

Überlegungen zur sprachstandbezogenen Relativierung von Wortschätzen. Ein theoretischer Rahmen und eine kleine empirische Studie

Rainer Perkuhn

1. Einleitung

Die zentrale Fragestellung des Beitrags ist, wie der Wortschatz einer bestimmten Personengruppe erhoben werden kann.¹ Der Einfachheit halber nehmen wir zunächst einen intuitiven Wortbegriff an, der aber an späterer Stelle noch einmal aufgerollt wird. Einen Wortschatz können wir somit vorläufig fassen als eine Menge von Wörtern, die die Personengruppe beherrscht. So glatt wie diese kurze Charakterisierung scheint, so schnell zeigt sich die Komplexität, wenn wir ‚beherrschen‘ – gerade im Zusammenhang mit ‚Gruppe‘ – genauer betrachten. ‚Beherrschen‘ umfasst in erster Linie ‚verstehen‘, d. h. ein Wort in einem bestimmten Verwendungszusammenhang bezüglich seiner Bedeutung zu interpretieren und eventuell hinsichtlich seiner Verwendungsspezifika richtig einzuordnen. Die Menge der Wörter, die diese Eigenschaft erfüllt, wird gemeinhin als rezeptiver bzw. passiver Wortschatz (auch: Vokabular) bezeichnet. In zweiter Linie zeigt sich ‚beherrschen‘ aber auch in der Fähigkeit, ein Wort richtig zu verwenden. An dieser Stelle wird gern dem rezeptiven Wortschatz der produktive bzw. aktive Wortschatz gegenübergestellt, insbesondere in Verbindung mit der Beobachtung von Sprachproduktionsprozessen. Unklar bleibt dabei allerdings, wie dem Umstand Rechnung getragen wird, dass eine Person – obwohl sie ein Wort kennt – seine Verwendung vermeidet, da das Wort nicht angemessen ist. Aus unserer Sicht ist dies Teil der produktiven Fertigkeit, die aber nicht mit einfachen Mitteln gemessen werden kann.

Ohne hier auf die Möglichkeiten und Tücken des Messens der Wortschätze einzugehen, wird der rezeptive Wortschatz eines Erwachsenen um Größenordnungen umfangreicher als der produktive Wortschatz eingeschätzt. Während ersterer im Bereich hoher fünfstelliger Zahlen angegeben wird, schwanken letztere zwischen hohen vierstelligen bis zu kleinen fünfstelligen Angaben (nach

1 Mein Dank geht an Daniel Wachter und Harald Längen für die Aufbereitung des Korpus, sowie an Elena Panzeter und Myriam Weiß für die ersten Auswertungen und Diskussionen.

Zimmer 1988, 24). Dieser Zusammenhang, ja sogar in Form einer Teilmen- genbeziehung erscheint plausibel, da man ein Wort nur dann sinnvoll einsetzen kann, wenn man es vorher (oft genug) gehört oder gelesen hat. Nur an wenigen Stellen finden sich Hinweise zur frühkindlichen Sprachentwicklung, dass es auch anders sein kann: „Die *early talkers* dagegen zeigen meist ein größeres produktives als rezeptives Vokabular“ (Bates et al. 1992, zit. nach Klann-Delius 1999, 36; Herv. i. O.).

2. Ermittlung von Wortschätzen

Gerade in diesem Forschungsbereich wurde vermutlich das umfassendste Re- pertoire an Methoden entwickelt, um Wortschätze zu ermitteln². Das aktive Vokabular wird meistens aus der Beobachtung natürlicher, nicht gestellter Sprachproduktionssituationen abgeleitet (etwa initial die Arbeiten von Stern/ Stern 1922 nach Kegel 1974, 52), seltener durch experimentelle Ergebnisse ge- stellter Impulse. Die Vielfalt der Verfahren für das Überprüfen des rezeptiven Vokabulars ist deutlich größer und umfasst Verständlichkeits- und Assozia- tionstests, sowie in neuerer Zeit die durch die moderne Technologie ermöglichte Messung von Augenbewegungen und Gehirnaktivitäten, Reaktionszeiten und Intonation, wodurch sogar pränatale Effekte auf gesprochene Sprache nach- gewiesen werden konnten (siehe hierzu Mampe et al. 2009, zit. nach Höhle 2010, 130). Allen diesen Ansätzen ist aber gemein, dass sie nicht garantieren können, dass selbst im positiven Fall ein Wort mit seinen sämtlichen Facetten beherrscht wird. Ein Wort wurde zwar jeweils in einem konkreten Zusammenhang kor- rekt produziert oder sinnvoll bzw. sinngemäß verstanden. Ob das Wort aber in allen seinen Mehrdeutigkeiten, Nebenlesarten und morphologischen, semanti- schen und pragmatischen Besonderheiten, etwa mitschwingenden Konnotatio- nen richtig eingeordnet wird, können diese Herangehensweisen nicht erfassen. Dies ist vermutlich eine naturgegebene Obergrenze aller Verfahren, die sowohl den Begriff ‚beherrschen‘ aber auch den Begriff ‚Wort‘ relativieren müssen.

Es macht die Sache nicht leichter, wenn wir den Wortschatz einer Gruppe von Personen ermitteln wollen. Die passiven Wortschätze der einzelnen Perso- nen werden als Mengen von Wörtern tendenziell eher über Schnittmengen zur Gesamtheit verknüpft werden. Würde man die Gesamtmenge als Vereinigung angehen, wäre eine Nebenbedingung unerlässlich, etwa in der Form, dass nur die Wörter zählen, die von den meisten/mindestens X% oder einem durch- schnittlichen/prototypischen Vertreter der Gruppe verstanden werden. Für die

2 Vgl. Kegel (1974, 52–59), für eine Übersicht der damaligen Zeit, Höhle (2010, 128), mit Verweis auf Kapitel 2 für neuere Methoden.

aktiven Wortschätze liegt es näher, von einer Vereinigung der Mengen auszugehen. Aber auch hier stellt sich die Frage, was wir uns unter einem (vereinigten) produktiven Wortschatz einer Personengruppe vorstellen sollen. Er kann nur als charakterisierend für die Gruppe gelten, wenn hier ebenfalls zumindest eine Mehrheit die Wörter auch versteht, sprich mit den individuellen passiven Wortschätzen geschnitten wird.

Von den oben genannten Verfahren zur Ermittlung von Wortschätzen ist genau genommen nur die ständige Beobachtung ein verlässlicher Weg, um einen Wortschatz so umfassend wie möglich zu erheben – unter den oben genannten Vorbehalten. Die Wörter, die geäußert wurden (wobei wir nicht zwischen mündlicher und schriftlicher Gestaltung unterscheiden), sind das aktive Vokabular. Die Wörter, die gehört oder gelesen wurden, stellen den Rahmen für das rezeptive Vokabular, sofern keine Hinweise darauf vorliegen, dass ein unbekanntes Wort nachgefragt oder nachgeschlagen wurde. In beiden Szenarien sind die Grenzen des Machbaren bedingt durch die Breite und Dauer der Beobachtung. Auch wenn sie gerade hinsichtlich individueller Sprachproduktivität zu sehr starker Streuung führen können, werden sie die für die dokumentierten Situationen üblichsten Wörter hervorbringen. In experimentellen Settings wird man hingegen meist mit (Maximal-)Erwartungen an die üblichen Wörter einsteigen und diese auf Beobachtung, Produzierbarkeit oder Verständlichkeit hin abprüfen. Die Alternative, alle Wörter einer Sprache der Reihe nach zu überprüfen, ist aus verständlichen Gründen nicht praktikabel. Nur für bestimmte Personengruppen sind die Anhaltspunkte aufgrund langjähriger Erfahrungen relativ klar: Kleinkinder und Kinder bis zum sechsten Lebensjahr (mit Ausnahmen auch bis zum zwölften Lebensjahr), für die viele Untersuchungen zur Entwicklung des Wortschatzes vorliegen, aus denen über Beobachtungen sprachlichen Verhaltens quasi Maximal-Wortschätze abgeleitet wurden³. Für Erwachsenenwortschätze werden Wortschätze oft über anderes Material abgeschätzt und hochgerechnet. Der „Bildungswortschatz“ nach Augst (2019) wird u. a. etwa aus den besonderen Formulierungen der Medien abgeleitet, die ihre Texte für die Zielgruppe „Bildungsbürger“ formulieren⁴.

