender Wörter ( $n+1/2$ ) beobachtbar ist, wohingegen immerhin die Eigenschaften von Wort  $n+1$  noch Effekte auf FPRTs, RPDs und teilweise auf TRTs zeigen. Dass keine *successor effects* (vgl. Kliegl et al. 2006) auf die Dauer der ersten Fixation nachgewiesen werden konnten, ist plausibel.

Man kann Tabelle 7.5 auch zeilenweise lesen. So lässt sich schnell erfassen, auf welche Maße ein bestimmter Faktor einen Einfluss hat und wie konsistent dieser in seiner Richtung ist. Die ersten drei Zeilen zeigen beispielweise, wie konsistent die Effekte von Wortlänge, Wortfrequenz und Wortfamiliarität wirken (die letzten beiden Maße sind jeweils residualisiert). Das war vorherzusehen und ist insofern beruhigend, als dass das Leseverhalten der VersuchsteilnehmerInnen im FLRC vergleichbar ist mit bereits existierenden Lesekorpora (vgl. Kliegl et al. 2004, 2006; Pynte/Kennedy 2006). Der verlangsamende Effekt der Familiarität wurde von mir bereits an anderem Ort (vgl. Wolfer et al. 2012b) dementsprechend interpretiert, dass residuale Familiarität kein Vertrautheitsmaß mehr darstellt, sondern ein Maß für die Interferenz von benachbarten Wörtern mit einer ähnlichen Form (gleiche Länge und das gleiche initiale Buchstabentrigramm). An dieser Stelle wollen wir nochmals auf diese Hypothese eingehen und die in Wolfer et al. (2012b) gezeigten Analysen vertiefen. Die Wortfamiliarität, wie sie hier gefasst und aus *dlexDB* (Heister et al. 2011) extrahiert wurde, definiert sich durch die kumulierte Frequenz aller Wörter, die das initiale Zeichentrigramm und die Wortlänge mit Wort  $n$  teilen. Unter dem Gesichtspunkt, dass die Familiarität fassen soll, wie oft LeserInnen in Texten ungefähr einer solchen Wortform begegnen, ist es berechtigt, dass auch die eigene Frequenz in das Maß mit eingeht. Über die Residualisierung wird nun hauptsächlich genau diese Komponente der eigenen Frequenz herausgerechnet. Dabei geht man aber das Risiko ein, dass durch die Residualisierung unvorhergesehene Effekte entstehen, die eventuell die statistischen Ergebnisse verzerren könnten. Dies lässt sich aber im vorliegenden Fall vermeiden. *dlexDB* stellt auch die **absolute** Familiarität und Frequenz von Wörtern bereit. In den bisher präsentierten Auswertungen waren diese Maße nicht zu gebrauchen, weil Frequenzen (und damit auch Familiaritätsmaße) typischerweise an der Korpusgröße normalisiert werden und

logarithmiert sind. Bei der vorliegenden Sachfrage kann es aber durchaus interessant sein, diese Maße miteinander zu verrechnen, denn die kumulierte Frequenz aller „Wettbewerber“ ergibt sich logischerweise durch die Differenz des absoluten Familiaritätsmaßes und der absoluten Frequenz des Wortes. Wörter, die für diese Differenz hohe Werte haben, haben viele bzw. hochfrequente Nachbarn. Die Differenzbildung ist zulässig, weil die zugrunde liegende Korpusgrundlage von *dlexDB* für beide Maße gleich groß ist.

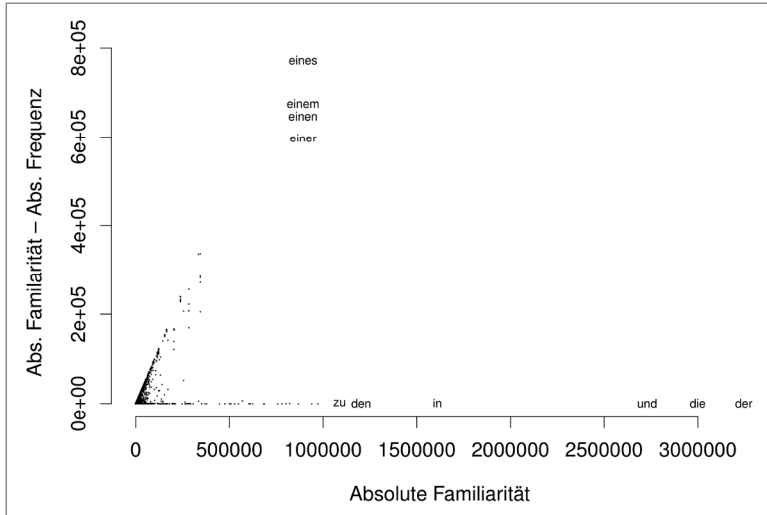


Abb. 7.5: Streudiagramm der absoluten Familiarität aller Wörter im FLRC gegen die Differenz aus der absoluten Familiarität und der absoluten Frequenz. Nur einige wenige Extrempunkte sind ausgeschrieben. Die anderen Wörter sind über Punkte symbolisch abgetragen.

Abbildung 7.5 gibt einen Überblick über den Zusammenhang der absoluten Familiarität und der Differenz zwischen absoluter Familiarität und absoluter Frequenz. Es sind einige interessante Dinge zu beobachten: Die Datenbasis scheint sich sehr stark aufzuspalten in zwei Arten von Wörtern: Jene, bei der die Familiarität fast nur aus der eigenen Frequenz besteht (die besten Beispiele sind hier die ausgeschriebenen Wörter ‘und’, ‘die’ und ‘der’. Die andere Gruppe von Wörtern schlägt sich in der nach rechts oben strebenden Punktwolke nieder, deren äußerster Rand aus dem Cluster ‘eines/einem/einen/einer’ besteht. Die Wörter dieses Clusters verhelfen sich **gegenseitig** zu einem hohen Differenzwert. Nur in relativ kleinen Bereichen finden sich mehrere Wörter abseits dieser beiden Hauptlinien im Diagramm. Das neu berechnete Maß fasst genau das, worauf die von uns (vgl. Wolfer et al. 2012b) zuvor formulierte Hypothese abzielte: Die kumulierte Frequenz der lexikalischen

Wettbewerber **ohne** die Frequenz des Wortes selbst. Die Residualisierung erübrigt sich insofern, dass dieses Differenzmaß nicht mit der absoluten Frequenz korreliert ( $r = 0,006$ ;  $p = 0,75$ ).<sup>107</sup> In die Nachanalyse geht das Differenzmaß logarithmiert ein. Wir wollen nun die Folgen dieser Verteilung auf die Verarbeitung analysieren und insbesondere wissen, ob diese vergleichbar sind mit den bereits vorgestellten Analysen, die anhand der residualen Familiarität durchgeführt wurden. In diesem Falle sollte der Verarbeitungsaufwand steigen, wenn das Differenzmaß steigt.

Als Datenbasis für diese Überprüfung dienen alle Wörter, für die die Differenz zwischen absoluter Familiarität und absoluter Frequenz größer 0 ist. Damit fallen alle Fälle weg, die in Abbildung 7.5 genau auf der  $x$ -Achse liegen. Diese Einschränkung wurde gewählt, um tatsächlich nur jenen Bereich an Werten abzudecken, in dem tatsächlich auch Varianz in der Datenmenge vorhanden ist. Außerdem stellt es einen qualitativen Unterschied dar, ob ein Wort im Sinne des hier verwendeten Familiaritätsmaßes überhaupt keine Nachbarn hat, der Differenzwert Familiarität minus Frequenz also 0 ist, oder ob Nachbarn vorhanden sind. Wir überprüfen den Einfluss der neuen Variable  $\log(\text{Abs. Familiarität} - \text{Abs. Frequenz})$ , indem wir für die FFDs, FPRTs, RPDs und TRTs je ein Modell mit den folgenden Faktoren berechnen: Wortlänge ( $w.len$ ), residualisierte Frequenz ( $res.freq.dlex.lm$ ),  $first.word.row$ ,  $last.word.row$ ,  $text.row$ ,<sup>108</sup> Position des Versuchsdurchgangs im Experiment ( $trial$ ) sowie das neue Maß  $fam.freq.diff$ . Die Zufallseffektstruktur besteht aus der Identität der VersuchsteilnehmerInnen ( $random\ intercept$  sowie der  $random\ slope$  in Abhängigkeit von  $trial$ ) und der Text-ID ( $random\ intercept$ ).

	<i>w.len</i>	<i>res.freq.dlex</i>	<i>fam.freq.diff</i>	<i>first.row</i>	<i>last.row</i>	<i>text.row</i>
FFD	8,7	-8,0	10,7	-15,3	-13,1	2,6
FPRT	119	-18,4	3,7	30,9	-13,4	-3,4
RPD	82,4	-15,1	6,4	10,5	24,6	7,8
TRT	150	-27,2	5,8	1,2	-31,2	-23,9

Tab. 7.6:  $t$ -Werte für die Modelle mit dem neuen Differenzmaß  $\log(\text{Abs. Familiarität} - \text{Abs. Frequenz})$ ,  $fam.freq.diff$ . Die Prädiktoren  $first.word.row$  und  $last.word.row$  sind aus Platzgründen abgekürzt.

<sup>107</sup> Diese Korrelation wurde nicht anhand des Blickbewegungsdatensatzes berechnet, sondern anhand eines Datensatzes, in den jedes Wort einmal einging. Das ist deshalb notwendig, weil in den Blickbewegungsdatensatz jedes Wort mehrfach eingeht, nämlich so oft wie es gelesen wurde. Die berichtete Korrelation macht aber nur eine Aussage über die Wörter.

<sup>108</sup> Zur Erinnerung: Das sind die Darstellungsfaktoren. Sie geben an, ob ein Wort das erste/letzte Wort in seiner Zeile ist und in welcher Textzeile auf dem Bildschirm es sich befindet.

Die Ergebnisse (siehe Tab. 7.6) sprechen dafür, dass es sich bei den zuvor berichteten Analysen nicht um bloße Artefakte der Residualisierungstechnik handelt. Der Übersichtlichkeit halber sind lediglich die *t*-Werte abgetragen. Diese geben aber einen guten Eindruck von der Effektstruktur. Alle Effekte, die in den zuvor berichteten Modellen signifikante Einflüsse zeigten, zeigen sie auch hier – und zwar in die jeweils gleiche Richtung (aufgrund der Beträge der *t*-Werte über 2 kann hier auf Signifikanztests verzichtet werden). Am interessantesten ist die Spalte für das neue Maß *fam.freq.diff*. Hier zeigen sich stabile, verlangsamende Einflüsse auf alle Maße. Das untermauert die These, dass hierdurch Wettbewerbseffekte operationalisiert werden, die von gleich langen Wörtern mit demselben Wortanfang ausgehen. Je häufiger diese in ihrer Gesamtzahl sind, desto länger wird Wort *n* gelesen.

Zurück zur Zusammenfassung der bis hier gezeigten Analysen. Es zeigen sich wiederholt deutliche Hinweise auf parafoveale Vorverarbeitung und eine Modulation der perzeptuellen Spanne durch die Eigenschaften des augenblicklich verarbeiteten Worts. Dies untermauert Effekte, die bereits in anderen Lesekorpora (vgl. Kliegl et al. 2006) und experimentellen Studien (vgl. Henderson/Ferreira 1990) gezeigt wurden. Ein Punkt bezüglich der Präsentation der Texte auf dem Bildschirm ließ sich aus der Kombination von FFDs und FPRTs ableiten: *First fixation durations* auf ersten Wörtern in einer Zeile sind geringer als für alle anderen Wörter, *first-pass reading times* hingegen länger. Dies deutet darauf hin, dass die Blickposition nach einem Zeilenwechsel (der mit weiteren Sakkadenlängen verbunden ist als das ‘normale’ Lesen innerhalb einer Zeile) oft korrigiert wird. Die erste Fixation ist dabei sehr kurz, weil die/der LeserIn direkt eine ‘Korrektursakkade’ nachschiebt, um die bevorzugte Blickposition auf dem Wort zu erreichen. Die damit assoziierte Verzögerung zeigt sich dann erst in den *first-pass reading times*.

Die Analysen auf der lexikalischen Ebene haben gezeigt, dass im FLRC Effekte zu finden sind, die auch in anderen Lesekorpora und einigen experimentellen Untersuchungen bereits gezeigt wurden. Das ist zwar beruhigend, aber nur in wenigen Punkten – dazu gehört der postulierte Wettbewerber-Effekt, der mit der Wortfamiliarität assoziiert ist – innovativ. Wir verlassen nun die Ebene rein lexikalischer Prädiktoren. Im folgenden Abschnitt wird die Verarbeitung von Mehrworteinheiten analysiert.

## 7.2 Bi- und Trigramme

Wie oft bestimmte Kombinationen von Wörtern in der natürlichen Sprache vorkommen, hat einen Einfluss auf die Verarbeitungszeit einzelner Wörter.<sup>109</sup>

<sup>109</sup> Für die vorliegende Arbeit wurden die normalisierten, logarithmierten Bi- und Trigrammfrequenzen auf Basis der Wortformen (nicht-lemmatisiert) ebenfalls aus *dlxDB* (Heister et al.

Ein wichtiger Punkt ist hier, dass nicht die Frequenzen innerhalb des FLRC selbst gemessen werden, sondern die Frequenzen in der Allgemeinsprache, wie sie über die Datenbank *dlexDB* relativ genau geschätzt werden können. Das gilt natürlich auch für die Einzelwortfrequenzen.

Naturgemäß enthalten hochfrequente Bigramme hochfrequente Wörter und hochfrequente Trigramme hochfrequente Bigramme. Diese hochfrequenten Wörter sind zum überwiegenden Teil Wörter aus der geschlossenen Klasse der Funktionswörter. Jene Wortkombination im FLRC, die am häufigsten in natürlicher Sprache vorkommt, ist 'in der'. Das in der Allgemeinsprache am seltensten vorkommende Bigramm aus dem FLRC ist 'im Beschluss' – ein interessantes Beispiel, weil gerade dieses Bigramm in juristischen Texten sicherlich überproportional vertreten ist, ist es doch, in der Terminologie von Towfigh (2009), ein linguistisches Instrument für die Herstellung externer Interdependenz. Das seltenste Trigramm aus dem FLRC ist die Wortfolge 'sich mit Hilfe'. Typisch an diesen Beispielen ist, dass die niedrigfrequenten Bi- und Trigramme Wörter aus der offenen Klasse der Inhaltswörter enthalten. Es ist leicht vorstellbar, dass Funktionswortfolgen, wie wir sie für hochfrequente Bi- und Trigramme gesehen haben, schneller verarbeitet werden – unter anderem, weil sie oft sehr gut vorhersagbar sind. Bei den oben genannten Beispielen ist zu beachten, dass es im FLRC auch Bi- und Trigramme gibt, die in *dlexDB* nicht gefunden werden konnten und somit wohl als noch seltener in der Allgemeinsprache einzuordnen sind. Dazu gehören Teile von Verweisen wie 'BVerfGE 90' oder '3 BtMG', aber auch Wortsequenzen, die der juristischen Fachsprache zuzuordnen sind, wie 'Prüfung gestellten Rechtsvorschriften' oder 'den entscheidungserheblichen Sachverhalt'. In den zu präsentierenden Analysen wird also von den textgattungsspezifischen Bi- und Trigrammen der Rechtssprache insofern abstrahiert, dass Bi- und Trigramme, die nicht in der Allgemeinsprache vorkommen, nicht in die Analyse mit eingehen.

Das Bigramm für Wort  $n$  besteht aus Wort  $n-1$  und Wort  $n$ . Entsprechend besteht das Trigramm für Wort  $n$  aus den Wörtern  $n-2$ ,  $n-1$  und  $n$ . Da das Bigramm für Wort  $n$  im Trigramm von Wort  $n$  immer komplett enthalten ist, sind diese beiden Maße untereinander korreliert. Tabelle 7.7 gibt einen Eindruck dieser Korrelationen untereinander. Grundlage ist der Datensatz, wie er in die Berechnung der Blickbewegungsdaten eingeht. Als Unigrammfrequenz wird die residuale Frequenz (residualisiert nach der Länge des Worts  $n$ ) angegeben, weil das das Maß ist, das in die Modelle eingeht.

---

2011) extrahiert. Um Missverständnisse zu vermeiden sei darauf hingewiesen, dass die Begriffe 'Bi-' und 'Trigramm' im Sinne von Wortkombinationen verwendet werden und **nicht** im Sinne von Buchstabenkombinationen (vgl. hierzu z.B. Gernsbacher 1984; Rice/Robinson 1975; White 2008).

	Res.Freq. 1-g	Freq. 2-g	Freq. 3-g	$p(n n-1)$
Freq. 2-g	0,40			
Freq. 3-g	0,22	0,52		
$p(n n-1)$	0,26	0,20	0,21	
$p(n n-2, n-1)$	0,11	0,03	0,10	0,60

Tab. 7.7: Pearson-Korrelationen ( $r$ ) zwischen Uni-, Bi- und Trigrammfrequenzen sowie bedingten Wahrscheinlichkeiten im FLRC. Alle Korrelationen sind höchst signifikant.

Wir können an dieser Stelle eine Baseline-Modellierung einsetzen, da uns die Analysen aus dem vorherigen Abschnitt bereits zur Verfügung stehen. Die Idee ist also folgende: Wir beachten alle Einflüsse auf die Lesezeit von Wort  $n$ , die wir bereits im letzten Abschnitt identifiziert haben. Die Lesezeiten werden von diesen Einflüssen bereinigt, indem wir die als ausschlaggebend identifizierten Effekte in ein Baseline-Modell aufnehmen und die Lesezeiten mit diesem Modell vorhersagen. Aus dieser Vorhersage werden die Residuen extrahiert. Wir berücksichtigen also nur noch jene Varianz, die nicht durch die bereits analysierten Effekte aufgeklärt wird. Diese residualen Lesezeiten sagen wir dann zunächst aus der Bigramm-Frequenz vorher. Zusätzlich nehmen wir die bedingte Wahrscheinlichkeit ( $p(n|n-1)$ ) des Worts  $n$ , gegeben Wort  $n-1$ , mit in das Modell auf. Tabelle 7.7 zeigt, dass auch die bedingten Wahrscheinlichkeiten erwartungsgemäß mit den jeweils zugehörigen  $n$ -Gramm-Frequenzen korreliert sind. Als Prädiktor geht also nicht die rohe bedingte Wahrscheinlichkeit ein, sondern ein nach der jeweiligen  $n$ -Gramm-Frequenz residualisiertes Maß. Nachdem wir die Effekte der Bigrammfrequenz und der zugehörigen, residualisierten bedingten Wahrscheinlichkeit analysiert haben, gehen diese Effekte mit in das Baseline-Modell ein. Für Trigramme wird das Vorgehen für Bigramme wiederholt. Das Vorgehen ist also – kurz zusammengefasst – folgendes:

- 1) Vorhersage der jeweiligen Lesezeit durch die Modelle aus Abschnitt 7.1 sowie Extraktion der Residuen.
- 2) Vorhersage dieser residualen Lesezeit durch Bigramm-Frequenz und residualisierter bedingter Wahrscheinlichkeit  $p(n|n-1)$ .
- 3) Bericht der Modelle sowie anschließende Extraktion der Residuen.
- 4) Vorhersage dieser neuen residualen Lesezeit durch Trigramm-Frequenz und nach Trigramm-Frequenz residualisierter bedingter Wahrscheinlichkeit  $p(n|n-2, n-1)$ .
- 5) Bericht der Modelle.



	Bigramm-Frequenz		residuale $p(n n-1)$	
	$\beta$ (SE)	$t$ -Wert	$\beta$ (SE)	$t$ -Wert
FFD	-0,003 (0,001)	-2,66**	-0,11 (0,02)	-6,50***
FPRT	-0,0004 (0,001)	-0,31	-0,08 (0,02)	-3,54***
RPD	-0,0004 (0,002)	-0,15	-0,05 (0,04)	-1,33
TRT	-0,005 (0,001)	-3,81***	-0,13 (0,02)	-5,52***

Tab. 7.8: Ergebnisse der Modelle auf Bigramm-Ebene. Signifikanzindikatoren beruhen auf MCMC-Simulationen und entsprechen den üblichen Schwellen.

Die Ergebnisse der Analysen für die zwei  $n$ -Gramm-Ebenen können kompakter aufbereitet werden, da jeweils nur zwei Prädiktoren in die Vorhersage eingehen. Tabelle 7.8 gibt einen Überblick über die Ergebnisse auf Bigramm-Ebene für alle Lesemaße. Grundlage sind auch hier gemischte Modelle, obwohl an dieser Stelle die Zufallseffektstruktur keine zusätzliche Varianz aufklärt. Das liegt an den abhängigen Variablen, die jeweils in die Modelle eingehen. Da diese die Residuen jener Modelle sind, die in Abschnitt 7.1 vorgestellt wurden, sind sie schon von Einflüssen der Zufallsfaktoren befreit.

Für *first fixation durations* (FFD) haben sowohl die Bigramm-Frequenz als auch die residuale bedingte Wahrscheinlichkeit einen deutlichen beschleunigenden Einfluss. Je häufiger die Wortsequenz Wort  $n-1$ , Wort  $n$  in der natürlichen Sprache ist, desto kürzer ist die erste Fixation auf dem zweiten Glied der Kette. Dasselbe gilt für den Fall, dass Wort  $n$  nach Wort  $n-1$  sehr wahrscheinlich ist. Bei *first-pass reading times* (FPRT) ist das Bild nicht ganz so klar, nur die bedingte Wahrscheinlichkeit zeigt auf diese einen signifikanten Einfluss. Für *regression path durations* (RPD) zeigen beide Faktoren keinen signifikanten Einfluss. Lediglich die bedingte Wahrscheinlichkeit geht numerisch in die vorhergesagte Richtung. Recht deutlich hingegen scheinen die Einflüsse der Bigramm-Frequenz und der bedingten Wahrscheinlichkeit auf die *total reading times* (TRT) zu sein. Die Effekte bewegen sich dort auf einem Niveau, das mit jenen der *first fixation durations* vergleichbar ist.

Der zweite Analyseschritt zeigt für die Trigramm-Frequenz vergleichbare Ergebnisse (siehe Tab. 7.9). Auf alle Lesezeitmaße hat die Trigramm-Frequenz einen beschleunigenden Einfluss. Je häufiger also die Wort-Dreiergruppe in der von *dlexDB* erfassten Allgemeinsprache ist, desto kürzer ist die erste Fixation auf dem letzten Element dieser Gruppe. Dasselbe gilt für den ersten Lesedurchgang des letzten Elements, den Regressionspfad und die Gesamtlesezeit des letzten Elements. Etwas anders verhalten sich die Einflüsse der residualen bedingten Wahrscheinlichkeiten. Je höher die residuale bedingte

Wahrscheinlichkeit für Wort  $n$ , desto länger die erste Fixation auf diesem Wort und je länger der entsprechende Regressionspfad. Beide Effekte sind mit vergleichsweise hohem Rauschen assoziiert.

	Trigramm-Frequenz		residuale $p(n n-2, n-1)$	
	$\beta$ (SE)	$t$ -Wert	$\beta$ (SE)	$t$ -Wert
FFD	-0,01 (0,002)	-4,07***	0,02 (0,01)	2,41*
FPRT	-0,01 (0,003)	-4,34***	0,008 (0,01)	0,61
RPD	-0,03 (0,006)		0,05 (0,02)	2,19*
TRT	-0,03 (0,003)	-8,16***	0,02 (0,01)	1,43

Tab. 7.9: Ergebnisse der Modelle auf Trigramm-Ebene. Signifikanzindikatoren beruhen auf MCMC-Simulationen und entsprechen den üblichen Schwellen.

Der positive Effekt der residualen bedingten Wahrscheinlichkeit (Trigramme) auf FFDs und RPDs ist so nicht vorhergesagt. Die Möglichkeit besteht, dass die multiplen vorhergehenden Residualisierungen eine Überkompensation darstellen, die zu den beobachteten Effekten führt. Zu lange Ketten von Residualisierungsvorgängen sollten also offenbar vermieden werden, da das 1) zu nur sehr schwer interpretierbaren Maßen führen kann<sup>110</sup> und 2) eventuell Überkompensationen nach sich ziehen kann, die vom „letzten Glied“ in der Kette ausgeglichen werden. Im vorliegenden Fall sollte man sich wohl für eine der beiden Variablen entscheiden:  $n$ -Gramm-Frequenz oder  $n$ -Gramm-Übergangswahrscheinlichkeit.

Außer dieses unerwarteten Effekts sind die Ergebnisse der Analysen auf Bi- und Trigrammebene hypothesenkonform: Wörter, die am Ende häufigerer Wortsequenzen stehen, werden im Allgemeinen schneller gelesen. Für Bigrammfrequenzen konnten das bereits Boston et al. (2008) in einer Untersuchung anhand des *Potsdam Sentence Corpus* zeigen. Die vorliegende Arbeit bestätigt deren Ergebnisse weitestgehend und erweitert die Aussage auf Trigramm-Frequenzen. McDonald/Shillcock (2003) finden Hinweise auf den beschleunigenden Einfluss von Übergangswahrscheinlichkeiten auf die Dauer der ersten Fixation (*single fixation durations* und *first fixation durations*) in Sätzen wie in (23). Interessanterweise können in ihrer Studie keine Effekte auf die *first pass reading times* (dort als *gaze durations* bezeichnet) gezeigt werden. Das kann einerseits daran liegen, dass sie einen stark experimentell kontrollierten Versuchsaufbau mit künstlichen Sätzen verwenden, in denen lediglich

<sup>110</sup> Vergleiche hierzu auch die Analysen zur residualen Anzahl der Levenshtein-Nachbarn in Abschnitt 7.1.1.

die Übergangswahrscheinlichkeiten zwischen einem Verb und Nomen variiert wurden. In Beispiel 23 ist das entsprechende Wortpaar kursiv hervorgehoben.

- (23) a. (hohe Übergangswahrscheinlichkeit)  
 One way to *avoid confusion* is to make the changes during vacation.
- b. (niedrige Übergangswahrscheinlichkeit)  
 One way to *avoid discovery* is to make the changes during vacation.

Die Ergebnisse der vorliegenden Arbeit (siehe Tab. 7.8) legen nahe, dass in natürlichen Texten auch ein Effekt der Bigramm-Übergangswahrscheinlichkeiten auf die Dauer des ersten Lesens eines Wortes gefunden werden kann.

Das *Potsdam Sentence Corpus* ist unter anderem aufgrund der darin enthaltenen Vorhersagbarkeitswerte attraktiv. Obwohl in der vorliegenden Arbeit keine Aussage bezüglich der empirischen Vorhersagbarkeit von Wörtern abgeleitet werden kann,<sup>111</sup> soll kurz auf eine Studie verwiesen werden, die den Zusammenhang von Bigramm-Übergangswahrscheinlichkeiten und Vorhersagbarkeit von Wörtern untersucht. Frisson/Rayner/Pickering (2005) kontrollieren in ihrem zweiten Experiment die empirische Vorhersagbarkeit des Wortes *n* im Satzkontext. Vorhersagbarkeit (*predictability*) operationalisieren sie dabei wie üblich über Lückentexte, in denen VersuchsteilnehmerInnen Satzfragmente vervollständigten. Wörter, die in bestimmten Kontexten häufiger genannt werden, gelten – gegeben den vorherigen Satzkontext – als vorhersagbarer. Frisson et al. (2005) finden keine Effekte der Übergangswahrscheinlichkeiten mehr, wenn die Vorhersagbarkeit des Wortes genau kontrolliert wird. Sie schließen daraus, dass Vorhersagbarkeit und Bigramm-Übergangswahrscheinlichkeit keine unabhängigen Effekte auf die Lesezeit haben, wie dies von McDonald/Shillcock (2003) behauptet wurde. Bigramm-Übergangswahrscheinlichkeiten und -frequenzen zeigen auch im FLRC einen soliden Einfluss auf die Lesezeiten. Trigrammfrequenzen haben unabhängig davon einen Einfluss. Zumindest teilweise werden damit – gegeben die Argumentation von Frisson et al. (2005) – Effekte der Vorhersagbarkeit abgefangen. Anders ausgedrückt: Die Hoffnung ist, dass durch die Aufnahme von Bi- und Trigrammfrequenzen in die Vorhersage von Lesezeitmaßen im FLRC das empirisch aufwändige Maß der *predictability* ersetzt wird.

<sup>111</sup> Es liegen keine Maße zur Vorhersagbarkeit der Wörter im FLRC vor. Die Erhebung dieser Variablen ist insbesondere für natürliche Texte überaus ressourcenaufwändig. Für das PSC war dies aufgrund der deutlich kleineren linguistischen Stichprobe eher möglich.

### 7.3 Analysen auf syntaktischer Ebene

Mehrere Forschungsgruppen wandten bereits Lesekorpus-Daten für die Erforschung von Phänomenen auf der Satzebene an. Die Arbeit von Boston et al. (2008) ist ein Beispiel für die Erforschung syntaktischer Effekte innerhalb von Lesekorpora, dort operationalisiert über zwei Arten von *surprisal*. Pynte et al. (2008a) und Pynte/New/Kennedy (2009) entwickeln syntaktische und semantische *constraint scores* und zeigen deren Einflüsse im *Dundee Corpus*. In der vorliegenden Arbeit wollen wir uns Effekten auf der syntaktischen Ebene zunächst dadurch nähern, dass die Einflüsse von Satzposition und syntaktischer Einbettung herausgearbeitet werden. Dies geschieht im folgenden Abschnitt. Im anschließenden Abschnitt wird dann ein kurzer Exkurs folgen. Dort wird ein Vorgehen beschrieben, das im Rahmen der Entstehung dieser Dissertation zum ersten Mal im Kontext von Lesekorpora erprobt wurde. Dabei wird von dem Wort als *region of interest*, also der grundlegenden Analyseregion, abstrahiert. Als neue *regions of interest* werden Phrasen angenommen. Nachdem dieses Vorgehen sehr kurz in Abschnitt 7.3.2 vorgestellt wird, wird in Abschnitt 7.3.3 der Einsatz für eine spezifische Fragestellung erprobt, nämlich ob sich die Relevanz von Textmaterial auf die Lesezeit von syntaktischen Phrasen niederschlägt. Eine mögliche Alternativerklärung wird im darauf folgenden Abschnitt 7.3.4 verfolgt: Diese nimmt Constraints – also die Einschränkung der Menge möglicher Satzfortsetzungen – als erklärendes Konzept an. Dort werden wir auch auf die oben angesprochenen Studien zurückkommen. Zuletzt wird dann noch auf Einflüsse ähnlichkeitsbasierter Interferenz eingegangen – ein Phänomen, das anhand der Freiburger Blickbewegungskorpora zum ersten Mal anhand natürlicher Daten untersucht wird.

#### 7.3.1 Position im Satz und syntaktische Einbettung

Das FLRC stellt eine Phrasenstrukturannotation bereit. Ein unmittelbar ableitbares Maß ist die Einbettungstiefe einzelner Wörter, operationalisiert über die Anzahl syntaktischer Knoten über dem Wort. Dieses Maß korreliert mit der Position eines Wortes im Satz ( $r = 0,43$  im Lesedatensatz des FLRC). Wir residualisieren daher die Einbettungstiefe nach der Stelle des Wortes im Satz, um die Kollinearität der beiden Faktoren aufzuheben. Für die Lesemaße werden gemischte Modelle berechnet, in die als feste Faktoren die relative Position im Satz,<sup>112</sup> die residuale Einbettungstiefe sowie die Interaktion der beiden Faktoren eingehen. Als Kontrollvariable geht außerdem der Korpusteil (Reformulierungen vs. Langtexte) mit in die Modelle ein. Als *random intercept*

<sup>112</sup> Die relative Position im Satz ist der Quotient Wortindex/Satzlänge.

wird die Textidentität aufgenommen. Die VersuchsteilnehmerInnen-ID als *random intercept* klärt keine zusätzliche Varianz auf, da sie bereits in die Modelle zur Residualisierung der Lesedaten einging. Als abhängige Variable dient das jeweilige residuale Lesezeitmaß. Herausgerechnet sind dabei bereits 1) Einflüsse der lexikalischen Ebene, die im entsprechenden Modell aus Abschnitt 7.1 beachtet sind, 2) die Einflüsse der Bigramm-Ebene sowie 3) Einflüsse der Trigramm-Ebene (zu den Punkten 2 und 3 siehe 7.2).

	Trigramm-Frequenz		residuale $p(n   n-2, n-1)$	
	$\beta$ (SE)	$t$ -Wert	$\beta$ (SE)	$t$ -Wert
FFD	-0,01 (0,002)	-4,07***	0,02 (0,01)	2,41*
FPRT	-0,01 (0,003)	-4,34***	0,008 (0,01)	0,61
RPD	-0,03 (0,006)		0,05 (0,02)	2,19*
TRT	-0,03 (0,003)	-8,16***	0,02 (0,01)	1,43

Tab. 7.10: Ergebnisse der Modelle für die relative Position des Wortes im Satz und die residuale Einbettungstiefe. Signifikanzindikatoren beruhen auf MCMC-Simulationen und entsprechen den üblichen Schwellen.

Der Teil des Korpus (Reformulierungen vs. komplette Texte) wurde zur Kontrolle in die Modelle mit aufgenommen, zeigt aber keinerlei signifikante Einflüsse. Dies ist insofern beruhigend, als dass die Phrasenstrukturannotation der beiden Korpusteile unterschiedlich entstanden ist. Unterschiede zwischen den beiden Korpusteilen hätten daher rühren können. In Tabelle 7.10 sind die restlichen Ergebnisse der gemischten Modelle kompakt zusammengefasst.<sup>113</sup> Am solidesten scheint der Effekt der Interaktion zwischen beiden Faktoren zu sein – was die Interpretation der Einzeleffekte schwierig macht. Abbildung 7.6<sup>114</sup> gibt einen Eindruck dieses Interaktionseffekts und zeigt für alle Lesezeitmaße ein vergleichbares Muster: Während die residuale Einbettungstiefe am Anfang eines Satzes einen verlangsamenden Einfluss auf die Lesezeiten hat, dreht sich dieser Einfluss gegen Ende des Satzes um. Am Satzende gilt: Je tiefer ein Wort eingebettet ist, desto schneller wird es gelesen.

Als häufig berichtetes Lesezeitmaß konzentrieren wir uns nun auf die *first-pass reading times* (FPRT). Es zeigt sich, dass der berichtete Interaktionseffekt hauptsächlich von Funktionswörtern gespeist wird – so zeigt es eine Einteilung des Datensets, wie sie beispielsweise Pynte et al. (2008a) vornehmen. Schränkt

<sup>113</sup> Aus Gründen der Übersichtlichkeit sind die Standardfehler der Schätzer nicht eingetragen. Da sich der  $t$ -Wert jedoch aus dem Verhältnis zwischen Effektschätzer ( $\beta$ ) und Standardfehler ergibt, kann die Größenordnung der Standardfehler abgeschätzt werden.

<sup>114</sup> In diesem Fall sind die Maße **nicht** aus log-transformierten Maßen zurücktransformiert, da es sich um Residuen handelt.

man die Analyse auf Inhaltswörter ein, was eine Einschränkung auf 54,4% des Datensets bedeutet, zeigen sich unabhängige beschleunigende Effekte für die relative Satzposition ( $t = -2,35$ ;  $p_{\text{MCMC}} = 0,019$ ) und die residuale Einbettungstiefe ( $t = -4,75$ ;  $p_{\text{MCMC}} < 0,001$ ). Die Interaktion ist nicht signifikant. Für Inhaltswörter scheint also zu gelten: Je weiter hinten ein Wort im Satz steht und je tiefer es eingebettet ist, desto schneller wird es gelesen. Beide Einflüsse haben einen unabhängigen Einfluss auf die FPRT, interagieren aber nicht.

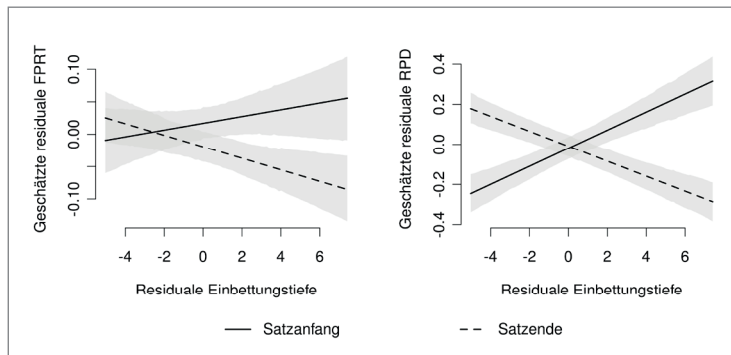


Abb. 7.6: Interaktionseffekt residuale Einbettungstiefe  $\times$  relative Satzposition für *first-pass reading times* (links) und *regression path durations* (rechts) inkl. 95%-Konfidenzintervallen.

Dieser unabhängige Effekt der relativen Satzposition widerspricht den Ergebnissen von Mitchell/Green (1978), die in ihren Experimenten 1 und 4 keinen Effekt der Position im Satz feststellen konnten. Sie argumentieren damit gegen einen Ansatz, wie ihn beispielsweise bereits Hochberg (1970) formuliert. Dieser besagt, dass LeserInnen in besser vorhersagbaren Textabschnitten schneller lesen sollten, weil sie genauere Hypothesen über das zukünftige Material aufstellen können, „as the materials become more constrained grammatically and semantically“ (Mitchell/Green 1978, S. 611). In jüngerer Zeit hat diese Idee wieder Auftrieb erhalten und Vorhersagbarkeits-effekte konnten mehrfach nachgewiesen werden (Ehrlich/Rayner 1981; Frisson et al. 2005; Kliegl et al. 2004, 2006; Konieczny 2000; Pynte et al. 2008a; Pynte/New/Kennedy 2008b). Vorhersagbarkeit wurde dabei auf unterschiedliche Arten operationalisiert – einige dieser Operationalisierungen wurden oder werden in der vorliegenden Arbeit noch detaillierter erläutert. In der Gesamtschau lassen sich die Vorhersagbarkeitskonzepte einteilen in syntaktisch motivierte Konzepte wie *surprisal* (Hale 2001) oder *syntactic constraint score* (Pynte et al. 2008b), semantisch orientierte Entwürfe wie dem LSA-basierten *semantic constraint score* von Pynte et al. (2008b) und empirisch erhobene Vorhersagbar-

keitsmaße (Binder/Pollatsek/Rayner 1999; Kliegl et al. 2004). In der vorliegenden Arbeit kann Vorhersagbarkeit zunächst nur über die relative Position des Wortes im Satz (meist kürzer als relative Satzposition bezeichnet) operationalisiert werden. Für diese zeigt sich ein unabhängiger beschleunigender Effekt auf die Lesezeit.

Betrachten wir nun den Effekt der Einbettungstiefe auf die *first-pass reading time* von Inhaltswörtern. Pynte et al. (2008a) schlagen vor, dass ein beschleunigender Effekt mit steigender Einbettungstiefe Ausdruck einer Lesestrategie der LeserInnen ist:

They [= deeply embedded words, S.W.] are more likely to function as members of a prepositional phrase, an adjectival phrase, a relative clause, etc., and will, by definition, be less central to the main topic of the sentence than less deeply embedded words. (Pynte et al. 2008a, S. 8)

Dieser Einschätzung inhärent (aber nicht direkt von Pynte et al. (2008a) geprüft) ist die Hypothese, dass der Einbettungseffekt nicht auf alle Inhaltswörter verallgemeinerbar ist. Am ehesten sollte sich ein solcher Relevanzeffekt auf Nomen und Adjektive niederschlagen. Auf die Verarbeitung von im Satz-zusammenhang relevanteren Elementen wie beispielsweise das finite Hauptverb eines Satzes sollte die Einbettungstiefe keinen beschleunigenden Einfluss haben. Um diese Hypothese zu überprüfen, können wir die Inhaltswörter in weitere Teilmengen unterteilen.

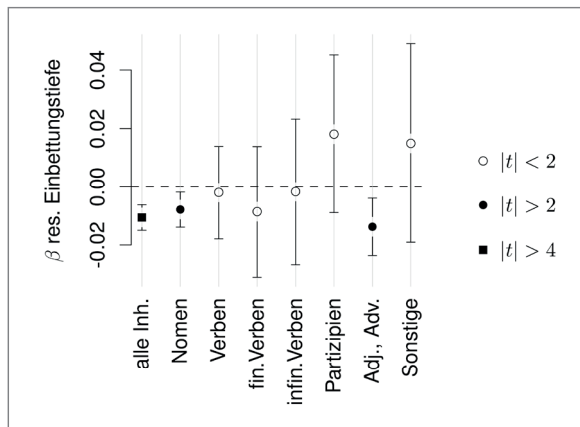


Abb. 7.7: Effektschätzer und deren Konfidenzintervalle für residuale Einbettungstiefe und FPRTs für verschiedene Teilmengen der Inhaltswörter. Assoziierte  $t$ -Werte sind über Symbole abgetragen. Konfidenzintervalle ergeben sich aus 1,96 Standardfehlern. Adj., Adv. = Adjektive und Adverbien.



Abbildung 7.7 zeigt eine Illustration der verschiedenen Effektschätzer für diese getrennten Analysen. Diese Darstellung ist angelehnt an jene von Boston et al. (2008, S. 10). Dort werden auf der  $x$ -Achse verschiedene abhängige Variablen abgetragen und pro Prädiktorvariable werden verschiedene Plots präsentiert. Die Grundidee der Effektschätzer-Visualisierung ist aber die gleiche. Sobald die Null-Linie im Konfidenzintervall eines Effektschätzers enthalten ist, ist der Effekt nicht signifikant. Das wird von den ebenfalls in Abbildung 7.7 über die Symbole abgetragenen  $t$ -Werte bestätigt. Ganz links ist der Effektschätzer der residualen Einbettungstiefe für alle Inhaltswörter abgetragen. Alle Effektschätzer rechts davon entstammen Modellen, die jeweils mit einer Teilmenge des Datensatzes operieren. Es zeigt sich, dass lediglich die Effektschätzer für alle Inhaltswörter sowie jene für Nomen und Adjektive/Adverbien einen signifikanten beschleunigenden Einfluss auf die FPRTs haben. Insbesondere bei Adjektiven und Adverbien kann diese Beschleunigung als gutes Indiz dafür gelten, dass LeserInnen in der Tat tiefer eingebetteten modifizierenden Strukturen weniger Beachtung schenken.

Es bietet sich eine weitere Untersuchung an, die die Ergebnisse der eben präsentierten Analyse untermauern könnte. Für einen Teil des FLRC haben wir die Möglichkeit, die Auswertung auf Phrasenebene vorzunehmen. Dieses Vorgehen soll im Folgenden zunächst kurz erläutert werden.

### 7.3.2 Exkurs: Syntaktische Knoten als Analyseregionen

In den bisherigen Darstellungen war die Analyseeinheit immer das Wort. Das heißt, der Text wird in Analyse-Regionen (*regions of interest*) zerlegt, in denen eine Region immer aus einem Wort besteht. Alle Lesezeitmaße werden dann anhand dieser Region berechnet. Dieses Vorgehen ist im Kontext der Lesekorpusforschung üblich. Interessanterweise steht es aber im Kontrast zu einem Standardvorgehen in sonstiger psycholinguistischer Forschung. Dort werden gezielt Stimulussätze entworfen, die minimal verändert werden, um Kontraste im experimentellen Design herzustellen. Die Sätze oder Texte bleiben also weitestgehend über die experimentellen Bedingungen hinweg identisch – meist ändert sich nur ein kleiner Teil des Satzes. Genau diese Region ist dann die *region of interest* (ROI) bzw. – je nach Sprachregelung – die *interest area* (IA), also die Region, wo gezielt nach Effekten der experimentellen Manipulation gesucht wird. Außerdem beruht die Berechnung der Lesevariablen auf der Einteilung der Regionen. Während die *total reading times* für eine Region, die aus beispielsweise drei Wörtern besteht, noch additiv zerlegbar sind (d.h. einfach auf die drei Wörter aufteilbar), gilt diese additive Zerlegbarkeit bereits für die *first-pass reading times*, die *regression path durations* oder auch die Wahr-

scheinlichkeit, dass eine regressive Sakkade während des Lesens dieser Region initiiert wird, schon nicht mehr. Die einer Analyse zugrunde gelegte Regioneneinteilung kann also weiter reichende Konsequenzen haben, als man zunächst meinen könnte.

Die Zuweisung einer ROI gestaltet sich bei natürlichen Texten, die keiner gezielten experimentellen Manipulation unterworfen sind, nicht so eindeutig wie bei der klassisch-experimentellen psycholinguistischen Forschung. Daher wählen wir einen Ansatz, der ausgehend von den Phrasenstruktur-Annotationen das sprachliche Stimulusmaterial in Regionen einteilt. Im Vergleich zur wortweisen Einteilung gilt es hier etwas wichtiges zu beachten: Ist das Wort die ausschlaggebende Einheit der Regioneneinteilung, ist der komplette Text mit Regionen abgedeckt und keine Region überschneidet sich mit einer anderen. Zieht man aber Knoten der Phrasenstrukturannotation als regions-einteilende Elemente heran, ist das nicht mehr der Fall, da Knoten unter anderen Knoten eingebettet sein können. Nehmen wir an, eine Leserin/ein Leser fixiert die Stelle, die in Beispiel 24 fett gedruckt ist (das 'öff' in 'veröffentlichen'). Die Dauer dieser Fixation trägt zur Lesezeit von mehreren Knoten bei. Dazu gehört der globale Satzknoten S, die erste Nominalphrase (NP) 'Auslöser ... wurde', die erste Präpositionalphrase (PP) 'für den ... Beschluss' sowie die erste Adjektivphrase (AdjP) 'am Freitag veröffentlichten'. Nutzt man die Knoten der Phrasenstrukturannotation zur Regioneneinteilung eines Textes, kann also eine Fixation zur Lesezeit mehrerer Knoten beitragen. Die zweite PP 'am Freitag' ist vor 'veröffentlichen' bereits wieder geschlossen – die Fixation würde nicht zur Lesezeit dieser PP zählen.

