THE MENTAL LEXICON IN ACQUISITION

- ASSESSMENT, SIZE & STRUCTURE-

Dissertation

zur Erlangung des akademischen Grades

Doktor der Philosophie

(Dr. phil)

eingereicht bei der

Humanwissenschaftlichen Fakultät der Universität Potsdam,

Department Linguistik

von

Jutta Trautwein, M.Sc.

Potsdam, Januar 2019

Published online at the
Institutional Repository of the University of Potsdam:
https://doi.org/10.25932/publishup-43431
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-434314

Erstgutachter: Prof. Dr. Sascha Schroeder, Universität Göttingen,
zuvor Max-Planck-Institut für Bildungsforschung Berlin

Zweitgutachterin: Prof. Dr. Christina Kauschke, Universität Marburg

Tag der Disputation: 12.08.2019

Danksagung

Mein herzlicher Dank gilt…

… Sascha Schroeder als meinem Doktorvater für die spannende, lehrreiche Zeit am MPI in Berlin und die weitere Begleitung bis zur Fertigstellung dieser Arbeit.

… Christina Kauschke für die kurzfristige und unkomplizierte Übernahme des Zweitgutachtens.

… Barbara Höhle für die Herstellung des Kontakts zur Universität Potsdam sowie der weiteren Prüfungskommission bestehend aus Isabell Wartenburger, Reinhold Kliegl und Nicole Stadie.

… dem gesamten Team REaD, insbesondere dem „coolen Büro“, für drei tolle Jahre, viele konstruktive Team-Meetings und die gegenseitige Unterstützung in dieser herausfordernden Zeit.

… allen Schulen, Lehrkräften, Kindern und Erwachsenen, die sich bereit erklärt haben, an meinen Datenerhebungen teilzunehmen.

… meinen Eltern Elisabeth und Ludwig Segbers dafür, dass ihr mir auf meinem bisherigen Lebensweg so vieles ermöglicht und mich stets mit großem Interesse und liebevoller Unterstützung begleitet habt und das auch weiterhin tut.

… meinem Mann Jan dafür, dass du mich immer wieder mit aufmunternden Worten und Gesten zum Weitermachen angespornt hast. Ohne dich wäre diese Arbeit nicht vollendet worden!

ZUSAMMENFASSUNG

Das mentale Lexikon wird als individueller Speicher, der semantische, orthographische und phonologische Informationen über alle bekannten Wörter enthält, verstanden. Die lexikalischen Einträge sind aufgrund von Ähnlichkeiten auf diesen Sprachebenen im Sinne einer Netzwerkstruktur verbunden. Bei der Sprachverarbeitung von Wörtern und Sätzen müssen die Informationen aus dem mentalen Lexikon abgerufen werden. Sind diese nicht oder nur teilweise vorhanden, ist der Prozess erschwert. Die Beschaffenheit des mentalen Lexikons ist damit zentral für sprachliche Fähigkeiten im Allgemeinen, welche wiederum essenziell für den Bildungserfolg und die Teilhabe an der Gesellschaft sind. Die Erfassung des mentalen Lexikons und die Beschreibung seiner Entwicklung ist demnach ein wichtiger Schwerpunkt linguistischer Forschung.

Im frühen Kindesalter ist es noch relativ einfach, den Umfang und Inhalt des mentalen Lexikons eines Individuums zu erfassen – dies kann beispielsweise durch Befragung der Eltern oder durch Aufzeichnung von Äußerungen erfolgen. Mit steigendem Inhalt wird diese Messung allerdings schwieriger: Umso mehr Wörter im mentalen Lexikon gespeichert sind, umso unmöglicher wird es, sie alle abzufragen bzw. zu testen. Dies führt dazu, dass es nur wenige Methoden zur Erfassung lexikalischer Eigenschaften nach Schuleintritt gibt. Aus diesem Grund bestehen auch nur wenige aktuelle Erkenntnisse über den Verlauf der lexikalischen Entwicklung in diesem Alter sowie deren spezifischen Einfluss auf andere (sprachliche) Fähigkeiten. Diese Lücke sollte in der vorliegenden Dissertation geschlossen werden. Dazu wurden zwei Ziele verfolgt: Zum einen sollte eine aussagekräftige Methode entwickelt werden, mit der Umfang und Inhalt des Wortschatzes von Kindern im Grundschulalter bis ins Erwachsenenalter bestimmt werden können. Zum anderen sollten die Ergebnisse der Methode dazu dienen, den lexikalischen Erwerb nach Schuleintritt genauer zu beschreiben und zu verstehen. Dabei wurde neben der Entwicklung der Lexikongröße auch die Struktur des Lexikons, d.h. die Vernetzung der Einträge untereinander, betrachtet.

Die grundsätzliche Idee der Arbeit beruht auf der Wörterbuch-Methode, bei der eine Auswahl an Wörtern aus einem Wörterbuch getestet und die Ergebnisse auf das gesamte Wörterbuch übertragen werden, um die Lexikongröße einer Person zu bestimmen. In der vorliegenden Dissertation diente das childLex Korpus, das die linguistische Umwelt von Grundschulkindern enthält, als Grundlage. Zunächst wurde ein Wortschatztest entwickelt, der auf dem Korpus basiert. Anschließend wurde das Testverhalten von virtuellen Versuchspersonen

simuliert, indem verschiedene Lexikongrößen aus dem Korpus gezogen wurden und überprüft wurde, welche der Items aus dem Wortschatztest in den Lexika enthalten waren. Dies ermöglichte die Bestimmung der Beziehung zwischen dem Verhalten im Wortschatztest und der absoluten Lexikongröße und ließ sich dann auf tatsächliche Studienteilnehmer übertragen. Neben der Wortschatzgröße konnten mit dieser Methode auch der wahrscheinliche Inhalt des mentalen Lexikons und so die Vernetzung des Lexikons zu verschiedenen Entwicklungszeitpunkten bestimmt werden.

Drei Studien wurden konzipiert, um die vorgestellten Ziele zu erreichen und die präsentierte Methode zu etablieren. Studie 1 diente der Entwicklung des Wortschatztests, der auf den childLex Korpus beruht. Hierzu wurde das Ja/Nein-Testformat gewählt und verschiedene Versionen für unterschiedliche Altersgruppen erstellt. Die Validierung mithilfe des Rasch-Modells zeigt, dass der Test ein aussagekräftiges Instrument für die Erfassung des Wortschatzes von Grundschuldkindern im Deutschen darstellt. In Studie 2 werden der darauf basierende Mechanismus zur Schätzung von Lexikongrößen sowie Ergebnisse zu deren Entwicklung vom Grundschul- bis ins Erwachsenenalter präsentiert. Es ergaben sich plausible Ergebnisse in Bezug auf die Wortschatzentwicklung, die einer quadratischen Funktion folgt und mit etwa 6000 Wörtern in der ersten Klasse beginnt und im Durschnitt 73.000 Wörter im jungen Erwachsenenalter erreicht. Studie 3 befasste sich mit den lexikalischen Inhalten in Bezug auf die Netzwerkstruktur des mentalen Lexikons in verschiedenen Altersgruppen. Dabei zeigt sich, dass die orthographische Vernetzung des mentalen Lexikons im Erwerb abnimmt.

Zusammengenommen liefert die Dissertation damit einen innovativen Ansatz zur Messung und Beschreibung der Entwicklung des mentalen Lexikons in der späteren Kindheit. Die Studien bieten aktuelle Ergebnisse zum lexikalischen Erwerb in einer Altersgruppe, in der dazu bisher wenige Erkenntnisse vorlagen. Die Ergebnisse zeigen eindrücklich, wie wichtig diese Phase für den Wortschatzerwerb ist und legen außerdem nahe, dass es starke interindividuelle Unterschiede im lexikalischen Erwerb gibt. Diesen entgegenzuwirken ist eines der Ziele zukünftiger Forschung und Bildung. Zudem ergeben sich aus der Dissertation vielfältige Möglichkeiten zur Anwendung der Methode sowohl zu Forschungszwecken, z.B. in Bezug auf die Übertragung auf andere Zielgruppen und den Effekt der Ergebnisse auf andere Fähigkeiten, als auch für die pädagogische Arbeit, z.B. für die Anpassung von Texten an bestimmte Zielgruppen.

ABSTRACT

The individual's mental lexicon comprises all known words as well related information on semantics, orthography and phonology. Moreover, entries connect due to similarities in these language domains building a large network structure. The access to lexical information is crucial for processing of words and sentences. Thus, a lack of information inhibits the retrieval and can cause language processing difficulties. Hence, the composition of the mental lexicon is essential for language skills and its assessment is a central topic of linguistic and educational research.

In early childhood, measurement of the mental lexicon is uncomplicated, for example through parental questionnaires or the analysis of speech samples. However, with growing content the measurement becomes more challenging: With more and more words in the mental lexicon, the inclusion of all possible known words into a test or questionnaire becomes impossible. That is why there is a lack of methods to assess the mental lexicon for school children and adults. For the same reason, there are only few findings on the courses of lexical development during school years as well as its specific effect on other language skills. This dissertation is supposed to close this gap by pursuing two major goals: First, I wanted to develop a method to assess lexical features, namely lexicon size and lexical structure, for children of different age groups. Second, I aimed to describe the results of this method in terms of lexical development of size and structure. Findings were intended to help understanding mechanisms of lexical acquisition and inform theories on vocabulary growth.

The approach is based on the dictionary method where a sample of words out of a dictionary is tested and results are projected on the whole dictionary to determine an individual's lexicon size. In the present study, the childLex corpus, a written language corpus for children in German, served as the basis for lexicon size estimation. The corpus is assumed to comprise all words children attending primary school could know. Testing a sample of words out of the corpus enables projection of the results on the whole corpus. For this purpose, a vocabulary test based on the corpus was developed. Afterwards, test performance of virtual participants was simulated by drawing different lexicon sizes from the corpus and comparing whether the test items were included in the lexicon or not. This allowed determination of the relation between test performance and total lexicon size and thus could be transferred to a sample of real participants. Besides lexicon size, lexical content could be approximated with this approach and analyzed in terms of lexical structure.

To pursue the presented aims and establish the sampling method, I conducted three consecutive studies. Study 1 includes the development of a vocabulary test based on the childLex corpus. The testing was based on the yes/no format and included three versions for different age groups. The validation grounded on the Rasch Model shows that it is a valid instrument to measure vocabulary for primary school children in German. In Study 2, I established the method to estimate lexicon sizes and present results on lexical development during primary school. Plausible results demonstrate that lexical growth follows a quadratic function starting with about 6,000 words at the beginning of school and about 73,000 words on average for young adults. Moreover, the study revealed large interindividual differences. Study 3 focused on the analysis of network structures and their development in the mental lexicon due to orthographic similarities. It demonstrates that networks possess small-word characteristics and decrease in interconnectivity with age.

Taken together, this dissertation provides an innovative approach for the assessment and description of the development of the mental lexicon from primary school onwards. The studies determine recent results on lexical acquisition in different age groups that were missing before. They impressively show the importance of this period and display the existence of extensive interindividual differences in lexical development. One central aim of future research needs to address the causes and prevention of these differences. In addition, the application of the method for further research (e.g. the adaptation for other target groups) and teaching purposes (e.g. adaptation of texts for different target groups) appears to be promising.

TABLE OF CONTENTS

I THEORETICAL FRAMEWORK

1 Introduction

How many words do you know?

This seemingly simple question appears to be almost impossible to answer on second thought. Moreover, we cannot precisely say which words we know and how well we know them. However, all this information is stored in our so-called mental lexicon and is crucial for language processing: To understand and produce words, sentences, texts and conversations, we need to retrieve information from our mental lexicon, on the meaning, the pronunciation and/or the spelling of words.

Because of its central role in language development and language processing, the study of the mental lexicon has occupied many researchers in the past decades, especially concerning language acquisition. Vocabulary partly reflects children's cognitive skills and their understanding of the world and is thus an important determinant for development in general. In addition, it is related to later reading skills and educational achievement. For these reasons, the measurement of vocabulary and the description of its development and interindividual differences as well as the investigation of effective training methods have been of increasing interest in linguistic, psychological and educational research.

However, there are still some open questions, especially concerning lexical development from later childhood to adulthood, mainly because of methodological issues. For very young children starting to learn a language, the amount of information stored in the mental lexicon is limited, thus its assessment is comparably easy. But with proceeding language development, vocabulary grows immensely and hence, its measurement becomes quite challenging. In fact, little is known about the development of the size and the structure of the mental lexicon during primary school and onwards. Only a few studies aimed to estimate total vocabulary and come to quite different results. Thus, until now, the course of lexical development from primary school onwards is still indistinct, although we know that school education strongly affects vocabulary development. Knowing how the mental lexicon enhances during this critical period of time, and being able to detect and counteract interindividual differences could help promote children's educational achievement, to enable them to reach their personal goals and thus to improve their possibilities to participate in society.

This thesis aims to be a first step towards the measurement of the mental lexicon and its development in late childhood. For this purpose, I develop an approach to estimate the size and content of the mental lexicon for different age groups from primary school to adulthood. Based on this method, I present results on the development of lexicon size and lexical structure at different time points of language acquisition.

The theoretical framework starts out with a definition of the mental lexicon, its organization and acquisition. This is followed by a description of previous approaches on the assessment of the mental lexicon including the introduction of graph theory to investigate lexical organization. Subsequently, I present former findings on the development of lexicon size, lexical structure and lexical quality. The theoretical framework ends with a brief summary and essential open research questions. Based on this background, I will derive my research aims assigned to three studies[1]. After describing the idea of the general method, I present the three studies separately. The final chapter includes a general discussion of the results as well as their theoretical and practical implications along with limitations and future lines of research. Finally, I will draw the central conclusions derived from this thesis.

2 The Mental Lexicon

Many studies have shown the impact of vocabulary on other language and literary skills (e.g. Rowe, Raudenbusch & Goldin-Meadow, 2012; Lee, 2011; Ouellette, 2006; Walley, Metsala & Garlock, 2003). Since language and literary skills are central for educational achievement (e.g. Hoff, 2014; Walker, Greenwood, Hart & Carta, 1994; Graham, 1987), the acquisition of vocabulary, developmental trajectories and external influences have gained a lot of attention within the last decades of linguistic research (e.g. McGregor, Oleson, Bahnsen & Duff, 2013; Szagun, Steinbrink, Franik & Stumper, 2006; Kauschke & Hofmeister, 2002; Anglin, Miller & Wakefield, 1993). In the following sections, I summarize the current status of research on the mental lexicon and derive open questions that need to be answered to fully understand lexical development and its impact on language processing.

[1] The three studies are published in peer-reviewed journals. Thus, each study is composed to be read independently from this dissertation. As a consequence, repetitions between studies and the overall thesis are not avoidable. Study 1 was published in a German journal whereas studies 2 and 3 were published in international journals.

2.1 DEFINITION, ORGANIZATION AND ACQUISITION

In the literature, vocabulary, vocabulary size or vocabulary knowledge are often used to describe the number of known words or the depth of semantic knowledge. The concept of the mental lexicon captures a broader framework including different aspects of word knowledge. Clark (1993) defines the mental lexicon as “a stock of established words speakers can draw on when they speak and have recourse to in understanding what they hear” (p.2) stored in memory. It thus comprises entries for all words a person knows. According to Clark (1993) each lexical entry has to at least contain information about the meaning, syntax, morphology and phonology of a word referring only to spoken language. In the lexical quality hypothesis (Perfetti & Hart, 2002; Perfetti, 2007), orthographic knowledge about the written form of a word is also part of the lexical entry when reading and writing skills are acquired. In this framework, the quality of entries can differ depending on the degree of knowledge in the different areas. Specification within the different parts of word knowledge, as well as strong connections between them are important for a high lexical quality that is a high degree of vocabulary knowledge (Perfetti, 2007).

Features of the mental lexicon are thus its size, in the literature commonly referred to as vocabulary breadth, and its quality of word knowledge, in the literature frequently referred to as vocabulary depth. Both features develop with age and vary among individuals (Perfetti & Stafura, 2014; Richter, Isberner, Naumann & Neeb, 2013; Perfetti, 2007). Some authors, however, claim that because they are highly interconnected, vocabulary size and quality address the same concept of vocabulary in general (Schmitt, 2014; Vermeer, 2001).

An additional feature of the mental lexicon is its organization. It is assumed that lexical entries are interconnected based on similarities in the specific language domains, e.g. concerning semantic (*hungry* – *food*), phonological (*hygrometer* – *hydrometer*) or orthographic (*widow* - *window*) relations (Aitchison, 2012). Findings regarding the effect of these assumed connections on word processing support this theory (e.g. Zielger, Muneaux & Grainger, 2003; Buchanan, Westbury & Burgess, 2001). As an alternative to vocabulary breadth and depth, Merea and Wolter (2004) even suggested to rather assess vocabulary size and organization than breadth and depth. They claimed that “vocabulary knowledge is rather more than the sum of the learners’ knowledge of the individual words in their vocabulary” (p. 88). The authors postulated that the network structure of the mental lexicon is essential for language processing and thus should be considered when assessing vocabulary. Taken together, the mental lexicon can be regarded as an individual network containing different kinds of personalized information on known words (for a summary see also Aitchison, 2012).

Language input is the primary source of vocabulary acquisition: In early language development, oral input provided by the child's social environment determines word learning (Hoff & Naigles, 2002; Huttenlocher et al., 1991). Acquiring entries for the mental lexicon first means isolating word forms in speech and mapping them to meaning, that is semantics. Morphological and syntactic information is added later as well as orthographic knowledge with developing reading skills. Carey (1978) differentiates between the processes of *fast mapping* and *slow mapping*. Fast mapping enables the children to build up a lexical representation (form – meaning) for a word with only a few exposures. That is, the number of known words increases quickly while the new lexical entries still lack deeper word knowledge. With the process of slow mapping, the children then add further information in terms of lexical quality (e.g. deeper semantic knowledge, connections to other words) to the already existing lexical entries (Clark, 1993; Perfetti, 2007). In later development, new words are more likely to occur in written language (Hayes & Aherns, 1988). Hence, reading becomes an important determinant for vocabulary acquisition (Duff, Tomblin & Catts, 2015).

In general, word frequency strongly affects lexical development. On the one hand, frequent words of a language are more likely to be encountered and are thus learned first (Bonin, Barry, Méot & Chalard, 2004; Goodman, Dale & Li, 2008; Rott, 1999). On the other hand, the frequency of exposures to a single word affects the inclusion and improvement of a lexical entry. Carey and Bartlett (1978) found out that for fast mapping, only one exposure to a word can suffice to incorporate it into the mental lexicon for children in early childhood. Similar results were found by Heibeck and Markman (1987) who generalized the finding for different semantic domains as well as Dollaghan (1985) who showed the word learning process for pseudowords. In the context of learning vocabulary through reading in later childhood, Jenkins, Stein and Wysocki (1984) found out that more than two exposures in text were necessary to acquire a word. This was also supported by the study of Nagy, Herman and Anderson (1985). However, to extend the lexical entry and add further word knowledge, more exposures are required (Hart & Perfetti, 2008; Perfetti, 2007; Clark 1993). Perfetti (2007) also claimed that the necessary numbers of exposures to learn words or improve lexical quality differs among individuals according to their language skills. He showed that high skilled individuals defined by reading comprehension had larger gains in lexical quality than low skilled individuals with the same exposure to the same set of stimuli.

To sum up, the mental lexicon can be regarded as a complex word storage containing entries with different aspects of word knowledge as well as connections between entries. Words are acquired via spoken or written language input whereas only a few exposures are sufficient to set up a basic lexical entry. Deeper word knowledge is subsequently added with further encounters of a word. The mental lexicon thus develops during language acquisition and differs among individuals in its size, quality of entries and interconnections. In the following sections, I will summarize previous approaches to assess the mental lexicon as well as findings on the development of these three lexical dimensions.

2.2 ASSESSMENT OF LEXICON SIZE AND STRUCTURE

Studies investigating the mental lexicon share the underlying problem of its assessment. One central question in the measurement of vocabulary is: when do we count a word as known, i.e. stored in the mental lexicon? As elaborated above, the mental lexicon comprises phonological, semantic, and later orthographic information. That is, there are different levels of word knowledge which are assessed in different kinds of vocabulary tests (Miller, 1999).

For young children vocabulary measurement is commonly done with parents' questionnaires where parents identify words their child understands and/or produces out of a list of possible words. The number of identified words marks the lexicon size of the child (e.g., Communicative Development Inventories, Fenson et al., 1993; for German: Elternfragebogen zur Erfassung von Risikokindern, Grimm & Doil, 2005). A comparable method to the parental questionnaires for older participants is the yes/no vocabulary test (Anderson & Freebody, 1983). A list of words is administered to the test taker who has to identify all words he or she knows. To prevent guessing, pseudowords are included in the list. Results indicate how many words are known, that is represent vocabulary size. Although the format has been shown to be useful for children as well as adults (e.g. Mochida & Harrington, 2006; Anderson & Freebody, 1983), there are no actual yes/no vocabulary tests available to assess the mental lexicon for different age groups in German as well as other languages.

In most studies with older children and adults, lexical knowledge is measured via vocabulary tests, e.g. picture naming or multiple choice tasks (Peabody Picture Vocabulary Test, Dunn & Dunn, 2007; for German: Aktiver Wortschatztest, Kiese-Himmel, 2005). Results for a set of items are calculated and compared to those of a large norming sample from the same age group[2]. These vocabulary tests mainly assess semantic knowledge and thus the score provides information on the strength of semantic representations for a certain set of items in comparison to the average test taker. The same goes for tests on phonological awareness (e.g. Test of Phonological Awareness, Torgesen & Bryant, 2004; for German: Test zur Erfassung der phonologischen Bewusstheit und der Benennungsgeschwindigkeit, Mayer, 2016) or orthographic knowledge (e.g. Test of Written Language, Hammill & Larsen, 2009; for German: Hamburger Schreib-Probe, May, 2012) for the respective language domain. Hence, they do not provide information on the total mental lexicon and are mainly intended to identify language difficulties.

With growing vocabulary size, the measurement of the total mental lexicon, its size and structure becomes more and more difficult. After a certain amount of words is stored in the mental lexicon, it is simply impossible to count or test them all. For this reason, two kinds of methods are available to assess the size and content of the whole mental lexicon, i.e. methods based on usage and sampling-based dictionary methods (Lorge & Chall, 1963). Both methods aim to approximate the total vocabulary size of a target group.

In usage-based methods, spoken or written language production of the group of interest is analyzed and the number of different words is counted (e.g. Pregel & Rickheit, 1987; Marah, 1872). However, this measure is costly and does not provide estimates for the vocabulary size of individuals. Hence, researchers have focused on sampling-based procedures (e.g. Anglin, Miller & Wakefield, 1993; Seashore & Eckerson, 1940). A sample of words is drawn from a dictionary or list of words is administered within a vocabulary test. The results are then projected to the whole dictionary or list. The method requires the construction of a valid vocabulary test based on a dictionary or word list comprising all words a language learner could possibly know (Nation, 1993a). None of the existing vocabulary tests available for children from primary school to adulthood in German meets these criteria or enables determination of the test taker's total lexicon size.

In linguistic research, it has also been of interest as on how the content of the mental lexicon is organized. This line of research regards the interconnections between entries and

[2] Additionally, norming samples can frequently be matched via other personal characteristics such as gender or mono- vs. bilingual acquisition.

their effect on word processing as well as the mechanisms behind the development of connections. It is generally assumed that lexical entries are connected due to semantic, phonological or orthographic similarities. Semantic similarities are commonly assessed via word associations (e.g. Zortea & Fumagalli de Salles, 2012) or co-occurrences in texts (e.g. Hills, Maouene, Riordan & Smith, 2010). Phonological and orthographic connections are defined as words that can be created by submitting, adding or deleting one phoneme or grapheme in a target word (e.g. Coltheart, Devalaar, Jonasson & Besner, 1977). Studies on the effect of these lexical "neighbors" on single word processing support the theory that they are related within the mental lexicon (e.g. Andrews, 1992; Zielger, Muneaux & Grainger, 2003; Buchanan, Westbury & Burgess, 2001).

The network structure of the mental lexicon has recently been examined under the use of graph theory, which has already been applied to other research fields such as social networks (e.g. Borgatti, Mehra, Brass & Labianca, 2009) or natural science (e.g. Mason & Verwoerd, 2007). In this framework, regarding the mental lexicon, words are represented by nodes and connections or neighborhoods via paths between nodes (e.g. Vitevitch, 2008; Zortea, Menegola, Villavicencio & Fumagalli de Salles, 2014). Graph theory allows the determination of different network characteristics presented in Table 2.1. They all aim to describe the structure, especially the interconnectivity of a network.

TABLE 2.1: Network measures in graph theory (adapted from Steyvers & Tenenbaum, 2005).

Measure	Definition
n	Number of nodes
L	Average path length of shortest paths between nodes
D	Diameter or maximum path length of the network
C	Clustering Coefficient
k	Degree or number of connections
<*k*>	Average degree
P(*k*)	Distribution of degrees

While the number of nodes provides information on the size of the network, *L* and *D* are measures of interconnectivity referring to the average, respectively maximum number of paths necessary to get from one node to another. In a highly connected network, both variables should thus be low. The clustering coefficient C is a measure of the probability of neighbors of a word to also be neighbors and is thus another measure of interconnectivity. It is calculated over all nodes i via the formula

$$C_i = T_i / \binom{ki}{2} = 2\ T_i / k_i\ (k_i - 1)$$

T_i denominates the number of links between the neighbors k of the node i and k_i (k_i - 1)/2 refers to the number of possible connections if all neighbors of a node were also related. C can thus range between 0 and 1 whereas a value closer to 1 indicates a high interconnectivity. The number of connections can be calculated per node with k also referred to as the degree, and averaged for the whole network with $<k>$. The distribution of degrees P(k) provides information on the occurrence of different values of k and thus gives an overview of neighborhoods in the network (Steyvers & Tenenbaum, 2005).

To investigate lexical organization, e.g. using graph theory, the size and content of the whole mental lexicon needs to be available so that the complete network structure can be generated and examined. In studies on lexical organization, corpus data (e.g. Vitevitch, 2008) or association norms (e.g. Steyvers & Tenenbaum, 2005) are commonly used to determine lexical networks. However, using the whole corpus or database of associations only provides a fully developed network and does not consider interindividual or developmental differences. Knowing how many and which words are stored in mental lexicon at different time points of development or for different individuals would allow the analysis of interindividual differences and developmental patterns. Hence, while this is possible for small lexicon sizes in early childhood (Beckage, Smith & Hills, 2011; Hills et al., 2009), it becomes more challenging with growing lexicon size since there is a lack of appropriate methods to measure total vocabulary. Although it has been shown that network structure is relevant for language development and language processing (see section 2.4), up to now, no methods to generate age-specific or individual lexical networks exist.

In conclusion, the assessment of the whole mental lexicon, i.e. its size and organization becomes more and more challenging with growing lexicon size. In the past, usage-based as well as sampling-based approaches have been employed for different age groups. However, no recent methods to determine lexicon size or lexical structure for children or adults in German are available.

2.3 DEVELOPMENT OF LEXICON SIZE

Despite the methodological issues presented in section 2.2, many researchers have attempted to measure and report numbers on total lexicon size for different age groups (see Table 2.2 with a classification of the used methods presented in section 2.2). Knowing how many words a typical language learner at a certain age has stored in his or her mental lexicon enables to understand the course of language acquisition (e.g. Anglin, Miller & Wakefield, 1993; Smith, 1941) as well as the relation of the mental lexicon to other language and cognitive skills (e.g. Lee, 2011; Ouellette, 2006). In addition, the comparison of lexicon size of individuals to those of peers of the same age or developmental status can provide information about the individual's language acquisition and can thus be used for screenings or diagnoses of language impairment (e.g. Rescorla, Mirak & Singh, 2000; Mayne, Yoshinaga-Itano, Sedey & Carey, 1998).

With eight to ten months, children begin to understand first words and start producing speech at the age of about twelve months (Fenson et al., 1994). With about 24 months, they can use about 200 words (Hoff, 2014). Lexicon size strongly increases onwards via fast mapping (see section 2.1). As Carey (1978) stated, children acquire about 9 words per day based on the assumption that children aged six have learned over 14,000 words. Anglin, Miller and Wakefield (1993) even estimated a growth rate of 20 words per day for children from grade 1 to grade 5. This increasing growth process is due to the development of reading skills and the receiving of school education as additional sources for vocabulary acquisition (Duff, Tombiln & Catts, 2015). However, due to methodological issues reported in section 2.2, studies on total lexicon sizes for children from primary school onwards and adults are limited and differ widely as presented in Table 2.2.

For English adults, Aitchison (2012) names a size of 50,000 words while Seashore and Eckerson (1940) calculated a mean total lexicon size of 155,000 words for undergraduate students. For English children, Anglin, Miller and Wakefield (1993) and Smith (1941) found similar results with about 10,000 lexical entries for first-graders, about 20,000 for third- and about 40,000 for fifth-graders. However, numbers for German children provided by Pregel and Rickheit (1987) are much smaller with an estimated vocabulary size up to 6900 words without any specifications for different age groups. Their results are more comparable to those of Biemiller and Slonim (2001) who estimated about 2700 words for children at the beginning of school, and about 8400 words in 5th grade.

TABLE 2.2: Total lexicon sizes reported from different authors for different age groups.

Authors	Target group	Estimates	Method
Marah (1872)*	Adults	3000 - 10,000 words	Usage
Seashore & Eckerson (1940)	Adults	155,000 words	Dictionary
Hartman (1946)	Adults	215,040 words	Dictionary
Goulden, Nation & Read (1990)	Adults	17,000 words	Dictionary
D'Anna, Zechmeister & Hall (1991)	Adults	16,785 words	Dictionary
Aitchison (2012)	Adults	50,000 words	unknown
Smith (1941)	Children, 1st grade	21,000 words	Dictionary
	Children, 3rd grade	38,000 words	
	Children, 5th grade	43,000 words	
Carey (1978)	Children, aged 6	14,000 words	unknown
Pregel & Rickheit (1987)**	Children, G, aged 6-10	6900 words	Usage
Anglin, Miller & Wakefield (1993)	Children, 1st grade	10,000 words	Dictionary
	Children, 3rd grade	20,000 words	
	Children, 5th grade	40,000 words	
Biemiller & Slonim (2001)	Children, 1st grade	2700 words	Dictionary
	Children, 2nd grade	5000 words	
	Children, 4th grade	6800 words	
	Children, 5th grade	8400 words	

** According to Seashore & Eckerson (1940)*

***This study addressed German-speaking participants. All other studies comprise English-speaking participants.*

Only few of the reported studies also address the issue of interindividual differences in lexicon size claiming that the number of known words strongly varies among individuals.

For example, Anglin, Miller and Wakefield (1993) found an increase of standard deviations for total lexicon sizes with entries from about 7,000 to 13,000 in first, 17,000 to 23,000 in third and even 27,000 to 53,000 in fifth grade. Biemiller and Slonim (2001) also discovered high standard deviations with up to 2,000 words per grade. According to Anglin, Miller and Wakefield (1993), the growth of interindividual differences could be explained by the so-called *Matthew effect*. It claims that individuals with high language skills tend to learn more and faster than individuals with low language leading to a widening gap between different skilled individuals (see also Stanovich, 2009). One central goal of education is to prevent this gap and thus to enable and facilitate participation in society for every individual. For this purpose, one must find a way to determine interindividual differences in vocabulary size, ascribe them to possible causes and eventually counteract these causes.

However, not only do most of the numbers presented vary strongly between studies, but most of them are outdated and address English-speaking or English-learning individuals. In addition, most of them do not provide numbers on interindividual differences for different age groups. That is, no recent numbers on lexicon size for children and adults in German are available although it is known that vocabulary grows strongly from school entry onwards. Nevertheless, these numbers are fundamental to understand and support developmental processes and to elaborate the effect of total vocabulary size on other cognitive skills.

2.4 Development of Lexical Structure and its Effect on Language Processing

Network structures in the mental lexicon have been examined regarding different language domains, mostly using graph theory (see section 2.2). For semantic networks, Steyvers and Tenenbaum (2005) used collected association data by Nelson, McEvoy and Schreiber (1999) as well as mappings between words and semantic categories (Roget, 1911) and mappings between words and meanings (WordNet, Miller, 1995) to establish a general semantic network for English adults. The authors found so-called *small-world characteristics*, that is a short average path length between words and a high possibility that neighbors of a word are also neighbors themselves. Furthermore, small-world networks possess a scale-free structure where a few nodes have many connections while many nodes have only a few connections. Beckage, Smith and Hills (2011) discovered comparable characteristics for lexicons of English children aged 15 to 36 months. They used parent’s checklists to determine children’s vocabularies and applied co-occurrence statistics to establish semantic connections.

For phonological networks of English adults, Vitevitch (2008) found one large component, several smaller connected components as well as lexical hermits, which were not connected to the rest of the network. For the large component, he also detected small-world characteristics. The author constructed the network by using a 20,000 word sample from a dictionary as an approximation of the adult lexicon. Connections were based on phonological similarity, that is a phonological neighbor was be defined by substituting, adding or deleting a phoneme of a given word. Arbesman, Strogatz and Vitevitch (2010) generalized the finding of small word characteristics for phonological networks to other languages, namely Spanish, Mandarin, Hawaiian and Basque. Unfortunately, no studies on the composition of orthographic networks in German are available up to now although they could give insight on the developmental processes of orthographic learning (but see for English: Siew, 2018).

The small-world structure has been of particular interest in network research because it can yield information on how the network develops. Barabási and Albert (1999) argue that new words are more likely to be connected to nodes with a high degree so that small-world characteristics arise. This process is named *preferential attachment* and provides information on which words are learned next. However, for early semantic networks, Hills et al. (2009) could show that words that possess many semantic neighbors in the learning environment are learned first and name this process *preferential acquisition*. They assume that these words represent prominent concept within the language and are thus important to learn. Zortea, Menegola, Villavicencio and Fumagalli de Salles (2014) compared semantic networks for children, adults and elderly in Portuguese. They found an increase of semantic neighbors and interconnectivity from childhood to adulthood indicating an improvement of lexical semantic structure.