3 Neben der Reflexion postulierter Wortschätze teilweise auch rein datengetrieben wie in Augst (1984).

4 Wobei sicher in Kauf genommen wird, dass ‚Bildung‘ auch prozedural gedeutet werden kann, indem Rezipienten unbekannte Begriffe nachschlagen und sich diese dadurch aneignen.

3. Referenzwortschätze

Somit sind wir bei einem bestimmten Verwendungszweck von Wortschätzen angelangt, der über die Beschreibung des IST-Zustandes hinausgeht. In verschiedenen Szenarien werden Wortschätze oft über eine Mischung aus empirischen und anwendungsbezogenen Kriterien ermittelt, die für bestimmte Fragestellungen als Bezugswortschatz fungieren. Sie werden als Zielgröße eines zu erlernenden Wortschatzes insbesondere in der Fremdsprachendidaktik vorgegeben und dienen im Weiteren dann auch zu Testzwecken (vgl. Nation 2016), ob ein bestimmtes Sprachniveau erreicht wurde. Im DaF-Bereich gibt es etwa Tests zu den verschiedenen Sprachniveau-Stufen des Gemeinsamen Europäischen Referenzrahmens⁵. Für die deutsche Sprache wurden die dafür zugrunde gelegten Ziel-Wortschätze größtenteils aus dem Zertifikatswortschatz Deutsch abgeleitet und thematisch angepasst und untergliedert. Die Vorläufer davon basierten wiederum auf einer sehr alten Worthäufigkeitsliste (vgl. Kaeding 1898/99), deren Spuren noch lange Zeit in verschiedenen Materialien sichtbar waren (wie z. B. das Lemma „Telefonapparat“ in Glaboniat et al. 2002).

Spätestens jetzt sind wir mit der Kaedingschen Liste gedanklich bei aus Textsammlungen erhobenen Referenzwortschätzen. Heutzutage werden diese nicht mehr von Hand, sondern aus elektronischen Sammlungen, aus Korpora (wie z. B. dem ‚Deutschen Referenzkorpus‘ DEREKO⁶) möglichst automatisch gewonnen. Diese umfassen das Vokabular und auch die gezählte Häufigkeit, da sie u. a. auch innerhalb korpuslinguistischer Methodik dazu verwendet werden, um abschätzen zu können, welche Wörter in anderen Korpora bzw. bestimmten Teilmengen besonders oft, gelegentlich auch: besonders selten vorkommen. Die große Herausforderung stellt sich dabei darin, neben der Verfügbarkeit einer geeigneten Datengrundlage, einen sinnvollen Wortbegriff festzulegen und zu kommunizieren, der über alle Ebenen der Abstraktion operationalisierbar ist.

4. Wortschätze in Entwicklung

Als letzten Gedanken an dieser Stelle möchten wir auf die zeitbezogenen Aspekte von Wortschätzen eingehen, die einen „dynamischen Wortschatzbegriff“ erforderlich machen (vgl. Schnörch 2015). Eine Sprache, insbesondere ihr Wortschatz, verändert sich im Laufe der Zeit. Dies wird besonders sichtbar an der am deutlichsten kodifizierten Form von Wortschätzen, an Wörterbüchern und ihren Lemmastrecken. Bei jeder Neuauflage werben die Wörterbücher mit

5 <http://www.europaeischer-referenzrahmen.de> (30.10.2019).

6 <http://www1.ids-mannheim.de/kl/projekte/korpora.html> (30.10.2019).

der Erweiterung um eine bestimmte Anzahl neuer Wörter. Eher im akademischen Bereich wird das Themengebiet beforscht und dokumentiert durch spezielle Neologismenprojekte (vgl. Herberg/Kinne/Steffens 2004; Steffens/al-Wadi 2015).

Auch der individuelle Wortschatz einer einzelnen Person verändert sich im Laufe ihres Lebens. Er verändert sich im Umfang, vermutlich auch im Status einzelner Elemente: In jungen Jahren von unbekannt zu rezeptiv, dann zu produktiv beherrscht, in älteren Jahren sicher auch in der umgekehrten Richtung, ganz unabhängig von pathologischen oder situativen Phänomen (wie ‚tip of the tongue‘⁷). Von den Lebensphasen ist mit Abstand der frühkindliche Spracherwerb am besten erforscht. Der Schwerpunkt liegt dabei meist auf der Entwicklung des produktiven Wortschatzes, der innerhalb gewisser Spielräume etwa vom Äußern des ersten Wortes mit 12 Monaten nach Erreichen eines Wortschatzes im Umfang von ca. 50 Wörtern mit ungefähr 18 Monaten einen Vokabelspurt dokumentiert, der bis zum zweiten Lebensjahr den Umfang auf etwa 300 Wörter wachsen lässt. Für die nächsten 250 Wörter brauchen Kinder ungefähr bis zum dritten Lebensjahr (vgl. Höhle 2010, 133). Für den rezeptiven Wortschatz, über dessen Messung und Messbarkeit viel diskutiert wurde, werden Zahlen für Sechsjährige genannt, die zwischen 2.500 und 3.000, bei anderen Untersuchungen zwischen 7.800 und 21.000 liegen (vgl. Kegel 1974, 54–57). Selbst wenn in diesen Zahlen viele Ungenauigkeiten verborgen liegen, da der Wortbegriff unklar bleibt, die Ableitbarkeit unterschiedlich gehandhabt wird, die Hochrechnung, gerade auch die Übertragbarkeit auf andere Sprachen angezweifelt werden kann, so deuten sie dennoch schon eine Herausforderung an: Der Wortschatz von Erwachsenen wird deutlich höher geschätzt, aktiv im Umfang großer vierstelliger bis kleiner fünfstelliger Zahlen, passiv von mittleren bis hohen fünfstelligen Zahlen, Augst nennt etwa 94.000 Wörter für gebildete Erwachsene (vgl. Augst et al. 1977 nach Zimmer 1988, 24). Die Wortschatzentwicklung von älteren Kindern und Jugendlichen wurde bisher wenig untersucht. Neben einem großen Freiraum an individueller Ausgestaltung wäre es sicher interessant, ob es einen Kern gibt, der sich gemeinsam zwischen dem sechsten und dem achtzehnten Lebensjahr entwickelt.

