

Brucher, L., Ugen, S., Müller, C., Cuntz, M., & Weth, C. (2023). Einfluss des Faktors Wortfrequenz auf die satzinterne Großschreibung bei Fünftklässler:innen. In I. Rautenberg & M. Bangel (Eds.), *Das Verhältnis zwischen Wissen und Können beim Lesen- und Schreibenlernen* (pp. 63–82). Hohengehren: Schneider Verlag.

Einfluss von des Faktors Wortfrequenz auf die satzinterne Großschreibung bei Fünftklässler:innen

Linda Brucher, Sonja Ugen, Claire Müller, Matthias Cuntz & Constanze Weth¹

Frequenz ist ein vielbeachteter Faktor, wenn es darum geht, zu erklären, warum manche Wörter und Teilbereiche von Wörtern eher fehleranfälliger sind als andere. In regelhaften Bereichen der Orthographie haben viele Studien zudem gezeigt, dass Frequenz einen größeren Einfluss auf Richtigschreibungen haben könnte als regelbasiertes Lernen.

Der vorliegende Beitrag nimmt dies zum Ausgangspunkt und untersucht, ob der Faktor der Wortfrequenz in der satzinternen Großschreibung eine Rolle spielt. Die Arbeit umfasst zwei Forschungsfragen: (1) Sagt die Wortfrequenz von Nomen deren Großschreibung voraus, wenn man Konkretheit kontrolliert? und (2) Sagt die Frequenz von Verben und Adjektiven in Bezug auf die Frequenz der entsprechenden Nominalisierungen die Großschreibung letzterer voraus?

Für die Untersuchung wurde die Großschreibung von 185 luxemburgischen Fünftklässler:innen in Lücken- und Satzdiktaten analysiert. Die Frequenz wurde aufgrund von drei Korpora berechnet. Die Regressionsanalysen zeigen das frequente Nomen eher großgeschrieben werden als seltene Nomen. Allerdings verschwindet dieser Effekt, wenn der lexikalisch-semantische Typ (Konkreta, Abstrakta, Nominalisierungen) in die Analysen einbezogen wird. Der ausschlaggebende Faktor für die Großschreibung in dieser Studie war schließlich Konkretheit, nicht Frequenz.

¹ Der Beitrag der Co-Autor:innen des Artikels umfasst: L. Brucher* Konzept, Methode, statistische Analysen, Schreiben und Überarbeiten; S. Ugen** statistische Analysen; C. Müller** statistische Analysen; M. Cuntz*** Erstellung childLux; C. Weth* Konzept, Schreiben und Überarbeiten, Betreuung. Ein großer Dank richtet sich an Reinold Funke für seine wertvollen Kommentare. Der Beitrag ist Ergebnis des vom Luxembourg National Research Fund (FNR) geförderten Projektes (PRIDE/15/10921377).

*Institut für Mehrsprachigkeitsforschung, Universität Luxemburg

** Luxemburgisches Zentrum für Bildungstests, Universität Luxemburg

***Université de Lorraine, AgroParisTech, INRAE, UMR Silva, Nancy, France

Kontakt : constanze.weth@uni.lu

Brucher, L., Ugen, S., Müller, C., Cuntz, M., & Weth, C. (2023). Einfluss des Faktors Wortfrequenz auf die satzinterne Großschreibung bei Fünftklässler:innen. In I. Rautenberg & M. Bangel (Eds.), *Das Verhältnis zwischen Wissen und Können beim Lesen- und Schreibenlernen* (pp. 63–82). Hohengehren: Schneider Verlag.

I Einleitung

Viele Studien haben gezeigt, dass Wortfrequenz ein wichtiger Faktor bei der Sprachverarbeitung ist (Ellis 2002). Sowohl beim Erkennen als auch in der Produktion von Wörtern gibt es Unterschiede in Bezug auf die Wortfrequenz. Je häufiger ein Wort im Input vorkommt, desto schneller und richtiger wird es erkannt und produziert (Griffin & Bock 1998; Eberhard et al. 2005). Seit vielen Jahren haben Studien zum Englischen und Französischen untersucht, ob manche Aspekte der Rechtschreibung mit der Automatisierung von Regelwissen oder mit frequenz-basiertem Wissen zu tun haben (Kemp & Bryant 2003; Sandra 2012; Pacton et al. 2013). Übertragen auf das Thema dieses Beitrags ließe sich die Frage formulieren, ob regelhafte Aspekte der satzinternen Großschreibung vor allem durch expliziten Unterricht und somit explizites Regelwissen erlernt werden oder ob das Können bzw. die Schwierigkeiten dieses Bereichs im Schrifterwerb vor allem durch in der Sprachverarbeitung implizite Frequenzeffekte erklärbar ist. Die vorliegende Studie untersucht den Einfluss der Wortfrequenz im Bereich der satzinternen Großschreibung. Genauer ermittelt der Beitrag, ob der Faktor Wortfrequenz unter Einbezug des Faktors Konkretheit (Konkreta vs. Abstrakta) die Großschreibung von Nomen bei Fünftklässler:innen beeinflusst. Zudem wird untersucht, ob der Faktor Wortfrequenz ausschlaggebend für die Schreibung von Nominalisierungen ist im Vergleich zur Schreibung der homophonen Verben und Adjektive.

Verschiedene Studien (für Zweit- und Sechstklässler:innen: Wahl et al. 2017; für Dritt- und Viertklässler:innen: Günther 2007; Mangelschots et al. [eingereicht]; für Fünftklässler:innen: Bîlici et al. 2019; Brucher et al. 2020; für Dritt- bis Sechstklässler:innen: Scheele 2006; für Fünft- bis Siebtklässler:innen: Betzel 2015) untersuchten den Einfluss des lexikalisch-semantischen Typs (Konkreta, Abstrakta und Nominalisierungen) auf die Leistung in der Großschreibung. Die Rechtschreibung wurde in Diktaten (Lückentexten) getestet. Alle Studien zeigten, dass der lexikalisch-semantische Typ die korrekte Schreibweise in Bezug auf die Großschreibung von Nomen in den untersuchten Altersgruppen signifikant beeinflusst. In allen Studien wurden Konkreta häufiger großgeschrieben als Abstrakta und Abstrakta wiederum häufiger als Nominalisierungen. Auch nach einem syntaktischen Training blieb dieses Schwierigkeitsmuster bestehen (Bîlici et al. 2019; Brucher et al. 2020; Wahl et al. 2017). Die Leistung bei der Großschreibung von Nomen könnte demnach von dem Faktor Konkretheit eines Items abhängen. Allerdings kommen Konkreta in Kinderkorpora häufiger vor als Abstrakta (Schroeder et al. 2015b). Darüber hinaus konzentriert sich der Deutschunterricht auch genau auf Wörter mit diesen lexikalisch-semantischen Eigenschaften. Daher könnte man ebenfalls annehmen, dass der hohe Grad an Richtigschreibungen bei Konkreta mit deren Frequenz im Unterricht und Kindertexten im Allgemeinen zusammenhängt.

Brucher, L., Ugen, S., Müller, C., Cuntz, M., & Weth, C. (2023). Einfluss des Faktors Wortfrequenz auf die satzinterne Großschreibung bei Fünftklässler:innen. In I. Rautenberg & M. Bangel (Eds.), *Das Verhältnis zwischen Wissen und Können beim Lesen- und Schreibenlernen* (pp. 63–82). Hohengehren: Schneider Verlag.

Wahrscheinlich sind Wörter, die auf etwas sinnlich Erfahrbares referieren, im Input der Kinder häufiger als Wörter, die auf abstrakte Begriffe referieren. Dies entspricht der theoretischen Modellierung eines orthographischen Lexikons im Zwei-Wege-Modell (Coltheart 2006; Houghton & Zorzi 2003), in dem häufige Wörter gesicherter gespeichert sind und schneller und korrekter abgerufen werden können.

Die hohe Fehlerhäufigkeit bei Nominalisierungen könnte ebenfalls mit der Wortfrequenz erklärt werden, da diese in der Regel weitaus seltener im Input der Kinder vorkommen als die entsprechenden Verben und Adjektive.