- (24) [S [NP Auslöser [PP für den [AdjP [PP am Freitag] veröffentlichten] Beschluss], [S [PP mit dem] [NP die „Haschisch-Entscheidung“ [PP von 1994]] [VP bestätigt] wurde]], war [NP eine Vorlage [NP des Amtsgerichts Bernau]]].

### 7.3.3 Relevanz auf Phrasenebene

Wir werden nun diese Phrasenstrukturannotation als ROI-Einteilung benutzen, um so weiter der oben aufgeworfenen Fragestellung nachzugehen, ob tiefer eingebetteten Strukturen tatsächlich weniger Relevanz beigemessen wird und sie damit schneller gelesen werden. Genauer formuliert besagt diese Hypothese, dass modifizierenden Elementen wie einigen Nomen, aber vor allem Adjektiven und Adverbien, weniger Relevanz beigemessen wird, wenn sie tiefer im Satz eingebettet sind. Was zuvor noch über Part-of-Speech-Tags operationalisiert wurde, kann nun auf Phrasentypen übertragen werden. Jene Phrasentypen, die am eindeutigsten mit modifizierenden Strukturen assozi-

iert sind, sind Adjektivphrasen (AdjP), Adverbialphrasen (AdvP) und Präpositionalphrasen (PP). So ist in Beispiel 24 für die Erschließung des Gesamtsinns des Satzes relativ unerheblich, dass der Beschluss am Freitag veröffentlicht wurde. Im Beispielsatz ist diese Information über eine AdjP auf Einbettungstiefe 4 realisiert, die 'Beschluss' pränominal modifiziert. Da die LeserInnen ohnehin bereits aufgrund mehrerer offener syntaktischer Klammern einer nicht zu unterschätzenden Beanspruchung von Ressourcen ausgesetzt sind, könnte es für die Verarbeitung des Satzes von Vorteil sein, der PP 'am Freitag' weniger Aufmerksamkeit zukommen zu lassen. Vermutlich ist die Information ohnehin irrelevant, sonst hätte die/der VerfasserIn des Textes die PP nicht so tief eingebettet – so könnte die (natürlich unbewusste) Raison der LeserInnen lauten. Verbalphrasen (VPs) hingegen kodieren meist Informationen, die für den Gesamtzusammenhang des Satzes relevanter sind. Es ist also nicht unbedingt davon auszugehen, dass diesen mit steigender Einbettungstiefe weniger Relevanz zukommt. Für die LeserInnen sollte es darum keinen Grund geben, der Verarbeitung von Material in VPs generell weniger Aufmerksamkeit zu widmen.

Zur Überprüfung dieser Hypothese werden die Lesemaße für die Regioneneinteilung auf Phrasenebene berechnet. Die rohen, logarithmierten FPRTs werden dann anhand eines einfachen linearen Modells residualisiert, in das die Länge der Region in Buchstaben, die Länge der Region in Wörtern sowie die Position der Region im Satz eingeht. Alle diese Faktoren haben höchstsignifikante Einflüsse auf die Lesezeiten der Regionen. Die auf diese Weise entstandenen residualen FPRTs werden nun der eigentlichen Analyse über ein lineares gemischtes Modell zugeführt, in das als fester Effekt lediglich die Einbettungstiefe des jeweiligen Knotens eingeht. Zusätzlich gehen als *random intercepts* die Text- und VersuchsteilnehmerInnen-Identität ein. Letztere wird über die *random slope* in Abhängigkeit der Position im Experimentalablauf modifiziert, um Ermüdungs- und Übungseffekte zu erlauben. Es werden gemischte Modelle mit der jeweils gleichen Prädiktorstruktur für verschiedene Phrasentypen berechnet.

In Abbildung 7.8 sind die Effektschätzer für die Einbettungstiefe aus diesen Modellen abgetragen. Für den Datenpunkt ganz links wurde der Effektschätzer aus dem Modell extrahiert, in das alle Knotentypen eingingen. Alle Knotentypen rechts davon sind Teilmengen dieser Datenreihe. Im verwendeten Annotationsschema gibt es für jeden der hier dargestellten Phrasentypen auch einen koordinierten Phrasentyp.<sup>115</sup> Diese koordinierten Phrasentypen werden

<sup>115</sup> Koordinierte Adjektivphrasen (CAAdjP), die im FLRC enthalten sind, sind beispielsweise 'geeignetes und erforderliches' oder 'weder unzulässig noch offensichtlich begründet'. Eine koor-

dem jeweils zugehörigen Phrasentyp subsumiert. Diese Maßnahme hat nur an einer Stelle eine Auswirkung, in diesem Fall ist diese für die Ergebnisse der Analyse sogar eher nachteilig.

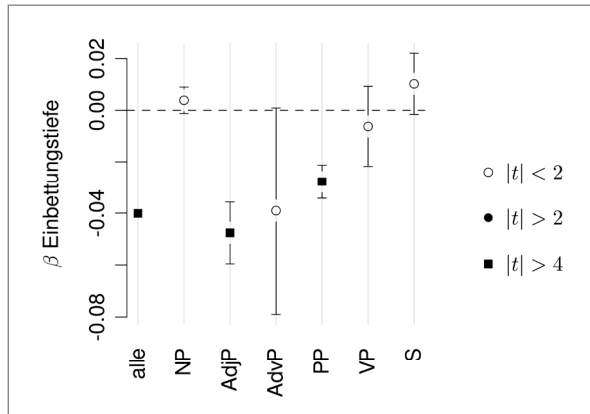


Abb. 7.8: Effektschätzer und deren Konfidenzintervalle für Einbettungstiefe und residuale FPRs für verschiedene Phrasentypen. Assoziierte t-Werte sind über Symbole abgetragen. Konfidenzintervalle ergeben sich aus 1,96 Standardfehlern.

In der Tat werden also aufgrund von Knoten definierte Regionen mit steigender Einbettungstiefe schneller gelesen. Wie sich in den Teilmengen-Analysen zeigt, speist sich dieser Effekt hauptsächlich aus Adjektiv- und Präpositionalphrasen. Der Effektschätzer für Adverbialphrasen liegt zwar in vergleichbaren Größenordnungen wie jener für alle Knoten, durch das große Konfidenzintervall wird dieser Effekt jedoch nicht signifikant. Das Konfidenzintervall ist für dieses Modell höchstwahrscheinlich deshalb so groß, weil das  $n$  für Adverbialphrasen lediglich bei 1.684 Fällen liegt. Im Vergleich zu Adjektivphrasen ( $n = 11.054$ ) und Präpositionalphrasen ( $n = 37.692$ ) ist dies sehr wenig. Trotzdem würde dieser Effekt signifikant, wenn man koordinierte Adverbialphrasen aus der Analyse ausschließen würde.

Um auszuschließen, dass das gefundene Effektmuster von einer systematischen Ungleichverteilung der Phrasentypen über verschiedene Einbettungstiefen hinweg herrührt, ist in Abbildung 7.9 eine normalisierte Heatmap der untersuchten Phrasentypen in Abhängigkeit der Einbettungstiefe abgedruckt. Je heller das entsprechende Feld, desto mehr Fälle umfasst diese Zelle. Die

dinierte Verbalphrase (CVP) ist 'die Verbreitung der Droge in der Gesellschaft einzuschränken und die damit verbundenen Gefahren zu verringern'. Ein Beispiel für eine koordinierte Präpositionalphrase (CPP) ist: 'aus sich heraus und ohne Beziehung der Akten'.

Normalisierung geschieht zeilenweise, das heißt, die Ungleichverteilung zwischen den verschiedenen Phrasentypen – die Nominalphrase ist beispielweise der mit Abstand häufigste Phrasentyp – fällt bei der Darstellung nicht mehr ins Gewicht. Eventuell ist für Adjektiv- und Präpositionalphrasen tatsächlich eine Verschiebung nach rechts (also in Richtung tiefere Einbettungen) erkennbar. Es sollten aber noch immer genug Nominal- und Verbalphrasen in tieferen Einbettungsebenen vorliegen, dass beschleunigende Effekte aufgefunden werden können.

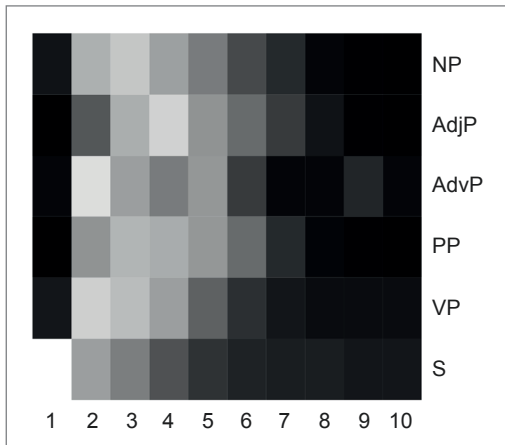


Abb. 7.9: Heatmap für die Verteilung der Phrasentypen (inkl. entsprechender Koordinatio-  
nen) über die Einbettungstiefen hinweg. Die abgetragenen Werte sind zeilenweise  
normalisiert. Einbettungstiefe 0 ist nicht abgetragen, weil dort nur Satzknotten  
möglich sind. Die Einbettungstiefen 11 und 12 sind nicht abgetragen, weil dort  
extrem wenig Fälle auftreten.

Insgesamt werden also tiefer eingebettete Knoten schneller gelesen. Betrachtet man nun Phrasentypen, die zu modifizierenden Zwecken im Sinne von Pynte et al. (2008a) eingesetzt werden können (AdjPs, AdvPs und PPs), zeigt sich zumindest für Adjektivphrasen und Präpositionalphrasen dieser beschleunigende Effekt der Einbettungstiefe. Der Effektschätzer für Adverbialphrasen ist zwar ähnlich groß, aufgrund der kleinen Datenbasis für diesen Phrasentyp kann aber keine statistisch abgesicherte Aussage gemacht werden. Wir erinnern uns, dass bei der Analyse anhand Wortregionen (siehe Abb. 7.7) auch für Nomen ein beschleunigender Effekt der Einbettungstiefe zu finden war. Im Lichte der eben vorgestellten Analyse auf Phrasenebene könnte dieser Effekt erklärbar sein: In PPs stehen oft Nomen (beispielsweise wie in Beispiel 24 die PP 'am Freitag' oder 'zur Erteilung der Duldung'). Sind es viel-

leicht eben diese Nomen, die den Einbettungseffekt zeigen? Eine Nachanalyse legt genau dies nahe: Schränken wir die Analyse auf jene Nomen ein, die in einer Präpositionalphrase stehen (26.091 Datenpunkte), bleibt der Effekt der Einbettungstiefe bestehen ( $t = -2,53$ ;  $p = 0,012$ ). Wählen wir jedoch jene Nomen, die **nicht** in einer Präpositionalphrase eingebettet sind (26.952 Datenpunkte), verschwindet der Effekt ( $t = -1,12$ ;  $p = 0,264$ ). Das bedeutet: Nomen, die in einer PP stehen, werden mit steigender Einbettung schneller gelesen. Für Nomen, die nicht in einer PP stehen, besteht dieser Effekt nicht.

Die Einschätzung, dass der beschleunigende Effekt von tieferen Einbettungen auf Lesezeiten von einer Lesestrategie stammt, wird durch die vorliegenden Ergebnisse untermauert. Diese Lesestrategie besagt, dass auf tiefer eingebettete Strukturen weniger Aufmerksamkeit verwendet werden muss, wenn diese Strukturen modifizierenden Charakter haben. Einer solchen Strategie ist die implizite Annahme inhärent, dass tiefer eingebettete Modifikationen unwichtiger für die zentrale Aussage des Satzes sind. Diese Strategie könnte als Kompensationsstrategie beschrieben werden. Demnach würden LeserInnen bei tieferen Einbettungen so viel Ressourcen auf die Aufrechterhaltung offener syntaktischer Klammern verwenden müssen, dass sie an anderen Stellen Ressourcen „einsparen“ müssen – und da bieten sich die ohnehin weniger relevanten Modifikationen an. Es wurden mehrere Indizien präsentiert, dass die von Pynte et al. (2008a) aufgeworfene These zur Lesestrategie tief eingebetteter Modifikationen tatsächlich ihre Berechtigung hat. Offenbar scheint es Korrelate bezüglich der Lesezeit ganz bestimmter Elemente zu geben. Bei anderen Elementen, wo vergleichbare Effekte laut der Hypothese nicht zu finden sein sollten, sind diese tatsächlich nicht zu finden. Das ist beispielsweise für Nomen der Fall, die nicht in PPs stehen. Aber auch Phrasentypen, auf die die Hypothese der Irrelevanz bei tiefer Einbettung unpassend erscheint, wie beispielsweise Verbal- und Nominalphrasen oder Satzknotten, zeigen den Effekt nicht – letztere beide Phrasentypen zeigen ihn sogar numerisch eher in die entgegengesetzte Richtung.

#### 7.3.4 Constraints als erklärendes Konzept: Relative Einbettungstiefe

Eine Alternativerklärung für den generell beschleunigenden Effekt der Einbettungstiefe würde – allgemein gesprochen – Constraints als erklärendes Konzept mit einbeziehen. Die Grundidee ist, dass durch syntaktische Einbettung die Menge der möglichen Fortsetzungen immer mehr eingegrenzt wird. Wenn die potenzielle Menge der möglichen Satzfortsetzungen immer kleiner wird, werden einzelne Elemente immer wahrscheinlicher und damit für die/ den LeserIn vorhersagbarer bzw. erwartbarer. Somit sinkt die Verarbeitungs-

zeit bei tiefen Einbettungen im Vergleich zu hohen Einbettungen. Doch dieser Effekt stammt nicht einfach daher, dass die Einbettung tiefer wird. Eine prominente Operationalisierung der Verarbeitungsschwierigkeit an einer ganz bestimmten Stelle im Satz legte Hale (2001) mit dem Konzept des *surprisal* vor (siehe auch 5.3.2), das – ganz knapp zusammengefasst – die Summe der Wahrscheinlichkeit aller zu einem Zeitpunkt verworfenen Fortsetzungen eines Satzes erfasst. Formale Beschreibungen von *surprisal* sind an mehreren Stellen verfügbar, u.a. bei Hale (2001) und Boston et al. (2008). Ausführliche Erklärungen sowie die Integration in ein Verarbeitungsmodell, das auf einer *Tree Adjoining Grammar* (TAG, auch Baumadjunktionsgrammatik, Joshi/Levy/Takahashi 1975) basiert, geben Demberg/Keller (2009) und Demberg/Keller/Koller (2013). Boston et al. (2008) finden für das PSC höhere Lesezeiten für Wörter mit höherem *surprisal* (also für Wörter, in deren Übergang viele bzw. sehr wahrscheinliche Satzinterpretationen oder *parses* verworfen werden müssen). Analysen von Demberg/Keller (2008) bestätigen diese Ergebnisse für den englischen Teil des *Dundee Corpus*. Fassen wir nun syntaktische Einbettung als eine bestimmte Art von Constraintsetzung auf, also als die Einführung von Einschränkungen, folgt daraus, dass das *surprisal* in tiefer eingebetteten Strukturen niedriger ausfallen müsste: „The more constraining the prior sentence fragment, the less will be the uncertainty regarding the way an incoming word can be integrated, and the lower the surprisal value associated with it.“ (Pynte et al. 2008a, S. 2). Nun ist die absolute Einbettungstiefe ein sehr grobes Maß für das Wirken syntaktischer Constraints. Sie ist aber das einzige Maß, das im FLRC (bisher) vorliegt. Pynte et al. (2008a, 2009) benutzen für ihre Untersuchungen am französischen Teil des *Dundee Corpus* ein Maß, das sie als *syntactic-constraint score* bezeichnen. Ausschlaggebend für den Wert dieses Maßes ist, wie oft eine bestimmte Kombination von offenen Phrasen vorkommt, gegeben die Kombination offener Phrasen am vorherigen Wort. Die Kombinationen werden dabei über numerische Vektoren gefasst, wobei jede Stelle im Vektor für einen Phrasentyp steht (für die folgende Darstellung vgl. auch Pynte et al. 2008a, S. 3f.). Die Übergangswahrscheinlichkeiten zwischen Vektoren müssen anhand eines großen Korpus geschätzt werden. Daran scheitert (bisher) die Errechnung eines vergleichbaren Maßes für das FLRC. Allerdings heißt das nicht, dass die Erweiterung des FLRC mit solch einem Maß völlig undenkbar wäre – die annotationstechnischen Voraussetzungen sollten soweit gegeben sein. Das gilt auch für den ebenfalls von Pynte et al. (2008a) eingesetzten *semantic-constraint score*. Dieser errechnet sich aus der LSA-basierten Ähnlichkeit (vgl. Foltz et al. 1998; Landauer/Dumais 1994; Landauer et al. 1998) zwischen dem Satzfragment bis Wort  $n-1$  und Wort  $n$ . Auch hier ist ein großes Vergleichskorpus notwendig, anhand dessen diese Ähnlichkeiten berechnet werden können. Pynte et al. (2008a) verwendeten hier ein Korpus mit



über 45 Millionen Wörtern, das sich aus Romanen und Filmdialogen zusammensetzte. Der Erweiterung des FLRC in diese Richtung steht grundsätzlich nichts entgegen, ist jedoch nicht Bestandteil dieser Dissertation.

Jedoch ist eine andere, ähnlich gelagerte Analyse mit der vorliegenden syntaktischen Annotation direkt möglich. Es wurde bereits in diesem Abschnitt gezeigt, dass die Einbettungstiefe einen deutlichen Effekt auf die Lesezeit hat. Jedoch wurde immer von der **absoluten** Einbettungstiefe ausgegangen, also der Anzahl der Knoten in der Phrasenstrukturannotation über dem aktuellen Wort bzw. der aktuellen Phrase. Doch wie steht es mit der **relativen** Einbettungstiefe? Diese kann auf mehrere Arten operationalisiert werden. Einerseits könnte lediglich der Unterschied zwischen der Einbettungstiefe von Wort  $n-1$  und Wort  $n$  berechnet werden (nennen wir dieses Maß  $\Delta_{emb}$ ). Im Sinne der syntaktischen Vorgänge ist dies allerdings ein sehr grobes Maß. Wenn  $\Delta_{emb}$  beispielsweise 1 beträgt (je nach Polarisierung des Maßes würde das bedeuten, dass Wort  $n$  eine Ebene tiefer oder höher eingebettet ist), ist unklar, wie viele syntaktische Klammern im Übergang von Wort  $n-1$  zu Wort  $n$  geschlossen oder geöffnet wurden. Es scheint unter dieser Perspektive sinnvoller zu sein, tatsächlich die Anzahl sich schließender und öffnender Klammern aus der syntaktischen Annotation abzuleiten und diese direkt als Prädiktoren in die Modelle aufzunehmen.

Welche Effekte sind unter Verarbeitungsperspektive zu erwarten? Kommt man als LeserIn an einen Punkt, an dem mindestens eine syntaktische Klammer geöffnet wird, sollte – zumindest ganz allgemein gesprochen – Verarbeitungsaufwand entstehen, weil die Repräsentation einer neuen Phrase initiiert werden muss. Um diese Hypothese zu überprüfen, werden wir zunächst eine Variable als Prädiktor aufnehmen, die nur überprüft, ob überhaupt eine Phrase geöffnet wird. Die Variable *phrase.geöffnet* erhält also den Wert *ja*, wenn die Anzahl sich öffnender Phrasen im Übergang zu Wort  $n$  größer 0 ist, ansonsten ist *phrase.geöffnet* gleich *nein*. Der Verarbeitungsaufwand für Wörter, die sich in ein und derselben Phrase befinden wie das vorherige Wort sollte geringer sein, da sich der Satz weiterhin auf derselben Einbettungsebene bewegt und keine neue Phrase repräsentiert werden muss.

Tabelle 7.11 zeigt, dass Wörter,<sup>116</sup> mit deren Auftreten eine neue Phrase geöffnet wird, unabhängig von relativer Position (*rel.pos*) und residualer Einbettungstiefe (*res.einb.tiefe*) eindeutig länger gelesen werden. Eine Angabe der Effektstärke ist dabei nicht sinnvoll, da es sich bei der abhängigen Varia-

<sup>116</sup> Es sei explizit darauf hingewiesen, dass hier wieder eine wortweise Regioneneinteilung benutzt wird.

blen um residualisierte FPRTs handelt. Der Effektschätzer bewegt sich aber über jenen für die relative Satzposition und der residualen Einbettungstiefe. Der Effekt ist also in seinen Ausmaßen nicht zu unterschätzen.

Groups	Name	Variance	Std.Dev.			
text.id	(Intercept)	0.00076528	0.027664			
Residual		0.17746548	0.421267			
		Estimate	Std. Error	t value	$P_{\text{MCMC}}$	
(Intercept)		0.012896	0.008753	1.473	0.1612	
rel.pos		-0.030624	0.010028	-3.054	0.0022	**
res.einb.tiefe		-0.004997	0.001667	-2.998	0.0036	**
phrase.geöffnet ja		0.058451	0.006129	9.537	0.0001	***

Tab. 7.11: Lineares gemischtes Modell für den Einfluss der Phrasenöffnung (ja/nein) auf die *first-pass reading times*, residualisiert nach Effekten der lexikalischen Ebene sowie Bi- und Trigrammfrequenzen und -übergangswahrscheinlichkeiten.

Hat es – wenn mindestens eine Phrase geöffnet wird – auch einen Einfluss, wie viele Phrasen geöffnet werden? Ein solcher Effekt würde auch eine Alternativklärung für den eben berichteten Effekt ausschließen. Diese Alternativklärung führt den Fakt, dass Wörter am Anfang einer Phrase schneller gelesen werden, darauf zurück, dass es innerhalb von Phrasen zu einer Beschleunigung der Lesezeiten kommt. Dieser Effekt wäre also das Spiegelbild des Effekts auf Satzebene. Die Existenz eines solchen Effekts würde automatisch nach sich ziehen, dass Wörter, die am Anfang einer Phrase stehen, langsamer gelesen werden als jene, die nicht am Anfang einer Phrase stehen. Schränkt man die Analyse jedoch nur auf phraseninitiale Wörter ein, kann ein Beschleunigungseffekt innerhalb von Phrasen keinen Einfluss mehr haben. Eine solche Analyse ist im Gegensatz zum Modell, das in Tabelle 7.11 einzusehen ist, mit einem erheblichen Verlust an Datenpunkten verbunden, da alle Wörter, die nicht das erste Wort einer Phrase sind ( $n = 121.781$ ), wegfallen und nurmehr 70.452 Datenpunkte übrig bleiben. Trotzdem zeigt sich, dass auch die Anzahl sich öffnender Phrasen einen signifikanten verlangsamenden Einfluss auf die *first-pass reading times* hat. Das entsprechende Modell ist in Tabelle 7.12 einzusehen. Es ist außerdem gut zu erkennen, dass der Einfluss der relativen Position des Wortes im Satz *rel.pos* fast verschwindet, und der Effekt der residualen Einbettungstiefe (*res.einb.tiefe*) nun mit einer Irrtumswahrscheinlichkeit von knapp 80% weit davon entfernt ist, signifikant zu sein.

Die gezeigten Effekte sind unter mehreren Gesichtspunkten in Perspektive zu setzen. Zunächst ist wichtig zu bemerken, dass keine direkte Verbindung gezogen werden sollte zwischen den Prozessen inkrementeller Textverarbeitung (hier gemessen durch *first-pass reading times*) und der Phrasenstruktur

eines Satzes, denn die Phrasenstruktur wird immer unter Bezugnahme des globalen Satzkontexts annotiert. Hier sind Konstrukte wie der *syntactic constraint score* bzw. *surprisal* wohl etwas näher am Gedanken der inkrementellen Satzverarbeitung. Wenn *surprisal* die Größe der Veränderung in der Wahrscheinlichkeitsverteilung bezüglich der globalen Satzinterpretation im Übergang zwischen zwei Wörtern misst, kann man sich zwar ausführlich darüber streiten, ob eben diese Wahrscheinlichkeitsverteilung ein „mentales Pendant“ hat. Trotzdem muss man aber anerkennen, dass dieses Maß eventuell näher am Gedanken einer inkrementellen Wort-für-Wort-Herleitung der globalen Satzinterpretation ist als die Anzahl sich öffnender Klammern an einem bestimmten Punkt im Satz. Dennoch ist das Maß in einem lokalen Kontext meines Erachtens nicht zu unterschätzen. So erhöht jede Phrase die syntaktische Komplexität eines Satzes, sobald sie eingeführt wird. Nominalphrasen können neue Referenten einführen oder bereits vorhandene wieder aufgreifen (siehe hierzu auch die Analysen aus Abschnitt 7.4), Verbalphrasen können neue Aktionen, ausgedrückt durch Verben, spezifizieren. Präpositionalphrasen können Diskursrepräsentationen mit zusätzlichen Informationen anreichern (was auch für Adjektivphrasen gilt) oder Beziehungen zwischen Konzepten ausdrücken. Die gefundenen Effekte deuten darauf hin, dass die Neueinführung solcher Phrasen Verarbeitungsaufwand bedeutet. Darüber hinaus gilt: Je mehr Phrasen auf einmal eingeführt werden, desto aufwändiger wird es.

Groups	Name	Variance	Std.Dev.			
text.id	(Intercept)	1.5277e-10	0.00001236			
Residual		1.5097e-01	0.38855042			
		<b>Estimate</b>	<b>Std. Error</b>	<b>t value</b>	<b>p<sub>MCMC</sub></b>	
(Intercept)		0.031365	0.015546	2.018	0.0340	*
rel.pos		-0.034594	0.018643	-1.856	0.0644	.
res.einb.tiefe		0.001028	0.002974	0.346	0.7872	
öffnende.klammern		0.025225	0.009186	2.746	0.0090	**

Tab. 7.12: Lineares gemischtes Modell für den Einfluss der Anzahl sich öffnender Klammern auf die *first-pass reading times*, residualisiert nach Effekten der lexikalischen Ebene sowie Bi- und Trigrammfrequenzen und -übergangswahrscheinlichkeiten.

Allerdings ist nicht zu erwarten, dass jede Phrase gleich erwartet oder unerwartet geöffnet wird. Und diese Erwartung sollte sich eigentlich auf die Verarbeitungsgeschwindigkeit auswirken. Betrachten wir Beispiel 25:

- (25) a. [S [NP Die Katze] jagt [NP die Maus]].  
 b. [S [NP Die Katze] schläft [NP einige Stunden]].

Es ist anzunehmen, dass die zweite NP 'die Maus' in Satz 25a relativ gut vorhersagbar ist. Das Verb 'jagen' kommt vermutlich häufiger mit Akkusativobjekt vor als ohne ein solches. Da 'die Katze' Agens der Jagdhandlung ist, ist eventuell auch noch die konkrete Lexikalisierung 'die Maus' dieser zweiten NP relativ gut vorhersagbar.<sup>117</sup> Im Vergleich zu Satz 25b sollte klar werden, dass die NP 'einige Stunden' nach dem Hauptverb 'schläft' weitaus weniger vorhersagbar ist als die NP 'die Maus' in Satz 25a. Ein Blick in das Google-Books-Korpus (Michel et al. 2011) bestätigt eine solche Verteilung für diese beiden konkreten Verbformen. Für eine Suche in deutschen Büchern von 1950 bis 2008 taucht die Form 'jagt' zwar durchgängig seltener auf als 'schläft'. Die Folge 'jagt <Artikel> <Nomen>' ist jedoch konsistent häufiger als die Formen 'schläft <Artikel> <Nomen>' oder 'schläft <Pronomen> <Nomen>'.<sup>118</sup> Damit ist davon auszugehen, dass die letzte NP in Satz 25a vorhersagbarer ist als jene in Satz 25b.

Solche Unterschiede in der Vorhersagbarkeit neu initiiertter Phrasen wurden in den obigen Analysen nicht beachtet, sollten aber ebenfalls Auswirkungen auf die Lesezeit haben. Der verlangsamende Effekt am phrasenöffnenden Lexem sollte dann für Phrasen, die unerwartet sind, stärker sein. Gegebenenfalls könnte sich dieser Effekt für stark vorhersagbare Phrasen umdrehen. Eine solche Argumentation ist nicht neu. So argumentieren beispielsweise MacDonald/Pearlmutter/Seidenberg (1994) im Kontext von Ambiguitätsauflösung und Verbarargumentstrukturen:

The argument structures associated with a word encode the relationships between the word and the phrases that typically occur with it (the word's arguments) and capture important facts about correlations between syntactic and semantic information. [...] For the verb *put*, the argument structure information captures English speakers' knowledge that not just any combination of a subject NP, an object NP and a PP is acceptable; the subject NP must take the role of agent (the thing doing the putting), the object NP must be the theme (the thing being put), and the PP must specify a location (where the theme is put). [Hervorhebung im Original, S.W.]

Diese Argumentation erinnert an die *subcategorization frames/rules*, die bereits von Chomsky (1965, S. 75ff.) in den Diskurs eingeführt wurden. Obligatorische Verbarargumente sind der prototypische Fall, in dem Phrasen hochvorhersagbar sind. Konieczny (2000) zeigt, dass Verben unter bestimmten Umständen erwartbarer werden, wenn die Argumente zuvor alle genannt wurden.

<sup>117</sup> Für eine konnektionistische Modellierung, die unter anderem dieses Phänomen in *simple recurrent networks* (SRNs) zeigt, siehe Elman (2009).

<sup>118</sup> Die entsprechende Abfrage ist einsehbar unter <http://goo.gl/eEu9Eo> (Stand: 16.3.2015).

Im Deutschen ist dies bei Nebensätzen mit Verb-Endstellung der Fall. Antizipation schlägt sich dann in schnelleren Lesezeiten auf dem Verb nieder. Der Effekt wirkt hier in die andere Richtung: Nicht das Verb öffnet einen *subcategorization frame*, der bestimmte Argumente erwarten lässt, sondern die Verb-Argumente helfen dabei, das Verb vorherzusagen. Eine letzte Verfeinerung der Analyse bietet sich hier also an: Insbesondere Nominalphrasen sollten zumindest unter bestimmten Bedingungen vorhersagbarer sein als andere Phrasentypen, da durch sie oft die Argumente von Verben realisiert werden, die gegebenenfalls durch die *subcategorization frames* von Verben vorhersagbar werden. Für eine solche Analyse werden zunächst alle Wörter selektiert, zu deren Zeitpunkt sich eine Phrase öffnet (im Modell aus Tabelle 7.11 sind das alle Wörter, für die *phrase.geöffnet* den Wert *ja* annimmt). Jedem Wort wird dann die Information hinzugefügt, ob im Übergang zu diesem Wort einer der folgenden Phrasentypen geöffnet wird: Nominalphrasen (NP), Adjektivphrasen (AdjP), Adverbialphrasen (AdvP), Präpositionalphrasen (PP), Verbalphrasen (VP) oder komplette Satzknotten (S). Diese Information kann aus der Phrasenstrukturannotation abgeleitet werden. Für jeden dieser Phrasentypen wird dann ein separates gemischtes Modell mit den Prädiktoren 'relative Position des Wortes im Satz', 'residuale Einbettungstiefe' und 'Öffnen einer Phrase des entsprechenden Typs' berechnet. Die Effektschätzer für diesen letzten Prädiktor sind in Abbildung 7.10 abgetragen. Wie sich relativ klar zeigt, sind die Effektschätzer fast aller Phrasentypen positiv.

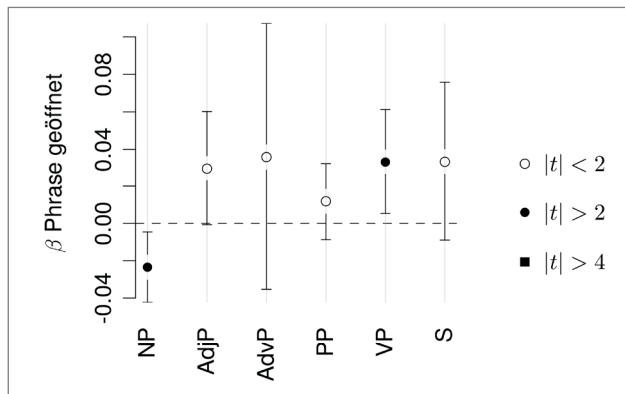


Abb. 7.10: Effektschätzer und deren Konfidenzintervalle für das Öffnen des jeweiligen Phrasentyps und residuale FPRTs. Assoziierte t-Werte sind über Symbole abgetragen. Konfidenzintervalle ergeben sich aus 1,96 Standardfehlern.

Der einzige Phrasentyp, der aus diesem Muster herausfällt, ist die Nominalphrase. Für Nominalphrasen gilt: Wenn an einem Wort mindestens eine NP

geöffnet wird, und mindestens eine dieser Phrasen eine NP ist, wird dieses Wort schneller gelesen ( $t = -2,44$ ;  $p_{\text{MCMC}} = 0,015$ ). Dies wiederum könnte lediglich ein Wortarten-Effekt sein, denn natürlich werden Nominalphrasen in der Mehrheit der Fälle von Artikeln eingeleitet. Das ist in 71,2% aller Fälle so, an zweiter Stelle rangieren hier normale Nomen mit 7,73%). Ein solcher Effekt der Wortart kann überprüft werden, indem man den Part-of-Speech-Tag (POS-Tag) des phrasenöffnenden Wortes zur Kontrolle als *random intercept*<sup>119</sup> mit in das Modell aufnimmt. Dadurch wird die Varianz, die durch die Wortart aufgeklärt wird, kontrolliert und steht für den Effekt des Phrasentyps nicht mehr „zur Verfügung“. Es zeigt sich, dass auch nachdem für den Effekt der Wortart kontrolliert wird, das Öffnen einer Nominalphrase einen signifikanten beschleunigenden Effekt auf die FPRTs hat ( $t = -1,98$ ;  $p_{\text{MCMC}} = 0,047$ ). Auch wenn nur jene Fälle ausgewählt werden, in denen tatsächlich **nur** eine NP geöffnet wird und die NP nicht eine von mehreren Phrasen ist, die an einem bestimmten Punkt geöffnet werden, zeigt sich dasselbe Muster. Das gilt sowohl für das Modell ohne POS-Tag als *random intercept* ( $t = -2,76$ ;  $p_{\text{MCMC}} = 0,006$ ), als auch für jenes mit POS-Tag als *random intercept* ( $t = -2,43$ ;  $p_{\text{MCMC}} = 0,015$ ).

Unter der Voraussetzung, dass Nominalphrasen oft Argumente von Verben kodieren, lassen sich die hier gefundenen Effekte als Antizipations- oder Vorhersageeffekte interpretieren. Im Gegensatz zu allen anderen Phrasentypen sind es nur Nominalphrasen, die bei deren Öffnung beschleunigende Effekte auf die *first-pass reading times* haben. Eine andere Art von Vorhersageeffekten wird außerdem am Ende von Abschnitt 7.5 vorgestellt. Dort wird für den Reformulierungsteil des FLRC gezeigt werden, welche Effekte die Distanz zwischen einem Verb und dem dazugehörigen Subjekt hat. Diese Frage ist ebenfalls von zentralem Interesse in der psycholinguistischen Forschung und ging als (Anti-)Lokalitätsdebatte in die Literatur ein.

### 7.3.5 Ähnlichkeitsbasierte Interferenz

In einer letzten Analyse in diesem Abschnitt wird der Fokus von strukturell geprägten Prädiktoren hin zu einem Ansatz verschoben, der die Architektur des menschlichen Arbeitsgedächtnisses und deren Einfluss auf die Sprachverarbeitung in das Zentrum stellt. Hauptsächlich stehen dabei Interferenzeffekte im Mittelpunkt, die aus der Ähnlichkeit gespeicherter Informationseinheiten entstehen. Gordon et al. (2001, 2006) zeigen dabei in unterschiedlichen experimentellen Setups, dass die Anzahl und der Typ von Nominalphrasen vor einem Verb einen messbaren Einfluss auf die Verarbeitung dieses Verbs haben.

<sup>119</sup> Die Aufnahme dieser Variable als *random intercept* ist im ursprünglichen Sinne von Zufallseffekten eigentlich nicht vorgesehen. Dennoch erfüllt sie den Zweck der Varianzkontrolle.

Die Stimuli sind so konstruiert, dass die Nominalphrasen, die die Verbargumente kodieren, beide vor dem Verb genannt sind. Sind diese Nominalphrasen beide nach der Form <Artikel> <Berufsbezeichnung> aufgebaut, dauert die Integration am Verb länger, als wenn eine der Nominalphrasen in Form eines Eigennamens bzw. eines Personalpronomens realisiert wird. Die Ähnlichkeit der präverbalen<sup>120</sup> Nominalphrasen wird bei ihrem Abruf relevant: Je ähnlicher sich die beiden Nominalphrasen sind, desto ähnlicher sind sich die mentalen Repräsentationen der von den NPs kodierten Referenten. Damit konkurrieren sie stärker um den Abruf, was den Abruf des korrekten, an das Verb zu bindenden Referenten erschwert, weil andere „gute Kandidaten“ im gleichen Satz genannt werden. Eine kritische Frage stellt sich dennoch: Welche Rolle spielt das Verb? Genauer: Welchen Einfluss haben die semantischen Kategorien, die ein Verb für einzelne thematische Rollen verlangt? Van Dyke/McElree (2006) gehen dieser Frage in einem Kontext nach, wo kritische Distraktoren, also die **nicht** an das Verb zu bindenden Referenten, nicht im Satz selbst genannt werden, sondern vor dem Lesen des Satzes auswendig gelernt werden mussten. Passt das Verb im Stimulussatz gut zu diesen Distraktoren, wird es signifikant langsamer gelesen. Van Dyke (2007) erweitert dieses Ergebnis in zwei Richtungen: 1) Sie benutzt Satzmaterial, in dem die interferierenden Nominalphrasen innerhalb des Satzes genannt werden und nicht vorher auswendig gelernt werden müssen. 2) Das Satzmaterial variiert die syntaktische und semantische Passung der interferierenden Nominalphrasen. Beispiel 26 zeigt das von Van Dyke (2007) verwendete Stimulusmaterial. Variiert wird die Form des Relativsatzes (hervorgehoben) sowie die Animiertheit der ersten im Relativsatz genannten NP (‘warehouse’ vs. ‘neighbor’). Der kritische Verbkomplex, an dem die Integrationskosten über Lesezeiten gemessen werden, ist durch Fettdruck hervorgehoben. Korrektes Agens der Beschwerdeaktion wäre in beiden Fällen ‘the (dangerous) resident’, interferierende NPs sind in beiden Fällen ‘the worker’ und ‘the (dangerous) warehouse/neighbor’, beachtet wird im folgenden lediglich die Variation der zweitgenannten NP.

- (26) a. The worker was suprised that the resident who *was living near the dangerous warehouse/neighbor* **was complaining** about the investigation.
- b. The worker was suprised that the resident who *said that the warehouse/neighbor was dangerous* **was complaining** about the investigation.

<sup>120</sup> Mir ist bewusst, dass das Wort ‘präverbal’ mit „vor dem Spracherwerb liegend“ besetzt ist. In Ermangelung einer ähnlich kurzen Alternative verwende ich den Begriff trotzdem. Hier ist er als „vor dem/einem Verb im Satz genannt“ zu interpretieren.



Variiert werden die syntaktische und semantische Passung der interferierenden Nominalphrase. Die syntaktische Variation besteht in Beispiel 26 zwischen den Sätzen 26a und 26b. Das Lagerhaus bzw. der Nachbar haben in Satz 26b im Satzkomplement ‘the warehouse/neighbor was dangerous’ Subjektstatus, im Relativsatz in Satz 26a dagegen nicht. Diesen Subjektstatus hat auch die korrekterweise an das Verb zu bindende Nominalphrase ‘the resident’. Die Interferenz sollte also aufgrund der höheren Ähnlichkeit bezüglich des Subjektstatus höher sein. Der zweite Faktor ist die semantische Passung, die in beiden Sätzen aus Beispiel 26 jeweils über die Variation zwischen ‘warehouse’ und ‘neighbor’ realisiert ist. Die Idee ist: Ein Lagerhaus ist kein gutes Agens einer Beschwerdeaktion, ein Nachbar hingegen schon. Global gesehen sollte dies keinen Unterschied machen, weil die syntaktische Position, in der diese NP steht, keine Anbindung an ‘was complaining’ zulässt. Das Mehr an Interferenz gegenüber dem Lagerhaus funktioniert in dieser Variation also nur über die semantische Ähnlichkeit (die Animiertheit) zwischen dem Nachbarn und dem Bewohner (‘resident’).

Erklärt werden die Effekte anhand eines kognitiven Modells von Lewis/Vasishth (2005), das innerhalb der kognitiven Architektur ACT-R (Anderson et al. 2004) realisiert ist. Ziel ist die Integration des Verbs mit seinen Argumenten. Wann immer eine NP gelesen wird, wird sie im deklarativen Gedächtnis abgelegt. Kodiert ist für diese abgelegten *chunks* dann unter anderem, ob die NP eine Subjektmarkierung aufweist, einen belebten Referenten kodiert<sup>121</sup> oder bereits an ein Verb gebunden ist. Sobald das Verb des Satzes gelesen wird, generiert dieses Abrufhinweise für seine Argumente, daher der Terminus *cue-based parsing* für Modelle dieser Klasse. Diese Abrufhinweise lösen eine Suche nach passenden Gedächtniseinheiten (*chunks*) im deklarativen Gedächtnis aus. Alle im deklarativen Gedächtnis abgelegten *chunks* konkurrieren um den Abruf.

Jene, die zusätzlich syntaktische oder semantische Selektionsbeschränkungen der Abrufhinweise erfüllen (beispielsweise Animiertheit), haben eine höhere Chance, abgerufen zu werden. Am Ende wird jene NP abgerufen, die im globalen Satzkontext die korrekte ist – denn nur bei deren Repräsentation im deklarativen Gedächtnis ist noch keine Bindung an ein Verb vermerkt. Ein Grundprinzip der zugrunde liegenden kognitiven Architektur ist, dass die Abrufgeschwindigkeit unter anderem von der Anzahl und Passung potenzieller Mitbewerber um den Abruf bestimmt wird. Ein anderer Einflussfaktor ist die Aktivierung des abgerufenen *chunks*. Diese Aktivierung zerfällt über die Zeit. In Beispiel 26 ist auch das ausschlaggebend, weil die korrekte NP ‘the

<sup>121</sup> Der semantische Aspekt ist bei Lewis/Vasishth (2005) noch nicht implementiert.

resident' weiter vom kritischen Verb entfernt ist als die Distraktor-NP 'the (dangerous) warehouse/neighbor'. Die Aktivierung für die Repräsentation der Distraktor-NP liegt deshalb leicht höher. Das heißt: Passt eine Distraktor-NP sowohl syntaktisch als auch semantisch zu den vom Verb generierten Abrufhinweisen, besteht maximale Interferenz – auch wenn sie eigentlich schon an ein Verb gebunden ist.

Anhand des FLRCs soll erstmals<sup>122</sup> die grundlegende Hypothese dieser Gruppe von Ansätzen bei der Verarbeitung von natürlichen Texten überprüft werden. Die bisherigen Arbeiten zur ähnlichkeitsbasierten Interferenz (Gordon et al. 2001; Gordon/Hendrick/Johnson 2004; Gordon et al. 2006) bzw. zur abrufbasierten Satzverarbeitung (Van Dyke 2007) zeigten entsprechende Effekte immer in hochkontrollierten experimentellen Stimulusätzen. Die interne Validität dieser Studien ist hoch. Jedoch stellt sich auch hier die Frage, inwieweit die Erkenntnisse auf die Verarbeitung natürlicher Texte übertragbar sind. Was die eben vorgestellten Ansätze grundsätzlich eint, ist die Annahme, dass präverbale Referenten einen Einfluss auf die Verarbeitungszeit finiter Verben haben sollten. Dabei ist zunächst auch nicht unbedingt ausschlaggebend, wie viele dieser Referenten semantisch passen oder die korrekte syntaktische Kategorie aufweisen. Sollten bestimmte Eigenschaften präverbaler Referenten nicht passen, fungieren diese lediglich als weniger gute Distraktoren.

Für das FLRC wurde für jedes finite Vollverb (POS-Tag VVFIN), Auxiliar (VAFIN) und finite Modalverb (VMFIN) annotiert, wie viele Nomen (NN), Eigennamen (NE), Personalpronomen (irreflexiv: PPER; reflexiv, PRF), substituierende Demonstrativpronomen (PDS) vor diesem Verb im Satz genannt wurden. All diese Elemente werden im Folgenden als „präverbale Referenten“ bezeichnet. Eine solche Sprachregelung kann gegebenenfalls die tatsächliche Anzahl an Referenten im engeren Sinne überschätzen. Einerseits gehen alle Abstrakta mit in die Zählung ein. Das ist in diesem Kontext gewünscht, denn auch Abstrakta können als Subjekte von Verben fungieren. Andererseits ist es auch möglich, dass ein Pronomen sich auf einen Referenten bezieht, der bereits im gleichen Satz genannt wurde. In einem solchen Fall wäre tatsächlich **ein** Referent vorhanden, das darauf referierende Pronomen fügt zur hier verwendeten Zählung allerdings eine Einheit hinzu (siehe hierzu auch Beispiel 27). Der Terminus „präverbale Referenten“ sollte also in diesem Sinne nicht zu streng aufgefasst werden. Theorien zur abrufbasierten Satzverarbeitung

---

<sup>122</sup> Von Müller-Feldmeth/Wolfer/Konieczny (2013) und Wolfer et al. (2013) wurden für das *PopSci*-Lesekorpus entsprechende Effekte ebenfalls für das Deutsche analysiert. Das FLRC entstand zeitlich vor dem *PopSci*-Lesekorpus, daher spreche ich in der vorliegenden Arbeit davon, dass es erstmals diese Möglichkeit bietet.

machen zu solchen Konstellationen keine expliziten Aussagen, es ist aber denkbar, dass ein Pronomen auch dann als Distraktor agieren kann, wenn sein Antezedens bereits genannt wurde. Durch die Zählung all dieser Wortarten kann die Anzahl relevanter präverbaler Elemente recht groß werden, insbesondere bei längeren Sätzen. Abbildung 7.11 gibt einen Eindruck, wie sich die finiten Verben im FLRC bezüglich der Anzahlen präverbaler Referenten verteilen.