Ähnlich wie das Projekt ‚Wortschatzwissen.de‘, das einen Wortschatz erheben möchte, den „Schulabgänger der 10. Klasse beherrschen (sollten)“⁸, möchten wir uns in einer deutlich bescheideneren, kleinen Studie mit Wortschatzen jüngerer Altersstufen beschäftigen. Da es sehr aufwändig wäre, Daten der in

7 Das – im Vergleich zu krankheitsbedingten Wortfindungsstörungen – Alltagsphänomen gesunder Menschen, bei dem „einem Sprecher ein Wort auf der Zunge liegt, er es aber nicht artikulieren kann“ (Höhle 2010, 55).

8 <https://www.wortschatzwissen.de> (30.10.2019).

den Altersstufen produzierter Texte zu erheben, und uns in unserem Rahmen auch keine Möglichkeit zur Verfügung steht, Experimente zum rezeptiven Wortverständnis durchzuführen, verfolgen wir einen anderen Ansatz. Wir wollen uns mit Texten beschäftigen, die gezielt für eine bestimmte Altersgruppe verfasst wurden. Dieser Gedanke ist zum Teil auch in dem oben genannten Projekt berücksichtigt. Es gibt sogar Hinweise darauf, dass ein ähnlicher Ansatz von Thorndike schon 1921 zur Ermittlung von Schulwortschätzen Anwendung gefunden hat (vgl. Thorndike 1921 nach Kegel 1974, 53), wobei die genauen altersstufen-bezogenen Quellen und ihre Bewertung im Zusammenspiel mit den anderen verwendeten Quellen nicht ganz nachvollziehbar sind. Einen grundsätzlich interessanten Datensatz bietet auch das Projekt ‚childLex‘⁹ an mit Häufigkeitsangabe und weiteren Kennzahlen zum Vokabular eines Kinderbuchkorpus. Unsere Datensammlung stellen wir allerdings nach anderen Kriterien zusammen und möchten weitere Kennzahlen verwenden bzw. entwickeln, um einen Eindruck der Wortschätze verschiedener Altersstufen zu bekommen, insbesondere auch, wie sich diese zueinander verhalten und sozusagen weiter-entwickeln.

5. Das ‚Carlsen-Korpus‘

Als Teil des ‚Deutschen Referenzkorpus‘ (DEREKO)¹⁰ baut das Leibniz-Institut für Deutsche Sprache seit kurzem ein Kinder- und Jugendbuchkorpus auf. Der Schwerpunkt ist dabei vorrangig zeitgemäße Literatur mit der Originalsprache deutsch, um den Kaedingschen „Telefonapparat“ (hier dann etwa „garst’ger Struwelpeter“) und Übersetzungseffekte zu vermeiden. In einer allerersten Fassung haben wir für diese Studie intern mit 35 Quellen arbeiten können. Diese wurden mit der DEREKO-Version 2019-II (Stand: 11.11.2019) freigegeben und sind damit auch über das aktuelle Recherchesystem verfügbar. Da bisher alle Quellen Publikationen aus dem Carlsen-Verlag sind, dem an dieser Stelle für die Überlassung der Nutzungsrechte gedankt sei, bezeichnen wir intern das Korpus in seiner aktuellen Form der Einfachheit halber als das ‚Carlsen-Korpus‘.

Alle Quellen dieses Korpus enthalten als Metadatum die Altersstufe, für die die Texte empfohlen sind. Im Vergleich zu Material, das von Didaktik-

9 ‚childLex‘ ist ein Kooperationsprojekt des Max-Planck-Instituts für Bildungsforschung mit der Universität Potsdam und der Berlin-Brandenburgischen Akademie der Wissenschaften (siehe dazu auch Schroeder et al. 2015). Weitere Informationen unter <https://www.mpib-berlin.mpg.de/de/forschung/max-planck-forschungsgruppen/mpfg-read/projekte/childlex> (30.10.2019).

10 <http://www1.ids-mannheim.de/kl/projekte/korpora.html> (30.10.2019).

Profis speziell für Schule und Unterricht der Altersstufen aufbereitet wurde, spekulieren wir darauf, dass die Texte auf eine unbefangene, quasi natürliche Art entstanden sind, wobei nicht mit allerletzter Sicherheit aufzuklären ist, wie viel Talent der Autorinnen und Autoren, wie viel redaktionelle Vorgaben oder sogar Überarbeitung bei der Verfertigung der Texte eingeflossen ist. Wie gut das Endprodukt tatsächlich gelungen ist und inwieweit es für die angegebene Altersstufe passt, wäre direkt nur mit größerem Aufwand zu prüfen – indirekt können wir das aber an dem Erfolg abschätzen, im Einzelfall an den Verkaufszahlen, auf Dauer aber auch daran, welche Autorinnen und Autoren sich in dem Markt zu einer festen Größe entwickeln. Wir hoffen, eine gute Auswahl getroffen zu haben, und werden das Korpus nach und nach ausbauen, für die kommenden Erweiterungen stehen bereits jetzt noch weitere 50 Bücher aus dem Carlsen-Verlag zu Verfügung.

Für die aktuelle Fassung können wir zurückgreifen auf 21 Bücher mit einer Altersempfehlung ab 7 Jahren, ein Buch ab 10 Jahren, drei ab 12 Jahren, vier ab 13 Jahren, fünf ab 14 Jahren, sowie eines ab 15 Jahren. Bei den Büchern der jüngsten Stufe handelt es sich durchweg um Conni-Erzählbände von Julia Boehme, die anderen Bücher stammen von verschiedenen Autorinnen und Autoren, die z. T. aber auch stufenübergreifend vertreten sind. Wir sind uns der Unausgewogenheit der Zusammenstellung durchaus bewusst, gerade auch bezüglich Erzählstil und inhaltlicher (Dis-)Kontinuität. Für unsere erste kleine Pilotstudie sehen wir deshalb auch das primäre Anliegen, einen Apparat zu schaffen, der u. a. hilft, diese Besonderheiten aufzudecken und ggf. zu kontrollieren – und der später auch auf umfassendere und ausgewogenere Datensätze angewendet werden kann.

Wir haben die Quellen zu Gruppen zusammengefasst, die sich vom Umfang nicht besonders unterscheiden sollten. Dabei haben die beiden am besten besetzten Schichten (7- und 13-Jährige) als Vorgabe gedient. Die dazwischen und darüber liegenden Angaben haben wir aufgrund der dünneren Datenlage zusammengenommen, sodass vier Gruppen entstanden sind (vgl. Abbildung 1).

Anhand einiger Kennzahlen, die wir aus den Daten berechnet haben, kann man die Herausforderung abschätzen, diese Textmengen miteinander vergleichen zu wollen. Wie erwartet nimmt der Umfang des Vokabulars (zunächst gemessen in Types, d. h. Wortformen) tendenziell zu. Allerdings wird auch die durchschnittliche Textlänge parallel dazu größer, was einerseits plausibel Kind angemessen ist, die Messung von lexikalischer Vielfalt aber erschwert.