Only few studies examined the development of phonological networks. Storkel (2004) could show that words with many phonological neighbors are also learned earlier (see also Vitevitch & Storkel, 2012; Stamer & Vitevitch, 2012). Charles-Luce and Luce (1990) demonstrated that the number of phonological neighbors per word increased from childhood to adulthood. However, further studies on the developmental processes of phonological networks, e.g. under the use of graph theory are still missing.

For orthographic network development, no comparable studies employing graph theory are available. Using an entirely different method, Castles, Davis and Letcher (1999) however, could show that children knew fewer orthographic neighbors than adults. They presented words with high and low orthographic neighborhoods to children and adults and asked them to identify neighbors in a list of actual orthographic neighbors as well as nonwords. Identified neighbors were considered as the “effective neighborhood size” and were smaller for children than for adults for all words. The findings implicate that orthographic networks emerge from childhood to adulthood. Yet, the developmental course is still unknown.

Besides the underlying processes of language development, the analysis of the structure of the mental lexicon can also be used to explain phenomena in language processing and vice versa. In particular, studies have shown that neighbors of a word influence word retrieval, e.g. in priming experiments with the neighbor as a prime for the target word. This holds for the semantic domain (e.g. Locker, Simpson & Yates, 2003; Holderbaum & Fumagalli de Salles, 2001), the phonological domain (e.g. Mulatti, Reynolds & Besner, 2006; Ziegler, Muneaux, 2007) as well as the orthographic domain (e.g. Andrews, 1992, Balota et al., 2007). However, results on the nature of the effect are controversial with some studies finding facilitative and others inhibitory effects (for a review regarding the effect of orthographic neighbors, see Andrews, 1997). In general, effects are explained by the network structure of the mental lexicon: Neighbors either boost or inhibit activation of the target word via paths between the words. Lately, the use of measure of graph theory has been shown to be fruitful in the study of the effect of network structure on language processing, mostly for the phonological domain (Chan & Vitevitch, 2009; Chan & Vitevitch, 2010; Siew & Vitevitch, 2016; for orthography see Siew, 2018).

Hence, neighborhoods are typically derived from corpora (e.g. by Kučera & Francis, 1967) and thus regard connections for an individual knowing all words included in the certain corpus. Since the size and composition of the mental lexicon develops over time and strongly varies between individuals (see sections 2.1 and 2.3), the use of the same neighborhood measures for all individuals might not be suitable. Still, no age-specific or individual measures of neighborhoods or networks are available for either of the language domains. To determine those, it is necessary to know how many and which words are stored in the mental lexicon at different time points in development.

To sum up, it has been of great interest how the mental lexicon is structured since the lexical organization provides information on developmental processes as well as the influence of the network structure on language processing. Graph theory has been shown to be beneficial for the description of lexical organization. However, up to now, studies have mostly focused on semantic or phonological networks and their formation and they mainly address very young children or adults with a fully developed lexicon. The description of networks and their development from primary school onwards, especially for orthographic networks, is still missing although it is known that orthographic similarities affect processing of single words. In addition, since the mental lexicon and its organization develop over time, age-specific or individual networks could be useful to describe effects of lexical structure on language processing as well as the courses of developmental processes.

2.5 Development of Lexical Quality

In comparison to the study of lexicon size and its development, there are more and more actual studies on aspects of lexical quality, i.e. lexical knowledge in different domains, and their acquisition[3]. Through more and more experiences with words, children are able to improve their lexical representations via the process of slow mapping as proposed by Carey (1978).

To tap into the development of lexical quality one can consider each domain – semantics, phonology and orthography according to Perfetti and Hart (2002) – separately and analyze developmental changes in each domain.

For semantic knowledge, evidence for developmental growth comes from studies with word definition tasks, which show an improvement of semantic word descriptions with age (e.g. Marinellie & Chan, 2006; Johnson & Anglin, 1995). In addition, studies have shown a twist in semantic categorization from complementary criteria to similarity criteria during childhood indicating a refinement of semantic entries within the mental lexicon (e.g. Roedder John & Sujan, 1990; Waxman & Gelman, 1986).

The study of phonological awareness can provide information on the development of phonological entries in the mental lexicon. Phonological awareness refers to the ability to recognize, identify or manipulate different phonological units such as whole words, phonemes,

[3] For reasons of completeness, this section provides as short overview on findings on the development of lexical quality, although its investigation is not a central aim of this dissertation. For more details on the topic please see Perfetti (2010) and Richter, Isberner, Nauber and Neeb (2013).

rimes or syllables (Anthony & Francis, 2005). Studies have shown that it develops from a shallow sensitivity of large phonological units to a deep awareness of small phonological units (e.g. Anthony & Lonigan, 2004; Anthony, Lonigan, Driscoll, Phillips & Burgess, 2003). This means an improvement of the phonological representations within the mental lexicon during language development.

Concerning orthographic representations, evidence is available from the study of reading development. In the framework proposed by Castles, Davis and Forster (2007), they claim a shift from a broader reading process to a more fine-grained word retrieval meaning an improvement in specificity of orthographic representations.

To sum up, knowledge in all three linguistic areas improves during development. However, the concept of lexical quality exceeds separate knowledge in the three domains; it rather captures the relation between knowledge within these three domains (Perfetti & Hart, 2002). Perfetti and Hart (2002), however, claim that knowledge develops separately at the beginning of language acquisition and becomes more and more associated later in development. A sophisticated lexical quality is necessary especially for reading comprehension, which has been shown mainly for adults (Perfetti & Hart, 2002; Perfetti, 2007). Richter, Isberner, Nauber and Neeb (2013) were also able to demonstrate this association for beginner readers: phonological, orthographic and semantic representations explained nearly 60% of the text comprehension skills of children aged six to ten. In addition, they showed that semantic representations served as a mediator between orthographic/phonological knowledge and text comprehension. The authors thus assume that during development, word form representations (orthographic/phonological) are established first whereas semantic knowledge is integrated later in acquisition. This fits the theory of Carey (1987) claiming that word forms are acquired quickly via fast mapping while semantic knowledge is added gradually via slow mapping.

In summary, there is strong evidence for the improvement of lexical quality during acquisition in that the language domains develop separately at the beginning and are gradually associated later in development.

3 SUMMARY AND RESEARCH QUESTIONS

The mental lexicon contains different aspects of word knowledge and thus plays a central role in language acquisition and language processing. The theoretical framework so far has demonstrated that that vocabulary size and structure affect other language and cognitive skills throughout development and are thus important to consider when investigating courses of interindividual differences or difficulties in these skills (section 2.1).

However, the study of lexical development bears several important methodological issues presented in section 2.2: While the mental lexicon for children at the beginning of language acquisition is comparably easy to asses, the growing number of words known makes counting or testing them all impossible. Usage-based as well as sampling-based approaches have been applied to approximate lexicon sizes for different age groups, but no actual findings on lexical development from primary school to adulthood in German are available. This leads to the central preceding research question:

How can we measure the size and content of the mental lexicon from primary school onwards? Sampling-based methods have been shown to be fruitful when used with a dictionary or list of high quality and an according valid vocabulary test measuring vocabulary size. The question is thus how existing methods could be enhanced to establish a way to measure an individual's total mental lexicon size, content and structure. It also includes the question of which test format is suitable for a valid vocabulary measurement.

On account of these methodological issues, there is a research gap on the course of lexical development, mainly in terms of lexicon size and lexicon structure: While it is known that the quality of lexical entries improves with language development (section 2.5), little is known about the development of lexicon size (section 2.3) and lexical organization (section 2.4) from primary school onwards, although it is indisputable that vocabulary strongly develops during this period due to school education and growing reading skills as a further source of acquisition. Studies on lexicon size are outdated and came to varying results whereas studies on lexical structure do not address the developmental changes of lexical networks and mainly address semantic and phonological structure. The central question derived from the presented theoretical framework is thus: How does the mental lexicon develop from primary school onwards? More precisely, the question bears two subordinate main questions:

How does the size of the mental lexicon develop from late childhood to adulthood? This question addresses the issue of how many words are learned during this critical period as well as the detection of interindividual differences in vocabulary size development. Knowing how many words the average language learner acquires at which point of time can help understanding developmental processes and adjusting teaching and training methods. Comparing individual children to larger norming samples in terms of lexicon size could be useful for screening purposes as well as the investigation of sources for difficulties vocabulary learning and according possible counteractions.

How does the network structure of the mental lexicon develop from primary school onwards? Lexical neighbors have been shown to affect word processing. Few studies considered the analysis of network structures in the mental lexicon, mainly for very young children or for adults and semantic or phonological networks. They found small-world characteristics for lexical networks indicated by a high interconnectivity and interpretations in terms of developmental processes such as preferential attachment. Yet, it is unclear whether these characteristics are persistent or change throughout language development, especially for orthographic networks. Knowing how the mental lexicon is organized and how lexical structure evolves could lead a more precise understanding of emerging connections within the mental lexicon and trajectories driving word acquisition.

A further but subsequent question addresses the effect of vocabulary size and structure: *How does lexical development in late childhood influence other (language) skills?* The effect of vocabulary on other skills such as reading comprehension (Joshi, 2005) or single word reading (Ouellette, 2006) has been shown, but traditionally, measures of semantic knowledge (e.g. definitions or multiple-choice tasks) have been used in these studies. Lexical neighbors have been demonstrated to affect language processing and are commonly determined via corpus measures (e.g. Andrews, 1992). Being able to determine total lexicon sizes and lexical structures for different age groups on an individual level allows measuring the effect of both variables on other skills. This could open new perspectives on the central role of vocabulary in language processing.

By conducting three studies, this dissertation is aimed to address important parts of these research questions to contribute to the understanding of the course of lexical development during late childhood.

4 AIMS OF THE STUDIES

As elaborated above, the preceding research question regarding the mental lexicon in late childhood is the issue of its assessment. The first central aim of this dissertation is thus to develop a method to assess the size and content of the mental lexicon from primary school onwards. The basic idea is to enhance the existing dictionary-based approach for lexicon size estimation by developing a new vocabulary test based on a linguistic language corpus and establishing a corpus-based sampling approach to estimate total vocabulary size.

The second aim is to describe the results according to the question of how lexicon size develops from primary school onwards. The pursuit of this aim is supposed to shed light on the courses of lexical development as well as interindividual differences.

The third aim comprises the assessment of the development of lexical structure using graph theory. It focuses on the emergence of orthographic networks from primary school onwards to provide a basis for research on developmental processes and their effect on language processing.

Three studies were conducted to address these aims. Studies 1 and 2 focus on the development of the method for lexicon size estimation (first aim). In addition, Study 2 contains results on the development of lexicon size (second aim). In Study 3, the development of lexical structure is examined under the use of graph theory (third aim). Since the main focus of the thesis is on late childhood, all studies include German children from first to sixth grade. In addition, Study 2 and Study 3 comprise a sample of children from eighth grade and moreover, Study 2 includes results for adults as well. The specific focus of each study is described below. In addition, Figure 4.1 summarizes the consecutive aims of the three studies.

STUDY 1. As discussed above, the use of the sampling-based approach to estimate total vocabulary size requires the application of a valid vocabulary size test based on a dictionary or word list. One test format that appears to be beneficial for the measurement of vocabulary size is the yes/no vocabulary test (Anderson & Freebody, 1983). The aim of Study 1 is thus to develop a yes/no vocabulary test that can be administered to different age groups but still be comparable between these groups of participants. The selection of items is based on the childLex corpus, a linguistic corpus for children in German (Schroeder, Würzner, Heister, Geyken & Kliegl, 2015a, see section 5.1). The validation of the test is performed via an Item-Response-Theory approach within the validity framework of Messick (1995). It contains several aspects of validity such as content, substantive, structural and external validity. Within this dissertation, Study 1 thus introduces the methodological basis for the following studies, which use of the vocabulary test to assess total vocabulary size and structure.

STUDY 2. The basic principle of sampling-based approaches is to project results from a vocabulary test to a whole dictionary or word list (e.g. Anglin, Miller & Wakefield, 1993; Seashore & Eckerson, 1940). An enhancement of this procedure is presented in Study 2 by projecting the results of the vocabulary test from Study 1 to the whole childLex corpus. The basic idea is to sample a group of virtual participants with different lexicon sizes from the corpus and let these virtual participants “take” the yes/no vocabulary test by checking whether the test items were included in their lexicons or not. This allows determining the relation between test performance and lexicon size, which then can be applied to a sample of real participants. For these real participants, age-specific lexicon sizes are determined so that the course of lexical development from primary school onwards can be obtained. The approach also allows the investigation of lexical content in terms of different parts of speech and morphological complexity of words. The study thus aims not only to provide a useful method to assess total mental lexicon size but also to generate actual and valid estimates of lexicon size development from primary school to adulthood.

STUDY 3. Study 3 examines the development of orthographic networks within the mental lexicon for different age groups under the use of graph theory. Based on the results on lexicon size from Study 2, the content of the average language learner at different points of time is simulated by sampling according lexicon sizes from the childLex corpus. This leads to a virtual sample of participants from different age groups for which networks based on orthographic neighborhoods can be constructed. For these networks, characteristics such as average and maximum path length as well as clustering coefficients are determined and analyzed in terms of developmental changes and the presence of small-world characteristics. This allows comparing results to studies from other language domains, to describe changes in network characteristics and is thus aimed to shed light on the courses of the development of orthographic lexical structure.

Taken together, the three studies are aimed to add to the limited research on the assessment and the development of the mental lexicon during primary school and onwards in German. On the one hand, the findings are aimed to extend the methodological approaches used to study the size and content of an individual’s mental lexicon. On the other hand, results are intended to resolve the question of how the mental lexicon emerges during the critical period of primary school and onwards.

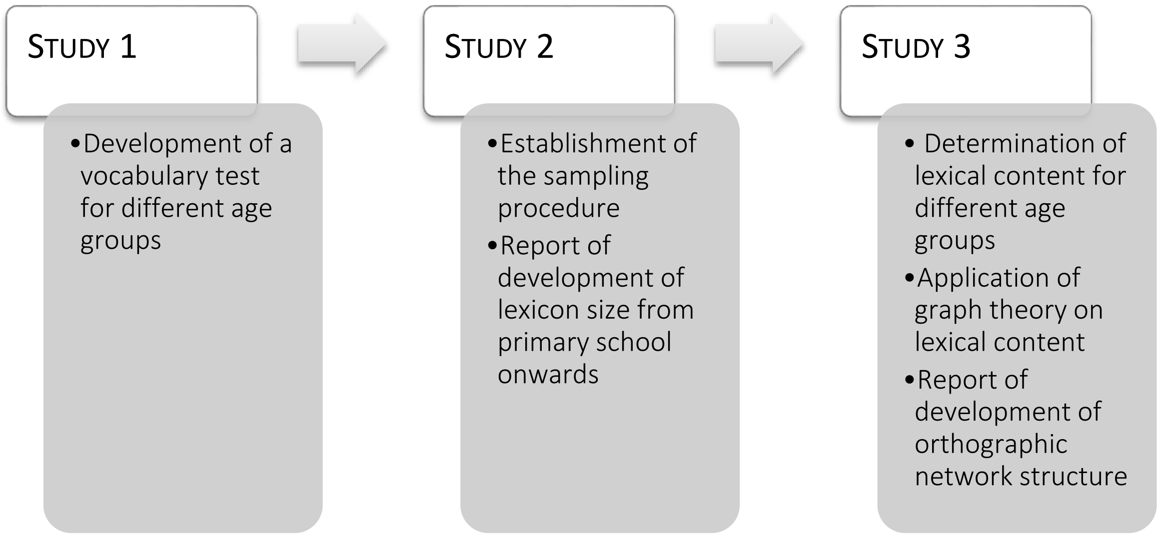

FIGURE 4.1: Consecutive aims of the three studies.

5 GENERAL METHODS

All three studies share the same general methods in that they are all based on the childLex corpus and studies 2 and 3 rely on the idea of a sampling approach using this corpus. Because they are so central to all studies, both methods will be presented here although their specific description is also part of the single studies (see section II).

5.1 THE CHILDLEX CORPUS

The childLex corpus (Schroeder et al., 2015a) is a written language corpus for children in German. It contains linguistic data from 500 German children's books with an intended reading age from 6 to 12 years. The books were selected via children and teacher questionnaires as well as library lending and online selling statistics to include the most popular books in the addressed age group. The corpus mostly comprises narrative texts and only a few formal or expository texts. It is intended to represent the linguistic environment which German children aged 6 to 12 encounter in their leisure time reading. Books were scanned manually, tokenization, lemmatization and parts-of-speech tagging were performed automatically using computational tools (WASTE, Jurish & Würzner, 2013; TAGH, Geyken & Hanneforth, 2006 and moot, Jurish, 2003). This resulted in about 10,000,000 tokens, 180,000 types and 117,000 lemmas. Available lexical variables include word length, word frequency and orthographic neighborhood size[4].

5.2 THE SAMPLING APPROACH

The sampling approach constitutes the central method of this dissertation and is aimed to reuse and extend the widely used dictionary approach (Nation, 1993a; Anglin, Miller & Wakefield, 1993) to estimate total individual lexicon size as well as lexical content. The idea is to project performance in a vocabulary size test on the whole corpus by using a sample of virtual participants. The aim of this section is to present the general idea whereas the detailed procedures and results are described in the particular studies (see section II).

A language corpus representative for the linguistic environment of the target group serves as the basis for the procedure, in this case the childLex corpus (section 5.1). Different lexicon sizes can be drawn from this corpus by sampling a certain number of words[5] from it. The sampling procedure needs to be frequency-sensitive since the word acquisition process is dependent on word frequency (Bonin et al., 2004)[6]. Each sampled lexicon represents a virtual participant for whom lexicon size (i.e. number of words) and lexical content (i.e. which words are included) is known. With repeating this procedure for several lexicon sizes and several times for each input size, a sample of virtual participants can be created for whom we know both, lexicon size and lexical content.

In a next step, test performance in a vocabulary test can be simulated with this virtual sample. For this purpose, the vocabulary test has to be based on the according corpus. In this dissertation, a new vocabulary test based on the childLex corpus was developed to comply with this condition (see Study 1, section II). By checking whether the vocabulary test items are included in the lexicons of the virtual participants or not, their test performance can be imitated. This allows the determination of the relation between test performance and lexicon size.

[4] More information and the corpus itself can be accessed via https://www.mpib-berlin.mpg.de/en/research/max-planck-research-groups/mprg-read/projects/childlex

[5] "Word" can refer to any entity of interest depending on the conclusions to be drawn. For example, if the number of known lemmas is supposed to be determined, the entity of lemmas is chosen.

[6] Other variables affecting language acquisition are discussed Study 2 as well as in the General Discussion.

In the final step, the same vocabulary test is administered to a sample of human participants. Applying the relation between test performance and the lexical variables from the virtual sample for the real sample allows determining their individual lexicon size per person. Knowing how many words an individual knows also enables to determine which words are most likely to be known. For this purpose, the probabilities of words to be included in the lexicon with a size *x* can be calculated by repeatedly sampling the lexicon size from the corpus and comparing lexical content. A schematic illustration of the process is displayed in Figure 5.1.

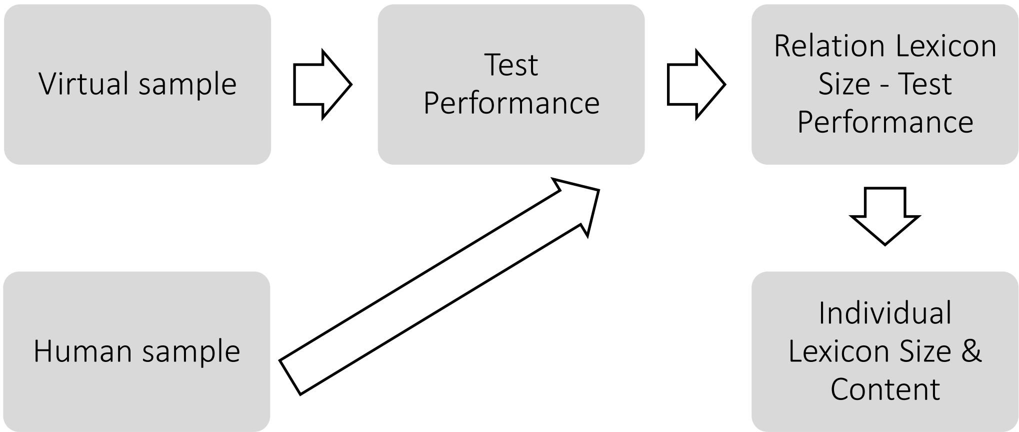

FIGURE 5.1: Schematic illustration of the sampling approach.

II STUDIES

6 STUDY 1: A YES/NO VOCABULARY TEST FOR CHILDREN

This chapter was published (in German):

TRAUTWEIN, J. & SCHROEDER, S. (2019). WOR-TE: Ein Ja/Nein-Wortschatztest für Kinder verschiedener Altersgruppen. Entwicklung und Validierung basierend auf dem Rasch Modell. *Diagnostica*, 65, 37-48. doi: 10.1026/0012-1924/a000212

6.1 ABSTRACT

In dem vorliegenden Artikel wird der Wortschatztest WOR-TE für deutsche Grundschulkinder vorgestellt. Der Test basiert auf der Ja/Nein-Methode, bei der die Teilnehmerinnen und Teilnehmer aus einer Liste von Wörtern und Pseudowörtern diejenigen ankreuzen sollen, die sie kennen. Er wurde für verschiedene Altersgruppen (1./2. Klasse, 3./4. Klasse, 5./6. Klasse) konzipiert und Item Response Theory-basiert mit dem Rasch-Modell validiert. Anhand des Validitätskonzepts nach Messick (1995) wurden verschiedene Aspekte von Konstruktvalidität untersucht: Inhaltliche Aspekte, Relevanz, Repräsentativität, Technische Qualität, substantielle Aspekte, Generalisierbarkeit und externe Aspekte. Die Ergebnisse zeigen, dass der Test ein valides Instrument zur Erfassung des orthographischen Wortschatzes im Grundschulalter darstellt und durch kleine Veränderungen, insbesondere in Bezug auf die Item-Auswahl, profitieren würde. Möglichkeiten des Einsatzes sowie Einschränkungen des Verfahrens werden diskutiert, ebenso wie die Verwendung des Validitätskonzeptes mit IRT für Ja/Nein-Test im Allgemeinen.

In this article we present a vocabulary test for German primary school children. The test is based on the yes/no-method where participants identify words they know out of a list of words and pseudowords. The test was developed for different age groups (Grade 1/2, Grade 3/4, Grade 5/6) and validated via Item Response Theory, namely the Rasch Model. Following the concept of suggested by Messick (1995), we analyzed several different aspects of construct validity: content aspects, relevance, representativity, technical quality, substantial aspects, generalizability and external aspects. Results show that the test is a valid instrument to measure the orthographic vocabulary of German primary school children but could also benefit from some minor changes concerning e.g. item selection. Possible applications and limitations of the instrument are discussed as well as the use of the validity concept and the validation via IRT for yes/no-vocabulary tests in general.

6.2 EINLEITUNG

Der Wortschatz stellt eine essenzielle Komponente der Sprachkompetenz dar und hängt eng mit der Lesefähigkeit und dem Schulerfolg zusammen (Biemiller, 2003, 2006). Er gilt daher in vielen frühen Screeningverfahren als Indikator für eine Sprachentwicklungsstörung (z. B. Elternfragebögen für die Früherkennung von Risikokindern - ELFRA; Grimm & Doil, 2006). Im späteren Spracherwerb wird der Wortschatz als Prädiktor für Lese- und Schreibfertigkeiten angesehen. So konnten Muter, Hulme, Snowling und Stevenson (2004) zeigen, dass die Wortschatzgröße zu Beginn der Grundschule das spätere Leseverständnis vorhersagte. Auch in querschnittlichen Untersuchungen wurde ein Zusammenhang zwischen Wortschatz und Leseverständnis nachgewiesen (Ricketts, Nation, & Bishop, 2007; Ouellette & Beers, 2010). Zudem wird die Lesegeschwindigkeit als Teil der Lesefähigkeit vom Wortschatz beeinflusst (z. B. Anderson, Wilson, & Fielding, 1988). Perfetti und Stafura (2014) nehmen an, dass ein besserer Wortschatz den lexikalischen Zugriff erleichtert, was wiederum das Leseverständnis begünstigt. Da die Lesefähigkeit zentral für den Schul- und daran anschließenden beruflichen Erfolg sowie die Teilhabe an der Gesellschaft ist, ist es entscheidend, Defizite und ihre Gründe früh aufzudecken, um effektive Trainingsmethoden einzuleiten (Biemiller, 2006).

Das Konstrukt *Wortschatz* ist nicht leicht zu definieren. Perfetti und Hart (2002) gehen in ihrer Hypothese der lexikalischen Qualität davon aus, dass der Wortschatz eine phonologische, eine orthographische und eine semantische Komponente umfasst. Die phonologische Komponente beinhaltet Wissen über die Aussprache, die orthographische über die Schreibung und die semantische über die Bedeutung eines Wortes. Die verschiedenen Wissensebenen können für ein Wort unterschiedlich stark ausgeprägt sein. Beim Lesen muss demnach zunächst die orthographische Form des Wortes abgerufen werden, für das Leseverständnis zudem das semantische Wissen und zum lauten Lesen die phonologische Komponente. So wird es auch häufig in Modellen zur visuellen Worterkennung angenommen (z. B. das Dual Route Model; Coltheart et al., 2001). Demnach ist insbesondere der orthographische Wortschatz für die Lesefähigkeit entscheidend.

Eine Möglichkeit zur Messung des orthographischen Wortschatzes ist die Ja/Nein-Methode von Anderson und Freebody (1983). Teilnehmende identifizieren alle ihnen bekannten Wörter innerhalb einer Wortliste. Um Raten zu vermeiden, enthält die Liste auch Pseudowörter. Anderson und Freebody (1983) fanden bei Fünftklässlerinnen und Fünftklässlern hohe Korrelationen mit mündlichen Definitionsaufgaben und Multiple-Choice-Wortschatztests und den Ergebnissen aus dem Ja/Nein-Test (r = .84 für Multiple Choice, r > .85 für Definitionsaufgaben). Die Autoren haben demnach ein valides Instrument zur Erfassung des orthographischen Wortschatzes entwickelt und zudem Zusammenhänge zu semantischem Wissen über Wörter gefunden. Ähnliche Ergebnisse erzielten auch andere Studien (z. B. Mochida & Harrington, 2006: r = .85 für Multiple Choice; Pellicer-Sánchez & Schmitt, 2012: r = .89 für mündliche Definitionen). Obwohl Anderson und Freebody (1983) den Wert von Ja/Nein-Tests auch für Kinder demonstriert haben, wird er bislang vornehmlich für die Messung des Wortschatzes von Erwachsenen in einer Zweitsprache verwendet (z. B. Eyckmans, 2004; Huibregtse, Admiraal, & Merea, 2002; Merea & Buxton, 1987; Mochida & Harrington, 2006; Lemhöfer & Broersma, 2012). Das Testformat unterscheidet sich von anderen bereits bestehenden Verfahren für Kinder im Deutschen, welche vielmals primär auf die semantische Ebene des Wortschatzes abzielen (z. B. Peabody Picture Vocabulary Test - PPVT-4; Lenhard, Lenhard, Segerer & Suggate, 2015; Patholinguistische Diagnostik bei Sprachentwicklungsstörungen - PDSS, Kauschke & Siegmüller, 2009). Hierbei wird häufig mit dem Benennen oder Zeigen von Bildern nach mündlicher Vorgabe gearbeitet. Zudem sind existierende Verfahren meist nur in Einzelerhebungen durchführbar und zielen oftmals auf die Diagnose semantischer Defizite ab (z. B. Wortschatz- und Wortfindungstest - WWT; Glück, 2007). Die meisten dieser bereits existierenden Instrumente sind außerdem für Vorschulkinder konzipiert worden. Ein weiteres Testformat für Schulkinder und Erwachsene ist die Auswahl von Synonymen nach schriftlicher Vorgabe (z. B. Grundintelligenztest - CFT-20; Weiß, 1998). Allerdings ist die Aufgabe sehr stark von den Distraktor-Items abhängig. Das Wissen über die Bedeutung der Distraktor-Items kann demnach die Lösung der Aufgabe beeinflussen (Anderson & Freebody, 1983). Mit steigendem Alter wird dies aber schwieriger, da der Wortschatz substanziell wächst, insbesondere im Schulalter (Segbers & Schroeder, 2017; Anglin, Miller & Wakefield, 1993). Geeignete Verfahren für Schulkinder, die für verschiedene Altersgruppen geeignet sind, gibt es im Deutschen aktuell nicht.

Da die Ergebnisse zum Ja/Nein-Testformat von Anderson und Freebody (1983) ermutigend sind und die Notwendigkeit besteht, den orthographischen Wortschatz im Grundschulalter zu messen, ist es vielversprechend, den Test für das Grundschulalter zu adaptieren. Tatsächlich birgt das Verfahren Vorteile für die praktische Anwendung: Wegen des geringen kognitiven Aufwands kann eine große Anzahl an Items in kurzer Zeit dargeboten werden. Zudem kann der Test in Gruppentestungen durchgeführt werden.

Bisher wurden die Verfahren lediglich über die Korrelationen zu Definitionsaufgaben oder Multiple-Choice-Verfahren validiert (z. B. Anderson & Freebody, 1983; Mochida & Harrington, 2006; Pellicer-Sánchez & Schmitt, 2012). Messick (1995) zufolge bezieht sich diese Art von Validierung auf die konvergente Validität, welche durch die Korrelation des Testscores mit externen Variablen, die dasselbe oder assoziierte Konstrukte messen, definiert ist. Er nennt allerdings noch weitere Aspekte, die zur Validierung eines Tests herangezogen werden sollten. Messick (1995) beschreibt inhaltliche Aspekte, die die Relevanz, die Repräsentativität und die technische Qualität des Testinhalts umfassen. Sie zielen damit darauf ab, zu überprüfen, inwiefern die Inhalte eines Tests zur Messung der entsprechenden Fähigkeit angemessen sind. Er beschreibt auch substanzielle Aspekte, die sich auf die Einbettung der Testergebnisse in ein nomologisches Netzwerk beziehen. Damit ist die Passung der Testergebnisse zu vorherigen Annahmen in Bezug auf die gemessene Fähigkeit gemeint. Des Weiteren nennt er strukturelle Aspekte, die sich auf Annahmen zur Struktur des zu messenden Konstrukts beziehen, Generalisierbarkeit, die die Adaption des Testformats für andere Items oder andere Teilnehmende meint, und externe Aspekte, die die Korrelation mit konvergenten und divergenten Variablen beinhaltet. Ein Ansatz zur Anwendung dieses Konzepts der Validierung für Wortschatztests wurde von Beglar (2010) sowie McLean, Kramer und Beglar(2015) vorgestellt. Sie untersuchten die verschiedenen Aspekte von Validität unter der Verwendung der Item Response Theory (IRT) anhand des Rasch-Modells. Shillaw (1996) zeigte zudem bereits, dass das Rasch-Modell für die Auswertung von Ja/Nein-Wortschatztests geeignet ist.

In der vorliegenden Studie wird ein Ja/Nein-Wortschatztest WOR-TE für deutsche Grundschulkinder verschiedener Altersgruppen vorgestellt. Beruhend auf dem Konzept der Validität von Messick (1995) sowie dem IRT-basierten Ansatz von Beglar (2010) und McLean et al. (2015) soll dabei gezeigt werden, dass es sich bei dem Test um ein valides Instrument zur Erfassung des orthographischen Wortschatzes von Grundschulkindern handelt. Dazu werden die Testergebnisse mithilfe des Rasch-Modells skaliert und auf die verschiedenen Aspekte der Validität nach Messick (1995) untersucht.

6.3 TESTENTWICKLUNG

Um eine breite Altersspanne von Grundschulkindern abzudecken, wurden drei Testversionen des WOR-TE (Wortschatz-Test) für verschiedene Altersgruppen (1./2. Klasse, 3./4.

Klasse, 5./6. Klasse) entwickelt. Da die Wortfrequenz die Itemschwierigkeit in Wortschatztests hauptsächlich bestimmt (z. B. Beglar, 2010), wurde die mittlere Lemmafrequenz[7] der Items (childLex Kinderkorpus; Schroeder, Würzner, Heister, Geyken & Kliegl, 2015b) in den verschiedenen Testversionen systematisch manipuliert (Tabelle 6.1). Die Auswahl passender Frequenzen für jede Altersgruppe basierte auf Ergebnissen von vorherigen Studien (u. a. Developmental Lexicon Study; Schröter & Schroeder, 2017). Die Materialien beinhalteten Nomen, Verben und Adjektive.

TABELLE 6.1: Verteilung der Frequenz und Itemschwierigkeit auf die drei Testversionen.

	Log Lemma Frequenz		Itemschwierigkeit
Testversion	*M* (*SD*)	Bereich	*M* (*SD*)
Klasse 1/2	1.5 (0.4)	2.7–1.0	-0.62 (0.88)
Klasse 3/4	0.6 (0.1)	0.9–0.4	1.05 (1.38)
Klasse 5/6	0.0 (0.1)	0.4–-0.2	2.56 (1.46)

Jede Testversion umfasste 100 Wörter. Um einen Vergleich der drei Testversionen zu ermöglichen, waren 20 Wörter in allen Testversionen identisch. Diese 20 Anker-Items wurden aus dem Frequenzbereich von allen drei Testversionen ausgewählt. Zusätzlich teilten sich aufeinanderfolgende Testversionen jeweils zehn Items. Das bedeutet, die Testversion für die 1. und 2. Klasse umfasste 70 unique Items, 20 Anker-Items, die in allen Testversionen enthalten waren, und 10 Anker-Items, die ebenfalls in der Version für die 3. und 4. Klasse enthalten waren. Die Testversion für die 3. und 4. Klasse enthielt 60 unique Items, die 20 allgemeinen Anker-Items, 10 geteilte Anker-Items mit der 1. und 2. Klasse und 10 geteilte Anker-Items mit der Version für die 5. und 6. Klasse. Die Testversion für die 5. und 6. Klasse enthielt 70 unique Items, die 20 allgemeinen Anker-Items sowie die 10 geteilten Items aus der Version für die 3. und 4. Klasse. Der Test umfasst damit insgesamt 240 Items (siehe Appendix A.1).