Der Zusammenhang zwischen den Faktoren Frequenz und Konkretheit und deren Einfluss auf die Rechtschreibung ist bisher noch nicht untersucht worden. Eine Studie mit bildgebenden Verfahren zeigte, dass die Unterschiede im Erkennen von konkreten und abstrakten Wörtern bei Erwachsenen durch den Faktor Konkretheit, aber nicht durch Wortfrequenz erklärt werden konnten (Jessen et al. 2000). Eine weitere Studie untersuchte die Korrelation zwischen beiden Faktoren und legt nahe, dass frequente Wörter als konkreter angesehen werden als nicht frequente Wörter (Levshina 2015). Bemerkenswert ist, dass in Rechtschreibuntersuchungen in anderen Sprachen Wortfrequenz als der wesentliche Einflussfaktor angesehen wird, um Richtig- oder Falschschreibung zu erklären (zusammenfassend für Englisch, Französisch und Niederländisch, siehe Sandra 2012). Dies lässt sich damit erklären, dass Kinder eine Sensibilität für die Regelmäßigkeiten einer Sprache in Bezug auf die Häufigkeit des Auftretens der Sprach- (und Rechtschreib-) Muster im Input entwickeln (Samara & Caravolas 2014; Treiman et al. 2006). Das bedeutet, dass die Wortformen, die häufiger gesehen werden, in der Regel besser im orthographischen Lexikon gespeichert sind und mit hoher Wahrscheinlichkeit häufiger richtig geschrieben werden (Pacton & Fayol 2003).

Einige Studien untersuchten Frequenzeffekte zwischen verschiedenen Wortklassen auf die Rechtschreibung im Französischen (Largy et al. 1996; Pacton & Fayol 2003). Die Studie von Largy et al. (1996) untersuchte Frequenzeffekte in Bezug auf die Rechtschreibung homophoner Nomen und Verben. Nach Largy et al. ist das Häufigkeitsverhältnis zwischen zwei homophonen Formen entscheidend für die Leistung, wenn Schreibende mit homophonen Formen (z.B. Nomen und Verb) konfrontiert werden, da die häufigere Form stärker aktiviert wird. Genauer stellten die Autoren die Hypothese auf, dass bei homophonen Wörtern verschiedener syntaktischer Kategorien (Nomen und Verben) die Pluralform der Kategorie aktiviert wird, die frequenter ist. Ein Beispiel ist das Nomen *timbres* („Briefmarken“) mit der Pluralendung *-s* und das homophone Verb *timbrent* („frankieren“ 3. Plural) mit der Pluralendung *-nt*. Da das Nomen *timbres* häufiger als das Verb (*ils*) *timbrent* ist, wurde erwartet, dass die nominale Pluralendung auch beim Schreiben des Verbs aktiviert würde. In den Fällen, in denen die Verbform häufiger ist, wurde andererseits erwartet, dass die verbale Pluralendung auch an Nomen geschrieben würde. Die durchgeführten Experimente bestätigten diese Erwartung. Die Studie von Largy et al.

Brucher, L., Ugen, S., Müller, C., Cuntz, M., & Weth, C. (2023). Einfluss des Faktors Wortfrequenz auf die satzinterne Großschreibung bei Fünftklässler:innen. In I. Rautenberg & M. Bangel (Eds.), *Das Verhältnis zwischen Wissen und Können beim Lesen- und Schreibenlernen* (pp. 63–82). Hohengehren: Schneider Verlag.

(1996) ist für das Deutsche aufschlussreich, da bei nominalisierten Verben und Adjektiven ebenfalls die Wortfrequenz als Einflussfaktor auf die Rechtschreibung untersucht werden kann.

Die vorliegende Studie untersucht zwei Aspekte hinsichtlich Wortfrequenz. Ein Ziel der vorliegenden Studie ist es, den Faktor Frequenz als Prädiktor für die Großschreibung von Nomen in Abhängigkeit vom Faktor Konkretheit zu untersuchen.

1. Sagt die Wortfrequenz von Nomen die Großschreibung von Nomen voraus, wenn man Konkretheit kontrolliert?

Dieser Frage liegt die Annahme zugrunde, dass (a) häufige Nomen häufiger korrekt geschrieben werden als seltene Nomen, unabhängig davon, ob sie dem lexikalisch-semantic Typ Konkreta, Abstrakta oder Nominalisierungen angehören. Außerdem wird angenommen (b), dass der Faktor Frequenz eine geringere Rolle spielt, wenn Konkretheit einbezogen wird. Der Effekt der Wortfrequenz auf die Großschreibung von Nomen könnte bei Konkreta geringer sein als bei Abstrakta und Nominalisierungen.

Ein zweites Ziel der Studie ist, die Auswirkung der Wortfrequenz bei Nominalisierungen in enger Anlehnung an die Forschung zu homophonen Intrusionsfehlern im Französischen (Largy et al. 1996) zu untersuchen.

2. Sagt die Frequenz der nicht-nominalisierten homophonen Form die Großschreibung bei Nominalisierungen voraus?

In Anlehnung an Largy et al. (1996) liegt dieser Frage die Annahme zugrunde, dass bei zwei konkurrierenden Wortformen (z.B. *schwimmen/Schwimmen*) Schreibende die häufigere orthographische Form verwenden.

Zur Beantwortung der beiden Forschungsfragen wurden in der vorliegenden Studie Sekundäranalysen zu den Daten einer Interventionsstudie (Brucher et al. 2020) durchgeführt und die Häufigkeit als Prädiktor für die Rechtschreibleistung von Fünftklässler:innen hinsichtlich der Großschreibung von Nomen untersucht. Für die Informationen zu Frequenz wurden verschiedene lexikalische Datenbanken verwendet. Die Datenbank Wortschatz Deutsch aus der *Corpora Collection* der Universität Leipzig (LCC 2011), die für die Kontrolle der Häufigkeit in der Testkonstruktion verwendet wurde, basiert auf deutschen Zeitungen und stellt ein Erwachsenenkorpora dar. Da diese Studie darauf abzielt, die Häufigkeit als Prädiktor für die Rechtschreibleistung von Nomen bei Grundschulkindern zu untersuchen, haben wir die Testitems *a posteriori* mit den Frequenzangaben in zwei Kinderkorpora

Brucher, L., Ugen, S., Müller, C., Cuntz, M., & Weth, C. (2023). Einfluss des Faktors Wortfrequenz auf die satzinterne Großschreibung bei Fünftklässler:innen. In I. Rautenberg & M. Bangel (Eds.), *Das Verhältnis zwischen Wissen und Können beim Lesen- und Schreibenlernen* (pp. 63–82). Hohengehren: Schneider Verlag.

abgeglichen. Die erste Datenbank für Kinder ist childLex (Schroeder et al. 2015b), die auf deutscher Literatur für 6- bis 12-jährige Kinder basiert. Zusätzlich wurde ein weiteres Korpus gemäß des childLex-Verfahrens für den luxemburgischen Kontext, wo die Studie verortet ist, erstellt: childLux. Grund für die Erstellung dieses Korpus ist, dass Deutsch in Luxemburg einen viel geringeren Stellenwert hat als in Deutschland, der Schweiz oder Österreich und für die meisten Kinder nur im Deutschunterricht präsent ist. Das Korpus childLux basiert deswegen auf den Schulbüchern der Kinder. Die Studie ist explorativ, da die Daten im Nachhinein in Bezug auf die vorliegenden Forschungsfragen hin analysiert wurden. Die Ergebnisse sind dadurch nicht verallgemeinerbar, geben aber Hinweise zum Stellenwert von Frequenz im Bereich der Großschreibung.