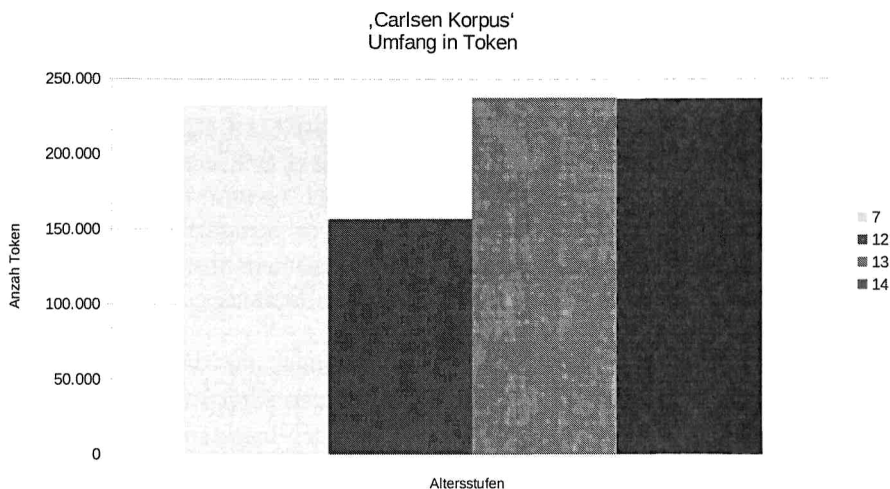


Abb. 1: Aufteilung des ,Carlsen-Korpus' in vier ungefähr gleich große Teilkorpora

	7	12	13	14	gesamt
Token	233.118	156.747	237.777	237.204	864.846
Texte	21	4	4	6	35
Textlänge Ø	11.101	39.187	59.444	39.534	24.710
Type	15.280	18.846	23.264	21.449	47.666
TTR	0,066	0,120	0,098	0,090	0,055
Hapaxe	7.516	10.862	12.854	11.669	24.245
Hapax %	49,19 %	57,64 %	55,25 %	54,40 %	50,86 %

Tab. 1: Kennzahlen der Teilkorpora des ,Carlsen-Korpus'

Aus mehreren Gründen wollen wir bei unseren ersten Betrachtungen auf der Wortformebene bleiben. Zunächst wollen wir die Eigenschaften der Texte der unterschiedlichen Gruppen besser verstehen. Dabei würde eine Lemmatisierung bestimmte Unterschiede ausblenden, die etwa auf Interaktionalität oder der Erzählperspektive beruhen. Darüber hinaus bringt die Operationalisierung, die wir für die Textmengen brauchen, Unwägbarkeiten ins Spiel, die wir erst in späteren Studien genauer untersuchen können. Mit unseren Altersstufen-Wortschätzen schließen wir eine Lücke zwischen dem frühkindlichen Stand etwa zu Beginn der Grundschule bis hin zu Abschnitten (12–14 Jahre), in denen die Wortschatzentwicklung eigentlich „in wesentlichen Zügen abgeschlossen ist“ (Klann-Delius 1999, 36) und somit sich Erwachsenenwortschätzen annähert. Der uns von dort vertraute Lemma-Begriff einer Grundform, wie er allgemein in der Lexikografie üblich ist, gilt für die Primarstufendidak-

tik aber nicht: Dort werden in vielen Fällen flektierte Wortformen angegeben, insbesondere natürlich unregelmäßige Bildungen, gerade der Hilfsverben (z. B. *bin, ist, war, fahren, fährt*¹¹).

„Zu den Grundformen sind relevante Ableitungen und flektierte Formen beispielhaft angefügt. Diese verdeutlichen, dass es keinen wortgenauen Grundwortschatz geben kann.“¹²

6. Studie zu den altersempfohlenen Texten

Für unser Korpus messen wir die Gesamthäufigkeit der Wortformen sowie auch getrennt die Häufigkeiten in den verschiedenen Altersstufen. Für die Altersstufen wiederum berechnen wir die Häufigkeiten insgesamt sowie in ihren verschiedenen Quellen, wobei wir diese Angaben erst in wenigen Einzelfällen zu Rate gezogen haben.

Die Gesamtverteilung folgt dem Zipf'schen Gesetz, wir sehen ganz wenige Wörter, die hochfrequent sind (6 über 10.000), einen mittelgroßen Bereich mittelfrequenter Wörter (113 über 1.000, 870 über 100) sowie einen großen Bereich seltener (5.240 über 10) und sehr seltener Wörter (41.437 1 bis 9 Mal).

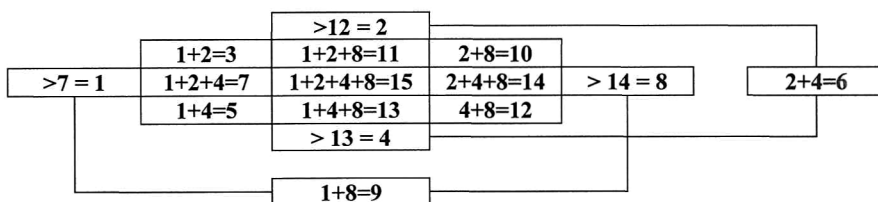
Für beide Ebenen interessiert uns neben der Häufigkeit auch die Streuung (Dispersion), d. h. wie sich die Vorkommen der Wörter über die Bestandteile der Korpora verteilen. Obwohl es zugegebenermaßen eine ausgiebige wissenschaftliche Auseinandersetzung mit dem Thema gibt (siehe Gries 2008), haben wir uns entschieden, ganz einfache Maße anzubieten, die die (Un)Ausgewogenheit auf eine intuitive Art darstellen und somit die schnelle Interpretation und Einordnung unterstützen sollen. Als einfachstes Maß kann die Anzahl der Teilkorpora gezählt werden, in denen die Wortform mindestens einmal enthalten ist. Außer der Maximalzahl ist allerdings jede andere Angabe mehrdeutig, da nicht zu ersehen ist, in welchen Teilen das Wort belegt ist. Als Abhilfe dazu bilden wir eine ‚binäre Summe‘: Jedes Teilkorpus bekommt als Wertigkeit eine Zweierpotenz zugewiesen (1, 2, 4, 8, usw.), sodass sich aus jeder Summe die beteiligten Summanden wieder eindeutig rekonstruieren lassen. Für unsere oberste Ebene erhalten wir somit Angaben von 1 bis 15. Da wir nicht ganz zufällig die Wertigkeiten aufsteigend zu den Altersstufen vergeben, können wir weitere Informationen sofort an den Zahlen ablesen: Jede Summe größer gleich 8 bedeutet, dass die Wortform auf jeden Fall auch im letzten Zeitabschnitt

11 <https://www.gut1.de/grundwortschatz/grundwortschatz-500.html> (30.10.2019).

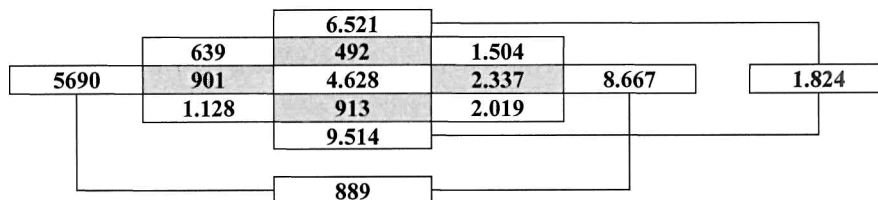
12 https://bildungsserver.berlin-brandenburg.de/fileadmin/bbb/unterricht/faecher/sprachen/deutsch/Grundschule/2011_11_25_GWS_1_WEB.pdf (30.10.2019).

vertreten ist. Das Vorhandensein im ersten Zeitabschnitt führt zu ungeraden Summen, eine gerade Summe indiziert das Fehlen im ersten Abschnitt.