Etwa 100 Verben (8,6%) haben keinerlei relevanten Elemente vor sich im gleich Satz. Die Mehrheit der Verben (25,4%) haben genau einen Referenten vor sich. Dann nimmt die Zahl stark ab. Die Hälfte aller Verben haben keine und bis zu drei präverbale Referenten. Zählt man Verben ohne präverbale Referenten nicht mit, liegt die Hälfte aller Verben in einem Bereich von einem bis fünf präverbale Referenten (in Abbildung 7.11 über gepunktete Linien gekennzeichnet). Wie hoch die Anzahl der relevanten präverbale Elemente nach der eingesetzten Zählung sein kann, zeigt sich am rechten Ende der Verteilung. Vor einem Verb stehen nach oben dargelegter Zählung 27 potenzielle Referenten. Im entsprechenden Satz, der aus dem Cannabis-Urteil stammt, wirken mehrere Faktoren zusammen, die zu einer so hohen Zählung führen können: 1) Der Satz enthält viele Verweise auf Gesetzesstellen. Absätze, Sätze und Nummern und diese Abkürzungen wurden mitgezählt. 2) Das entsprechende Verb steht in einem Konditionalsatz mit Verbletzstellung. 3) Der Satz enthält viele Nominalisierungen, die – würden sie als verbale Strukturen realisiert – nicht in die Zählung einfließen würden.

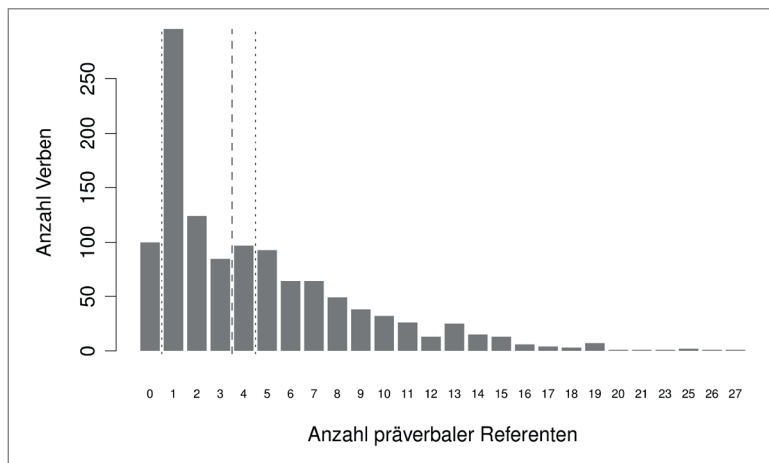


Abb. 7.11: Verteilung der Anzahlen der präverbale Referenten im FLRC. Vor der gestrichelten Linie befinden sich 52,1% der Verben. Zwischen den beiden gepunkteten Linien befinden sich 51,8% der Verben.

Zunächst werden nun alle Verben im Datensatz selektiert, die mindestens einen präverbalen Referenten haben, wobei auch im Falle von einem präverbalen Referenten nicht sichergestellt ist, dass dies ein Referent ist, der an das entsprechende Verb gebunden werden muss. Finite Verben ohne präverbale Referenten treten logischerweise in (Neben-)Sätzen und Fragen mit V1-Stellung auf (siehe Beispiel 27 aus der Pressemitteilung zum unerlaubten Aufenthalt in Deutschland (Bundesverfassungsgericht 2003a), die Anzahl relevanter präverbaler Elemente ist jeweils indiziert). 'Besitzt' hat keine, 'bleibt' hat zwei ('er', 'Duldung') und 'gilt' hat vier präverbale Referenten ('er', 'Duldung', 'er', 'Bundesgebiet'). Auch in bestimmten Adverb- oder Pronominaladverb-Konstruktionen kann der Fall auftreten, dass keine Referenten vor dem finiten Verb stehen (siehe Beispiel 28, das ebenfalls der Pressemitteilung zum unerlaubten Aufenthalt in Deutschland entnommen ist). Da Pronominaladverbien ebenso wie Pronomen die grammatische Funktion von einzelnen Satzgliedern übernehmen können, stellen sie einen Grenzfall in der Zählung präverbaler Referenten dar. Das FLRC bietet daher die Möglichkeit, diese in die Zählung mit aufzunehmen oder nicht. Es wird jene Variante gewählt, bei der sie nicht in die Zählung relevanter präverbaler Elemente aufgenommen werden.

- (27) *Besitzt*<sub>0</sub> er keine Duldung und *bleibt*<sub>2</sub> er gleichwohl im Bundesgebiet, *gilt*<sub>4</sub> die strafrechtliche Regelung des § 92 Abs. 1 Nr. 1 AuslG.
- (28) Hiergegen *setzte*<sub>0</sub> er sich mit seiner Vb [= Verfassungsbeschwerde] zur Wehr.

Um Extremfälle auszuschließen, selektieren wir nur finite Verben, die höchstens zehn relevante Elemente vor sich im Satz haben. Das umschließt 81,1% der Fälle. Die Grenze von zehn präverbalen Referenten wurde gewählt, um zumindest ansatzweise eine Verallgemeinerbarkeit über die hier untersuchte Textsorte zu ermöglichen. Das oben dargelegte Beispiel zeigte, dass zumindest zwei der Gründe für extrem viele präverbale Referenten (viele Nominalisierungen sowie viele Verweise auf Gesetzesstellen) ein Merkmal der Textgattung Rechtsprache sein könnten. Sollte diese Selektion das Ergebnis einer hier vorgestellten Analyse beeinflussen, wird das an entsprechender Stelle erwähnt.

Van Dyke (2007) findet mehr oder weniger starke, aber stets signifikante Effekte für alle untersuchten Lesezeitmaße. Das gilt zumindest für die Manipulation der syntaktischen Interferenz. Semantische Interferenz (animierte vs. nicht-animierte Distraktor-NP) scheint einen weniger starken Einfluss auf die von ihr gemessenen Lesezeiten gehabt zu haben (vgl. ebd., S. 416f. sowie S. 422f.). Zunächst werde ich die Lesezeiten beim ersten Lesen (*first-pass reading times*) der finiten Verben analysieren. Ich räume zusätzlich die Möglichkeit ein, dass sich der Effekt der präverbalen Referenten über die Einbettungstiefe des Verbs

hinweg verändert. Hierzu nehmen wir die Interaktion der Faktoren Einbettungstiefe  $\times$  präverbale Referenten mit in das Modell auf. Zunächst wird der Prädiktor, der die Anzahl der präverbale Referenten erfasst, nach Position und Einbettungstiefe residualisiert, da diese Faktoren kollinear sind: Logischerweise befinden sich vor Verben, die weiter hinten im Satz stehen, mehr relevante präverbale Elemente ( $r = 0,72$ ;  $p = \min p$ ). Tiefer eingebettete Verben stehen weiter hinten im Satz, also ist die Anzahl präverbale Referenten höher, wenn Verben tiefer eingebettet sind ( $r = 0,57$ ;  $p = \min p$ ).

Groups	Name	Variance	Std.Dev.			
text.id	(Intercept)	0.0028203	0.053106			
Residual		0.1723724	0.415178			
		Estimate	Std. Error	<i>t</i> value	$p_{\text{MCMC}}$	
(Intercept)		0.022755	0.040441	0.563	0.5258	
rel.pos		-0.014975	0.068575	-0.218	0.8320	
einb.tiefe		-0.011791	0.012908	-0.913	0.3538	
res.n.pverb.ref		0.061299	0.015859	3.865	0.0002	***
einb.tiefe:res.n.pverb.ref		-0.021994	0.004927	-4.464	0.0001	***

Tab. 7.13: Gemischtes Modell für residuale *first-pass reading times* auf finiten Verben in Abhängigkeit von relativer Position im Satz (*rel.pos*), Einbettungstiefe (*einb.tiefe*), residueller Anzahl präverbale Referenten (*res.n.pverb.ref*) sowie der Interaktion Einbettungstiefe  $\times$  präverbale Referenten.

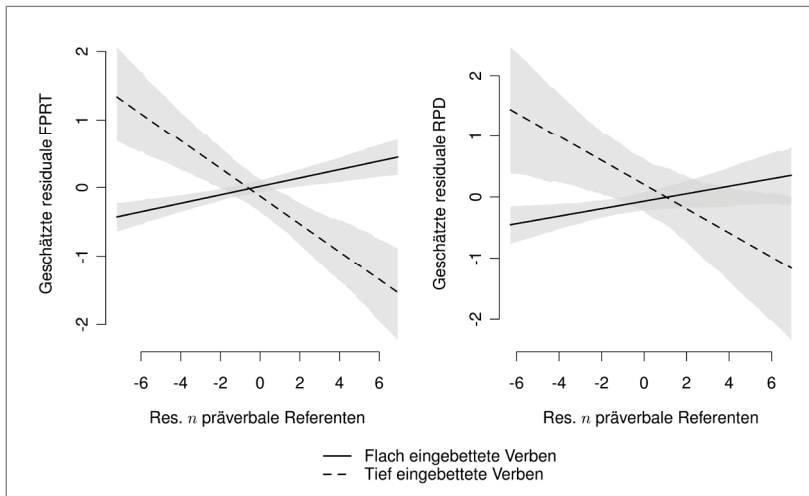


Abb. 7.12: Interaktionseffekte auf FPRTs (links) und RPDs (rechts): Residuale Anzahl relevanter präverbale Elemente  $\times$  Einbettungstiefe.

Das Modell für residuale FPRTs ist in Tabelle 7.13 zu sehen. Die residuale Anzahl präverbaler Referenten zeigt einen deutlich verlangsamenden Effekt: Je mehr relevante Elemente vor einem finiten Verb im Satz stehen, desto länger wird dieses Verb gelesen. Um diesen Effekt vollständig interpretieren zu können, muss man jedoch die Modulation durch die Einbettungstiefe beachten. Das Muster des Interaktionseffekts, das auch aus Abbildung 7.12 ersichtlich ist, ist recht interessant: Erwartet wurde ein verlangsamender Effekt der Anzahl präverbaler Referenten. Je mehr Referenten vor einem Verb stehen, desto mehr potenzielle Distraktoren sind vorhanden, wenn die korrekten Argumente abgerufen werden müssen. Diese Vorhersage scheint sich allerdings nur für flach eingebettete Verben zu bewahrheiten (durchgezogene Linien in Abb. 7.12). Für tief eingebettete Verben scheint sich dieser Effekt umzudrehen, oder zumindest zu nivellieren (gestrichelte Linien in Abb. 7.12). Abbildung 7.12 zeigt bereits, dass dieses Effektmuster auch für *regression path durations* zu beobachten ist, aber etwas verrauschter ist.<sup>123</sup>

Groups	Name	Variance	Std.Dev.		
text.id	(Intercept)	8.5158e-01	0.922812		
vtn	(Intercept)	1.6037e-01	0.400463		
vtn	trial	1.3966e-05	0.003737		
	<b>Estimate</b>	<b>Std. Error</b>	<b>z value</b>	<b>p value</b>	
(Intercept)	-1.762653	0.151577	-11.629	< 2e-16	***
rel.pos	1.756744	0.133748	13.135	< 2e-16	***
einb.tiefe	0.005422	0.027645	0.196	0.844510	
res.n.pverb.ref	0.105937	0.029284	3.618	0.000297	***
trial	-0.002266	0.001263	-1.794	0.072834	.
einb.tiefe:res.n.pverb.ref	-0.029647	0.009499	-3.121	0.001803	**

Tab. 7.14: Logistisches gemischtes Modell für die Wahrscheinlichkeit einer regressiven Sakkade auf finiten Verben in Abhängigkeit von relativer Position im Satz, Einbettungstiefe, residualer Anzahl präverbaler Referenten (*res.n.pverb.ref*) sowie der Interaktion Einbettungstiefe  $\times$  präverbale Referenten.

Eine weitere abhängige Variable ist noch interessant, nämlich die Wahrscheinlichkeit, dass eine regressive Sakkade initiiert wird, wenn das Verb fixiert wird. Wenn mehr präverbale Referenten während des Argumentabrufs aus dem deklarativen Gedächtnis interferieren, sollte das zu einer wachsenden Verunsicherung bei der Leserin/dem Leser führen. Dies könnte sich auf die Wahrscheinlichkeit auswirken, dass ein regressiver Blicksprung initiiert wird.

<sup>123</sup> Der Effektschätzer bei RPDs für die residuale Anzahl präverbaler Referenten liegt bei  $\beta = 0,06$  ( $SE = 0,03$ ). Das ergibt  $t = 2,18$  mit  $p_{MCMC} = 0,03$ . Für die Interaktion *einb.tiefe*  $\times$  *res.n.pverb.ref* gilt:  $\beta = -0,02$ ;  $SE = 0,008$ ;  $t = -2,59$ ;  $p_{MCMC} = 0,01$ .

Eventuell könnte auch dieser Effekt von einer Interaktion mit der Einbettungstiefe des Verbs moduliert werden. Tabelle 7.14 zeigt die Ergebnisse für die Wahrscheinlichkeit, dass auf einem finiten Verb eine regressive Sakkade initiiert wird. Erwartbar ist der Effekt der relativen Position des Verbs im Satz (*rel.pos*). Je weiter hinten ein Wort im Satz steht, desto mehr Material steht vor ihm, auf das zurückgesprungen werden kann. Die anhand des Modells geschätzte Wahrscheinlichkeit, dass von einem Verb im Satz zurückgesprungen wird, steigt mit der relativen Position dieses Verb von rund 13% am Anfang des Satzes auf rund 47% am Ende des Satzes. Interessant ist wiederum die letzte Zeile aus Tabelle 7.14. Es zeigt sich, dass auch für diese Variable der Effekt der Anzahl relevanter präverbaler Elemente durch die Einbettungstiefe des Verbs moduliert wird.

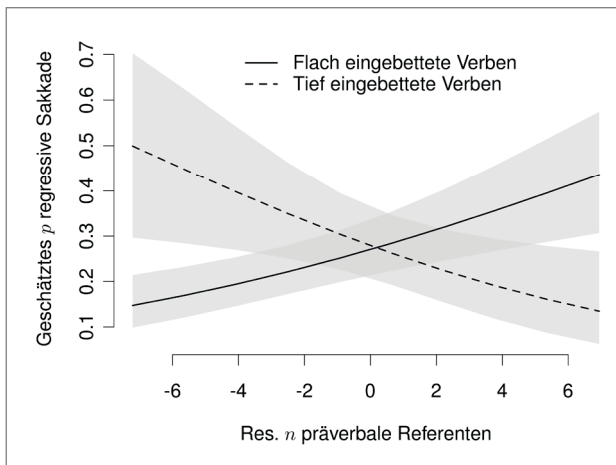


Abb. 7.13: Interaktionseffekt auf die Wahrscheinlichkeit einer regressive Sakkade: Residuale Anzahl relevanter präverbaler Elemente  $\times$  Einbettungstiefe.

Abbildung 7.13<sup>124</sup> gibt einen Eindruck dieser Interaktion. Wie der Skalierung der  $y$ -Achse zu entnehmen ist, sind diese Effekte recht groß: Für flach eingebettete Verben steigt die geschätzte Wahrscheinlichkeit einer regressive Sakkade von rund 15% für Verben mit sehr wenigen relevanten Elementen vor ihnen im Satz auf 44% für Verben mit vielen präverbale Referenten. Eine Konfundierung mit der Position der entsprechenden Verben im Satz ist relativ unwahrscheinlich, da dieser Effekt im logistischen Regressionsmodell bereits berücksichtigt und die Anzahl relevanter präverbaler Elemente unter anderem anhand dieses Faktors residualisiert wurde.

<sup>124</sup> Die Effektschätzer und Konfidenzintervalle sind mithilfe des R-Pakets *effects* (Fox 2003) extrahiert.



Alle vorgestellten Analysen zur Anzahl präverbaler Referenten deuten darauf hin, dass Verben in der Tat schwieriger zu verarbeiten sind, wenn mehr mögliche Argumente vor ihnen im Satz stehen. Dieser Effekt wird vorhergesagt von Theorien der ähnlichkeitsbasierten Interferenz bzw. abrufbasierten Satzverarbeitung, wie sie in diesem Abschnitt vorgestellt wurden. Diese Modelle sagen jedoch nicht vorher, dass diese Effektrichtung nur für flach eingebettete Verben gilt. Die Ergebnisse zu tief eingebetteten Verben sind etwas schwierig zu interpretieren. Es ist anhand der vorliegenden Analysen nicht gesichert, ob sich der Effekt für diese tatsächlich umkehrt, dass sie also mit steigender Anzahl präverbaler Referenten schneller gelesen werden. Die Abbildungen 7.12 und 7.13 legen dies zwar nahe, zeigen aber auch, dass dieser Effekt mit mehr Rauschen behaftet ist als jener für flach eingebettete Verben. Um diesem Teil der Interaktion auf den Grund zu gehen, kann die Analyse auf tief eingebettete Verben eingeschränkt werden. Die beschleunigenden Effekte für Verben mit einer Einbettung tiefer als der vierten Ebene können in dieser Analyse bestätigt werden.

Eine Hypothese, die dieses Muster erklären könnte, besagt: Satzverarbeitungsprozesse laufen für flach eingebettete finite Verben „normal“ ab. Zu dieser Normalität gehört unter anderem die Interferenz möglicher Abrufkandidaten bei der Argument-Verb-Integration. Im Allgemeinen läuft auch die Verarbeitung tiefer eingebetteter Verben in diesem „normalen Rahmen“ ab. Sobald aber die Interferenz der präverbaler Referenten zu hoch wird, ist die Ressourcenbelastung aufgrund tieferer Einbettung **und** erhöhter Interferenz während der Argumentintegration zu hoch. Der menschliche Satzverarbeitungsmechanismus operiert dann in einem Modus, der als „oberflächliche Satzverarbeitung“ (*shallow parsing*) bezeichnet werden kann. Diese oberflächliche Satzverarbeitung kann zu einer Repräsentation des Satzes führen, die „gut genug“ (*good enough*) ist (siehe hierzu Arbeiten von Ferreira/Bailey/Ferraro 2002; Ferreira/Patson 2007; Sanford/Sturt 2002). Oberflächliche Verarbeitung äußert sich dann in sinkenden *first-pass reading times*. Um diese Hypothese empirisch untermauern zu können, wären weitere Datenerhebungen notwendig. Es leitet sich direkt die Vorhersage ab, dass die Repräsentationen, die bei der oberflächlichen Verarbeitung der Sätze entstehen, in Teilen falsch oder unvollständig sind. Das würde bedeuten, dass Argumente falsch oder unvollständig an Verben gebunden werden. Um dies überprüfen zu können, wären allerdings bessere Verständnistests notwendig, als sie das FLRC zur Verfügung stellt. Ferreira/Patson (2007, S. 71) umschreiben die Grundidee des „Good Enough“-Ansatzes folgendermaßen:

In 2002, Ferreira and colleagues suggested that the representations created during language comprehension were just ‘good enough’ [...]. The fundamental

idea was to challenge the assumption that the overarching goal of the language comprehension system is to deliver an accurate and detailed representation of the speaker's utterance. Ferreira et al. argued that instead the system's responsibility is to create representations that are suitable for the task the listener wants to perform with the linguistic input. Typically, the listener's task is to provide an appropriate followup to the speaker's contribution (i.e., to maintain a dialogue), but of course there are other possibilities – for example the listener might be required to execute a motor action, or merely nod politely at appropriate moments. Their point was that listeners are rarely required to prove the accuracy or detailed nature of their understanding of some utterance, and this might have important implications for the architecture of the language processing system.

„HörerInnen“ können hier auf „LeserInnen“ erweitert werden. Ferreira et al. (2002) verwendeten ebenfalls Materialien in Schriftform wie beispielsweise den Holzwegsatz in Beispiel 29. Interessanterweise haben die meisten TeilnehmerInnen auf die Frage, ob Mary das Baby gebadet habe, positiv geantwortet. Hier liegt eine Falschanbindung von Verbargumenten vor, die noch bis zum „gefühlten Verstehen“ des Satzes anhält. 'The baby' wird also als Objekt an 'bathed' gebunden. Tatsächlich ist es jedoch das Subjekt von 'played'. Mary badet und das Baby spielt – Mary badet aber nicht das Baby.

(29) While Mary bathed the baby played in the crib.

Es gibt weitere Ansätze mit einer ähnlichen Stoßrichtung, die hier aber nicht ausführlich vorgestellt werden können. Es sei aber kurz auf den Überblicksartikel von Sanford/Sturt (2002) verwiesen, die unter anderem Studien mit sogenannten *Depth-charge*-Sätzen (Wason/Reich 1979) präsentieren. Dabei können korrekte syntaktische Analysen von widersprechendem Weltwissen „überschrieben“ werden. Die Untersuchungen von Baker/Wagner (1987) legen nahe, dass offensichtlich falsche Informationen eher nicht detektiert werden, wenn sie in Testsätzen tiefer eingebettet sind.

Sanford/Sturt (2002) führen mehrere Argumente ins Feld, warum es einen solchen oberflächlichen Verarbeitungsmodus geben könnte: „Underspecified representations are more tolerant in that they are not so likely to be disturbed by changes in interpretation forced by subsequent linguistic input.“ (ebd., S. 385). Gerade bei der Verarbeitung gesprochener Sprache könnte eine solche Toleranzschwelle essenziell für das Verstehen sein, so Sanford und Sturt weiter. Gesprochene Sprache ist oft weit entfernt von vollkommener grammatikalischer Korrektheit und muss unter Zeitdruck aufgenommen und verarbeitet werden. Ein anderes Argument wurde in der vorliegenden Arbeit bereits eingeführt:

Subordination and focus represent ways of signalling what is important in a discourse, and we have seen how they play a role in determining processing

depth of processing. [...] Signalling importance is a general issue, and a theory of discourse comprehension must address this and its relation to processing. (Sanford/Sturt 2002, S. 386)

Diese Argumentation kehrt das Argument der Lesestrategie um, das von Pynte et al. (2008a) ins Spiel gebracht wurde, und verschiebt die Perspektive von den TextrezipientInnen auf die TextproduzentInnen. Während Pynte et al. (ebd., S. 8) noch eher den LeserInnen eine Strategie zuschreiben, irrelevanten (tiefer eingebetteten) Textstellen weniger Aufmerksamkeit zukommen zu lassen, sehen Sanford/Sturt (2002) eher die TextproduzentInnen in der Rolle, weniger relevante Strukturen tiefer einzubetten.

Bezieht man diese Ausführungen auf die hier präsentierten Analysen zur abrufbasierten Satzverarbeitung, muss man ebenfalls mit einem Ausgleich einer hohen Ressourcenbelastung argumentieren: Ist die Anzahl offener syntaktischer Klammern hoch und herrscht hohe Interferenz durch zahlreiche präverbale Referenten, wechselt der menschliche Satzverarbeitungsmechanismus in einen flachen Verarbeitungsmodus, der eventuell unterspezifizierte, unvollständige oder unkorrekte Argumentanbindungen zur Folge haben kann.<sup>125</sup> Ob während der Lesevorgänge der hier präsentierten natürlichen Texte tatsächlich ein solches Phänomen aufgetreten ist, müsste gezielt über die Erhebung textferner Verstehensprodukte überprüft werden. Möglich wäre das beispielsweise über Tests oder Verständnisfragen, die gezielt auf diese Hypothese zugeschnitten sein müssten. Gefragt würde gezielt nach der Anbindung von Argumenten an die entsprechenden Verben. Die Einbettungstiefe der Verben sowie die Anzahl präverbaler Referenten würde gezielt variiert. Dabei müsste darauf geachtet werden, dass die Kapazitätsbelastung tatsächlich so hoch ist, dass VersuchsteilnehmerInnen zumindest in einigen Fällen in einen flachen Verarbeitungsmodus „gezwungen“ werden. Wenn ein solches Experiment mit der Erfassung von Blickbewegungen kombiniert würde, müsste der hier vorgestellte beschleunigende Effekt für tief eingebettete Verben mit vielen präverbalen Referenten repliziert werden. Auch der (erwartete) verlangsamende Effekt auf weniger tief eingebettete Verben sollte hervortreten.

### 7.3.6 Zusammenfassung

Zu Beginn dieses Abschnitts wurde gezeigt, dass ein Interaktionseffekt zwischen der relativen Satzposition eines Worts und der Einbettungstiefe besteht.

<sup>125</sup> Diese Formulierung legt nahe, dass es zwei Modi gibt: Normale und oberflächliche Verarbeitung. Das muss nicht unbedingt der Fall sein. Es ist auch denkbar, dass es sich um ein Kontinuum handelt.

Dieser Interaktionseffekt ist höchstwahrscheinlich auf Funktionswörter zurückzuführen. Inhaltswörter zeigen diesen Effekt nicht. Sie zeigen unabhängige Effekte der relativen Position im Satz und der Einbettungstiefe. Im Weiteren ging ich von der Hypothese aus, dass der Effekt der Einbettungstiefe für Inhaltswörter auf eine Lesestrategie zurückzuführen ist, laut derer tiefer eingebetteten Strukturen weniger Relevanz beigemessen wird. Dies führt dazu, dass sie weniger aufmerksam und damit schneller gelesen werden. Eine Aufteilung der Analysen für verschiedene Wortarten innerhalb der Inhaltswörter zeigte, dass lediglich für Nomen, Adjektive und Adverbien beschleunigende Effekte der Einbettungstiefe zu zeigen sind. Anhand einer Regioneneinteilung, die nicht auf Wörtern, sondern auf Phrasen beruht, wurde dieser Effekt weiter untermauert. Lediglich Adjektiv- und Präpositionalphrasen zeigten einen Einbettungseffekt. Für Adverbialphrasen befindet sich der Schätzer in ähnlichen Regionen, die Datenbasis ist aber zu gering. Da genau dies jene Phrasentypen sind, die potenziell unwichtige Modifikationen enthalten können, ist dies ein weiterer Hinweis auf die Relevanzhypothese. In einem weiteren Schritt wurde außerdem gezeigt, dass sich der zuvor berichtete Effekt für Nomen nur in Präpositionalphrasen zeigt. Betrachtet man Nomen, die nicht in Präpositionalphrasen stehen, ist der Einbettungs-/Relevanzeffekt nicht zu beobachten.

Im weiteren Verlauf des Abschnitts wurde gezeigt, dass Einbettungstiefe nicht nur global, sondern auch lokal wirkt – dass also nicht nur die absolute, sondern auch die relative Einbettungstiefe einen Einfluss auf die Verarbeitungszeit hat. Hierzu wurde die Anzahl der sich öffnenden Klammern als Prädiktor hinzugezogen. Zunächst wurde gezeigt, dass das Öffnen einer neuen syntaktischen Phrase Verarbeitungsaufwand bedeutet. Für Fälle, in denen eine Phrase geöffnet wird, hat außerdem die Anzahl sich öffnender Phrasen einen verlangsamenden Effekt auf die Lesezeit. Außerdem wurde gezeigt, dass dieser Effekt nicht für alle Phrasentypen gleich ist. Unter Gesichtspunkten der Vorhersagbarkeit wurde argumentiert, dass einige Phrasen – insbesondere Nominalphrasen – oft hoch vorhersagbar sind. Sie sollten also beim Öffnen weniger Verarbeitungsaufwand erfordern als andere Phrasentypen. In der Tat sind es denn auch Nominalphrasen, die als einziger Phrasentyp einen allen anderen Phrasentypen entgegengesetzten Effekt zeigen.

Mit der Analyse von Phänomenen, die von der Theorie des *cue-based parsing* vorhergesagt werden, bewegten wir uns im Laufe des Kapitels weg von linguistisch-strukturellen Prädiktoren. Es wurden Elemente gezählt, die vor einem Verb im Satz stehen und potenziell als Argument für dieses Verb in Frage kämen, ließe man den globalen Satzkontext außer Acht. Es wurde gezeigt, dass Verben in der Tat schwieriger zu verarbeiten sind, wenn mehr mögliche

Argumente vor ihnen im Satz stehen. Dies gilt insbesondere für flach eingebettete Verben; für tief eingebettete Verben scheint sich dieser Effekt umzudrehen. Es wurde daher argumentiert, dass auch Effekte des *shallow parsings* eine Rolle spielen könnten.

Wir wollen nun noch einen Schritt weitergehen und im vorletzten Ergebniskapitel einige Verarbeitungsphänomene auf der Textebene analysieren. Das FLRC bietet hierzu erstmals die Möglichkeit, da es das erste Lesekorpus mit zusammenhängenden deutschen Texten ist.

## 7.4 Analysen auf Textebene

Für die Analysen der Effekte auf der Textebene steht die manuelle Koreferenz-Annotation bereit (siehe 5.2.3), anhand derer die Einflüsse verschiedener Eigenschaften referenzieller Ausdrücke auf die Lesezeit untersucht werden können. Dabei werden wir uns auf die in Abschnitt 2.2 vorgestellten Modelle beziehen. Die Verbindung zu diesen Modellen wird zu Beginn des augenblicklichen Abschnitts wieder aufgenommen, nachdem einige einfache Verteilungsstatistiken bezüglich der referenziellen Ausdrücke im FLRC vorgestellt wurden. Die erste Variable, deren Vorhersagekraft dann untersucht wird, ist der Typ des referenziellen Ausdrucks (7.4.1). Dann wird die Referenzdistanz als Einflussfaktor untersucht (7.4.2) bevor die Prädiktoren 'Abstraktheit' bzw. 'zusätzliche Information' mit in die Untersuchung aufgenommen werden (7.4.3).

Im FLRC wurden 1.295 Anaphern annotiert. Kataphern werden aufgrund der geringen Zahl im FLRC ( $n = 46$ ) nicht ausgewertet. Tabelle 7.15 gibt einen Eindruck der Verteilung der anaphorischen Ausdrücke bezüglich ihrer Form (Proform oder volle Nominalphrase) und ob die referenzielle Beziehung eine Satzgrenze überschreitet oder nicht. Wie der assoziierte  $\chi^2$ -Test zeigt, weicht die Verteilung signifikant von einer angenommenen gleichmäßigen Verteilung ab. Die standardisierten Pearson-Residuen legen außerdem nahe, dass jede Zelle der Kreuztabelle zu diesem Effekt beiträgt.<sup>126</sup> Das Muster ist dabei nicht überraschend: Nominalphrasen werden eher dazu eingesetzt, wenn auf Diskursentitäten referiert wird, die in einem Satz vor dem aktuellen genannt wurden. Antezedenzien von Proformen befinden sich tendenziell eher im gleichen Satz wie die Proform selbst. Erwartungsgemäß gibt es hierbei innerhalb der Gruppe der Proformen ein Ungleichgewicht: Die Antezedenzien substituierender Demonstrativpronomen (POS-Tag PDS, 'dieser', 'jener')

<sup>126</sup> Die standardisierten Pearson-Residuen sind als z-Werte zu interpretieren. Sobald das Residuum einen Betrag von 2 überschreitet, kann von einer signifikanten Abweichung vom Erwartungswert in der jeweiligen Zelle ausgegangen werden.

befinden sich eher außerhalb des aktuellen Satzes.<sup>127</sup> Das Muster für die häufigeren Personalpronomen ist jedoch so deutlich, dass dieser Effekt nicht in Tabelle 7.15 zu sehen ist.

Hypothesen für die Verarbeitung von anaphorischen Ausdrücken leiten sich direkt aus Abschnitt 2.2 ab, wo einige Theorien und Studien zur Verarbeitung von Anaphern vorgestellt wurden. Zunächst soll einschränkend angemerkt werden, dass nicht alle Aspekte der Anaphernauflösung in dieser Arbeit untersucht werden können. So wird beispielsweise die Rolle des Diskursfokus, die unter anderem im bereits angesprochenen *Scenario Mapping and Focus Model* (Garrod et al. 1994; Sanford/Garrod 1981, 1994, 1998) von zentralem Interesse ist, keine explizite Beachtung finden. Ein erster Faktor, dessen Einfluss in der vorliegenden Arbeit untersucht werden soll, ist die Form des anaphorischen Ausdrucks. Gemäß der u.a. von Garrod et al. (1994, S. 41) propagierten Hierarchie sollte die Auflösung von Proformen im Allgemeinen länger dauern als jene von definiten Nominalphrasen oder wiederholten Namensnennungen. Pronomen „passen“ oft zu mehreren Antezedenzien, da sie unterspezifiziert sind und im Vergleich zu expliziten Nominalphrasen weniger semantischen Gehalt haben (vgl. z.B. Gernsbacher 1989).

	volle NP	Proform	Summe
innerhalb Satz	178 (-4,5)	126 (9,4)	304
über Satzgrenze	879 (2,5)	112 (-5,2)	991
Summe	1.057	238	1.295

Tab. 7.15: Kreuztabelle für anaphorische Ausdrücke, kreuztabelliert nach dem Typ der Referenz (volle NP vs. Proform) und ob die anaphorische Beziehung eine Satzgrenze überschreitet. In Klammern sind die standardisierten Pearson-Residuen vermerkt.

Zweiter Faktor (neben der Form der Anapher) ist die Distanz zwischen anaphorischem Ausdruck und dem entsprechenden Antezedens. Die hier gemessene räumliche Distanz zwischen Anapher und Antezedens ist nicht zu verwechseln mit der konzeptuellen Distanz, wie sie Almor (1999) beschreibt (siehe ebenfalls 2.2). In den folgenden Analysen ist die Anzahl intervenierender Wörter zwischen referenziellem Ausdruck und dem Anfang des Antezedens ausschlaggebend. Relevant ist immer die letzte Nennungsinstanz des

<sup>127</sup> Neun substituierende Demonstrativpronomen stellen die Referenz innerhalb eines Satzes her, 28 über die Satzgrenze hinweg.



Antezedens. Welche das ist, ist teilweise nicht ganz einfach zu entscheiden. So ist beispielsweise das Urteil zum unerlaubten Aufenthalt in Deutschland stark von einem Wechsel zwischen Beschreibungen des konkreten Einzelfalls und rechtlichen Erörterungen geprägt. Teilweise wird über mehrere Absätze hinweg referiert, wann und unter welchen Umständen der konkrete Beschwerdeführer (ein Syrer) in Deutschland eingereist ist. Hinzu kommen Beschreibungen der Entscheidungen bisheriger Instanzen. Solche Abschnitte sind geprägt von langen, relativ dichten Referenzketten, wo immer wieder auf den konkreten Syrer referiert wird. Solche beinahe als narrativ zu bezeichnenden Abschnitte werden unterbrochen von rechtstechnischen Ausführungen, in denen die Entscheidung des Bundesverfassungsgerichts vorbereitet und begründet wird. In diesen Abschnitten wird – um beim bereits eingeführten Beispiel zu bleiben – nicht mehr auf den syrischen Beschwerdeführer referiert. In diesen Teilen wird auf einer abstrakten Ebene argumentiert. Es wird vom konkreten Einzelfall abstrahiert. Ist in den narrativen Abschnitten noch vom Syrer oder vom Beschwerdeführer die Rede, ist in den rechtlichen Abhandlungen und Begründungen eher von 'einem Ausländer' die Rede. An den Grenzen dieser Abschnitte wurde die Koreferenz zwischen 'dem Beschwerdeführer' und 'einem Ausländer' in der Annotation aufrecht erhalten. Um das Urteil als Ganzes verstehen zu können, muss ein Leser/eine Leserin diese Diskursentitäten miteinander verbinden – versteht man nicht, dass der konkrete Beschwerdeführer die Rolle des abstrakten Ausländers einnimmt, wird man den Text nicht verstehen. Das Aufrechterhalten der Koreferenzkette in diesen Fällen ist daher meines Erachtens gerechtfertigt.

Die Distanz zwischen einem anaphorischen Ausdruck und seinem Antezedens sollte im Allgemeinen einen verlangsamenden Effekt auf die Lesezeiten haben (siehe hierzu auch 2.2). Die Grundprinzipien des menschlichen Arbeitsgedächtnisses wie Interferenz und Zerfall tragen dazu bei, dass Material, dessen Rezeption länger zurück liegt, schwieriger zugänglich ist. Wird also auf eine Diskursentität referiert, deren Nennung länger zurück liegt, sollte die Repräsentation dieser Entität schwächer aktiviert sein. Dies führt zu erschwertem Abruf, was sich wiederum in längeren Lesezeiten ausdrücken sollte. In manchen Fällen sind eventuell auch orientierende Sakkaden zurück im Text nötig, um die korrekte Anbindung der Anapher herstellen zu können.

Als abhängige Variablen werden in diesem Abschnitt wiederum die residualisierten Lesezeiten dienen. Die Einflüsse der Effekte auf lexikalischer Ebene sind ebenso kontrolliert wie jene der Bi- und Trigramm-Ebene. Somit ist die Struktur der abhängigen Variablen dieselbe wie jene im letzten Kapitel, in dem Einflüsse auf der syntaktischen Ebene herausgearbeitet wurden.



### 7.4.1 Referenztyp

Die erste Hypothese besagt, dass anaphorische Proformen länger verarbeitet werden sollten als volle Nominalphrasen, die anaphorisch eingesetzt werden. Selektiert wurden alle Wörter, die als anaphorische Ausdrücke markiert wurden. Der Datensatz umfasst somit knapp 20.000 Zeilen von 80 VersuchsteilnehmerInnen. Tabelle 7.16 gibt einen Überblick über die gemischten Modelle für die residualisierten Lesezeitmaße. Die relative Position des referenziellen Ausdrucks im Satz sowie die Einbettungstiefe sind als Kontrollvariablen aufgenommen. Die Einbettungstiefe ist wiederum residualisiert nach der relativen Position im Satz. Die Interaktion dieser beiden Faktoren wurde nicht in die Modelle aufgenommen. Es zeigt sich jedoch, dass keiner der hier berichteten Effekte verschwindet, wenn man auch die Interaktion relative Position  $\times$  residuale Einbettungstiefe aufnimmt. Es zeigt sich für jedes der vorgestellten Lesezeitmaße, dass der Referenztyp einen deutlichen Einfluss hat. Der Kontrast zwischen einer vollen Nominalphrase und einer Proform (letzte zwei Spalten in Tab. 7.16) ist deutlich ausgeprägt: Proformen werden länger gelesen (bzw. haben einen längeren Regressionspfad) als volle Nominalphrasen.

	relative Position im Satz		residuale Einbettungstiefe		Ref.typ == Proform	
	$\beta$	<i>t</i> -Wert	$\beta$	<i>t</i> -Wert	$\beta$	<i>t</i> -Wert
FFD	-0,008	-0,36	-0,004	-1,04	0,08	3,73***
FPRT	-0,004	-0,14	-0,004	-0,78	0,12	3,97***
RPD	-0,001	-0,03	-0,02	-2,86**	0,15	2,83**
TRT	-0,01	-0,33	-0,0006	-0,06	0,19	5,54***

Tab. 7.16: Ergebnisse der gemischten Modelle für die relative Position im Satz, die residuale Einbettungstiefe und den Typ der Referenz (volle NP vs. Proform). Abhängige Variablen sind residualisiert (Erklärung im Text).

Der Unterschied zwischen dem Effekt auf *first pass reading times* (FPRTs) und jenem auf *total reading times* (TRTs) deutet darauf hin, dass nach dem ersten Lesen der referenziellen Ausdrücke noch Effekte bestehen. Um dies detaillierter zu untersuchen, können die *re-reading times* (RRTs) analysiert werden. Ein weiteres Modell für alle referenziellen Ausdrücke, in das die Wortlänge, die residuale Wortfrequenz, die residuale Wortfamiliarität, die residuale Einbettungstiefe, die relative Position im Satz, die Position im Experiment sowie die

*random intercepts* für VersuchsteilnehmerInnen<sup>128</sup> und Text-IDs als Kovariaten eingehen, bestätigt die Vermutung. Proformen werden auch bei wiederholtem Lesen länger gelesen als volle Nominalphrasen ( $\beta = 0,16$ ;  $SE = 0,05$ ;  $t = 3,16$ ;  $p_{\text{MCMC}} = 0,002$ ). Bezüglich der weitergehenden Analysen gibt es mehrere Probleme. Wie in Tabelle 7.15 abgelesen werden kann, hängt die Form des referenziellen Ausdrucks stark davon ab, ob über eine Satzgrenze hinweg referiert wird oder nicht. Das heißt, die Prädiktoren ‘Referenztyp’ und ‘Referenz-über-Satzgrenze’ sind kollinear.

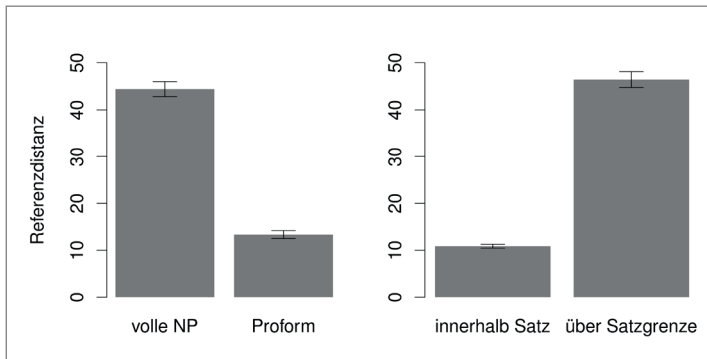


Abb. 7.14: Referenzdistanz gemessen in Wörtern zwischen Antezedens und Anapher in Abhängigkeit von Referenztyp (links) und Referenz-über-Satzgrenze (rechts). Die Fehlerbalken symbolisieren Standardfehler.

#### 7.4.2 Referenzdistanz

Ein dritter Faktor, der diesem Variablen-Konglomerat angehört, ist die Referenzdistanz. Diese wurde über die Anzahl der Wörter zwischen Anapher und Antezedens operationalisiert. Logischerweise korreliert auch dieser Prädiktor mit Referenztyp und Referenz-über-Satzgrenze. Die entsprechenden Zusammenhänge sind in Abbildung 7.14 abgetragen. Es zeigt sich deutlich, dass die durchschnittliche Referenzdistanz für Proformen deutlich niedriger ist als für volle Nominalphrasen. Da innerhalb eines Satzes meist gar nicht so viel Platz ist, um über große Distanzen zu referieren, ist die Referenzdistanz innerhalb von Sätzen ebenfalls deutlich geringer als für Referenzen, die über die Satzgrenze hinaus gehen.<sup>129</sup> Eine Möglichkeit, diese Kollinearität in den Griff zu bekommen, ist die Aufteilung des Datensatzes für die weiteren Analysen. So

<sup>128</sup> Zusätzlich geht hier auch die *random slope* in Abhängigkeit der Position im Experiment mit ein.

<sup>129</sup> Für die in Abbildung 7.14 gezeigten Unterschiede erübrigen sich statistische Tests, da die Unterschiede offensichtlich sind.

kann zunächst anhand des Referenztyps geteilt werden (volle NP vs. Proform). Für diese Datensätze kann der Einfluss des Faktors 'Referenz-über-Satzgrenze' überprüft werden. Beide Datensätze werden dann wiederum geteilt in jene Fälle, in denen die Koreferenzbeziehung über eine Satzgrenze aufgebaut wird, und in die anderen Fälle, in denen das nicht geschieht. Innerhalb dieser vier Subdatensätze kann dann der Effekt der Referenzdistanz untersucht werden. Ein solches Vorgehen ist relativ konservativ, da die Anzahl der Fälle pro Datensatz drastisch eingeschränkt wird.

Zunächst werden getrennte Analysen für volle Nominalphrasen und Proformen durchgeführt. Prädiktoren sind wiederum die relative Position im Satz sowie die residuale Einbettungstiefe. Neu hinzu kommt der Prädiktor 'Referenz-über-Satzgrenze'. Ein signifikanter Effekt der Überschreitung einer Satzgrenze zeigt sich lediglich für die residualen *first fixation durations* auf Proformen ( $\beta = -0,17$ ;  $SE = 0,08$ ;  $t = -2,135$ ;  $p_{\text{MCMC}} = 0,04$ ). Die ersten Fixationen auf anaphorischen Proformen, die eine Referenzbeziehung über die Satzgrenze aufbauen, sind signifikant kürzer. Im nächsten Schritt werden die beiden Teildatensätze für Proformen und volle Nominalphrasen erneut geteilt. Es ergeben sich vier Datensätze: 1) Volle Nominalphrasen, die eine Referenzbeziehung innerhalb desselben Satzes aufbauen, 2) Proformen, die auf ein Element im selben Satz referieren sowie 3) volle Nominalphrasen, die eine Referenz über die Satzgrenze hinweg aufbauen und 4) Proformen, die über die Satzgrenze hinweg referieren. Die Kombinationen sind auch in Tabelle 7.17 über die Zeilen- und Spaltennamen ersichtlich. Für residuale FFDs und residuale FPRTs ergeben sich in keiner der Subanalysen signifikante Effekte. Für die residuale Dauer des Regressionspfads (RPDs) ergibt sich ein marginal signifikanter Effekt für Proform-Referenzen innerhalb eines Satzes. Tendenziell gilt: Je weiter die Distanz zwischen Proform und Antezedens, desto länger der Regressionspfad ( $\beta = 0,038$ ;  $SE = 0,023$ ;  $t = 1,70$ ;  $p_{\text{MCMC}} = 0,091$ ).