II Material und Methode

1 Teilnehmende

Eine Teilstichprobe von 185 Schüler:innen (92 Jungen und 93 Mädchen; Durchschnittsalter 11 Jahre), die an einer Interventionsstudie teilnahmen (Brucher et al. 2020), wurde in dieser Studie berücksichtigt. Alle Schüler:innen besuchten die fünfte Klasse in Luxemburg und wurden aus fünfzehn verschiedenen Schulklassen rekrutiert. Die Schulen befanden sich in Bezirken mit einem unterdurchschnittlichen sozioökonomischen Status (HISEI zwischen 35 und 45). Die Genehmigung für die Studie wurde von der Forschungsethikkommission der Universität Luxemburg und dem luxemburgischen Zentrum für Datenschutz (CNPD) erteilt. Es wurden nur die Daten der Kinder analysiert, die die Zustimmung der Erziehungsberechtigten hatten. Alle Teilnehmer:innen besuchten das luxemburgische Schulsystem seit der ersten Klasse.

2 Materialien und Aufbau des Tests

Zur Beurteilung der Großschreibung von Nomen wurde ein Test konstruiert. Er bestand aus zwei verschiedenen Aufgaben in Form von Diktaten. Bei der Lückentextaufgabe mussten die Schüler:innen 48 Test- und 16 Kontrollwörter einsetzen, um die Sätze zu vervollständigen. Kontrollwörter dienten zur Kontrolle von

Brucher, L., Ugen, S., Müller, C., Cuntz, M., & Weth, C. (2023). Einfluss des Faktors Wortfrequenz auf die satzinterne Großschreibung bei Fünftklässler:innen. In I. Rautenberg & M. Bangel (Eds.), *Das Verhältnis zwischen Wissen und Können beim Lesen- und Schreibenlernen* (pp. 63–82). Hohengehren: Schneider Verlag.

einem übermäßigen Gebrauch von Großschreibung. Eine Analyse der Kontrollwörter ergab, dass Schüler:innen mehr als 90% der Kontrollwörter richtig mit einem kleinen Anfangsbuchstaben schrieben (Brucher et al. 2020).

Für die Testdurchführung erhielt jedes Kind ein Heft mit vorgedruckten Sätzen mit einer Lücke, die im folgenden Beispiel unterstrichen ist: *Der Clown packt den Koffer*. Bei der Satzaufgabe mussten die Schüler:innen 24 unabhängige Sätze schreiben, die ebenfalls diktiert wurden und jeweils ein Testwort enthielten.

Schüler:innen, die die Testwörter mit einem großen Anfangsbuchstaben schrieben, erzielten jeweils einen Punkt. Eine Varianzanalyse (ANOVA) zur korrekten Großschreibung mit Aufgabentyp als Inner-Subjekt-Faktor zeigte keinen signifikanten Leistungsunterschied zwischen den Aufgaben ($F(1, 178) = 2,57, p > 0,05, \eta_p^2 = 0,014$).

Die Testwörter umfassten Konkrete, Abstrakta und Nominalisierungen, die Kontrollwörter Verben im Infinitiv und Adjektive. Die Testwörter waren in vier Nominalphrasen (NP) Strukturen eingebettet: Determinans, Nomen (DN); Determinans, Adjektiv, Nomen (DAN); Adjektiv, Nomen (AN) und Nomen (N). Darüber hinaus wurde die Frequenz über die verschiedenen lexikalisch-semantischen Typen mit den Frequenzklassen der LCC (2011) kontrolliert. Für alle Kombinationen von lexikalisch-semantischen Typen und NP-Strukturen wählten wir jeweils 6 Wörter aus verschiedenen Frequenzklassen aus, so dass Wörter mit niedriger, mittlerer und hoher Frequenz vertreten waren. Wir wählten Wörter aus den Frequenzklassen 7, 9, 10, 11, 12 und 13 (die Frequenzklasse 0 umfasst die häufigsten Wörter; Wörter der Frequenzklasse 1 sind halb so häufig wie Wörter der Frequenzklasse 0; die Frequenzklasse 20 umfasst sehr seltene Wörter). Hinsichtlich der morphologischen Merkmale wurde die Anzahl der (nicht) pluralfähigen Nomen ausgeglichen. Zusätzlich wurden Faktoren wie die Position der NP im Satz (initial vs. nicht initial) und die Funktion der NP (Subjekt vs. Objekt) innerhalb jeder Kombination von lexikalisch-semantischem Typ und NP-Struktur ausgeglichen. Die interne Konsistenz der Items wurde mit Hilfe von Cronbachs Alpha gemessen ($\alpha = 0,92$ für Konkrete; $\alpha = 0,88$ für Abstrakta; $\alpha = 0,84$ für Nominalisierungen).

3 Lexikalische Datenbanken

LCC (2011). Zum Zeitpunkt der Testkonstruktion entschieden wir uns, die LCC (2011) zu verwenden, die auf Zeitungsartikeln für Erwachsene basiert, da keine

Brucher, L., Ugen, S., Müller, C., Cuntz, M., & Weth, C. (2023). Einfluss des Faktors Wortfrequenz auf die satzinterne Großschreibung bei Fünftklässler:innen. In I. Rautenberg & M. Bangel (Eds.), *Das Verhältnis zwischen Wissen und Können beim Lesen- und Schreibenlernen* (pp. 63–82). Hohengehren: Schneider Verlag.

Datenbank für luxemburgische Schüler:innen zur Verfügung stand. Die LCC (2011) umfasst insgesamt 720 Millionen Token und 8 Millionen Types.

Die Verwendung eines Korpus für Erwachsene zur Kontrolle der Frequenz von Wörtern bei Kindern könnte jedoch problematisch sein (Schroeder et al. 2015a). Schroeder et al. (2015a) analysierten, wie die Frequenzklassen von zwei Erwachsenenkorpora (DWDS und CELEX) und einem Kinderkorpus (childLex) korrelieren. Während die Korrelationen zwischen den drei Korpora in der höchsten und der niedrigsten Frequenzklasse hoch waren, war die Korrelation zwischen Erwachsenen- und Kinderkorpora in den mittleren Frequenzklassen wesentlich geringer. Schroeder et al. (2015a) kamen zu dem Schluss, dass die Vertrautheit mit hochfrequenten Wörtern bei Erwachsenen und Kindern vergleichbar ist, während die Vertrautheit mit nicht hochfrequenten Wörtern sich deutlich unterscheidet. Ein weiterer Nachteil bei der Verwendung der LCC (2011) ist, dass bei der Wortzählung zwar die Groß- und Kleinschreibung berücksichtigt, aber keine Unterscheidung zwischen Wortklassen vorgenommen wird. Dies erschwert die Zählung für Nomen und insbesondere für Nominalisierungen, da die Frequenzschätzung für ein Wort mit großem Anfangsbuchstaben (z.B. *Grün*) Nomen (z.B. *das schöne Grün*), Eigennamen (z.B. *Frau Grün*) und Adjektive, die am Anfang von Sätzen vorkommen (z.B. *Grün ist die Wand*), umfasst.

Um den Einfluss der Frequenz auf die Großschreibung von Nomen im Deutschen zu untersuchen, haben wir die Frequenzen auf der Grundlage von zwei Korpora, die auf Kindertexten basieren und Wortklassenunterscheidungen enthalten, neu analysiert: childLex und childLux.

childLex (2015b). Die lexikalische Datenbank childLex (Schroeder et al. 2015b) basiert auf deutscher Literatur (Kinderbücher und Schulbücher) für 6- bis 12-jährige Kinder. Sie umfasst 7,8 Millionen Token, die sich auf 180 000 Types und 120 000 Lemmata verteilen. Die Einträge unterscheiden zwischen Groß- und Kleinschreibung, sind lemmatisiert und syntaktisch mit *Part-of-Speech-Tagging* annotiert, was es erlaubt, auch nominalisierte Verben und Adjektiven als Wortkategorie zu erfassen (Schroeder et al. 2015b).

childLux. Das zweite Kinderkorpus wurde in Analogie zu childLex (Schroeder et al. 2015b) erstellt und daher childLux genannt. Es wurde erstellt, um eine genauere Annäherung an die Frequenzen für luxemburgische Fünftklässler:innen zu erhalten. Das Korpus basiert auf den vom luxemburgischen Bildungsministerium empfohlenen Schulbüchern. Die linguistische Verarbeitung (Tokenisierung und Lemmatisierung) für dieses Korpus lehnt sich eng an das Verfahren an, das für die lexikalische Datenbank childLex verwendet wurde (Jurish & Würzner 2013; Schroeder et al.