Tabelle 2 bietet eine Orientierung für die Angaben der verschiedenen Teil- bzw. Schnittmengen: Links ist das Feld für die Angaben zu Wörtern, die nur in der Altersstufe > 7 belegt sind, oben analog das Feld für die Altersstufe > 12 (sowie rechts > 14 , unten > 13). Das Feld schräg dazwischen indiziert die Angaben zu ihrer Schnittmenge, sowie analog alle anderen Felder die Schnittmengen zu den den Altersstufen entsprechenden Summanden. In der Mitte ist die Anzahl der Wörter angegeben, die in allen Altersstufen belegt sind.



Tab. 2: Zuordnung der binären Summen zu dedizierten Teil- bzw. Schnittmengen



Tab. 3: Wortformenhäufigkeiten in den verschiedenen Teil- bzw. Schnittmengen (vgl. Tab. 2)

Mit dieser Kodierung lässt sich die Liste gezielt untersuchen und auswerten. Tabelle 3 kann man z. B. die Häufigkeitsangaben der einzelnen Teilwortschätze entnehmen. In jeder Altersstufe gibt eine mittlere bis hohe vierstellige Zahl an Wörtern, die nur bei dieser Altersstufe beobachtet wurden. Durch Umsortieren können wir uns die Wörter dieser Teilwortschätze gezielt anschauen, Tabelle 4 zeigt die jeweils 20 häufigsten Wörter der altersstufen-spezifischen Teilmengen. Beim jeweils 20. Eintrag sind dessen Häufigkeit und sein Rang (#) in der Gesamthäufigkeitsliste angegeben.

Stufe > 7	Stufe > 12	Stufe > 13	Stufe > 14
<i>Conni</i>	<i>Yurik</i>	<i>Glass</i>	<i>Aurelia</i>
<i>Anna</i>	<i>Paps</i>	<i>Sophie</i>	<i>Mim</i>
<i>Billi</i>	<i>Nora</i>	<i>Nicholas</i>	<i>Sasa</i>
<i>Mau</i>	<i>Erasmus</i>	<i>Tereza</i>	<i>Darian</i>
<i>Annika</i>	<i>Mams</i>	<i>Caro</i>	<i>Victoria</i>
<i>Behrens</i>	<i>Delphine</i>	<i>Robin</i>	<i>Asaf</i>
<i>Beppo</i>	<i>Margarethe</i>	<i>Janine</i>	<i>Leyla</i>
<i>Connis</i>	<i>Dandelion</i>	<i>Paula</i>	<i>Sina</i>
<i>Nicki</i>	<i>Pola</i>	<i>Samuel</i>	<i>Elric</i>
<i>Liska</i>	<i>Husch</i>	<i>Visible</i>	<i>Alexander</i>
<i>Celina</i>	<i>Schröders</i>	<i>Gable</i>	<i>Miros</i>
<i>Reisig</i>	<i>Natascha</i>	<i>Manja</i>	<i>Mascha</i>
<i>Lars</i>	<i>Sabrina</i>	<i>Pascal</i>	<i>Tanita</i>
<i>Sandulescu</i>	<i>Claus</i>	<i>Diannes</i>	<i>Tabea</i>
<i>Torben</i>	<i>Arnie</i>	<i>Annie</i>	<i>Nele</i>
<i>Elton</i>	<i>Carla</i>	<i>Katis</i>	<i>Ay</i>
<i>staunt</i>	<i>Tara</i>	<i>Paleiko</i>	<i>Sofia</i>
<i>Ponys</i>	<i>Kowalski</i>	<i>Valerie</i>	<i>Yeniköy</i>
<i>Amadeus</i>	<i>Samuele</i>	<i>Inge</i>	<i>Elfie</i>
<i>Billis</i> (71 = # 1.269.)	<i>Döller</i> (45 = # 1.870.)	<i>Joint</i> (45 = # 1.870.)	<i>Rasih</i> (55 = # 1.563.)

Tab. 4: Die 20 häufigsten Wortformen, die ausschließlich in den jeweiligen Teilkorpora vorkommen

Alle vier Bereiche werden dominiert von Eigennamen, meist Personennamen der Akteurinnen und Akteure der jeweiligen Geschichten, vereinzelt dann auch Ortsnamen. Auch bei *Erasmus* und *Delphine* handelt es sich um Eigennamen. Als ergänzende Information können wir hier auch die Dispersion der unteren Ebene verwenden, bei der die Eigennamen der letzten drei Spalten ungleiche Verteilungen zeigen – nur die erste Spalte kann da nicht weiter aufgelöst werden, da es sich durchgängig um Conni-Bücher mit weitestgehend denselben Akteurinnen und Akteuren handelt.

Etwas aus der Reihe tanzen *staunt* und *Ponys* in der ersten Spalte, *Paps* und *Mams* in der zweiten und *Joint* in der dritten Spalte. Neben den Eigennamen und der konkreten Verbform haben wir mit den anderen Beispielen Hinweise auf weitere Einflüsse auf die Zusammensetzung der verschiedenen Wortschätze. Diese sind natürlich geprägt durch die Handlungen und Themenbereiche, die mit den Büchern abgedeckt werden, z. B. die Unternehmungen mit kleineren Kindern in Spalte 1 oder der Umgang mit Drogen in Spalte 3. *Mams* und *Paps* spiegelt dann doch ein wenig den Erzählstil wider, wie die Autorinnen und Autoren die Kinder ihrer Erzählungen ihre Eltern ansprechen lassen.

Um das Phänomen der Verbform weiter aufschlüsseln zu können, haben wir das Dispersionsmaß für eine weitere Form der Filterung eingesetzt. Wir haben nachgeschaut, welche Wortformen zwar insgesamt häufig sind, aber noch nicht

in Altersstufe > 7 beobachtet wurden, d. h. die Formen mit Kennziffer 14 nach Häufigkeit absteigend sortiert (siehe Tabelle 5).

Wortform	> 7	> 12	> 13	> 14	Summe	binär summierte Dispersion
<i>fragte</i>	0	101	39	144	284	14
<i>hielt</i>	0	85	25	124	234	14
<i>nahm</i>	0	54	32	148	234	14
<i>begann</i>	0	51	19	109	179	14
<i>nickte</i>	0	50	26	98	174	14
<i>stellte</i>	0	51	26	92	169	14
<i>versuchte</i>	0	39	12	116	167	14
<i>legte</i>	0	50	21	94	165	14
<i>fühlte</i>	0	36	26	94	156	14
<i>setzte</i>	0	42	18	90	150	14
<i>schüttelte</i>	0	47	20	75	142	14
<i>fuhr</i>	0	42	13	86	141	14
<i>spürte</i>	0	29	10	99	138	14
<i>drehte</i>	0	47	5	82	134	14
<i>warf</i>	0	54	13	61	128	14
<i>tat</i>	0	44	12	71	127	14
<i>lasse</i>	0	23	96	1	120	14
<i>öffnete</i>	0	34	17	64	115	14
<i>Wut</i>	0	21	22	69	112	14
<i>flüsterte</i>	0	19	25	66	110	14

Tab. 5: Die 20 häufigsten Wortformen, die noch nicht in Altersstufe > 7 beobachtet wurden

Bis auf *lasse* und *Wut* enthält der Abschnitt ausschließlich Präteritum-Verbformen und auch bei den danach folgenden Einträgen verhält es sich nicht anders – lediglich unterbrochen von z. B. *Fluss* (da Conni wohl nur Urlaub am Meer, aber nicht am Fluss gemacht hat) und *Großvater* (da Conni ihren Großvater anders nennt). Es entsteht der Eindruck, dass die Präteritumbildungen systematisch für die jüngste Stufe vermieden werden. Zu den Gründen können wir nur Vermutungen anstellen: Sei es, weil der Erzählstil für die Jüngeren dadurch lebendiger wirken soll, oder sei es, um die Sprache im Wortschatz (Grundschul-)Kind angemessener zu gestalten, so wie es die oben erwähnte differenzierte Darstellung von Grundschulwortschätzen nahelegt. Um zu prüfen, ob andere Formen der Verbparadigmen sich anders, womöglich komplementär verhalten, könnten wir gezielt nach diesen suchen, eventuell unterstützt durch die Ermittlung der gemeinsamen Grundform. Eine derartige, aufgrund der Datenmenge zwingend automatische Lemmatisierung hat aber durchaus ihre Unwägbarkeiten. Wir wollen zunächst mit einer zweiten Form der Beschreibung der Streuung arbeiten, die uns in der Fragestellung nicht ganz unverhofft genauso gut weiterbringt.