	<b>volle NP</b>	<b>Proform</b>
innerhalb Satz	$\beta = 0,004$ ; $SE = 0,003$	$\beta = 0,021$ ; $SE = 0,011$
	$t = 1,55$ ; $p_{\text{MCMC}} = 0,12$	$t = 1,92$ ; $p_{\text{MCMC}} = 0,06$
über Satzgrenze	$\beta = 0,001$ ; $SE = 0,0002$	$\beta = -0,025$ ; $SE = 0,014$
	$t = 2,99$ ; $p_{\text{MCMC}} = 0,003$	$t = -1,83$ ; $p_{\text{MCMC}} = 0,11$

Tab. 7.17: Modellergebnisse der gemischten Modelle für die vier Subdatensätze, die sich aus den Kombinationen von Zeilen und Spalten ergeben. Abgetragen sind die Kennwerte für den Effekt des Prädiktors Referenzdistanz auf residuale TRTs.

Die analogen Ergebnisse für die Gesamtlesezeiten (TRTs) sind in Tabelle 7.17 zusammengefasst.<sup>130</sup> Für volle Nominalphrasen, deren Antezedens nicht im gleichen Satz ist (in Tab. 7.17 die Zelle links unten), zeigt sich ein signifikant verlangsamender Effekt der Referenzdistanz: Je weiter das Antezedens einer vollen Nominalphrase von dieser entfernt ist, desto länger wird die referierende Nominalphrase gelesen. Innerhalb von Sätzen (in Tab. 7.17 links oben) kann dieser Effekt nicht gezeigt werden. Für Proformen deutet sich ein Effektmuster an, das entgegengesetzt ist: Hier kann ein marginal verlangsamender Effekt der Referenzdistanz gezeigt werden – allerdings nur, wenn die Proform innerhalb des Satzes referiert (in Tab. 7.17 rechts oben). Ein beschleunigender Effekt der Referenzdistanz, wenn das Antezedens der Proform außerhalb des aktuellen Satzes liegt (Tab. 7.17: rechts unten), deutet sich nur numerisch an, reicht jedoch nicht an die Schwelle zur marginalen Signifikanz heran. Interessanterweise zeigt sich, dass (zumindest marginale) Effekte in jenen Zellen zu zeigen sind, die für den jeweiligen Referenztyp die meisten Fälle im Blickbewegungsdatensatz enthalten.

	volle NP	Proform
innerhalb Satz	2.442	1.703
über Satzgrenze	12.108	1.509

Tab. 7.18: Anzahl Datenpunkte im Blickbewegungsdatensatz für die vier Subdatensätze, die sich aus der Kombination von 'Referenztyp' und 'Referenz-über-Satzgrenze' ergeben.

Analog zu Tabelle 7.15 zeigt Tabelle 7.18, wie viele Fälle im Blickbewegungsdatensatz in die jeweilige Kombination fallen. Es ist darüber hinaus zu beachten, dass die real in die Modelle eingehenden Fälle in der Regel noch weniger sind, weil nicht jede/r VersuchsteilnehmerIn jedes Wort überhaupt fixiert. Auch können fehlende Werte in einer der Prädiktorvariablen (und zwar auch bei den vorgeschalteten Residualisierungen nach textueller und *n*-Gramm-Ebene) dazu führen, dass ein Datenpunkt nicht ins Modell eingeht. Die Vermutung liegt nahe, dass die Aufspaltung in vier Subdatensätze und die damit einhergehende Reduzierung der jeweiligen Fallzahl dazu führt, dass die Standardfehler (SE) der Effektschätzer ansteigen. Dies wiederum kann dazu führen, dass relativ große Effektschätzer ( $\beta$ ) nicht zu signifikanten Effekten führen.

<sup>130</sup> Obwohl in dieser Analyse keine Effekte auf *first-pass reading times* festgestellt werden können, dafür aber auf *total reading times*, können keine Effekte auf *re-reading times* nachgewiesen werden.

Jene Effekte, die bisher auf Referenzebene zu Tage gefördert wurden, sind größtenteils leicht erklärbar und hypothesenkonform: Proformen werden generell langsamer gelesen – dieser Effekt zeigt sich auf FFDs, FPRTs, RPDs und TRTs (siehe Tab. 7.16). Nach einer Teilung des Datensets in volle Nominalphrasen und Proformen kann nur für die FFDs auf Proformen ein signifikanter Effekt für die Überschreitung einer Satzgrenze gezeigt werden. Dieser geht in eine eher unerwartete Richtung: Wenn Proformen über die Satzgrenze referieren, sind die ersten Fixationen auf ihnen kürzer. Wie soeben gezeigt wurde, führt eine weitere Teilung des Datensets dazu, dass Effekte der Referenzdistanz nur auf TRTs gezeigt werden können. Diese Effekte gehen in die erwartete Richtung: Je weiter die Referenzbeziehung, desto länger die Lesezeit. Für Proformen gilt das aber nur innerhalb von Sätzen und bei vollen Nominalphrasen nur, wenn die Referenz die Satzgrenze überschreitet.

In den Modellen, die diesen Effekten zugrunde liegen, gehen – wie oben erwähnt – die relative Position im Satz und die residuale Einbettungstiefe als feste Faktoren mit ein. In den Abschnitten 7.3.1 und 7.3.5 wurde gezeigt, dass Einbettungstiefe eine Rolle als interagierender Faktor spielen kann. U.a. gilt das für den Effekt der Anzahl präverbaler Referenten, der in tieferen Einbettungstiefen nachzulassen scheint. Dort wurde als mögliche Erklärung für dieses Effektmuster ein *shallow parsing*-Prozess angeführt. Unter diesem Gesichtspunkt liegt es nahe, auch in der aktuellen Analyse einen solchen Interaktionseffekt zu vermuten. Es werden daher jene Zellen aus Tabelle 7.17 selektiert, in denen signifikante Effekte der Referenzdistanz gezeigt werden konnten. Dies sind – zur Wiederholung – volle Nominalphrasen, die über die Satzgrenze referieren und Proformen, die innerhalb ihres Satzes referieren. In die Ausgangsmodelle wird zusätzlich die Interaktion zwischen residualer Einbettungstiefe und Referenzdistanz aufgenommen. Die daraus resultierenden Modelle sind in den Tabellen 7.19 (für Nominalphrasen, die über die Satzgrenze hinweg referieren) und 7.20 (für Proformen, die innerhalb desselben Satzes referieren) abgetragen.

Es zeigt sich, dass sowohl für Nominalphrasen als auch für Proformen die Referenzdistanz noch immer einen signifikanten Einfluss hat. Dieser Effekt wird jedoch von der Einbettungstiefe der jeweiligen Wörter moduliert. Wie in den Visualisierungen dieser Interaktionen (Abb. 7.15) abzulesen ist, gibt es zwar einen relativ deutlichen Effekt der Referenzdistanz für flach eingebettete Strukturen, der aber mit zunehmender Einbettungstiefe abnimmt. Abbildung 7.15 legt nahe, dass sich der Effekt mit zunehmender Einbettung in die Gegenrichtung verkehrt – das ist aber nur numerisch der Fall.

Groups	Name	Variance	Std.Dev.			
text.id	(Intercept)	0.0017289	0.04158			
Residual		0.3222069	0.56763			
		Estimate	Std. Error	t value	$p_{\text{MCMC}}$	
(Intercept)		-0.1045972	0.0298568	-3.503	0.0044	**
rel.pos		-0.1299408	0.0447176	-2.906	0.0056	**
res.einb.tiefe		0.0296512	0.0133517	2.221	0.0262	*
ref.distanz		0.0013197	0.0003688	3.578	0.0001	***
res.einb.tiefe:ref.distanz		-0.0004502	0.0002178	-2.067	0.0368	*

Tab. 7.19: Lineares gemischtes Modell für residuale TRTs auf referierenden Nominalphrasen, deren Antezedens außerhalb desselben Satzes liegt, in Abhängigkeit von relativer Position im Satz, residualer Einbettungstiefe, Referenzdistanz und der Interaktion residuale Einbettungstiefe  $\times$  Referenzdistanz.

Groups	Name	Variance	Std.Dev.			
text.id	(Intercept)	0.033997	0.18438			
Residual		0.243870	0.49383			
		Estimate	Std. Error	t value	$p_{\text{MCMC}}$	
(Intercept)		0.102715	0.127230	0.807	0.4456	
rel.pos		-0.364460	0.174083	-2.094	0.0440	*
res.einb.tiefe		0.128462	0.044848	2.864	0.0064	**
ref.distanz		0.029141	0.011505	2.533	0.0100	*
res.einb.tiefe:ref.distanz		-0.014513	0.005426	-2.675	0.0102	*

Tab. 7.20: Lineares gemischtes Modell für residuale TRTs auf Proformen, deren Antezedens im selben Satz liegt, in Abhängigkeit von relativer Position im Satz, residualer Einbettungstiefe, Referenzdistanz und der Interaktion residuale Einbettungstiefe  $\times$  Referenzdistanz.

Die Erklärung, die nach dem vorherigen Abschnitt wiederum herangezogen werden könnte, um diesen Interaktionseffekt zu erklären, geht in eine ähnliche Richtung wie dort. Wenn man voraussetzt, dass ab einer gewissen Einbettungstiefe die ansonsten „normal“ ablaufenden Verarbeitungsprozesse in eine Art „oberflächlichen Modus“ wechseln, könnte ein solches Muster entstehen (für eine ausführlichere Diskussion der Ansätze siehe 7.3.5). Wie müsste man sich einen solchen oberflächlichen Referenzanbindungsmodus vorstellen? Es würde bedeuten, dass die TeilnehmerInnen, während sie die Texte verarbeiten, in manchen Fällen – nämlich wenn die referenziellen Ausdrücke sehr tief eingebettet sind – das Antezedens beispielsweise einer Proform teilweise unterspezifiziert lassen. Eine solche Aussage muss natürlich mit aller Vorsicht getroffen werden. Es ist ein relativ indirekter Schluss, wenn man aus

der Tatsache, dass die Referenzdistanz keinen Einfluss mehr hat, ableitet, dass die VersuchsteilnehmerInnen wohl die Referenzanbindung nicht mehr vollständig vorgenommen haben. Dies müsste in Experimenten überprüft werden, anhand derer gezielt getestet werden könnte, ob die Repräsentation der Referenzbeziehung vorhanden ist oder nicht. Ein solches experimentelles Setup ist nicht trivial zu entwerfen. Beim Test, ob die Referenz tatsächlich repräsentiert wird, muss vermieden werden, dass das Herstellen der Referenz induziert wird – eine bloße Abfrage nach dem Muster ‘Auf wen oder was wird mit diesem Pronomen referiert?’ würde daher vermutlich scheitern. Wenn man davon ausgeht, dass die Repräsentation einer Diskursentität reinstanciiert wird, wenn darauf referiert wird, sollte man eher die Aktivierung der entsprechenden Diskursentität abprüfen. Dies sollte beispielsweise über ein Priming-Paradigma möglich sein. Die Aktivierung von Diskursentitäten sollte dann – sofern die Erklärung der oberflächlichen Verarbeitung aufrecht zu erhalten ist – schwächer sein, wenn die Proform, die eben darauf referiert hat, tiefer eingebettet ist.

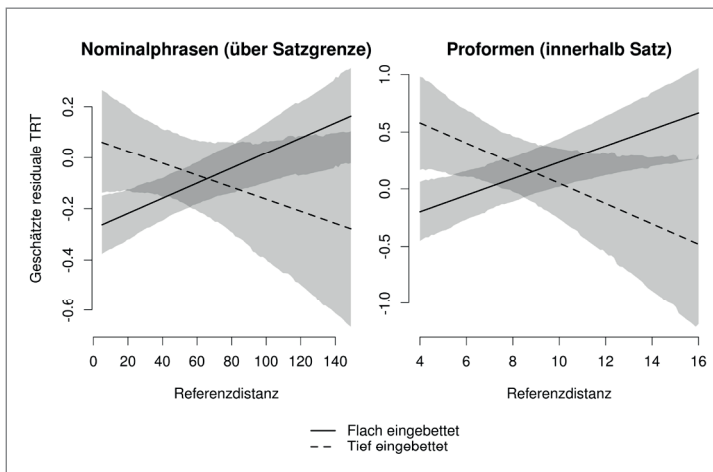


Abb. 7.15: Interaktionseffekte auf residuale TRTs von referierenden vollen NPs, deren Antezedens außerhalb ihres Satzes liegt (links) und Proformen, deren Antezedens im selben Satz liegt (rechts).

### 7.4.3 Abstraktheit und zusätzliche Informationen

Für referierende Nominalphrasen steht eine manuelle Annotation zur Verfügung, die einige der Merkmale umfasst, die in der *Informational Load Hypothesis* von Almor (1999) relevant sind (siehe hierzu Abschnitt 2.2 der vorliegenden Arbeit). Diese Eigenschaften sind 1) die Abstraktheitsbeziehung zwischen referierender Nominalphrase und dem Antezedens und 2) das Hinzufügen



von Informationen zur Diskursrepräsentation des Antezedens durch den referenziellen Ausdruck. Tabelle 7.21 zeigt die Verteilung der referierenden Nominalphrasen bezüglich dieser beiden Variablen.

	<b>abstrakter</b>	<b>gleich</b>	<b>spezifischer</b>	<b>Summe</b>
keine neue Info.	117 (-0,81)	440 (2,22)	97 (-3,04)	654
neue Info.	34 (1,83)	34 (-5,00)	61 (6,85)	129
Summe	151	474	158	783

Tab. 7.21: Anzahl referierender Nominalphrasen, kreuztabelliert danach, ob durch den referierenden Ausdruck neue Information hinzugefügt wird und ob sich der referierende Ausdruck auf einer abstrakteren, derselben oder spezifischeren Ebene befindet. In Klammern sind die standardisierten Pearson-Residuen vermerkt.

In der Mehrzahl der Fälle (83,5%) wird keine neue Information hinzugefügt. Der Standardfall (mit insgesamt 56,2% der Fälle) scheint zu sein, dass keine neue Information hinzugefügt wird und sich die referierende NP auf derselben Ebene befindet wie ihr Antezedens. Man könnte dies als die einfachste Form der Referenz sehen, weil dort einfach das Antezedens wieder aufgegriffen und auch gleich benannt wird. In 391 (88,9%) dieser Fälle ist die Oberflächenform der Referenz auch exakt gleich wie jene des Antezedens (diese Information ist nicht in Tab. 7.21 abgetragen). Eindeutig unterrepräsentiert sind entsprechend jene Fälle, in denen neue Information durch einen referierenden Ausdruck hinzugefügt wird, der sich auf derselben Taxonomie-Ebene befindet wie sein Antezedens. Zusammenfassend kann man aus der in Tabelle 7.21 dargestellten Verteilung erkennen, dass neue Information häufig dadurch eingeführt wird, dass ein Referenzausdruck auf einer taxonomisch spezifischeren Ebene gewählt wird. Das ist der Fall, wenn im Urteil zum unerlaubten Aufenthalt in der Bundesrepublik auf eine zuvor eingeführte Botschaft (im Sinne von Landesvertretung) referiert wird, durch ein Adjektiv aber klar gemacht wird, dass es sich um die syrische Botschaft handelt (Referenzbeziehung 'Botschaft' ← 'syrische Botschaft'). In dieselbe Kategorie fällt die im Urteil zum AWACS-Einsatz verwendete Referenzbeziehung 'Soldaten' ← 'deutsche Soldaten'. Taxonomisch abstrakter, aber ebenfalls neue Information hinzufügend ist die Referenzbeziehung 'die Bundeswehr' ← 'ein Parlamentsheer' in Beispiel 30. Die Bundeswehr wird als Parlamentsheer bezeichnet, was eine neue Information darstellt. Taxonomisch ist die Bezeichnung 'Parlamentsheer' über der Bundeswehr anzusiedeln, die Bundeswehr ist eine Instantiierung des Konzepts 'Parlamentsheer'.<sup>131</sup>

<sup>131</sup> Die Grenze zu einer nicht-taxonomischen Charakterisierung ist hier sehr dünn. Das einzige andere Heer bzw. die einzige andere Armee, die unter dieser Bezeichnung geführt wird bzw.

- (30) Der konstitutive Parlamentsvorbehalt hat ein hohes Gewicht, weil [die Bundeswehr]<sub>i</sub>, [ein Parlamentsheer]<sub>i</sub> ist.

Doch auch referierende Ausdrücke auf derselben taxonomischen Ebene können weitere Information zur Diskursrepräsentation hinzufügen. So wird in Beispiel 31 (aus dem Reformulierungskorpus) die Diskursrepräsentation der Eltern erst im letzten Satz mit der Eigenschaft angereichert, dass sie nicht miteinander verheiratet sind (Referenzbeziehung 'die Eltern/diese/sich' ← 'nicht miteinander verheiratete[.] Eltern'<sub>i</sub>). Eine taxonomische Unter- oder Überordnung ist hier nicht zu erkennen. Dies ist das Standardmuster von anaphorischen NPs, die zwar im Vergleich zum Antezedens nicht in ihrer Spezifität wechseln, aber trotzdem neue Information hinzufügen. In solchen Fällen wird meist das gleiche Wort wiederverwendet, aber durch ein Adjektiv modifiziert.

- (31) Paragraph 1626 BGB ist mit Artikel 6 Grundgesetz nicht vereinbar. Eine Übergangsregelung fehlt. Diese müsste eine gerichtliche Einzelfallprüfung unter zwei Bedingungen vorsehen. Erstens müssten [die Eltern]<sub>i</sub>, mit dem Kind zusammengelebt haben. Zweitens müssten [diese]<sub>i</sub>, [sich]<sub>i</sub>, vor dem Inkrafttreten des Kindschaftsrechtsreformgesetzes am 1. Juli 1998 getrennt haben. In diesem Fall könnte das Wohl des Kindes einer gemeinsamen elterlichen Sorge der [nicht miteinander verheirateten Eltern]<sub>i</sub> entgegenstehen.

Die oben eingeführten Beispiele 'syrische Botschaft' und 'deutsche Soldaten' sind auf der linguistischen Oberfläche nicht davon zu unterscheiden. Es geht darum, aus dem Kontext zu erschließen, ob eine Beziehung im Sinne einer Taxonomie oder Meronymie vorliegt. Im Falle der 'Eltern' ist das nicht der Fall, da zuvor nicht von Eltern im Allgemeinen die Rede war. Die 'nicht miteinander verheirateten Eltern' sind keine Teilmenge der zuvor genannten 'Eltern', sondern identisch mit diesen. Es wird aber die Information hinzugefügt, dass sie nicht miteinander verheiratet sind. Ähnliches gilt beispielsweise für die Referenzbeziehung 'einstweilige Anordnung' ← 'begehrte einstweilige Anordnung'. Auch hier geht es um dieselbe einstweilige Anordnung, nämlich die konkret zur Verhandlung stehende. Siehe hierzu auch die Ausführungen zu Beispiel 5 in Abschnitt 2.2. Insbesondere der Wechsel auf eine taxonomisch spezifischere Ebene wird dazu eingesetzt, eine Diskursrepräsentation mit Informationen anzureichern. Unterrepräsentiert hingegen ist der Fall, dass dies auf derselben Spezifitätsebene geschieht – in diesem Falle wird eher keine neue Information hinzugefügt.

---

wurde, ist die 'New Model Army', das Parlamentsheer während des englischen Bürgerkriegs von 1645 bis 1660.

Wenn das Antezedens wortwörtlich wiederholt wird, was von den 781 Vorkommen anaphorischer NPs in der Hälfte der Fälle geschieht (391; 49,9%), zählt dies selbstverständlich auch in diese Zelle, da bei exakter Wiederholung einer Nominalphrase weder auf eine andere Spezifitätsebene gewechselt noch neue Information hinzugefügt werden kann. Um die Einflüsse der Faktoren 'neue Information hinzugefügt' *neue.info* und 'Spezifitätsrichtung' *spez.richtung* auf die Verarbeitungsgeschwindigkeit von anaphorischen Nominalphrasen untersuchen zu können, kann man zunächst jene Fälle außen vor lassen, in denen die Anapher genau dieselbe Oberflächenform hat wie ihr Antezedens. Es werden zunächst alle anaphorischen Nominalphrasen selektiert, die in ihrer Oberflächenform von ihrem Antezedens abweichen. Das sind 392 Formen, die die Hälfte (50,1%) aller anaphorischen Nominalphrasen ausmachen. Wir werden uns zunächst auf die *first-pass reading times* auf anaphorischen Nominalphrasen konzentrieren. Die Einflüsse der Referenzdistanz und der Einfluss der Referenz über die Satzgrenze wurden bereits in den vorherigen Analysen beachtet, ebenso – im Abschnitt zuvor – die relative Position im Satz und die Einbettungstiefe. Daher wird die abhängige Variable von den Einflüssen dieser Faktoren bereinigt.

Groups	Name	Variance	Std.Dev.			
text.id	(Intercept)	0.00059981	0.024491			
Residual		0.16002351	0.400029			
		<b>Estimate</b>	<b>Std. Error</b>	<b>t value</b>	<b>p<sub>MCMC</sub></b>	
(Intercept)		-0.09580	0.03227	-2.969	0.0042	**
neue.info T		0.17176	0.04985	3.446	0.0002	***
spez.richtung abst		0.08709	0.03723	2.340	0.0180	*
spez.richtung spez		0.09916	0.03863	2.567	0.0092	**
neue.info T:spez.richtung abst		-0.15670	0.07623	-2.056	0.0444	*
neue.info T:spez.richtung spez		-0.24197	0.06888	-3.513	0.0006	***

Tab. 7.22: Lineares gemischtes Modell für referierende Nominalphrasen, die nicht denselben Wortlaut haben wie ihr Antezedens. Als feste Faktoren gehen die Spezifitätsrichtung (*spez.richtung*) und das Hinzufügen neuer Information (*neue.info*) sowie deren Interaktion in das Modell ein.

Tabelle 7.22 zeigt die Ergebnisse dieses Modells. Abbildung 7.16 vereinfacht die Interpretation des Interaktionseffekts. Betrachten wir zunächst die Fälle, in denen keine neue Information hinzugefügt wird. Hier zeigt sich, dass jene Nominalphrasen, die auf der gleichen Abstraktheitsebene wie ihr jeweiliges Antezedens sind, am schnellsten gelesen werden. Die beiden Kontraste zu den spezifischeren und abstrakteren Nominalphrasen sind deutlich signifikant. Ein Unterschied zwischen diesen spezifischeren und abstrakteren Nominal-

phrasen kann nicht festgestellt werden. Interessanterweise kehrt sich diese Abfolge für referierende Nominalphrasen, über die neue Informationen hinzugefügt werden, um. Hier werden nun Wörter in referierenden Nominalphrasen, die auf derselben Spezifitätsebene sind, am langsamsten gelesen. Der Unterschied zu abstrakteren Nominalphrasen ist nicht signifikant, jener zu spezifischeren deutlich. Der in Abbildung 7.16 angedeutete Effekt für spezifischere Nominalphrasen (Datensymbol: Dreieck, Linienart: gepunktet) im Übergang von 'keine neue Information' zu 'neue Information' ist nicht signifikant.<sup>132</sup> Es lässt sich zusammenfassen: Nur für jene Nominalphrasen, die sich auf derselben Spezifitätsebene befinden wie ihre Antezedenzen, kann ein Unterschied dahingehend festgestellt werden, ob sie neue Information zur Diskursrepräsentation ihres Antezedens hinzufügen oder nicht. Während diese Nominalphrasen am schnellsten gelesen werden, wenn keine Information hinzugefügt wird, werden sie am langsamsten gelesen, wenn dies geschieht.

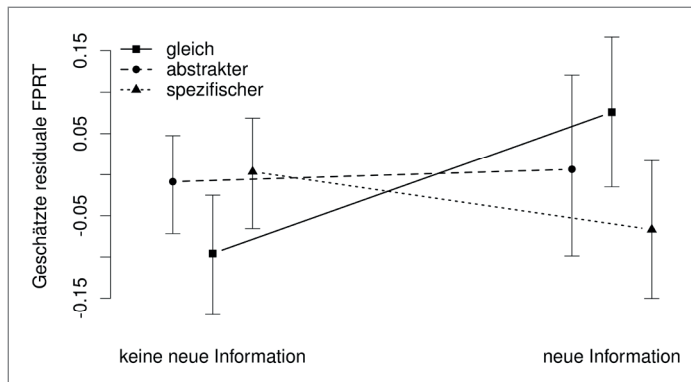


Abb. 7.16: Interaktionseffekt Spezifitätsrichtung × Hinzufügen neuer Informationen auf residuale FPRTs. Es sind nur Fälle beachtet, in denen die anaphorische Nominalphrase einen anderen Wortlaut hat wie ihr Antezedens.

Wie in Tabelle 7.21 bereits angedeutet, ist der Standardfall einer referierenden Nominalphrase, wenn sie sich auf derselben Spezifitätsebene befindet wie das Antezedens und keine Information hinzufügt. Interessanterweise ist das – wie eben gezeigt wurde – auch jene Ausprägung, die am schnellsten verarbeitet wird. Dies kann nun einfach daran liegen, dass diese Art der Referenz in der Tat kognitiv am unaufwändigsten ist: Es muss weder die Ebene der Repräsentation gewechselt werden, noch muss eine Diskursrepräsentation mit

<sup>132</sup> In einem separaten Modell für diesen Kontrast gilt:  $t = -1,57$ ;  $p_{\text{MCMC}} = 0,15$ .

neuen Informationen angereichert werden. Sollten diese beiden Faktoren unabhängig voneinander zu kognitivem Aufwand beitragen, sollten NPs auf der gleichen Spezifitätsebene auch dann am einfachsten sein, wenn neue Information hinzugefügt wird. Dies ist aber nicht der Fall. Im Gegenteil: Sie sind die am längsten gelesenen referierenden Nominalphrasen, die neue Information bereitstellen. Dies kann 1) durch eine Erwartungsverletzung und 2) durch eine andere Art der Informationshinzufügung in diesem Fall erklärt werden. Die in 1) angesprochene Erwartungsverletzung würde sich dadurch ergeben, dass LeserInnen so oft mit dem Standardfall „gleiche Spezifitätsebene/keine neue Information“ konfrontiert werden, dass der unübliche Verwendungskontext zu einer verlangsamten Verarbeitung führt. Möglichkeit 2), die von Interpretation 1) nicht ausgeschlossen wird, besagt, dass die Verwendung eines Adjektivs für den beobachteten Effekt verantwortlich ist. Ein zusätzliches Adjektiv sollte im Falle der gleichen Spezifitätsebene häufiger sein, weil die Information weder über das Einordnen des Antezedens in eine Oberkategorie noch über die Zuordnung des Antezedens in eine bestimmte Unterklasse hinzugefügt werden kann.

Es zeigt sich aber auch, dass es nicht grundsätzlich mit mehr Verarbeitungsaufwand verbunden zu sein scheint, wenn durch einen anaphorischen Ausdruck neue Information zu einer Diskursrepräsentation hinzugefügt wird. Sowohl für abstraktere als auch für spezifischere anaphorische Nominalphrasen konnte kein Effekt auf die *first-pass reading times* festgestellt werden. Deutlich ist der Effekt für anaphorische Nominalphrasen auf gleicher Abstraktionsebene: Die FPRTs steigen hier deutlich für jene NPs, die neue Informationen hinzufügen. Für *total reading times* ist der Effekt für NPs auf der gleichen Spezifitätsebene nur noch numerisch zu beobachten (siehe Abb. 7.17, Linienart: durchgezogen, Datensymbol: Quadrat). Der für FPRTs nicht einmal marginal signifikante Effekt für NPs auf einer spezifischeren Ebene (gepunktet, Dreieck) hingegen ist für *total reading times* deutlich ausgeprägt ( $t = -2,79$ ;  $p_{\text{MCMC}} = 0,005$ ). Dieser Effekt ist nicht nur durch die *re-reading times* zu erklären.

Erinnern wir uns an die in Abschnitt 2.2 eingeführten Theorien zu Anaphernauflösung. Insbesondere in der *Accessibility Theory* von Ariel (2000) und der *Informational Load Hypothesis* von Almor (1999) taucht immer wieder ein Prinzip auf, nach dem der Einsatz eines bestimmten anaphorischen Ausdrucks gerechtfertigt sein muss. Ist der Einsatz nicht gerechtfertigt, so beispielsweise für anaphorische Ausdrücke mit hohem *informational load*, bei denen keine Information hinzugefügt werden soll und das Antezedens leicht zugänglich ist, ist mit steigenden Verarbeitungszeiten zu rechnen. Almor (1999) formuliert dieses Rechtfertigungsprinzip (*justification*) explizit für den eben genannten Fall:

Anaphors with high informational load with respect to a particular antecedent, which do not add new information about their antecedents [...], are easier to process when their antecedent is not focused than when it is. Only in the former case does the high informational load have a justification in aiding the identification of the antecedent. (ebd., S. 753)

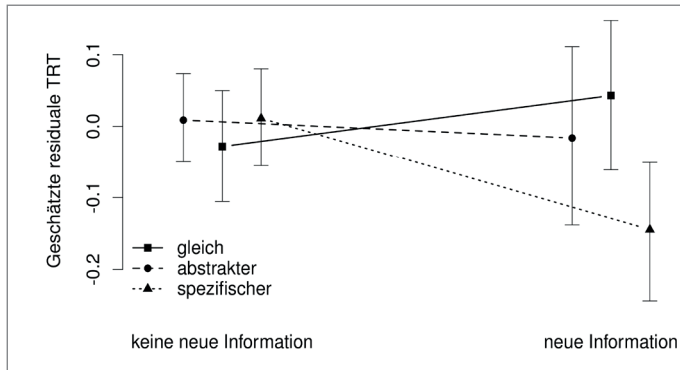


Abb. 7.17: Interaktionseffekt Spezifitätsrichtung  $\times$  Hinzufügen neuer Informationen auf residuale TRTs. Es sind nur Fälle beachtet, wo die anaphorische Nominalphrase einen anderen Wortlaut hat als ihr Antezedens.

Die hier vorgestellten Analysen unterstützen ein solches Rechtfertigungsprinzip. Ein bestimmter Diskursreferent befindet sich immer auf einer bestimmten Spezifitätsebene. Will man nun später im Text auf diesen Referenten referieren und benutzt dabei eine volle Nominalphrase, ist die Standardannahme, dass man auf derselben Spezifitätsebene bleibt. Tut man dies nicht, muss der Einsatz einer abstrakteren oder spezifischeren NP gerechtfertigt sein – beispielsweise durch das Hinzufügen neuer Information. Ist der Einsatz nicht gerechtfertigt, kann dies zu Verarbeitungsschwierigkeiten führen. Wenn allerdings der Einsatz tatsächlich gefordert ist, weil neue Information hinzugefügt werden soll, dann ist 1) der Einsatz der Standardvariante (gleiche Ebene) defizitär (siehe steigende Lesezeiten in Abb. 7.16 im Übergang von ‘keine neue Information’ zu ‘neue Information’ und 2) der Einsatz der spezifischeren Variante profitabel (siehe fallende Lesezeiten in Abb. 7.17 im Übergang von ‘keine neue Information’ zu ‘neue Information’). Eine (hier) nicht durch Korpusstudien unterstützte Beobachtung aus dem Zeitungsjournalismus passt zu diesem Muster: In Zeitungstexten finden wir oft anaphorische Nominalphrasen, die im Vergleich zum Antezedens auf eine andere Spezifitätsebene wechseln (beispielsweise ‘Merkel’  $\leftarrow$  ‘Kanzlerin’  $\leftarrow$  ‘CDU-Chefin’).<sup>133</sup> Dort

<sup>133</sup> Dieses Beispiel ist dem dritten Absatz des Artikels entnommen, der auf [www.spiegel.de/politik/deutschland/konjunkturkrise-cdu-vorstand-stuetzt-merkel-im-steuerstreit-a-593611.html](http://www.spiegel.de/politik/deutschland/konjunkturkrise-cdu-vorstand-stuetzt-merkel-im-steuerstreit-a-593611.html) abrufbar ist (Stand: 16.3.2015).

hat dies – neben der Vermeidung von wörtlichen Wiederholungen – häufig die Funktion, zusätzliche Informationen über den augenblicklichen Diskursreferenten einzuführen. Auch von Polenz (1998, S. 143) beschreibt dieses Vorgehen:

So wird in Nachrichtentexten und Kommentaren oft beim Wiederbezug auf die gleiche Person oder Institution der Bezugsausdruck variiert, um Hintergrund- oder Nebenbei-Informationen zu geben, um die Vielfalt von Rollen und Aspekten anzudeuten oder um Stellungnahmen, Einschätzungen und Bewertungen nebenbei anzubringen. [Hervorhebung entfernt, S.W.]

Das hier herausgearbeitete Verarbeitungsmuster mag gut zu einem Prinzip der Rechtfertigung passen. Diese Idee hat allerdings eine Schwäche: Das Prinzip scheint nur für spezifischere Nominalphrasen zu gelten. Für abstraktere Nominalphrasen ist in keinem der Lesezeitmaße zu beobachten, dass neue Information hinzufügende NPs schneller gelesen werden als NPs, die das nicht tun. Das passt auf eine gewisse Weise zu den linguistischen Daten (siehe Tab. 7.21), die nahelegen, dass die Verteilung der referierenden Nominalphrasen, die abstrakter als ihre Anzedenzien sind, nicht eindeutig in Richtung 'neue Information' tendieren wie dies spezifischere NPs tun. Eine Korpusstudie, die ein solches Muster untermauern könnte, müsste allerdings mit mehr Sprachmaterial wiederholt werden.

Nachdem wir uns im Verlauf dieses Kapitels von der lexikalischen über die Bi- und Trigramm-Ebene zur syntaktischen und textuellen Ebene „vorgearbeitet“ haben, werden wir nun ausgewählte Konsequenzen der Reformulierungen analysieren. Die Auswahl der Gegenstandsbereiche Nominalisierungen und komplexe Nominalphrasen ist dabei sehr nah an den Reformulierungsdimensionen des FRLCRef-Teils. Außerdem ist zu zeigen, wie sich die Reformulierung der syntaktisch komplexen Texte auf die Auflösung von referenziellen Bezügen auswirkt. Wo dies sinnvoll erscheint, wird die Analyse auf die jeweils anderen Komplexitätstypen beziehungsweise das komplette *Freiburg Legalese Reading Corpus* ausgeweitet. Die Analysen bezüglich der Reformulierungen werden dann nicht mehr am Ende dieses Kapitels zusammengefasst, sondern direkt am Anfang des nächsten und letzten Kapitels dieser Dissertation (8.1).



## 7.5 Auswirkungen der Reformulierungen auf die Rezeption

In Abschnitt 6.2.1 wurde bereits gezeigt, dass für die Lesedauer kein einfacher Effekt der Reformulierungsversion gezeigt werden kann. Es scheint also nicht so einfach zu sein, dass reformulierte Texte immer schneller gelesen werden als die Originaltexte.

Ein anderes interessantes Maß ist die Anzahl an Regressionen pro Text, also wie oft jede Leserin/jeder Leser während des Lesens im Text zurück sprang. Auch für dieses Maß ist der Zusammenhang mit der Reformulierungsversion nicht trivial. In einem gemischten Modell (nicht abgedruckt) zeigt sich zwar, dass die Anzahl von Regressionen sowohl für moderat reformulierte Texte als auch für stark reformulierte Texte **höher** ist als für die Originaltexte – sobald jedoch der Komplexitätstyp (viele Nominalisierungen vs. komplexe Nominalphrasen vs. komplexe Syntax) mit in dieses Modell aufgenommen wird, zeigt sich bereits, dass dies insbesondere auf die Textlänge zurückzuführen ist. In den längsten Texten (Komplexitätstyp *Synt.*) treten die meisten Fixationen auf, am zweithäufigsten in den mittellangen Texten (Komplexitätstyp *NP*). Alle Kontraste zwischen Komplexitätstypen sind dabei signifikant (alle  $p_{\text{MCMC}} = 0,0001$ ). Da die Textlänge hier offenbar einen sehr großen Einfluss hat, nehmen wir diese mit in das Modell auf. Da sich die Reformulierungen für die unterschiedlichen Komplexitätstypen unterschiedlich stark auswirken könnten, wird außerdem die Interaktion zwischen Reformulierungsversion und Komplexitätstyp mit ins Modell aufgenommen. Das resultierende Modell ist in Tabelle 7.23 einzusehen.

Außer den *random intercepts* für VersuchsteilnehmerInnen (*vtn*) wurde eine *random slope* in Abhängigkeit des Versuchsdurchgangs (*trial*) aufgenommen. Die Aufnahme als *fixed effect* ist gerechtfertigt: Je später im Experiment, desto weniger regressive Sakkaden treten auf. Ebenso – und das ist mit Abstand der stärkste Effekt – zeigt sich, dass die Aufnahme der Textlänge berechtigt war. Dieser Effekt erklärt so viel Varianz, dass die im vorherigen Modell noch höchst signifikanten Kontraste der Komplexitätstypen nicht mehr signifikant hervortreten. Um einen besseren Eindruck der Interaktion zu erhalten, sind die Schätzwerte für die Interaktion *typ* × *version* in Abbildung 7.18 abgetragen.

Insbesondere die stark reformulierten syntaktisch komplexen Texte (Dreieck-Symbol, ganz rechts) fallen aus dem Rahmen. Dieses Muster fand sich bereits in Abschnitt 6.2.1, wo diese Texte die am längsten gelesenen waren. Das Teilen von Sätzen scheint also auch hinsichtlich der Anzahl an Regressionen nicht dazu zu führen, dass der Leseprozess effizienter vonstatten geht – eher

das Gegenteil ist der Fall. Bei den beiden anderen Komplexitätstypen ist kein Effekt der Reformulierungsversion festzustellen. Auch in der stark reformulierten Version gruppieren *Nom.*- und *NP*-Texte noch zusammen.

Groups	Name	Variance	Std.Dev.			
vtn	(Intercept)	1.2994	1.1399			
vtn	trial	0.0000	0.0000			
Residual		5.0724	2.2522			
		Estimate	Std. Error	t value	$p_{MCMC}$	
(Intercept)		0.5027876	0.3160490	1.591	0.1106	
typ NP		-0.4700868	0.2789355	-1.685	0.1018	
typ Synt.		-0.1556346	0.2931334	-0.531	0.5908	
version moderat		0.2745294	0.2755284	0.996	0.3230	
version stark		-0.0718272	0.2754600	-0.261	0.7972	
text.len		0.0228710	0.0008108	28.208	0.0001	***
trial		-0.0357453	0.0075655	-4.725	0.0001	***
typ NP:version moderat		0.4977581	0.3898355	1.277	0.2024	
typ Synt.:version moderat		0.3136039	0.3919066	0.800	0.4242	
typ NP:version stark		0.0845730	0.3925478	0.215	0.8412	
typ Synt.:version stark		0.7819087	0.3899840	2.005	0.0438	*

Tab. 7.23: Gemischtes Modell für die Anzahl an Regressionen pro Text. Enthalten sind die Prädiktoren Komplexitätstyp (*typ*), Reformulierungsversion (*version*) sowie Textlänge (*text.len*) und Position des Texts im Experiment (*trial*). Siehe Text für Zufallseffektstruktur.

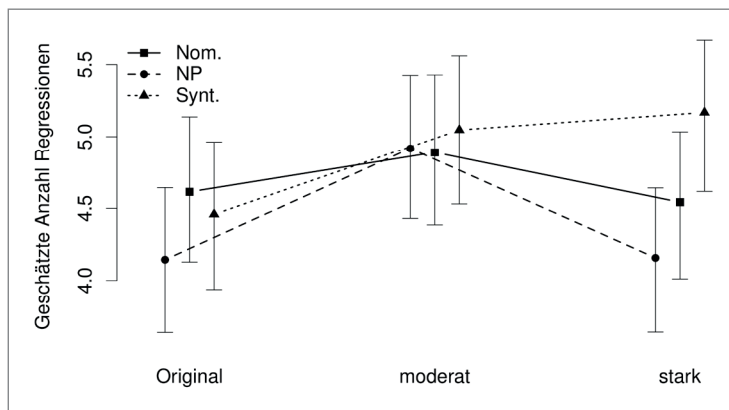


Abb. 7.18: Interaktion zwischen Reformulierungsversion und Komplexitätstyp bezüglich der Anzahl der Regressionen pro Text. Abgetragen sind Schätzwerte anhand des gemischten Modells sowie 95%-Konfidenzintervalle (ermittelt auf Basis einer MCMC-Simulation).

### 7.5.1 Nominalisierungen

Wir wollen nun gezielt auf die verschiedenen Komplexitätsquellen fokussieren, anhand derer die Texte im Reformulierungskorpus ausgewählt wurden. Zunächst soll der Einfluss von Nominalisierungen auf den Rezeptionsprozess untersucht werden. Nominalisierungen bzw. Nominalstil (vgl. Oksaar 1988, S. 193f.) wurde(n) als Merkmal von Rechtssprache identifiziert (vgl. Hansen-Schirra/Neumann 2004). Auch andere Fachsprachen, wie beispielsweise die deutsche Wissenschaftssprache (vgl. Müller-Küppers/Zöllner 2007, S. 12) sind bekannt dafür, dass viele Nominalisierungen verwendet werden. Auer/Baßler (2007) sehen im „moderne[n] westliche[n] Wissenschaftsstil“ eine „starke Tendenz zur Nominalisierung“ (ebd., S. 16). Insofern betrifft die Analyse der Verarbeitungsprozesse solcher Nominalisierungen mehrere Fachsprachen. Wir wollen mit einer ersten sehr einfachen Analyse einsteigen. Wir selektieren dabei alle Nomen aus allen Texten des kompletten FLRC. Der Blickbewegungsdatensatz umfasst dann  $n = 70.742$  Fälle. Innerhalb dieser Teilmenge des FLRC werden alle \*ung-Nominalisierungen identifiziert. Dies geschieht zunächst über eine Abfrage anhand eines regulären Ausdrucks.<sup>134</sup> Die erwartete Effektrichtung ist relativ klar: Nominalisierungen sollten auch für geübte LeserInnen von Wissenschaftssprache im Vergleich zu anderen Nomen Schwierigkeiten bereiten. Es sollten sich also für *first-pass* und *total reading times* Effekte für den Kontrast ‘Nominalisierungen’ vs. ‘alle anderen Nomen’ zeigen. Um sicher zu gehen, dass keine Einflüsse der lexikalischen, Bi- und Trigramm- sowie syntaktischen Ebene in der Analyse interferieren, werden als abhängige Variablen die nach diesen Ebenen residualisierten Lesezeitmaße verwendet.

	$\beta_{\text{nomin}}$	Std.fehler	t-Wert	$p_{\text{MCMC}}$
residuale FPRT	0,063	0,011	5,243	0,0001
residuale TRT	0,131	0,015	8,911	0,0001

Tab. 7.24: Zusammenfassung der Modellergebnisse für den Einfluss von Nominalisierungen auf residuale FPRTs und TRTs.

Tabelle 7.24 fasst die Ergebnisse dieser Analysen zusammen. Es zeigt sich sehr deutlich, dass Nominalisierungen sowohl beim ersten Lesen als auch insgesamt länger gelesen werden als alle anderen Nomen, die in dieser Analyse die Vergleichsgruppe sind. Eine weitere Analyse, in der lediglich Darstellungs-

<sup>134</sup> Dieser lautet `^ [[:upper:]] [[:alpha:]] *ung$`.

und lexikalische Faktoren<sup>135</sup> als Baseline-Faktoren beachtet wurden, zeigt die Größe dieses Effekts. Dafür wurden rohe logarithmierte Lesezeiten vorhergesagt und die Schätzer des Modells zurücktransformiert. Der Effekt für *first-pass reading times* ist mit lediglich 5,7 ms ziemlich klein (aber trotzdem signifikant). Für *total reading times* ist der Effekt mit 31,1 ms erwartungsgemäß etwas größer. Ein gemischtes logistisches Regressionsmodell für die Übersprungswahrscheinlichkeit zeigt außerdem, dass Nominalisierungen signifikant seltener übersprungen werden als normale Nomen ( $z = -7,71$ ;  $p < 0,0001$ ). Nach Bereinigung von Wortlänge, Wortfrequenz, Wortfamiliarität sowie allen Darstellungsfaktoren wird die Übersprungswahrscheinlichkeit für normale Nomen auf 28,3% geschätzt.<sup>136</sup> Für Nominalisierungen sinkt der Schätzwert für die Übersprungswahrscheinlichkeit auf 21,1%.

Für die *total reading times* wollen wir nun eine Einschätzung treffen, wie groß dieser Effekt im Vergleich zu anderen Unterschieden zwischen bestimmten Wortklassen ist. Wir kodieren daher folgende Informationen im Gesamtdatensatz: 1) Ist das Wort ein Inhaltswort oder nicht? 2) Ist das Wort ein Funktionswort oder nicht? 3) Ist das Wort ein finites Verb oder nicht? 4) Ist das Wort ein Nomen oder nicht? 5) Ist das Wort eine Nominalisierung oder nicht? All diese Informationen gehen in ein gemischtes Modell ein. Dabei sind einige Klassen ineinander verschachtelt: Alle Nomen sind auch Inhaltswörter und alle Nominalisierungen auch Nomen (und damit auch Inhaltswörter).

Abbildung 7.19 zeigt für jeden dieser Faktoren den jeweiligen Effektschätzer aus einem gemischten Modell, das ebenfalls die logarithmierten *total reading times* unter Hinzunahme der lexikalischen Faktoren und Darstellungsfaktoren (siehe Anm. 32) vorhersagt. Jeder dieser Schätzer erfasst den Unterschied zwischen der Gruppe der Wörter, die der jeweiligen Wortklasse  $x$  angehören und der Gruppe, die dieser Klasse nicht angehört, also allen anderen Wörtern. Der Punkt ganz rechts (assoziiertes  $t = 5,21$ ) zeigt, dass Nominalisierungen länger gelesen werden als alle anderen Wörter. Das ist im Lichte der vorhergehenden Analyse keine Überraschung. Die Mehrinformation, die in Abbildung 7.19 liegt, ist folgende: Dieser Effekt der Nominalisierungen ist deutlicher in die positive Richtung ausgeprägt als jener für finite Verben. Das ist umso aussagekräftiger, als der Effekt für alle Nomen, Nominalisierungen mit einbezogen, eigentlich in die entgegengesetzte Richtung zeigt. Obwohl Nomen also generell schneller gelesen werden als alle anderen Wörter, werden Nominalisierungen als Teilmenge von Nomen deutlich langsamer gelesen.