Brucher, L., Ugen, S., Müller, C., Cuntz, M., & Weth, C. (2023). Einfluss des Faktors Wortfrequenz auf die satzinterne Großschreibung bei Fünftklässler:innen. In I. Rautenberg & M. Bangel (Eds.), *Das Verhältnis zwischen Wissen und Können beim Lesen- und Schreibenlernen* (pp. 63–82). Hohengehren: Schneider Verlag.

2015b). Die Datenbank umfasst 71 667 Token. Types und Lemmata wurden nur für die untersuchten Test- und Kontrollwörter berechnet.

4 Erstellung des childLux-Korpus

Grund für die Erstellung von childLux war es, zu kontrollieren, ob die Wortfrequenzen im Input luxemburgischer Kinder sich von denen in anderen DACHL-Ländern unterscheiden. Die Textauswahl für das Korpus stellt das Mindestmaß an deutschem Input dar, dem die Schüler:innen im luxemburgischen Schulsystem begegnen, d.h. den Lehrbüchern im Deutschunterricht.

Linguistische Auswahl. Das childLux Korpus umfasst die vom luxemburgischen Bildungsministerium (MENJE) herausgegebenen deutschen Lehrbücher von Klasse 1 bis 5 (*Mila 1, Mila 2, Sprachfuchs 3 Band 1, Sprachfuchs 3 Band 2, Sprachfuchs 4 Band 1*). Ziel war eine Annäherung an die Frequenz von Wörtern, denen luxemburgische Schüler:innen begegnen. Das einzige Kriterium für die Aufnahme eines Buches in das Korpus war, dass es sich um ein vom Bildungsministerium empfohlenes Schulbuch für den Deutschunterricht handelte.

Linguistische Verarbeitung. Die digitalen Versionen der Lehrbücher wurden verwendet, um die Wortformen zu isolieren. Der Text wurde mit Hilfe von pdftotext (Glyph & Cog, LLC, Petaluma, CA, USA) aus den PDF-Versionen extrahiert und mit eigens dafür entwickelten Skripten bereinigt. Der Webservice CAB (www.deutschestextarchiv.de/cab) des Deutschen Textarchivs wurde dann verwendet, um den bereinigten Rohtext zu tokenisieren, die Token morphologisch zu analysieren, Typen zu identifizieren und die Token zu lemmatisieren (Jurish 2012).

5 Verfahren zum Vergleichen der drei Korpora

Frequenzschätzung. Um die drei Korpora (LCC, childLex, childLux) zu vergleichen, haben wir zunächst logarithmierte Werte für alle Korpora berechnet (in Analogie zu childLex von Schroeder et al. 2015b): Die jeweilige Anzahl der Typen und Lemmata im Korpus wurde bestimmt. Normalisierte Frequenzen wurden berechnet, indem die Anzahl durch die Gesamtzahl der Token (in Millionen) im Korpus geteilt wurde. Der

Brucher, L., Ugen, S., Müller, C., Cuntz, M., & Weth, C. (2023). Einfluss des Faktors Wortfrequenz auf die satzinterne Großschreibung bei Fünftklässler:innen. In I. Rautenberg & M. Bangel (Eds.), *Das Verhältnis zwischen Wissen und Können beim Lesen- und Schreibenlernen* (pp. 63–82). Hohengehren: Schneider Verlag.

Logarithmus zur Basis 10 der normalisierten Frequenzen wurde für die weiteren Analysen verwendet. Die LCC hat beispielsweise 720 421 868 Token. Ein Token, das 720 mal in der LCC vorkommt, hat demnach einen Log-Wert von 0, ein Token, das 7,204 mal vorkommt, einen Log-Wert von 1, ein Token, das 72,042 mal vorkommt, einen Log-Wert von 2 und ein Token, das 720,422 mal vorkommt, einen Log-Wert von 3. Ein Log-Wert von 6 ist der maximale Log-Wert. Um einen besseren Vergleich zwischen den Korpora zu ermöglichen, wurden 6 Frequenzklassen definiert.

6 Erstellen der Prädiktor-Variablen

Für *childLex* und *childLux* wurden sechs Frequenzklassen definiert, die auf den berechneten Log-Werten der normalisierten Frequenzen des jeweiligen Korpus basieren (siehe Tabelle 1).

Neben den Prädiktorvariablen *Frequenz childLex* und *Frequenz childLux* gibt es den Prädiktor *lexikalisch-semantischer Typ*, der die drei lexikalisch-semantischen Typen Konkreta, Abstrakta und Nominalisierungen unterscheidet.

Frequenzklasse	Log-Werte
1	0-0,5
2	>0,5-1
3	>1-1,5
4	>1,5-2
5	>2-2,5
6	>2,5

Tabelle 1: *Frequenzklassen und Log-Werte für childLex und childLux*

III Ergebnisse

Zunächst untersuchten wir, welche Frequenzschätzungen der drei Korpora am stärksten mit der korrekten Großschreibung von Nomen korrelieren. Hierfür haben wir für jedes Korpus die Korrelation zwischen korrekter Großschreibung der Schüler:innen und den Frequenzklassen berechnet. Hier wählten wir die Frequenzen der Wortform (Type) aus, da für die LCC nur diese Frequenzen verfügbar waren. Bei

Brucher, L., Ugen, S., Müller, C., Cuntz, M., & Weth, C. (2023). Einfluss des Faktors Wortfrequenz auf die satzinterne Großschreibung bei Fünftklässler:innen. In I. Rautenberg & M. Bangel (Eds.), *Das Verhältnis zwischen Wissen und Können beim Lesen- und Schreibenlernen* (pp. 63–82). Hohengehren: Schneider Verlag.

der LCC ist die Korrelation zwischen korrekter Großschreibung (in Prozent) und Frequenzklasse signifikant ($r = 0,267$; $p < 0,05$; $n = 72$). Nach Cohen (1992) steht die Korrelation 0,1 für einen kleinen Effekt, 0,3 für einen mittleren und 0,5 für einen großen Effekt. Beim childLex ist die Korrelation zwischen korrekter Großschreibung und Frequenz des Nomens ebenfalls signifikant ($r = 0,49$; $p < 0,001$; $n = 72$). Auch beim childLux korrelieren korrekte Großschreibung von Nomen und Frequenz des Nomens signifikant ($r = 0,46$; $p < 0,001$; $n = 72$). Die signifikanten Korrelationen zeigen, dass je höher die Frequenzklasse (und somit die Frequenz eines Nomens) ist, desto höher ist auch die korrekte Großschreibung eines Nomens. Da die Frequenzen von childLex und childLux stärker mit der korrekten Großschreibung korrelierten als die der LCC, wurden weitere Regressionsanalysen getrennt für die beiden Kinderkorpora durchgeführt.

Zieht man Parallelen zu der Forschung zum Französischen und Niederländischen (Largy et al. 1996; Pacton & Fayol 2003; Sandra 2012), könnte man vermuten, dass die Großschreibung des Nomens durch das allgemeine Vorkommen in Texten, die die Lernenden bisher gelesen haben, beeinflusst wird, unabhängig von der Wortform, in der es auftritt: Je häufiger ein Wort in einem Text großgeschrieben vorkommt, desto häufiger wird es von den Proband:innen korrekt großgeschrieben. Genauer gesagt ist die Berücksichtigung der Frequenz von Lemmata relevant für die Untersuchung der Großschreibung. Folglich basieren die Analysen in beiden Korpora auf den Frequenzen der Lemmata.

1 Analyse zur Forschungsfrage 1

Hinsichtlich unserer ersten Forschungsfrage, ob der Faktor Frequenz unter Berücksichtigung des Faktors lexikalisch-semantischer Typ die korrekte Großschreibung von Nomen vorhersagt, haben wir eine Regressionsanalyse in drei Schritten durchgeführt.