Unser bisher verwendetes Maß berücksichtigt das Vorhandensein in einer Altersstufe nur binär, d. h. ein einzelnes Vorkommen trägt genauso dazu bei wie eine hohe Frequenz. Wir haben eine weitere Darstellung entwickelt, die in einer übersichtlichen und intuitiven Form einen Überblick gibt, in welchen Frequenzregionen ein Wort in den verschiedenen Teilkorpora vertreten ist. Absolute oder relative Frequenzangaben sind dazu nicht gut geeignet. Häufigkeitsklassen beruhen auf der Idee, absolute Frequenzen zu meist exponentiell/logarithmisch definierten Klassen zusammenzufassen, wie wir es mit unserem Beispiel oben mit Zehnerpotenzen bereits verwendet haben. Alternativ kann man auch einen entsprechenden Abstand zum Maximalwert betrachten: Bezogen auf z. B. die Basis 2 hieße dies, dass eine Häufigkeitsklasse der Anzahl der Verdopplungsschritte entspricht, die man braucht, um von der Frequenz des betrachteten Wortes zu der des höchstfrequenten Wortes zu gelangen. Der Nachteil dieser Maße ist, dass ihre Skala mit zunehmender Korpusgröße wächst, d. h. dass es in einem größeren Korpus mehr Häufigkeitsklassen gibt als in einem kleinen. Insbesondere die niedrigen Klassen lassen sich nicht mehr aufeinander beziehen, die Einmalvorkommen (Hapaxe) bekommen unterschiedliche Kennzahlen. Um sowohl die niedrigste als auch die höchste Frequenzklasse in verschiedenen großen Korpora gleich kennzeichnen zu können, müssen wir statt des Verdopplungsfaktors ‚2‘ einen dynamischen Faktor ermitteln, der sich aus der jeweils maximalen Frequenz ergibt. Für unser ‚Carlsen-Korpus‘ wollen wir als Kennzahlen nur einstellige Ziffern (mit intuitiver aufsteigender Orientierung) einsetzen, sodass sich der Faktor als $b = \sqrt[9]{max}$ ergibt, mit dem sich alle Frequenzen als logarithmische Klassen zwischen $hk(0) = 0$, $hk(1) = 1$ bis $hk(max) = 9$ abbilden lassen ($hk(x) = c$, wenn $x * b^{9-c} = max$)¹³. Für unser viergeteiltes Korpus erhalten wir somit eine vierstellige Ziffernfolge, die in ihrer Reihenfolge die jeweiligen Häufigkeitsklassen in der Abfolge der Altersstufen wiedergibt (siehe Tabelle 6, rechte Spalte).

Token	> 7	> 12	> 13	> 14	gesamt	bin. Summe	HK-Kennzahl
<i>und</i>	4.496	4.082	5.888	6.196	20.662	15	9999
<i>die</i>	4.038	3.105	4.420	4.931	16.494	15	9999
<i>ich</i>	1.391	2.796	6.055	3.681	13.923	15	8998
<i>sie</i>	4.901	2.143	2.421	2.738	12.203	15	9888

Tab. 6: Die 4 häufigsten Wortformen mit Kennzeichnung der Häufigkeitsklassen

13 Rundungseffekte werden durch kleine Aufschläge abgefangen.

Dass das Personalpronomen für die erste Person nur geringen Schwankungen unterliegt, ist ein Hinweis darauf, dass die Erzählperspektive nur einen geringen Einfluss auf die Häufigkeitsverteilung hat.

Der Ziffernfolge fehlt zwar der visuelle Charme einer grafischen Darstellung (wie Säulendiagrammen), sie lässt sich aber auch direkt als Zahl interpretieren, mit deren Hilfe wir wiederum über Sortierung Eigenschaften verschiedener Abschnitte herausstellen können. So sehen wir jeweils zu Anfang einer der (auch schon durch die binäre Summe erfassten) Teilmengen – sofern möglich und vorhanden – aufsteigende Verläufe, dann gleichbleibende, zum Schluss fallende. Die oben betrachtete ungleiche Verteilung der binären Summe 14 (nicht in Stufe 7, sonst überall) drückt sich dann als 0xxx aus. Beispiele mit einem absolut komplementären Verhalten („x000“) haben wir schon gesehen, dies sind die Top 20 der binären Summe 1 (siehe Tabelle 4, linke Spalte), die eine HK-Kodierung von 9000 bis 5000 aufweisen. Als Hinweis auf unser Phänomen der ungleichen Verbformverteilung finden wir dort allerdings nur ein Beispiel. Genau genommen wäre es auch überraschend, wenn bestimmte Verbformen bei den höheren Altersstufen komplett verschwinden, wahrscheinlicher ist maximal ein gewisser Kompensationseffekt, somit ein Rückgang in Maßen. Mit einem kleinen Trick können wir auch dies über die Kodierung herausfiltern: Wir schauen uns alle Einträge an, die mit einem hohen Wert in Stufe 7 (bis zu HK 6) anfangen, der dann in mindestens einer der folgenden Stufe deutlich unterschritten wird (Differenz > 2, siehe Tabelle 7, die beiden linken Spalten).

<i>fragt</i>	7464	<i>fragte</i>	0545
<i>ruft</i>	7353	<i>rief</i>	1435
<i>meint</i>	7342	<i>meinte</i>	1333
<i>zieht</i>	6453	<i>zog</i>	1545
<i>hält</i>	6453	<i>hielt</i>	0545
<i>lacht</i>	6352	<i>lachte</i>	0434
<i>grinst</i>	6351	<i>grinste</i>	0424
<i>erklärt</i>	6343	<i>erklärte</i>	0434
<i>denkt</i>	6343	<i>dachte</i>	4556
<i>nickt</i>	6325	<i>nickte</i>	0545

Tab. 7: Absteigende (Präsens-)Verbformen mit Kennzeichnung der Häufigkeitsklassen

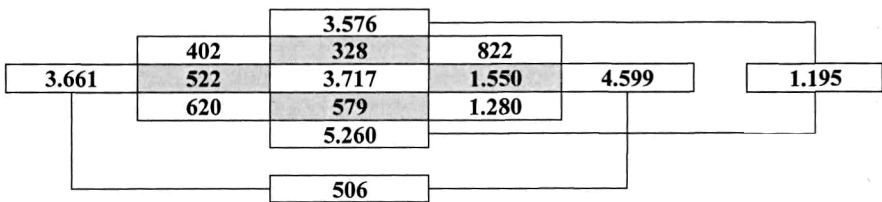
Die Liste der absteigenden Wortformen lässt sich noch um einige Präsens-Verbformen fortführen. Für alle lassen sich durchgängig Präteritum-Formen angeben, die nahezu komplementär aufsteigen (außer *meinte* und *dachte* in Tabelle 7 sowie *hakte*, *staunte*, *bekam*, *guckte*, *versprach*, *kicherte*, *entdeckte*, *schluckte* mit gleichmäßige(re)m Verhalten). Die übrigen Elemente der Liste sind meist Eigennamen, sonst finden sich nur wenige Substantive: *Papa*, *Mama*, *Opa*, *Oma*, *Onkel*, *Pony(s)*, *Kater*, *Berg*, *Zelt*, *Knochen*. Neben den Inhalten der Ge-

schichten zeigt sich hier erneut der unterschiedliche Stil bei der Anrede der Familienrollen.