Um das Raten zu minimieren, wurden zu jeder Testversion 24 Pseudowörter hinzugefügt. Diese wurden durch das Austauschen von mindestens einem Buchstaben in einem realen Wort bzw. die Aneinanderreihung von Morphemen konstruiert und waren in jeder Testversion

[7] Als Lemma wird die zitierfähige Grundform eines Wortes bezeichnet.

gleich. Für jede Altersgruppe wurden zwei Pseudoparallel-Versionen A und B mit randomisierter Item-Reihenfolge erstellt.

Die wortwörtliche Instruktion für die teilnehmenden Kinder lautete: „Im Folgenden seht ihr eine Liste von Wörtern. Ihr sollt die Wörter markieren, die ihr kennt. Dabei dürft ihr nicht raten, denn die Liste enthält auch Wörter, die es gar nicht gibt. Wenn ihr ratet, merken wir das sofort. Kreuzt nur die Wörter an, die ihr wirklich kennt.“ Drei Beispielitems (2 Wörter und 1 Pseudowort) wurden zur Veranschaulichung der Aufgabe besprochen. Abhängig von der Altersgruppe dauerte die Testdurchführung 5 bis 15 Minuten.

6.4 METHODE

6.4.1 STICHPROBE

Insgesamt nahmen *N* = 422 Kinder (Klassen 1–6) von fünf Berliner Grundschulen an der Studie teil. Vierundzwanzig Kinder (6 %) füllten den Wortschatztest unvollständig aus und wurden daher aus den weiteren Analysen ausgeschlossen, sodass die Daten von *N* = 398 Kindern (198 weiblich, 197 männlich, 3 ohne Geschlechterangabe) verwendet werden konnten. Ein Großteil der Kinder (233, 59 %) gab Deutsch als ihre einzige Muttersprache an, während 150 Kinder (38 %) angaben, mindestens eine weitere Muttersprache gelernt zu haben. Eigenschaften der Stichprobe sind in Tabelle 6.2 enthalten.

6.4.2 INSTRUMENTE

Der *Wortschatz* wurde mit dem Subtest *Sprachverständnis* (Kognitiver Fähigkeitstest - KFT1-3, Heller & Geisler, 1983) bzw. „Wortschatz“ (Kognitiver Fähigkeitstest - KFT 4-12+ R, Heller & Perleth, 2000) untersucht. In der Version für die 1. bis 3. Klassenstufe wählen die Kinder nach auditiver Vorgabe ein passendes Bild aus fünf Bildern aus. Für die 4. bis 12. Klasse handelt es sich um ein Multiple-Choice-Verfahren, wobei zu einem fettgedruckten Wort das passende Synonym gesucht werden muss. Die Rohwerte wurden in jahrgangsspezifische T-Werte überführt. Die Reliabilität wurde mit Cronbachs α von .57 (1.-3. Klasse) bzw. .71 (4.-6. Klasse) bestimmt. Für die früheren Klassen ist sie damit zu gering, in den höheren Altersstufen akzeptabel.

Die *Lesegeschwindigkeit* wurde mit dem SLS 1-4 (Mayringer & Wimmer, 2003) bzw. 5-8 (Auer, Gruber, Mayringer & Wimmer, 2005) erfasst. Dabei sollen die Kinder innerhalb von drei Minuten für möglichst viele Sätze angeben, ob sie wahr oder falsch sind. Der Testscore ergibt sich aus den korrekt markierten Sätzen. Es werden alterskorrigierte Normwerte verwendet. Cronbachs α zur Überprüfung der Reliabilität lag bei .96 und ist somit sehr gut.

Orthographische Fähigkeiten wurden mit der Hamburger Schreibprobe 1-9 (May, 2002) ermittelt. Dabei werden Wörter und Sätze diktiert und anschließend die richtigen Grapheme gezählt. Die Ergebnisse werden als alterskorrigierte T-Werte berichtet. Zur Berechnung der Reliabilität wurde die Anzahl richtiger Grapheme pro Wort verwendet. Da verschiedene Wörter pro Altersgruppe eingesetzt werden, wurde Crohnbachs α separat berechnet. Der Mittelwert war mit M = .81 sehr zufriedenstellend. Da die orthographischen Fähigkeiten in der 1. Klasse noch sehr stark schwanken, fand hier keine Erfassung statt.

Die *nonverbale Intelligenz* der Teilnehmenden wurde mit dem Matrizen-Subtest des CFT 1 (Cattell, Weiß & Osterland, 1997) bzw. CFT 20-R (Weiß, 2006) erhoben. Die Aufgaben bestehen jeweils aus einem Muster, welches mithilfe einer Auswahl von fünf Möglichkeiten vervollständigt werden muss. Testteilnehmerinnen und Testteilnehmer haben dafür sechs (CFT 1 für die Erstklässler) bzw. drei Minuten (CFT 20-R, ab Klasse 2) Zeit. Da lediglich ein Subtest durchgeführt wurde, können nur die Rohwerte (Anzahl richtiger Antworten) für die Analyse verwendet werden. Für die Überprüfung der Reliabilität wurde ein zufriedenstellender Wert von Cronbach‘s α mit .81 (1. Klasse) bzw. .68 (Klasse 2-6) berechnet.

6.4.3 Prozedur

Das schriftliche Einverständnis der Eltern war notwendig für die Studienteilnahme. Alle Aufgaben wurden während der Schulzeit innerhalb von zwei Schulstunden (à 45 Minuten) im Klassenverband durchgeführt. Zusätzlich wurden demographische Daten (Alter, Geschlecht und Muttersprache) mit einem Fragebogen ermittelt. Mithilfe von Identifikationsnummern wurden die Daten anonymisiert. Für die Teilnahme erhielten die Kinder ein kleines Dankeschön.

6.4.4 Analysen

Zur Analyse wurde eine Item-Response-Analyse unter Einsatz des Rasch-Modells (Embretson & Reise, 2000) durchgeführt. Um Unterschiede zwischen den Altersgruppen zu berücksichtigen, wurde ein Multiple-Group-Modell gewählt, bei dem die verschiedenen Altersgruppen als separate Gruppen behandelt wurden (Bock & Zimowski, 1997). Die 20 Ankeritems ermöglichten dabei eine Schätzung der Itemparameter von allen Testversionen auf einer gemeinsamen Skala (Embretson & Reise, 2000). Die Modelle wurden mit dem TAM Paket für R

(Kiefer, Robitzsch & Wu, 2016) geschätzt, welches Margninal Maximum Likelihood (MML) für die Parameterschätzung verwendet (Mislevy & Stocking, 1989). Für die Modellschätzung wurde *vertical linking* und *concurrent calibration* genutzt (für einen Überblick über Skalierungsmethoden siehe Kolen & Brennan, 2004). Die Modelle wurden identifiziert, indem der erste Itemparameter auf 0 fixiert wurde. Personenparameter, die das latente Personenmerkmal des Wortschatzes repräsentieren, wurden ebenfalls mit MML geschätzt. Aufgrund fehlender korrekter Antworten musste ein Item (*äsen*, Version 5./6. Klasse) zuvor ausgeschlossen werden. Auf die Prüfung der Modellpassung wird im Ergebnisteil eingegangen.

TABELLE 6.2: Stichprobenbeschreibung und mittlere Hit- und False-Alarm-Raten pro Altersgruppe.

			Geschlecht			Muttersprache				
Klasse	*N*	*M* Alter (*SD*)	männl.	weibl.	NA	D	ND	NA	*M* Hit Rate (*SD*)	*M* False Alarm Rate (*SD*)
1	37	6.6 (0.5)	13	23	1	18	16	3	.56 (.17)	.16 (.10)
2	49	7.3 (0.7)	24	25	0	33	14	2	.53 (.20)	.11 (.11)
3	75	8.0 (0.6)	38	35	2	49	23	3	.38 (.18)	.05 (.07)
4	107	9.0 (0.6)	65	42	0	67	37	3	.52 (.17)	.05 (.07)
5	62	10.0 (0.6)	22	40	0	35	26	1	.30 (.14)	.05 (.05)
6	68	11.2 (0.6)	35	33	0	31	34	3	.44 (.16)	.05 (.07)
total	398	8.9 (1.6)	197	198	3	233	150	15	.45 (.19)	.06 (.08)

Anmerkungen: NA = Keine Angabe, D = Deutsch als einzige Muttersprache, ND = weitere Muttersprachen neben Deutsch.

6.5 ERGEBNISSE

Die Raten der Hits und False Alarms sind in Tabelle 6.2 dargestellt. Im Folgenden werden in Bezug auf den Ja/Nein-Wortschatztest fünf verschiedene Aspekte von Konstruktvalidität nach Messick (1995) in Betracht gezogen. Im Anschluss werden die Passung des Rasch-Modells und die Validität des Tests bewertet und die Nützlichkeit der IRT-basierten Validierung diskutiert.

6.5.1 Inhaltliche Aspekte

Zunächst wurde überprüft, inwieweit der Inhalt des Ja/Nein-Wortschatztests angemessen ist, um den Wortschatz der Testteilnehmenden zu messen.

Inhaltliche Relevanz

Messick (1995) definiert die inhaltliche Relevanz als eine Auswahl von Aufgaben, die relevant für die Messung des Konstruktes sind. Für den vorliegenden Ja/Nein-Wortschatztest ist dies dadurch gegeben, dass die Wörter aus einem spezifischen Korpus für Kindersprache ausgewählt wurden. Diese Wörter werden daher mit hoher Wahrscheinlichkeit von den Kindern im Alltag rezipiert. Zur Anpassung an die jeweiligen Altersgruppen wurde zudem die Frequenz der Wörter für die verschiedenen Testversionen systematisch manipuliert (Tabelle 6.1). Zusätzlich dienten vorherige Studien (u.a. DeveL; Schröter & Schroeder, 2017) dazu, Wörter auszuwählen, die eine ausreichende Variabilität in den Erkennensraten bei der Zielgruppe hatten. Die inhaltliche Relevanz ist somit durch die Testkonstruktion gegeben.

Repräsentativität

Messick (1995) betont, dass ein Test alle wichtigen Teile des Konstrukts enthalten muss, um repräsentativ zu sein. Dies beinhaltet eine ausreichende Anzahl von Items, eine adäquate Streuung der Itemschwierigkeit und das Fehlen von Lücken in der Item-Hierarchie (Beglar, 2010). Abbildung 6.1 zeigt eine Item-Personen-Zuordnung für die Itemschwierigkeit und den Personenparameter aus den Testergebnissen. Links ist die Verteilung der Itemschwierigkeit dargestellt. Die rechte Seite repräsentiert die Verteilung der Personenparameter. Bezüglich der Anzahl von Items empfiehlt Beglar (2010) zehn Items pro Schwierigkeitsstufe. In Abbildung 6.1 ist zu sehen, dass dieses Kriterium für die meisten Schwierigkeitsstufen erfüllt wurde, lediglich an den Rändern der Verteilung ist die Anzahl etwas geringer. Der Test würde also profitieren, wenn man besonders einfache und besonders schwere Items hinzufügt. Die Streuung der Itemschwierigkeit erscheint ausreichend. Sie rangiert zwischen -2.95 und 5.37, wobei 97 % der Personenparameter zwischen -2.5 und 5 lag. Es können keine Lücken in der Itemhierarchie beobachtet werden. Allerdings können vier Items aus der Version für die 5./6. Klasse als zu schwer angesehen werden (*brüsk*: 5.37; *süffisant*: 5.05; *Häme*: 4.59; *schartig*: 4.59). In einer neuen Testversion sollten diese Items ausgelassen bzw. ersetzt werden. Die Verteilung der Personenparameter zeigt keinen Boden- oder Deckeneffekt und der mittlere Standardfehler *SE* = .03 (*SD* = .005) lässt auf eine präzise Messung der Personenfähigkeit schließen.

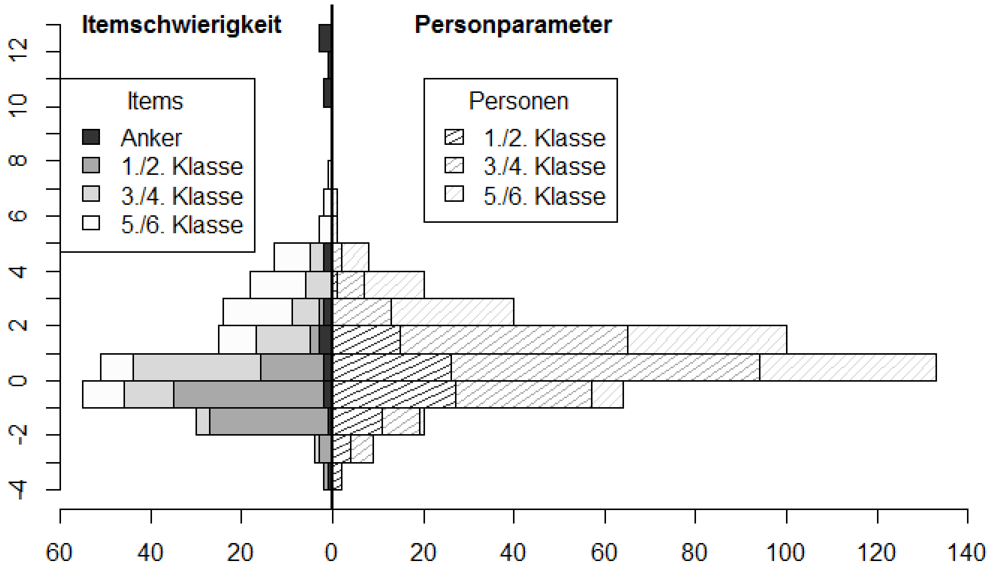

ABBILDUNG 6.1: Verteilung von Personen- und Itemparametern für den WOR-TE.

Um zusätzlich zu überprüfen, ob die Items repräsentativ für den gesamten Korpus sind, wurden die drei Maße Lemmafrequenz, Anzahl der orthographischen Nachbarn und Wortlänge der Items mit denjenigen der Wörter aus dem gesamten Korpus verglichen. Dafür wurden diejenigen Wörter aus dem gesamten Korpus herausgenommen, die nur einmal vorkamen unter der Annahme, dass sie nicht ausreichend repräsentativ für den Wortschatz eines Sprechers der Sprache sind. Der Vergleich der Maße ergab keinen Unterschied in der mittleren Frequenz der Items und der Wörter aus dem Gesamtkorpus, t = -0.10, p = .92. Die Wortlänge unterschied sich dahingehend, dass im Gesamtkorpus insgesamt mehr längere Wörter enthalten waren, t = -19.80, p < .001. Dies lässt sich dadurch erklären, dass im Gesamtkorpus auch Komposita enthalten sind, die für die Auswahl der Items nicht beachtet wurden. Bedingt durch die Wortlänge, die mit der Anzahl der orthographischen Nachbarn zusammenhängt, ergab sich auch bezüglich dieses Merkmals ein signifikanter Unterschied, t = 3.71, p < .001. Schränkt man allerdings die Länge der Wörter im Gesamtkorpus entsprechend der Länge der Items ein, verschwindet diese Differenz, t = 1.10, p = 0.27. Insgesamt kann man davon ausgehen, dass die ausgewählten Items repräsentativ für den Korpus und damit für die verwendete Sprache in deutscher Kinderliteratur sind.

Technische Qualität

Technische Qualität meint die Passung der Items zum verwendeten Modell (Beglar, 2010). Um diese zu messen, wurde der Itemfit zum Rasch-Modell mit dem Maß der *Rasch infit mean-square statistic (MNSQ)*[8] bestimmt. Angelehnt an McNamara (1996) wurde ein Kriterium von ±2 Standardabweichungen vom Mittelwert der Infit Statistik (= 1) gewählt, um eine fehlende Passung zu identifizieren. Da die Standardabweichung 0.12 betrug, wird ein Infit-Wert zwischen 0.76 und 1.24 als akzeptabel für die Item-Passung behandelt. Kein Infit-Wert kleiner als 0.76 kann beobachtet werden, jedoch wird für zehn Items (4 %) ein zu hoher Wert gemessen (1./2. Klasse: *Planet*, Infit MNSQ = 1.38; *Backe*, Infit MNSQ = 1.31; *Statue*, Infit MNSQ = 1.31; *starren*, Infit MNSQ = 1.27; *passieren*, Infit MNSQ = 1.25; 5./6. Klasse: *sengen*, Infit MNSQ = 1.46; *Spind*, Infit MNSQ = 1.30; Ankeritems: *Tresse*, Infit MNSQ = 1.53; *wähnen*, Infit MNSQ = 1.34; *konstatieren*, Infit MNSQ = 1.30). Nach Ausschluss der unpassenden Items wurde ein neues Modell geschätzt und analog erneut der Itemfit überprüft. Dieser Prozess wurde so häufig wiederholt, bis keine unpassenden Items mehr vorhanden waren. Dabei wurden weitere 16 Items ausgeschlossen (1./2. Klasse: *Schüssel, Museum, knirschen, Strahl, ignorieren, reagieren, Stapel, Gegend, grell;* 3./4. Klasse: *artig*; 5./6. Klasse: *Galosche, Fanfare, schwelen, graziös, imitieren*; Ankeritems: *Reuse*).

Zusammenfassend lässt sich bezüglich der inhaltlichen Aspekte festhalten, dass die Items relevant und repräsentativ für das zu messende Konstrukt des orthographischen Wortschatzes sind, bis auf vier zu schwere Items. Durch die Analyse der technischen Qualität wurden zehn nicht passende Items identifiziert. Kleine Veränderungen an der Zusammensetzung der Items könnten zur Verbesserung des Tests beitragen. In einem nächsten Schritt wurden die unpassenden Items entfernt und ein neues Modell mit den 209 verbleibenden Items wurde geschätzt. Die neu geschätzten Parameter korrelierten hoch mit denen aus dem vorherigen Modell (Items: r = .99; Personen: r = .99). Die generellen Ergebnisse wurden durch die

[8] Mean Squares geben die chi-Quadrat Statistik geteilt durch deren Freiheitsgerade an und zeigen inwiefern die tatsächlichen Antworten mit denen des Modells übereinstimmen.

neue Modellschätzung also nicht verändert. Für die folgenden Analyseschritte wurde das Modell mit der reduzierten Itemanzahl verwendet. Um die Passung des Modells zusätzlich zu überprüfen, wurde zudem mithilfe der Q3-Statistik (Yen, 1984) evaluiert, ob die Antworten auf die Items unabhängig voneinander sind. Dazu wird die Residualkorrelation für alle Itempaare berechnet. Sie sollte bei lokaler Unabhängigkeit ungefähr 0 betragen. Im vorliegenden Modell ist dies der Fall, der Mittelwert der Q3-Statistik lag bei M = -0.01, SD = 0.09. Lediglich 3 % der gesamten Residualkorrelationen weichen mehr als zwei Standardabweichungen vom Mittelwert ab und können damit als Ausreißer angesehen werden. Die lokale Unabhängigkeit der Antworten ist damit gegeben, was zusätzliche Evidenz für die Passung des Rasch-Modells auf die Daten darstellt.

6.5.2 SUBSTANZIELLE ASPEKTE

Der substanzielle Aspekt von Validität betrifft die Passung der Testergebnisse zu vorherigen Theorien in Bezug auf Prozesse, die die Testleistung beeinflussen (Messick, 1995). Es ist bereits bekannt, dass Wortfrequenz die Leistung in Aufgaben bei Wortschatztest beeinflusst. Um dies für die vorliegenden Daten zu überprüfen, wurde die Itemschwierigkeit mit der logarithmierten Lemmafrequenz der Items (childLex, Schroeder et al., 2015a) korreliert. Mit r = -.74 kann die Korrelation als sehr hoch bezeichnet werden, was zeigt, dass hochfrequente Wörter einfacher zu erkennen sind.

In Bezug auf die Testergebnisse der Kinder wurden Altersgruppe, Geschlecht und Muttersprache als relevante Faktoren, die die Testleistung beeinflussen können, betrachtet. Um den Effekt dieser drei Variablen zu messen, wurde jeweils eine einfaktorielle ANOVA mit dem Personenparameter als abhängige Variable und Altersgruppe, Geschlecht, Muttersprache als unabhängige Variablen gerechnet. Dazu wurden die Personen, die keine Angaben zu Geschlecht oder Muttersprache gemacht haben, ausgeschlossen (N = 11). Die Modelle zu Muttersprache und Geschlecht enthielten zusätzlich das Alter (zentriert an der jeweiligen Altersgruppe) als Kontrollvariable. Die Ergebnisse zeigen einen signifikanten Effekt der Altersgruppe, $F(2,378) = 57.77$, $p < .001$, $\eta^2 = .27$. Post-hoc Analysen zeigen einen signifikanten Unterschied zwischen allen Altersgruppen, alle $p < .001$. Dies entspricht vorherigen Studien, da der Wortschatz mit dem Alter ansteigt (Segbers & Schroeder, 2017). Es gibt keinen Effekt des Geschlechts auf den orthographischen Wortschatz, $F(1, 378) = 2.3$, $p = .103$. Dies passt zu Ergebnissen aus vorherigen Studien, die keinen Unterschied im Wortschatz zwischen Jungen und

Mädchen im Grundschulalter beobachtet haben (z. B. Anglin, Miller & Wakefield, 1993). Der Effekt der Muttersprache ist signifikant, $F(1, 378) = 6.15$, $p = .013$, $\eta^2 = .02$, und zeigt, dass monolinguale Kinder ein signifikant höheres Testergebnis hatten als bi- und multilinguale Kinder. Dies steht im Einklang mit früheren Ergebnissen zu besserer Wortschatzfähigkeit von monolingualen im Gegensatz zu bilingualen Personen (z. B. Bialystok, Luk, Peets & Yang, 2010).

Als ein weiterer Indikator für substanzielle Validität wurde das Rateverhalten betrachtet. Das Ja/Nein-Testformat beinhaltet Pseudowörter, um Rateverhalten zu minimieren. Für jedes Kind wurde der Anteil falsch ausgewählter Pseudowörter berechnet, um den Zusammenhang zwischen dem Rateverhalten und dem Testwert zu ermitteln. Der durchschnittliche Anteil angekreuzter Pseudwörter lag bei $M = .07$ ($SD = .09$). Für die einzelnen Klassenstufen ist sie in Tabelle 2 dargestellt. Sie ist damit in allen Klassen sehr niedrig, lediglich in der 1. und 2. Klasse war sie ein wenig erhöht. Allgemein raten Kinder bei der Durchführung nicht, was die allgemeine Konstruktvalidität des Instruments unterstützt. Zudem korreliert die False-Alarm-Rate nur mit $r = -.15$, $t = -3.1$, $p = .002$, mit dem Personenparameter aus dem Modell. Die Testleistung ist also weitgehend unabhängig von dem Antwortverhalten bei den Pseudowörtern. Der Einbezug eines Rateparameters in das Modell erscheint damit nicht indiziert. Im Diskussionsteil wird dieser Punkt noch einmal aufgegriffen.

Zusammenfassend ist festzuhalten, dass die Testergebnisse zu substanziellen Theorien auf der Item-Ebene bezüglich Korrelationen mit Wortfrequenz und auf der Personen-Ebene bezüglich des Einflusses von Alter, Geschlecht und Muttersprache passen. Zudem wird das Rateverhalten durch den Einbezug von Pseudowörtern erfolgreich minimiert.

6.5.3 Strukturelle Aspekte

Laut Messick (1995) ist es für die Testvalidität entscheidend, dass der Inhalt des Tests ein zugrundeliegendes Konzept misst. Im Fall des Ja/Nein-Tests handelt es sich hierbei um den orthographischen Wortschatz, der ein Konstrukt bzw. eine Dimension darstellt. Um zu überprüfen, ob der Test tatsächlich nur diese eine Dimension misst, wurde das Modell auf Eindimensionalität getestet. Dazu wurden zwei Modelle mit verschiedenen Dimensionen geschätzt und jeweils mit dem eindimensionalen anhand des Log Likelihoods verglichen. Im ersten Vergleichsmodell wurden die Dimensionen durch die drei Testversionen für verschiedene Altersgruppen definiert, sodass spezifische Items einer Testversion auf eine Dimension abgebildet wurden. Die Analyse ergab keinen signifikanten Unterschied zwischen den beiden Modellen,

p = 1. In einem zweiten Vergleichsmodell wurden die Dimensionen nach Wortarten (Nomen, Verben, Adjektive) definiert, sodass jede Wortart eine Dimension darstellte. Auch hier zeigte sich kein signifikanter Unterschied zwischen den Modellen, p = 1.

Daraus lässt sich schlussfolgern, dass die Hinzunahme weiterer Dimensionen das Modell nicht verbessert, was die Annahme von Eindimensionalität stützt. Somit stehen die Ergebnisse in Einklang mit der Annahme über die Struktur des zugrundeliegenden Konstrukts und erfüllen damit dieses Kriterium für Validität nach Messick (1995).

6.5.4 Generalisierbarkeit

Die Generalisierbarkeit eines Tests lässt sich sowohl auf Item-Ebene als auch auf Personen-Ebene bestimmen und stellt ebenfalls einen Aspekt von Konstruktvalidität nach Messick (1995) dar. Auf der Item-Ebene wird betrachtet, inwiefern die Testergebnisse auf andere Items, die das gleiche Konstrukt messen, generalisiert werden können. Auf der Personen-Ebene wird betrachtet, inwiefern die Testergebnisse auf andere Populationen generalisiert werden können. Eine Möglichkeit, Generalisierbarkeit zu messen, ist die Kreuzvalidierung von Ergebnissen mithilfe verschiedener Teilungskriterien. Auf der Item-Ebene wurden die 204 Items in zwei Gruppen geteilt und je ein neues Modell pro Gruppe geschätzt, wie bereits zur Prüfung der Modellpassung angegeben. Anschließend wurden die Personenparameter beider Modelle verglichen. Mit einer Korrelation von r = .94 zeigt sich ein starker Zusammenhang, was darauf hinweist, dass die Personenparameter auch mit verschiedenen Item-Gruppen hergestellt werden können. Auf der Personen-Ebene wurden die teilnehmenden Kinder in zwei Gruppen geteilt und jeweils ein neues Modell für jede Gruppe berechnet. Anschließend wurde die Korrelation der Itemparameter beider Modelle berechnet. Mit r = .97 kann diese als sehr hoch bewertet werden. Mit verschiedenen Stichproben werden demnach sehr ähnliche Itemparameter gemessen.

Die Testergebnisse sind demnach durchaus generalisierbar, sowohl auf der Item- als auch auf der Personen-Ebene, was wiederum die Validität des Tests laut Messicks Definition (1995) unterstreicht.

6.5.5 Externe Aspekte

Der Zusammenhang zwischen Testergebnissen und anderen externen Variablen ist ein weiterer Aspekt von Validität. Messick (1995) schlägt vor, dabei sowohl konvergente Variablen, die eng mit dem zu messenden Konstrukt zusammenhängen, als auch divergente Konstrukte, die nur schwach oder gar nicht mit den Testergebnissen in Verbindung stehen, in Betracht zu ziehen.

In vorherigen Studien ist der starke Zusammenhang zwischen mündlichen Definitionen und den Ergebnissen aus Ja/Nein-Wortschatztests häufig gezeigt worden (z. B. Anderson & Freebody, 1983; Mochida & Harrington, 2006). Für die vorliegende Untersuchung wurden in einer Pilotstudie Daten zur mündlichen Definition von Kindern erhoben. Die teilnehmenden Kinder (N = 27, Alter M = 10.3, SD = 0.57) wurden nach Durchführung des Wortschatztests WOR-TE aufgefordert, mündliche Definitionen, sowohl zu einem Teil der angekreuzten als auch zu einem Teil der nicht angekreuzten Wörter, zu geben. Die Definitionen wurden auf Grundlage ihres semantischen Gehalts auf einer Skala von 0 bis 3 Punkten in Anlehnung an Gutierrez-Cheflen und DeCurtis (1999) bewertet. Die Ergebnisse zeigten zum einen, dass angekreuzte Wörter besser definiert werden konnten (M = 1.07, SD = 0.93) als nicht angekreuzte (M = 0.30, SD = 0.68). Zudem erwies sich ein hoher Zusammenhang zwischen dem summierten Definitionsergebnis und dem Personenparameter im WOR-TE, r = .69. Dies ist vergleichbar mit vorherigen Studien und weist darauf hin, dass der mit dem WOR-TE gemessene orthographische Wortschatz auch eng mit semantischen Kenntnissen über Wörter verbunden ist.

Zusätzlich enthielt die vorliegende Studie zur Messung der konvergenten Validität mehrere (standardisierte) Instrumente, die Variablen, die eng mit dem orthographischen Wortschatz verknüpft sind, erheben. Dazu wurden der Wortschatz mit einem Multiple-Choice-Verfahren (KFT) gemessen, die Leseflüssigkeit (SLS) und die Schreibfähigkeit (HSP) erhoben.

Tabelle 6.3 zeigt die Interkorrelationen der Personenvariablen (manifeste im oberen, minderungskorrigierte im unteren Dreieck). Es konnten wie erwartet moderate bis hohe Korrelationen des Personenparameters aus dem Ja/Nein-Wortschatztest mit den anderen Konstrukten gemessen werden. Ein größerer orthographischer Wortschatz hängt somit eng mit dem Wortschatz, der Leseflüssigkeit und der Schreibfähigkeit zusammen. Dies steht in Einklang mit vorherigen Studien zum Zusammenhang des Wortschatzes zu anderen Variablen (Anderson, Wilson, & Fielding, 1988; Aarnoutse, van Leeuwe, Voeten, & Oud, 2001) und zeigt, dass der WOR-TE tatsächlich den Wortschatz erfasst.

TABELLE 6.3: Interkorrelationen (Pearsons *r*) der Personenvariablen für konvergente und divergente Validität.

	Interkorrelationen				
	1	2	3	4	5
1 WOR-TE	[.90]	.51 (.14)	.40 (.17)	.37 (.21)	.21 (.12)
2 Multiple Choice-Wortschatz (KFT)	.64 (.14)	[.64]	.28	.43	.22
3 Leseflüssigkeit (SLS)	.41 (.17)	.36	[.96]	.49	.04
4 Schreibfähigkeit (HSP)	.38 (.21)	.54	.50	[.99]	.19
5 Nonverbale Intelligenz (CFT)	.26 (.14)	.34	.05	.23	[.67]

Anmerkungen: Die obere Dreiecksmatrix enthält die manifesten, die untere die minderungskorrigierten Korrelationen. Die Reliabilität ist in der Diagonale in eckigen Klammern angegeben. Da die standardisierten Instrumente altersspezifische Werte ergeben, wurden die Korrelationen mit dem WOR-TE für jede Klassenstufe separat berechnet und anschließend gemittelt. Standardabweichungen sind in runden Klammern angegeben.

Zur Messung der divergenten Validität wurde die nonverbale Intelligenz mithilfe eines CFT-Subtests (Matrizen) erhoben. Die Korrelation mit dem WOR-TE ist ebenfalls in Tabelle 3 dargestellt. Wie erwartet fällt sie relativ gering aus, $r = .26$. Ähnliche Ergebnisse wurden in anderen deutschen Wortschatztests gefunden (z. B. WWT, Glück, 2011).

Bezüglich der externen Aspekte der konvergenten Validität konnten plausible Korrelationen für den WOR-TE mit anderen, dem orthographischen Wortschatz nahen Konstrukten gefunden werden. Zudem zeigen Daten aus einer Pilotstudie mit mündlichen Definitionen ähnliche Ergebnisse wie frühere Studien zu Ja/Nein-Wortschatztests. Für die divergente Validität wurde ein geringer Zusammenhang zwischen Wortschatz und nonverbaler Intelligenz gezeigt. Die Ergebnisse zur externen Validität sind damit zufriedenstellend.

6.5.6 GÜLTIGKEIT DES RASCH-MODELLS

Ein wesentlicher Aspekt bei der Verwendung des Rasch-Modells für eine Testanalyse ist die Prüfung der Gültigkeit des Modells. Zwar liegt dafür kein allgemeingültiges Verfahren vor, dennoch können verschiedene Analysen, die die Annahmen des Modells bestätigen, zur Prüfung der Passung herangezogen werden (Rost, 1999). Viele dieser Analysen sind bereits im vorgestellten Validitätskonzept enthalten.

Zum einen betrifft dies Analysen, die sich auf die Passung des Modells auf den Datensatz beziehen. In der vorliegenden Analyse sind dazu die Split-Half-Korrelationen heranzuziehen. Sowohl eine Aufteilung der Items in zwei Gruppen als auch eine Aufteilung der Personen in zwei Gruppen ergab eine hohe Korrelation der jeweiligen korrespondierenden Parameter. Die Modellannahme der Stichprobenunabhängigkeit ist damit bestätigt und spricht für die Modellpassung. Die Modellannahme der lokalen stochastischen Unabhängigkeit konnte zudem mit der Q3-Statistik unterstrichen werden. Zum anderen können zur Prüfung des Modells Vergleiche mit anderen Modellen, die aus theoretischer Sicht sinnvoll sind und ebenfalls auf die Daten passen könnten, in Betracht gezogen werden (Rost, 1999). Hierzu wurde das vorliegende Modell mit Modellen mit mehreren Dimensionen verglichen, zum einen auf Ebene der Testversionen, zum anderen auf Ebene der Wortarten. Aus theoretischer Sicht liegt darin die Annahme, dass in den verschiedenen Altersgruppen (Testversionen) unterschiedliche Fähigkeiten zur Lösung des Tests benötigt werden bzw. für die verschiedenen Wortarten jeweils andere Kompetenzen gefragt sind. Beide Modelle zeigten keinen signifikanten Unterschied zum ursprünglichen Modell, was dessen Passung ebenfalls unterstreicht. Zusätzlich wurde zur Überprüfung der Modell-Passung der Likelihood-Ratio-Test nach Andersen (Glas & Verhelst, 1995) einzeln für ein Modell pro Klassenstufe geschätzt. Lediglich für die 2. Klasse ergab sich ein leicht signifikantes Ergebnis (p = .04), in allen anderen Klassenstufen war der Test nicht signifikant (alle p >.1). Zusammenfassend kann davon ausgegangen werden, dass das Modell ausreichend auf die Daten passt.