1.1 childLux

In einem ersten Schritt untersuchten wir, wie die Lemmafrequenzen von childLux die Leistung hinsichtlich der Großschreibung von Nomen vorhersagen. Die Regressionsanalyse zeigt, dass die Lemmafrequenzen die Großschreibung von Nomen signifikant vorhersagen ($p < 0,001$). Das Modell erklärt einen signifikanten

Brucher, L., Ugen, S., Müller, C., Cuntz, M., & Weth, C. (2023). Einfluss des Faktors Wortfrequenz auf die satzinterne Großschreibung bei Fünftklässler:innen. In I. Rautenberg & M. Bangel (Eds.), *Das Verhältnis zwischen Wissen und Können beim Lesen- und Schreibenlernen* (pp. 63–82). Hohengehren: Schneider Verlag.

Anteil der Varianz der Großschreibung ($r^2 = 0,25$; $F(1, 70) = 23,07$; $p < 0,001$). Ein Anstieg um eine Einheit der Frequenzklasse in childLux führt zu einem Anstieg der korrekten Großschreibung um 5,26 Prozentpunkte. Dies bedeutet, dass Nomen, die einer höheren Frequenzklasse angehören, und denen Kinder somit häufiger begegnen, häufiger korrekt großgeschrieben werden.

Im zweiten Schritt nahmen wir den lexikalisch-semantischen Typ als Prädiktor in das Regressionsmodell auf. Es war zu beobachten, dass der lexikalisch-semantische Typ einen erheblichen Anteil der zusätzlichen Varianz der korrekten Großschreibung erklärt ($r^2 = 0,65$; bereinigtes $r^2 = 0,63$). Es zeigte sich ebenfalls, dass sich die durch die Frequenz und den lexikalisch-semantischen Typ erklärte Varianz bis zu einem gewissen Grad überschneidet, da bei Schritt 2 die Frequenz kein signifikanter Prädiktor mehr war ($p = 0,1$).

Im Vergleich zu Konkreta wurden Abstrakta und Nominalisierungen seltener richtig geschrieben: korrekte Großschreibung in 47,7% der Fälle für Konkreta im Vergleich zu korrekter Großschreibung in 28,5% bzw. 14% der Fälle für Abstrakta und Nominalisierungen.

In Schritt 3 haben wir zusätzlich Interaktionsterme zwischen lexikalisch-semantischen Typen und Frequenzklassen in das Modell eingefügt. Die Interaktionsterme waren nicht signifikant, was darauf hindeutet, dass der Einfluss der Frequenz auf die Großschreibung nicht durch den lexikalisch-semantischen Typ bestimmt wird. Frequenz ist kein Prädiktor für die Großschreibung innerhalb der verschiedenen lexikalisch-semantischen Typen. Dies bedeutet, dass, beispielsweise innerhalb der Kategorie der konkreten Nomen, die Frequenzklasse die korrekte Großschreibung nicht vorhersagt. Das gleiche gilt für die Kategorie der abstrakten Nomen und Nominalisierungen. Die Zusammenfassung der Regressionsanalyse ist in Tabelle 1 zu finden.

Brucher, L., Ugen, S., Müller, C., Cuntz, M., & Weth, C. (2023). Einfluss des Faktors Wortfrequenz auf die satzinterne Großschreibung bei Fünftklässler:innen. In I. Rautenberg & M. Bangel (Eds.), *Das Verhältnis zwischen Wissen und Können beim Lesen- und Schreibenlernen* (pp. 63–82). Hohengehren: Schneider Verlag.

	B	SE B	Signifikanz
Schritt 1²			
(Konstante)	17,42	4,16	< 0,001
childLux_Lemmafrequenz	5,26	1,09	< 0,001
Schritt 2²			
(Konstante)	47,65	4,61	< 0,001
childLux_Lemmafrequenz	1,49	0,88	0,10
Abstrakta	-19,12	3,56	< 0,001
Nominalisierung	-33,70	3,87	< 0,001
Schritt 3²			
(Konstante)	50,48	10,28	< 0,001
childLux_Lemmafrequenz	0,86	2,22	0,70
Abstrakta	-25,15	11,37	0,03
Nominalisierung	-34,43	11,058	0,003
childLux_Lemmafrequenz: Abstrakta	1,63	2,58	0,53
childLux_Lemmafrequenz: Nominalisierung	-0,27	2,63	0,92

Tabelle 2: Zusammenfassung der Regressionsanalyse zur Vorhersage von Großschreibung von Nomen durch den Faktor Frequenz (Lemmata aus childLux) (Schritt 1); Vorhersage von Großschreibung von Nomen durch den Faktor Frequenz und den Faktor lexikalisch-semantischer Typ (Konkreta, Abstrakta, Nominalisierung) (Schritt 2); Vorhersage von Großschreibung von Nomen durch den Faktor Frequenz, den Faktor lexikalisch-semantischer Typ und den Interaktionsterm zwischen Frequenz und lexikalisch-semantischem Typ (Schritt 3).

1.2 childLex

Wie für childLux untersuchten wir, wie die Lemmafrequenzen von childLex die korrekte Großschreibung von Nomen vorhersagen. Die Regressionsanalyse zeigt, dass auch hier die Lemmafrequenzen die Großschreibung von Nomen signifikant vorhersagen ($p < 0,001$). Das Modell erklärt einen signifikanten Anteil der Varianz der Ergebnisse bezüglich der Großschreibung von Nomen ($r^2 = 0,23$; $F(1, 70) = 20,57$; $p < 0,001$). Ein Anstieg um eine Einheit in der Frequenzklasse führt zu einem Anstieg der korrekten Großschreibung um 6,42 Prozentpunkte. Im zweiten Schritt nahmen wir den lexikalisch-semantischen Typ als Prädiktor in das Regressionsmodell auf. Die Analyse zeigte, dass der lexikalisch-semantische Typ einen erheblichen

² Anmerkung. Schritt 1: $r^2 = 0,25$; $F(1, 70) = 23,07$; $p < 0,001$; Schritt 2: bereinigtes $r^2 = 0,63$; $F(3, 68) = 41,23$; $p < 0,001$; Schritt 3: bereinigtes $r^2 = 0,62$; $F(5, 66) = 24,61$; $p < 0,001$

Brucher, L., Ugen, S., Müller, C., Cuntz, M., & Weth, C. (2023). Einfluss des Faktors Wortfrequenz auf die satzinterne Großschreibung bei Fünftklässler:innen. In I. Rautenberg & M. Bangel (Eds.), *Das Verhältnis zwischen Wissen und Können beim Lesen- und Schreibenlernen* (pp. 63–82). Hohengehren: Schneider Verlag.

Anteil der Varianz der Großschreibung erklärt ($r^2=0,63$; bereinigtes $r^2=0,61$). Es ist festzustellen, dass die Frequenz nach der Aufnahme des lexikalisch-semantic Typs in das Modell kein signifikanter Prädiktor mehr ist ($p > 0,05$). In Schritt 3 fügten wir zusätzlich einen Interaktionsterm hinzu. Die Interaktionsterme waren nicht signifikant, was darauf hindeutet, dass der Einfluss der Frequenz auf die Großschreibung nicht durch den lexikalisch-semantic Typ bestimmt wird.