Wir hatten eingangs darauf hingewiesen, wie wichtig es ist, sich Gedanken zu machen über die sprachliche Einheit, die betrachtet wird, und die Dateneinteilung. Deshalb wollen wir abschließend noch zwei Varianten der Betrachtungsperspektive andeuten. Zum einen zeigen wir einen Überblick über die Zusammensetzung nach Lemmata, so wie sie ein automatischer Lemmatisierer (ohne Kontextbetrachtung) vorgeschlagen hat (vgl. Belica 1994).

	> 7	> 12	> 13	> 14	gesamt
Texte	21	4	4	6	35
Token	233.118	156.747	237.777	237.204	864.846
Type	15.280	18.846	23.264	21.449	47.666
Lemmata	10.335	12.112	14.723	13.381	28.617

Tab. 8: Ausgewählte Kennzahlen im Vergleich zu Lemmata



Tab. 9: Lemmahäufigkeiten in den verschiedenen Teil- bzw. Schnittmengen (vgl. Tab. 2)

Das Verhältnis Lemma zu Wortform ist bei den verschiedenen Altersstufen-Gesamtwortschätzen etwa 1: 1,5. Bei den Schnitt-/Teilmengen sieht es sehr unterschiedlich aus: die Gesamtschnittmenge ist nur auf 80 % reduziert, bei ‚Stufe 7-spezifisch‘ auf 64 %, die übrigen Einzelwortschätze auf 53–55 %, alle anderen Mengen dazwischen. Diese Abweichungen können zwei mögliche Erklärungen haben: Entweder verwenden die Altersstufen unterschiedliche Ausschnitte der Flexionsparadigmen (hier eher Tempus als Person¹⁴) oder der Anteil der flektierenden Wortklassen ist unterschiedlich.

Die Parallelen zwischen den höheren Altersstufen deuten aber auch darauf hin, dass deren Unterschiede weniger in dem vermuteten Sprachstand liegen als vielmehr in den Handlungen der Geschichten. Die Altersstufe (10-) 12 enthält erste Hinweise auf die „erste Liebe“ und Romantik, die Stufe 13 thematisiert

14 Der Lemmatisierer kann Formen, die formgleich zu einer flektierten Form des Partizips sind, fälschlicherweise als Adjektiv ein- und somit einem anderen Lemma zugeordnet haben.

Probleme im Jugendalter, u. a. mit Drogen, Stufe 14 zeigt Spuren von Fantasy-Romanen. Vielleicht spielt hier eher der Geschmack oder die Vorliebe der Altersstufe (auch in Verbindung mit dem hier nicht betrachteten Geschlecht) eine größere Rolle für die Altersempfehlung als das antizipierte Wortschatzverständnis.

	7-9	10-15
Token	233.118	631.728
Texte	21	14
Textlänge Ø	11.101	45.123
Type	15.280	41.976
TTR	0,066	0,066
Hapaxe	7.516	21.799
Hapax %	49,19 %	51,93 %
Lemmata	10.335	24.956
Lemma-Hapaxe	4.677	11.337
L-Hapax %	45,25 %	45,43 %

Tab. 10: Kennzahlen anhand anderer Dateneinteilung

Gruppieren wir die Daten anders, verlieren wir zwar unsere Feinkörnigkeit, jedoch deutet sich anhand der Kennzahlen eine bessere Vergleichbarkeit an (vgl. Tabelle 10).

Trotz allem bleibt natürlich die spannendste Frage, der außerhalb der beschriebenen Phänomene weiter nachzugehen wäre: Welches sind die 5.690 Token bzw. 3.661 Lemmata, die später nicht mehr beobachtet werden?

7. Fazit

Um den rezeptiven Wortschatz einer bestimmten Altersstufe zu erheben, ist der beschriebene Ansatz eine gute, und wenn nicht die einzige, so auf jeden Fall eine sinnvoll ergänzende Methode zu anderen Vorgehen. Um dem Verdacht genauer nachgehen zu können, dass die Texte für die höheren Altersstufen aus anderen Gründen als dem Wortschatzverständnis so empfohlen wurden, bedarf es einer breiteren Datengrundlage und weitergehenden Auswertungen.

Mit den durch die Kennzahlen ermöglichten Sortier- und Filtermöglichkeiten haben wir gute Voraussetzungen geschaffen, um die Eigenschaften der verschiedenen Teilwortschätze zu explorieren. Dabei hat auch geholfen, sich nicht zu früh auf einen Dispersions- und Lemmabegriff festzulegen, der genau genommen für diese sich entwickelnden Wortschätze noch ausgearbeitet werden muss. Aber auch mit diesen konzeptuell geklärten Begriffen ist eine vollstän-

dige Operationalisierung der Wortschatzermittlung nicht zu erwarten, sie wird immer ein großer Teil intellektueller Interpretation und „Handarbeit“ bleiben.

8. Literatur

- Augst, Gerhard (1984): *Kinderwort. Der aktive Kinderwortschatz (kurz vor der Einschulung)*. Frankfurt a. M.: Peter Lang.
- Augst, Gerhard (2019): *Der Bildungswortschatz. Darstellung und Wörterverzeichnis*. Hildesheim: Olms.
- Belica, Cyril (1994): *A German Lemmatizer. Final Report MLAP93-21/WP2*. Luxemburg.
- Glaboniat, Manuela/Müller, Martin/Rusch, Paul/Schmitz, Helen/Wertenschlag, Lukas (2002): *Profile deutsch: gemeinsamer europäischer Referenzrahmen; Lernzielbestimmungen, Kannbeschreibungen, kommunikative Mittel, Niveaus A1, A2, B1, B2*. Berlin: Langenscheidt.
- Gries, Stefan Th. (2008): *Dispersions and adjusted frequencies in corpora*. In: *IJCL*, 13 (4), 403–437.
- Herberg, Dieter/Kinne, Michael/Steffens, Doris (2004): *Neuer Wortschatz. Neologismen der 90er Jahre im Deutschen. Unter Mitarbeit von Elke Tellenbach und Doris al-Wadi. (= Schriften des Instituts für Deutsche Sprache; 11)*. Berlin/New York: de Gruyter.
- Höhle, Barbara (Hrsg.) (2010): *Psycholinguistik. (Studienbuch Sprachwissenschaft)*. Berlin: Akademie Verlag.
- Kaeding, Friedrich Wilhelm (Hrsg.) (1897/98): *Häufigkeitwörterbuch der deutschen Sprache 1, 2*. Berlin-Steglitz: Selbstverlag. (Teilreprint in: *Grundlagenstudien aus Kybernetik und Geisteswissenschaft* 4/1963, Beiheft).
- Kegel, Gerd (1974): *Sprache und Sprechen des Kindes. (rororo studium; 56)*. Reinbek: Rowohlt.
- Klann-Delius, Gisela (1999): *Spracherwerb. (Sammlung Metzler; 321)*. Stuttgart: Metzler.
- Nation, Ian S. Paul (2016): *Making and Using Word Lists for Language Learning and Testing*. Amsterdam: Benjamins.
- Schnörch, Ulrich (2015): *Wortschatz*. In: Ulrike Haß, Petra Storjohann (Hrsg.): *Handbuch Wort und Wortschatz. (= Handbücher Sprachwissen; 3)*. Berlin/Boston: de Gruyter, 3–26.