<sup>135</sup> Als Darstellungsfaktoren ging die Information mit ein, ob das Wort das erste oder letzte Wort einer Bildschirmzeile war sowie die Zeile auf dem Bildschirm. Als lexikalische Faktoren gingen die Wortlänge, -frequenz und -familiarität mit ein.

<sup>136</sup> Zum Vergleich: Die Übersprungswahrscheinlichkeit für Artikel liegt bei 60,1%.

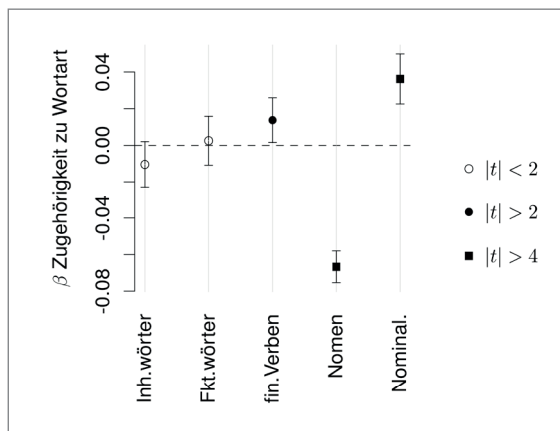


Abb. 7.19: Effektschätzer und deren Konfidenzintervalle für Zugehörigkeit zu einer bestimmten Wortklasse auf *total reading times*. Konfidenzintervalle ergeben sich aus 1,96 Standardfehlern.

Wir konnten bisher zeigen, dass \*ung-Nominalisierungen länger gelesen werden als andere Nomen (wobei Einflüsse bis zur und einschließlich der syntaktischen Ebene kontrolliert sind) und signifikant seltener übersprungen werden. Außerdem konnte gezeigt werden, dass sich die Effekte auf die *total reading time* in höheren Regionen bewegen als jene für finite Verben, die im Vergleich zu allen anderen Wörtern ebenfalls länger gelesen werden.

Das Ziel von Hansen et al. (2006), Nominalisierungen in verbale Strukturen umzuwandeln, hatte eindeutige Auswirkungen auf der linguistischen Ebene: Der Anteil verbaler Strukturen stieg im Übergang von Originalfassung zu moderat reformulierten Texten an (siehe 5.2.1). Es wäre unglücklich für die Reformulierungen, wenn die Verarbeitung von Verben in reformulierten Texten schwieriger würde – denn dann würde man auf der einen Seite schwierige Elemente (Nominalisierungen) entfernen und auf der anderen Seite für Schwierigkeiten bei der Verarbeitung der „Ersatzelemente“ sorgen. Verbale Strukturen sollten also in den reformulierten Texten nicht langsamer verarbeitet werden. Nomen sollten schneller verarbeitet werden. Zur Analyse selektieren wir alle Texte (Originale, moderate und starke Reformulierungen) des Komplexitätstyps *Nom.* aus dem FLRC. Da es oft keinen Unterschied zwischen moderater und starker Reformulierungsstufe gibt (siehe hierzu 5.1.1), kodieren wir für jedes Wort, ob es in einem reformulierten Text vorkommt oder nicht. Nun selektieren wir alle interessanten verbalen Elemente: finite Verben (POS-Tags  $V[A/M/V]$  FIN), Infinitivformen ( $V[A/V]$  INF,  $VVIZU$ ) sowie Partizipformen ( $VVPPP$ ). In das gemischte Modell für logarithmierte FPRTs

gehen lexikalische und Darstellungsprädiktoren mit ein. Die Zufallseffektstruktur besteht aus *random intercepts* für Text-IDs und VersuchsteilnehmerInnen (für diese wird außerdem die *random slope* in Abhängigkeit der Position im Versuchsdurchlauf aufgenommen). Als experimenteller Prädiktor geht der Faktor 'reformuliert oder nicht' mit in das Modell ein. Ein zweites Modell besteht aus exakt derselben Prädiktorstruktur. Es werden dabei die FPRTs auf Nomen (POS-Tag NN) geschätzt. Tabelle 7.25 zeigt alle *fixed effects* für die beiden Modelle.

	verbale Strukturen			Nomen		
	$\beta$	<i>t</i> -Wert	$p_{MCMC}$	$\beta$	<i>t</i> -Wert	$p_{MCMC}$
(Intercept)	5,02	56,5	0,0001***	-5,26	66,8	0,0001***
Länge	0,05	8,39	0,0001***	-0,04	10,7	0,0001***
Res.Freq.	-0,07	-2,95	0,0046**	-0,004	-0,24	0,81
Res.Famil.	0,002	0,07	0,95	-0,17	6,00	0,0001***
Letzt.Wort	0,09	2,38	0,02*	-0,05	1,29	0,19
Erst.Wort	0,08	1,28	0,19	-0,46	4,83	0,0001***
Zeile	0,01	0,61	0,56	-0,04	-3,22	0,0008***
Reform.	-0,01	-0,20	0,80	-0,08	-2,05	0,03*
Trial	-0,003	-1,28	0,20	-0,001	-0,74	0,45

Tab. 7.25: Zusammenfassung der Modellergebnisse für FPRTs auf verbalen Strukturen und Nomen in Texten mit vielen Nominalisierungen.

Interessant ist im aktuellen Kontext Zeile *Reform.*, in der der Schätzer dafür abgetragen ist, ob es sich um einen reformulierten Text handelt oder nicht. Negative Schätzer ( $\beta$ ) und *t*-Werte deuten dabei auf einen beschleunigenden Effekt der Reformulierungen auf die FPRTs hin. Es kann kein signifikanter Effekt für Verben nachgewiesen werden. Nomen in reformulierten Texten werden hingegen signifikant schneller gelesen als in den Originaltexten. Die Hypothese, dass die Reformulierungen einen positiven Effekt auf die Verarbeitung von Nomen und keinen negativen Effekt auf die Verarbeitung von verbalen Strukturen hatten, kann somit anhand dieser Analyse untermauert werden.

Es gibt zwei Dinge zu bemerken. 1) Die Fallzahl in Modell 2 (Nomen) ist höher (1.501 vs. 769 für Verben). Es ist also rein aufgrund der Stichprobengröße wahrscheinlicher, einen Effekt zu finden. Allerdings ist es nicht so, dass der Effekt für Verben knapp an der Signifikanzschwelle liegt – die Irrtumswahr-

scheinlichkeit liegt hier bei 80%. Es ist also nicht ohne Weiteres davon auszugehen, dass doppelt so viele Fälle hier einen signifikanten Effekt zu Tage fördern würden. 2) Der anhand der oben genannten Hypothese erwartete Effekt sollte sich eigentlich in einer signifikanten Interaktion Reformulierung  $\times$  Wortart niederschlagen. Diese Interaktion ist im entsprechenden Modell (nicht abgedruckt) aber nicht signifikant ( $t = -1,34$ ;  $p = 0,19$ ).

Eine einfache letzte Analyse soll zeigen, dass die Auslagerung der Komplexität heraus aus den Nominalisierungen nicht nur keine Nachteile für die verbalen Elemente bringt, sondern auch nicht für das Lesen des Textes als Ganzes. Die reformulierten Texte des Komplexitätstyps *Nom.* sind im Schnitt um etwa fünf Zeichen kürzer als die Originaltexte,<sup>137</sup> von daher ist ohnehin bereits zu erwarten, dass sie etwas schneller gelesen werden. Das ist auch der Fall. Durchschnittlich werden moderat reformulierte Nominalisierungs-Texte 1002 ms schneller gelesen als die Originale, für starke Reformulierungen beträgt dieser Unterschied 1471 ms. Auch wenn man für die Zeichenanzahl kontrolliert, ergibt sich ein Effekt der Reformulierungen. Moderat reformulierte Texte werden marginal schneller gelesen ( $\beta = -930$ ;  $SE = 415$ ;  $t = -2,24$ ;  $p_{\text{MCMC}} = 0,06$ ), bei stark reformulierten Texten ist der Effekt signifikant ( $\beta = -1154$ ;  $SE = 416$ ;  $t = -2,78$ ;  $p_{\text{MCMC}} = 0,02$ ).<sup>138</sup> Das heißt zusammengefasst: Die Reformulierungen der Nominalisierungen führten dazu, dass die Nomen schneller gelesen wurden, Verben nicht länger und die Texte als Ganzes ebenfalls schneller.

An dieser Stelle wäre der analytische Brückenschlag zurück ins Textmaterial interessant, denn es liegt eine explizite Hypothese vor, welche Strukturen besonders lang gelesen werden sollten. Ein solcher Rückbezug auf das linguistische Ausgangsmaterial ist nicht unmittelbar über die rohen Lesezeiten möglich, da diese – wie durch die statistischen Modelle gezeigt – durch eine Vielzahl an Faktoren beeinflusst werden. Anhand der Modelle kann aber für jeden Lesedurchgang jedes Wortes eine Lesezeit geschätzt werden. Diese Schätzung ergibt sich direkt aus den Modellen selbst und beachtet neben den festen Faktoren auch die Zufallseffektstruktur. Es wird im Folgenden ein Vorgehen vorgeschlagen, wie Wörter identifiziert werden können, die besonders schwierig sind (hier: eine besonders lange geschätzte *first-pass reading time* haben). Dabei werden die Modellanpassungswerte für den entsprechenden Wortlesevorgang in eine Rangfolge gebracht. Jene Wörter, die in dieser Rangfolge die vorderen Plätze einnehmen, werden gemäß den entsprechenden

<sup>137</sup> Moderate Reformulierungen sind im Schnitt 5,3 Zeichen kürzer als die Originaltexte, starke Reformulierungen 5,7 Zeichen.

<sup>138</sup> Auch der *fixed effect* für die Textlänge ist erwartungsgemäß signifikant ( $\beta = 54,9$ ;  $SE = 11,7$ ;  $t = 4,68$ ;  $p_{\text{MCMC}} = 0,002$ ).



Modellen als besonders schwierig eingeschätzt. Gehen wir so für das zuletzt berichtete Modell für Nomen (Tab. 7.25, drei rechte Spalten) vor und extrahieren die oberen 50 Wortlesevorgänge, tauchen die in Tabelle 7.26 gezeigten 15 Wörter in dieser Reihenfolge auf den vorderen Rängen auf.

	Vorkommen	Kompositum	Rechtsspr.	Nomin.
Rechtsfolge	2	ja	ja	nein
Beschwerdeführerin	11	ja	ja	nein
Richtigstellung	6	ja	?	ja
Gebührenermäßigung	1	ja	nein	ja
Aufgabenwahrnehmung	8	ja	?	ja
Entfaltungsfreiheit	11	ja	nein	nein
Berufsfreiheit	2	ja	nein	nein
Veröffentlichung	1	nein	nein	ja
Persönlichkeitsrecht	2	ja	ja	nein
Gebührenordnung	1	ja	nein	?
Elternverantwortung	1	ja	?	?
Rechtsverordnung	1	ja	ja	?
Aufrechterhaltung	1	ja	nein	ja
Verbindung	1	nein	nein	ja
Rahmengesetzgebung	1	ja	ja	?

Tab. 7.26: Wörter unter den ersten 50 Rängen der vom Modell geschätzten FPRT und die Anzahl der Vorkommen unter den ersten 50 Rängen.

Es lassen sich einige Muster in dieser Wortliste feststellen: Unter den 15 Wörtern sind 13 Komposita. Diese sind in den allermeisten Fällen rechtssprachliche Fachbegriffe. Interessant ist außerdem, dass auch vergleichsweise einfache bzw. alltagsnahe Wörter wie ‘Verbindung’ und ‘Richtigstellung’ enthalten sind. Dabei handelt es sich um Nominalisierungen<sup>139</sup> im klassischen Sinne. Dies gilt auch für die ‘Gebührenermäßigung’, die ‘Aufgabenwahrnehmung’, die ‘Aufrechterhaltung’. Auch die ‘Rechtsverordnung’ und ‘Rahmengesetzge-

<sup>139</sup> Ein Wort wurde immer dann als Nominalisierung kategorisiert, wenn eine verbal formulierte Variante denkbar ist. Diese muss nicht unbedingt höherfrequent sein. Einige Fälle sind unklar.

bung' sind Nominalisierungen, wobei hier die verbalen Konstruktionen 'Recht verordnen' und 'Gesetz geben' eher ungewohnt erscheinen.<sup>140</sup>

Das vorgestellte Vorgehen stellt eine attraktive Methode dar, von stark verallgemeinernden Modellen, die über ein gesamtes Korpus angepasst wurden, den Bezug zu einzelnen sprachlichen Belegen herzustellen. Wie eben erläutert, ergeben sich so Listen von sprachlichen Strukturen, die in ihrer Zusammenstellung aus linguistischer Sicht durchaus nachvollziehbar sind. Diese Methode sollte in Zukunft jedoch dahingehend weiterentwickelt werden, dass offensichtliche Effekte wie die Wortlänge eine kleinere Rolle in der aufgestellten Rangfolge spielen. Als erste Annäherung für verarbeitungsaufwändige Strukturen funktioniert das Verfahren aber offenbar recht gut. Bei der Vorstellung des Verfahrens im kognitionswissenschaftlichen Kolloquium an der Universität Freiburg wurde eingewandt, dass nicht die geschätzten Lesezeiten, sondern die Residuen aus dem Modell extrahiert werden sollten. Residuen fassen aber jene Varianz in den Lesezeiten, die **übrig bleibt**, wenn man den Einfluss der ins Modell aufgenommenen Faktoren beachtet. Hier sollte mit dem Vorgehen aber nicht der Einfluss der Faktoren ausgeschlossen werden, sondern anhand des konkreten linguistischen Materials „greifbar“ gemacht werden.

## 7.5.2 Komplexe Nominalphrasen

Um die Konsequenzen der Reformulierungen für die Verarbeitung von Nominalphrasen zu überprüfen, bietet es sich an, auf den Datensatz zurückzugreifen, in dem nicht die Wörter die kleinste Einheit der Analyse sind, sondern syntaktische Knoten in der Phrasenstrukturannotation (siehe hierzu den Exkurs zu Regionen in 7.3.2).

Als Datengrundlage werden alle Nominalphrasen in Texten mit dem Komplexitätstyp *NP* selektiert. Die Hypothese ist relativ klar: Wenn die Nominalphrasen durch die Reformulierung linguistisch weniger komplex werden, und sich das auf die Verarbeitung niederschlägt, sollten sich kürzere Lesezeiten auf Nominalphrasen in reformulierten Textversionen zeigen. Tabelle 7.27 fasst die Ergebnisse für drei gemischte Modelle zusammen. Die Prädiktorstruktur war für die Vorhersage der drei Lesezeitmaße *first-pass reading times*, *regression path durations* und *total reading times* immer gleich: Die Regionenlänge (gemessen in Buchstaben und gegebenenfalls Leerzeichen), die relative Po-

<sup>140</sup> Das ist meines Erachtens ein Hinweis darauf, wie fest diese Begriffe in der juristischen Fachsprache verankert sind. Eine Rückführung in eine verbale Formulierung scheint hier bereits unangemessen.

sition im Satz,<sup>141</sup> die Einbettungstiefe der Nominalphrase, die beiden Kontraste für die Reformulierungsversion sowie die Position des Versuchsdurchgangs im Gesamtexperiment ('Trial'). Es gehen außerdem der *random intercept* für VersuchsteilnehmerInnen sowie die *random slope* für VersuchsteilnehmerInnen in Abhängigkeit des Trials mit ein.

	FPRT		RPD		TRT	
	$\beta$	t-Wert	$\beta$	t-Wert	$\beta$	t-Wert
(Intercept)	6,33	53,71***	6,64	63,29***	7,19	94,59***
Reg.länge	0,01	22,6***	0,01	21,66***	0,01	45,52***
rel.Reg.position	0,27	2,7**	0,34	3,86***	-0,12	-2,24*
Knoteneinb.	-0,05	-3,74**	-0,01	-1,01	0,0007	0,10
Reform. moderat	-0,13	-2,07*	-0,03	-0,55	-0,04	-1,00
Reform. stark	-0,16	-2,62**	-0,09	-1,58	-0,02	-0,59
Trial	0,003	-1,17	-0,01	-2,07*	-0,01	-3,45**

Tab. 7.27: Zusammenfassung der Modellergebnisse für logarithmierte FPRTs, RPDs und TRTs auf Nominalphrasen in Texten mit komplexen Nominalphrasen. Signifikanzindikatoren kodieren die üblichen Schwellen (Ermittlung über MCMC-Simulation).

Die Reformulierungen haben einen signifikant beschleunigenden Einfluss auf die FPRTs auf Nominalphrasen. Weder für die Dauer des Regressionspfads noch die Gesamtlesezeit lässt sich dieser Effekt zeigen.

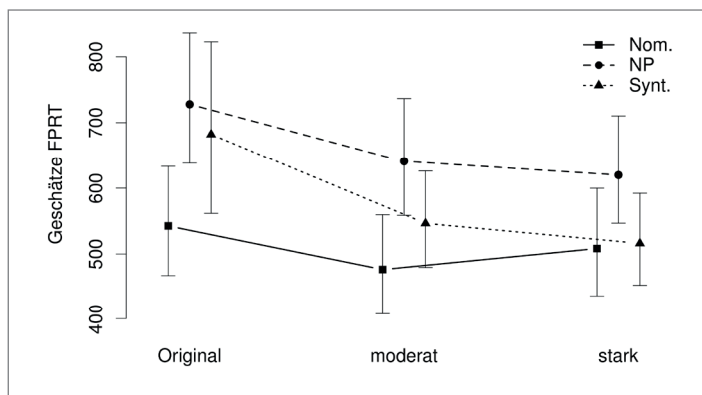


Abb. 7.20: Geschätzte, zurücktransformierte FPRTs auf Nominalphrasen in Abhängigkeit der Reformulierungsversion (je Komplexitätstyp ein Modell).

<sup>141</sup> Die relative Position des Endes der Region (im Vergleich zu Mitte und Anfang) hat hier die größte Vorhersagekraft.

Abbildung 7.20 (gestrichelte Linie/ausgefüllte Kreise) zeigt, dass dieser Effekt auf die FPRTs insbesondere zwischen den Originaltexten und den moderaten Reformulierungen besteht. Nominalphrasen in reformulierten Texten werden ca. 90 ms schneller gelesen. Der erwartete Effekt tritt also ein: Reformulierte Nominalphrasen werden in der Tat kürzer gelesen. Der Effekt zeigt sich aber ausschließlich beim ersten Durchlesen der Nominalphrasen.

In Abschnitt 5.2.1 wurde gezeigt, dass die Anzahl der Kinderknoten unter Nominalphrasen über die Reformulierungsversionen hinweg drastisch abnimmt – allerdings gilt das nur für syntaktisch komplexe Texte (siehe Abbildung 5.5 rechts). Abbildung 7.20 zeigt auch die Schätzer für die syntaktisch komplexen Texte (gestrichelte Linie, Dreieck). Auch dort zeigt sich ein beschleunigender Effekt der Reformulierungsversion auf die Dauer des ersten Lesens (FPRTs).

Die Texte mit komplexen Nominalphrasen stellen einen besonders interessanten Fall dar, denn wie der rechte Teil von Abbildung 5.5 zeigt, werden die Kinderknoten unter den Nominalphrasen nicht weniger, obwohl das durch die Reformulierungen zu erwarten wäre. Trotzdem steigt mit der Reformulierung der Texte die Geschwindigkeit des ersten Lesens. Doch welche linguistische Variable ist für diesen Effekt verantwortlich, wenn es nicht die Anzahl der Kinderknoten unter Nominalphrasen ist?

Betrachtet man die verwendeten Texte, fällt auf, dass insbesondere das Vorkommen postmodifizierender Genitivattribute (Nominalphrasen, die mit dem Kantenlabel GR<sup>142</sup> gekennzeichnet sind) für die Phrasenkomplexität in den Nominalphrasen der Originaltexte verantwortlich ist. Für die Rechtssprache als Ganzes ist das keine Seltenheit. Für die Verwaltungssprache beschreibt beispielsweise auch Becker-Mrotzek (1999), dass „bei den ergänzenden Satzgliedern [...] vor allem die partizipialen und substantivischen Attribute durch Klammerbildung eine hohe Informationsdichte [ermöglichen]“ (S. 1396). Auch Hansen (2009) greift diesen Aspekt im Zusammenhang mit Verwaltungssprache auf (vgl. ebd., S. 297). Beispiel 32a ist diesbezüglich besonders eindrücklich.

(32) a. (Komplexitätstyp *NP*, Originalversion)

Bei der Umsetzung [<sub>1</sub> der Vorgaben [<sub>2</sub> der Gerichte] für eine verfassungskonforme Regelung [<sub>3</sub> der Überführung von Ansprüchen und Anwartschaften aus den Zusatz- und Sonderversor-

<sup>142</sup> *Genitive Right*, vgl. Bohnet (2003). Für das Reformulierungskorpus (FLRCRef und Originaltexte) liegt eine Annotation vor, die Kantenlabels umfasst. Für die vollständigen natürlichen Texte ist das aus Ressourcengründen nicht der Fall.

gungssystemen [<sub>4</sub> der ehemaligen DDR]] lässt sich der Gesetzgeber von der befriedenden Wirkung [<sub>5</sub> dieser Entscheidungen] leiten.

b. (Komplexitätstyp *NP*, starke Reformulierung)

Der Gesetzgeber lässt sich, wenn er die Vorgaben [der Gerichte] umsetzt, wie verfassungskonform zu regeln ist, dass Ansprüche und Anwartschaften aus den Zusatz- und Sonderversorgungssystemen [der ehemaligen DDR] überführt werden, von der befriedenden Wirkung [dieser Entscheidungen] leiten.

Alle rechtsangeschlossenen Genitivattribute sind über Klammerung hervorgehoben. Die Indizes stehen der Übersichtlichkeit halber am Anfang der zu nummerierenden Phrase. Modifiziert werden 'Umsetzung', 'Vorgaben', 'Regelung', 'Sonderversorgungssysteme' sowie 'Wirkung'. Dabei sind drei der fünf Genitivattribute (nämlich die Phrasen 2, 3 und 4) in ein anderes Genitivattribut eingebettet (in Phrase 1). Die Anzahl der Genitivattribute nimmt in Texten des Komplexitätstyps *NP* im Übergang von den Originaltexten zu moderat reformulierten Texten deutlich ab (von insgesamt 25 auf 16), im Übergang von moderat reformulierten Texten zu stark reformulierten Texten fällt nur noch ein Genitivattribut weg.

Sollte die Verarbeitung dieser teilweise exzessiv eingesetzten Genitivattribute tatsächlich Schwierigkeiten bereiten, müssten diese Genitivattribute langsamer verarbeitet werden als andere Knoten. Dabei sollte die Vergleichsmenge nicht einfach aus allen anderen Phrasenstrukturknoten bestehen, sondern nur aus Nominalphrasen, die keine Genitivattribute sind. Die in Tabelle 7.27 aufgeführten und in Abbildung 7.20 (gestrichelte Linie, Datensymbol: Kreise) dargestellten Effekte wären dann darauf zurückzuführen, dass weniger schwer zu verarbeitende Genitivattribute in den Texten sind – die Verarbeitung von Nominalphrasen insgesamt sollte dann einfacher werden.

Zur Überprüfung dieser Hypothese wird ein gemischtes Modell mit den Prädiktoren 'Regionenlänge in Buchstaben', 'relativer Position der Region im Satz', 'Knoteneinbettung', 'Reformulierungsversion', 'zeitliche Position im Experiment', sowie dem Faktor 'Genitivattribut (ja/nein)' herangezogen. Außerdem wird die Interaktion Genitivattribut × Reformulierungsversion aufgenommen. Darüber könnten eventuelle Unterschiede in der Auswirkung der Reformulierung auf die Genitivattribute im Vergleich zu anderen Nominalphrasen erfasst werden. Es gehen, wie gehabt, ein *random intercept* für VersuchsteilnehmerInnen sowie eine *random slope* in Abhängigkeit der zeitlichen Position im Experiment in das Modell ein.

Abbildung 7.21 visualisiert den Interaktionseffekt Reformulierungsversion  $\times$  Genitivattribut. Die Interaktion selbst zeigt keine signifikanten Effekte ( $t_1 = -0,43$ ;  $t_2 = -0,19$ ). Der Kontrast zwischen Genitivattributen und Nicht-Genitivattributen hingegen ist signifikant ( $t = 2,38$ ;  $p_{\text{MCMC}} = 0,019$ ) und insbesondere in den Originaltexten sehr ausgeprägt. Das ist zu erwarten, denn insbesondere in den Originaltexten kommen lange, verschachtelte Genitivattribute vor. In den reformulierten Versionen finden sich weniger Genitivattribute. Nominalphrasen mit dem Kantenlabel GR werden – so die geschätzte Effektgröße laut gemischtem Modell, wenn alle anderen Faktoren konstant gehalten werden – 150 ms länger gelesen als andere Nominalphrasen.

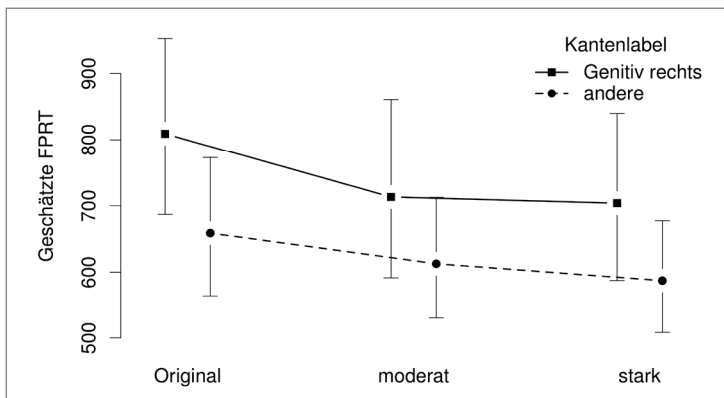


Abb. 7.21: Geschätzte, zurücktransformierte FPRTs auf Nominalphrasen in Abhängigkeit von Genitivattribut (ja/nein) und Reformulierungsversion (Interaktionseffekt).

Die vorgestellten Ergebnisse zur Nominalphrasenkomplexität sind auf den ersten Blick nicht mit den zuvor präsentierten Analysen zur Verarbeitung von Nominalisierungen verbunden. Stellen wir uns jedoch die Frage, **wie** die komplexen Genitivattribute in den Reformulierungen „aufgebrochen“ werden, wird eine Verbindung offensichtlich: Häufig werden in den Genitivattributen verwendete Nominalisierungen in verbale Strukturen umformuliert. Das Genitivattribut wird dabei in einen subordinierten Satz umgewandelt. Das erklärt auch das in Abbildung 5.2 (linkes Schaubild) gefundene Muster: Auch in Texten des Komplexitätstyps *NP* sinkt über alle Texte hinweg der Anteil an Nominalisierungen. Exemplarisch ist dies auch an den Reformulierungen des Originalauszugs aus Beispiel 32a zu beobachten. Zwei post-modifizierende Genitivattribute sind im Vergleich zum Originaltext weggefallen und die übrig gebliebenen Attribute sind nicht mehr ineinander verschachtelt. Außerdem wurden die Nominalisierungen ‘Umsetzung’, ‘Re-

gelung' und 'Überführung' allesamt im Laufe des Reformulierungsprozesses in verbale Strukturen umgewandelt, was mit der Einführung subordinierter Teilsätze einher geht.

Auch hier soll noch die Gesamtlesezeit der Texte untereinander verglichen werden. Denn was würde es den LeserInnen nützen, wenn zwar die Nominalphrasen einfacher werden, dafür aber die Sätze insgesamt länger werden und deshalb auch länger gelesen werden müssten. Zunächst: Die Texte werden tatsächlich im Schnitt etwas länger: Die durchschnittliche Textlänge steigt von 205 (Originale) auf 214 (moderate Reformulierungen) und 227 (starke Reformulierungen) Zeichen. Trotzdem werden die Texte nicht signifikant langsamer gelesen. Zwar steigt die durchschnittliche Lesezeit der Texte um 776 ms (Original zu moderat) bzw. 2201 ms (Original zu stark). Das scheint aber laut des gemischten Modells jeweils im Bereich einer Zufallsschwankung zu liegen. Die Unterschiede sind nicht signifikant.

### 7.5.3 Syntaktisch komplexe Texte

In Abschnitt 5.2.1 (insbesondere Abb. 5.5) wurde gezeigt, dass in den syntaktisch komplexen Texten der Anteil von Demonstrativ- und Indefinitpronomen über die Versionen hinweg steigt. Es wurde außerdem gezeigt, dass die Sätze analog dazu immer kürzer werden (Abb. 5.4). Es wurde argumentiert, dass dies einer Komplexitätsverschiebung von der syntaktischen auf die textuelle Ebene gleich kommt. Beziehungen, die zwischen Personen, Sachverhalten und/oder Ereignissen bestehen, können entweder innerhalb eines Satzes über Sub- oder Koordination oder eben zwischen Sätzen ausgedrückt werden. Wird die Information über mehrere Sätze verteilt, muss die Textkohärenz über referenzielle Ausdrücke aufrecht erhalten werden (siehe dazu auch Beispiel 12 und die dazugehörigen Ausführungen in 5.2.1). An dieser Stelle werde ich nun eine Auswertung auf linguistischer Ebene nachreichen, die zwar nah an die in Abbildung 5.5 gezeigten Kontraste herankommt, aber noch näher am zu untersuchenden Phänomen ist.

Abbildung 7.22 zeigt die Anteile der referenziellen Ausdrücke gemäß der in Abschnitt 5.2.3 eingeführten Koreferenzannotation. Das linke Schaubild zeigt dabei recht deutlich, dass es insbesondere die syntaktisch komplexen Texte sind, in denen der Anteil der referenziellen Ausdrücke über die Reformulierungsversionen hinweg ansteigt. Das repliziert ziemlich parallel das Muster, das in Abbildung 5.5 über die Part-of-Speech-(POS)-Annotation eingeführt wurde. Das rechte Schaubild unterteilt die rechte Balkengruppe und spaltet die referenziellen Ausdrücke auf in volle, referierende Nominalphrasen und Pro-



formen (Pronomen und Demonstrativa). Diese Unterscheidung wurde bereits in Abschnitt 7.4.1 eingeführt. Im Übergang von Originalversion zu moderater Reformulierung sind zunächst mehr Proformen zu beobachten.<sup>143</sup> Der Anteil von Proformen steigt dann im Übergang von moderater zu starker Reformulierungsversion kaum mehr (0,3%). Dafür werden mehr referierende NPs eingesetzt (Anstieg: 3,9%). Die referenzielle Komplexität scheint also tatsächlich zu steigen. Das lässt sich überprüfen, indem wir als weitere Variable für die referenziellen Ausdrücke das Überschreiten einer Satzgrenze hinzunehmen.

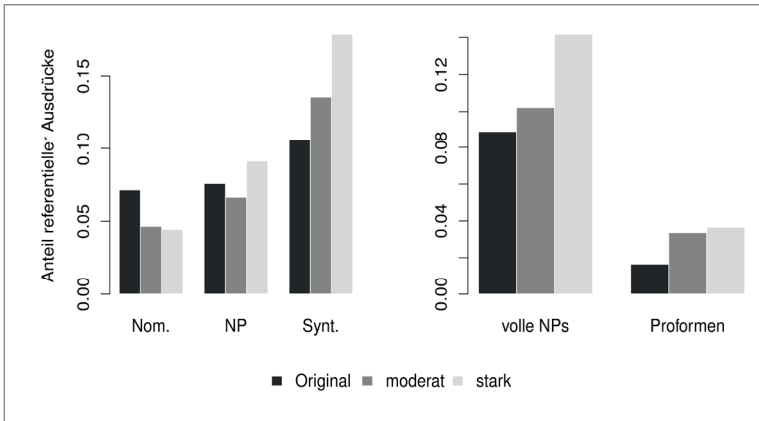


Abb. 7.22: Anteil referenzieller Ausdrücke über alle Komplexitätstypen hinweg (links) sowie der Anteil der verschiedenen Referenztypen in syntaktisch komplexen Texten (rechts).

Das Ergebnis ist in Abbildung 7.23 in Form eines Mosaik-Plots abgebildet. Dieser ist auf Anhieb nicht ganz leicht zu lesen. Es gilt immer: Der Flächeninhalt eines Rechtecks ist proportional zur Anzahl der Fälle, die in den entsprechenden Block fallen. Die verschiedenen Reformulierungsversionen sind in Blöcken von oben nach unten voneinander abgesetzt. In jedem Block befinden sich vier Rechtecke.<sup>144</sup> Die oberen zwei Rechtecke symbolisieren jeweils die Anzahl der referierenden Nominalphrasen, die unteren die Anzahl der Proformen. Vertikal unterteilt sich jeder Block in referenzielle Ausdrücke, die die Satzgrenze nicht überschreiten (links) und jene, die es tun (rechts). Es wird sehr deutlich, dass im Übergang von Original zu starker Reformulierung die

<sup>143</sup> Der Anstieg für Proformen (1,7%) ist nur minimal größer als jener für die vollen NPs (1,3%). Im Verhältnis zum Ausgangspunkt (Proformen: 1,7%; NPs: 8,9%) muss der Anstieg für Proformen aber als größer gewertet werden.

<sup>144</sup> Im obersten Block befinden sich nur zwei Rechtecke und zwei gestrichelte Linien, die gestrichelten Linien geben an, dass in die entsprechenden Zellen keine Fälle fallen.

referenziellen Ausdrücke eher die Satzgrenze überschreiten. Während in den Originalen nur Referenzen innerhalb von Sätzen auftauchen, kehrt sich dieses Verhältnis für die starken Reformulierungen um. Insgesamt gilt also: Während in den syntaktisch komplexen Originaltexten noch alle Referenzen innerhalb eines Satzes realisiert sind, gewinnen in den stark reformulierten Texten die Referenzen über die Satzgrenze – und dort besonders die Nominalphrasen – die Oberhand. In diesem Sinne sinkt also die Komplexität auf der Satzebene (siehe beispielsweise die Abb. 5.3 und 5.4 sowie Tab. 5.2), sie wird aber – zumindest was die referenziellen Beziehungen angeht – auf die textuelle Ebene verschoben.

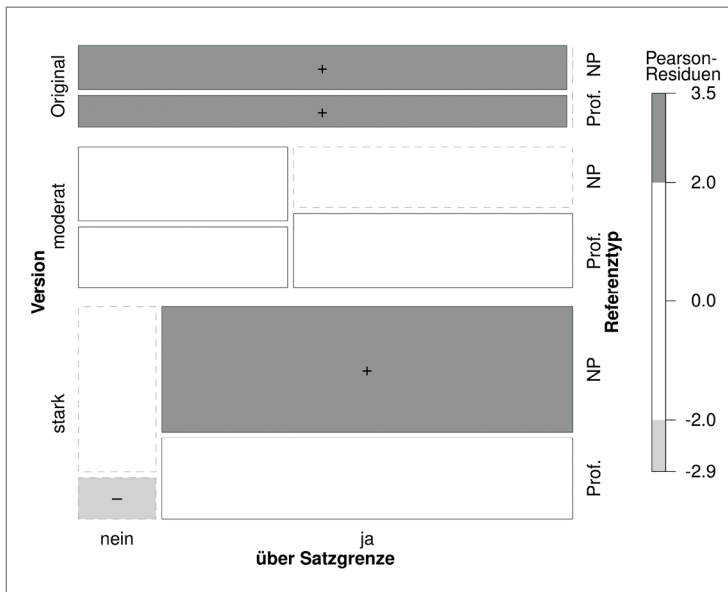


Abb. 7.23: Mosaikplot für die Verteilung referenzieller Ausdrücke in syntaktisch komplexen Texten. Eingefärbte Flächen deuten auf eine signifikante Abweichung von der Erwartung hin. Die Richtung der Abweichung (nach oben/unten) ist über das Vorzeichen kodiert. Für die dem Plot zugrunde liegende Kontingenztabelle gilt:  $\chi^2(11) = 83,4; p = 3,28 \times 10^{-13}$ .

Zunächst soll nun überprüft werden, ob diese linguistische Komplexitätsverlagerung einen Effekt auf die Verarbeitung hat. In Abschnitt 7.3 wurde gezeigt, dass in bestimmten Fällen tiefere syntaktische Einbettungen mit schnelleren Lesezeiten bzw. kürzeren Regressionspfaden assoziiert sind. Wie wird dies aber von den Reformulierungen beeinflusst? Gerade weil die Reformu-

lierungen dazu geführt haben, dass Sätze weniger tief eingebettet werden, sollte hier ein Effekt zu erwarten sein. Man könnte beispielsweise erwarten, dass sich der beschleunigende Effekt der Einbettungstiefe für reformulierte Texte verstärkt. Wenn Sätze syntaktisch weniger komplex werden, könnte eventuell die Einbettungstiefe von Wörtern von den LeserInnen noch besser als Signal der Relevanz oder der syntaktischen Constraints genutzt werden.

Groups	Name	Variance	Std.Dev.			
vtn	(Intercept)	3.4600e-12	1.8601e-06			
Residual		6.7171e-01	8.1958e-01			
		Estimate	Std. Error	t value	$P_{MCMC}$	
(Intercept)		0.045653	0.117147	0.390	0.7144	
version moderat		0.106732	0.145614	0.733	0.4482	
version stark		0.333983	0.173017	1.930	0.0506	
einb.tiefe		-0.001239	0.015801	-0.078	0.9496	
version moderat:einb.tiefe		-0.030017	0.023777	-1.262	0.2118	
version stark:einb.tiefe		-0.139674	0.041927	-3.331	0.0006	***

Tab. 7.28: Lineares gemischtes Modell für residuale RPDs in Abhängigkeit der Reformulierungsversion (*version*) und Einbettungstiefe (*einb.tiefe*). Nur syntaktisch komplexe Texte; nur Wörter, die nicht als referenzieller Ausdruck annotiert wurden.

In der Tat scheint sich ein solches Effektmuster zu zeigen. In Abbildung 7.24 ist zu sehen, wie sich der Effekt der Einbettungstiefe auf residuale *regression path durations* mit steigender Reformulierungsversion verstärkt. Das gemischte Modell dazu ist in Tabelle 7.28 einzusehen. Als abhängige Variable dient dabei die von lexikalischen Einflüssen befreite RPD, die zusätzlich nach Bi- und Trigramm-Frequenz sowie nach relativer Satzposition und der Interaktion relative Satzposition  $\times$  Einbettungstiefe residualisiert ist. Als Prädiktoren in das eigentliche Modell gehen dann die Reformulierungsversion, die Einbettungstiefe sowie die Interaktion beider Variablen mit ein. Um Effekte der referenziellen Ebene weitestgehend ausblenden zu können, wurden nur Wörter in das Baseline-Modell und das Analysemodell aufgenommen, die nicht als referenzielle Ausdrücke annotiert wurden.

Wie die Breite der Konfidenzintervalle in Abbildung 7.24 bereits vermuten lässt, gibt es in stark reformulierten Texten der Komplexitätsklasse *Synt.* logischerweise weniger Wörter, die besonders tief eingebettet sind. Um zu überprüfen, ob der beobachtete Effekt auf diesen Umstand zurückzuführen ist, replizieren wir das gemischte Modell aus Tabelle 7.28 nur mit Wörtern, die auf Einbettungstiefe 6 oder höher angesiedelt sind. Diese Grenze wurde ge-

wählt, da hier noch in allen Reformulierungsversionen Wörter vorkommen. Auch anhand dieses Modells kann der kritische Interaktionseffekt gezeigt werden ( $t = -2,19$ ;  $p_{\text{MCMC}} = 0,023$ ). Wir können also schließen, dass die Reformulierungen einen Einfluss auf die Rezeption der syntaktisch komplexen Texte haben. Folgt man der Argumentation, dass tiefere syntaktische Einbettung als Hinweis auf geringere Relevanz genutzt werden kann, wirken sich die Reformulierungen positiv auf den Verständnisprozess aus. Nachdem wir die eine Seite der Komplexitätsverschiebung betrachtet haben (die syntaktische Ebene), wollen wir uns nun Effekten auf der Diskursebene zuwenden – genauer: den Effekten der referenziellen Komplexität innerhalb der Texte.

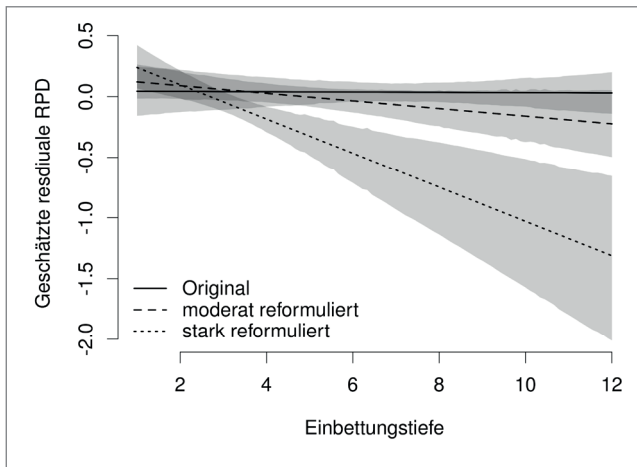


Abb. 7.24: Interaktion Reformulierungsversion  $\times$  Einbettungstiefe für syntaktisch komplexe Texte laut dem Modell aus Tabelle 7.28; beachtet sind nur Wörter, die nicht als referenzieller Ausdruck annotiert wurden.

Hierzu werden alle Wörter in referenziellen Ausdrücken selektiert. Als *fixed effects* geht die Information mit ein, ob das Wort das erste oder letzte Wort der Zeile ist (beide Variablen können jeweils entweder den Wert WAHR oder FALSCH annehmen) und die Textzeile. Als Variablen auf lexikalischer Ebene dienen die Wortlänge, die residuale Wortfrequenz und die residuale Familiarität. Als Designfaktor wird die Reformulierungsversion aufgenommen. Die Zufallseffektstruktur besteht aus zwei *random intercepts* für VersuchsteilnehmerInnen- und Textidentität sowie einer *random slope* für VersuchsteilnehmerInnen in Abhängigkeit des Versuchsdurchlaufs (der Versuchsdurchlauf selbst ist deshalb auch als fester Faktor aufgenommen).

Groups	Name	Variance	Std.Dev.			
text.id	(Intercept)	4.2283e-02	0.2056292			
vtn	trial	2.5138e-05	0.0050137			
vtn	(Intercept)	3.1511e-02	0.1775138			
Residual		5.6147e-01	0.7493107			
		<b>Estimate</b>	<b>Std. Error</b>	<b>t value</b>	<b><math>P_{\text{MCMC}}</math></b>	
(Intercept)		5.3237141	0.1044263	50.98	0.0001	***
w.len		0.0455696	0.0041124	11.08	0.0001	***
res.freq.dlex.lm		-0.0148637	0.0182062	-0.82	0.4212	
res.famil.lm		0.0851895	0.0401180	2.12	0.0368	*
last.word.row yes		0.1615995	0.0555888	2.91	0.0042	**
first.word.row yes		-0.0466057	0.0704699	-0.66	0.5268	
text.row		0.0053952	0.0102615	0.53	0.6536	
version moderat		0.0363424	0.0832444	0.44	0.6518	
version stark		0.2190609	0.0808458	2.71	0.0058	**
ref.typ pro-form		0.1707732	0.0532431	3.21	0.0012	**
typ NP		0.0560946	0.0945986	0.59	0.5518	
typ Synt.		-0.1125225	0.0924207	-1.22	0.2042	
trial		-0.0003573	0.0020905	-0.17	0.8620	

Tab. 7.29: Lineares gemischtes Modell für die *regression path durations* auf referierenden Ausdrücken.

Tabelle 7.29 zeigt das resultierende Modell. Außer der Wortlänge hat u.a. der Referenztyp einen Effekt – der Regressionspfad von Proformen ist länger als jener von referierenden NPs. Dass Proformen schwieriger zu verarbeiten sind als referierende Nominalphrasen, wurde bereits im vorhergehenden Abschnitt 7.3 gezeigt. In der Tat ergibt sich auch ein signifikanter Kontrast zwischen starken Reformulierungen und den beiden anderen Versionen.<sup>145</sup> Um diesem Effekt auf die Spur zu kommen, können wir noch zusätzlich die Interaktion Version  $\times$  Komplexitätstyp mit in das Modell aufnehmen. Diese wird zwar selbst nicht signifikant, erlaubt es uns aber, separate Modellschätzer für den Effekt der Version innerhalb der verschiedenen Komplexitätstypen zu erhalten. Die geschätzten Werte sind in Abbildung 7.25 abgetragen.

Das Ergebnis ist einigermaßen überraschend. Es sind offenbar nicht die syntaktisch komplexen Texte, bei denen die Einführung von zusätzlichen referierenden Elementen durch die Reformulierungen zu verlängerten Regressionspfaden führt. Es sind die Texte der beiden anderen Komplexitätstypen, die den im Modell aus Tabelle 7.29 gezeigten Effekt speisen. Bei den Texten mit

<sup>145</sup> Nur der Kontrast original vs. stark ist in Tabelle 7.29 vermerkt. Eine Umordnung des Faktors ergibt für den Kontrast moderat vs. stark einen *t*-Wert von 2,31.

vielen Nominalisierungen zeigt sich der größte Unterschied (Effektgröße: 107 ms), obwohl die Anzahl referenzieller Ausdrücke bei den Texten mit vielen Nominalisierungen über die Reformulierungsstufen hinweg sinkt. Der gezeigte Effekt scheint also von einigen wenigen referenziellen Ausdrücken zu stammen. Bei den Texten mit komplexen Nominalphrasen (Effektgröße Original vs. stark: 96 ms) ist nicht verwunderlich, dass wir den Effekt genau an dieser Stelle beobachten können. Wir erinnern uns (siehe 5.2.1): Im Übergang von Originaltexten zu moderaten Reformulierungen wurde der Inhalt komplexer Nominalphrasen zunächst in subordinierte Strukturen ausgelagert. Diese wurden im zweiten Reformulierungsschritt in eigenständige Sätze transformiert. Das Vorgehen war hier das gleiche wie bei Reformulierungen des Komplexitätstyps *Synt*. Bei den Texten mit komplexen Nominalphrasen scheint der erwartete Effekt jedoch um einiges stärker zu sein als bei den syntaktisch komplexen Texten selbst.