6.6 Diskussion

In diesem Artikel wurde der Ja/Nein-Wortschatztest WOR-TE für Grundschulkinder vorgestellt und anhand des Rasch-Modells validiert. Der Test enthält drei Versionen für verschiedene Altersgruppen und kann im Gruppensetting innerhalb von kurzer Zeit angewendet werden. Gegenüber anderen Verfahren (z. B. PPVT- 4; Lenhard, Lenhard, Segerer & Suggate, 2015; PDSS, Kauschke & Siegmüller, 2009) hat er damit den klaren Vorteil, dass er in einer Gruppensituation mit mehreren Kindern angewendet werden kann. Zudem ist er für eine Altersgruppe konzipiert, für deren Messung im Bereich Wortschatz bisher wenige Verfahren vorlagen. Bereits existierende Verfahren (z. B. WWT; Glück, 2009) zielen eher auf die Diagnostik semantisch-lexikalischer Defizite ab. Der WOR-TE hingegen ist eher ressourcenorientiert und zur Messung des orthographischen Wortschatzes von Kindern geeignet. Gegenüber anderen Verfahren, die beispielsweise das Finden von Synonymen beinhalten (z. B. CFT-20-R; Weiß, 1986) hat der WOR-TE den Vorteil, dass die Abhängigkeit von den Distraktor-Items, in diesem Fall die Pseudowörter, relativ gering ist, was sich in den geringen Korrelationen mit dem Testverhalten gezeigt hat. Allen anderen Testverfahren hat der WOR-TE zudem die hohe Anzahl an Test-Items, die durch das einfache Testformat begründet sind, voraus.

In der vorliegenden Studie wurde versucht, anhand des Validitätskonzepts nach Messick (1995) Evidenz für die Validität des WOR-TE zu finden. Diese Aspekte umfassen inhaltliche, substanzielle, strukturelle und externe Aspekte sowie Generalisierbarkeit. Die Gültigkeit des Rasch-Modells wurde anhand verschiedener Aspekte als ausreichend betrachtet.

Zusammengefasst liegen starke Hinweise für die Validität des Verfahrens zur Messung des kindlichen orthographischen Wortschatzes vor. Es wurden zufriedenstellende Ergebnisse für alle von Messick (1995) vorgeschlagenen Aspekte der Validität erzielt. Die Analysen gaben zudem Hinweise auf Möglichkeiten zur weiteren Verbesserung des Verfahrens, insbesondere bezüglich der Anzahl der Items und der Itemauswahl. Alles in allem besteht eine starke Evidenz dafür, dass es sich bei dem WOR-TE um ein valides Instrument zur Erfassung des orthographischen Wortschatzes bei Grundschulkindern im Deutschen handelt. Insbesondere in forschungsbezogenen Kontexten stellt er damit eine gute Option zur Erfassung des kindlichen Wortschatzes dar. Da keine Normwerte vorliegen, ist eine Individualdiagnose derzeit mit dem Instrument jedoch nicht möglich.

Während die Auswertung von Ja/Nein-Wortschatztests in vorherigen Studien häufig anhand der Hits und False-Alarm-Raten erfolgte (z. B. Eyckmans, 2004; Huibregtse, Admiraal, & Merea, 2002), wurde in der vorliegenden Studie lediglich auf die Hits zurückgegriffen, um die Auswertung mit dem Rasch-Modell zu ermöglichen. Mochida und Harrington (2006) konnten bereits zeigen, dass die alleinige Auswertung der Hits am besten mit anderen Wortschatzmaßen korrelierte. Auch unsere Ergebnisse sprechen dafür, dass die Korrektur mithilfe der False-Alarm-Rate nicht notwendig ist. Es bestand nur eine schwache Beziehung zwischen Rateverhalten und Personenparameter, die sogar tendenziell darauf hinwies, dass Kinder mit höherer Ratetendenz einen geringeren Personenscore hatten und damit weniger Wörter angekreuzt haben. Korrigiert man anhand der False-Alarm-Rate, geht man davon aus, dass Personen, die mehr raten, auch generell *zu viele* Wörter angekreuzt haben (Mochida & Harrington, 2006), was in den vorliegenden Daten nicht der Fall ist. Dies rechtfertigt zunächst die alleinige Verwendung der Hit-Raten. Dennoch werden weiterhin verschiedene Methoden der Auswertung und Korrektur von Ergebnissen von Ja/Nein-Wortschatztests diskutiert (vgl. Huibregtse, Admiraal, & Merea, 2002; Pellicer-Sánchez & Schmitt, 2012). Eine weitere Analyse der vorliegenden Daten könnte hier weitergehende Erkenntnisse liefern.

Einschränkend lässt sich zudem festhalten, dass der Einsatz des Tests bei Leseanfängerinnen und Leseanfängern kritisch zu sehen ist. Dies zeigt sich durch die erhöhten Ratetendenzen in der 1. und 2. Klasse. Die Lesefähigkeit ist in diesen Klassenstufen wohlmöglich noch zu gering, sodass eine Erhebung des orthographischen Wortschatzes erst später möglich ist. Limitierend ist für diese junge Altersgruppe auch die Reliabilität im standardisierten Wortschatztest aus dem KFT zu nennen. Möglicherweise ist der Wortschatz in diesem Alter von Kind zu Kind sehr unterschiedlich (siehe auch Segbers & Schroeder, 2017), was eine reliable Messung mithilfe einer kleinen Item-Anzahl erschwert. Ein Vergleich der Testdaten mit auditiv vorgegebenen Wörtern in höherer Anzahl als Ja/Nein-Verfahren könnte hier eine sinnvolle Ergänzung sein.

Die Analyse des Effekts von Mehrsprachigkeit auf das Testergebnis zeigte, dass es Unterschiede zwischen ein- und mehrsprachigen Kindern im Testverhalten gibt. Eine detailliertere Analyse dieser Unterschiede könnte Aufschluss darüber geben, inwiefern der Einsatz des Tests bei mehrsprachigen Kindern sinnvoll ist bzw. die Ergebnisse mit denen der einsprachigen Kinder vergleichbar sind. Weiterhin lässt sich anmerken, dass das Verfahren nicht zur Erfassung von detailliertem Wortschatzwissen, insbesondere auf der semantischen Ebene, geeignet ist. Zwar sind die Ergebnisse aus der Pilotstudie mit den mündlichen Definitionsaufgaben vielversprechend, dennoch können mit dem WOR-TE keine detaillierten Aussagen über das semantische Wissen gemacht werden. Um dieses zu erfassen und eine differenzierte Individualdiagnose zu erstellen, sind aufwändigere Testverfahren von Nöten. Bei dem vorgestellten Instrument handelt es sich also um eine Möglichkeit zur Erfassung des orthographischen Wortschatzes, der substanziell mit dem semantischen Wortschatz zusammenhängt.

Zusätzlich konnte gezeigt werden, dass die Verwendung des Validitätskonzepts nach Messick (1995) die Möglichkeiten zur Validierung eines Ja/Nein-Wortschatztests über die üblichen Korrelationen mit mündlichen Definitionen oder Multiple-Choice-Fragen hinaus erweitert. Die vorliegenden Analysen beinhalteten relevante Schritte zur Sicherung von Evidenz für die Validität eines Verfahrens und zur Absicherung und Verbesserung der Qualität eines Instruments. Das vorgestellte Vorgehen zur Validierung kann damit als wichtiger Beitrag für die

Entwicklung von Ja/Nein-Wortschatztests angesehen werden und sollte für die zukünftige Konstruktion ähnlicher Instrumente in Betracht gezogen werden.

7 Study 2: Lexicon Size Estimation

This chapter was published:

Segbers, J. (now Trautwein) & Schroeder, S. (2017). How many words do children know? A corpus-based estimation of children's lexicon size. *Language Testing, 34*(3), 297-320.
doi: 10.1177/0265532216641152

7.1 ABSTRACT

In this article we present a new method for estimating children's total vocabulary size based on a language corpus in German. We drew a virtual sample of different lexicon sizes from a corpus and let the virtual sample "take" a vocabulary test by comparing whether the items were included in the virtual lexicons or not. This enabled us to identify the relation between test performance and total lexicon size. We then applied this relation to the test results of a real sample of children (grades 1–8, aged 6 to 14) and young adults (aged 18 to 25) and estimated their total vocabulary sizes. Average absolute vocabulary sizes ranged from 5900 lemmas in first grade to 73,000 for adults, with significant increases between adjacent grade levels except from first to second grade. Our analyses also allowed us to observe parts of speech and morphological development. Results thus shed light on the course of vocabulary development during primary school.

7.2 Introduction

How many words does a person know? This question has interested many researchers within the last decades and led to very different approaches to solve it (e.g. Seashore & Eckerson, 1940; Goulden, Nation, & Read, 1990). However, it is difficult to answer. While the measurement of vocabulary size in young children is relatively easy since they do not know a lot of words, determining lexicon size in older children or even adults is rather challenging. Because of the amount of words they know, it is simply impossible to assess them all directly. As a consequence, the estimation of vocabulary size is often based on dictionaries or frequency lists: a subset of words is tested and the results are projected to the total number of words in the dictionary or list (Nation, 1993a). Other authors estimate lexicon size by analyzing (written) language production (Pregel & Rickheit, 1987). However, because of the variation in methods, estimation results differ and often only address adults' vocabulary size (e.g. D'Anna, Zechmeister, & Hall, 1991). Owing to these methodological difficulties, no reliable estimates for children's vocabulary size in primary school are available. Yet, they are necessary to describe language acquisition processes and growth rates and thus to enrich theories of vocabulary development. In addition, they enable researchers and educators to investigate different causes for vocabulary deficits in children.

In this paper, we reuse and expand existing methods to estimate children's vocabulary size on the basis of a written corpus for children (*childLex*, Schroeder et al., 2015a) using a corpus-based sampling approach. First, we point out the importance of vocabulary in (written) language development and describe its development and assessment. Then we present previous studies and methods to estimate total vocabulary size. Finally, we introduce our approach, its methods and results, and discuss it according to previous findings and theories on vocabulary development.

7.2.1 Vocabulary: Implications, development, and assessment

Vocabulary is a crucial component of language competence and language use (Nation, 1993b). It has been shown to be related to other language domains such as grammar and phonology during language development (Bates & Goodman, 1999; Gathercole & Baddeley, 1989) and is strongly connected to auditory and reading comprehension (Tannenbaum, Torgesen, & Wagner, 2006; Ouellette, 2006). Early vocabulary predicts later reading ability and school success (Muter, Hulme, Snowling, & Stevenson, 2004; Grimm & Doil, 2005), and vocabulary and reading performance stay connected throughout the lifespan (Braze, Tabour, Schankweiler, & Mencl, 2007; Landi, 2010). Thus, measuring vocabulary is common in diagnosing language impairment (Hoff, 2014). It is also necessary for further specifying the relation of vocabulary to other cognitive and language-related abilities and for conducting and planning training and intervention programs (Nation, 2012).

In describing early vocabulary development one often differentiates between receptive and productive vocabulary since language comprehension develops prior to production. Fenson, Dale, Reznick, Thal, & Pethik (1994) found that children at the age of 8 to 10 months began to understand first words. At 16 months, they comprehended more than 150 words. At the age of 12 months, children start producing their first words and are able to speak about 50 words on average at 18 months. Following that, the growth rate increases and at the age of 24 months they use about 200 words (Hoff, 2014). Vocabulary development progresses but estimates for total vocabulary size and growth for older children or young adults can rarely be found and if so, they vary substantially owing to methodological differences (see following section). However, it is commonly assumed that receptive vocabulary exceeds productive vocabulary throughout the lifespan (Clark, Hutcheson, & van Buren, 1974). For English adults, total lexicon size is currently estimated to comprise about 50,000 words (Aitchison, 2012). Yet, it is unclear how children's vocabulary actually develops to finally reach this "goal."

As Hoff (2014) points out, early vocabulary of young children contains mostly nouns (45%). One reason for this is that they represent actual things in the children's environment; that is, they are perceptible for the child and thus their meaning is more transparent than for verbs for example. Analyzing the development of parts of speech distributions with growing vocabulary size is challenging for the same reasons lexicon size estimation itself is complicated. In their study with German school children, Pregel and Rickheit (1987) found that children's vocabularies contain about 55% nouns, about 35% verbs and 10% adjectives, based on language production of 6- to 10-year old children. They compared these numbers to Ruoff's (1981) results for adults, who estimated about 60% of nouns, 30% of verbs and 10% of adjectives. The reason for the increase in the proportion of nouns during vocabulary development is the fact that nouns are particularly likely to show effects of semantic differentiation (Clark, 1993). However, very little is known about how the prevalence of different parts of speech develops within primary school.

Another important question in vocabulary research is the nature of the relationship between lexical and morphological development. For example, Anglin, Miller & Wakefield (1993) found that, in English, vocabulary development is mainly driven by derivational processes and that bimorphemic words are most frequent in fifth graders' mental lexicons. By contrast, mono- and multimorphemic words are less frequent. It is unclear, however, whether these findings generalize to morphologically rich languages such as German.

In general, frequency of occurrence in a language determines which words are learned first. The more a child is exposed to a certain word the more likely he or she will be able to store it in his or her mental lexicon (Goodman, Dale, & Li, 2008; Naigles & Hoff-Ginsberg, 1998). While early language development is mostly driven by spoken language input, reading becomes more and more important in learning new words. As unknown words are more likely to appear in books (Hayes & Ahrens, 1988), the roles of print exposure and leisure time reading increase with age (Cunningham & Stanovich, 1991).

For very young children, vocabulary size is commonly tested by asking their parents to report which words out of a list their children understand (receptive vocabulary) and produce (productive vocabulary) (e.g. CDI, Fenson et al., 1993). Preschool children's vocabulary is mostly tested via picture naming (e.g. EVT-2, Williams, 2007) or picture choice after an auditory stimulus (e.g. PPVT-4, Dunn & Dunn, 2007). For school children as well as for adults, multiple–choice methods (e.g. finding a synonym out of a set of candidates) are often used regarding both their L1 and L2 vocabulary (e.g. Nation & Beglar, 2007). Another procedure introduced by Anderson and Freebody (1983) is the yes/no method, where test takers have to identify all words they know out of a list. To prevent guessing, there are pseudowords included. The authors found high correlations with actual knowledge of word meanings measured by definitions. Besides the advantages of multiple-choice methods, the yes/no tests afford less cognitive engagement and many items can be administered within a short period of time. Since the first introduction of the method, several studies have applied it in L1- as well as in L2-language testing (Mochida & Harrington, 2006; Lemhöfer & Broersma, 2012). We introduced a German version for primary school children, which was used in this study (Trautwein & Schroeder, 2018).

7.2.2 Previous studies on total lexicon size: Results and methodological issues

Many existing vocabulary tests focus on the measurement of relative vocabulary size compared to a norming sample (e.g. PPVT-4, Dunn & Dunn, 2007). Unfortunately, they provide no information on the total number of known words although these are relevant to describe vocabulary development on an average and individual level and to relate it to other developmental processes. However, various authors have tried to estimate people's total vocabulary size in different ways. Lorge and Chall (1963) distinguish between methods based on usage and sampling-based methods. In usage-based methods, spoken or written language production of the group of interest is analyzed and the number of different words is counted. With this method, Pregel and Rickheit (1987) estimated the vocabulary size of German school children aged from 6 to 10 years as consisting of up to 6900 words. However, they do not differentiate between age groups as the focus of their study was to obtain frequency norms. According to Seashore and Eckerson (1940), Marah (1872) estimated adults' vocabulary with this method to comprise from 3000 to 10,000 words. Since this approach is costly and does not provide estimates for the vocabulary size of individuals, many researchers have focused on sampling procedures. Here, a dictionary or a frequency list represents all possible known words in a language. A representative sample of words is then drawn from the dictionary or list and administered within a vocabulary test. The results are finally projected to the whole dictionary or list. One of the first attempts to estimate vocabulary size with this procedure is the study by Seashore and Eckerson (1940). They calculated a mean total vocabulary size of about 155,000 words for undergraduate college students. With the same method, Smith (1941) tested children's vocabulary size and estimated about 21,000 words for first-grade children, 38,000 words for third grade children and 43,000 for fifth-grade children. Anglin, Miller and Wakefield (1993) determined a lexicon size of about 10,000 in first, 20,000 in third, and 40,000 in fifth grade and also calculated an average growth rate of 20 words per day .

The differences between the reported results are caused by some important methodological issues as described by Nation (1993a). First, the size of the dictionary or list used is crucial. According to Lorge and Chall (1963), a larger dictionary provides a better basis for vocabulary estimation since it is more likely to contain all possible words a certain person might know. Nation (1993a) points out that the dictionary has to include more words than the average test taker is believed to know to ensure that vocabulary size is not underestimated. A second and very important methodological issue is the definition of a word within the dictionary and thus within the vocabulary. Therefore, researchers have to decide whether they count derivations (e.g. *drink* vs. *drinkable*) and inflections (e.g. *walk* vs. *walked*) as well as compounds (e.g. *main station*) as one or multiple entries. This decision reflects the assumption about representations of these morphological complex words within the mental lexicon and influences the conclusions that can be drawn from the results as well as the comparisons with other studies. Third, Nation (1993a) emphasizes the size and the compilation of the sample of items to be tested. He stresses that a larger sample leads to a smaller confidence interval for testing and thus to more accurate results. He therefore suggests that using a simple test design where a lot of items can be answered within a short amount of time without a lot of cognitive engagement. Furthermore, Nation highlights the importance of word frequency among the test items since high-frequency words are more likely to be known. He suggests ordering words by frequency classes of the same size and then taking the same number of words from every frequency level so that neither high- nor low-frequency words are overrepresented within the sample. Finally, Nation points out the necessity for authors to report clearly all the decisions described above, so that other researchers can evaluate and replicate the findings. In a later review, Nation (2012) advises, owing to technological progress and the emergence of language corpora, to prefer a frequency-based sampling over the dictionary-based method. He therefore suggests building up a corpus that contains a representative sample of the words of the language of interest. For German read by children, such a corpus was introduced by Schroeder and colleagues (2015a) and will be described below.

7.2.3 THE CHILDLEX CORPUS AND THE GERMAN LANGUAGE

The childLex corpus (Version 0.16.03; Schroeder et al., 2015a) is a written language corpus for German read by children and contains linguistic data for words from 500 children's books. It comprises about 10,000,000 tokens, 180,000 types and 117,000 lemmas. The corpus was intended to include books that are frequently read by children aged 6 to 12 years in school and in their leisure time. Both teacher and children questionnaires and library lending statistics were considered as part of the book selection process. Thus, we assume that it is representative for the written language exposure of German school children and that the relative frequencies of the corpus can be used to approximate the order in which words are learned

(Naigles & Hoff-Ginsberg, 1998). It therefore meets the criteria for the basis of vocabulary size estimation described by Nation (1993a, 2012).

As Nation (1993a, 2012) pointed out, the definition of the unit of analysis is crucial for vocabulary size estimation. In contrast to English, German is a morphologically rich language (Fleischer, Barz, & Schroder, 2012). Concerning its inflection, for example, a verb such as *lachen* ("to laugh") can appear in 13 different forms depending on person and tense (*ich lache, du lachst*, etc.). In comparison, in English there do exist four different forms of the word (*laugh, laughs, laughed, laughing*). Nouns and adjectives are also inflected according to number and case in German. Furthermore, German is a very productive language. Especially compounding is very common and, in contrast to English, compounds are mostly written without spacing (e.g. *Bahnhof* means *train station*). Also, derivation is very frequent in German, e.g. the prefix "un-" can be combined with adjectives to form an antonym (e.g. *glucklich* – *unglucklich*, *happy* – *unhappy*). While inflection in German is supposed to happen post-lexically, it is unclear whether compositions and derivations are stored as whole units within the mental lexicon or combined after retrieval of the single constituents (Fleischer et al., 2012). We therefore decided to use the lemma as the base unit of our analysis. In the following, a lemma is defined as the abstracted base form of a word. Thus, all inflectional forms of a word are represented by the same lemma whereas compounds and derivations are counted as different lemmas. D'Anna and colleagues (1991) argue that a lemma represents a base word in a language and thus is the best count for different words known. Thus, the 117,000 lemmas of the childLex corpus served as the basis for our vocabulary estimation method. Due to the fact that a lot of words in a language do only occur very infrequently (e.g.in childLex, 48.30% of lemmas occur only once within the corpus), a frequency-level classification and sampling scheme as suggested by Nation is not feasible: A lot of very infrequent words would have to be tested to project the results to the whole corpus as it was done in previous studies. We therefore decided to apply a sampling-based method which allowed us to draw item and person characteristics from the corpus.

7.2.4 Our approach: A corpus-based estimation of vocabulary size

In the present study, we estimate the vocabulary size of school children at different ages and of young adults. We created a vocabulary test based on the yes/no method introduced by Anderson and Freebody (1983). To determine total vocabulary size, we then reused and expanded the dictionary method described by Nation (2012) using the childLex corpus as our basis. Based on the assumption that the relative frequencies in childLex are representative for children's written language exposure, we drew virtual lexicons of different sizes from the corpus and let them "take" a vocabulary test. To this end, we repeatedly sampled different lexicon sizes from the corpus and checked whether or not the test items were included in the lexicon. Thus, given a specific lexicon size, we know the probability that a particular test item can be solved. This allowed us to identify the relation between test results and lexicon size. We then let a real sample of German school children and young adults take our vocabulary test and used the results from the virtual dataset for the estimation of participants' total vocabulary size. Our method also enabled us to compute vocabulary growth rates and to estimate the development of parts of speech proportions and morphological categories within the mental lexicon.

7.3 Method

7.3.1 Sample

Table 7.1: Sample Characteristics.

			Gender		
Grade	*N*	*M* Age	Male	Female	NA
1	37	6.6 (0.5)	13	23	1
2	49	7.3 (0.7)	24	25	0
3	75	8.0 (0.6)	38	35	2
4	107	9.0 (0.6)	65	42	0
5	62	10.0 (0.6)	22	40	0
6	68	11.2 (0.6)	35	33	0
8	73	12.8 (0.5)	21	51	1
Adults	30	22.4 (2.1)	14	16	0

Note: SD provided in parentheses.

A sample of 495 children took part in the study. Twenty-four children (4.84%) did not complete the vocabulary test, resulting in a final sample of 471 children (249 female, 217 male, five not reported). Children's data were collected in primary (grades 1 to 6) and secondary (grade 8) schools in Berlin. Thirty young adults (16 female, 14 male) were tested at the

Max Planck Institute for Human Development Berlin. The study was approved by the ethics committee of the Max Planck Institute for Human Development and by the school administration of Berlin. Participation was voluntary and based on parental consent if necessary. Children received candy for their participation and adults were reimbursed with 12 Euros. The number of participants in each grade as well as mean age and gender distribution are provided in Table 7.1.

7.3.2 Vocabulary Test

The self-developed Vocabulary Test was based on the yes/no method introduced by Anderson and Freebody (1983; Segbers & Schroeder, 2018). In this test, participants were presented with a list of 100 words and had to identify all words they knew. To prevent guessing, the list also contained 24 pseudowords. We created five test versions for different age groups (first/second grade, third/fourth grade, fifth/sixth grade, eighth grade, adults). Because item difficulty mainly depends on word frequency, we decreased mean log lemma frequency systematically in order to ensure an optimal level of test difficulty in each version (see Table 7.2). To link the different test versions, a subset of 20 words was used in all age groups and subsequent test versions each shared 10 overlapping link items. Owing to a technical error, there were 11 shared items between Version 5/6 and Version 8. Thus, the total number of items was 379. The number of items in each test version is provided in Table 7.2.

Table 7.2: Frequency distribution and number of items in different test versions.

	Log Frequency		*N* Items*		
Test Version	*M* (*SD*)	Range	Unique	Overlap	Link
1./2. Grade	1.5 (0.4)	2.7 – 1.0	70	10	20
3./4. Grade	0.6 (0.1)	0.9– 0.4	60	10	20
5./6. Grade	0.0 (0.1)	0.4 – -0.2	59	11	20
8. Grade	-0.3 (0.2)	0 – -0.6	59	10	20
Adults	-0.7 (0.2)	-0.1 – -1.0	70	---	20

** "Unique" indicates the number of items that appear only in the test version of this age group. "Overlap" indicates the number of items that are shared between the test version of this age group and the next higher age group. "Link" indicates the number of items shared between all test versions.*

Pseudowords were created by exchanging the vowels of a different list of real words (e.g. *schwach* to *schwich*) or by combining two existing morphemes (e.g. *Fuhrtum*) and were identical in all test versions. For each version, two randomized pseudoparallel forms A and B with different word orders were created. Participants were instructed to identify all known words. They were told explicitly that the list also comprised pseudowords and that thus guessing could easily be detected. Depending on participant's age, the test took between 5 and 15 minutes.

7.3.3 ANALYSIS

All analyses were performed with the Software R (R Core Team, 2015). Data analysis comprised of a sequence of four interconnected steps: First, we drew virtual samples of different lexicon sizes from the corpus. Second, we let the virtual samples "take" the vocabulary test and estimated corresponding item and person parameters using item response theory (IRT). This allowed us to determine the relationship between lexicon size and person parameters. Third, we used the virtual item parameters to estimate a person parameter for each participant in our empirical sample. As the relationship between person parameters and lexicon size is known, it is therefore possible to estimate an individual's vocabulary size and compute growth rates between age groups. In a final step, we additionally analyzed the development of different parts of speech and morphological categories. Each step is explained in detail in the Results section.

7.4 RESULTS

7.4.1 SAMPLING OF LEXICON SIZES

In a first step, we sampled virtual lexicons of different sizes from the childLex corpus. The sampling procedure was based on the list of all lemmas included in childLex (version 0.16.03; approx. 117,000 lemmas) and sensitive to the frequency of each lemma, i.e., high-frequency lemmas were more likely to be included in a sample than low-frequency lemmas. We varied lexicon sizes from 1000 to 115,000 lemmas. Between 1000 and 70,000 lemmas, lexicon size was increased in steps of 100. Between 70,000 and 115,000 lemmas, lexicon size was increased in steps of 5000 lemmas. This resulted in 700 different lexicon sizes which were sampled 100 times each by drawing the according number of lemmas from the corpus. All 70,000 virtual lexicons were used for further analyses. Owing to the sampling procedure, virtual lexicons of the same size could potentially comprise completely different lemmas. However, since sampling was based on word frequency, high-frequently used lemmas were more likely to be included in several lexicons at the same time. As a consequence, small lexicon sizes shared a substantial proportion of their lemmas with each other while large lexicon sizes were more heterogeneous. Thus, small lexicon sizes were more likely to contain high-frequency lemmas but could also comprise lemmas with low frequencies. With growing lexicon size the amount of low frequent lemmas increased.

7.4.2 Estimation of virtual item and person parameters

In a next step, we examined whether the 379 words of our vocabulary test were included in each virtual lexicon or not. In other words, we let our virtual lexicons "take" each of the five versions of the vocabulary test by comparing the sampled lexicons with our test items. If a test item was included in the lexicon, we considered it as known by the virtual participant with the according lexicon size. In case a test item was not included in the lexicon, we assumed that is was not known by this virtual participant. Thus, we were able to compute the probability that a particular test item can be solved as a function of the size of a lexicon.

The relationship between item solving probability and lexicon size is just a special case of the dependency between person ability and item difficulty. In order to analyze such relationships, item-response models are ideal and thus commonly used in educational testing. Model estimation was executed with the ltm package for R (Rizopoulos, 2006). In the present study, we analyzed our virtual data using the two-parameter-itemresponse model (2 PL-Model; Embretson & Reise, 2000; Bock & Zimowski, 1997). In the 2 PL-Model, the two item parameters of difficulty and discrimination as well as the person parameter representing the latent ability are estimated. The model fitted our data, $LL = -8307611$, and a comparison with the simpler one-parameter model ($LL = -8437834$) revealed a significant improvement of model fit for the two-parameter model, $\Delta\ \chi 2\ (378) = 260445$, $p < .01$.

Because the fit of the model was adequate, we fitted item parameters using Conditional Maximum Likelihood (CML) and saved them for further analyses. All 70,000 virtual lexicons were included in the analysis. Item difficulty ranged from −4.85 and 3.19 ($M = -0.36$, $SD = 1.98$); its distribution is displayed in Figure 7.1 A. Item discrimination ranked between 0.79 and 4.18 ($M = 1.63$, $SD = 0.78$) and is presented in Figure 7.1 B.

In addition, person parameters for the different lexicon sizes were calculated via expected-a-priori (EAP) estimation using the PP package for R (Reif, 2014). Estimated person parameters ranged from −4.10 to 5.52 ($M = 0.03$, $SD = 1.63$).

Because lexicon size (and, therefore, latent ability) was known a priori, this enabled us to relate person parameters and virtual lexicon sizes (see Figure 7.2). The relationship was very strong and could nearly perfectly be captured by a cubic function (R_2 = .99). Since we know the relation between person parameters and lexicon size, we are able to transform person parameters to lexicon sizes (and vice versa). This procedure can also be applied to person parameters derived from empirical samples.

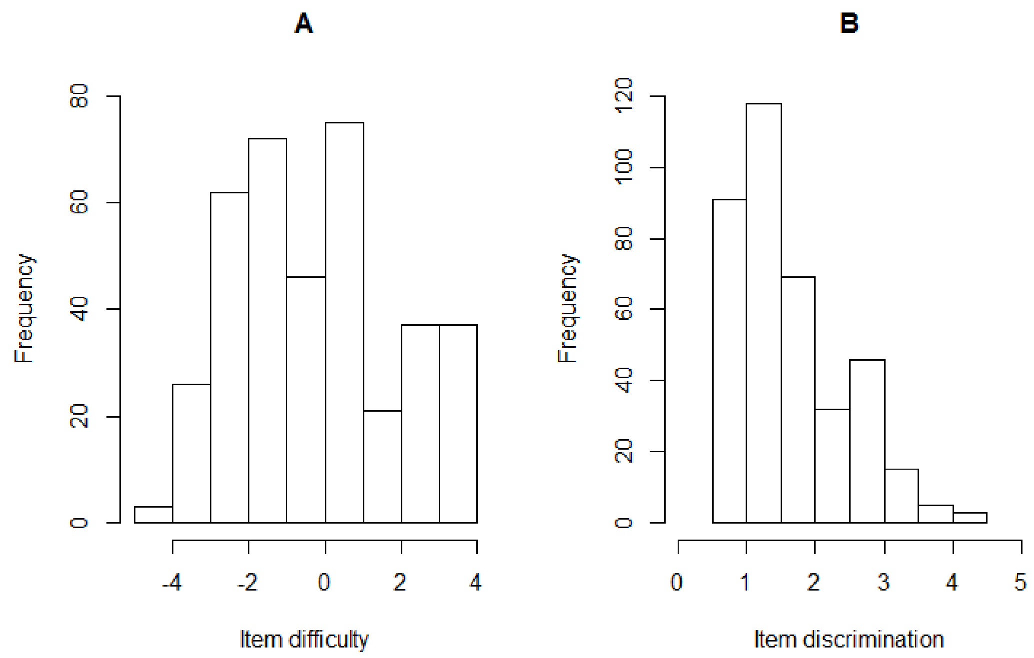

FIGURE 7.1: Distribution of item difficulty (A) and item discrimination (B) derived from the virtual lexicons.

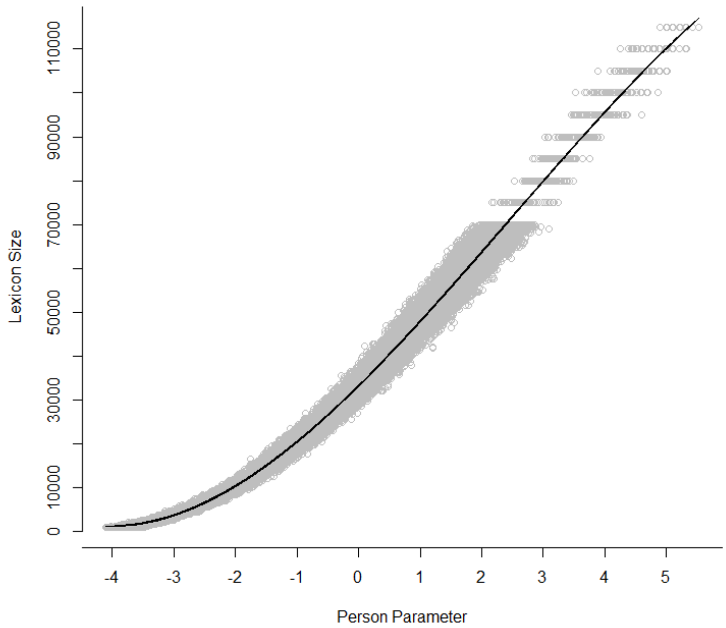

FIGURE 7.2: Relation of lexicon size and person parameter.

TABLE 7.3: Mean lexicon sizes and standard deviations per grade.

Grade	*M* Person Parameter (*SD*)	*M* Lexicon Size (*SD*)
1	-2.66 (0.44)	5925 (2481)
2	-2.68 (0.61)	6097 (4063)
3	-1.96 (0.54)	11,182 (4647)
4	-1.56 (0.54)	14,819 (5643)
5	-1.19 (0.60)	18,812 (6796)
6	-0.61 (0.67)	25,694 (8706)
8	0.30 (0.77)	38,029 (11,107)
Adults	2.60 (1.14)	73,625 (17,593)

7.4.3 ESTIMATION OF EMPIRICAL PERSON PARAMETERS AND LEXICON SIZES

In a last step, data from our empirical sample was analyzed. Since we know the relationship between item difficulty and person ability (see above), virtual item parameters were used to examine participants' actual performance on the Vocabulary Test and to estimate corresponding person parameters. Again, parameter estimation was based on EAP. The distribution of person parameters is provided in Figure 7.3. The overall mean was $M = -1.12$ (SD = 1.46) and parameters ranged between −3.92 and 4.51. Mean person parameters and standard deviations per grade are provided in the first column of Table 7.3.