	B	SE B	Signifikanz
Schritt 1³			
(Konstante)	14,41	4,96	0,005
childLex_Lemmafrequenz	6,42	1,42	< 0,001
Schritt 2³			
(Konstante)	53,12	5,69	< 0,001
childLex_Lemmafrequenz	0,29	1,23	0,815
Abstrakta	-20,78	3,59	< 0,001
Nominalisierung	-36,32	4,26	< 0,001
Schritt 3³			
(Konstante)	64,75	10,11	< 0,001
childLex_Lemmafrequenz	-2,50	2,36	0,293
Abstrakta	-39,40	12,15	0,002
Nominalisierung	-47,09	11,41	< 0,001
childLex_Lemmafrequenz: Abstrakta	4,88	3,02	0,111
childLex_Lemmafrequenz: Nominalisierung	2,39	3,23	0,463

Tabelle 3: Zusammenfassung der Regressionsanalyse zur Vorhersage von Großschreibung von Nomen durch den Faktor Frequenz (Lemmata aus childLex) (Schritt 1); Vorhersage von Großschreibung von Nomen durch den Faktor Frequenz und den Faktor lexikalisch-semantic Typ (Konkreta, Abstrakta, Nominalisierung) (Schritt 2); Vorhersage von Großschreibung von Nomen durch den Faktor Frequenz, den Faktor lexikalisch-semantic Typ und den Interaktionsterm zwischen Frequenz und lexikalisch-semantic Typ (Schritt 3).

³ Anmerkung. Schritt 1: $r^2 = 0,23$; $F(1, 70) = 20,57$; $p < 0,001$; Schritt 2: angepasstes $r^2 = 0,61$; $F(3, 68) = 38,69$; $p < 0,001$; Schritt 3: $r^2 = 0,62$; $F(5, 66) = 24,61$; $p < 0,001$

Brucher, L., Ugen, S., Müller, C., Cuntz, M., & Weth, C. (2023). Einfluss des Faktors Wortfrequenz auf die satzinterne Großschreibung bei Fünftklässler:innen. In I. Rautenberg & M. Bangel (Eds.), *Das Verhältnis zwischen Wissen und Können beim Lesen- und Schreibenlernen* (pp. 63–82). Hohengehren: Schneider Verlag.

2 Analyse zur Forschungsfrage 2

Zur Beantwortung unserer zweiten Forschungsfrage, ob die Frequenz der getesteten Verben und Adjektive in verbalem bzw. adjektivischem Gebrauch die Richtigschreibung der entsprechenden Nominalisierungen voraussagt, überprüften wir zunächst, ob nicht-nominalisierte Wortformen häufiger im childLux vorkommen als nominalisierte Wortformen. Ein *t*-Test bestätigte, dass die durchschnittliche Frequenzklasse der nicht-nominalisierten Formen ($M = 4,91$; $SD = 1,22$) höher war als die durchschnittliche Frequenzklasse der Nominalisierungen ($M = 2,36$; $SD = 1,75$). Der Unterschied ist mit $t(18) = 3,96$, $p < 0,001$ signifikant.

Zusätzlich untersuchten wir die Rechtschreibleistungen (in Prozent) bezüglich der nominalisierten Formen und ihrer jeweiligen nicht-nominalisierten Formen (Verben und Adjektive) mit einem *t*-Test für gepaarte Stichproben. Die Ergebnisse zeigen, dass die nicht-nominalisierten Formen häufiger richtig geschrieben wurden ($M = 96,31$; $SD = 2,58$) als die nominalisierten Formen ($M = 19,97$; $SD = 17,29$). Diese Ergebnisse sind mit $t(10) = 13,07$ $p < 0,001$ signifikant.

Außerdem führten wir eine Regressionsanalyse durch, um herauszufinden, ob die Frequenz der nicht-nominalisierten Form die korrekte Großschreibung von Nominalisierungen vorhersagt. Wir analysierten dazu die korrekte Großschreibung bei der nominalisierten Form, bezogen aber sowohl die Frequenzklassen der nicht-nominalisierten als auch die der nominalisierten Formen in das Modell ein. Die Analyse zeigte, dass kein signifikanter Anteil der Varianz durch das Modell erklärt wurde ($F < 1$).

Schließlich führten wir eine weitere Regressionsanalyse durch und nahmen das Verhältnis der Frequenzklasse der nominalisierten Formen zu der Frequenzklasse der nicht-nominalisierten Form in das Modell auf. Auch dieses Modell konnte keinen signifikanten Anteil der Varianz erklären ($F < 1$).

IV Diskussion

In der Studie wurde untersucht, inwieweit die Frequenz von Wörtern die Rechtschreibleistung von Fünftklässler:innen in Luxemburg bei der Großschreibung von Nomen vorhersagt. Zunächst wurde der Frage nachgegangen, ob die Frequenz die Großschreibung von Nomen an sich vorhersagt. Anschließend untersuchten wir, ob die Frequenz die Großschreibung von Nomen auch dann noch vorhersagt, wenn man den Prädiktor lexikalisch-semantischer Typ (Konkreta, Abstrakta und Nominalisierungen) hinzufügt. Im Rahmen einer zweiten Forschungsfrage

Brucher, L., Ugen, S., Müller, C., Cuntz, M., & Weth, C. (2023). Einfluss des Faktors Wortfrequenz auf die satzinterne Großschreibung bei Fünftklässler:innen. In I. Rautenberg & M. Bangel (Eds.), *Das Verhältnis zwischen Wissen und Können beim Lesen- und Schreibenlernen* (pp. 63–82). Hohengehren: Schneider Verlag.

konzentrierten wir uns auf den Effekt der Frequenz auf die Großschreibung von Nominalisierungen.

Wir haben drei verschiedene Korpora, einen Erwachsenenkorporus (*LCC*) und zwei Kinderkorpora, betrachtet, um die Frequenzen unserer Testitems zu schätzen. Bei den beiden Kinderkorpora *childLex* (Schroeder et al. 2015a) und *childLux* war die Korrelation zwischen Frequenz und Rechtschreibleistung bei der Großschreibung von Nomen ähnlich hoch. Die Frequenzen aus dem *childLux*-Korpus, das den sprachlichen Input luxemburgischer Schulbücher widerspiegelt, korrelierten nicht höher mit der Großschreibung als die Frequenzen aus dem *childLex*-Korpus, das die von Kindern gelesene Literatur auf Deutsch wiedergibt. Die Korrelation zwischen den Frequenzen der Erwachsenenkorpora und der korrekten Großschreibung fiel jedoch geringer aus.

Der Vergleich der Korpora in Bezug auf die Rechtschreibleistungen von Fünftklässler:innen führt die Relevanz der Korpusauswahl vor Augen. Da der sprachliche Input von Kindern sich von Erwachsenen unterscheidet, scheint ein Kinderkorpus geeigneter für die Analysen zu sein. Ein weiteres Ergebnis liegt im Vergleich von *childLex* und *childLux*. Die Motivation für die Erstellung von *childLux* war die Annahme, dass in den luxemburgischen Schulbüchern andere Wörter bzw. Wortformen häufiger vorkommen könnten als in dem viel breiter angelegten Korpus *childLex*, der auf deutschen Kinderbüchern basiert. Dies wäre für die Wortfrequenzen relevant gewesen, da in Luxemburg aufwachsende Kinder mit Deutsch häufig nur im Deutschunterricht in Kontakt kommen, jedoch sonst noch mit Luxemburgisch und weiteren Sprachen. Zumindest in Bezug auf die Frequenz von Nomen scheint *childLex* für die Einschätzung der Wortfrequenzen im Luxemburger Kontext aber angemessen zu sein.

Die Regressionsanalysen wurden mit den beiden Kinderkorpora durchgeführt. Bei der Verwendung von *childLux*-Frequenzen im Vergleich zu *childLex*-Frequenzen konnten wir keinen auffälligen Unterschied in den Regressionsanalysen feststellen. Die Ergebnisse der Regressionsanalysen mit *childLux*- bzw. *childLex*-Frequenzen bestätigen die gleichen Tendenzen. Es wäre denkbar, dass die Korrelation mit *childLux* höher ausgefallen wäre, wenn das Korpus größer wäre. Da das *childLux*-Korpus jedoch auf einer verhältnismäßig kleinen Datenbasis beruht, haben die für dieses Korpus ermittelten Frequenzen eine geringere Reliabilität als die, die für *childLex* ermittelt wurden.