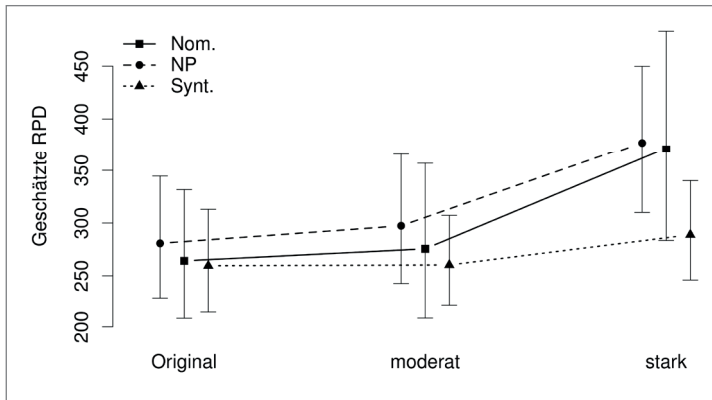


Abb. 7.25: Geschätzte, zurücktransformierte RPDs auf referierenden Ausdrücken in Abhängigkeit der Textversion und des Komplexitätstyps.

Das bedeutet: Tatsächlich steigt mit stärkerer Reformulierungsversion der Regressionspfad für referenzielle Ausdrücke. Dieser Effekt wird jedoch hauptsächlich von den Texten mit vielen Nominalisierungen und komplexen Nominalphrasen gespeist. Dort kann also tatsächlich von einer Verlagerung der Verarbeitungskomplexität von der syntaktischen auf die textuelle Ebene gesprochen werden. Für syntaktisch komplexe Texte konnte dieses Muster zwar auf linguistischer Ebene gezeigt werden, die Effekte auf Verarbeitungsebene sind jedoch nur numerisch zu beobachten und sehr viel kleiner als für die anderen Komplexitätstypen. Leider ist die Datenbasis für eine weitergehende Analyse, die sich auf syntaktisch komplexe Texte beschränken würde, zu klein. Wünschenswert wäre hier natürlich noch die detaillierte Aufschlüsse-

lung nach Anaphern über die Satzgrenze vs. Referenzen innerhalb eines Satzes. Auch die anderen in Abschnitt 7.4 untersuchten Eigenschaften wie das Hinzufügen von Informationen und das Wechseln auf eine andere Spezifitätsebene könnten noch miteinander gekreuzt werden. Mit jedem Faktor, der zusätzlich beachtet wird, sinkt jedoch die Anzahl der Datenpunkte pro Datenzelle. Das ist unbefriedigend und deutet bereits auf ein Problem hin, das für die Bearbeitung von sehr detaillierten Sachfragen anhand des FLRC besteht: Das Blickbewegungskorpus umfasst – obwohl Lesedaten von 80 VersuchsteilnehmerInnen verfügbar sind – für solche Detailfragen dann eben doch zuwenig Datenpunkte. Müssen viele (kategoriale) Prädiktoren eingesetzt und/oder eine Vorauswahl ganz bestimmter linguistischer Strukturen getroffen werden, stößt man auch mit Daten von 80 VersuchsteilnehmerInnen schnell an eine kritische Grenze.

Die letzte Analyse bezüglich der syntaktisch komplexen Texte soll wiederum mit möglichst einfachen Mitteln die Auswirkungen der Reformulierungen auf die Gesamtlesezeit beleuchten. Die syntaktisch komplexen Texte werden im Übergang von Originaltext (durchschnittliche Länge in Zeichen: 273) zu moderater Reformulierung etwas kürzer (262 Zeichen). Die stark reformulierten Texte sind wieder deutlich länger (291 Zeichen). Es ist aufgrund dieser Verteilung zunächst nicht überraschend, dass die moderat reformulierten Texte am schnellsten gelesen werden (Originaltexte: 17,1 s; moderate Reformulierungen: 14,2 s; starke Reformulierungen: 17,2 s). Doch auch wenn man die Textlänge kontrolliert, ist der Unterschied zu den Originalen noch immer signifikant ( $\beta = 1842$ ;  $SE = 712$ ;  $t = 2,589$ ;  $p_{\text{MCMC}} = 0,014$ ). Der Unterschied zu den stark reformulierten Texten ist lediglich marginal signifikant ( $\beta = 1248$ ;  $SE = 706$ ;  $t = 1,767$ ;  $p_{\text{MCMC}} = 0,081$ ). Moderat reformulierte Texte scheinen hier also am schnellsten gelesen zu werden.

#### 7.5.4 Fernabhängigkeiten: Dependenzdistanz

Im letzten Teil dieses Abschnitts wollen wir uns mit einem Phänomen beschäftigen, das in der Psycholinguistik von großem Interesse ist: Fernabhängigkeiten. Es wird dabei insbesondere die Abhängigkeit zwischen einem finiten Verb und seinen Argumenten betrachtet. Hier werden wir uns mit dem Subjekt beschäftigen. Wir erinnern uns: In Abschnitt 7.3.5 spielte die Beziehung zwischen einem Verb und seinen Argumenten bereits eine Rolle. Dort wurde – ohne Berücksichtigung der korrekten Anbindung – gezeigt, dass die Verarbeitung von Verben länger dauert, wenn mehr potenzielle Referenten vor dem Verb im Satz stehen. Für die nun folgenden Analysen wurden die Annotationen des Reformulierungsteils des FLRC benutzt, um die Abhän-



gigkeit zwischen einem Verb und dessen Subjekt zu extrahieren. Der Einfluss der Distanz zwischen dem Verb und dem dazugehörigen Subjekt wurde prominent von Gibson (1998, 2000) untersucht. In der *Dependency Locality Theory* (DLT) wird besonderer Wert auf die Inkrementalität menschlicher Satzverarbeitung gelegt. Da Wörter nach und nach in das Satzverarbeitungssystem eingehen, so argumentiert Gibson (2000, S. 95), wird auch die mentale Repräsentation des Satzes nach und nach aufgebaut. Dabei spielen zwei komputationale Ressourcen eine tragende Rolle: Strukturelle Integration und das Aufrechterhalten der bereits herausgebildeten Strukturen im Gedächtnis. Die erste wird dazu verwendet, ein Wort in die bis zu diesem Zeitpunkt aufgebaute Struktur einzubauen. Die zweite muss unter anderem dafür sorgen, dass offene Abhängigkeiten aufrecht erhalten werden. „One of the key ideas underlying [DLT] is *locality*, such that the cost of integrating two elements [...] depends on the distance between the two. [Hervorh. i. Orig.]“ (Gibson 2000, S. 95f.). Die postulierte Richtung dieses Zusammenhangs ist relativ klar: Je größer die Distanz oder Lokalität<sup>146</sup> zwischen zwei zu integrierenden Elementen ist, desto komplexer ist deren Integration, was sich entsprechend auf die Lesezeiten niederschlagen sollte. Eben dieser Effektrichtung widerspricht Konieczny (2000). Er zeigt anhand deutscher Sätze mit Verbendstellung, dass sich dieser Lokalitätseffekt umdrehen kann, wenn das zwischen Verb und Argument intervenierende Material das Verb hoch vorhersagbar macht. Später im aktuellen Abschnitt wird diese Studie nochmals aufgegriffen.

In Abschnitt 5.2.1 wurde gezeigt, dass sich die Abhängigkeitsdistanzen im FLRC eindeutig rechtsschief verteilen. Rund die Hälfte aller Verben stehen direkt neben ihrem Subjekt (Distanz = 1). Die Subjekt-Verb-Abfolge verteilt sich ebenfalls nicht symmetrisch: Rund drei Viertel aller Verben stehen nach dem Subjekt. Weiterhin wurde gezeigt, dass sich insbesondere die Reformulierungen von syntaktisch komplexen Texten auf die durchschnittliche Abhängigkeitsdistanz auswirken: Über die verschiedenen Versionen hinweg sinkt die durchschnittliche Abhängigkeitsdistanz deutlich. Es tritt außerdem eine Verschiebung der Verb-Subjekt-Abfolge ein: Durch die Einführung von Fragen und das Trennen von Sätzen „rutschen“ mehr Verben vor ihre Subjekte.

In der oben genannten (Anti-)Lokalitätsdebatte sind die Lesezeiten direkt auf dem Verb von Interesse. Auch im kognitiven Modell zu *cue-based parsing* von Lewis/Vasishth (2005) sind die Abrufzeiten am und damit die Verarbeitungszeiten auf dem finiten Verb von Interesse, denn zu diesem Zeitpunkt findet die Integration mit den Argumenten statt. Die vorliegende Annotation berücksich-

<sup>146</sup> Gibson (2000) benutzt die beiden Begriffe oft synonym. Für eine feinere Unterscheidung sei auf die entsprechenden Texte verwiesen.

tigt dies insofern, als die Lokalitätsinformation immer am finiten Verb angeschlossen ist und nicht am zugehörigen Subjekt. Die Datenselektion für die folgende Analyse definiert sich also folgendermaßen: Es werden alle Lesezeiten auf finiten Verben selektiert. Für das Reformulierungskorpus sind das 3.356 Beobachtungen. Als lokalitätsrelevante Prädiktoren stehen die Distanz zwischen Verb und Subjekt (wie in 5.2.1 beschrieben) sowie die Abfolge von Verb und Subjekt zur Verfügung. Die Hypothesen sind relativ klar: Überwiegen Lokalitätseffekte, wie Gibson (2000) sie beschreibt, sollte mit steigender Distanz zwischen Verb und Subjekt die Lesezeit auf dem Verb steigen. Überwiegen jedoch Vorhersagbarkeitseffekte, wie sie von Konieczny (2000) beschrieben werden, sollte steigende Subjekt-Verb-Distanz mit kürzeren Lesezeiten auf dem Verb assoziiert sein – zumindest gilt das allgemein betrachtet über alle Verben hinweg. Es ist außerdem zu erwarten, dass die kanonische Reihenfolge im Hauptsatz Subjekt – Verb vorteilhaft für die Verarbeitung ist. In einer Voranalyse bestätigt sich dies für FPRTs und die Wahrscheinlichkeit einer regressiven Sakkade. Da die Ergebnisse dieser Voranalyse ansonsten eher ohne starke Hinweise bleiben, wollen wir uns auf Fälle beschränken, in denen das Verb **hinter** dem Subjekt steht. Ohnehin kann man eigentlich nur bei dieser Konstellation davon ausgehen, dass die am Verb angeschlossene Dependenzdistanz tatsächlich das operationalisiert, was sie operationalisieren soll.

Nach der Selektion von „Postsubjekt-Verben“ stehen 2.629 Datenpunkte zur Verfügung. Als Kontrollprädiktoren fungieren Wortlänge, Wortfrequenz, Wortfamiliarität, Darstellungsfaktoren und die relative Position des Verbs im Satz, weil hier in der kurz erwähnten Voranalyse relevante Einflüsse gezeigt werden konnten. Dieser Prädiktor variiert zwischen nahe 0 (erstes Wort im Satz)<sup>147</sup> und 1 (letztes Wort im Satz). Um unterschiedlich starke Einflüsse der Dependenzdistanz für Verben an unterschiedlichen Stellen im Satz zu erlauben, wurde außerdem die Interaktion Dependenzdistanz  $\times$  relative Position mit in das Modell aufgenommen.

Für FPRTs sind keine Effekte zu beobachten. Dasselbe gilt für TRTs (alle relevanten  $ts < 1$ ). Das Modell für RPDs ist in Tabelle 7.30 zu sehen. Es wird ein signifikanter negativer Einfluss der Dependenzdistanz (*dep.dist*) auf die *regression path durations* identifiziert. Dieser Effekt wird jedoch durch die Interaktion mit der relativen Einbettungstiefe (*rel.pos*) moduliert. Der Schätzer (*Estimate*) für diese Interaktion liegt in ganz ähnlichen Regionen wie jener für den einfachen Effekt der Dependenzdistanz, hat aber das entgegengesetzte

<sup>147</sup> Die relative Satzposition  $x$  wird über  $x = \text{Position des Verbs/Satzlänge}$  berechnet. 0 kann somit nie erreicht werden. Ein Wert für  $x$  nahe 0 bedeutet, dass das Verb sehr nah am Satzanfang steht.

Vorzeichen. Dies lässt darauf schließen, dass die relative Position des Verbs im Satz den Dependenzdistanz-Effekt für Verben, die weit hinten im Satz stehen, nivelliert. Abbildung 7.26 visualisiert diese Interaktion auf Basis der Schätzwerte des Modells.

Groups	Name	Variance	Std.Dev.			
vtn	trial	1.0517e-05	0.003243			
vtn	(Intercept)	6.6597e-02	0.258065			
item.id	(Intercept)	1.3075e-01	0.361588			
Residual		9.1775e-01	0.957994			
		Estimate	Std. Error	t value	$p_{MCMC}$	
(Intercept)		5.311058	0.149230	35.59	0.0001	***
w.len		0.044036	0.016606	2.65	0.0072	**
res.freq.dlex.lm		-0.096690	0.044099	-2.19	0.0274	*
res.famil.lm		0.090457	0.081027	1.12	0.2654	
trial		-0.007078	0.003145	-2.25	0.0316	*
text.row		0.047337	0.020328	2.33	0.0260	*
first.word		-0.162974	0.176092	-0.93	0.3492	
last.word		0.387892	0.082911	4.68	0.0001	***
rel.pos		0.423939	0.162291	2.61	0.0100	*
dep.dist		-0.115192	0.019265	-5.98	0.0001	***
rel.pos:dep.dist		0.123665	0.021925	5.64	0.0001	***

Tab. 7.30: Lineares gemischtes Modell für die *regression path durations* auf finiten Verben.

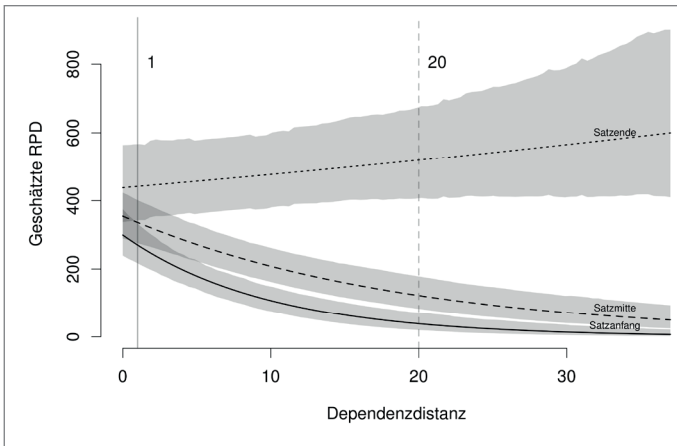


Abb. 7.26: Interaktion Dependenzdistanz × relative Position des Verbs im Satz für RPDs (zurücktransformiert aus logarithmierten RPDs). 95%-Konfidenzintervalle sind über MCMC-Simulationen ermittelt. Zur Bedeutung der senkrechten Linien siehe die Erklärungen im Text.

Für die relative Position des Verbs im Satz wurden für Abbildung 7.26 drei Werte gewählt, die einigermaßen einfach interpretierbar sind. Die durchgezogene Linie ist jene für den Wert  $rel.pos = 0,1$ , also für jene Verben, die relativ weit vorne im Satz stehen. Die gestrichelte Linie repräsentiert  $rel.pos = 0,5$ , stellt also den geschätzten Einfluss der Dependenzdistanz für jene Wörter dar, die genau in der Mitte ihres Satzes liegen. Die gepunktete Linie stellt den Wert für Wörter dar, die am Satzende liegen, für die also gilt:  $rel.pos = 1$ . Alle anderen denkbaren Positionen im Satz zwischen 0,1 und 1 liegen zwingend zwischen den entsprechenden Linien, da die Interaktion kontinuierlich geschätzt wird.<sup>148</sup> Die senkrechten Linien in Abbildung 7.26 geben den jeweiligen Maximalwert der Dependenzdistanz für die entsprechenden Werte von  $rel.pos$  an. Logischerweise kann die Dependenzdistanz für Wörter, die am Satzanfang stehen, noch nicht sehr groß sein, denn dort ist noch nicht viel Material, in dem das Subjekt überhaupt stehen könnte. Konsequenterweise ist auch die längste Dependenzdistanz, die für Wörter mit einer relativen Position im Satz  $\geq 0,1$  auftritt, gleich 1. Das heißt, dass hier das Subjekt immer direkt vor dem Verb steht. Die abfallende durchgezogene Linie schätzt also die Dauer des Regressionspfades für Fälle, die im Datensatz gar nicht vorliegen.<sup>149</sup> Nicht so die Linie für Verben, die in der Satzmitte liegen. Hier treten Distanzen von bis zu 20 Wörtern zwischen Verb und Subjekt auf. Beispiel 33 zeigt den Satz, in dem die Dependenzdistanz 20 beträgt.

- (33) Es verstoße gegen Verfassungsgrundsätze, dass ein Ehemann auch dann an der gemeinsamen elterlichen Sorge für sein Kind beteiligt werde, wenn **er** sich schon vor der Geburt des Kindes von der Mutter trenne und daher für das Kind tatsächlich keine Bedeutung **gewinne**, dass aber umgekehrt der Vater eines nicht-ehelichen Kindes auch dann nicht an der elterlichen Sorge beteiligt werde, wenn er lange Jahre mit der Mutter zusammengelebt und für das Kind Verantwortung getragen habe, wie das bei ihm der Fall gewesen sei.

Das relevante 'gewinne' steht in der Mitte des Satzes. Zugehöriges Subjekt ist das ebenfalls fett gedruckte 'er', das auf 'ein Ehemann' zurückreferiert. Die breiten Konfidenzintervalle um die gepunktete Linie (Verben am Satzende) sowie die geringe Steigung der Gerade lassen vermuten, dass keine Effekte

<sup>148</sup> Eine unterschiedliche Anpassung für verschiedene Abschnitte von  $dep.dist$  wäre über sogenannte *generalized additive mixed models* (GAMMs) möglich. Im vorliegenden Fall ist es aber durchaus gewünscht, dass die Interaktion kontinuierlich angepasst wird, da keine Hypothesen für nicht-kontinuierliche Auswirkungen von  $rel.pos$  auf  $dep.dist$  vorliegen.

<sup>149</sup> Durch das Modell wird auf Fälle extrapoliert, die vom Datensatz nicht abgedeckt sind. Das ist in den meisten Fällen eine wünschenswerte Eigenschaft von Regressionsmodellen. Hier suggeriert die Extrapolation aber meines Erachtens einen Zusammenhang, der so aufgrund des Datensatzes nicht haltbar ist.

der Dependenzdistanz am Satzende zu finden sind. Ein zum Modell aus Tabelle 7.30 analoges Modell, in dem nur die Wörter aus der hinteren Hälfte der Sätze aufgenommen wurden, bestätigt das. Die relative Position des Verbs im Satz hat noch immer einen signifikanten Einfluss, nicht aber die Dependenzdistanz oder die Interaktion der beiden Prädiktoren. Wie aber eben gesehen und in Abbildung 7.26 abgetragen, hat die Dependenzdistanz für Wörter in der ersten Satzhälfte einen Einfluss auf die Dauer des Regressionspfads – und zwar einen beschleunigenden. Die Wahrscheinlichkeit von regressiven Sakkaden verhält sich analog (ohne Abb.): Insgesamt liegt die Wahrscheinlichkeit für eine regressive Sakkade am Satzende zwar deutlich höher. Aber je weiter hinten ein Verb im Satz steht, desto geringer wird der Einfluss der Dependenzdistanz zwischen Subjekt und Verb darauf, ob eine regressive Sakkade initiiert wird oder nicht.

Wie ist diese anscheinend recht robuste Interaktion zwischen der Position im Satz und der Dependenzdistanz zwischen Subjekt und Verb zu erklären? Am auffälligsten ist zunächst die Effektrichtung, die offenbar gegen die *Dependency Locality Theory* spricht. Es sind keine Effekte der Dependenzdistanz zu finden, die einen erhöhten Verarbeitungsaufwand für längere Dependenzdistanzen nahelegen. Für *first-pass reading times* und *total reading times* wurden keinerlei Effekte (auch nicht in Form von Interaktionen) gefunden. Die Effektrichtung für *regression path durations* für Verben in der ersten Satzhälfte legt allerdings Antilokalitätseffekte nahe.

Eine Erklärung könnte mit dem Typ des finiten Verbs zusammenhängen. In die Analysen gingen sowohl finite Vollverben als auch finite Hilfs- und Modalverben ein. Es ist sicherlich zu einfach gedacht, dass sich diese Verbgruppen bezüglich der Dependenzdistanz zum jeweiligen Subjekt gleich verhalten sollten. Betrachten wir die Beispiele 34 und 35, wird klar, dass es eben Hilfs- und Modalverben sind, die oft am Ende (oder im hinteren Bereich) eines Satzes stehen.

- (34) Das Deutsche Institut für Jugendhilfe und Familienrecht hält es für gerechtfertigt, dass der Gesetzgeber der Mutter folgende Möglichkeit eingeräumt **hat**<sub>VAFIN</sub>: Sie kann den Vater in die Mitverantwortung für das gemeinsame Kind nehmen, wenn dieser zu deren Übernahme bereit **ist**<sub>VAFIN</sub>.
- (35) Nach diesen Grundsätzen ergibt sich keine Zuständigkeit des Bundesgesetzgebers aus Artikel 125a Absatz 2 Grundgesetz, die sich nach Artikel 125a Absatz 2 Satz 3 Grundgesetz auch auf Bundesrecht erstreckt, das vor dem 15. November 1994 erlassen worden **ist**<sub>VAFIN</sub> aber heute auf der Grundlage von Artikel 75 Absatz 2 Grundgesetz nicht mehr erlassen werden **könnte**<sub>VMFIN</sub>.

In Beispiel 34 stehen die beiden finiten Hilfsverben 'hat' und 'ist' jeweils am Ende ihres Satzes (der Doppelpunkt zählt als Satzgrenze, wenn er vollständige Sätze trennt). In Beispiel 35 befindet sich das finite Hilfsverb 'ist' im hinteren Bereich des langen Satzes (*rel.posist* = 0,70). Hilfsverben gelten zwar im Satz als die finiten Verben, der semantische Gehalt des Prädikats geht jedoch vom abhängigen Vollverb aus. Als eine Eigenschaft von Hilfsverben wird daher auch oft eine „Desemantisierung“ angeführt (vgl. z.B. Remberger 2006, S. 12f.). Für das Hilfsverb 'hat' aus Beispiel 34 ist das zugehörige Vollverb 'eingeräumt'. Das Modalverb 'könnte' in Beispiel 35 steht innerhalb des Verbclusters 'erlassen werden könnte'. Wenn man nun annimmt, dass die Distanz zum Subjekt nicht am Hilfsverb relevant wird, sondern bereits am entsprechenden Vollverb, würde man auf Hilfsverben keinen messbaren Einfluss der Dependenzdistanz erwarten. Wenn man ferner nachweisen könnte, dass Hilfsverben in den hinteren Regionen von Sätzen häufiger sind als in vorderen Regionen, könnte das eine plausible Erklärung für den oben dargestellten Interaktionseffekt sein.

Um genau diese Ungleichverteilung von finiten Hilfs-, Modal- und Vollverben bezüglich der Satzregion zu untersuchen, wurde jedem finiten Verb die Information hinzugefügt, ob es vorne (*rel.pos* ≤ 0,7) oder hinten (*rel.pos* > 0,7) in seinem Satz steht. Kreuztabelliert man nun diese Information mit der Verbart, lässt sich auf Gleichverteilung der Verbarten über die Satzregionen testen. Abbildung 7.27 zeigt den Mosaikplot dieser Kontingenztabelle für den Blickbewegungsdatensatz des FLRCRef-Korpus.

Wie der assoziierte  $\chi^2$ -Test nahelegt, ist die Gleichverteilungsannahme deutlich verletzt. Die Vorzeichen der Residuen, die in Abbildung 7.27 über die Vorzeichen in den Zellen und farblich kodiert sind, zeigen das oben beschriebene Muster: In vorderen Regionen von Sätzen sind finite Vollverben deutlich überrepräsentiert, Hilfs- und Modalverben sind dort unterrepräsentiert. Das entgegengesetzte Muster zeigt sich für hintere Satzregionen: Hier sind finite Vollverben deutlich unterrepräsentiert. Hinten in Sätzen sind finite Hilfs- und Modalverben mehr als doppelt so häufig wie finite Vollverben (716 zu 306 Datenpunkte im Blickbewegungsdatensatz; Faktor 2,3). Vorne in Sätzen (in den vorderen 70%) halten sich die Verbtypen nahezu die Waage (826 zu 781; Faktor 1,1). Die oben beschriebene Interaktion ist also höchstwahrscheinlich auf diese Ungleichverteilung zurückzuführen. Am finiten **Hilfsverb** selbst scheint die Dependenzdistanz gar nicht oder in viel geringerem Ausmaß relevant für die Lesezeit zu sein. Auch Konieczny (2000) variiert nicht die Distanz zwischen dem Argument und dem Hilfsverb, sondern jeweils zum Vollverb.

Stimulussatz 36 ist jener mit der längsten Distanz zwischen dem relevanten Vollverb ‘gelegt’ und dem Subjekt ‘er’ sowie dem Akkusativobjekt ‘die Rose’.

- (36) Er hat die Rose, die auffällig schön gewachsen und ganz besonders farbenprächtig war, auf den kleinen runden Tisch **gelegt**, und ...

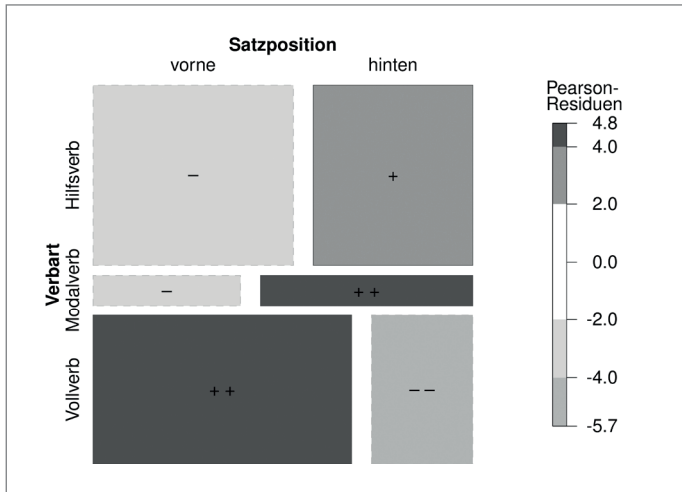


Abb. 7.27: Mosaikplot für die Verteilung verschiedener Verbtypen über Satzregionen. Die Richtung der Abweichung (nach oben/unten) ist über das Vorzeichen kodiert. Doppelte Vorzeichen kodieren signifikantere Abweichungen. Für die dem Plot zugrunde liegende Kontingenztabelle gilt:  $\chi^2(2) = 106,9; p = \min(p)$ .

Konieczny (2000) konnte zeigen, dass in dieser Bedingung das den Teilsatz abschließende Vollverb (hier: ‘gelegt’) im Vergleich zu den anderen Versuchsbedingungen am schnellsten gelesen wurde. Dieser Effekt konnte in der vorliegenden Arbeit insbesondere für finite Verben weiter vorne im Satz nachgewiesen werden – mutmaßlich deshalb, weil in den hier untersuchten Texten in hinteren Satzregionen eher Hilfs- und Modalverben stehen. Eine Nachanalyse zeigt, dass der Effekt der Dependenzdistanz für finite Vollverben in der Tat ein anderer ist als für Hilfsverben. Abbildung 7.28 zeigt die Interaktion zwischen den Prädiktoren Dependenzdistanz und der Verbart ( $\beta = -0,019; SE = 0,01; t = -1,967; p_{MCMC} = 0,05$ ). Zwar ist der Effekt offensichtlich mit großem Rauschen behaftet, trotzdem ist die Interaktion signifikant, die eine andere Wirkungsrichtung der Dependenzdistanz auf finite Vollverben als für Hilfs- oder Modalverben nahelegt.



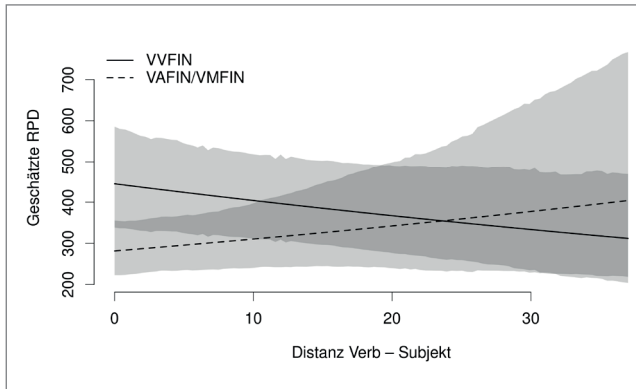


Abb. 7.28: Interaktion Dependenzdistanz  $\times$  Verbtyp für *regression path durations*. 95%-Konfidenzintervalle sind über eine MCMC-Simulation ermittelt.

Wir wollen die Analysen nun näher an die von Konieczny (2000) untersuchten Phänomene rücken. In einem Lesekorpus können zwar die intervenierenden Strukturen nicht so genau kontrolliert werden, wie es in der oben zitierten Studie der Fall war. Dafür können mehr Typen von sprachlichen Strukturen beachtet werden. Die Distanzinformation, die oben für finite Hilfs- und Modalverben analysiert wurde, kann auch für die entsprechenden Vollverben<sup>150</sup> annotiert werden. Diese Vollverben sind – wie oben kurz ausgeführt und über Beispiel 36 illustriert – jene sprachlichen Strukturen, die die eigentliche semantische Information tragen. Die Dependenzdistanz zum Subjekt wird dann nicht vom Hilfsverb aus gemessen, sondern vom assoziierten Vollverb aus. In Beispiel 34 wäre das nicht die Distanz von Hilfsverb ‘hat’, sondern jene vom Partizip ‘eingeräumt’. In die Analyse geht dann das Vollverb ein, nicht das finite Hilfsverb. Nehmen wir nun diese verbalen Elemente mit in die Vorhersage der *regression path durations* auf, schließen dafür aber die finiten Hilfs- und Modalverben aus, bleibt die Interaktion Dependenzdistanz relative Position im Satz trotzdem bestehen ( $\beta = 0,053$ ;  $SE = 0,014$ ;  $t = 3,90$ ). Auf eine erneute Abbildung kann an dieser Stelle verzichtet werden, weil die Interaktion dasselbe Muster aufweist wie die in Abbildung 7.26 zeigte.

Die Hinweise auf Antilokalitätseffekte, insbesondere in den vorderen Bereichen von Sätzen, sind im Laufe der hier gezeigten Analysen recht deutlich. Es wäre nun natürlich noch wünschenswert, zwischen verschiedenen intervenie-

<sup>150</sup> Wann immer in den folgenden Analysen von Vollverben oder Partizipien die Rede ist, meine ich damit die von den Hilfsverben abhängigen verbalen Strukturen. Dies müssen nicht immer Partizipien sein (sind es aber in der Mehrzahl der Fälle). Es können auch Adjektive (‘ist ... vereinbar’) oder Infinitive (‘müsste ... *vorsehen*’) sein.

renden Strukturen zu unterscheiden – zwischen Subjekt und Verb (bzw. Partizip) können sehr unterschiedliche Arten sprachlichen Materials liegen. Um diese intervenierenden Strukturen weiter aufzuschlüsseln, fehlt es im FLRC allerdings leider an Material. So gingen in die zuletzt gezeigten Analysen nur noch 2.062 Datenpunkte ein. Diese Effekte sollten anhand größerer Lesekorpora verifiziert und die Analysen dementsprechend erweitert werden.

Zuletzt soll nun der Bogen zurück zu den Reformulierungen geschlagen werden. Die Analysen bezüglich der Dependenzdistanz, die bisher präsentiert wurden, basieren zwar alle auf dem Reformulierungskorpus, die verschiedenen Stufen der Reformulierungen wurden bisher jedoch noch nicht beachtet. In Abschnitt 5.2.1 (siehe Abb. 5.6) wurde gezeigt, dass die durchschnittliche Dependenzdistanz zwischen Verb und Subjekt über die Reformulierungsstufen hinweg sinkt. Dieser Effekt ist auf die syntaktisch komplexen Texten (Komplexitätstyp *Synt.*) zurückzuführen. Das in Abbildung 5.6 (rechtes Diagramm, rechte Balkengruppe) gezeigte Muster bleibt auch bestehen, wenn die Datenmenge auf jene Verben eingeschränkt wird, die nach dem Subjekt stehen und man Hilfs- und Modalverben ausschließt und dafür die bedeutungstragenden Partizipien beachtet (analog zu den zuletzt präsentierten Analysen). Zunächst: Die Dependenzdistanz zu verringern, muss im Lichte der oben präsentierten Analysen nicht unbedingt als Vorteil gewertet werden, denn es wurde gezeigt, dass im vorderen Bereich von Sätzen eine höhere Dependenzdistanz zu schnelleren Verarbeitungszeiten auf den Verben und Partizipien zu führen scheint. Reformuliert man die Texte so, dass nur noch sehr geringe Dependenzdistanzen auftreten, verliert man diesen Antilokalitätseffekt. Das soll im Umkehrschluss aber nicht bedeuten, dass man Texte absichtlich so formulieren sollte, dass die Dependenzdistanzen besonders hoch sind, nur damit Vollverben schneller verarbeitet werden können. Vielmehr wäre es interessant, ob es gelingen kann, einen Text so zu formulieren, dass, wenn längere Subjekt-Verb-Distanzen auftreten, dieser Antilokalitätseffekt von den LeserInnen besser genutzt werden kann. Ist dies mit den Reformulierungen gelungen? Übersetzt in einen statistisch messbaren Effekt suchen wir nach einer Interaktion zwischen Dependenzdistanz und Reformulierungsversion. Wäre diese Interaktion signifikant, würde dies nahelegen, dass sich die Distanz zwischen Subjekt und Verb je nach Reformulierungsversion unterschiedlich auswirkt.

Tabelle 7.31 zeigt das entsprechende statistische Modell für die Vorhersage von *regression path durations*. Die Modellparameter für die Interaktion sind in den letzten zwei Zeilen abgetragen. Es zeigt sich, dass der Effekt der Dependenzdistanz in der moderaten Reformulierungsversion marginal signifikant moduliert wird und in den starken Reformulierungen hoch signifikant. Die Vorzeichen sind jeweils negativ. Die Interaktion ist in Abbildung 7.29 abgetragen.

In der Tat scheint die Dependenzdistanz also über Reformulierungsstufen hinweg einen unterschiedlichen Einfluss zu haben. Auch wenn Abbildung 7.29 einen verlängernden Einfluss der Dependenzdistanz in den Originaltexten nahelegt, kann dies statistisch nicht bestätigt werden. Zwar ist der Effekt der Dependenzdistanz in Originaltexten relativ groß (362 ms), aber nicht signifikant ( $t < 1$ ). Anders in reformulierten Texten: Fasst man moderate und starke Reformulierungen zusammen (was aufgrund Abbildung 7.29 sinnvoll erscheint), zeigt sich ein signifikant beschleunigender Einfluss der Dependenzdistanz auf die RPDs:  $\beta = -0,011$ ;  $SE = 0,005$ ;  $t = -2,45$ ;  $p_{\text{MCMC}} = 0,013$ . Die Effektgrößen liegen bei 140 ms für moderate und 177 ms für starke Reformulierungen. Offenbar ist es durch die Reformulierungen der Texte tatsächlich gelungen, dass LeserInnen den Antilokalitätseffekt besser ausnutzen konnten. „Ausnutzen“ klingt dabei natürlich nach einem sehr bewussten Vorgang – so ist dies aber nicht zu verstehen. Viel mehr wurde die Satzkomplexität offenbar auf vielen Ebenen so reduziert, dass die LeserInnen beim Herausbilden von Erwartungen, welches Verb im Satz verwendet wird, besser wurden. Das ist eine ziemlich vage Aussage, denn es ist nicht letztendlich klar, auf welcher Ebene die Satzkomplexität so vereinfacht wurde, dass die Antilokalitätseffekte „durchschlagen“ konnten. Es kann hier zwar keine letztendliche Lösung angeboten werden, auf welche Art der Reformulierung dieser Effekt zurückzuführen ist, ein Vorschlag soll aber gemacht werden.

In Abschnitt 7.3 konnte gezeigt werden, dass Verben unter Umständen länger gelesen werden, wenn mehr mögliche Referenten vor ihnen im Satz stehen (für eine detailliertere Darstellung der Effekte sei auf die entsprechenden Textteile verwiesen). Dies wird von den bereits vorgestellten Theorien zum *cue-based parsing* (vgl. Lewis/Vasishth 2005; Van Dyke/McElree 2006) vorhergesagt. Bringen wir dies nun in Verbindung mit den eben vorgestellten Analysen. Nehmen wir einmal an, dass viele mögliche Referenten vor dem Verb dazu führen, dass die Verarbeitung des Verbs erschwert wird. Dann könnte der menschliche Satzverarbeitungsmechanismus so sehr damit beschäftigt sein, die korrekten Argumente für das Verb zu finden und sie anzubinden, dass die gezeigten Antilokalitätseffekte nicht mehr wirken könnten. Die Anzahl der präverbalen Referenten sinkt in reformulierten Texten deutlich.<sup>151</sup> Das ist nicht überraschend, denn insbesondere in syntaktisch komplexen Texten werden die Sätze deutlich kürzer – es ist daher einfach viel weniger Textmaterial vorhanden, in dem interferierende Referenten vor dem Verb stehen

<sup>151</sup> Paarweise *t*-Tests mit Signifikanzkorrektur nach Holm ergeben höchst signifikante Kontraste zwischen Originalen und moderaten sowie starken Reformulierungen. Zwischen den beiden Reformulierungsstufen kann kein Unterschied nachgewiesen werden.

könnten. Um nun zu überprüfen, ob die Anzahl präverbaler Referenten tatsächlich einen Einfluss auf den Dependenzdistanz-Effekt hat, bilden wir vier Gruppen von Verben (in Klammern ist die jeweilige Anzahl an Messpunkten im Blickbewegungsdatensatz angegeben, die Gruppen sind über Median-splits hergestellt): (1) Verben mit niedriger Dependenzdistanz und wenigen präverbalen Referenten ( $n = 585$ ), (2) Verben mit niedriger Dependenzdistanz und vielen präverbalen Referenten ( $n = 298$ ), (3) Verben mit hoher Dependenzdistanz und wenig präverbalen Referenten ( $n = 109$ ) sowie (4) Verben mit hoher Dependenzdistanz und vielen präverbalen Referenten ( $n = 415$ ). Die jeweiligen Kombinationen sind naturgemäß ungleich besetzt. Verben, die eine hohe Distanz zu ihrem Subjekt aufweisen, haben typischerweise auch viele präverbale Referenten, die relevante Drittvariable ist hier die Länge des Satzes, in dem das Verb steht. Berechnen wir nun die Interaktion dieser beiden Gruppierungsvariablen, ergibt sich das in Abbildung 7.30 gezeigte Bild.

Groups	Name	Variance	Std.Dev.			
vtn	trial	0.000000	0.00000			
vtn	(Intercept)	0.034188	0.18490			
item.id	(Intercept)	0.072051	0.26842			
Residual		0.947952	0.97363			
		<b>Estimate</b>	<b>Std. Error</b>	<b>t value</b>	<b>p<sub>MCMC</sub></b>	
(Intercept)		6.0234054	0.2381792	25.289	0.0001	***
w.len		-0.0459209	0.0195336	-2.351	0.0208	*
res.freq.dlex.lm		0.1753914	0.0622227	2.819	0.0080	**
res.famil.lm		0.2460511	0.0749595	3.282	0.0016	**
trial		-0.0119078	0.0039190	-3.038	0.0042	**
text.row		0.0382493	0.0181299	2.110	0.0406	*
first.word		0.3747851	0.2111351	1.775	0.0706	.
last.word		0.3565575	0.0703330	5.070	0.0001	***
rel.pos		0.6591998	0.1322647	4.984	0.0001	***
dep.dist		0.0002802	0.0095406	0.029	0.9812	
version moderat		0.0953190	0.1291437	0.738	0.4460	
version stark		0.2443904	0.1350027	1.810	0.0698	.
dep.dist:version moderat		-0.0247951	0.0139320	-1.780	0.0778	.
dep.dist:version stark		-0.0594606	0.0192483	-3.089	0.0030	**

Tab. 7.31: Gemischtes Modell für *regression path durations* auf finiten Vollverben und Partizipien. Von Interesse sind insbesondere die Faktoren Dependenzdistanz (*dep.dist*) und Reformulierungsversion (*version*) sowie die Interaktion der beiden Variablen.

Der Haupteffekt der Dependenzdistanz-Gruppe ist signifikant ( $\beta = -0,536$ ;  $SE = 0,196$ ;  $t = -2,731$ ;  $p_{\text{MCMC}} = 0,007$ ), jener für die Gruppe der Anzahl präverbaler Referenten nicht ( $\beta = -0,217$ ;  $SE = 0,137$ ;  $t = -1,584$ ;  $p_{\text{MCMC}} = 0,129$ ). Leider ist auch der Interaktionseffekt nicht signifikant, den wir uns aufgrund der oben ausgeführten Interpretation versprochen hätten ( $\beta = 0,325$ ;  $SE = 0,24$ ;  $t = 1,351$ ;  $p_{\text{MCMC}} = 0,173$ ). Die Effektgröße für die Gruppe mit wenigen präverbalen Referenten ist zwar deutlich höher (158 ms gegenüber 58 ms für die Gruppe mit vielen präverbalen Referenten), das Rauschen ist hier aber ebenfalls recht hoch. Eventuell könnte dieser Effekt in einem größeren Lesekorpus nachgewiesen werden, denn Abbildung 7.30 legt eine Interaktion nahe. Auch eine experimentelle Studie, die der Interaktion von (anti-)lokalitätsbasierten Effekten in Kombination mit Theorien zum *cue-based parsing* auf den Grund geht, könnte interessant sein. In einem experimentellen Design hätte man es nicht mit ungleich besetzten Zellen zu tun, die sich hier aus dem natürlichen Sprachmaterial ergeben.

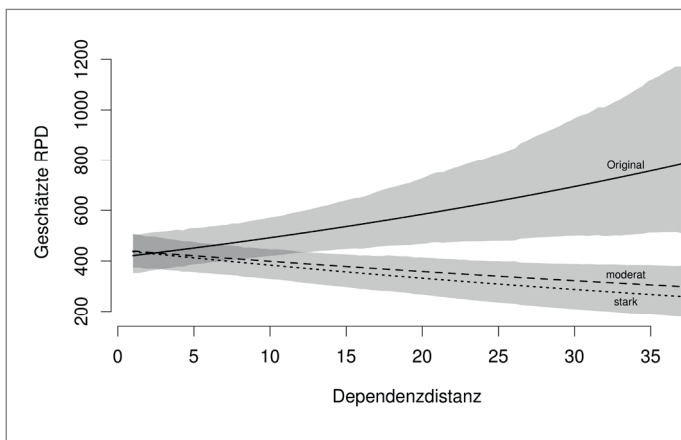


Abb. 7.29: Interaktion Dependenzdistanz  $\times$  Komplexitätstyp für *regression path durations*, keine Hilfs- und Modalverben, mit Partizipien. 95%-Konfidenzintervalle sind über eine MCMC-Simulation ermittelt. Der Übersichtlichkeit halber sind nur die Konfidenzintervalle für die Originale sowie die starken Reformulierungen abgetragen.

Die letzte Analyse zeigte, wie sich Effekte auf unterschiedlichen linguistischen Ebenen integrieren lassen, um Einsichten über die Interaktion sprachlicher Eigenschaften zu gewinnen. Auch wenn die zuletzt gezeigten Ergebnisse vorsichtig interpretiert werden sollten, da oft eine größere Datenbasis wünschenswert wäre, scheint eine solche Integration überaus vielversprechend für die Verständlichkeitsforschung.

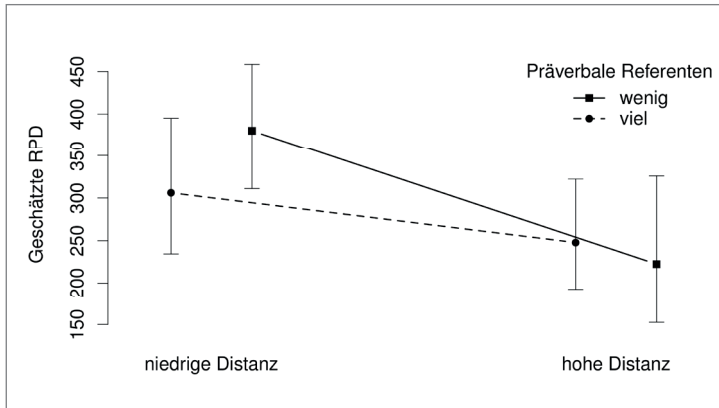


Abb. 7.30: Interaktion der Gruppenvariablen niedrige/hohe Dependenzdistanz × wenige/viele präverbale Referenten für *regression path durations*.