To investigate whether the item parameters derived from the virtual lexicons were appropriate to fit our real data, we also estimated person parameters using item parameters derived from the empirical sample directly. The person parameters from both analyses correlated highly ($r = .93$) indicating that using the virtual item parameters was appropriate.

Finally, we were able to transform the empirical person parameters into individual lexicon sizes via the cubic function derived above. Mean lexicon sizes and standard deviations for each grade are provided in the second column of Table 7.3. As expected, our lexicon size estimation shows a growing trend, with about 6000 lemmas in first grade and about 73,000 lemmas in young adults.

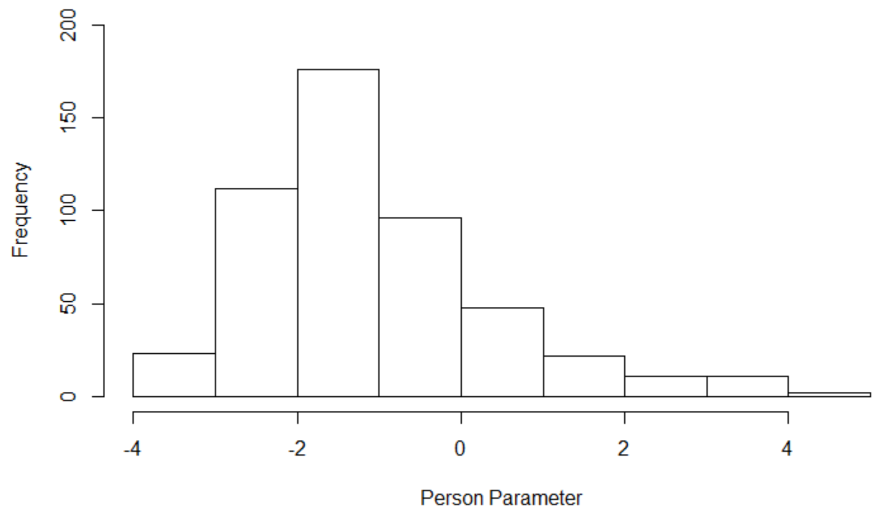

FIGURE 7.3: Distribution of person parameters in the empirical sample.

Figure 7.4 shows the growth of lexicon size between grades by plotting average lexicon size per grade against mean age per grade. An ANOVA revealed a significant effect of *grade* on lexicon size, $F(7, 493) = 302.8$, $p < .05$. Post-hoc analyses showed no significant difference between first and second grade. All other differences were significant (all $p < .01$). The high standard deviations, however, point to great interindividual differences within each grade.

Generally, vocabulary growth could nearly perfectly be described as a quadratic function of age, *R2* = .99, also displayed in Figure 7.4:

$$(1)\quad \textit{Lexicon Size} = -54.59 * \textit{Age}^2 + 6032.58 * \textit{Age} - 33938.01$$

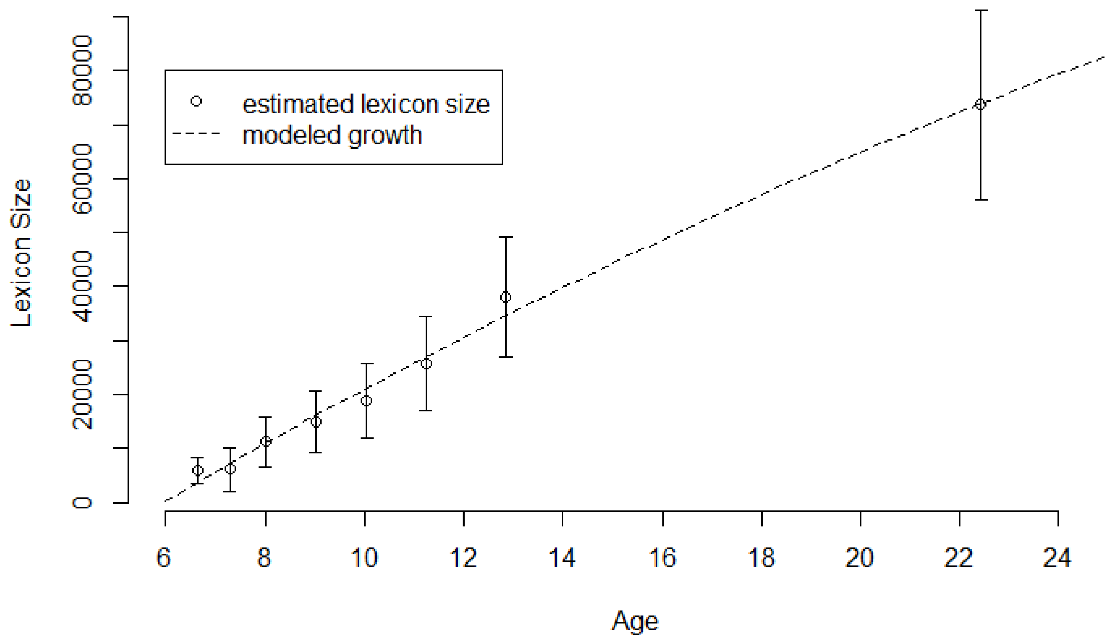

FIGURE 7.4: Development of lexicon size by age with the modeled quadratic function (bars represent standard deviations).

7.4.4 FURTHER ANALYSES: PARTS OF SPEECH DEVELOPMENT AND VOCABULARY GROWTH RATES

Our estimation method also allowed us to analyze the development of different parts of speech within the vocabulary (e.g. nouns, verbs and adjectives) as childLex contains parts-of-speech tagging. The sampling of different virtual lexicon sizes from childLex enabled us to count word classes within the lexicon. The investigated categories were *nouns, verbs, adjectives, function words*, and *others* (containing, e.g., proper names). The development of different word classes with growing lexicon size is displayed in Figure 7.5. Clearly, the number of words is increasing in every category (Figure 7.5 A). Regarding the proportions of the different parts of speech, only the noun category is growing while the proportions for the other parts of speech are slightly decreasing with growing lexicon size (Figure 7.5 B).

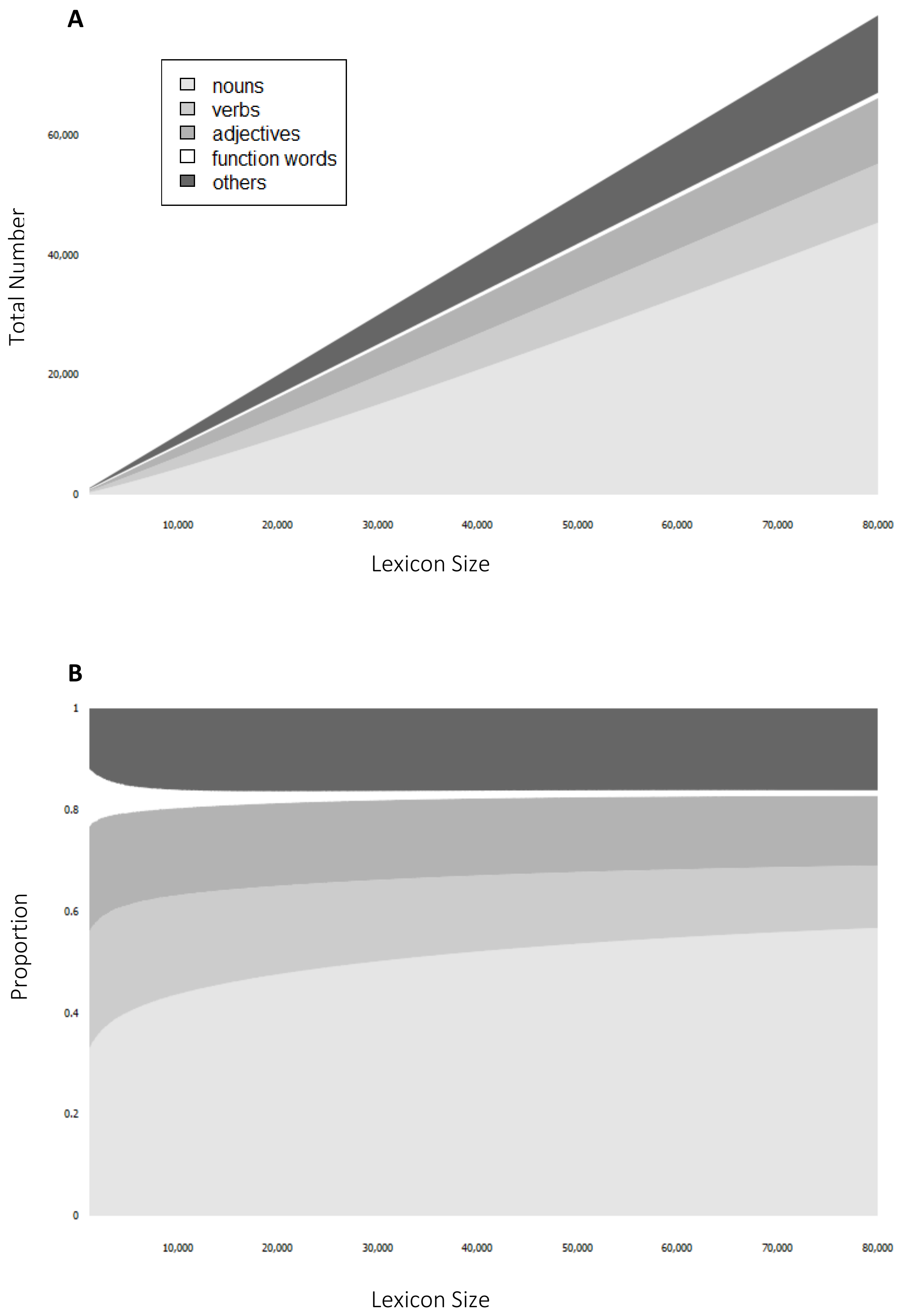

FIGURE 7.5: Development of parts of speech with lexicon size in total numbers (A) and proportions (B).

By identifying the function between lexicon size and of speech displayed in Figure 7.5 A, we were able to estimate the average numbers of parts of speech within the vocabularies of our empirical sample (Table 7.4). Results show that nouns dominate vocabulary at all stages of lexical development, followed by verbs, adjectives, other words (e.g. proper names), and function words.

Similar to Anglin, Miller and Wakefield (1993), we also calculated average vocabulary growth rates, that is the number of words learned per day. To this end, we computed the difference between grades for both total lexicon size and for the different word categories and divided it by the average age difference between grades (assuming one year per grade and 365 days per year). Results are displayed in Table 7.5. Whereas growth rates from first to second grade are relatively small, they range between about 10 and 20 words per day from grade 2 to grade 8. Again, most of children's vocabulary growth is driven by learning new nouns.

TABLE 7.4: Mean numbers of parts of speech per grade.

Grade	Nouns	Verbs	Adjectives	Function Words	Others
1	1697 (1426)	1655 (296)	1286 (338)	327 (20)	960 (401)
2	1795 (2335)	1676 (486)	1309 (554)	329 (33)	988 (656)
3	4718 (2671)	2284 (555)	2002 (633)	369 (37)	1809 (750)
4	6808 (3243)	2718 (674)	2498 (769)	399 (45)	2396 (911)
5	9103 (3906)	3195 (812)	3042 (926)	431 (55)	3041 (1097)
6	13,058 (5003)	4018 (1040)	3980 (1186)	486 (70)	4152 (1406)
8	20,147 (6383)	5492 (1327)	5661 (1514)	585 (89)	6144 (1794)
Adults	40,604(10,111)	9746 (2103)	10,512 (2398)	871 (141)	11,892 (2841)

Note: SD provided in parentheses.

Similarly, we examined the development of different morphological categories within our sample. To this end, all lemmas in childLex were analyzed using the morphological tagger SMOR (Schmid, Fitschen, & Heid, 2004). Here, we concentrate on two different variables: Morphemic complexity (mono-, bi- or multimorphemic, i.e. words that consisted of three or more morphemes) and morphological category (monomorphemic, derivation, composition, derivation & composition).

Table 7.5: Estimated vocabulary growth rates (in words per day).

	Parts of Speech					
	Nouns	Verbs	Adjectives	Function Words	Others	Total
Grade 1 – Grade 2	0.27	0.06	0.06	0	0.08	0.47
Grade 2 – Grade 3	8.01	1.67	1.90	0.11	2.25	13.93
Grade 3 – Grade 4	5.73	1.19	1.36	0.08	1.61	9.96
Grade 4 – Grade 5	6.29	1.31	1.49	0.09	1.77	10.94
Grade 5 – Grade 6	10.83	2.25	2.57	0.15	3.04	18.85
Grade 6 – Grade 8	9.71	2.02	2.30	0.14	2.73	16.90
Grade 8 - Adults	14.01	2.91	3.32	0.20	9.94	24.38

Since our analyses are based on lemmas, inflection was not included in the morphological categorization. Figure 6 shows the development of words with different morphemic complexity with increasing lexicon size. Clearly, the number of lemmas in each category increases (Figure 7.6 A). However, the corresponding proportions (Figure 7.6 B) show that the percentage of monomorphemic lemmas consistently decreases while the percentage of bi- and multimorphemic lemmas increases with growing lexicon size. Thus, most of children’s lexicon growth is driven by the acquisition of morphologically complex words.

The same trend can be seen in Table 7.6, which provides the average number of words in each complexity category separately for each grade level. While the number of bi- and multimorphemic lemmas is relative small in first-graders, it strongly increases during lexical development. In adults, bi- and multimorphemic words constitute the majority of words in the lexicon.

TABLE 7.6: Means of numbers of words by morphemic complexity per grade.

Grade	Monomorphemic	Bimorphemic	Multimorphemic
1	4,865 (1,334)	974 (1,011)	87 (136)
2	4,940 (2,131)	1,054 (1,690)	102 (243)
3	7,624 (2,396)	3,159 (1,960)	399 (293)
4	9,478 (2,809)	4,706 (2,443)	635 (393)
5	11,475 (3,330)	6,439 (2,995)	916 (504)
6	14,730 (3,971)	9,514 (4,001)	1,450 (742)
8	20,202 (4,693)	15,274 (5,340)	2,552 (1,084)
Adults	33,005 (6,052)	33,785 (9,373)	6,835 (2,259)

Note: SD provided in parentheses.

To analyze further the development of morphological complex words, we classified the bi- and multimorphemic lemmas into different categories depending on whether they are formed by derivation, composition or a combination of both processes. Figure 7.7 shows the development of the different categories with increasing lexicon size. Although all categories generally increase during lexical development, the relative growth is most pronounced for compound words.

Table 7.7 shows the means of different morphological categories for each grade separately. In grade 1 and grade 2, derivation is the most prominent morphological category. After this, however, compounds become more frequent than derivations. The proportion of combinations of derivation and composition is generally relatively small.

Average growth rates (words/day) for the different morphological categories are provided in Table 7.8. Results again show that vocabulary development is mostly driven by monomorphemic words in early grades, but morphologically complex words and particularly compounds are becoming increasingly important during later lexical development.

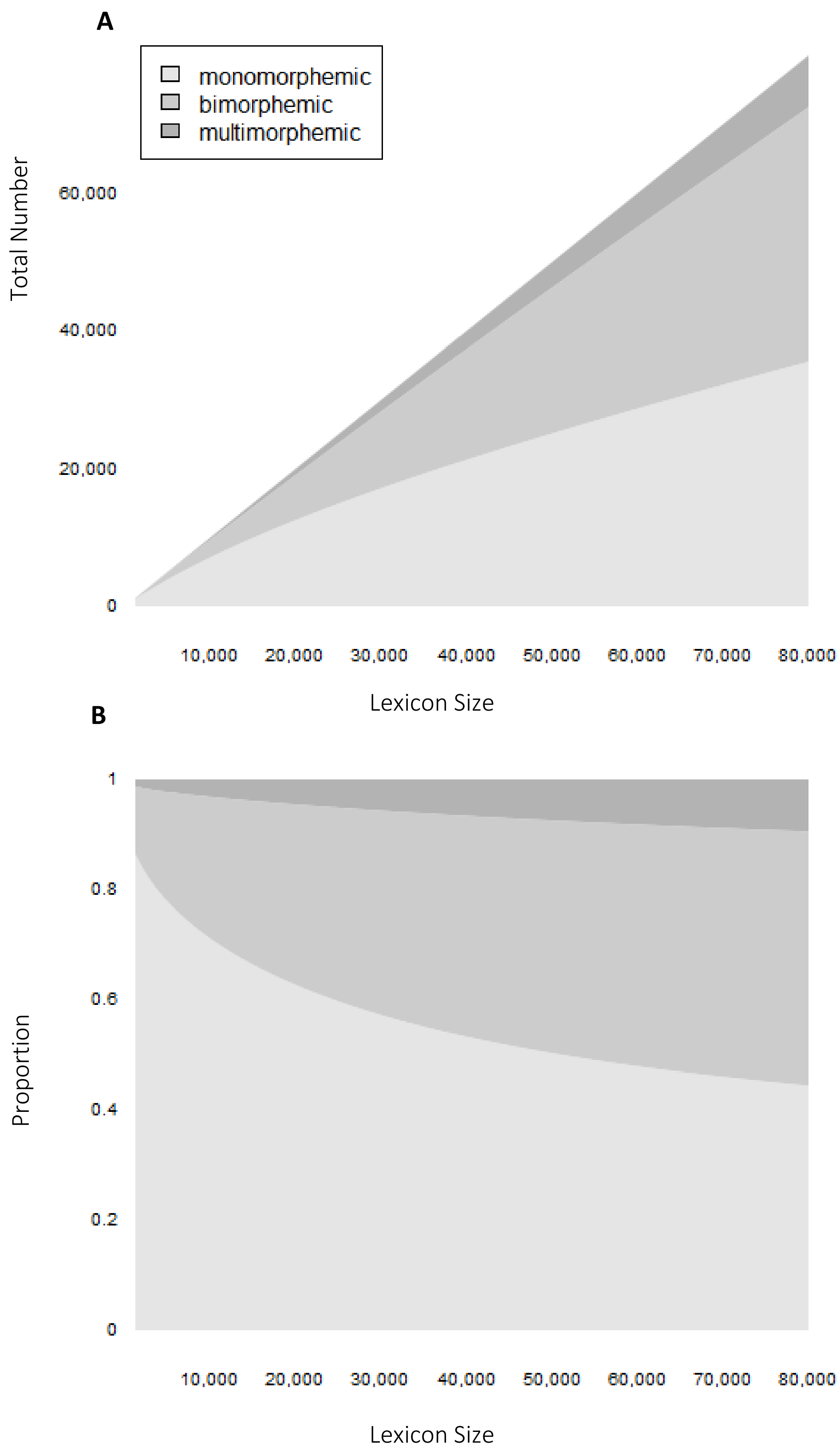

FIGURE 7.6: Development of morphological complexity with lexicon size in total numbers (A) and proportions (B).

TABLE 7.7: Means of morphological word types per grade.

Grade	Monomorphemic	Derivation	Composition	Derivation + Composition
1	4,865 (1,334)	602 (410)	427 (679)	32 (57)
2	4,940 (2,131)	628 (665)	490 (1,165)	38 (104)
3	7,624 (2,396)	1,462 (756)	1,931 (1,372)	165 (126)
4	9,478 (2,809)	2,052 (906)	3,023 (1,761)	267 (170)
5	11,475 (3,330)	2,692 (1,082)	4,275 (2,199)	388 (219)
6	14,730 (3,971)	3,780 (1,357)	6,563 (3,063)	620 (325)
8	20,202 (4,693)	5,685 (1,688)	11,037 (4,259)	1,104 (478)
Adults	33,005 (6,052)	10,840 (2,511)	26,770 (8,126)	3,010 (1,008)

Note: SD provided in parentheses.

TABLE 7.8: Growth rates for the morphological categories in words per day.

	Monomorphemic	Derivation	Composition	Derivation + Composition
Grade 1 – Grade 2	0.21	0.07	0.17	0.02
Grade 2 – Grade 3	7.35	2.29	3.95	0.35
Grade 3 – Grade 4	5.08	1.61	2.99	0.28
Grade 4 – Grade 5	5.42	1.75	3.43	0.33
Grade 5 – Grade 6	8.97	2.98	6.27	0.64
Grade 6 – Grade 8	7.50	2.61	6.13	0.66
Grade 8 - adults	8.77	3.53	10.78	1.31

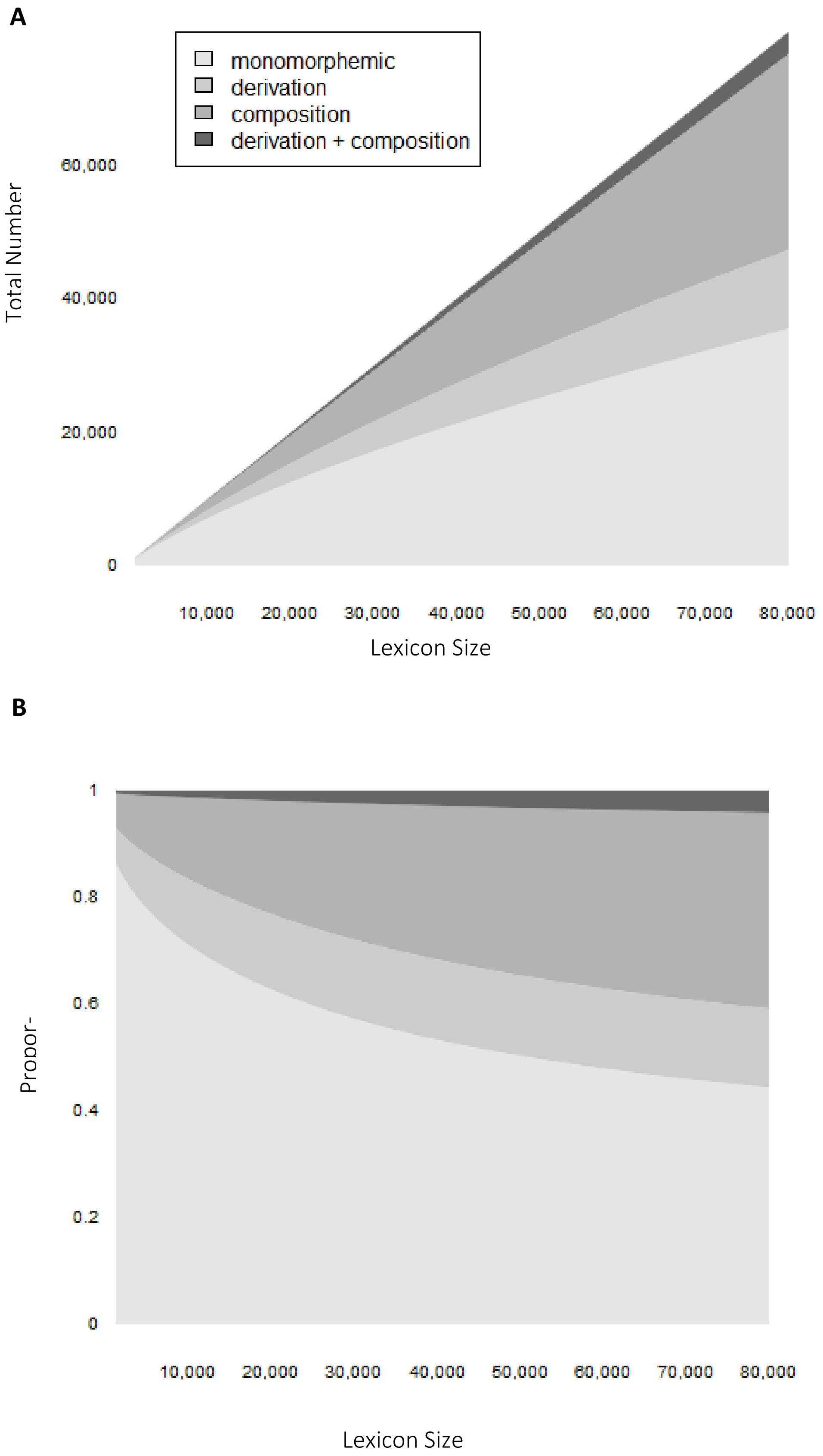

FIGURE 7.7: Development of morphological word types with lexicon size in total numbers (A) and proportions (B).

7.5 Discussion

In this study, we introduced a corpus-based method to estimate children's vocabulary size. We used a vocabulary test based on the yes/no method (Anderson & Freebody, 1983). From the childLex corpus, we drew a virtual sample of different lexicon sizes and estimated item and person parameters for our test with this sample by using the 2 PL-IRT Model (Embretson & Reise, 2000). This enabled us to identify the function between person parameters and lexicon size. We then let a real sample of school children from grades 1 to 6, 8 and young adults take the test. Person parameters were estimated using item difficulties derived from the virtual sampling procedure. As a consequence, we were able to compute the total lexicon sizes for each participant. Based on these estimates, we analyzed average vocabulary sizes and growth rates per grade as well as the proportion of different parts of speech and morphological categories with increasing lexicon size.

Our method is an extension of the previously used dictionary methods and meets the criteria for vocabulary size estimation described by Nation (2012). It is based on a corpus which is representative for the frequency structure of German read by children. Furthermore, we used a vocabulary test which contains a lot of items but requires only little cognitive effort. The fit of the 2 PL-Model was satisfying and the correlation of person parameters derived from the virtual item parameters with person parameters derived from the real data was high. We therefore think it is a useful method for the estimation of people's total vocabulary size.

Our estimation of school children's vocabulary fills the gap between the known vocabulary size of young children and the assumptions about adult lexicon size in German. Our results show lexicon size increases from approximately 6000 lemmas in first grade to about 73,000 lemmas in young adulthood. In comparison to previous studies on total vocabulary size in English children (see, e.g., Anglin, 1993), our estimates are considerably smaller. These differences may be due to methodological disparities: Anglin's estimation, for example, included root words as well as inflected and derived words, literal compounds and idioms. By contrast, our estimation was based only on lemmas, thus inflected words are excluded. When the number of inflected words is subtracted from total lexicon size, the estimates reported by Anglin, Miller and Wakefield (1993) and here are more similar (and the same holds for the study of Smith, 1941). Relatedly, the estimates provided by our method are substantially larger than the estimates that have previously been reported for German school children (e.g. Pregel &

Rickheit, 1987). Again, these discrepancies might be driven by methodological differences. While Pregel and Rickheit (1987) analyzed written and spoken language production, our method is based on a receptive vocabulary measure. Since receptive vocabulary is assumed to be larger than productive vocabulary (Clark et al., 1974), our results do not contradict but complement the findings of Pregel and Rickheit by also providing estimates for children's receptive vocabulary development. With regard to adults' vocabulary size, our estimate of 73,000 lemmas exceeds the commonly assumed number of 50,000 words in English (Aitchison, 2012). This difference might be ascribed to cross-linguistic disparities: German is a morphologically rich language and compounding is particularly frequent. As a consequence, the number of different words or lemmas is likely to be higher than in English. This, in turn leads to larger vocabulary size estimates for German adults. Surprisingly, we did not observe the same language effect for children's vocabularies which were generally smaller, but not larger than in English. However, none of the previous studies has tried to estimate the development of vocabulary from child- to adulthood as it was done in the present investigation, leading to comparable estimates for all age groups as in our study. This is important, because our results for the growth rates of different parts of speech and morphological categories showed that most of children's vocabulary growth was related to the acquisition of nouns and particularly compounds. This is in line with findings on early vocabulary acquisition in German (Hoff, 2014; Pregel & Rickheit, 1987).

With regard to vocabulary growth, we found strong differences between almost all age groups indicating a remarkable development of vocabulary from primary school to adulthood. Only between first and second grade, no significant increase in vocabulary size was observed. This finding might be explained by the fact that children are still learning to read during that time. Vocabulary growth in school is mostly driven by reading activity since new words are more likely to occur in written than in spoken language (Hayes & Ahrens, 1988; Nagy, Herman, & Anderson, 1985). During first grade, children's reading ability is at a low level, thus their reading input is relatively small, leading to limited vocabulary growth. After this initial phase, however, children's vocabulary grows by several thousand lemmas a year as has been reported in other studies (e.g. Anglin, Miller & Wakefield, 1993; Smith, 1941). Overall, the developmental trajectory could well be described by a quadratic function which is in line with prior findings on early vocabulary development (Kauschke & Hofmeister, 2002; Huttenlocher, Haight, Byrk, Seltzer, & Lyons, 1991).

The accelerating dynamic of children's vocabulary growth is also demonstrated in our average growth rates of learned words per year. Again, we only find small growth rates in grade 1. From grade 2 to grade 8, however, growth rates are substantial and vary between 10 and 20 learned words per day. After grade 8, growth rates are even higher with approximately 25 new words per day. Thus, vocabulary growth is not completed after the end of children's compulsory school education but is likely to increase owing to further education and experiences during young adulthood. Future research on older adult's vocabularies using our estimation method could shed light on the determinants of this process.

It is important to note that the standard deviations in all age groups were very high indicating substantial variability in vocabulary size. In grade 4, for example, low-performing children (−1SD) have functional vocabulary sizes that are similar to the average vocabulary size in grade 2 and high-performing children (+1SD) have vocabulary sizes that are similar to the average vocabulary size in grade 6. This highlights the importance of investigating interindividual differences in vocabulary development. In addition, it emphasizes the necessity to be able to assess the vocabulary size of each participant individually in language assessments, and not to rely on grade-level averages.

Our findings also confirm another crucial point: Given the enormous growth rates per year, vocabulary cannot only be taught in school alone (Nation, 1993b). As Jenkins, Stein, and Wysocki (1984) suggested, other activities such as leisure time reading, practicing hobbies or watching movies are similarly important. According to Nation (1993b), teachers should encourage their students to engage in such activities and thus support indirect vocabulary learning. Regarding the growth rates of parts of speech within the lexicon, we found that vocabulary development is mostly driven by the increasing number of known nouns. This leads to further implications for vocabulary growth and vocabulary teaching: Teaching nouns, directly or indirectly, plays an important role in supporting vocabulary growth.

Our investigation of morphological development showed children's lexical development is strongly driven by the acquisition of morphologically complex words. In early grades, most newly acquired complex words are derivations but in later grades the acquisition of compounds dominates lexical development. Anglin, Miller and Wakefield (1993) also observed a decline in the proportion of monomorphemic words and an increasing percentage of bi- and multimorphemic words with vocabulary development in English. However, in their study this trend was mostly driven by the acquisition of new derivations but not compounds. Again, this discrepancy might partly be explained by cross-linguistic differences. Compounding is very frequent in German and most compounds can be generated and understood spontaneously (e.g. the non-existing compound “Kleinkinderbaum” [small children’s tree] can easily be interpreted as a kind of tree that has been especially designed, planted, etc. for small children). This example demonstrates that orthographic and methodological differences might also contribute to the diverging findings between English and German. In contrast to German, compounds are usually written using spaces or hyphens in English. As a consequence, they are not recognized by algorithms that solely rely on white-space segmentation for tokenization (Jurish & Wurzner, 2013). In sum, our findings fit to the assumption that morphological processes become more and more important during vocabulary development. This has important educational implications and shows that it is essential to call attention to morphological processes such as derivation and composition to enrich the learner’s vocabulary.

In the present study, we used lemmas as the basic unit of analysis as it is particularly suited for an inflectionally rich language such as German. However, the method can be easily applied to other linguistic entities (e.g. inflected word types or stems) depending on the assumption on how words are stored in the mental lexicon. For example, if one assumes that inflected word forms constitute distinct lexical entries, the virtual sampling procedure as well as the selection of test items would simply be based on this unit of analysis. Similarly, the method can easily be extended to other corpora (e.g. for adults or other languages).

7.6 LIMITATIONS AND CHALLENGES

Although our approach appears very promising there are obviously also some limitations and challenges concerning the interpretation of our results and the application of the method for further research. First, we have to point out that our method crucially depends on the corpus which is used. As Kornai (2002) emphasizes, a language does not contain a finite number of words, mostly because of productive morphological processes which are especially important for languages such as German. In addition, as any corpus is only sample, its quality depends on its representativeness of the unobserved target population. This issue is particularly important with regard to frequency counts since our sampling method is sensitive to the frequency distribution in the corpus. In this regard, Fengxiang (2010) stresses that especially the number and frequency of rare words varies substantially with corpus size and might be underestimated.

We believe that the childLex corpus generally meets the necessary requirements regarding both quantity and quality. However, it also has some obvious limitations. The corpus mostly comprises narrative but not expository texts. Since children spend much more time reading narratives as opposed to non-narratives, this is consistent with the goal of representing the words that children have most likely encountered in their leisure-time reading (Topping, 2015). However, this necessarily limits its predictive value for students' performance on expository reading assignments which are more common in school settings. Furthermore, we also used the childLex frequencies to estimate adults' vocabulary size. Although there is a strong relationship between the frequency counts in childLex and corresponding adult corpora (Schroeder et al. 2015a), there are also clear discrepancies, especially for low-frequency words. As a consequence, the results for the adults might be less precise than for children. In summary, the methods employed for corpus construction are crucial for the application of our vocabulary estimation method and thus have to be carefully evaluated before using the approach.

We also have to emphasize that our method leads only to an approximation of total lexicon size and that there might be other factors than word frequency that drive vocabulary acquisition which were not considered here. Our estimation does not involve semantic information, for example concerning homonymy or polysemy. Also, our results do only contain information about the number of known words, and not on the quality with which the words are represented within the mental lexicon (Perfetti & Hart, 2002).

7.7 Conclusions

In conclusion, the reported findings make an important contribution to the discussion about vocabulary development in German across the school years. The proposed method cannot only be used to estimate total vocabulary size but also allows analyzing other linguistic phenomena such as the development of different parts of speech or morphological categories within the mental lexicon. The resulting estimates of total lexicon size are useful for describing and understanding language acquisition processes in German. Most importantly, in contrast to other methods, our approach enables researchers to estimate the lexicon size for each individual separately. Thus, differential developmental trajectories and their effects on children's reading performance can be investigated on an individual level. This is important in order to gain further insights about the relation between vocabulary and cognitive development, which might be used to improve existing training and intervention methods.