Hinsichtlich unserer ersten Forschungsfrage hat die Regressionsanalyse gezeigt, dass Frequenz kein signifikanter Prädiktor für die Rechtschreibleistung bei der Großschreibung von Nomen ist, wenn wir den Faktor „lexikalisch-semantischer Typ“ (Konkreta, Abstrakta, Nominalisierungen) kontrollieren. Im Gegensatz dazu ist der lexikalisch-semantische Typ ein signifikanter Prädiktor für die Rechtschreibleistung

Brucher, L., Ugen, S., Müller, C., Cuntz, M., & Weth, C. (2023). Einfluss des Faktors Wortfrequenz auf die satzinterne Großschreibung bei Fünftklässler:innen. In I. Rautenberg & M. Bangel (Eds.), *Das Verhältnis zwischen Wissen und Können beim Lesen- und Schreibenlernen* (pp. 63–82). Hohengehren: Schneider Verlag.

bei der Großschreibung. Der lexikalisch-semantiche Typ erklärt einen signifikanten Anteil der Varianz in der Rechtschreibleistung.

Der Faktor Konkretheit scheint hier ausschlaggebend zu sein. Dass Konkretheit (vs. Abstraktheit) einen zentralen Stellenwert beim Verarbeiten von Wörtern hat und dass diese im Deutschen ausschlaggebender ist als der Faktor Frequenz, haben schon Arbeiten im Bereich der Worterkennung gezeigt (Jessen et al. 2000). Dieser Befund weist darauf hin, dass die in vielen empirischen Studien vorgenommene Einteilung in Konkreta und Abstrakta sowie Nominalisierungen sinnvoll ist. Ob das bessere Abschneiden bei Konkreta unabhängig von der relativ hohen Frequenz von Konkreta im Unterricht ist, müsste untersucht werden.

Unsere zweite Forschungsfrage konzentrierte sich auf die Auswirkung der Häufigkeit von Nominalisierungen und den entsprechend nominalisierten Verben und Adjektiven. Diese Fragestellung stand in engem Zusammenhang mit der Forschung über Intrusionsfehler bei Homophonen im Französischen (Largy et al. 1996). In den von uns verwendeten Korpora erwiesen sich die Frequenzen der nicht-nominalisierten Formen als signifikant höher als die Frequenzen der nominalisierten Formen. Außerdem zeigten die Ergebnisse, dass die Rechtschreibkorrektheit für die Verb- oder Adjektivform mit kleinem Anfangsbuchstaben signifikant höher war als für die homophone nominalisierte Form mit Großbuchstaben. Tatsächlich werden Nominalisierungen in der vorliegenden Auswahl nur selten korrekt großgeschrieben und weder die Häufigkeit von nominalisierten und nicht-nominalisierten Wortformen noch ein Häufigkeitsverhältnis sagten die Rechtschreibgenauigkeit signifikant voraus. Bei der Untersuchung der Auswirkungen der Frequenz innerhalb von Nominalisierungen sagten weder die Frequenz der nominalisierten Formen noch die Frequenz der nicht-nominalisierten Formen noch die Frequenzdifferenz zwischen beiden Formen die Rechtschreibleistung für Nominalisierungen voraus. Die Frequenz innerhalb von Nominalisierungen scheint also keinen Einfluss auf die Rechtschreibleistung der Kinder in unserer Stichprobe zu haben.

Der vorliegende Beitrag leistet einen Beitrag in Bezug auf die Beantwortung der Frage, inwiefern explizites (Regel-)Wissen oder implizite Faktoren im Schreibprozess die Richtigschreibungen im Bereich der satzinternen Großschreibung beeinflussen. Die Ergebnisse zeigen, dass die korrekte Großschreibung eines Nomens in dieser Studie nicht durch die Häufigkeit des Auftretens des Rechtschreibmusters im Input und somit durch in der Sprachverarbeitung implizite Frequenzeffekte erklärbar ist, sondern durch den Faktor der Konkretheit. Konkrete Wörter referieren auf einen konkreten Gegenstand mit klar erfassbaren semantischen Merkmalen. Inwiefern der Konkretheitseffekt bei der Großschreibung durch einen Verarbeitungsvorteil konkreter Wörter bedingt ist, bleibt ungeklärt. Denn genau auf dieser Eigenschaft liegt im Regelunterricht dieses Samples zur Großschreibung der Fokus („lexikalischen

Brucher, L., Ugen, S., Müller, C., Cuntz, M., & Weth, C. (2023). Einfluss des Faktors Wortfrequenz auf die satzinterne Großschreibung bei Fünftklässler:innen. In I. Rautenberg & M. Bangel (Eds.), *Das Verhältnis zwischen Wissen und Können beim Lesen- und Schreibenlernen* (pp. 63–82). Hohengehren: Schneider Verlag.

Ansatz“). Somit könnte das im Unterricht erworbene explizite Regelwissen den Konkretheitseffekt verstärken.

Die Aussagekraft der vorliegenden Studie wird durch einige Einschränkungen begrenzt. Die Studie ist explorativ, da wir Prätestdaten einer früheren Interventionsstudie erneut auf die Fragestellung der vorliegenden Arbeit hin analysiert haben. Obwohl die Frequenz bei der Testkonstruktion berücksichtigt und zwischen den Testaufgaben ausgeglichen wurde, wurde der Test nicht mit dem Ziel konstruiert, Frequenzeffekte zu analysieren. Um sich den Frequenzschätzungen des deutschen Inputs von Grundschulkindern in Luxemburg angemessener zu nähern, haben wir *post-hoc* den Korpus childLex verwendet und zusätzlich einen Kinderkorpus auf der Basis luxemburgischer Schulbücher (childLux) erstellt.

Zweitens konzentrierte sich die Studie auf den Faktor Frequenz in einem spezifischen Rechtschreibbereich, nämlich den der Großschreibung von Nomen im Deutschen. Die Frequenz scheint kein so entscheidender Faktor zu sein wie in syntaxbezogenen Rechtschreibbereichen in anderen Sprachen, wie z.B. im Französischen (Pluralmarker bei Nomen und Verben). Die geringere Bedeutung der Frequenz liegt allerdings an dem zweiten Prädiktor, dem Faktor lexikalisch-semantischer Typ, ein Faktor, der in den Studien zum Französischen nicht zum Tragen kam. Darüber hinaus könnte die Frequenz bei der Großschreibung im Deutschen auch anders gewichtet sein, da sie zwar Nomen bzw. den Kern von Nominalphrasen kennzeichnet, der Großbuchstabe jedoch zusätzlich auch andere Bereiche, zum Beispiel den Satzanfang, markiert. Diese mehrfache Verwendung der Majuskel könnte ein Aspekt sein, der den Vergleich zwischen Pluralmarkierung von Homophonen im Französischen und Großschreibung im Deutschen doch weniger vergleichbar macht als angenommen. Denn der Einfluss der Wortfrequenz auf die Großschreibung könnte geringer ausfallen, wenn die Großschreibung auch bei Wörtern im mentalen Lexikon abgespeichert wird, die keine Nomen sind, aber am Satzanfang stehen.

Die Ergebnisse unserer Studie bei Nominalisierungen sind mit den von Largy et al. (1996) experimentell untersuchten homophonen Nomen und Verben auch aufgrund der verwendeten Methoden nur bedingt vergleichbar. Im Vergleich zu dieser ist unsere Untersuchung als explorative Vorstudie zu werten. Die Analysen zeigten jedoch, dass die Frequenz der Nominalisierungen in Bezug auf die entsprechenden Verben und Adjektive keinen Einfluss auf die Rechtschreibung hatten. Deswegen scheint ein Intrusionseffekt wie im Französischen weniger wahrscheinlich.

Dass die vorliegende Studie insgesamt keine Frequenzeffekte gefunden hat, könnte auf die Korrelation zwischen beiden Prädiktoren, d.h. Frequenz und lexikalisch-semantischer Typ, zurückzuführen sein. Da wir auf Kinderkorpora zurückgegriffen

Brucher, L., Ugen, S., Müller, C., Cuntz, M., & Weth, C. (2023). Einfluss des Faktors Wortfrequenz auf die satzinterne Großschreibung bei Fünftklässler:innen. In I. Rautenberg & M. Bangel (Eds.), *Das Verhältnis zwischen Wissen und Können beim Lesen- und Schreibenlernen* (pp. 63–82). Hohengehren: Schneider Verlag.

haben, könnte die Korrelation zwischen Frequenz und lexikalisch-semanticem Typ sogar noch ausgeprägter sein als in Korpora von Erwachsenen, die häufig auf Zeitungstexten basieren. Weitere Untersuchungen sollten versuchen, die Bezüge zwischen Frequenz und Konkretheit sowie Frequenz und Nominalisierungen zu entschlüsseln. Insbesondere bedarf es dazu kontrollierter Experimente.