- Schroeder, Sascha/Würzner, Kay-Michael/Heister, Julian/Geyken, Alexander/Kliegl, Reinhold (2015): childLex: A lexical database of German read by children. *Behavior Research Methods*, 47, 1085–1094.
- Steffens, Doris/al-Wadi, Doris (2015): *Neuer Wortschatz. Neologismen im Deutschen 2001–2010*. 3. Aufl. Mannheim: Institut für Deutsche Sprache.
- Thorndike, Edward L. (1921): *The teacher's word book*. New York: Teachers College, Columbia University.
- Zimmer, Dieter E. (1988): *So kommt der Mensch zur Sprache*. (Haffmans TaschenBuch 16, Erstveröffentlichung 1986). Zürich: Haffmans.

Anhang: Zusammensetzung des ‚Carlsen-Korpus‘

(> 7):

- Boehme, Julia (2011): *Conni auf dem Reiterhof*. Conni-Erzählbände, Band 1, (Erstveröffentlichung 2002). Hamburg: Carlsen-Verlag.
- Boehme, Julia (2011): *Conni und der Liebesbrief*. Conni-Erzählbände, Band 2, (Erstveröffentlichung 2002). Hamburg: Carlsen-Verlag.
- Boehme, Julia (2011): *Conni geht auf Klassenfahrt*. Conni-Erzählbände, Band 3, (Erstveröffentlichung 2003). Hamburg: Carlsen-Verlag.
- Boehme, Julia (2011): *Conni feiert Geburtstag*. Conni-Erzählbände, Band 4, (Erstveröffentlichung 2003). Hamburg: Carlsen-Verlag.
- Boehme, Julia (2011): *Conni reist ans Mittelmeer*. Conni-Erzählbände, Band 5, (Erstveröffentlichung 2004). Hamburg: Carlsen-Verlag.
- Boehme, Julia (2011): *Conni und der verschwundene Hund*. Conni-Erzählbände, Band 6, (Erstveröffentlichung 2004). Hamburg: Carlsen-Verlag.
- Boehme, Julia (2011): *Conni rettet Oma*. Conni-Erzählbände, Band 7, (Erstveröffentlichung 2005). Hamburg: Carlsen-Verlag.
- Boehme, Julia (2011): *Conni und das Geheimnis der Kois*. Conni-Erzählbände, Band 8, (Erstveröffentlichung 2006). Hamburg: Carlsen-Verlag.
- Boehme, Julia (2011): *Conni und die Jungs von nebenan*. Conni-Erzählbände, Band 9, (Erstveröffentlichung 2007). Hamburg: Carlsen-Verlag.
- Boehme, Julia (2011): *Conni und das ganz spezielle Weihnachtsfest*. Conni-Erzählbände, Band 10, (Erstveröffentlichung 2007). Hamburg: Carlsen-Verlag.

- Boehme, Julia (2011): Conni und das Hochzeitsfest. Conni-Erzählbände, Band 11, (Erstveröffentlichung 2008). Hamburg: Carlsen-Verlag.
- Boehme, Julia (2011): Conni in der großen Stadt. Conni-Erzählbände, Band 12, (Erstveröffentlichung 2008). Hamburg: Carlsen-Verlag.
- Boehme, Julia (2011): Conni und die verflixte 13. Conni-Erzählbände, Band 13, (Erstveröffentlichung 2009). Hamburg: Carlsen-Verlag.
- Boehme, Julia (2011): Conni und der Dinoknochen. Conni-Erzählbände, Band 14, (Erstveröffentlichung 2009). Hamburg: Carlsen-Verlag.
- Boehme, Julia (2011): Conni und das tanzende Pony. Conni-Erzählbände, Band 15, (Erstveröffentlichung 2010). Hamburg: Carlsen-Verlag.
- Boehme, Julia (2010): Conni und der große Schnee. Conni-Erzählbände, Band 16. Hamburg: Carlsen-Verlag.
- Boehme, Julia (2011): Conni rettet die Tiere. Conni-Erzählbände, Band 17. Hamburg: Carlsen-Verlag.
- Boehme, Julia (2011): Conni und die Detektive. Conni-Erzählbände, Band 18. Hamburg: Carlsen-Verlag.
- Boehme, Julia (2012): Conni und der Ferienzirkus. Conni-Erzählbände, Band 19, (Erstveröffentlichung 2011). Hamburg: Carlsen-Verlag.
- Boehme, Julia (2015): Conni geht zum Film. Conni-Erzählbände, Band 26. Hamburg: Carlsen-Verlag.
- Boehme, Julia (2016): Conni und das Ponyabenteuer. Conni-Erzählbände, Band 27. Hamburg: Carlsen-Verlag.

(> 10/> 12):

- Rotfuß, Veronika (2009): Mücke im März, (Erstveröffentlichung 2008). Hamburg: Carlsen-Verlag.
- Steinhöfel, Andreas (2010): Paul Vier und die Schröders, (Erstveröffentlichung 1992). Hamburg: Carlsen-Verlag.
- Steinhöfel, Andreas (2010): Trägerische Stille, (Erstveröffentlichung 1993). Hamburg: Carlsen-Verlag.
- Steinhöfel, Andreas (2011): Defender, (Erstveröffentlichung 2001). Hamburg: Carlsen-Verlag

(> 13):

- Ottenschläger, Madlen (2013): High Love. Hamburg: Carlsen-Verlag.

- Steinhöfel, Andreas (2010): Die Mitte der Welt. Jubiläumsausgabe, (Erstveröffentlichung 1998). Hamburg: Carlsen-Verlag.
- Vogel, Maja von (2010): Mein neues Leben und ich. Hamburg: Carlsen-Verlag.
- Wich, Henriette (2010): Frosch oder Prinz? Du entscheidest!. Sieben Chancen für die Liebe, (Erstveröffentlichung 2008). Hamburg: Carlsen-Verlag.

(> 14/> 15):

- Dürig, Regina (2011): Katertag. oder: Was sagt der Knopf bei Nacht?. Hamburg: Carlsen-Verlag.
- Fast, Valentina (2015): Royal, Band 1. Ein Leben aus Glas. Hamburg: Carlsen-Verlag.
- Hasse, Stefanie (2015): Darian & Victoria, Band 2. Düstere Schatten. Hamburg: Carlsen-Verlag.
- Kreller, Susan (2012): Elefanten sieht man nicht. Hamburg: Carlsen-Verlag.
- Schulz, Hermann (2010): Iskender, (Erstveröffentlichung 1999). Hamburg: Carlsen-Verlag.
- Steinhöfel, Andreas (2010): Der mechanische Prinz, (Erstveröffentlichung 2003). Hamburg: Carlsen-Verlag.