8 Study 3: Orthographic Networks in the Mental Lexicon

This chapter was published:

Trautwein, J. & Schroeder, S. (2018). Orthographic Similarities in the Developing Mental Lexicon. Insights from Graph Theory and Implications for Orthographic Development. *Frontiers in Psychology*, *2552* (9). doi: 10.3389/fpsyg.2018.02252

8.1 Abstract

In this study, we examine the development of orthographic networks in the mental lexicon using graph theory. According to this view, words are represented by nodes in a network and connected as a function of their orthographic similarity. With a sampling approach based on a language corpus for German school children, we were able to simulate lexical development for children from Grade 1–8. By sampling different lexicon sizes from the corpus, we were able to analyze the content of the orthographic lexicon at different time points and examined network characteristics using graph theory. Results show that, similar to semantic and phonological networks, orthographic networks possess small-word characteristics defined by short average path lengths between nodes and strong local clustering. Moreover, the interconnectivity of the network decreases with growth. Implications for the study of the effect of network measures on language processing are discussed.

8.2 INTRODUCTION

The study of the structure of the mental lexicon and its effect on lexical access has been of interest for several authors in the past. However, although it is unquestionable that orthographic neighborhoods affect word processing during reading development, the development of orthographic similarities in the mental lexicon has rarely been analyzed. The purpose of this study is to examine the development of orthographic similarities in the mental lexicon during reading acquisition by applying graph theory to simulated data of lexical development. We first highlight the importance of orthographic knowledge in reading development and then define neighborhoods in different language domains as well as their effect on language processing, also regarding developmental changes in effects. Subsequently we report studies on the development of neighborhoods in the mental lexicon and point out that there is a lack of studies on orthographic development. Afterward we present graph theory and its advantages to examine the neighborhood structure of the mental lexicon. Before presenting our methods and results on the development of orthographic networks, we describe the necessity of simulated data for the age group of interest. In the Discussion section, we relate our findings to current theories of orthographic development and effects of neighborhoods on visual word recognition. In addition, we discuss implications for future studies of language processing during lexical development.

8.2.1 ORTHOGRAPHIC KNOWLEDGE IN READING DEVELOPMENT

Sophisticated orthographic knowledge is crucial for reading competence. When decoding print, we compare the read word form with words already stored in our orthographic lexicon, which allows us to access semantic information as well as phonological information for reading out loud. This process is implemented in models of the reading process and is necessary for an efficient reading competence (e.g., the DRC, Coltheart et al., 2001). During reading development, a learner has to establish and improve his or her orthographic lexicon as well as the process of word retrieval from it. In models of reading development, it is assumed that children shift from letter-by-letter reading to a more word-based process of lexical access, probably because more and more words are stored as a whole in their orthographic lexicons (Acha and Perea, 2008). Castles et al. (2007) suggest that the recognition process shifts from a more broadly to a more finely tuned mechanism and support their theory with form-priming experiments. They showed form priming effects for developing readers but no effects for proficient readers. Their explanation involves the composition of the orthographic lexicon: Beginning readers only know a few words that are similar to the form prime. That is, the form prime eases activation of the target word. Proficient readers, however, know a lot of words similar to the form prime and so activation cannot concentrate on the target word alone. That is, in this framework, orthographic similarities within the orthographic lexicon play an important role in the reading process.

8.2.2 Neighborhoods in the Mental Lexicon: Structure and Effects

The mental lexicon comprises information on phonological, orthographic and semantic features of words. It is assumed that entries are interconnected due to shared features in these domains (Perfetti and Hart, 2002). That is, the mental lexicon can be conceptualized as a large network with nodes and connections. Directly connected words are usually referred to as neighbors. Semantic neighbors are words with similar semantic characteristics (e.g., salt – pepper; Aitchison, 2012). Phonological neighbors are defined as words that can be created by exchanging, deleting or adding one phoneme from another word (e.g., cat /kæt/– hat /hæt/; Yates, 2005). Similarly, orthographic neighbors are defined as words that can be created by exchanging, deleting, or adding a single letter from another word. Since the mental lexicon grows in size during language development (Segbers and Schroeder, 2017), the neighborhoods within the mental lexicon might also change. The study of neighborhoods in the mental lexicon is particularly interesting because neighbors have shown to influence language processing. For example, semantic neighbors often ease processing of target words in semantic priming experiments (e.g., Sánchez-Casas et al., 2006; Holderbaum and Fumagalli de Salles, 2011), although some studies also show an inhibitory effect depending on nature of semantic relation between prime and target (Abad et al., 2003; for a review see Neely, 2012). Furthermore, it has been shown that words with a lot of semantic neighbors can be retrieved faster (e.g., Buchanan et al., 2001), that is the effect of semantic neighbors is facilitative. For phonological neighbors, a study by Yates (2005) has led to similar results with a facilitative effect of phonological neighbors on visual word recognition. Mulatti et al. (2006) also found this effect for reading aloud. In this framework, the activation of neighbors boosts the activation of the target word. For orthographic neighborhoods, the results on the effect are controversial as summarized in the review by Andrews (1997). Although many studies also found facilitative effects, Andrews (1997) points out that the frequency of the neighbors also have an important influence. That is, the presence of high frequency neighbors inhibits the access of low frequency target words (see also Grainger, 1990; Sears et al., 1995; Grainger and Jacobs, 1996; Pollatsek et al., 1999; Grainger et al., 2005). In this framework, the activation of high frequency neighbors impedes the activation of the target word since they compete with each other.

8.2.3 Developmental Changes in Neighborhood Effects and Lexical Structure

Only a few studies addressed developmental changes in the effect of neighborhoods on word recognition. Holderbaum and Fumagalli de Salles (2011) found higher semantic priming effects for children in visual word recognition than for adults. This indicates that they rely more on semantic information than skilled readers, probably because their orthographic lexicon is still developing. For phonological neighborhoods, Metsala (1997) found developmental differences in the processing especially for words from sparse neighborhoods and low frequency words. He ascribes these findings to developmental changes in the mental lexicon which is refined during language acquisition. Castles et al. (2007) examined masked form-priming using orthographic primes that either differed in one letter from the target word (*rlay – play*) or where two letters were transposed (*lpay – play*). They found priming effects only for beginning readers which dissolved during development and also attributed these findings to a shift from a more broadly to a refined processing mechanism. Thus, the composition of the orthographic lexicon in terms of neighborhoods is directly linked to the development of reading competence. That is, although the number of studies analyzing developmental changes in neighborhood effects is limited, they all ascribe their results on developmental patterns to changes in the mental lexicon and its access during language and reading acquisition.

To some extent, developmental trajectories in lexical development have been assessed. For example, for semantic neighborhoods in the mental lexicon, Steyvers and Tenenbaum (2005) could show that new words enter the lexicon when they already have a lot of neighbors in the vocabulary. However, Hills et al. (2009) also tested further developmental mechanisms that might drive semantic neighborhood development. They conclude that words with many semantic neighbors in the learning environment are more noticeable and represent key words in the network which makes them important. Similar results have also been found in several further studies investigating semantic networks (Hills et al., 2010; Bilson et al., 2015). For phonological neighbors, a similar pattern has been reported. In particular, Storkel

(2004) showed that age of acquisition and phonological density influence phonological neighborhood growth and that words from dense phonological neighborhoods are learned earlier. Vitevitch and Storkel (2012) found the same pattern using computational models of network learning. Further evidence comes from Stamer and Vitevitch (2012) who showed that words from dense phonological neighborhoods are acquired earlier in second language learning. The only evidence for developmental changes in orthographic neighborhood size has been provided by Castles et al. (1999) who, however, used a completely different approach. They selected words with a high and a low orthographic neighborhood size and presented the target words as well as the neighbors on a list together with nonwords to children and adults. The participants were asked to identify all words they know out of the list. The authors considered identified words as existing neighbors in the participant's lexicon and called this measure their "effective neighborhood size." They then compared knowledge of neighbors of children and adults for words with dense and sparse neighborhoods. Importantly, children knew fewer neighbors than adults for words with dense and sparse neighborhoods. This is in line with the notion that children's effective neighborhood size is small for all words. Further analyses on orthographic network development are still missing. However, to understand and predict effects and developmental changes of neighborhoods during language acquisition, those analyses are necessary.

To sum up, the study of neighborhood effects in lexical access has been of interest in several different approaches, also regarding developmental changes. They are often ascribed to developmental changes in neighborhoods in the mental lexicon and are connected to changes in reading development. However, these developmental changes have not been determined for orthographic development yet. That is, the properties of orthographic networks are still unclear. However, since orthographic neighbors influence orthographic processing (see Andrews, 1997), the examination of the neighborhood structure and its influence on reading and writing is highly important. The aim of this study is thus to shed light on the courses of lexical development regarding orthographic neighborhoods. Results could be used to predict and explain effects of neighborhoods in reading development and processes in reading acquisition.

8.2.4 Analyzing Networks through Graph Theory

One approach to investigate connections in the mental lexicon and their development is graph theory. It has been used in some studies in order to analyze semantic (Steyvers and Tenenbaum, 2005; Zortea et al., 2014) and phonological networks (Vitevitch, 2008; Chan and Vitevitch, 2010) and is also applicable to other fields of network research such as brain interconnectivity in neuroimaging (e.g., Rubinov and Sporns, 2010; Van Wijk et al., 2010). Besides, network models can be used to identify conditional (in-)dependencies between variables or competencies, e.g., with regard to reading ability (Colé et al., 2018). According to the graph approach concerning the mental lexicon, words are represented as nodes and connections (= neighborhoods) as paths between nodes. Several measures can be used to describe the network. The number of nodes n represents the number of words in the mental lexicon. The number of links of a node *ki* is equal to the number of direct neighbors and is also referred to as the degree. It can be averaged across the whole network with <*k*>. The distribution of the degree P(*k*) represents the probability that a randomly chosen node has the degree *k* and is thus another measure of connectivity of the network. The average path length L and the maximum path length between two nodes (also referred to as the diameter) *D* represents the number of steps needed to get from one node to the other. The clustering coefficient *C* measures the probability that two neighbors of a node are neighbors themselves and is thus a measure of graph connectivity (Steyvers and Tenenbaum, 2005).

For all these measures, words that are not connected to the network are excluded. However, these so-called "lexical hermits" also provide information on how well the network is interconnected and should be considered as well. Thus, all measures yield information on the interconnectivity of the network which might influence language processing. They thus can be regarded as an extension of the traditional neighborhood measure.

For phonological networks, Chan and Vitevitch (2009, 2010) already demonstrated the use of network measures to analyze neighborhood effects in language processing above the traditional measure of neighborhood size. In this study, we will thus determine network measures for orthographic networks during lexical development. Furthermore, network measures allow the comparison of networks in different (language) domains. Several studies have shown that many networks possess small world characteristics (Watts and Strogatz, 1998). That is, they exhibit a high interconnectivity between nodes as indicated by a short average path length and a high clustering coefficient. Furthermore, such networks have a scale-free structure with a power-law degree distribution. This means that few nodes have many connections while many nodes only have few connections. This structure appears to be ideal for language processing since it allows a high local interconnectivity (= clusters) as well as easy global access through "bridges" that connect clusters (Beckage et al., 2011). Steyvers and Tenenbaum (2005) as well as Hills et al. (2009) demonstrated that this also holds for semantic networks in natural language. In addition, Vitevitch (2008) showed that this finding generalizes to phonological networks in English as well as in other languages (Arbesman et al., 2010). For orthographic networks, no comparable studies have been conducted yet. In particular, it is unclear at present, whether orthographic networks possess small-world characteristics and a scale-free structure similar to other language domains. Another aim of this study is thus to examine, whether orthographic networks are structured similar to other language domains with small world characteristics.

8.2.5 Challenges of Lexical Measurement

To examine changes in orthographic neighborhoods, it appears to be reasonable to analyze mental lexicons of children during orthographic development. During this phase, children add a lot of new entries to their vocabulary (Segbers and Schroeder, 2017), thus great developmental differences can be expected. In addition, several authors have assumed a lexical restructuring from a broader to a more fine-grained access process in this phase (e.g., Castles et al., 2007) which could also be due to developmental changes in lexical content. We thus decided to analyze orthographic lexicons for children from Grade 1 to 8. However, the measurement of vocabulary and thus orthographic networks is challenging. While the number of known words in young children is limited and thus relatively easy to estimate, the orthographic lexicon grows rapidly after children enter school (Anglin et al., 1993; Segbers and Schroeder, 2017). As a consequence, it is impossible to analyze the complete vocabulary by testing every word a child might possibly know. However, one way to approximate the size and content of children's orthographic lexicon is the dictionary method (e.g., Nation, 1993). In this method, words sampled from a dictionary are tested and then the results are projected onto the whole lexicon. Using a variant of this method, we (Segbers and Schroeder, 2017) have been able to estimate the average vocabulary size in grades 1–8 in German. In that study, word frequency was used as a proxy variable to simulate language learning, although other factors might also influence language development (see Discussion section). In the present study, we wanted to add to these findings and further analyzed the structure of the orthographic networks and the development of their characteristics. Since lexicon size increases dramatically between grades 1 and 8 (Segbers and Schroeder, 2017) we assumed a change in lexical structure supporting an enhancement of language processing for an improving, more efficient reading process. Since network characteristics have been shown to influence language processing, the findings were aimed to lead to implications for further research on the effect of network measures on lexical access.

For this purpose, we used the average vocabulary sizes per grade to simulate data on the content of vocabularies for 50 virtual participants. We did this by using the childLex corpus, a written language corpus which represents the reading environment of German children aged from 6 to 12 (Schroeder et al., 2015b). By sampling words from the corpus we conducted the content of lexical development for 50 virtual participants. As in the former study (Segbers and Schroeder, 2017) we used word frequency as a proxy variable which drives language learning. This enabled us to analyze the simulated data in terms of network measures and their development with growing lexicon size.

8.3 MATERIALS AND METHODS

8.3.1 SAMPLING PROCEDURE

We simulated 50 prototypical language learners in German who we subsequently refer to as virtual participants. The question whether effects are significant is less important in simulation studies because sample size can be arbitrarily increased. Instead, it is more useful to focus on overall developmental differences and the shape of the effects. As a consequence, the present sample size was chosen so that medium to large effects (r_0.3–0.4) could be detected with a power of 0.80 using a significance level of a = 0.05. Sampling was based on the childLex corpus (version 0.16, Schroeder et al., 2015b) which is a corpus consisting of 500 German children's books for a reading age from 6 to 12 years. It is thus representative for children's reading environment when they start to read. The complete corpus was treated as the fully developed adult network. The childLex corpus comprises ca. 10 million tokens which are distributed over approximately 180,000 types (distinct word forms including inflection etc.) and 120,000 lemmas (syntactic base forms; see Schroeder et al., 2015b, for details). As linguistic networks are typically analyzed on the type level, we used types (distinct orthographic sequences) in the following analysis (see Table 8.1). However, analyses on the lemma level lead to a very similar pattern of results concerning developmental changes. They are depicted in Table 8.2 of the Appendix (A.2). The sampling procedure was sensitive to type frequency, i.e., types that occur more often in the corpus were more likely to be drawn. We assumed that there were no differences in the overall size of the mental lexicon between children. That is, the size of the networks of all 50 virtual participants in each grade were identical and corresponded to the average grade-specific lexicon size reported by Segbers and Schroeder (2017) which are provided in Table 8.1.

The sampling procedure worked as follows: The estimated average lexicon size in grade 1 is 31,570 types. In a first step, we sampled 31,570 types from the childLex corpus for each of the 50 virtual participants. This set represented the initial state of their orthographic network and was different for each virtual participant. As the sampling procedure was sensitive to frequency, high-frequency types (function words, etc.) were likely to be included in all virtual lexicons. However, because the virtual lexicons were sampled independently, they also differed from each other – particularly in the low-frequency range. After this initial sampling step, all remaining types in the childLex corpus (i.e., 180,000–31,750) were used as the basis for the second step of the sampling procedure which represents the growth of the orthographic lexicon between grade 1 and 2. This set represents the learning environment, i.e., it comprises all words in children's print environment that are still left to be learned. As the initial lexicons differed between virtual participants, their learning environments were also different. Again, the sampling procedure was sensitive to word frequency, i.e., words which were more frequent in the learning environment were more likely to be drawn.

In a next step, 1,036 new types were added to each of the 50 virtual lexicons. This number is the difference between the average size of the lexicon in grade 1 (31,750 types) and the average size of the lexicon in grade 2 (32,606; see Table 8.1). The number of newly learned words was the same for all virtual participants and we thus assumed that there were no differences in the rate of lexicon growth. The new types were sampled from the individual learning environment of each virtual participant and the sampling procedure was again sensitive to frequency. Figure 8.1 schematically illustrates the sampling procedure for one virtual participant. The right column shows the learning environment at each time point: A the beginning, the total childLex corpus was used for sampling. For the following steps, the already learned words were excluded from the learning environment since they do not need to be learned anymore. The middle column shows the number of learned words for each time point. The left column illustrates the particular lexicon size on the type level for each grade.

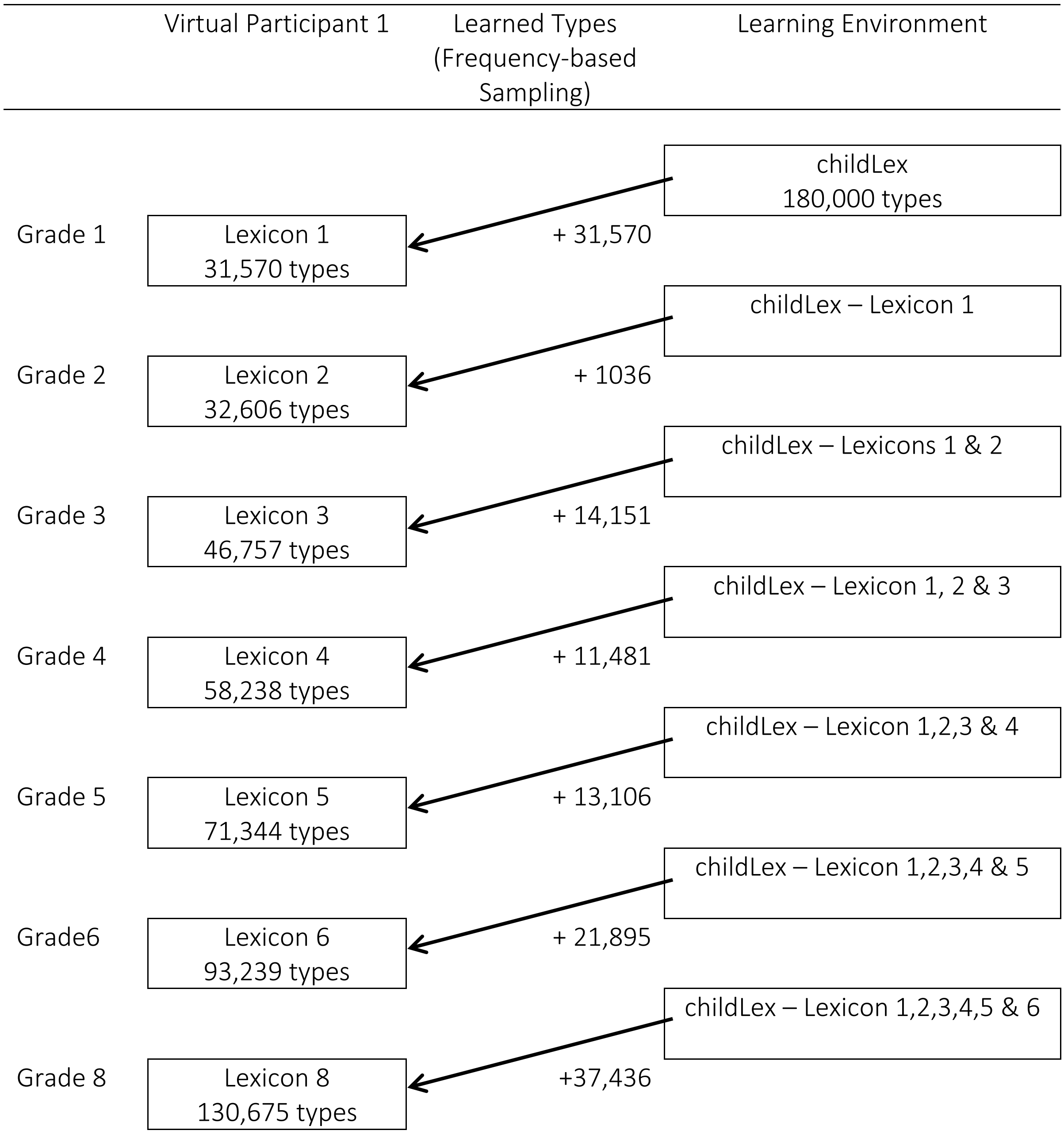

FIGURE 8.1: Schematic Illustration of the sampling procedure for one virtual participant.

This same procedure was repeated for the following grades (3, 4, 5, 6, and 8) until the lexicon had the size of the average language learner in grade 8. This resulted in seven orthographic lexicons for each of the 50 virtual participants (grades 1–6 and 8), that is, 350 lexicons in total. We compared the mean frequencies of the newly learned words in each age group with the mean frequencies of words per grade as determined by Segbers and Schroeder (2017). Both frequency trajectories showed a decreasing pattern, i.e., high-frequency words were acquired first and low-frequency words later. This pattern is also observed in the childLex corpus (see Schroeder et al., 2015b) indicating that the sampling procedure of the present study reflects the actual learning process of children in language acquisition.

8.3.2 Analyses

We computed important network characteristics for each grade specific lexicon of each virtual participant. These analyses were performed with the igraph (Csardi and Nepusz, 2006) and the vwr package (Keuleers, 2013) in R. For each virtual participant in each grade, an unweighted orthographic network was created in which types served as nodes and were connected with each other via paths if they were orthographic neighbors. Traditionally, orthographic neighbors have been defined by the substitution of a single letter (Coltheart et al., 1977). More recent approaches, however, assume that words created via deletions or insertions are also orthographically related (Yarkoni et al., 2008). Our definition of orthographic neighbors is thus based on the orthographic Levenshtein distance, a measure to quantify the (dis)similarity between two letter strings which takes substitutions, deletions, and insertions into account (Levenshtein, 1966). Orthographic neighbors were defined as words with a Levenshtein distance of 1, i.e., words that can be created by substituting, deleting, or inserting a single letter in a source word (e.g., for the word “hat”, not only words such as “hot” are neighbors but also words such as “hate”).

Following Vitevitch (2008) “lexical hermits” (i.e., nodes that do not connect to any other node in the lexicon) were excluded for the construction of the networks since some of the graph measures can only be applied to fully connected networks (also see Discussion for the role of lexical hermits). The number and proportion of lexical hermits were determined. For each network, we calculated the graph measures (for the explanation see also Introduction) n (number of nodes in the network), $<k>$ (mean degree, that is the mean number of neighbors per word), $P(k)$ (distribution of degree, that is the distribution of degrees for the whole network), L (average path length, that is the average number of paths to get from one node to another), D (maximum path length, that is the maximum number of paths to get from one node to another) and C (clustering coefficient, that is the probability of neighbors of a word to also be neighbors). C is calculated over all nodes i using the formula

$$C_i = T_i \,/\, \left(\frac{ki}{2}\right) = 2\ T_i \,/\, k_i\ (k_i - 1)$$

T_i can be referred to as the number of links between the neighbors, *k* of the node *i* and k_i ($k_i - 1$)/2 stands for the number of connections that would be assumed if all neighbors of a node were also neighbors (Steyvers and Tenenbaum, 2005).

Each network measure served as the outcome variable in a repeated measurement ANOVAs using grade (1, 2, 3, 4, 5, 6, 8) as a factor varying within virtual participants (which also stands for varying network size). Significant overall effects were complemented by computing t tests (using Tuckey's correction for multiple comparisons) between consecutive grades. In addition, the shape of the overall effect of grade on each network measure was analyzed by fitting different functions to the data.

8.4 RESULTS

In our analyses we were interested in the properties of orthographic lexical networks and how they develop over time in our virtual simulation of children's reading acquisition. Important descriptive statistics for the network measures in each grade are provided in Table 8.1. The developmental patterns for these measures are displayed in Figure 8.2. Lines represent the shape of the effect of grade on each measure.

The mean degree of the network's nodes $<k>$ increased with age, $F(6,343) = 164.8$, $p < 0.001$. Thus, the number of neighbors per node increases overall as expected. The results showed a mean number of orthographic neighbors between 6 and 7 across all age groups. Post hoc analyses showed a significant difference only between second and third grade, $p < 0.001$, and third and fourth grade, $p < 0.05$. The shape of the effect could be described with a quadratic function indicated by a significant linear trend= -0.005, $t = -7.66$, as well as a significant quadratic trend = -0.079, $t = 13.28$ (intercept = 6.69, $t = 637.77$).With an $R^2 = 0.70$, the fit of the function was sufficient, that is it adequately represents the data. The curve thus shows a deaccelerating trend with a strong increase of neighbors in the beginning which levels out later in development after fifth grade. That is, overall, the number of neighbors per node increases with network development, mostly between second and fourth grade where a lot of new neighbors are added to the existing network. From grade 5 onwards, the development is slower: although the number of neighbors still increases, the growth grade becomes smaller.

TABLE 8.1: Lexicon sizes and network measures in different age groups.

	M Lexicon Size		Network Measures *M (SD)*					Lexical hermits *M* (*SD*)	
Grade	Lemmas	Types	*n*	*<k>*	*L*	*D*	*C*	*n*	*proportion*
1	5,925	31,570	16,027	6.78	9.84	46.04	.49	15,543	49%
			(84)	(0.06)	(0.29)	(5.51)	(.02)	(84)	
2	6,097	32,606	16,580	6.79	9.85	45.44	.49	16,026	49%
			(86)	(0.06)	(0.30)	(5.37)	(.02)	86)	
3	11,182	46,757	24,155	6.90	9.99	49.22	.47	22,602	48%
			(87)	(0.05)	(0.28)	(6.45)	(.01)	(87)	
4	14,819	58,238	30,368	6.93	10.25	51.64	.45	27,870	48%
			(94)	(0.04)	(0.25)	(6.24)	(.01)	(94)	
5	18,812	71,344	37,479	6.95	10.52	51.30	.43	33,865	48%
			(115)	(0.04)	(0.19)	(4.92)	(.01)	(115)	
6	25,694	93,293	49,465	6.96	10.56	47.48	.41	43,828	47%
			(118)	(0.02)	(0.13)	(3.35)	(.00)	(118)	
8	38,029	130,675	70,123	6.98	10.51	45.18	.38	60,552	46%
			(109)	(0.02)	(0.08)	(3.86)	(.00)	(109)	

Note: Lexicon sizes were determined via a sampling procedure established by Segbers & Schroeder (2017). In this framework, we developed a vocabulary test (Trautwein & Schroeder, in press). We then determined the relation between vocabulary test results and total vocabulary size by drawing a virtual sample of different lexicon sizes from the childLex corpus and checking whether the items of the vocabulary test were included or not. We then adapted this relation to a sample of real children and determined their total vocabulary sizes for different grades.

The average path length *L* ranged between 9.84 and 10.56, that is between 9 and 11 paths were necessary to get from one node to another. The effect of grade was significant,

$F(6,343) = 96.92$, $p < 0.001$. Post hoc analyses revealed a significant increase between second and third grade, $p < 0.05$, between third and fourth grade, $p < 0.001$, and between fourth and fifth grade, $p < 0.001$. The effect of grade on the average path length could be described by a cubic function indicated by a significant linear trend = -0.016, $t = -7.08$, a significant quadratic trend = 0.18, $t = 6.69$ and a cubic trend = -0.44, $t = -4.54$ (intercept = 10.11, $t = 106.17$). The fit of the function was sufficient, $R^2 = 0.63$, indicating an adequate representation of the data. That is, in the beginning the development of the average path length is weak, from second to fifth grade it increases rapidly and afterward the curve levels out. Again, we observed a deaccelerating trend with larger differences in the early grades and smaller increases in later development. Overall, the growth of the average path length indicates that the interconnectivity of the network decreases with growth since more paths are necessary to get from one node to another. The loss of interconnectivity is thus strongest between second and fifth grade. Afterward, the average path length still increases but with a decreased growth rate. Overall, that is, although the number of neighbors increases, the interconnectivity of the orthographic network decreases.

The maximum path length *D*, that is the maximum number of paths between two nodes, ranged between 45 and 52. From first to second grade, the diameter decreased, then increased until fourth grade and decreased again up to eighth grade. The overall effect of grade was significant, $F(6,343) = 13.58$, $p < 0.001$. Post hoc analyses, however, showed a significant increase only between second and third grade, $p < 0.01$, and a decrease between fifth and sixth grade, $p < 0.01$. The shape of the effect could be displayed as a quartic polynomial (linear trend = 0.13, $t = 3.68$, quadratic trend = -2.17, $t = -3.92$, cubic trend = 12.03, $t = 3.95$, quartic trend = -23.41, $t = -3.57$, intercept = 59.47, $t = 13.35$) but the fit of the function was very low ($R^2 = 0.19$) in comparison to the fits of the other functions (all $R^2 > 0.6$). That is, the function did not sufficiently represent the data. Thus, concerning the diameter, we could not obtain a clear pattern of network development since a rising diameter means a loss of interconnectivity but a decreasing diameter means a higher degree of interconnectivity.

The clustering coefficient *C* ranged between 0.38 and 0.49, which indicates a high probability of the neighbors of a word to also be neighbors of each other. The effect of grade was significant, $F(6,343) = 631.8$, $p < 0.001$. While the difference between first and second grade was not significant, $p > 0.90$, C decreased from second grade onwards, all $p < 0.01$. The shape of the effect could easily be described as a quadratic function evident via a significant linear trend = -0.002, t = -8.43 and a significant quadratic trend =-0.007, t =4.205 (intercept = 0.51, t = 185.04). The fit statistic of the function was very high, R^2 = 0.99, indicating that the quadratic function adequately describes the data. The curve shows a small decrease in the beginning with no significant difference between first and second grade and a steady strong decline of the clustering coefficient onwards. We thus observe an accelerating trend with small differences in the beginning of lexical development and larger differences later in development.

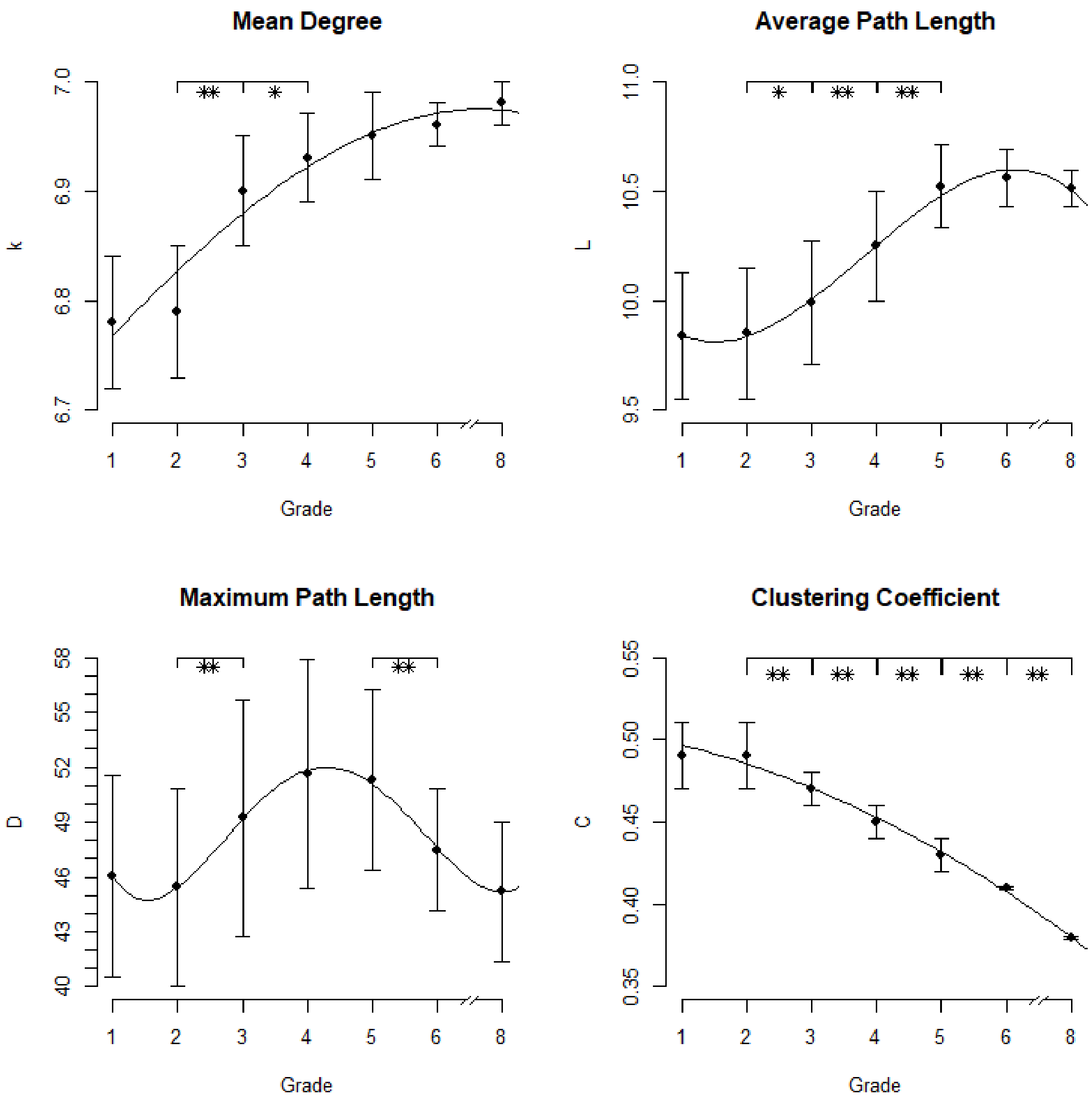

FIGURE 8.2: Means and standard deviations for the network measures in the different grades with overall effects depicted as lines. * = $p < .05$, ** = $p < .01$.