## 8. Implikationen

Das *Freiburg Legalese Reading Corpus* (FLRC) ist das einzige Lesekorpus, das neben natürlichem Textmaterial auch reformulierte Textteile enthält. Somit vereint das FLRC Charakteristika anderer Lesekorpora mit klassisch-experimentellen Ansätzen, wo Stimulusmaterial ganz gezielt variiert wird, um den Einfluss dieser Variation auf die Verarbeitung zu messen. Ausgehend von Texten des Komplexitätstyps „Nominalisierungen“ konnte gezeigt werden, dass Nominalisierungen in der Tat langsamer verarbeitet und seltener übersprungen werden als alle anderen Nomen. Es konnte außerdem gezeigt werden, dass dieser Effekt relativ stark ist, wenn man ihn mit dem Effekt für andere Arten von Wörtern, wie z.B. finiten Verben, vergleicht. Eine Befürchtung war, dass sich die Verarbeitungskomplexität im Zuge der Reformulierungen verschiebt: weg von den Nominalisierungen hin zu den verbalen Strukturen, die an Stelle der Nominalisierungen eingeführt werden.

Das ist – bei aller Vorsicht, die man bei der Interpretation eines Nicht-Effekts walten lassen muss – nicht der Fall. In den reformulierten Texten wurden zwar alle Nomen schneller gelesen, verbale Strukturen aber nicht signifikant langsamer. Wir können also davon ausgehen, dass eine Umformung von Nominalisierungen in verbale Strukturen in der Tat einen Vorteil für die Verarbeitung der Texte bringt.

Über die Extraktion der Schätzwerte für die entsprechenden Modelle wurde über ein Rangordnungsverfahren gezeigt, welche Wörter jene sind, die am langsamsten gelesen wurden (als Maß diente hier die *first-pass reading time*). 15 Wörter befanden sich unter diesen langsamsten 50 Wortlesedurchgängen, davon 13 Komposita, fünf Wörter, die eindeutig der rechtssprachlichen Fachterminologie zuzuordnen sind (sowie drei Zweifelsfälle) und sechs Nominalisierungen (sowie vier Zweifelsfälle). In der Gesamtschau sind verlangsamende Effekte für Nominalisierungen also recht deutlich und die Umformulierung in verbale Konstruktionen scheint keine eindeutig messbaren Nachteile mit sich zu bringen.

Ähnlich effektiv scheinen die Reformulierungen komplexer Nominalphrasen gewesen zu sein: Die *first-pass reading times* für Nominalphrasen sinken über die Reformulierungsstufen hinweg. Für syntaktisch komplexe Texte konnte gut gezeigt werden, wie die abfallenden Lesezeiten parallel verlaufen zur sinkenden Zahl der Kinderknoten unter Nominalphrasen. Für die Texte des Komplexitätstyps NP konnte dieser direkte Zusammenhang zwischen linguistischer Struktur und Verarbeitungszeit jedoch nicht gezeigt werden.

Hier identifizierte ich die Anzahl rechtsangeschlossener Genitivattribute (die auch selbst Nominalphrasen sind) als linguistischen Einflussfaktor. Diese Konstruktionen werden in den Originaltexten teils exzessiv eingesetzt und ineinander verschachtelt. In den Reformulierungen wurde dieses Muster aufgebrochen, was in der Folge zu sinkenden Lesezeiten auf Nominalphrasen insgesamt führte. Dieser Effekt kann dadurch erklärt werden, dass die Lesezeit von Genitivattributen generell über jener aller anderen Nominalphrasen liegt. Formuliert man die Texte so um, dass diese Genitivattribute wegfallen, führt dies zwangsläufig zu schnelleren Lesezeiten auf Nominalphrasen. Zusätzlich konnte erklärt werden, warum auch für Texte des Komplexitätstyps NP die Anzahl an Nominalisierungen über die Reformulierungsstufen hinweg sinkt: Auch hier wurden Nominalisierungen in verbale Strukturen umgewandelt, was zur Einführung von subordinierten Teilsätzen führte. Auch dies trug zu den schnelleren Lesezeiten auf Nominalphrasen über die Reformulierungsstufen hinweg bei.

In syntaktisch komplexen Texten wurden während der Reformulierungen besonders lange Sätze getrennt. Auf linguistischer Ebene führte dies zu einer Reduktion der syntaktischen Komplexität, wie über mehrere entsprechende Maßzahlen (5.2.1) gezeigt werden konnte. Trennt man Sätze, müssen trotzdem relevante Referenzen aufrecht erhalten werden. Ich habe gezeigt, dass die referenzielle Komplexität auf linguistischer Ebene in der Tat über die Reformulierungsversionen hinweg zunimmt. Unter anderem lässt sich das an vermehrten Referenzen über die Satzgrenzen hinweg festmachen. Für die Verarbeitung der Texte hatte dies zwei Konsequenzen: 1) Der beschleunigende Effekt der Einbettungstiefe ist in reformulierten Texten stärker als in den Originaltexten. Interpretiert habe ich diesen Effekt so, dass Relevanzhinweise und syntaktische Constraints in den reformulierten Texten von den LeserInnen besser genutzt werden konnten. 2) Für die Verarbeitung von Proformen und referierenden Nominalphrasen konnten keine Effekte in syntaktisch komplexen Texten festgestellt werden. Entsprechende Effekte zeigten sich aber für Texte mit vielen Nominalisierungen und komplexen Nominalphrasen. In diesen Texten führte höhere referenzielle Komplexität (operationalisiert durch eine größere Anzahl an referenziellen Ausdrücken) zu höheren Verarbeitungszeiten auf referenziellen Ausdrücken.

In den Analysen zu Lokalitäts- oder Dependenzdistanzeffekten rückte wieder die Beziehung zwischen Verben und deren Argumenten ins Blickfeld. Diese Beziehung spielte bereits in den Analysen auf der syntaktischen Ebene eine Rolle, wurde hier aber auf die Abfolge und Distanz zwischen Subjekt und Verb fokussiert. Zusammenfassend muss für diese Analysen betont werden, dass sich in keiner Analyse Lokalitätseffekte im Sinne der *Dependency Locality*

*Theory* (DLT, Gibson 1998, 2000) ergeben haben. An einigen Stellen fanden sich allerdings Antilokalitätseffekte, wie sie auch Konecny (2000) für das Deutsche zeigen konnte. Auffällig ist, dass weder für *first-pass* noch für *total reading times* Effekte der Dependenzdistanz gefunden werden konnten. Der einzige stabile Effekt ist jener der Verb-Subjekt-Abfolge. Hier konnte an mehreren Stellen gezeigt werden, dass die kanonische Subjekt-Verb-Abfolge eines deutschen Hauptsatzes (Subjekt vor finitem Verb) dazu führt, dass das Verb schneller gelesen wird, weniger regressive Sakkaden initiiert werden und kürzere Regressionspfade hat. Für *regression path durations* und die Wahrscheinlichkeit von regressiven Sakkaden zeigte ich Effekte der Dependenzdistanz: Je weiter das finite Verb vom Subjekt entfernt ist, desto seltener und kürzer wird vom Verb in den vorherigen Text zurückgesprungen – das gilt allerdings hauptsächlich für den Anfang und die Mitte des Satzes.

Da viele der untersuchten finiten Verben Hilfs- und Modalverben waren, habe ich in der Folge die Analyse auf die zugehörigen Vollverben erweitert, was die Analyse näher an die Studie von Konecny (2000) gerückt hat. Auch dort waren die relevanten Elemente die nicht-finiten Vollverben. Der oben bereits angesprochene Effekt der Dependenzdistanz bestätigte sich hier.

Zuletzt präsentierte ich eine Analyse zu den Reformulierungsversionen. Es konnte gezeigt werden, dass der beschleunigende Effekt der Dependenzdistanz nur in reformulierten Texten nachzuweisen ist. Dieser Interaktionseffekt ist sehr robust. Es ist davon auszugehen, dass dies mit der sinkenden Satzkomplexität zusammenhängt. Die Interpretation lautet: Nicht immer sind Antilokalitätseffekte (das heißt hier: beschleunigende Effekte der Dependenzdistanz zwischen Subjekt und Verb) zu beobachten – es müssen begünstigende Faktoren hinzutreten. Ein Vorschlag in der vorliegenden Arbeit ist, dass einer dieser begünstigenden Faktoren die geringere Anzahl präverbaler Referenten sein könnte. Damit lässt sich der Antilokalitätseffekt allerdings nur numerisch in Zusammenhang bringen: Das Muster lässt darauf schließen, dass in der Tat die Dependenzdistanzeffekte in jenen Fällen stark sind, wo wenig präverbale Referenten die Anbindung stören. Die Datenbasis der vorliegenden Arbeit genügt allerdings nicht, diesen angenommenen Effekt statistisch zu untermauern. Weitere Forschung ist hier notwendig. Beispielsweise könnte eine experimentelle Untersuchung die in natürlicher Sprache auftretenden Ungleichverteilungen aufbrechen und so den Effekt klarer zeigen.

Dies schließt die Zusammenfassung der Konsequenzen der Reformulierungen ab und bringt uns zur allgemeinen Zusammenfassung und Rückschau der in dieser Arbeit vorgestellten Ergebnisse. Effekte auf der lexikalischen Ebene, wie sie bereits in anderen Blickbewegungskorpora und experimentel-

len Untersuchungen gezeigt wurden, replizierte und erweiterte ich anhand des *Freiburg Legalese Reading Corpus*. Insbesondere den Zusammenhang zwischen Wortfrequenz und -familiarität konnte ich genauer aufschlüsseln. Alternative Maße wie beispielsweise die Differenz zwischen absoluter Familiarität und absoluter Frequenz besitzen Vorhersagekraft für Wort-Lesezeiten: Ich interpretierte dieses Differenzmaß als Konkurrenzmaß, das die kumulierte Frequenz von Konkurrenten um den lexikalischen Abruf operationalisiert. Sowohl experimentelle Untersuchungen als auch Analysen anhand von Lesekorpora können von dieser Vorhersagekraft profitieren. Dies gilt für jene Fälle, wo diese lexikalischen Eigenschaften selbst im Zentrum des Interesses stehen, aber auch für die Fälle in denen das Konkurrenzmaß als Kovariante dient.

Auf syntaktischer Ebene wurden ebenfalls Effekte repliziert, die bereits in anderen Lesekorpora gefunden wurden. Dies gilt insbesondere für den Effekt der Einbettungstiefe von Wörtern und Phrasen. Ich überprüfte die These, dass die Einbettungstiefe von LeserInnen als Hinweis auf die Relevanz der Information genutzt wird. In der Tat werden insbesondere Adjektive, Adverbien und Nomen in modifizierenden Strukturen schneller gelesen, wenn sie tiefer im Satz eingebettet sind. Auch auf Phrasenebene zeigte sich ein entsprechender Effekt: Adjektivphrasen, Präpositionalphrasen und Adverbialphrasen werden ebenfalls schneller gelesen, wenn sie tiefer eingebettet sind – im Gegensatz zu Verbalphrasen, Sätzen und Nominalphrasen. LeserInnen scheinen also die Einbettungstiefe von modifizierenden Strukturen als Hinweis auf deren Relevanz zu verstehen. Aus verarbeitungsökonomischen Gründen widmen sie dann jenen Strukturen, die als weniger zentral für die Aussage des Satzes angesehen werden, weniger Aufmerksamkeit.

Ein Feld der Psycholinguistik, das bisher hauptsächlich anhand kontrollierter Experimente untersucht wurde, ist die Diskursverarbeitung. Insbesondere die *Informational Load Hypothesis* (Almor 1999), die ausführlich in Abschnitt 2.2 behandelt wurde, diente hier zur Hypothesengenerierung und der Definition von Prädiktoren. Im FLRC ist ein Rechtfertigungsprinzip von referenziellen Ausdrücken zu beobachten, wie es von einigen Theorien (vgl. auch Ariel 1990) der Anaphernauflösung und verwendung postuliert wird. Das gilt insbesondere für referierende Nominalphrasen, die sich auf einer spezifischeren Taxonomie-Ebene befinden als ihr Antezedens. Die vorgestellten Analysen zeigen, dass LeserInnen erwarten, dass mit spezifischeren Nominalphrasen neue Information zur Diskursrepräsentation hinzugefügt wird. Ist dies nicht der Fall, wird das mit längeren Lesezeiten „bestraft“.

## 8.1 Implikationen für die Grundlagenforschung

Für die psycholinguistische Grundlagenforschung zum Textlesen kann die vorliegende Arbeit insofern wertvolle Hinweise geben, als einige Phänomene, die bereits in Blickbewegungskorpora gezeigt wurden, anhand eines deutschen Lesekorpus mit natürlichem Sprachmaterial repliziert und an manchen Stellen erweitert wurden. Das gilt insbesondere für die meisten Effekte, die auf der lexikalischen Ebene gezeigt wurden. Die Beziehung zwischen der Familiarität und Frequenz eines Lexems wurde dahingehend erweitert, dass Familiarität als eine Art Konkurrenzmaß verstanden werden kann, das über die Effekte lexikalischer Nachbarn hinauszugehen scheint. Auch die Extrahierung bestimmter Darstellungsfaktoren (Zeilenanfang und -ende) und ihre Effekte auf das Leseverhalten sind interessant für weitere Lesestudien. Oftmals wurden Stimuli so konstruiert, dass Sätze gar nicht oder nur an irrelevanten Stellen in die nächste Zeile umbrechen. Hier wurde gezeigt, dass diese Darstellungsfaktoren stabile Einflüsse auf Lesezeitmaße haben, dass sie bis zu einem gewissen Grad aber auch kontrolliert werden können, wenn sie als Faktoren in statistische Modelle mit aufgenommen werden. Doch nicht nur bereits vorhandene Erkenntnisse aus Lesekorpora konnten teilweise im Rahmen der Arbeit repliziert werden, sondern auch Ergebnisse, die bisher nur anhand streng kontrollierter Experimentalstimuli nachgewiesen wurden. Dies erweitert die Relevanz von Phänomenen wie beispielsweise der dynamischen Anpassung der perzeptuellen Spanne, weil sie offenbar auch in einem ökologisch valideren Setting innerhalb natürlicher Sprachdaten gefunden werden können.

Für syntaktische Faktoren habe ich einige Operationalisierungen (wie beispielsweise das Öffnen einer syntaktischen Klammer) vorgeschlagen und evaluiert. Diese Operationalisierung ist relativ billig (gegeben eine syntaktische Annotation der Texte) und klärt spezifische Teile der Varianz in Blickbewegungen auf. Durch die Kombination verschiedener Ebenen der sprachlichen Annotation (Part-of-Speech und Phrasenstruktur) habe ich die in der obigen Zusammenfassung beschriebenen Relevanzeffekte gezeigt. Weitere Untersuchungen anhand von Blickbewegungskorpora sowie anhand speziell darauf ausgerichteter experimenteller Untersuchungen sollten diese Lesestrategie „widme unwichtigeren Dingen weniger Aufmerksamkeit“ evaluieren und näher beschreiben. Auch weitere Operationalisierungen von Relevanz auf sprachlicher Ebene können die psychologische Wirkungsweise dieses Effekts elaborieren. Auf syntaktischer Ebene habe ich außerdem zum ersten Mal eine phrasenstrukturbasierte Einteilung von *regions of interest* in Lesekorpora angewendet. Die Ergebnisse, die anhand dieser Regioneneinteilung gewon-

nen wurden, sprechen dafür, dass diese Methode der „dynamischen“ Regioneneinteilung, die von der Wortebene abstrahiert, eine vielversprechende Analysemöglichkeit in Lesekorpora ist.

Zum ersten Mal wurden außerdem referenzielle Faktoren anhand natürlichen Textmaterials untersucht. Hier habe ich eine Kategorisierung von referenziellen Ausdrücken vorgenommen, die psychologische Relevanz in ökologisch validen Studien besitzt. Das oben bereits angesprochene Rechtfertigungsprinzip von anaphorischen Ausdrücken konnte anhand dieser Kategorisierung gezeigt werden. Es ist also davon auszugehen, dass auch andere Effekte auf referenzieller Ebene in Lesekorpora nachgewiesen werden können.

## 8.2 Angewandte Sprachwissenschaft

Lesezeitmaße lassen sich nicht eins zu eins auf Verständlichkeit abbilden. Untersucht man das menschliche Textverstehen, müssen (zumindest) auf methodischer Ebene der Verstehensprozess und die Verstehensprodukte unterschieden werden. Lesezeitmaße werden dabei als beobachtbares Phänomen der Prozess-Seite gesehen. Wenn ein Text langsamer gelesen wird, heißt das nicht gleichzeitig, dass er weniger verständlich ist. Der einzige Hinweis in dieser Arbeit, der auf die Verbindung von Lesezeit und Verständniserfolg abzielt, legt eher eine positive Korrelation der beiden Konstrukte nahe: Je länger eine fragenrelevante Textstelle gelesen wird, desto eher wird die dazugehörige Frage korrekt beantwortet. Dieser Effekt stammt nicht daher, dass der gesamte Text aufmerksamer gelesen wurde, denn als Kontrollfaktor habe ich hier die Gesamtlesezeit aller nicht-fragenrelevanten Textabschnitte aufgenommen. Eine Gleichung, die besagt, dass schneller gelesene Texte besser verstanden werden, ist also zu vereinfachend. Vom Rücksprungverhalten der LeserInnen konnte kein Einfluss auf die Antwortperformanz festgestellt werden. Die in dieser Arbeit gefundenen Effekte lassen sich also hauptsächlich auf den Verstehensprozess beziehen und zeigen, wie sich linguistische Strukturen innerhalb der Texte auf den Leseprozess auswirken. Will man wissen, inwiefern sich beispielsweise Verzögerungen im Verstehensprozess auf Verstehensprodukte auswirken, muss man Fragen zu allen oder möglichst vielen Textabschnitten stellen. Die in dieser Arbeit verwendeten Verstehensfragen waren hauptsächlich darauf ausgerichtet, die LeserInnen aufmerksam zu halten. Sie eignen sich daher nur eingeschränkt für Rückschlüsse auf die mentalen Repräsentationen, die beim Lesen der Texte entstanden sind.

Welche Schlüsse lassen sich für die Verarbeitung juristischer Fachsprache ziehen? Hier sind insbesondere die Ergebnisse des Reformulierungskapitels interessant: Nominalisierungen sind in deutschen Fachsprachen ein weit ver-



breitetes sprachliches Phänomen. In der Rechtssprache ist insbesondere die mit dem Einsatz einer Nominalisierung verbundene Deagentivierung attraktiv. Dies gilt auch für die Wissenschaftssprache, wie Schwarze (2007, S. 196) anspricht. Verwendet man eine Nominalisierung, kann man vermeiden, dass das Agens genannt werden muss. Durch die deagentivierende Konstruktion der Nominalisierung in Kombination mit inhaltsleeren Verben ist es möglich, sprachlich keinen Wert auf die/den UrheberIn einer Handlung zu legen – genau das, was in Konstruktionen wie ‘Abschiebung durchführen’ oder ‘Duldung erteilen’ gewünscht ist. In der vorliegenden Arbeit habe ich gezeigt, dass sich Nominalisierungen deutlich verzögernd auf den Lesefluss auswirken. Die Nominalisierungen wurden in zwei Stufen reformuliert – die von den Nominalisierungen ausgedrückte Information musste aber trotzdem im Text enkodiert werden. Die verbalen Strukturen, die stattdessen eingeführt wurden, wurden nicht länger gelesen, und die Gesamtlesezeit der Texte verringerte sich signifikant. Der Reformulierungseffekt der Nominalisierungen schlägt also bis auf die Ebene des Gesamttextes durch. Der Verarbeitungsprozess von Nomen wird spürbar beschleunigt, während an keiner anderen Stelle Nachteile beim Verarbeitungsprozesses entstehen.

Als besonders komplexe (also den Verarbeitungsprozess verlangsamende) Nominalphrasen haben sich rechtsangeschlossene (postmodifizierende) Genitivattribute erwiesen. Deren Verwendung bietet sich an, wenn man möglichst viel Information in einer Phrase sprachlich kodieren will. Obwohl hier die Texte mit steigender Reformulierungsversion etwas länger wurden, stiegen die Gesamtlesezeiten der Texte nicht signifikant. Auch wenn die Reformulierungen der Texte mit komplexen Nominalphrasen nicht ganz von jenen mit vielen Nominalisierungen getrennt werden können, brachten die Reformulierungen der Genitivattribute für die Verarbeitung von Nominalphrasen einen Vorteil. An dieser Stelle wäre es sehr interessant, die Verarbeitung von Genitivattributen weiter zu untersuchen. Es bleibt noch weitgehend unklar, warum gerade diese in ihrer Verarbeitung so komplex sind. Es liegt nahe, dass es nicht an den Genitivattributen an sich liegt, sondern daran, dass teils exzessiv ineinander verschachtelt wurden.

Mit der Einschränkung, dass Lesezeiten nicht eins zu eins auf die Verständlichkeit eines Texts abgebildet werden können, lassen sich einige Formulierungsempfehlungen aussprechen. Nominalisierungen scheinen den Lesefluss deutlich zu verzögern. Gleichzeitig scheint die/der LeserIn nicht von den alternativ eingeführten Konstruktionen „bestraft“ zu werden. Einer Umformung von Nominalisierungen in die entsprechenden verbalen Konstruktionen steht also nichts im Wege. Eine Frage ist jedoch stets, inwieweit auf die Ausdrucksmacht von Nominalisierungen verzichtet werden kann. Will man beispiels-



weise aus 'die Abschiebung wird vollzogen' eine aktiv-verbale Struktur formulieren, benötigt man zwingend ein Agens der Abschiebe-Handlung ('die Behörde schiebt ... ab'). Wenn das an dieser Stelle nicht gewünscht ist, muss eventuell auf eine Passivkonstruktion zurückgegriffen werden, wobei das Patiens der Handlung genannt werden müsste ('der Ausländer wird abgeschoben'), wenn man nicht auf die Konstruktion 'es wird abgeschoben' zurückgreifen möchte. Einschränkend muss noch angemerkt werden, dass die Reformulierungen der Texte mit vielen Nominalisierungen nicht zu besserer Antwortperformanz der LeserInnen geführt hat. Ob so ein Effekt mit den vorliegenden Mitteln nicht nachweisbar war oder ob er tatsächlich nicht vorhanden ist, kann hier nicht abschließend entschieden werden. Immerhin führten die Reformulierungen aber nicht zu schlechteren Leistungen. Vergleichbare Antwortperformanz bei schnellerem Lesen scheint ein „guter Deal“ zu sein.

Dasselbe gilt für die Texte mit komplexen Nominalphrasen. Die Antwortperformanz bleibt über die Reformulierungsversionen hinweg auf dem gleichen Niveau. Nachdem die Anzahl der ineinander verschachtelten Genitivattribute, die hauptsächlich für die hohe Komplexität verantwortlich waren, reduziert wurde, verringerte das die Lesezeit auf Nominalphrasen deutlich. Die Genitivattribute wurden häufig in verbale Strukturen umgeformt. Obwohl die Texte dadurch im Schnitt etwas länger wurden, wurden sie nicht signifikant länger gelesen. Eine weitere Formulierungsempfehlung besagt somit, dass Genitivattribute möglichst nicht ineinander verschachtelt werden sollten. Entsprechende Inhalte können z.B. in Teilsätzen ausgedrückt werden. Das macht Texte zwar etwas länger – dieser Nachteil überwiegt aber nicht den Vorteil, den man aus der sparsamen Verwendung und Verschachtelung von Genitivattributen gewinnt.

Eine oft ausgesprochene Formulierungsempfehlung ist, möglichst kurze Sätze zu formulieren. In der vorliegenden Arbeit wurde dies ebenfalls in den Reformulierungen berücksichtigt. Das Teilen von Sätzen führt tatsächlich dazu, dass die syntaktische Einbettung von Wörtern besser dazu benutzt werden kann, in tief eingebetteten Satzregionen schneller zu lesen. Hier wurde dies dahingehend interpretiert, dass die LeserInnen die Relevanzsignale, die der Text bereitstellt, besser nutzen können, wenn die Sätze kürzer sind. Für jene Texte, in denen die Sätze gezielt geteilt wurden, konnten keine verlangsamenden oder beschleunigenden Effekte auf referenziellen Ausdrücken festgestellt werden. Moderat reformulierte Texte, in denen nur einige wenige Sätze geteilt wurden, wurden insgesamt schneller gelesen als die Originale. Ein ganz bestimmtes Intervall an Satzlängen, das als optimal gilt, kann daraus aber nicht abgeleitet werden. Die Ergebnisse für die syntaktisch komplexen Texte deuten nämlich auch darauf hin, dass eine zu starke „Zerstückelung“ der

Sätze uneffektiv ist. Dies zeigte sich unter anderem daran, dass innerhalb dieser Texte auch sehr viele regressive Sakkaden während des Leseprozesses auftraten. Beim Teilen von Sätzen gibt es also einen *trade-off*: Einerseits werden Texte mit kürzeren Sätzen schneller verarbeitet, andererseits können zu viele kurze Sätze dazu führen, dass die LeserInnen mehr Aufwand bei der Verarbeitung der referenziellen Struktur des Texts betreiben müssen.

Ich möchte kurz auf einen Punkt zurückgreifen, den ich bereits in der Einleitung der Arbeit betont habe: So erklären viele in der (Re-)Formulierung von Rechtstexten beteiligte Personen, dass das Ergebnis ihrer Arbeit umso besser würde, je früher im Ausarbeitungsprozess sie beteiligt sind. Im Lichte der vorherigen Ausführungen scheint dies plausibel. Will man z.B. die Länge der verwendeten Sätze optimieren, ist es nicht effektiv, einen bereits bestehenden Text, der viele lange Sätze enthält, zu „retten“. Vielmehr sollte man die Chance nutzen, frühzeitig im Formulierungsprozess gute Formulierungen zu finden. Eine von Anfang an ausgeglichene Verteilung der zu vermittelnden Information über mehrere Sätze sowie deren effektive referenzielle Verknüpfung muss frühzeitig sichergestellt werden. Effektiv heißt hierbei insbesondere, nicht zu viele überschneidende referenzielle Beziehungen zu benutzen und insbesondere ambige Referenzbeziehungen zu vermeiden.

### **8.3 Schluss**

In Zukunft gilt es anhand von Lesedaten auf natürlichem Textmaterial die Verbindung zwischen Blickbewegungsverhalten und der Verständlichkeit eines Textes konzeptuell noch stärker zu unterfüttern. Genaue und umfassende Operationalisierungen, die die mentale Repräsentation beim Lesen von Texten besser überprüfbar machen, sind erstrebenswert. Auch die Verknüpfung psycholinguistischer Grundlagenforschung mit Fragen der angewandten Linguistik ist gewinnbringend. Erst durch die interdisziplinäre Verknüpfung beider Ansätze können wir noch effektivere Werkzeuge entwickeln, um komplexe Stellen in Texten zu identifizieren und diese zu vermeiden.



## 9. Literatur

- Administrative Procedure Act (1986): In: Svensk författningssamling (SFS), S. 223.
- Almor, Amit (1999): Noun-phrase anaphora and focus: The informational load hypothesis. In: *Psychological Review* 106, 4, S. 748-765.
- Altmann, Erik M./Gray, Wayne D. (2002): Forgetting to remember: The functional relationship of decay and interference. In: *Psychological Science* 13, 1, S. 27-33.
- Anderson, Anne H./Garrod, Simon C./Sanford, Anthony J. (1983): The accessibility of pronominal antecedents as a function of episode shifts in narrative text. In: *The Quarterly Journal of Experimental Psychology. Section A: Human Experimental Psychology* 35, 3, S. 427-440.
- Anderson, John R. (1974): Retrieval of propositional information from long-term memory. In: *Cognitive Psychology* 6, S. 451-474.
- Anderson, John R. (1983): *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Anderson, John R./Bothell, Daniel/Douglass, Scott (2004): Eye movements do not reflect retrieval processes: Limits of the eye-mind hypothesis. In: *Psychological Science* 15, 4, S. 225-231.
- Anderson, John R./Bothell, Daniel/Byrne, Michael D./Douglass, Scott/Lebiere, Christian/Qin, Yulien (2004): An integrated theory of the mind. In: *Psychological Review* 111, 4, S. 1036-1060.
- Andrews, Mark/Vigliocco, Gabriella/Vinson, David (2009): Integrating experiential and distributional data to learn semantic representations. In: *Psychological Review* 116, 3, S. 463-498.
- Ariel, Mira (1988): Referring and accessibility. In: *Journal of Linguistics* 24, 1, S. 65-87.
- Ariel, Mira (1990): *Accessing noun-phrase antecedents*. London: Routledge.
- Ariel, Mira (2000). The development of person agreement markers: From pronouns to higher accessibility markers. In: Barlow, Michael/Kemmer, Suzannne (Hg.): *Usage-based models of language*. Stanford: Center for the Study of Language and Information, S. 197-260.
- Ariel, Mira (2001): Accessibility theory: An overview. In: Sanders, Ted J.M./Schliperoord, Joost/Spooren, Wilbert (Hg.): *Text representation*. Amsterdam u.a.: Benjamins, S. 29-87.
- Auer, Peter/Baßler, Harald (2007): *Der Stil der Wissenschaft*. In: Auer/Baßler (Hg.), S. 9-30.
- Auer, Peter/Baßler, Harald (Hg.) (2007): *Reden und Schreiben in der Wissenschaft*. Frankfurt a.M./New York: Campus.

- Ausubel, David P. (1960): The use of advance organizers in the learning and retention of meaningful verbal material. In: *Journal of Educational Psychology* 51, 5, S. 267-272.
- Baayen, R. Harald (2008): *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Baayen, R. Harald (2011): *languageR: Data sets and functions with „Analyzing linguistic data: A practical introduction to statistics“*. <http://CRAN.R-project.org/package=languageR> [= Handbuch zu R package version 1.4].
- Baayen, R. Harald/Piepenbrock, Richard/van Rijn, Hedderik (1993): *The CELEX lexical data base*. Philadelphia, PA: Linguistic Data Consortium. Universität Pennsylvania. [CD-ROM].
- Baddeley, Alan D. (1992): Working memory. In: *Science* 255, S. 556-559.
- Baker, Linda/Wagner, Jody L. (1987): Evaluating information for truthfulness: The effects of logical subordination. In: *Memory and Cognition* 15, 3, S. 247-255.
- Ballansat-Aebi, Suzanne (2000): „Attendu que“ – französische Gerichtsurteile als Herausforderung für den Übersetzer. In: *La traduction juridique: Histoire, théorie(s) et pratique*. Bern: ASTTI, S. 713-736.
- Barsalou, Lawrence W. (1999): Perceptual symbol systems. In: *Behavioral and Brain Sciences* 22, S. 577-660.
- Bates, Douglas M./Maechler, Martin/Bolker, Ben (2012): *lme4: Linear mixed-effects models using Eigen and Eigen++*. [= Handbuch zu R package version 0.999999-0].
- Baumann, Peter (2012): The role of hierarchical structure in syntactic dependency integration. In: *Bradley/Fernández/Fodor (Hg.)*, S. 91.
- Becker-Mrotzek, Michael (1999): Die Sprache der Verwaltung als Institutionensprache. In: *Hoffmann, Lothar/Kalverkämper, Hartwig/Wiegand, Herbert Ernst (Hg.): Fachsprachen: Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft*. 2. Halbbd. (= Handbücher zur Sprach- und Kommunikationswissenschaft 14.2). Berlin/New York: de Gruyter, S. 1391-1402.
- Bierwisch, Manfred (1965): Poetik und Linguistik. In: *Gunzenhäuser, Rul/Kreuzer, Helmut (Hg.): Mathematik und Dichtung*. München: Nymphenburger, S. 49-66.
- Binder, Katherine S./Pollatsek, Alexander/Rayner, Keith (1999): Extraction of information to the left of the fixated word in reading. In: *Journal of Experimental Psychology: Human Perception and Performance* 25, 4, S. 1162-1172.
- Bohnet, Bernd (2003): Mapping phrase structures to dependency structures in the case of (partially) free word order languages. In: *Proceedings of the First International Conference on Meaning-Text Theory*. Paris, S. 239-249
- Börsenverein des Deutschen Buchhandels (2012): *Markt mit Perspektiven – das E-Book in Deutschland 2011*. Pressemappe. [www.boersenverein.de/sixcms/media.php/976/E-Book-Studie%202012](http://www.boersenverein.de/sixcms/media.php/976/E-Book-Studie%202012) (Stand: 4.3.2013).

- Boston, Marisa F./Hale, John T./Kliegl, Reinhold/Vasishth, Shrvan (2008): Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. In: *Journal of Eye Movement Research* 2, 1, S. 1-12.
- Bradley, Dianne/Fernández, Eva/Fodor, Janet Dean (Hg.) (2012): *Proceedings of the 25th Annual CUNY Conference on Human Sentence Processing*. New York: CUNY Graduate School.
- Buckley, Paul B./Gillman, Clifford B. (1974): Comparisons of digits and dot patterns. In: *Journal of Experimental Psychology* 103, 6, S. 1131-1136.
- Bundesministerium der Justiz (2011): Bundesjustizministerium mit dem Preis für gute Gesetzgebung ausgezeichnet. Pressemitteilung ohne Nummerierung. [www.bmj.de/SharedDocs/Pressemitteilungen/DE/2011/20111026\\_Preis\\_fuer\\_gute\\_Gesetzgebung.html](http://www.bmj.de/SharedDocs/Pressemitteilungen/DE/2011/20111026_Preis_fuer_gute_Gesetzgebung.html) (Stand: 25.10.2012).
- Bundesverfassungsgericht (2003a): Zur Strafbarkeit des unerlaubten Aufenthalts in der Bundesrepublik Deutschland. Pressemitteilung 19/2003. [www.bundesverfassungsgericht.de/pressemitteilungen/bvg19-03.html](http://www.bundesverfassungsgericht.de/pressemitteilungen/bvg19-03.html) (Stand: 15.2.2012).
- Bundesverfassungsgericht (2003b): Zur Beteiligung deutscher Soldaten an NATO AWACS-Einsatz in der Türkei. Pressemitteilung 26/2003. [www.bundesverfassungsgericht.de/pressemitteilungen/bvg26-03.html](http://www.bundesverfassungsgericht.de/pressemitteilungen/bvg26-03.html) (Stand: 15.2.2012).
- Bundesverfassungsgericht (2004): Zur Verfassungsmäßigkeit der Strafvorschriften über den Umgang mit Cannabis. Pressemitteilung 67/2004. [www.bundesverfassungsgericht.de/pressemitteilungen/bvg04-067.html](http://www.bundesverfassungsgericht.de/pressemitteilungen/bvg04-067.html) (Stand: 12.2.2012).
- Calvo, Manuel G./Meseguer, Enrique/Carreiras, Manuel (2001): Inferences about predictable events: Eye movements during reading. In: *Psychological Research* 65, S. 158-169.
- Caplan, David/Waters, Gloria S. (1999): Verbal working memory and sentence comprehension. In: *Behavioral and Brain Sciences* 22, S. 77-126.
- Carpenter, Patricia A./Just, Marcel Adam (1975): Sentence comprehension: A psycholinguistic processing model of verification. In: *Psychological Review* 82, 1, S. 45-73.
- Chomsky, Noam R. (1965): *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, Noam R. (1981): *Lectures on government and binding: The Pisa lectures*. Dordrecht: Foris.
- Christmann, Ursula (2004): Verstehens- und Verständlichkeitsmessung: Methodische Ansätze in der Anwendungsforschung. In: Lerch (Hg.), S. 33-62.
- Christmann, Ursula/Groeben, Norbert (1996): Textverstehen, Textverständlichkeit – Ein Forschungsüberblick unter Anwendungsperspektive. In: Krings, Hans Peter (Hg.): *Wissenschaftliche Grundlagen der technischen Kommunikation*. Tübingen: Narr, S. 129-189.

- Clark, Herbert H. (1973): The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. In: *Journal of Verbal Learning and Verbal Behavior* 12, S. 335-359.
- Clifton, Chris, Jr./Staub, Adrian/Rayner, Keith (2007): Eye movements in reading words and sentences. In: van Gompel, Roger/Fischer, Martin H./Murray, Wayne/Hill, Robin L. (Hg.): *Eye movements: A window on mind and brain*. Amsterdam u.a.: Elsevier, S. 341-371.
- Clinton, William J. (1998): Memorandum on plain language in government writing, June 1, 1998. In: Peters, Gerhard/Woolley, John T.: *The American Presidency Project*. [www.presidency.ucsb.edu/ws/index.php?pid=56067](http://www.presidency.ucsb.edu/ws/index.php?pid=56067) (Stand: 19.5.2017).
- Consten, Manfred/Schwarz-Friesel, Monika (2009): Anapher. In: Hoffmann, Ludger (Hg.): *Handbuch der deutschen Wortarten*. Berlin: de Gruyter, S. 263-292.
- Cook, Anne E./Myers, Jerome L./O'Brien, Edward J. (2005): Processing an anaphor when there is no antecedent. In: *Discourse Processes* 39, 1, S. 101-120.
- Corbett, Albert T. (1984): Prenominal adjectives and the disambiguation of anaphoric nouns. In: *Journal of Verbal Learning and Verbal Behavior* 23, S. 683-695.
- Daudaravičius, Vidas/Marcinkevičiene, Ruta (2004): Gravity counts for the boundaries of collocations. In: *International Journal of Corpus Linguistics* 9, 2, S. 321-348.
- De Vogelaer, Gunther (2007): Extending Hawkins' comparative typology: Case, word order, and verb agreement in the Germanic languages. In: *Nordlyd* 34, S. 167-182.
- Demberg, Vera/Keller, Frank (2008): Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. In: *Cognition* 109, S. 193-210.
- Demberg, Vera/Keller, Frank (2009): A computational model of prediction in human parsing: Unifying locality and surprisal effects. In: *Proceedings of the 29th Meeting of the Cognitive Science Society (CogSci-09)*. Amsterdam: Cognitive Science Society, S. 947-952.
- Demberg, Vera/Keller, Frank/Koller, Alexander (2013): Incremental, predictive parsing with psycholinguistically motivated tree-joining grammar. In: *Computational Linguistics* 9, 4, S. 1025-1066.
- Dickes, Paul/Steiner, Laure (1977): Ausarbeitung von Lesbarkeitsformeln für die deutsche Sprache. In: *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 9, 1, S. 20-28.
- Diessel, Holger (1999): *Demonstratives: Form, function, and grammaticalization*. Amsterdam: Benjamins.
- Duffy, Susan A./Rayner, Keith (1990): Eye movements and anaphor resolution: Effects of antecedent typicality and distance. In: *Language and Speech* 33, 2, S. 103-119.
- Dürscheid, Christa (2007): *Syntax: Grundlagen und Theorien*. 4., überarb. und erg. Aufl. Göttingen: Vandenhoeck und Ruprecht.



- Ehrlich, Stefan F./Rayner, Keith (1981): Contextual effects on word perception and eye movements during reading. In: *Journal of Verbal Learning and Verbal Behavior* 20, S. 641-655.
- Eichhoff-Cyrus, Karin M./Antos, Gerd (Hg.) (2009): *Verständlichkeit als Bürgerrecht? Die Rechts- und Verwaltungssprache in der öffentlichen Diskussion*. Bd. 9. Mannheim: Bibliographisches Institut.
- Eichhoff-Cyrus, Karin M./Strobel, Thomas (2009): Einstellungen der Justiz zur Rechts- und Verwaltungssprache: Eine Trendumfrage. In: *Der Sprachdienst* 53, 5, S. 133-151.
- Eichhoff-Cyrus, Karin M./Antos, Gerd/Schulz, Rüdiger (2009): *Wie denken die Deutschen über die Rechts- und Verwaltungssprache? Eine repräsentative Umfrage der Gesellschaft für deutsche Sprache*. Wiesbaden: Gesellschaft für deutsche Sprache.
- Elman, Jeffrey L. (2009): On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. In: *Cognitive Science* 33, S. 1-36.
- Engbert, Ralf/Nuthmann, Antje/Richter, Eike/Kliegl, Reinhold (2005): SWIFT: A dynamical model of saccade generation during reading. In: *Psychological Review* 112, S. 777-813.
- Enzensberger, Hans-Magnus (2001): Von den Vorzügen der Unverständlichkeit. In: *Rechtshistorisches Journal* 20, S. 523-524.
- Ericsson, K. Anders/Kintsch, Walter (1995): Long-term working memory. In: *Psychological Review* 102, 2, S. 211-245.
- Federal Plain Language Guidelines (March 2011, Rev. 1, May 2011), Plain Language Action and Information Network (PLAIN). [www.plainlanguage.gov/howto/guidelines/FederalPLGuidelines/FederalPLGuidelines.pdf](http://www.plainlanguage.gov/howto/guidelines/FederalPLGuidelines/FederalPLGuidelines.pdf) (Stand: 8.11.2012).
- Fedorenko, Evelina/Piantadosi, Steve/Gibson, Edward (2012): Processing relative clauses in supportive contexts. In: *Cognitive Science* 36, 3, S. 471-497.
- Ferreira, Fernanda/Bailey, Karl D. G./Ferraro, Vittoria (2002): Good-enough representations in language comprehension. In: *Current Directions in Psychological Science* 11, S. 11-15.
- Ferreira, Fernanda/Patson, Nikole D. (2007): The 'good enough' approach to language comprehension. In: *Language and Linguistics Compass* 1, 1-2, S. 71-83.
- Field, Andy/Miles, Jeremy/Field, Zoe (2012): *Discovering statistics using R*. Los Angeles: Sage.
- Fluck, Hans-Rüdiger (1997): *Fachdeutsch in Naturwissenschaft und Technik: Einführung in die Fachsprachen und die Didaktik/Methodik des fachorientierten Fremdsprachenunterrichts (Deutsch als Fremdsprache)*. 2., neu bearb. Aufl. Heidelberg: Groos.
- Foltz, Peter W./Kintsch, Walter/Landauer, Thomas K. (1998): The measurement of textual coherence with latent semantic analysis. In: *Discourse Processes* 25, 2, S. 285-307.

- Forbes, Fritz W./Cottle, William C. (1953): A new method for determining the readability of standardized tests. In: *Journal of Applied Psychology* 37, S. 185-190.
- Fox, John (2003): Effect displays in R for Generalised Linear Models. In: *Journal of Statistical Software* 8, 15, S. 1-27.
- Frank, Nicola (2007): Tagungsbericht zum Symposium der GfdS in Zusammenarbeit mit dem germanistischen Institut der Martin-Luther-Universität Halle-Wittenberg: Verständlichkeit als Bürgerrecht? Die Rechts- und Verwaltungssprache in der öffentlichen Diskussion. In: *Der Sprachdienst* 6/07, S. 243-250.
- Frank, Stefan L./Monsalve, Irene Fernandez/Thompson Robin L./Vigliocco, Gabriella (2013): Reading time data for evaluating broad-coverage models of English sentence processing. In: *Behavior Research Methods* 45, S. 1182-1190.
- Frazier, Lyn/Fodor, Janet Dean (1978): The sausage machine: A new two-stage parsing model. In: *Cognition* 6, S. 291-325.
- Frenck-Mestre, Cheryl (2005): Eye-movement recording as a tool for studying syntactic processing in a second language: A review of methodologies and experimental findings. In: *Second Language Research* 21, 2, S. 175-198.
- Frisson, Steven/Rayner, Keith/Pickering, Martin J. (2005): Effects of contextual predictability and transitional probability on eye movements during reading. In: *Journal of Experimental Psychology: Learning, Memory and Cognition* 31, 5, S. 862-877.
- Garrod, Simon C./Sanford, Anthony J. (1977): Interpreting anaphoric relations: The integration of semantic information while reading. In: *Journal of Verbal Learning and Verbal Behavior* 16, S. 77-90.
- Garrod, Simon C./Sanford, Anthony J. (1994): Resolving sentences in a discourse context: How discourse representation affects language understanding. In: Gernsbacher (Hg.), S. 675-698.
- Garrod, Simon C./Freudenthal, Daniel/Boyle, Elizabeth (1994): The role of different types of anaphor in the on-line resolution of sentences in a discourse. In: *Journal of Memory and Language* 33, S. 39-68.
- Gernsbacher, Morton Ann (1984): Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. In: *Journal of Experimental Psychology: General* 113, 2, S. 256-281.
- Gernsbacher, Morton Ann (1989): Mechanisms that improve referential access. In: *Cognition* 32, S. 99-156.
- Gernsbacher, Morton Ann (Hg.) (1994): *Handbook of psycholinguistics*. San Diego, CA u.a.: Academic Press.
- Gernsbacher, Morton Ann/Foertsch, Julie (1999): Three models of discourse comprehension. In: Garrod, Simon C./Pickering, Martin J. (Hg.): *Language processing*. Hove: Psychology Press, S. 283-299.