Literaturverzeichnis

- Betzel, D. (2015): *Zum weiterführenden Erwerb der satzinternen Großschreibung: Eine Längsschnittstudie in der Sekundarstufe I*. Baltmannsweiler: Schneider Verlag Hohengehren.
- Bilici, N., Ugen, S., & Weth, C. (2019). The effects of a syntactic training on multilingual fifth graders' spelling patterns of noun capitalisation in German. *Writing Systems Research* 11(2), 95–109.
- Brucher, L., Ugen, S., & Weth, C. (2020): The impact of syntactic and lexical trainings on capitalization of nouns in German in grade five. *L1 Educational Studies in Language and Literature* 20(Running Issue), 1–23.
- Cohen, J. (1992): Statistical Power Analysis. *Current Directions in Psychological Science* 1(3), 98–101.
- Coltheart, M. (2006). Dual route and connectionist models of reading: an overview. *London Review of Education* 4(1), 5–17.
- Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: Number agreement in sentence production. *Psychological Review* 112(3), 531–559.
- Ellis, N. C. (2002). Frequency effects in language processing. A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24(2), 143–188.
- Griffin, Z. M., & Bock, J. K. (1998). Constraint, word frequency, and levels of processing in spoken word production. *Journal of Memory and Language* 38(3), 313–338.
- Günther, H. (2007): Der Vistembor brelhte dem Luhr Knotten auf den bänken Leuster. Wie sich die Fähigkeit zur satzinternen Großschreibung entwickelt. *Zeitschrift für Sprachwissenschaft* 26(spec), 155–179.
- Houghton, G., & Zorzi, M. (2003). Normal and impaired spelling in a connectionist dual-route architecture. *Cognitive Neuropsychology* 20(2), 115–162.
- Jessen, F., Heun, R., Erb, M., Granath, D. O., Klose, U., Papassotiropoulos, A., & Grodd, W. (2000): The concreteness effect: Evidence for dual coding and context availability. *Brain and Language* 74(1), 103–112.

Brucher, L., Ugen, S., Müller, C., Cuntz, M., & Weth, C. (2023). Einfluss des Faktors Wortfrequenz auf die satzinterne Großschreibung bei Fünftklässler:innen. In I. Rautenberg & M. Bangel (Eds.), *Das Verhältnis zwischen Wissen und Können beim Lesen- und Schreibenlernen* (pp. 63–82). Hohengehren: Schneider Verlag.

- Jurish, B. (2012): *Finite-state Canonicalization Techniques for Historical German*. Universität Potsdam.
- Jurish, B., & Würzner, K.-M. (2013): Word and Sentence Tokenization with Hidden Markov Models. *Journal for Language Technology and Computational Linguistics* 28(2), 61–83.
- Kemp, N., & Bryant, P. (2003): Do Beez Buzz? Rule-Based and Frequency-Based Knowledge in Learning to Spell Plural -s. *Child Development* 74(1), 63–74.
- Largy, P., Fayol, M., & Lemaire, P. (1996): The homophone effect in written French: The case of verb-noun inflection errors. *Language and Cognitive Processes* 11(3), 217–255.
- LCC = *Leipzig Corpora Collection*. (2011): https://corpora.uni-leipzig.de?corpusId=deu_newscrawl_2011
- Levshina, N. (2015): *How to Do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam: John Benjamins.
- Mangelschots, K., Ugen, S., & Weth, C. [eingereicht]: *Spelling and Detection of German Nominal Capitalization by Poor and Good Capitalizers in Fourth Grade*.
- Pacton, S., & Fayol, M. (2003): How Do French Children Use Morphosyntactic Information When They Spell Adverbs and Present Participles? *Scientific Studies of Reading* 7(3), 273–287.
- Pacton, S., Sobaco, A., Fayol, M., & Treiman, R. (2013). How does graphotactic knowledge influence children’s learning of new spellings? *Frontiers in Psychology* 4(October), 1–10.
- Samara, A., & Caravolas, M. (2014): Statistical learning of novel graphotactic constraints in children and adults. *Journal of Experimental Child Psychology* 121, 137–155.
- Sandra, D. (2012): Spelling strategies in alphabetic scripts: Insights gained and challenges ahead. In: Libben, G., Jarema, G. & Westbury, C. (eds.): *Methodological and Analytic Frontiers in Lexical Research*. Amsterdam, Philadelphia: John Benjamins, 110–140.
- Scheele, V. (2006): *Entwicklung fortgeschrittener Rechtschreibfertigkeiten*. Frankfurt a.M.: Peter Lang.
- Schroeder, S., Würzner, K.-M., Heister, J., Geyken, A., & Kliegl, R. (2015a): childLex: Eine lexikalische Datenbank zur Schriftsprache für Kinder im Deutschen. *Psychologische Rundschau* 66, 155–165.
- Schroeder, S., Würzner, K.-M., Heister, J., Geyken, A., & Kliegl, R. (2015b): childLex: A lexical database of German read by children. *Behavior Research Methods* 47(4), 1085–1094.
- Treiman, R., Kessler, B., Zevin, J. D., Bick, S., & Davis, M. (2006): Influence of consonantal context on the reading of vowels: Evidence from children. *Journal of Experimental Child Psychology* 93(1), 1–24.

Brucher, L., Ugen, S., Müller, C., Cuntz, M., & Weth, C. (2023). Einfluss des Faktors Wortfrequenz auf die satzinterne Großschreibung bei Fünftklässler:innen. In I. Rautenberg & M. Bangel (Eds.), *Das Verhältnis zwischen Wissen und Können beim Lesen- und Schreibenlernen* (pp. 63–82). Hohengehren: Schneider Verlag.

Wahl, S., Rautenberg, I., & Helms, S. (2017): Evaluation einer syntaxbasierten Didaktik zur satzinternen Großschreibung. *Didaktik Deutsch* 22(42), 32–51.

Anhang

	B	SE B	Sign
Intercept	20,66	14,81	0,178
childLux_Lemma_lc	-0,75	2,65	0,781

Tabelle 4: Zusammenfassung der Regressionsanalyse (Lemmafrequenz der nicht-nominalisierten Form des childLux) (lc = kleingeschrieben).

Anmerkung: $R^2 = 0$, $F(1, 21) < 1$; $p > 0,05$

	B	SE B	Sign
Abfangen	37,23	12,73	0,008
childLex_Lemma_klein	-4,13	2,48	0,111

Tabelle 5: Zusammenfassung der Regressionsanalyse (Lemmafrequenzen der nicht-nominalisierten Form des childLex).

Anmerkung: $R^2 = 0$, $F(1, 21) = 2,775$; $p > 0,05$

	B	SE B	Sign
Intercept	15,91	6,83	0,03
childLux_Lemma_Diff	0,20	1,85	0,916

Tabelle 6: Zusammenfassung der Regression (Differenz zwischen der Lemmafrequenz der nicht-nominalisierten Form und der Lemmafrequenz der nominalisierten Form des childLux).

Anmerkung: $R^2 = 0$, $F(1, 21) < 1$; $p > 0,05$

	B	SE B	Sign
Intercept	23,19	5,79	< 0,01
childLex_Lemma_Diff	-2,35	1,78	0,201

Tabelle 6: Zusammenfassung der Regression (Differenz zwischen der Lemmafrequenz der nicht-nominalisierten Form und der Lemmafrequenz der nominalisierten Form des childLex).

Anmerkung: $R^2 = 0$, $F(1, 21) = 1,746$; $p > 0,05$