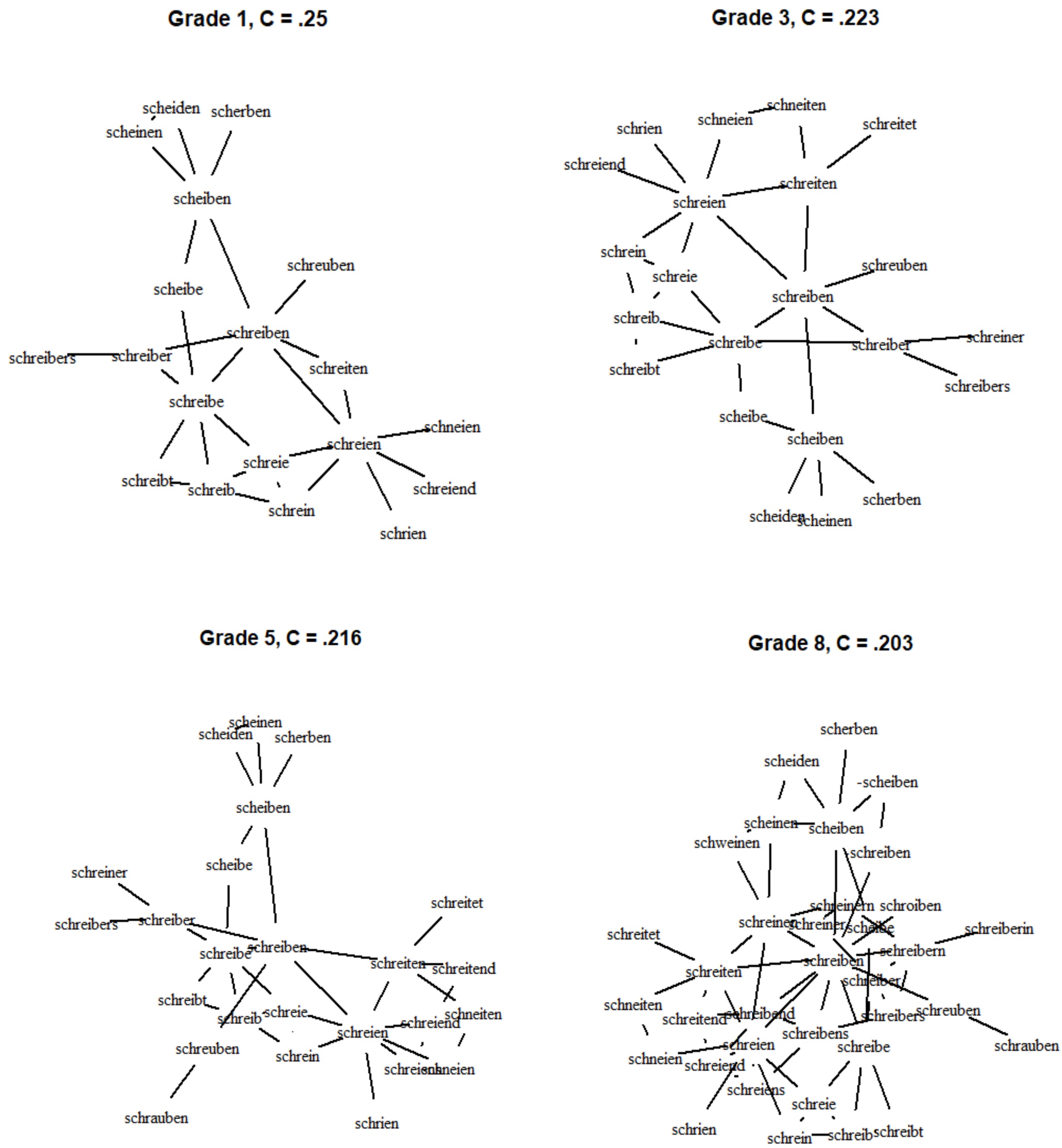

FIGURE 8.3. The network and according clustering coefficient for the word „schreiben“ – „to write“. Note that for reasons of comprehensibility only neighbors of a maximal Levenshtein distance of 2 are depicted.

To illustrate this, an excerpt of the growing network is depicted in Figure 8.3. It shows the network of the word "schreiben" ("to write") with neighbors of a maximum Levenshtein distance of 2. As evident, more neighbors are added but the clustering coefficient decreases. Overall, the decrease of the clustering coefficient indicates that with network growth, the probability of neighbors to also be neighbors decreases. Thus, with growth, the interconnectivity of the network declines, especially from grade 2 onwards.

The number and proportion of lexical hermits, that is words that do not relate to the lexical network, are also displayed in Table 8.1. While the total number of hermits increases with lexical growth, their proportion in relation to the total lexicon stays constantly high at almost 50% for all age groups. That is, a high number of words is not connected to the mental lexicon at all.

Another important question is whether the number of links between the nodes follows a power-law distribution similar to networks in other language domains. This is typically demonstrated by plotting k against its relative frequency P(*k*) on a log-log scale. If this relationship is linear, the network has a structure that follows a power-law distribution. The degree distribution for one exemplary virtual participant for each grade is displayed in Figure 8.4. As can be seen, the relationship between log *k* and log P(*k*) appears to be linear. A linear function provided a good fit of $R^2 > 0.85$ in for all virtual participants at each time point. That is, the degree distribution follows a constantly descending power-law function and the number of nodes with a lot of connections decreases.

To sum up, the network measures show highly interconnected orthographic networks in orthographic development. However, although the number of neighbors per node slightly increased with development, the interconnectivity of the network decreased with network growth. Apart from the diameter, this pattern could be obtained in all network measures. The proportion of hermits stays constant throughout development. Furthermore, there are a lot of nodes with a few connections and only a few nodes with a lot of connections as shown by the descending function of the degree distribution.

8.5 Discussion

In this study, we analyzed orthographic networks and their development from grade 1 to grade 8 in German. With a frequency-sensitive sampling approach, we simulated orthographic lexical development for 50 virtual participants from grade 1 to 8 and examined their networks' characteristics and their development by using graph theory. Findings indicate that the number of orthographic neighbors per word increases but the interconnectivity of the network decreases with network growth. Overall, the orthographic networks are dense, indicated by a high clustering coefficient as well as a high mean degree.

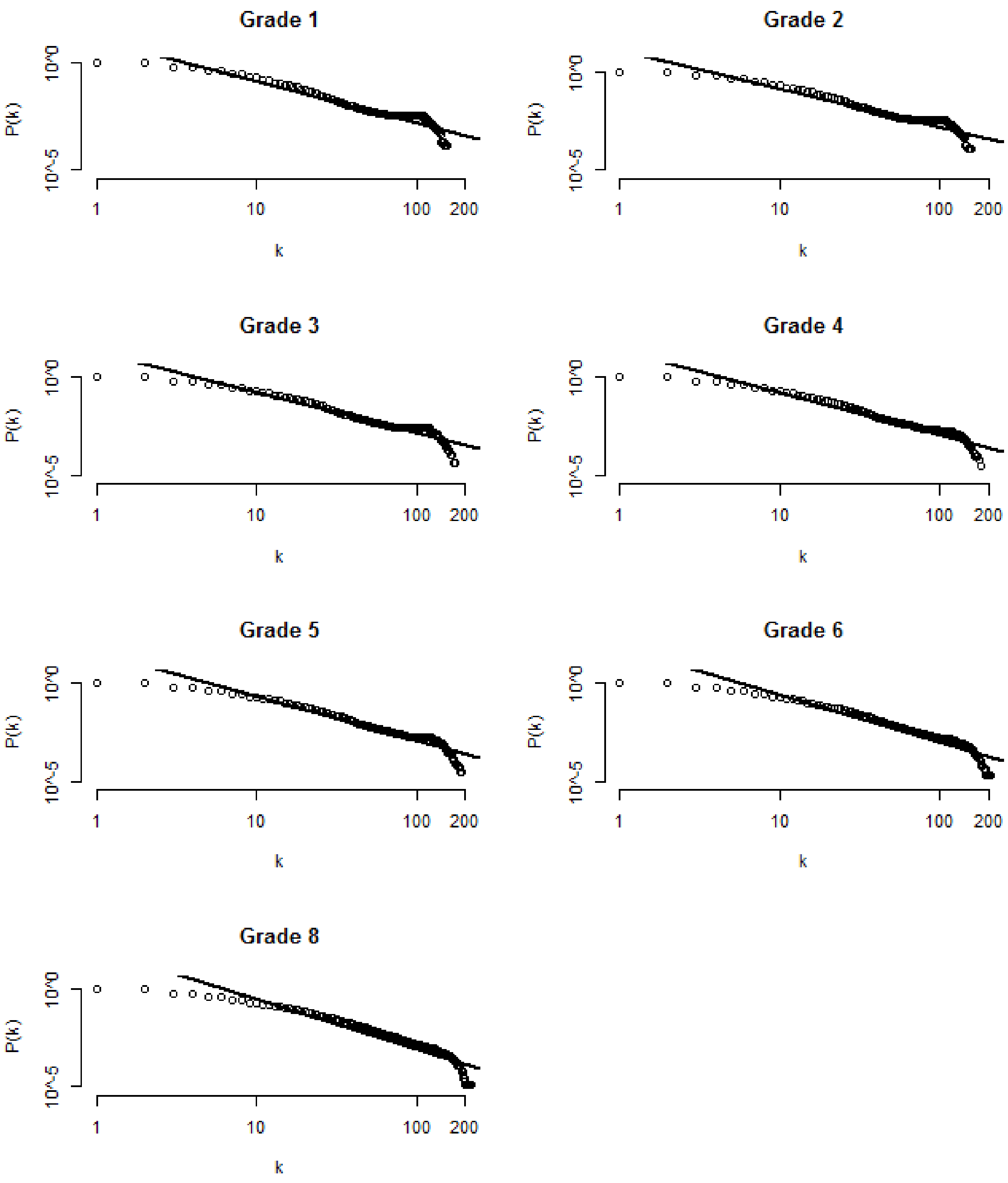

Figure 8.4. Log-log plot of the degree distribution for one exemplary virtual participant at each time points. Lines represent the fit of a linear function to the data.

8.5.1 Loss of Interconnectivity during Orthographic Learning

In particular, although the number of neighbors per word (mean degree) increases, the overall connectivity of the network decreases as shown by the average path length as well as the clustering coefficient. The interconnectivity of the network slowly decreases between first and second grade, then strongly declines until sixth grade and levels off onwards. That is, in the beginning networks contain a lot of words that are interconnected, thus neighbors of each other. With growing lexicon size, more neighbors are added to already known words but are not necessarily interconnected. This finding appears to be counterintuitive since one would assume a growth of interconnectivity with growing neighborhood sizes. However, regarding orthographic development and learning this makes sense: At the beginning of primary school, children know mostly words that are similar to one another. During development, less similar words are added gradually to the mental lexicon. That is, the mental lexicon becomes more and more differentiated with age.

In addition, the findings can also explain aspects of reading development: At the beginning, in a highly interconnected network, a lot of similar words compete with each other while reading. That is, activation spreads across the network and cannot be focused on the particular target word. It leads to a higher probability of mistakes as well as longer reading times. With growing orthographic lexicon size, the process becomes more refined since the competition between words is minimized due to fewer interconnections in the network. Activation can be centered to the target word, reading thus becomes faster and fewer mistakes are made. This underlines the theories of Metsala (1997) as well as Castles et al. (2007) who suggested a shift from a broader to a more refined process of lexical access. Our findings suggest that the lexicon itself becomes more sophisticated which leads to a more refined access mechanism.

8.5.2 Implications for the Effect of Neighborhood Size during Development

The change of network measures over time could also imply changes in the effect of neighborhood size and neighborhood frequency over time. In fact, the neighborhoods themselves change with development, the mean degree of our networks increased while the interconnectivity of the neighbors decreased. That is, fitting the findings of Castles et al. (1999), children have smaller neighborhoods than adults which might affect language processing of target words and should be considered in future studies on developmental changes of neighborhood effects.

However, as Chan and Vitevitch (2009, 2010) demonstrated, network measures above the traditional neighborhood sizes measure can lead to important findings on influences of lexical structure on language processing. Thus, e.g., the clustering coefficient should also be considered when examining developmental changes in neighborhood effects. While the growth of neighborhoods could lead to the assumption of an increase of neighborhood effects, the decrease of interconnectivity could imply a decline of neighborhood effects. Future research should address this question.

In this regard, one important feature of our simulation method is that network measures per grade can be derived, e.g., the mean degree or clustering coefficient for a certain word in a certain age group. Since the estimation of an individual's lexicon size is possible (see Segbers and Schroeder, 2017), even individual neighborhood sizes could be determined and used for the study of effects in language processing. It has also been shown that the frequency of the neighbors has a crucial moderating effect with high-frequency neighbors having inhibitory and low-frequency neighbors having facilitative effects (Sears et al., 1995; Pollatsek et al., 1999). In our network approach, this frequency information could be implemented using a weighted network approach in which the weights of the paths depend on frequency of the corresponding nodes. This could further help to predict effects of neighbors on lexical access.

8.5.3 Small World Characteristics in Orthographic Networks

Since we observed a relatively short average path length and a high clustering coefficient, we assume that orthographic networks possess small world characteristics as defined by Watts and Strogatz (1998). As evident from the degree distribution, we also found a scale-free organization of the networks. That is, the orthographic lexicon is structured like semantic and phonological networks with a small number of well-connected nodes which could be denominated as key entries (Borgatti, 2006; Vitevitch and Goldstein, 2014). This fits the findings of several authors who have discussed the existence of a core lexicon which contains words with a lot of connections that are important for communication (e.g., Ferrer et al., 2001; Siew, 2013; Stella, Beckage, Brede & De Domenico, 2018). Identifying such key players and core lexicons could be promising for teaching strategies since our simulations suggest that they play an important role in language learning.

Throughout all grades, the number of words without orthographic neighbors called "lexical hermits" is constantly high (almost 50%). That is, these words do not connect to the mental lexicon via orthographic similarity. The high proportion reflects the distribution of orthographic neighbors in the German language itself: While because of its morphological richness and orthographic transparency a part of words in German possesses a lot of neighbors, a high proportion of words does not have any orthographic neighbors at all (in childLex, 62.8% of all types do not have neighbors, Schroeder et al., 2015a). One possible explanation for this in terms of lexical development is that they are learned and integrated via one of the other language domains. Semantic neighbors for example may also play an important role in orthographic learning since words that are semantically connected often appear in the same contexts in written language. Burgess and Lund (2000) have proposed a measure of semantic neighborhood provided by co-occurrence in texts (see also Durda et al., 2006). Analyzing the present data with such a measure of semantic neighborhood could shed more light on the development of the lexical hermits.

Compared to the semantic network described by Steyvers and Tenenbaum (2005) we observed higher values for average and maximum path length as well as for the clustering coefficient and the mean degree in all age groups. This might be due to the fact that the orthographic networks we examined were generally larger than the semantic networks analyzed by Steyvers and Tenenbaum (2005). Furthermore, the results reflect the difference in the definition of neighborhoods: In our study, orthographic neighbors were defined as words with a Levenshtein distance of 1. That is, the probability of a word's neighbors to also be neighbors is very high in the orthographic domain (e.g., the neighbors hat and fat for the word cat are also neighbors). In the semantic domain, however, neighborhoods are defined by the word's meaning and are thus more restricted (e.g., although the words dog and mouse are both semantic neighbors of the word cat, they are not necessarily also neighbors of each other). This also results in a higher degree for orthographic then for semantic networks.

In comparison to Vitevitch's (2008) phonological network analysis we found higher average path lengths as well as a higher clustering coefficient for all age groups. However, our networks also comprise more nodes than Vitevitch's analyses. Furthermore, we analyzed the network on a type level which increases the number of neighbors of a word. In addition, differences between languages (English by Vitevitch, German in our study) might also have influenced the results. German is a morphologically rich language (Fleischer et al., 2012) and has a sophisticated inflectional system. This might increase the number of orthographic neighbors. This also holds for the results on the lemma level, since even after lemmatization, German is morphologically rich because of derivation and compounds. A study explicitly comparing different languages – similar to Arbesman et al. (2010) for phonological networks – would be able to address this issue. In general, because of differences in network size and network quality, the comparison of our results to earlier studies on semantic networks (e.g., Steyvers and Tenenbaum, 2005) as well as phonological networks (e.g., Vitevitch, 2008) is not straight forward. For future research, the analysis of networks in different language domains with the same size and quality could lead to more comparable results and findings. Especially the comparison between phonological and orthographic networks could be interesting since both domains share a lot of characteristics in a transparent language such as German.

8.5.4 Limitations and Future Prospects

Although our results provide an important contribution to the study of orthographic learning, they also have important limitations. Maybe the most important caveat of our study is that the presented findings are not based on real empirical data but on simulations of children's orthographic development using a corpus sampling approach. Relying on simulations methods is necessary for the analysis of orthographic networks because most of children's orthographic development takes place after they have entered school and children's vocabularies are already quite extensive (over 5,000 lemmas or 30,000 types according to our approximations). In contrast to studies focusing on language development in infants (e.g., Hills et al., 2009), it is thus not feasible anymore to collect data for every potential word in the mental lexicon.

The crucial question is whether it is likely that our results will generalize to children's real orthographic development. Of course, the answer to this question depends on how well the orthographic learning mechanism in the real world is approximated by our sampling procedure and its underlying assumptions. In this context, several points have to be discussed. First, a fundamental assumption of our simulations is that children's orthographic learning is influenced by probability that a word is encountered in the learning environment, i.e., word frequency. This assumption seems to be plausible given that word frequency is also a major determinant in children's earlier lexical development (see, e.g., Goodman et al., 2008) and age-of-acquisition norms usually correlate highly with word frequency (e.g., Kuperman et al., 2012). To confirm this, we collected age-of-acquisition data for 1152 words in the childLex corpus (Schröter and Schroeder, 2017) and also found a high correlation between log type frequency and age-of-acquisition norms provided by adults, $r = -0.51$, $p < 0.001$. This confirms that written word frequency is indeed a major factor affecting the time when a word is acquired by children. Naturally, high frequency words are short and possess a lot of neighbors due to the principle of economy in linguistics, that is the aspiration to transport as many information as possible with the smallest effort possible (Vicentini, 2003). Thus, frequency and orthographic neighborhood size highly correlate ($r = 0.27$ in childLex, see also: Landauer and Streeter, 1973; Frauenfelder et al., 1993; Siew, 2013). That is, in our modeled learning process words with a lot of neighbors are also learned first because they also have a higher frequency. The mechanism thus fits other theories on language learning which have shown that words from dense neighborhoods are learned earlier (Storkel, 2004; Hills et al., 2009).

However, there are clearly other factors that influence word learning such as cognitive development, education and personal experience. In addition, other linguistic characteristics such as phonological similarity or semantic concreteness might also influence orthographic learning (Kyte and Johnson, 2006; Ouellette, 2010). These factors should be taken into account in future studies. Another assumption of the reported simulations is that the childLex corpus that served as the basis for the sampling is a realistic approximation of children's real print-related learning environment. The childLex corpus is quite extensive compared to other corpora for children and comprises approximately 5 times as many words as an average child is likely to read in grades 1–6. As any corpus, however, it is just a sample from the population of the many books that children potentially can read. During the assembly of the corpus, we took great care to include books that are actually read by the children and based the selection on library loan statistics, teacher ratings, and children's self-reports (see Schroeder et al., 2015a, for a description of the corpus selection). We are thus confident, that the corpus is an ecologically valid sample for children's print-related learning environment in German at the beginning of the twenty-first century.

Relatedly, we implicitly assumed that the learning environment stays constant during children's orthographic learning because we used the same corpus for all grades. This assumption is certainly a simplification, because young children beginning to read are likely to read different books than older children who have different interests and better reading skills. One way to refine our sampling procedure is thus to adapt the learning environment by sampling from different subcorpora for different age groups. However, at least in a transparent orthography such as German, in which most children are able to read rather fluently at the end of grade 1, books for younger and older children are actually not that different in terms of their linguistic characteristics (see Schroeder et al., 2015b, for a summary). Most differences are related to the lexical level, i.e., books for older children have a more varied vocabulary and introduce additional expressions for the same entities. This shift in lexical diversity, however, is taken into account by the frequency-sensitive sampling mechanism. That is, even with the sampling from the subcorpora, we would expect the same pattern of results reported in this study.

Finally, we assumed that our virtual participants did not differ in their size of their orthographic networks and all showed the same rate of orthographic growth. Thus, the size of their initial mental lexicon and the number of types that are acquired in each grade were fixed to the average lexicon size and growth rate that has been reported for German (Segbers and Schroeder, 2017). However, it is well known that there are large interindividual differences in children's print exposure and orthographic development (Stanovich, 2009; Pfost et al., 2014; Schroeder et al., 2015b). It is thus very likely that the size of children's orthographic networks will show great variability and grow with differential rates. Just simulating orthographic development for average readers is clearly only a starting point for future investigations. An important finding from our study is, however, the mechanisms underlying children's orthographic network development are remarkably stable and did not differ between children in grades 1 and 8. It is thus rather unlikely that the qualitative nature of children's orthographic network growth is strongly influenced by mere quantitative aspects of their lexicons. With our findings, we thus provided average numbers on orthographic lexical development for different age groups.

However, the next step for future research needs to be the modeling of individual network growth for certain children at different age points over time. Longitudinal data on lexical growth would be necessary to perform these analyses. In addition, in combination with the measurement of the effect of orthographic neighborhood sizes at these different time points these findings could lead to a more sophisticated understanding of orthographic development during primary school.

One further application of our study addresses recommendations for the content of children's books or texts for children in different age groups. As implicated by our study, young children tend to know a lot of similar words while networks for older children are more differentiated and less connected. Including these known words (and their neighbors) into texts for different age groups could ease the reading process as well as orthographic learning. Clearly, more research on the effect of known neighbors on single word reading is necessary to address this issue.

8.5.5 CONCLUSIONS

To sum up, this study reports data from a simulation study analyzing the development of the orthographic lexicon using graph theory. Our results demonstrate that orthographic networks exhibit small-world characteristics similar to phonological or semantic networks. In addition, we found that the interconnectivity of the network decreases with growth while the neighborhood size itself increases. The results support theories of reading development which claim a shift from a broader to a more fine-grained reading process. Moreover, by showing that the network characteristics and thus neighborhoods change with age, developmental differences in language processing could be explained with our results.

Analyzing orthographic networks using graph theory is thus a promising approach for further research on (individual) language development. In addition, the presented method enables the derivation of age-specific or individual network characteristics which in turn can be used for studies of language processing, namely the effect of network measures on word retrieval.

III GENERAL DISCUSSION

9 Summary of Results

The present dissertation examined lexical development from primary school onwards by applying a newly developed method to estimate the size and content of an individual's mental lexicon. The mental lexicon plays a central role in language development and language processing since it comprises all words a language learner knows. Lexical entries contain different kinds of information in the language domains and are interconnected due to similarities in these domains. However, up to now, little was known about lexical development after school entry because with increasing size, the assessment of the total mental lexicon becomes more and more challenging. The central goals of the dissertation were thus to develop a procedure that leads to substantiated results on size and content of the individual mental lexicon and to describe the results, more precisely the development of lexicon size and lexical structure for German primary school children. To this end, three studies were conducted. In Study 1, I focused on the establishment of a new vocabulary test, which I then used in Study 2 to apply a corpus-based sampling procedure for lexicon size estimation. Based on these estimates, in Study 3, I concentrated on the investigation of lexical structure and its development in terms of orthographic networks in the developing mental lexicon.

In Study 1, I developed a yes/no vocabulary test to assess vocabulary for primary school children. Three versions for different age groups (grade 1/2, grade 3/4, grade 5/6) were conducted based on word frequencies derived from the childLex corpus. Each test version comprised 100 words as well as 24 nonwords to prevent guessing. The evaluation of the test's validity using Item Response Theory as well as the validity framework proposed by Messick (1995) appeared to be fruitful and showed that the test successfully measures vocabulary knowledge in the participant group. The newly developed instrument could be useful for other scientific contexts since the measurement of vocabulary from primary school onwards is challenging and only few practical instruments exist. For this dissertation, it represents the basic method for the following studies and the applied sampling approach.

In Study 2, I used the test results from the yes/no vocabulary test to estimate individual total lexicon sizes by applying a corpus sampling approach. Test performance of virtual participants with different lexicon sizes was simulated by drawing them from the corpus. This enabled to determine the function between lexicon size and test performance. In a final step, test performance of real participants was projected on the results of the virtual sample to determine individual lexicon sizes. It resulted in plausible values for the absolute number of known words in different age groups with a vocabulary development from about 6000 words in first grade to about 73,000 words for young adults and large interindividual differences. Moreover, the development could be described as a quadratic function and the acquisition of different parts of speech and morphological categories were reported. They indicated that vocabulary development is strongly driven by the acquisition of nouns and the increase of morphological complex words. The study thus provides an innovative approach to measure lexicon size and contributes to the limited findings on lexical development during the crucial period from primary school to adulthood.

In STUDY 3, I regarded lexical structure and its development from a network perspective with words represented by nodes and connections based on orthographic neighborhoods. Lexical networks were determined for average language learners via the sampling approach and based on the average numbers of lexicon size derived in Study 2. Graph theory was applied to determine several network characteristics such as average and maximum path length and clustering coefficient to evaluate the network's interconnectivity. Results showed that orthographic networks in the mental lexicon possess small-world characteristics similar to other language domains, more precisely a small average path length and a high clustering coefficient, as well as a scale-free structure with a few nodes with many connections and many nodes with a few connections. Furthermore, the study revealed that network interconnectivity decreased with growing lexicon size while the average number of neighbors per word increased.

Figure 9.1 summarizes the main results of the three studies. Taken together, they provide an extensive insight into lexical acquisition from primary school onwards in terms of lexicon size and lexical structure. By presenting a novel and sophisticated approach to estimate an individual's lexicon size and lexical structure, the thesis thus advances the investigation of suitable methods for lexical assessment. In addition, it sheds light on the course of lexical development in late childhood and emphasizes the importance of this period for language acquisition. Below, I will discuss the contribution and implications of this dissertation to the fields of lexical assessment and lexical development along with limitations and prospects for future studies.

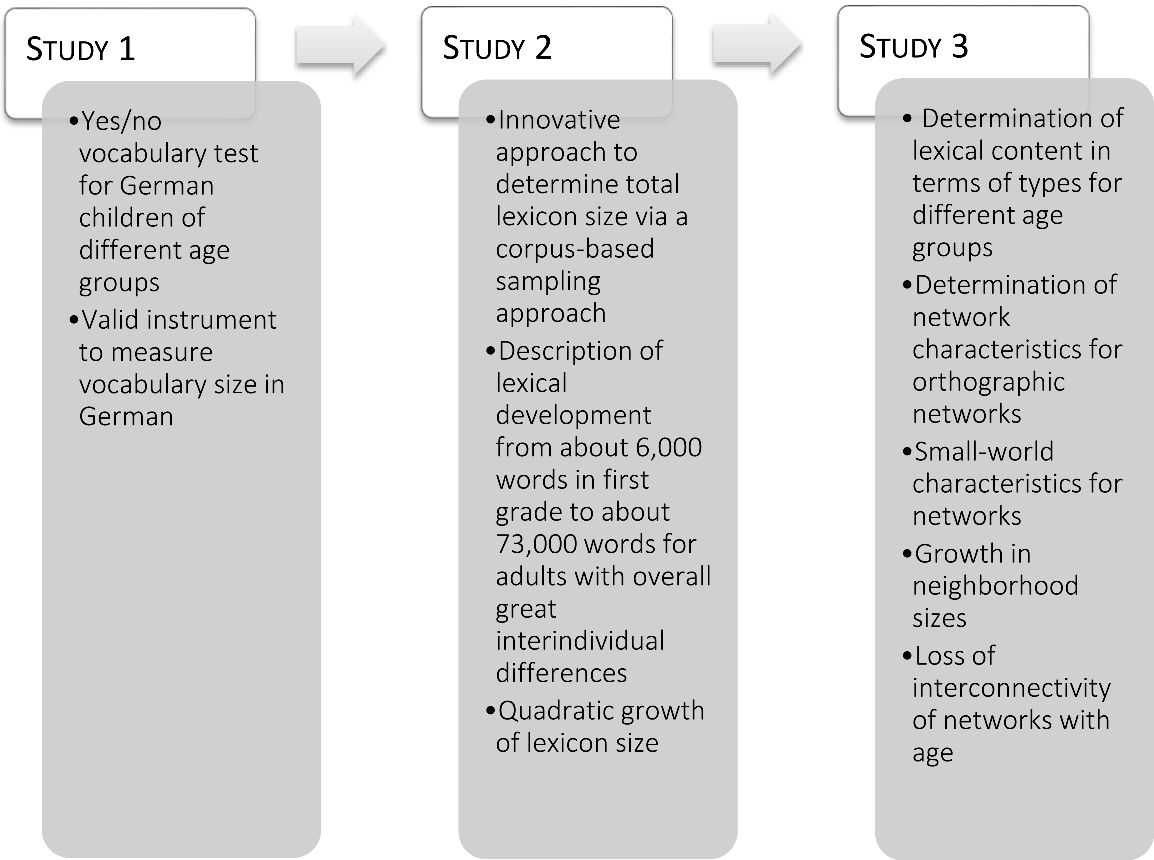

Figure 9.1: Summarized results of the three studies.

10 Methods of Lexical Assessment

The central reason for the lack of findings on lexical development from school entry onwards was the absence of valid methods to assess the mental lexicon with growing size, i.e. after the simple counting or testing of all known words is not possible anymore. The first essential aim of this dissertation was thus to develop a new method to measure the individual mental lexicon from primary school to adulthood (first research aim). The presented approach includes the combination of a yes/no vocabulary test (Study 1) and a corpus-based sampling procedure (Study 2) as well as graph theory (Study 3) and lead to plausible values on individual lexicon size (Study 2) as well as simulated data on lexical content and structure (Study 3). In the following, I discuss benefits of the test format, the sampling procedure and the network approach as well as their combination along with limitations and prospects for future research.

10.1 THE YES/NO VOCABULARY TEST

One general question in vocabulary assessment is when do we count a word as known? In other words – how do we test whether a person knows a word? As elaborated in the theoretical framework, entries in the mental lexicon comprise different aspects of word knowledge, mainly on phonology, semantics and later orthography (Perfetti & Hart, 2002; see section 2.1). According to this framework, we may only know the pronunciation of a word without being able to write it down or we may have only a vague idea of a word's meaning although we clearly know how to spell it. What kind of knowledge do we measure in vocabulary assessment then?

For my approach to measure vocabulary, I chose the yes/no test format introduced by Anderson and Freebody (1983). It bears several advantages: it can be administered within a group setting and performance as well as the evaluation of results is feasible without high effort. In addition, it is comparable to parental checklists, which are commonly applied for the assessment of young children's vocabulary size. Moreover, it includes a high number of items and is less reliant on the alternative answers as in multiple-choice tasks (Anderson & Freebody, 1983). However, the test format clearly assesses the orthographic representation of lexical entries since it is administered as a reading version. Moreover, it is a measure of receptive vocabulary and does not consider productive vocabulary. Although these orthographic representations are an important part of lexical entries, particularly for reading competence, this is also a central limitation of the presented approach, especially concerning beginner readers. For them, a listening version of the test could be more suitable since problems with reading might lead to an underestimation of vocabulary size (see also McLean, Kramer & Beglar, 2015). In general, it would be interesting to compare results of a reading and listening version of the test for the same participants to investigate whether the number of phonological and orthographic entries in the mental lexicon are comparable.

In addition, one could argue whether knowing the orthographic form of a word is sufficient to count the word as known. Miller (1999) argues that a person must know a word's meaning to count it as stored in vocabulary. He also notes that there are different levels of knowledge of meaning between not knowing at all and completely knowing a word. However, in line with previous studies (Mochida & Harrington, 2006; Anderson & Freebody, 1983), I found a high correlation between knowledge of meaning assessed via multiple-choice questions and the yes/no vocabulary test in Study 1. This indicates that the measure taps into vocabulary knowledge in terms of semantics. In addition, the pilot study of oral definitions reported in Study 1 showed a high correlation between word definitions and test results (*r* = .69). This result indicates that the yes/no format indeed taps into the knowledge of word meaning. However, conducting further test formats based on the childLex corpus such as a multiple-choice task or word definitions and projecting the results on the overall corpus using the sampling approach presented here could lead to further insight on the interplay of the different language domains in the mental lexicon.

One important feature of the yes/no vocabulary test I presented is its reliance on the childLex corpus. It is not only necessary for the sampling approach building on it but is also the first vocabulary test in German relying on a language corpus for the specific age group. Most corpora focus on the language of adults (e.g. CELEX, Baayen, Piepenbrock & Gulikers, 1995; DWDS, Geyken, 2007). However, their characteristics are partly applicable for children since the comparison between adult and children corpora show important differences (see Schroeder et al., 2015b). Besides its application for the estimation of total lexicon size, the yes/no test is thus a promising instrument for vocabulary assessment of primary school children who were clearly underrepresented in methods of vocabulary measurement before. In addition, the application of the Rasch Model as proposed by Beglar (2010) and the use of the validation framework presented by Messick (1995) have been shown to be beneficial for the validation of the test format and can also be implemented into future studies on vocabulary test validation.

In summary, the newly developed yes/no test represents a valid instrument to measure the orthographic receptive vocabulary of primary school children. Within this thesis, it serves as the basis for the applied sampling approach. Although the test format holds several advantages, a comparison with results of other formats to be implemented in the sampling approach could be beneficial for future research.

10.2 The Sampling Approach

Based on the vocabulary test, I developed a sampling approach to estimate total vocabulary sizes and contents for different individuals. The sampling approach is grounded on the dictionary method used in the past to estimate total lexicon sizes (e.g. Anglin, Miller& Wakefield, 1993; Nation, 1993a). By applying recent technological methods (the childLex corpus, Item Response Theory), it can be regarded as an advancement of the dictionary approach. The dictionary method had several difficulties in that the number of entries was vague and the selection of items to be tested (e.g. every nth item on every nth page) was often biased towards items that occupy a lot of space in the dictionary (Nation, 1993a). The sampling approach presented here overcomes these difficulties by using the childLex corpus as the basis instead of a dictionary. It is clearly known how many words it comprises and the selection of items to be tested was based on word frequency, which is known to approximate item difficulty (Tamayo, 1987).