- Gibson, Edward (1998): Linguistic complexity: Locality of syntactic dependencies. In: *Cognition* 68, S. 1-76.
- Gibson, Edward (2000): The dependency locality theory: A distance-based theory of linguistic complexity. In: Marantz, Alec/Yasushi, Miyashita/O'Neil, Wayne (Hg.): *Image, language, brain: Papers from the first mind articulation project symposium*. Cambridge, MA: MIT Press, S. 94-126.
- Gibson, Edward/Wu, H.-H. Iris (2011): Processing Chinese relative clauses in context. In: *Language and Cognitive Processes*, S. 1-31. <http://dx.doi.org/10.1080/01690965.2010.536656> (Stand: 20.3.2017).
- Göpferich, Susanne (2006): *Textproduktion im Zeitalter der Globalisierung: Entwicklung einer Didaktik des Wissenstransfers*. 2. Aufl. Tübingen: Stauffenburg. [Unveränd. Nachdr. der 1. Aufl. 2002].
- Göpferich, Susanne (2008): Textverstehen und Textverständlichkeit. In: Janich, Nina (Hg.): *Textlinguistik: 15 Einführungen*. Tübingen: Narr, S. 291-312.
- Göpferich, Susanne (2009): Comprehensibility assessment using the Karlsruhe Comprehensibility Concept. In: *The Journal of Specialised Translation* 11, S. 31-52.
- Gordon, Peter C./Hendrick, Randall/Johnson, Marcus (2001): Memory interference during language processing. In: *Journal of Experimental Psychology: Learning, Memory and Cognition* 27, 6, S. 1411-1423.
- Gordon, Peter C./Hendrick, Randall/Johnson, Marcus (2004): Effects of noun phrase type on sentence complexity. In: *Journal of Memory and Language* 51, S. 97-114.
- Gordon, Peter C./Hendrick, Randall/Johnson, Marcus/Lee, Yoonhyoung (2006): Similarity-based interference during language comprehension: Evidence from eye tracking during reading. In: *Journal of Experimental Psychology: Learning, Memory and Cognition* 32, 6, S. 1304-1321.
- Götz-Votteler, Katrin (2008): *Aspekte der Informationsentwicklung im Erzähltext*. (= Tübinger Beiträge zur Linguistik 505). Tübingen: Narr.
- Graesser, Arthur C./Millis, Keith K./Zwaan, Rolf A. (1997): Discourse comprehension. In: *Annual Review of Psychology* 48, S. 163-89.
- Graesser, Arthur C./Singer, Murray/Trabasso, Tom (1994): Constructing inferences during narrative text comprehension. In: *Psychological Review* 101, 3, S. 371-395.
- Grice, Herbert Paul (1975): Logic and conversation: Speech acts. In: Cole, Peter/Morgan, Jerry L. (Hg.): *Syntax and semantics*. Bd. 3. New York: Academic Press, S. 41-58.
- Gries, Stefan T. (2013): 50-something years of work on collocations: What is or should be next ... In: *International Journal of Corpus Linguistics* 18, 1, S. 137-165.
- Groeben, Norbert (1982): *Leserpsychologie: Textverständnis – Textverständlichkeit*. Münster: Aschendorff.

- Groeben, Norbert/Christmann, Ursula (1989): Textoptimierung unter Verständlichkeitsperspektive. In: Antos, Gerd/Krings, Hans Peter (Hg.): Textproduktion: ein interdisziplinärer Forschungsüberblick. (= Konzepte der Sprach- und Literaturwissenschaft 48). Tübingen: Niemeyer, S. 165-196.
- Groeben, Norbert/Christmann, Ursula (1996): Textverstehen und Textverständlichkeit aus sprach-/denkpsychologischer Sicht. In: Börner, Wolfgang/Vogel, Klaus (Hg.): Texte im Fremdsprachenerwerb: verstehen und produzieren. (= Tübinger Beiträge zur Linguistik 418). Tübingen: Narr, S. 67-89.
- Grosz, Barbara J. (1977): The representation and use of focus in a system for understanding dialogues (Report). Tech. Note 151, Artificial Intelligence Center, SRI International. Menlo Park, CA: SRI International.
- Hale, John T. (2001): A probabilistic earley parser as a psycholinguistic model. In: Proceedings of the 2nd Conference of the North American Chapter of the Association for Computational Linguistics. Bd. 2. Pittsburgh, PA: Association for Computational Linguistics, S. 159-166.
- Hale, John T. (2003): The information conveyed by words in sentences. In: Journal of Psycholinguistic Research 32, 2, S. 101-123.
- Hansen, Sandra (2009): Verwaltungssprache in Erpresserbriefen. In: Beiträge zur Fremdsprachenvermittlung, Sonderheft 15, S. 295-312.
- Hansen, Sandra/Dirksen, Ralf/Kunz, Kerstin/Küchler, Martin/Neumann, Stella (2006): Comprehensible legal texts – utopia or a question of wording? On processing rephrased German court decisions. In: Hermes – Journal of Language and Communication Studies 36, S. 15-40.
- Hansen, Silvia/Teich, Elke (1999): Kontrastive Analyse von Übersetzungskorpora: ein funktionales Modell. In: Gippert, Jost (Hg.): Sammelband der Jahrestagung der GLDV 99. Prag: enigma corporation, S. 311-322.
- Hansen-Schirra, Silvia/Neumann, Stella (2004): Linguistische Verständlichmachung in der juristischen Realität. In: Lerch (Hg.), S. 167-184.
- Hasselrot, Anne-Marie (2012): The Swedish approach to clear legislation. In: Proceedings of Clarity 68 (November 2012). Washington, DC: Clarity, S. 36-37.
- Haviland, Susan E./Clark, Herbert H. (1974): What's new? Acquiring new information as a process in comprehension. In: Journal of Verbal Learning and Verbal Behavior 13, S. 512-521.
- Hawkins, John A. (1986): A comparative typology of English and German: Unifying the contrasts. London: Croom Helm.
- Heister, Julia/Würzner, Kay-Michael/Bubenzer, Johannes/Pohl, Edmund/Hanneforth, Thomas/Geyken, Alexander/Kliegl, Reinhold (2011): dlexDB – eine lexikalische Datenbank für die psychologische und linguistische Forschung. In: Psychologische Rundschau 62, 1, S. 10-20.

- Hemforth, Barbara/Konieczny, Lars (2008): Sätze und Texte verstehen und produzieren. In: Müssele, Jochen (Hg.): *Allgemeine Psychologie*. 2., neu bearb. Aufl. Heidelberg: Spektrum Akademie Verlag, S. 504-549.
- Hemforth, Barbara/Konieczny, Lars/Scheepers, Christoph/Colonna, Savéria/Schimke, Sarah/Baumann, Peter/Pynte, Joël (2010): Language specific preferences in anaphor resolution: Exposure or gricean maxims? In: Ohlsson, Stellan/Catrambone, Richard (Hg.): *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*. Portland, OR: Cognitive Science Society, S. 2218-2223.
- Henderson, John M./Ferreira, Fernanda (1990): Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. In: *Journal of Experimental Psychology: Learning, Memory and Cognition* 16, S. 417-429.
- Heutger, Viola (2006): Der Platz der juristischen Fachsprache in der Experten-Laien-Kommunikation. In: *Hermes – Journal of Language and Communication Studies* 36, S. 55-63.
- Hochberg, Julian (1970): Components of literacy: Speculations and exploratory research. In: Levin, Harry/Williams, Joanna P. (Hg.): *Basic studies on reading*. New York u.a.: Basic Books, S. 74-89.
- Hüttner, Manfred (1986): *Prognoseverfahren und ihre Anwendung*. Berlin/New York: de Gruyter.
- Hyönä, Jukka/Lorch, Robert F., Jr./Rinck, Mike (2003): Eye movement measures to study global text processing. In: Hyönä, Jukka/Radach, Ralph/Deubel, Heiner (Hg.): *The mind's eye: Cognitive and applied aspects of eye movement research*. Amsterdam u.a.: Elsevier, S. 313-334.
- Hyönä, Jukka/Tommola, Jorma/Alaja, Anne-Mari (1995): Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. In: *The Quarterly Journal of Experimental Psychology. Section A: Human Experimental Psychology* 48, 3, S. 598-612.
- Jaeger, T. Florian (2008): Categorical data analysis: Away from ANOVAs (transformations or not) and towards logit mixed models. In: *Journal of Memory and Language* 59, S. 434-446.
- Janisse, Michael P. (1974): Pupil size, affect and exposure frequency. In: *Social Behavior and Personality* 2, 2, S. 125-146.
- Johnson, Wendell (1944): *Studies in language behaviour: I. A program of research*. (= *Psychological Monographs: General and Applied* 56, 2). Washington, DC u.a.: American Psychological Association, S. 1-15.
- Joshi, Aravind K./Levy, Leon S./Takahashi, Masako (1975): Tree adjunct grammars. In: *Journal of Computer and System Sciences* 10, 1, S. 136-163.
- Just, Marcel Adam/Carpenter, Patricia A. (1976): Eye fixations and cognitive processes. In: *Cognitive Psychology* 8, S. 441-480.

- Just, Marcel Adam/Carpenter, Patricia A. (1980): A theory of reading: From eye fixations to comprehension. In: *Psychological Review* 87, 4, S. 329-354.
- Just, Marcel Adam/Carpenter, Patricia A. (1992): A capacity theory of comprehension: Individual differences in working memory. In: *Psychological Review* 99, 1, S. 122-149.
- Kaan, Edith (2002): Investigating the effects of distance and number interference in processing subject-verb dependencies: An ERP study. In: *Journal of Psycholinguistic Research* 31, 2, S. 165-193.
- Kahnemann, Daniel/Beatty, Jackson (1966): Pupil diameter and load on memory. In: *Science* 154, 3756, S. 1583-1585.
- Kaiser, Elsi (2010): Investigating the consequences of focus on the production and comprehension of referring expressions. In: *International Review of Pragmatics* 2, S. 266-297.
- Kaiser, Elsi (2011): Focusing on pronouns: Consequences of subjecthood, pronominalisation, and contrastive focus. In: *Language and Cognitive Processes* 26, 10, S. 1625-1666.
- Kennedy, Alan (2000): Parafoveal processing in word recognition. In: *The Quarterly Journal of Experimental Psychology. Section A: Human Experimental Psychology* 53, 2, S. 429-455.
- Kennedy, Alan (2003): *The Dundee Corpus*. Dundee: The University of Dundee, School of Psychology. [CD-ROM].
- Kennedy, Alan/Pynte, Joel (2005): Parafoveal-on-foveal effects in normal reading. In: *Vision Research* 45, S. 153-168.
- Kennedy, Alan/Hill, Robin L./Pynte, Joel (2003): *The Dundee Corpus*. In: van Gompel, Roger P.G. (Hg.): *Proceedings of the 12th European Conference on Eye Movements*. Dundee: Dundee University.
- Kennedy, Alan/Radach, Ralph/Heller, Dieter/Pynte, Joel (Hg.) (2000): *Reading as a perceptual process*. Amsterdam: Elsevier.
- Kennison, Shelia M./Clifton, Charles, Jr. (1995): Determinants of parafoveal preview benefit in high and low working memory capacity readers: Implications for eye movement control. In: *Journal of Experimental Psychology: Learning, Memory and Cognition* 21, 1, S. 68-81.
- Kintsch, Walter (1974): *The representation of meaning in memory*. New York: Halsted Press.
- Kintsch, Walter (1988): The role of knowledge in discourse comprehension: A construction-integration model. In: *Psychological Review* 95, 2, S. 163-182.
- Kintsch, Walter (1998): *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

- Kintsch, Walter/Greeno, James G. (1985): Understanding and solving word arithmetic problems. In: *Psychological Review* 92, 1, S. 109-129.
- Kintsch, Walter/van Dijk, Teun A. (1978): Toward a model of text comprehension and production. In: *Psychological Review* 85; 5, S. 363-394.
- Klare, George R. (1963): *The measurement of readability: Useful information for communicators*. Ames, IA: Iowa State University Press.
- Klare, George R. (2000): *The measurement of readability: Useful information for communicators*. In: *ACM Journal of Computer Documentation* 24, 3, S. 107-121.
- Kliegl, Reinhold/Nuthmann, Antje/Engbert, Ralf (2006): Tracking the mind during reading: The influence of past, present, and future words on fixation durations. In: *Journal of Experimental Psychology: General* 135, 1, S. 12-35.
- Kliegl, Reinhold/Olson, Richard K./Davidson, Brian J. (1983): On problems of unconfounding perceptual and language processes. In: Rayner, Keith (Hg.): *Eye movements in reading: Perceptual and language processes*. New York: Academic Press, S. 333-343.
- Kliegl, Reinhold/Grabner, Ellen/Rolfs, Martin/Engbert, Ralf (2004): Length, frequency, and predictability effects of words on eye movements in reading. In: *European Journal of Cognitive Psychology* 16, 1-2, S. 262-284.
- Konieczny, Lars (2000): Locality and parsing complexity. In: *Journal of Psycholinguistic Research* 29, 6, S. 627-645.
- Konieczny, Lars/Müller, Daniel/Hachmann, Wibke/Schwarzkopf, Sarah/Wolfer, Sascha (2009): Local syntactic coherence interpretation. Evidence from a visual world study. In: Taatgen, Nils A./van Rijn, Hedderik/Schomaker, Lambert/Nerbonne, John (Hg.): *Proceedings of the 31st Annual Conference of the Cognitive Science Society*. Amsterdam: Cognitive Science Society, S. 1133-1138.
- Krefeld, Thomas (1985): *Das französische Gerichtsurteil in linguistischer Sicht: Zwischen Fach- und Standessprache*. Frankfurt a. M. u.a.: Lang.
- Krippendorf, Klaus (1994): *Der verschwundene Bote: Metaphern und Modelle der Kommunikation*. In: Merten, Klaus/Schmidt, Siegfried J./Weischenberg, Siegfried (Hg.): *Die Wirklichkeit der Medien: Eine Einführung in die Kommunikationswissenschaft*. Opladen: Westdeutscher Verlag, S. 79-113.
- Lakoff, George/Johnson, Marcus (1980): *Metaphors we live by*. Chicago, IL/London: University of Chicago Press.
- Landauer, Thomas K./Dumais, Susan T. (1994): Latent semantic analysis and measurement of knowledge. In: Kaplan, Randy M./Burstein, Jill C. (Hg.): *Educational testing service conference on natural language processing techniques and technology in assessment and education*. Princeton, NJ: Educational Testing Service.



- Landauer, Thomas K./Foltz, Peter W./Laham, Darrell (1998): An introduction to latent semantic analysis. In: *Discourse Processes* 25, 2-3, S. 259-284.
- Lang, Duncan Temple (2012): Rstem: Interface to snowball implementation of porter's word stemming algorithm. [Handbuch zu R package version 0.4-1].
- Langer, Inghard/Schulz von Thun, Friedemann/Tausch, Reinhard (1974): *Verständlichkeit in Schule, Verwaltung, Politik und Wissenschaft: Mit einem Selbsttrainingsprogramm zur verständlichen Gestaltung von Lehr- und Informationstexten.* München/Basel: Reinhardt.
- Langer, Inghard/Schulz von Thun, Friedemann/Tausch, Reinhard (2006): *Sich verständlich ausdrücken.* 8. Aufl. München/Basel: Reinhardt.
- Language Act (2009): *Svensk författningssamling (SFS).* Stockholm: Sveriges riksdag, S. 600.
- Lashöfer, Jutta (1992): *Zum Stilwandel in richterlichen Entscheidungen: Über stilistische Veränderungen in englischen, französischen und deutschen zivilrechtlichen Urteilen und in Entscheidungen des Gerichtshofs der Europäischen Gemeinschaften.* (= Internationale Hochschulschriften 61). Münster/New York: Waxmann.
- Lerch, Kent D. (Hg.) (2004): *Recht verstehen: Verständlichkeit, Missverständlichkeit und Unverständlichkeit von Recht.* Bd. 1: Die Sprache des Rechts. Berlin/New York: de Gruyter.
- Levy, Roger (2008): Expectation-based syntactic comprehension. In: *Cognition* 106, 3, S. 1126-1177.
- Lewis, Richard L./Vasishth, Shravan (2005): An activation-based model of sentence processing as skilled memory retrieval. In: *Cognitive Science* 29, 3, S. 375-419.
- Lewis, Richard L./Vasishth, Shravan/Van Dyke, Julie A. (2006): Computational principles of working memory in sentence comprehension. In: *Trends in Cognitive Sciences* 10, 10, S. 447-454.
- Limbach, Jutta (2008): Sprachzucht ist ein Beitrag zur Demokratie. In: Eichhoff-Cyrus/Antos (Hg.), S. 371-377.
- Liversedge, Simon P./Paterson, Kevin B./Pickering, Martin J. (1998): Eye movements and measures of reading time. In: Underwood (Hg.), S. 55-75.
- Locke, Joanne (2004): *A history of plain language in the United States Government.* [www.plainlanguage.gov/whatisPL/history/locke.cfm](http://www.plainlanguage.gov/whatisPL/history/locke.cfm) (Stand: 9.11.2012).
- Lorch, Robert F., Jr./Myers, Jerome L. (1990): Regression analyses of repeated measures data in cognitive research. In: *Journal of Experimental Psychology: Learning, Memory and Cognition* 16, 1, S. 149-157.
- MacDonald, Margellan C./Pearlmutter, Neal J./Seidenberg, Mark S. (1994): Lexical nature of syntactic ambiguity resolution. In: *Psychological Review* 101, 4, S. 676-703.

- Masson, Michael E. J./Waldron, Marie Anne (1994): Comprehension of legal contracts by non-experts: Effectiveness of plain language redrafting. In: *Applied Cognitive Psychology* 8, S. 67-85.
- Mayer, Richard E. (1997): Multimedia learning: Are we asking the right questions? In: *Educational Psychologist* 32, S. 1-19.
- Mayer, Richard E. (2005a): Cognitive theory of multimedia learning. In: Mayer (Hg.), S. 31-48.
- Mayer, Richard E. (2005b): Introduction to multimedia learning. In: Mayer (Hg.), S. 1-16.
- Mayer, Richard E. (Hg.) (2005): *The Cambridge handbook of multimedia learning*. Cambridge u.a.: Cambridge University Press.
- McConkie, George W./Rayner, Keith (1975): The span of the effective stimulus during a fixation in reading. In: *Perception & Psychophysics* 17, 6, S. 578-586.
- McConkie, George W./Kerr, Paul W./Reddix, Michael D./Zola, David (1988): Eye movement control during reading: The location of initial eye fixations on words. In: *Vision Research* 28, 10, S. 1107-1118.
- McDonald, Scott A./Shillcock, Richard C. (2003): Eye movements reveal the on-line computation of lexical probabilities during reading. In: *Psychological Science* 14, 6, S. 648-652.
- McKoon, Gail/Ratcliff, Roger (1980): The comprehension processes and memory structures involved in anaphoric reference. In: *Journal of Verbal Learning and Verbal Behavior* 19, S. 668-682.
- McKoon, Gail/Ratcliff, Roger (1988): Contextually relevant aspects of meaning. In: *Journal of Experimental Psychology: Learning, Memory and Cognition* 14, 2, S. 331-343.
- McKoon, Gail/Ratcliff, Roger (1992): Inference during reading. In: *Psychological Review* 99, 3, S. 440-466.
- McKoon, Gail/Ratcliff, Roger (1998): Memory-based language processing: Psycholinguistic research in the 1990s. In: *Annual Review of Psychology* 49, S. 25-42.
- McKoon, Gail/Gerrig, Richard J./Greene, Steven B. (1996): Pronoun resolution without pronouns: Some consequences of memory-based text processing. In: *Journal of Experimental Psychology: Learning, Memory and Cognition* 22, 4, S. 919-932.
- Michel, Jean-Baptiste et al. (2011): Quantitative analysis of culture using millions of digitized books. In: *Science* 331, 6014, S. 176-182.
- Mielliet, Sebastien/Sparrow, Laurent/Sereno, Sara C. (2007): Word frequency and predictability effects in reading French: An evaluation of the E-Z Reader model. In: *Psychonomic Bulletin & Review* 14, 4, S. 762-769.

- Mitchell, Don C./Green, David W. (1978): The effects of context and content on immediate processing in reading. In: *Quarterly Journal of Experimental Psychology* 30, 4, S. 609-636.
- Mulkern, Ann E. (1996): The game of the name. In: Fretheim, Thorstein/Gundel, Jeanette K. (Hg.): *Reference and referent accessibility*. Amsterdam u.a.: Benjamins, S. 235-250.
- Müller-Feldmeth, Daniel (2013): *Modeling frequency effects in local syntactic coherence processing (Arbeitstitel)*. Diss. (unveröff.). Freiburg i.Br.: Universität Freiburg.
- Müller-Feldmeth, Daniel/Wolfer, Sascha/Konieczny, Lars (2013): The influence of preverbal referents on verb processing: Evidence for similarity-based interference from a new German reading corpus. In: *Frenck-Mestre, Cheryl/Alario, F.-Xavier/Nguyen, Noël/Blache, Philippe/Meunier, Christine (Hg.): Proceedings of the 19th Architectures and Mechanisms for Language Processing (AMLaP)*. Marseille: Aix-Marseille Université.
- Müller-Küppers, Evelyn/Zöllner, Inge (2007): *Leseverstehen: Fachtexte mit Übungen und methodischen Hinweisen*. 2. Aufl. Ismaning: Hueber.
- Neumann, Stella/Hansen-Schirra, Silvia (2004): Der Konjunktiv als Verständnisproblem in Rechtstexten. In: *Zeitschrift für Angewandte Linguistik* 41, S. 67-87.
- Nicol, Janet L./Swinney, David A. (2003): The psycholinguistics of anaphora. In: *Barss, Andrew (Hg.): Anaphora: A reference guide*. Malden, MA u.a.: Blackwell, S. 72-104.
- Nussbaumer, Markus (2001): Von Schwärmern und Skeptikern und ein Versuch, Realist zu sein. In: *Rechtshistorisches Journal* 20, S. 644-655.
- Nussbaumer, Markus (2008): Der Verständlichkeit eine Anwältin! Die Redaktionskommission der schweizerischen Bundesverwaltung und ihre Arbeit an der Gesetzssprache. In: *Eichhoff-Cyrus/Antos (Hg.)*, S. 301-323.
- O'Brien, Edward J./Albrecht, Jason E. (1991): The role of context in accessing antecedents in text. In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17, 1, S. 94-102.
- O'Brien, Edward J./Duffy, Susan A./Myers, Jerome L. (1986): Anaphoric inference during reading. In: *Journal of Experimental Psychology: Learning, Memory and Cognition* 12, 3, S. 346-352.
- O'Brien, Edward J./Raney, Gary E./Albrecht, Jason E./Rayner, Keith (1997): Processes involved in the resolution of explicit anaphors. In: *Discourse Processes* 23, S. 1-24.
- O'Brien, Edward J./Shank, Dolores M./Myers, Jerome L./Rayner, Keith (1988): Elaborative inferences during reading: Do they occur on-line? In: *Journal of Experimental Psychology: Learning, Memory and Cognition* 14, 3, S. 410-420.
- Oksaar, Els (1988): *Fachsprachliche Dimensionen*. (= *Forum für Fachsprachen-Forschung* 4). Tübingen: Narr.

- Osgood, Charles E./Luria, Zella (1954): A blind analysis of a case of multiple personality using the semantic differential. In: *The Journal of Abnormal and Social Psychology* 49, S. 579-591.
- Osgood, Charles E./Suci, George J./Tannenbaum, Percy (1957): *The measurement of meaning*. Urbana, IL: University of Illinois Press.
- Pampel, Fred C. (2000): *Logistic regression: A primer*. Thousand Oaks, CA u.a.: Sage.
- Patil, Umesh/Vasishth, Shraavan/Kliegl, Reinhold (2009): Compound effect of probabilistic disambiguation and memory retrievals on sentence processing: Evidence from an eye-tracking corpus. In: Howes, Andrew/Peebles, David/Cooper, Richard (Hg.): *Proceedings of the 9th International Conference on Cognitive Modeling*. Manchester: University of Manchester, S. 92-97.
- Perkuhn, Rainer/Keibel, Holger/Kupietz, Marc (2012): *Korpuslinguistik*. Paderborn: Fink.
- Petrov, Slav/Klein, Dan (2007): Improved inference for unlexicalized parsing. In: *Proceedings of the NAACL HLT 2007*. Rochester, NY, S. 404-411.
- Petrov, Slav/Barrett, Leon/Thibaux, Romain/Klein, Dan (2006): Learning accurate, compact, and interpretable tree annotation. In: *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL*. Sydney: Association for Computational Linguistics, S. 433-440.
- Pinheiro, Jose C./Bates, Douglas M. (2000): *Mixed-effects models in S and S-PLUS*. New York: Springer.
- Pinheiro, Jose C./Bates, Douglas M./DebRoy, Saikat/Sarkar, Deepayan/R Development Core Team (2010): *nlme: Linear and nonlinear mixed effects models*. [Handbuch zu R package version 3.1-97].
- Plaehn, Oliver (1998): *Annotate Programm-Dokumentation (negra project report)*. Saarbrücken: Universität des Saarlandes. [www.coli.uni-saarland.de/projects/sfb378/negra-corpus/annotate-manual.ps.gz](http://www.coli.uni-saarland.de/projects/sfb378/negra-corpus/annotate-manual.ps.gz) (Stand: 19.5.2017).
- Plain Regulations Act of 2012, H.R. 3786, 112th Congress (2012).
- Plain Writing Act of 2010, H.R. 946, 111th Congress, 2nd Session (2010).
- Pollatsek, Alexander/Reichle, Erik D./Rayner, Keith (2006): Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. In: *Cognitive Psychology* 52, S. 1-56.
- Porac, Clare/Coren, Stanley (1976): The dominant eye. In: *Psychological Bulletin* 83, 3, S. 880-897.
- Polenz, Peter von (1988): *Deutsche Satzsemantik: Grundbegriffe des Zwischen-den-Zeilen-Lesens*. 2., durchges. Aufl. Berlin/New York: de Gruyter.
- Potts, George R./Keenan, Janine M./Golding, Jonathan M. (1988): Assessing the occurrence of elaborative inferences: Lexical decision versus naming. In: *Journal of Memory and Language* 27, S. 399-415.

- Pynte, Joel/Kennedy, Alan (2006): An influence over eye movements in reading exerted from beyond the level of the word: Evidence from reading English and French. In: *Vision Research* 46, 22, S. 3786-3801. doi: 10.1016/j.visres.2006.07.004.
- Pynte, Joel/New, Boris/Kennedy, Alan (2008a): A multiple regression analysis of syntactic and semantic influences in reading normal text. In: *Journal of Eye Movement Research* 2, 1, S. 1-11.
- Pynte, Joel/New, Boris/Kennedy, Alan (2008b): On-line contextual influences during reading normal text: A multiple regression analysis. In: *Vision Research* 48, 21, S. 2172-2183.
- Pynte, Joel/New, Boris/Kennedy, Alan (2009): On-line syntactic and semantic influences in reading revisited. In: *Journal of Eye Movement Research* 3, 1, S. 1-12.
- Radach, Ralph/Heller, Dieter (2000): Relations between spatial and temporal aspects of eye movement control. In: Kennedy et al. (Hg.), S. 165-192.
- Radach, Ralph/McConkie, George W. (1998): Determinants of fixation positions in words during reading. In: Underwood (Hg.), S. 77-100.
- Ratcliff, Roger (1978): A theory of memory retrieval. In: *Psychological Review* 85, 2, S. 59-108.
- Rayner, Keith (1979): Eye guidance in reading: Fixation locations within words. In: *Perception* 8, S. 21-30.
- Rayner, Keith (1998): Eye movements in reading and information processing: 20 years of research. In: *Psychological Bulletin* 124, 3, S. 372-422.
- Rayner, Keith/Fischer, Martin H. (1996): Mindless reading revisited: Eye movements during reading and scanning are different. In: *Perception & Psychophysics* 58, 5, S. 734-747.
- Rayner, Keith/Pollatsek, Alexander (2006): Eye movement control in reading. In: Traxler, Matthew J./Gernsbacher, Morton Ann (Hg.): *Handbook of psycholinguistics*. 2. Aufl. Amsterdam u.a.: Elsevier, S. 613-657.
- Rayner, Keith/Reichle, Erik D./Pollatsek, Alexander (2000): Eye movement control in reading: Updating the E-Z Reader model to account for initial fixation locations and refixations. In: Kennedy et al. (Hg.), S. 701-719.
- Rayner, Keith/Well, Arnold D./Pollatsek, Alexander/Bertera, James H. (1982): The availability of useful information to the right of fixation in reading. In: *Perception & Psychophysics* 31, 6, S. 537-550.
- Reboul, Anne (1997): What (if anything) is accessibility? A relevance-oriented criticism of Ariel's accessibility theory of referring expressions. In: Connolly, John H. (Hg.): *Discourse and pragmatics in functional grammar*. Berlin/New York: de Gruyter, S. 91-108.

- Reddy, Michael J. (1979): The conduit metaphor: A case of frame conflict in our language about language. In: Ortony, Andrew (Hg.): *Metaphor and thought*. Cambridge, MA u.a.: Cambridge University Press, S. 284-324.
- Reichle, Erik D./Rayner, Keith (2000): Comparing the E-Z reader model to other models of eye movement control in reading. <http://cogprints.org/1169/> (Stand: 20.3.2017).
- Reichle, Erik D./Rayner, Keith/Pollatsek, Alexander (1999): Eye movement control in reading: Accounting for initial fixation locations and refixations within the E-Z Reader model. In: *Vision Research* 39, S. 4403-4411.
- Reichle, Erik D./Rayner, Keith/Pollatsek, Alexander (2003): The E-Z reader model of eye-movement control in reading: Comparisons to other models. In: *Behavioral and Brain Sciences* 26, 4, S. 477-526.
- Reichle, Erik D./Warren, Tessa/McConnell, Kerry (2009): Using E-Z reader to model the effects of higher level language processing on eye movements during reading. In: *Psychonomic Bulletin & Review* 16, 1, S. 1-21.
- Reichle, Erik D./Pollatsek, Alexander/Fisher, Donald L./Rayner, Keith (1998): Toward a model of eye movement control in reading. In: *Psychological Review* 105, 1, S. 125-157.
- Remberger, Eva-Maria (2006): *Hilfsverben: Eine minimalistische Analyse am Beispiel des Italienischen und Sardischen*. (= Linguistische Arbeiten 504). Tübingen: Niemeyer.
- Reynolds, Ralph E./Anderson, Richard C. (1982): Influence of questions on the allocation of attention during reading. In: *Journal of Educational Psychology* 74, 5, S. 623-632.
- Rice, Glenn A./Robinson, David Owen (1975): The role of bigram frequency in the perception of words and nonwords. In: *Memory and Cognition* 3, S. 513-518.
- Sander, Gerald G. (2004): *Deutsche Rechtssprache: Ein Arbeitsbuch*. Tübingen: Francke.
- Sanford, Anthony J./Garrod, Simon C. (1981): *Understanding written language: Explorations of comprehension beyond the sentence*. Chicester/New York: Wiley.
- Sanford, Anthony J./Garrod, Simon C. (1989): What, when and how?: Questions of immediacy in anaphoric reference resolution. In: *Language and Cognitive Processes* 4, 3-4, S. 235-262.
- Sanford, Anthony J./Garrod, Simon C. (1994): Selective processing in text understanding. In: Gernsbacher (Hg.), S. 699-719.
- Sanford, Anthony J./Garrod, Simon C. (1998): The role of scenario mapping in text comprehension. In: *Discourse Processes* 26, 2-3, S. 159-190.
- Sanford, Anthony J./Sturt, Patrick (2002): Depth of processing in language comprehension: Not noticing the evidence. In: *Trends in Cognitive Sciences* 6, 9, S. 382-386.
- Sanford, Anthony J./Moar, K./Garrod, Simon C. (1988): Proper names as controllers of discourse focus. In: *Language and Speech* 31, 1, S. 43-56.

- Schank, Roger C./Abelson, Robert P. (1977): *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Erlbaum.
- Schmalhofer, Franz/McDaniel, Mark A./Keefe, Dennis (2002): A unified model for predictive and bridging inferences. In: *Discourse Processes* 33, 2, S. 105-132.
- Schmidt, Siegfried J. (1976): *Texttheorie: Probleme einer Linguistik der sprachlichen Kommunikation*. 2., verb. und ergänzte Aufl. München: Fink.
- Schnell, Rainer/Hill, Paul B./Esser, Elke (2008): *Methoden der empirischen Sozialforschung*. 8. Aufl. München: Oldenbourg.
- Schnotz, Wolfgang (2005): An integrated model of text and picture comprehension. In: Mayer (Hg.), S. 49-69.
- Schnotz, Wolfgang (2006): Was geschieht im Kopf des Lesers? Mentale Konstruktionsprozesse beim Textverstehen aus der Sicht der Psychologie und der kognitiven Linguistik. In: Blühorn, Hardarik/Breindl, Eva/Waßner, Ulrich Hermann (Hg.): *Text – Verstehen: Grammatik und darüber hinaus*. (= Jahrbuch des Instituts für Deutsche Sprache 2005). Berlin/New York: de Gruyter, S. 222-238.
- Schnotz, Wolfgang/Bannert, Maria (1999): Einflüsse der Visualisierungsform auf die Konstruktion mentaler Modelle beim Text- und Bildverstehen. In: *Zeitschrift für Experimentelle Psychologie* 46, 3, S. 217-236.
- Schnotz, Wolfgang/Bannert, Maria (2003): Construction and inference in learning from multiple representation. In: *Learning and Instruction* 13, S. 141-156.
- Schnotz, Wolfgang/Bannert, Maria/Seufert, Tina (2002): Toward an integrative view of text and picture comprehension: Visualization effects on the construction of mental models. In: Otero, José/León, José A./Graesser, Arthur C. (Hg.): *The psychology of science text comprehension*. Mahwah, NJ: Erlbaum, S. 385-416.
- Schriver, Karen A. (1989): Evaluating text quality: The continuum from text-focused to reader-focused methods. In: *IEEE Transactions on Professional Communication* 32, 4, S. 238-255.
- Schröder, Ole/Würdemann, Christian (2008): Rechtstexte verständlich formulieren. Implementierung einer Sprachanalyse im Gesetzgebungsverfahren. In: Eichhoff-Cyrus/Antos (Hg.), S. 324-335.
- Schwarze, Sabine (2007): Wissenschaftsstile in der Romania: Frankreich/Italien. In: Auer/Baßler (Hg.), S. 185-210.
- Shannon, Claude E. (1948): A mathematical theory of communication. In: *Bell System Technical Journal* 27, 3-4, S. 379-423, 623-656.
- Shannon, Claude E. (1951): Prediction and entropy of printed English. In: *Bell System Technical Journal* 30, S. 50-64.
- Shannon, Claude E./Weaver, Warren (1963): *The mathematical theory of communication*. Urbana, IL/Chicago, IL: University of Illinois Press.



- Shepard, Roger N./Metzler, Jacqueline (1971): Mental rotation of three-dimensional objects. In: *Science* 171, 3972, S. 701-703.
- Siegle, Greg J./Steinhauer, Stuart R./Carter, Cameron S./Ramel, Wiveka/Thase, Michael E. (2003): Do the seconds turn into hours? Relationships between sustained pupil dilation in response to emotional information and self-reported rumination. In: *Cognitive Therapy and Research* 27, 3, S. 365-382.
- Simon, Dieter (2004): Rechtsverständlichkeit. In: Lerch (Hg.), S. 405-411.
- Singer, Murray (2007): Inference processing in discourse comprehension. In: Gaskell, Gareth (Hg.): *Oxford handbook of psycholinguistics*. Oxford: Oxford University Press, S. 343-360.
- Singer, Murray/Andrusiak, Peter/Reisdorf, Paul/Black, Nancy L. (1992): Individual differences in bridging inference processes. In: *Memory and Cognition* 20, 5, S. 539-548.
- Sperber, Dan/Wilson, Deirde (1995): *Relevance: Communication and cognition*. 2. Aufl. Oxford: Blackwell.
- Starr, Matthew S./Rayner, Keith (2001): Eye movements during reading: Some current controversies. In: *Trends in Cognitive Sciences* 5, 4, S. 156-163.
- Steinbach, Ingrid/Langer, Inghard/Tausch, Reinhard (1972): Merkmale von Wissens- und Informationstexten im Zusammenhang mit der Lerneffektivität. In: *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 4, S. 130-139.
- Sturt, Patrick (2003): The time-course of the application of binding constraints in reference resolution. In: *Journal of Memory and Language* 48, S. 542-562.
- Sweller, John (2005): Implications of cognitive load theory for multimedia learning. In: Mayer (Hg.), S. 19-30.
- Tabor, Whitney/Galantucci, Bruno/Richardson, Daniel (2004): Effects of merely local syntactic coherence on sentence processing. In: *Journal of Memory and Language* 50, S. 355-370.
- Tagliamonte, Sali A./Baayen, R. Harald (2012): Models, forests and trees of York English: *was/were* variation as a case study for statistical practice. In: *Language Variation and Change* 24, 2, S. 135-178.
- Tanenhaus, Michael K./Spivey-Knowlton, Michael J./Eberhard, Kathleen M./Sedivy, Julie C. (1995): Integration of visual and linguistic information in spoken language comprehension. In: *Science* 268, S. 1632-1634.
- Tannen, Deborah/Wallat, Cynthia (1986): Medical professionals and parents: A linguistic analysis of communication across contexts. In: *Language in Society* 15, 3, S. 295-311.

- Tapiero, Isabelle/Denhière, Guy (1995): Simulating recall and recognition by using Kintsch's construction-integration model. In: Weaver, Charles A./Mannes, Suzanne/Fletcher, Charles R. (Hg.): *Discourse comprehension: Essays in honor of Walter Kintsch*. Kap. 12. Hillsdale, NJ: Erlbaum.
- Taylor, Wilson (1953): Cloze procedure: A new tool for measuring readability. In: *Journalism Quarterly* 30, S. 415-433.
- Till, Robert E./Mross, Ernest F./Kintsch, Walter (1988): Time course of priming for associate and inference words in a discourse context. In: *Memory and Cognition* 16, 4, S. 283-298.
- Towfigh, Emanuel Vahid (2009): Komplexität und Normenklarheit – oder: Gesetze sind für Juristen gemacht. In: *Der Staat* 48, 1, S. 29-73.
- Tulving, Endel (1974): Cue-dependent forgetting: When we forget something we once knew, it does not necessarily mean the memory trace has been lost; it may only be inaccessible. In: *American Scientist* 62, 1, S. 74-82.
- Underwood, Geoffrey (Hg.) (1998): *Eye guidance in reading and scene perception*. Amsterdam u.a.: Elsevier.
- van Dijk, Teun A. (1977): Semantic macro-structures and knowledge frames in discourse comprehension. In: Just, Marcel Adam/Carpenter, Patricia A. (Hg.): *Cognitive processes in comprehension*. Hillsdale, NJ: Erlbaum, S. 3-32.
- van Dijk, Teun A. (1980): *Macrostructures: An interdisciplinary study of global structures in discourse, interaction and cognition*. Hillsdale, NJ: Erlbaum.
- Van Dyke, Julie A. (2007): Interference effects from grammatically unavailable constituents during sentence processing. In: *Journal of Experimental Psychology: Learning, Memory and Cognition* 33, 2, S. 407-430.
- Van Dyke, Julie A./McElree, Brian (2006): Retrieval interference in sentence comprehension. In: *Journal of Memory and Language* 55, 2, S. 157-166.
- van Gompel, Roger/Majid, Asifa (2004): Antecedent frequency effects during the processing of pronouns. In: *Cognition* 90, S. 255-264.
- Vitu, Françoise/O'Regan, J. Kevin/Mittau, M. (1990): Optimal landing position in reading isolated words and continuous text. In: *Perception & Psychophysics* 47, 6, S. 583-600.
- Vitu, Françoise/O'Regan, J. Kevin/Inhoff, Albrecht W./Topolski, Richard (1995): Mindless reading: Eye-movement characteristics are similar in scanning letter strings and reading texts. In: *Perception & Psychophysics* 57, 3, S. 352-364.
- Wagner, Hildegard (1984): *Die deutsche Verwaltungssprache der Gegenwart: eine Untersuchung der sprachlichen Sonderform und ihrer Leistung*. 3. Aufl. (= Sprache der Gegenwart 9). Düsseldorf: Schwann.

- Washburne, Carleton/Morphett, Mabel Vogel (1938): Grade placement of children's books. In: *The Elementary School Journal* 38, 5, S. 355-364.
- Wason, Peter C./Reich, Shuli S. (1979): A verbal illusion. In: *Quarterly Journal of Experimental Psychology* 31, S. 591-597.
- Weaver, Warren/Shannon, Claude E. (1949): *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.
- Wein, Margarethe/Kison, Silvio (2007): „Second Life“ – „Second Lex“: Gesellschaft für deutsche Sprache agiert in Halle. In: *Scientia Halensis – Das Magazin der Martin-Luther-Universität Halle-Wittenberg*. [www.unimagazin.uni-halle.de/index.php?id=392](http://www.unimagazin.uni-halle.de/index.php?id=392) (Stand: 24.10.2012).
- Weldle, Helmut (2011): Syntaktische und referenzielle Sprachverstehensprozesse aus konnektionistischer Perspektive. Freiburg i.Br.: Freiburger Dokumentenserver (Frei-Dok). [www.freidok.uni-freiburg.de/volltexte/8325/](http://www.freidok.uni-freiburg.de/volltexte/8325/) (Stand: 4.1.2013).
- Weldle, Helmut/Konieczny, Lars/Wolfer, Sascha/Baumann, Peter/Müller-Feldmeth, Daniel (2010): Local syntactic coherences violate structural binding domains in online anaphora resolution. In: *Proceedings of the 23rd Annual Meeting of the CUNY 2010 Conference on Human Sentence Processing*. New York.
- Weldle, Helmut/Konieczny, Lars/Wolfer, Sascha/Müller-Feldmeth, Daniel/Baumann, Peter (2010): Anaphora and local coherences. In: *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.
- White, Sarah J. (2008): Eye movement control during reading: Effects of word frequency and orthographic familiarity. In: *Journal of Experimental Psychology: Human Perception and Performance* 34, 1, S. 205-223.
- Wilson, Deirde/Sperber, Dan (2005): Relevance theory. In: Horn, Laurence R./Ward, Gregory (Hg.): *The handbook of pragmatics*. Oxford: Blackwell, S. 249-290.
- Wolfer, Sascha/Hansen, Sandra/Konieczny, Lars (2012a): The effect of reference distance on anaphora processing: Evidence from a reading corpus of German texts. In: *Proceedings of the 18th AMLaP (Architectures and Mechanisms for Language Processing)*. Riva del Garda.
- Wolfer, Sascha/Hansen, Sandra/Konieczny, Lars (2012b): Familiarity and frequency disentangled: An eye-tracking corpus study with German texts. In: Bradley/Fernández/Fodor (Hg.), S. 85.
- Wolfer, Sascha/Hansen, Sandra/Konieczny, Lars (2013): Are shorter sentences always easier? Discourse level processing consequences of reformulating texts. In: *Proceedings of the 7th EST Congress*. Germersheim.
- Wolfer, Sascha/Müller-Feldmeth, Daniel/Held, Uli/Konieczny, Lars/Auer, Peter (2013): Investigating comprehensibility of intra-language translations: Introducing a reading corpus of popular science texts. In: *Proceedings of the 7th EST Congress*. Germersheim.

- Wolfer, Sascha et al. (2013): PopSci: A reading corpus of popular science texts with rich multi-level annotations. A case study. In: Holmqvist, Kenneth/Mulvey, Fiona/Johansson, Roger (Hg.): Book of Abstracts of the 17th European Conference on Eye Movements. *Journal of Eye Movement Research* 6, 3, S. 46.
- Zwaan, Rolf A. (2004): The immersed experiencer: Toward an embodied theory of language comprehension. (= *The Psychology of Learning and Motivation* 44). New York: Academic Press, S. 35-62.
- Zwaan, Rolf A./Radvansky, Gabriel A. (1998): Situation models in language comprehension and memory. In: *Psychological Bulletin* 123, 2, S. 162-185.

# **Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache Corpus Linguistics and Interdisciplinary Perspectives on Language (CLIP)**

herausgegeben von / edited by  
Marc Kupietz, Harald Lungen, Christian Mair

## **Bisher sind erschienen / Already published:**

### **Band/Vol. 1**

Marek Konopka / Jacqueline Kubczak /  
Christian Mair / František Štícha /  
Ulrich H. Waßner (Hgg.)

#### **Grammatik und Korpora 2009**

Dritte Internationale Konferenz  
2011, 604 Seiten/pages  
€[D] 108,-  
ISBN 978-3-8233-6648-5

### **Band/Vol. 2**

Vera Marková

#### **Synonyme unter dem Mikroskop**

Eine korpuslinguistische Studie  
2012, 269 Seiten/pages  
€[D] 88,-  
ISBN 978-3-8233-6689-8

### **Band/Vol. 3**

Paul Bennett / Martin Durrell /  
Silke Scheible / Richard J. Whitt (eds.)

#### **New Methods in Historical Corpora**

2013, 284 Seiten/pages  
€[D] 88,-  
ISBN 978-3-8233-6760-4

### **Band/Vol. 4**

Noah Bubenhofer / Marek Konopka /  
Roman Schneider

#### **Präliminarien einer Korpusgrammatik**

2013, 248 Seiten/pages  
€[D] 88,-  
ISBN 978-3-8233-6701-7

### **Band/Vol. 5**

Jost Gippert / Ralf Gehrke (eds.)

#### **Historical Corpora**

Challenges and Perspectives  
2015, 380 Seiten/pages  
€[D] 98,-  
ISBN 978-3-8233-6922-6

### **Band/Vol. 6**

Max Möller

#### **Das Partizip II von Experienter-Objekt- Verben**

Eine korpuslinguistische Untersuchung  
2015, 394 Seiten/pages  
€[D] 98,-  
ISBN 978-3-8233-6964-6

### **Band/Vol. 7**

Sascha Wolfer

#### **Verstehen und Verständlichkeit juristisch- fachsprachlicher Texte**

2017, 312 Seiten/pages  
€[D] 98,-  
ISBN 978-3-8233-8152-5

Juristische Texte sind schwer zu verstehen, insbesondere – aber nicht nur – für juristische Laien. Dieser Band beleuchtet diese These ausgehend von linguistischen Verständlichkeitsmodellen und kognitionswissenschaftlichen Modellen der menschlichen Textverarbeitung. Anhand von Aufzeichnungen von Blickbewegungen beim Lesen, einem sogenannten Lesekorpus, werden umfangreiche statistische Modelle berechnet. Diese geben Auskunft über Fragen psycholinguistischer Grundlagenforschung auf der Wort-, Satz- und Textebene. Ferner untersucht er, wie sich Auswirkungen von Reformulierungen auf den Verstehensprozess auswirken. Dabei stehen bekannte Komplexitätsmarker deutscher juristischer Texte im Fokus: Nominalisierungen, komplexe Nominalphrasen und syntaktisch komplexe Texte.

ISBN 978-3-8233-8152-5



9 783823 381525