A central unique feature of the presented approach is the establishment of a virtual sample of participants with different simulated lexicon sizes. It takes word frequency as a proxy for word learning into account by using it to establish different virtual lexicon sizes. Moreover, it enables the simulation of test performance in dependence on lexicon size and uses it to project test results on the overall corpus. In traditional dictionary methods, test results were projected via percentages, for example if 50% of the tested words were known it was assumed that 50% of the whole dictionary was known (Seashore & Eckerson, 1940). Although frequency levels were also taken into account in some previous studies (e.g. Goulden, Nation & Read, 1990), the present approach offers an enhanced account by considering the actual word learning process driven by word frequency (e.g. Naigles & Hoff-Ginsberg, 1998) through the sampling of different lexicon sizes and the simulation of the according test performance. It is thus an innovative approach to the question of how results from a vocabulary test can be projected to the whole dictionary or corpus.

Compared to usage-based methods of vocabulary size (e.g. Pregel & Rickheit, 1987), the approach is less costly and leads to individual results per person. Furthermore, the approach can be adapted to other corpora (e.g. for adults or other languages) as well as for other language domains (e.g. phonology or semantics) and entities (e.g. types, nouns, verbs etc.). Generally, since we cannot count the actual number of known words neither which words are known, it is important to remark that the approach generates only estimates which are supposed to approximate the actual mental lexicon.

Along with this, the approach also bears some important limitations. The first one addresses the dependency of the approach to the corpus, in this case the childLex corpus.

Although the quality of the corpus appears to be sufficient, the use of another corpus or for another language domain, e.g. for spoken language, could lead to different results. That is, the procedure mainly addresses receptive orthographic vocabulary and can only partly be projected on productive vocabulary or other language domains.

Furthermore, I used the childLex corpus to estimate the mental lexicon for children and adults although a corpus for adult language might be more suitable here (for a comparison of childLex with an adult corpus see Schroeder et al., 2015b). Combining different corpora to estimate lexicons for different age groups is thus one important challenge for future research. A further limitation of the procedure is the simulation of the learning process based on word frequency. The sampling of words from the corpus for the virtual sample was only frequency-sensitive. Although it can be assumed that word frequency is a main determinant for language learning (Hoff & Naigles, 2002; Huttenlocher et al., 1991), there are other factors that drive lexical development as well, such as cognitive processing, education and personal experience (e.g. Rowe, Jacobson & Van der Oord, 1999; Hawkins & Bender, 2002). These factors were not considered in the present sampling approach but could be implemented in prospective improvements of the procedure.

One important benefit for future research is the possible application of the sampling approach to other contexts. For example, it is commonly assumed that receptive vocabulary exceeds productive vocabulary in early childhood (Kauschke & Hofmeister, 2002). The comparison of both lexicon sizes by using a receptive measure of vocabulary as presented in this thesis as well as a productive measure of vocabulary (e.g. picture naming, see Dunn & Dunn, 2007) and projecting both on the whole corpus for different age groups would shed light on the question whether this difference persists throughout development.

In addition, the application of the procedure could allow the assessment of vocabulary size and structure for other languages. Since languages differ, for example in their morphological complexity and inflectional systems (Miestamo, 2017), it would be interesting to investigate whether vocabulary sizes and networks vary among speakers of different languages. Another field of research where the assessment of the mental lexicon plays a central role is second language learning. For teaching in a second language, it is important to know the students' language level and to plan and adapt content and instructions of teaching accordingly. In addition, teaching methods can be evaluated by pre- and post-tests of language proficiency (Nation, 2012). Applying the presented sampling approach to investigate the size and structure of the mental lexicon of second language learners could thus be one promising advancement of the introduced procedure.

A further promising adaptation of the procedure is the development of the mental lexicon during later adulthood. While the present studies show a strong increase of word acquisition up to young adulthood, it could be assumed that lexical entries diminish with cognitive decline (Burke & Shafto, 2004) which would result in a decreasing lexicon size and possibly a decline of connections between entries. Applying the sampling approach under the use of an adult language corpus and comparing different age groups in late adulthood could inform theories of language and cognitive shifts with age.

Moreover, in clinical research, especially in clinical linguistics, the adaptation of the method for patients suffering from language difficulties (e.g. specific language impairment in children, see McGregor, Oleson, Bahnsen & Duff, 2013; aphasia or dementia for adults, see Kemper, Greiner, Marquis, Prenovost & Mitzner, 2001) could update theories on the causes of language impairment in relation to the size and structure of the mental lexicon and could additionally be used for screening or diagnostic purposes.

To sum up, the sampling approach not only provides substantial results on lexical development for the group of interest in this thesis but also is adaptable for other fields of research and important research questions.

10.3 The Network Approach

The network perspective introduced in Study 3 is one example of the application of the sampling approach to tap into the content and structure of the mental lexicon. Although graph theory has already been used to analyze lexical structure for adults and very young children in other language domains (Vitevitch, 2008; Steyvers & Tenenbaum, 2005; Hills et al., 2009), its application for orthographic networks under a developmental perspective in late childhood is a novel approach introduced in the present thesis (see also Siew, 2018 for English). Study 3 thus proposes a way to apply graph theory to assess the development of lexical structure in late childhood. Clearly, for this idea the estimated size and content of the mental lexicon at different time points derived from the sampling approach from Study 2 were necessary. The study used simulated data on average lexical networks and thus reported results for the average language learner. It thus gives a first insight on how the amount of words stored in the mental lexicon reported in Study 2 might be organized in terms of connections between words based on orthographic similarities. Again, it is important to note that the content of the lexicons derived from the sampling approach relies on probabilities of words to be known based on word frequency. The determined networks are thus approximates of the actual lexical content which cannot be identified as a whole.

One important question in this context is how connections within the mental lexicon are established and whether these persist throughout development. As elaborated in chapter 2, we assume connections between words based on studies of language processing. For example, when a neighbor prime (a semantic, phonological or orthographic neighbor) affects word retrieval of a target word in a priming experiment, we consider a connection between the two words within the mental lexicon (e.g. Zielger, Muneaux & Grainger, 2003; Buchanan, Westbury & Burgess, 2001). In the network approach presented in Study 3, all neighborhoods were thus defined via the Levenshtein distance of 1 meaning that words are considered as neighbors when they differ in one letter (Levenshtein, 1966). However, there is a lack of studies on neighborhood effects for children, especially concerning orthographic neighbors. In conclusion, we cannot be sure whether connections between similar words already arise in childhood and persist throughout language development. The establishment of orthographic connections might be related to reading proficiency. As I concluded in Study 3, changes in the neighborhoods of the mental lexicon might result in a shift from a more broadly to a more fine-tuned reading process as suggested by Castles et al. (2007). Additional evidence comes from Yap, Balota, Sibley and Ratcliff (2012), who have shown that adults with high vocabulary knowledge were less sensitive to orthographic neighborhood effects than adults with low rates of vocabulary knowledge. However, the relation might also be reciprocal: A more fine-tuned retrieval mechanism could result in more and more specific connections within the mental lexicon. Studies on the developmental changes of neighborhood effects in word recognition and their causes would be beneficial to tap into this line of research in order to examine whether the definition of neighborhoods used here is applicable for children in the same way as for adults.

Study 3 analyzed orthographic networks since the yes/no vocabulary test mainly addresses the receptive orthographic vocabulary. It was conducted on a type level because orthographic neighborhoods are traditionally derived from similarities between types whereas the total lexicon sizes reported in Study 2 are determined on the level of lemmas. The sampling approach, however, allows the application of any other entity as well as other language domains. As a consequence, the developmental network perspective can also be employed for the development of semantic or phonological neighborhoods, on a lemma or word stem level. In addition, it can be used to analyze growth processes to improve the understanding of developmental trajectories addressing the question of how connections in the mental lexicon arise and develop (see for example Hills et al., 2009, see also section 11).

A further future line of research includes the individual measure of network development. While the study concentrated on simulation data and thus described the average development of lexical network structure, the method could also be used to examine an individual's orthographic network. For this purpose, the actual individual lexicon size as estimated in Study 2 could be used to approximate lexical content for single participants. Results could provide network characteristics for the individual mental lexicon. These could be used to describe actual growth processes for orthographic networks, similar to those of semantic networks described by Hills et al. (2009), as well as the effect of individual network measures on language processing (see section 11.2). Figure 11.2 (section 11.2) illustrates how the network measures can be derived for different individuals with different lexicon sizes. A further way to refine the network approach on an individual level is to not only take lexicon size into account for network determination. Other factors such as the underlying growth processes, growth rates or language skills could be considered in future applications of the measure. Taken together, along with the sampling approach, the network approach did not only lead to substantial results on network development for the present research question but enables the application for other future lines of research.

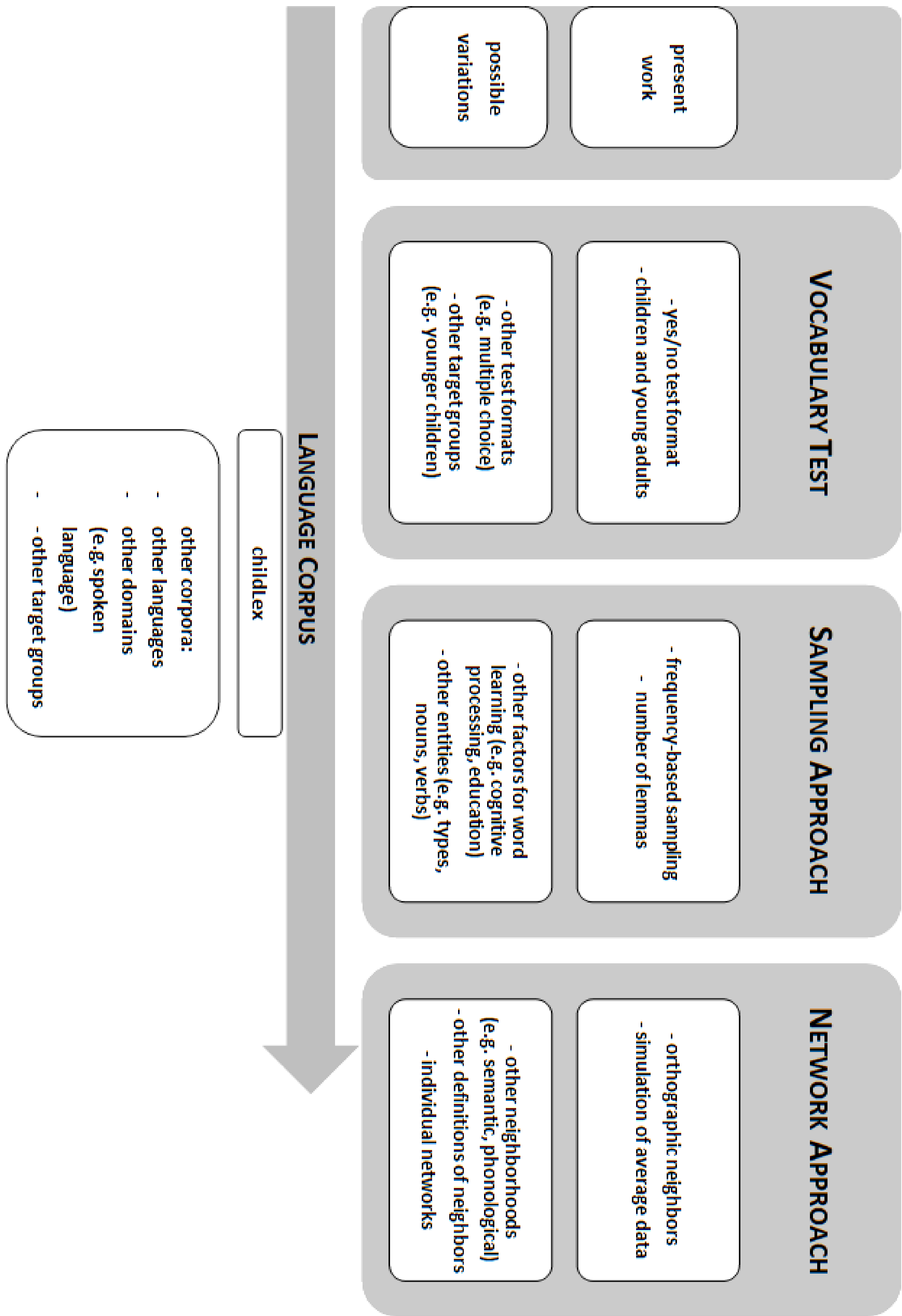

FIGURE 10.1. Combination of methods in the present work and possible variations for future applications.

10.4 The Combination of Methods and Further Applications

One unique feature of this dissertation is the combination of the particular methods discussed above – the yes/no vocabulary test, the sampling approach and the network approach – to provide an extensive procedure for lexical assessment.

While previous studies mostly addressed one lexical aspect, the present work considers lexicon size and structure and thus enables the relation of findings for both lexical features. The use of the childLex corpus as a common basis supports this approach.

Figure 10.1 summarizes the consecutive methods used in the present work. Moreover, it shows that the combination of the methods is flexible with different possible variations for each method, as discussed in sections 10.1, 10.2 and 10.3.

Besides these possible variations, there are further applications of the approach, especially regarding the amount of language exposure. Through an extension of the sampling procedure, the according amount of words encountered can be calculated. By using a vector which contains all words exactly as many times as they occur in the corpus, we can simulate the word learning process by sampling different amounts of words encountered from the corpus. We can then determine the according lexicon size by counting how many different words have been encountered. For example, if someone has encountered 10,000 words, we take a 10,000-word sample from the vector. Subsequently, we can count the number of different words in the 10,000-word sample which stands for the according lexicon size[9]. With this procedure, we can determine the relation between lexicon size and amount of words encountered which is depicted in Figure 10.2.

With this relation, it is thus possible to calculate how many words have been encountered by an individual or on average for a certain age group. For example, regarding the participants of Study 1, children in first grade have encountered about 240,000 words, children in fourth grade about 1,000,000 words and young adults about 14,000,000 words.

Knowing how many words have been encountered leads to possible calculations on which words have been encountered, and more importantly, how many times they have

[9] It is necessary to base this calculation on an assumption on how many word encounters are needed to store a word in the mental lexicon. In the presented analysis, this assumption was set to three encounters based on previous research findings (e.g. Jenkins, Stein & Wysocki, 1984). However, other assumptions could also be possible in the application of the approach.

been encountered. For this purpose, one can count the occurrence of each word in the sample of words encountered and thus receives the frequency with which the word has been encountered. Figure 10.3 A schematically illustrates the application of the approach for language exposure and individual frequencies. Figure 10.3 B shows an example of the application with letters (X, Y, Z) standing for different words. In this framework, a person who has encountered 10 words has a lexicon size of 3 (= different words occurring in the encountered words) and individual frequencies for these three words can be determined: word X has been encountered three times, word Y likewise and word Z four times.

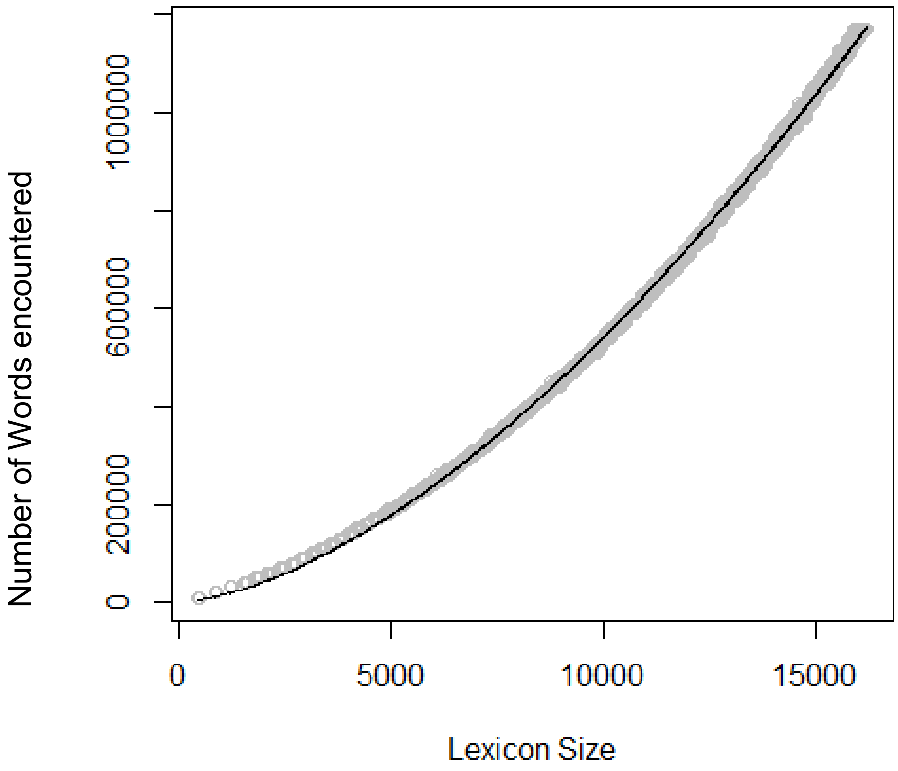

Figure 10.2: Relation between lexicon size and number of encountered words.

Why would it be important to know how many words have been encountered and how many times each word has been experienced? Studies of reading development and reading competence particularly show that language or print exposure affects reading skills (for a review see Mol & Bus, 2011). In this context, print exposure is traditionally assessed via author or book title checklists (e.g. Stanovich & West, 1989). Participants who know a lot of authors or book titles are assumed to have encountered more texts and thus have a higher degree of print exposure. With the application of the sampling approach presented here, the determination of absolute numbers of language or print exposure would be possible. This new measure could support the understanding of the relation of print exposure and reading competence.

The measure of individual frequencies suggested here could shed more light on the influence of frequency on language processing. The effect of word frequency on word recognition has been reported by many studies. It shows that high frequent words are processed faster and more accurate than low frequent words (see Monsell, 1991, for a review). It is assumed that high frequent words have better representations in the mental lexicon through more exposures and thus can be retrieved easier than low frequent words. Moreover, the effect has been shown to interact with language skills: High skilled individuals show weaker effects of word frequency than low skilled individuals (e.g. Chateau & Jared, 2000). This phenomenon is interpreted in that high skilled individuals have more language exposure and thus more individual encounters or higher individual frequencies for words (Diependaele, Lemhöfer & Brysbart, 2013). Up to now, these individual frequencies are assessed via subjective frequency ratings where a large group of participants judges how many times they have encountered certain words. Results are subsequently averaged across participants (e.g. Kuperman & van Dyke, 2013). Although this measure has been fruitful, its assessment is costly and does not provide individual measures per person. The proposed application of the sampling procedure presented here would result in individual frequency measures per participant and could thus be used to measure the effect of individual frequency on word processing in detail. This could open up new perspectives on the frequency by skill interaction.

In summary, the presented approach gives an answer to the question of how vocabulary size and structure can be assessed for individuals with a growing mental lexicon from primary school onwards. It enhances prior approaches and provides recent numbers on lexical development from late childhood to adulthood. The combination of a yes/no vocabulary test and the corpus sampling of virtual participants represents an innovative procedure which leads to estimates on total lexicon sizes for different age groups. In addition, further applications of the method could provide information on measures of print exposure and individual frequencies which could in turn be used for future studies on language processing. The methods established in this thesis can thus be regarded as an important contribution to the field of language assessment.

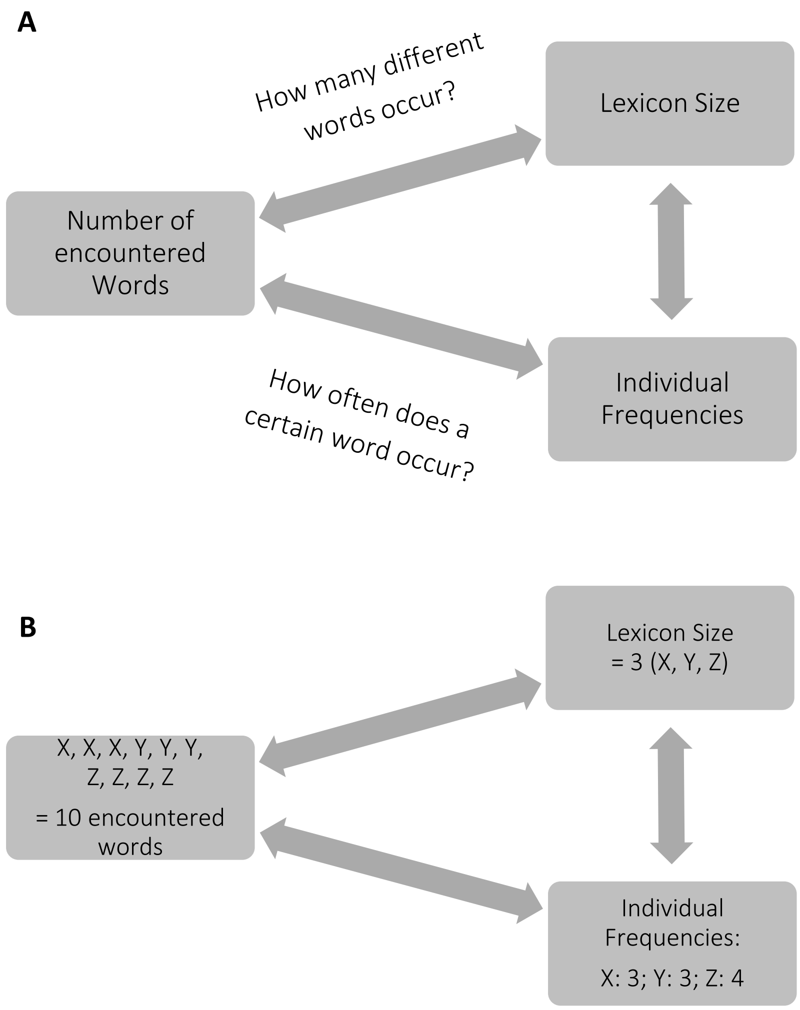

FIGURE 10.3: **A**: Schematic illustration of the application of the approach for language exposure and individual frequencies.
B: Exemplary illustration of the application with letters standing for words.

11 LEXICAL DEVELOPMENT FROM PRIMARY SCHOOL ONWARDS

The central motivation of this dissertation was the lack of findings on the courses of lexical development from primary school to adulthood. Justified by the lack of valid procedures to measure the mental lexicon with growing size, little was known about lexical development in late childhood. While it was clear that the quality of lexical entries in terms of linked knowledge in different language domains improves during this period (Richter et al, 2013, see section 2.3), the courses of later development of lexicon size and lexical structure were uncertain up to now. Previous studies on lexicon size were outdated and varied strongly (see Table 2.2, section 2) whereas no studies on lexical structure regarding orthographic neighborhoods existed for this age group. The thesis thus aimed to investigate the development of lexicon size (second research aim) and lexicon structure (third research aim). It is thus a sophisticated approach to provide insight into lexical acquisition in late childhood. In the following, I discuss the value of the presented results on lexical development and inter-individual differences as well as limitations and future lines of research derived from this thesis.

11.1 Development of Lexicon Size and Structure

Several key findings regarding the acquisition of vocabulary size and structure can be derived from the conducted studies:

First, as expected, the mental lexicon grows immensely from primary school to adulthood. While at the beginning of school entry, the average lexicon contains about 6,000 entries, young adults know about 73,000 words on average. The growth rate is accelerating as evident by the quadratic growth function (see Figure 7.4, chapter 7) as well as the increasing rates of words learned per day reported in Study 2 (see Table 7.5, chapter 7). The growth in lexicon size is mostly driven by an increasing amount of nouns and multimorphemic words.

These results on vocabulary growth underline the process of fast mapping where many words are added to vocabulary with only a few exposures within a short period of time. However, many authors described this process mainly for preschool children (e.g. Carey, 1978; Regier, 2005). The present results show, that the mechanism is still viable for older children as well. Although children acquire many important language skills in early childhood (Hoff, 2014), there is still a high rate of language development especially concerning vocabulary size and structure in late childhood. The growth of multimorphemic words within vocabulary indicates that a sophisticated morphological knowledge supports vocabulary development. As shown by Hasenäcker (2016), morphological processing improves during reading development in primary school for German children in that affixes gain orthographic representations, which can be used for whole-word representations in the mental lexicon and vice versa. The improvement of morphological skills may thus be one important factor to the substantial increase of vocabulary size, especially in a morphological rich language such as German.

The present results also demonstrate the central role of reading and school education for vocabulary development. Previous studies have shown that exposure to print enhances language skills such as vocabulary (for a review see Mol & Bus, 2011). In addition, new words tend to become more likely to appear in texts than in oral language (Hayes & Ahrens, 1988). An increase of reading activities may thus be one reason for the immense vocabulary growth and besides provides an explanation for the missing growth from first to second grade where reading ability is at the beginning of development.

As a second key finding, a large part of the mental lexicon is strongly interconnected via orthographic similarities throughout this development. On average, words possess six to seven orthographic neighbors with a growing trend during acquisition. The interconnectivity of the network, however, decreases in that the probability of neighbors to also be neighbors declines. This indicates that in the beginning of primary school, many very similar words are known and organized within a highly interconnected network. During development, new neighbors are added to existing entries but do not necessarily share similarities with many other words. With age, the mental lexicon thus becomes more and more differentiated supporting theories of reading development shifting from broader to more fine-tuned processes (Castles et al., 2007).

The results of Study 3 give a first insight on the possible underlying growth processes regarding network structure: As the neighborhoods of lexical entries increase, words appear to be added to already existing entries. As Barabási and Albert (1999) propose, networks with small-world characteristics grow via preferential attachment, i.e. new words are added to words that already possess many neighbors. This growth process could also be evident for orthographic development in primary school in the presented work: Children might tend to add words in to their mental lexicon to which they already know similar other words (i.e. orthographic neighbors) as indicated by the small-world characteristics found in the networks for all age groups as well as the growth in total neighborhood size. Yet, the decline of interconnectivity, especially the decrease of the clustering coefficient, does not clearly support this theory. Hills et al. (2009) proposed preferential acquisition as an alternative growth process and showed that, early in development, young children learned words with many semantic neighbors in the linguistic environment. They argue that these words are very prominent for children so that they can acquire them easily. To differentiate these particular growth processes for orthographic development via the presented network approach could be one goal for future studies. Results would help inform theories of language acquisition and improve teaching and training methods.

Yet, Study 3 has also shown that many words do not possess connections within the mental lexicon as evident by a high proportion of lexical hermits (about 50%). Developmental processes that assume the docking of new words to the existing network cannot explain the acquisition of these hermits, at least in terms of orthographic networks. Further studies on the course of acquisition on the individual level that take lexical hermits as well as other language domains such as semantics and phonology into account are necessary to tap into the underlying growth processes of vocabulary acquisition in late childhood. The present results could thus serve as a basis for this future goal.

11.2 Interindividual Differences

One central implication of the findings from Study 2 and Study 3 is the individuality of the mental lexicon. Study 2 demonstrated great interindividual differences in lexicon size within each age group with high standard deviations. For example, in first grade vocabulary size can range from 3,500 (-1 *SD*) to 8,500 (+1 *SD*) words, in fourth grade even between 9,000 (-1 *SD*) and 20,000 words (+1 *SD*).

In this context, previous studies have postulated the so-called Matthew effect, which proposes that good readers have higher learning rates than poor readers. This leads to an increasing gap between both groups (Duff, Tomblin & Catts, 2015; Mol & Bus, 2011; Stanovich, 2009; Anglin, Miller & Wakefield, 1993, see also section 2.3). The effect is strengthened by the assumption that good readers tend to read more in their leisure time than poor readers do since it is requires less effort for them (e.g. Anderson, Wilson & Fielding, 1988) and thus can even improve their vocabulary knowledge by learning words from texts. Previous studies have also shown that skilled readers tend to learn more words from contexts than poor readers do, also explaining the growing gap between both groups (e.g. Perfetti, Wlotko & Hart, 2005).

My findings support the occurrence of the Matthew effect in that interindividual differences in lexicon size increase with age. While children below average (-1 *SD*) start with about 3,500 words and know about 56,000 words as young adults, children above average (+1 *SD*) begin primary school with about 8,500 lexical entries and know about 91,000 words as young adults. The below average child thus gains about 52,000 entries while the lexicon of the above average child grows by circa 82,500 entries. Figure 11.1 exemplary illustrates the Matthew effect and thus the growing gap for lexicon size between a poor (-1 *SD*) and a proficient (+1 *SD*) language learner with the according lexicon sizes and quadratic growth functions with data provided in Study 2. It remarkably shows the differences in vocabulary growth rates for these exemplary language learners.

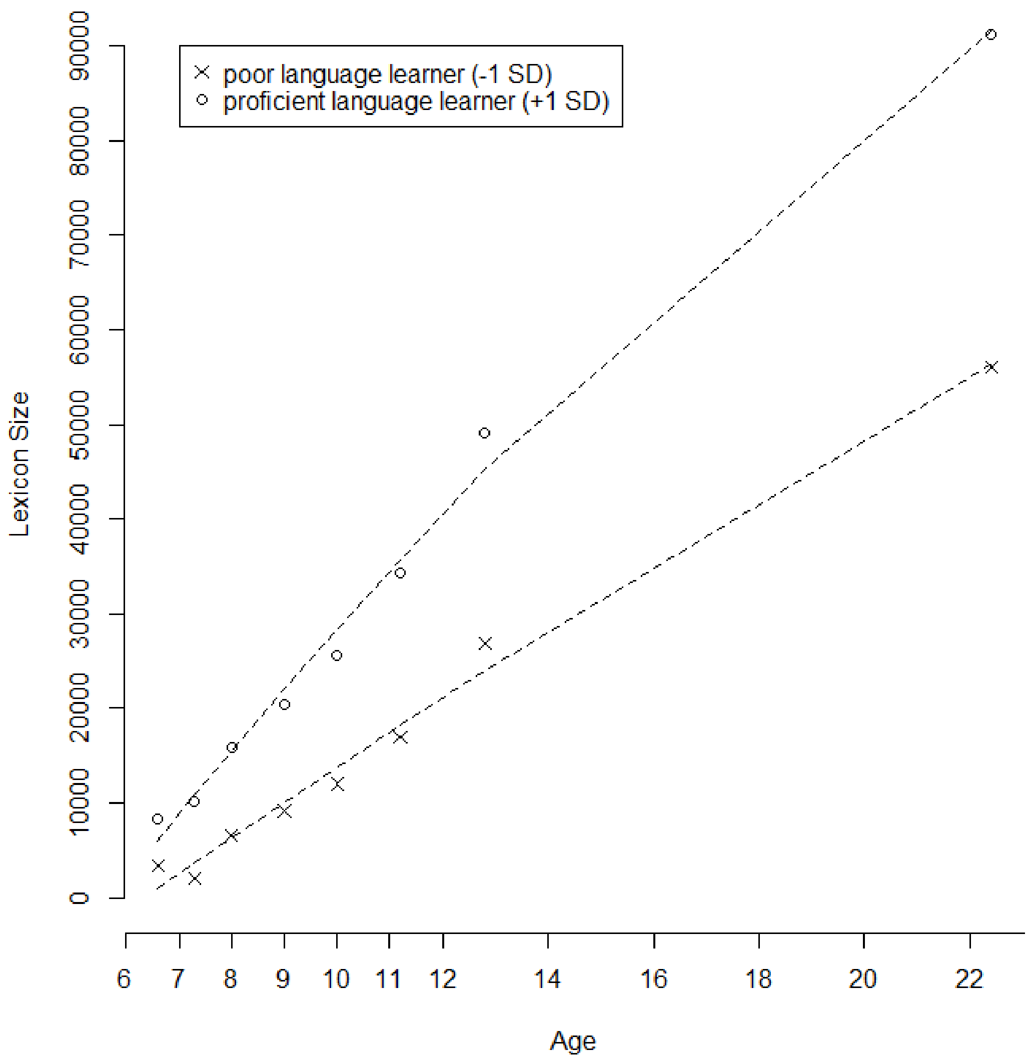

Figure 11.1: Mean lexicon sizes and growth curves for a poor (-1 *SD*) and a proficient (+1 *SD*) language learner.

Note: The figure contains mean lexicon sizes derived from Study 2.Lines represent the according quadratic growth functions.

Although Study 3 used simulated average data and did not directly report interindividual differences, the results allow similar conclusions: The network characteristics change with lexicon size, which implies that the exemplary same aged children (-1/+1 *SD*) differ substantially in their network measures and thus their network connectivity – their measures can be located at different points of the curves presented in Figure 8.2 (chapter 8). Even though these interindividual differences have been assumed before and have been found for younger children (e.g. Kauschke, 1999; Kidd, Donnelly & Christiansen, 2018), the present studies provide an impressive insight on the extent of these differences.

With an additional analysis, I illustrate these differences in Figure 11.2 for the clustering coefficient *C*. For this supplementary examination, I used the average lexicon size of a virtual poor (-1 *SD*) and a virtual proficient (+1 *SD*) language learner derived from Study 2 and depicted in Figure 11.1[10]. With the network sampling approach presented in Study 3 I then determined the orthographic network for these different virtual participants and subsequently calculated their clustering coefficient *C*. Figure 11.2 shows that the clustering coefficient is generally higher for the poor language learner since a smaller lexicon comes along with a higher interconnectivity (see Study 3). While *C* is similar for both virtual participants at the beginning of development, it diverges in continuative language acquisition, again underlining the Matthew effect. In terms of lexical development, the network of the proficient language learner thus becomes more and more differentiated. This additional analysis on the one hand provides an outlook on further applications of the network approach on the individual level (see also section 10.3). On the other hand, regarding the results on differences in lexical structure, it impressively depicts interindividual differences in network characteristics dependent on lexicon size.

In this regard it might also be interesting to evaluate the lexical structure of individuals with different age and the same lexicon size – for example a young proficient language learner and an older poor language learner. The underlying question is whether the different rates of vocabulary growth or possible different growth processes and experience lead to differences in lexical structure despite the equality of lexicon size. However, to address this question, the determination of networks presented in Study 3 needs to be refined in that it

[10] For this purpose, the lexicon sizes counted in lemmas (Figure 11.1) were transferred into types as it was done in Study 3.

does not only rely on lexicon size but also takes differences in growth processes and experience into account (see also section 10.3).

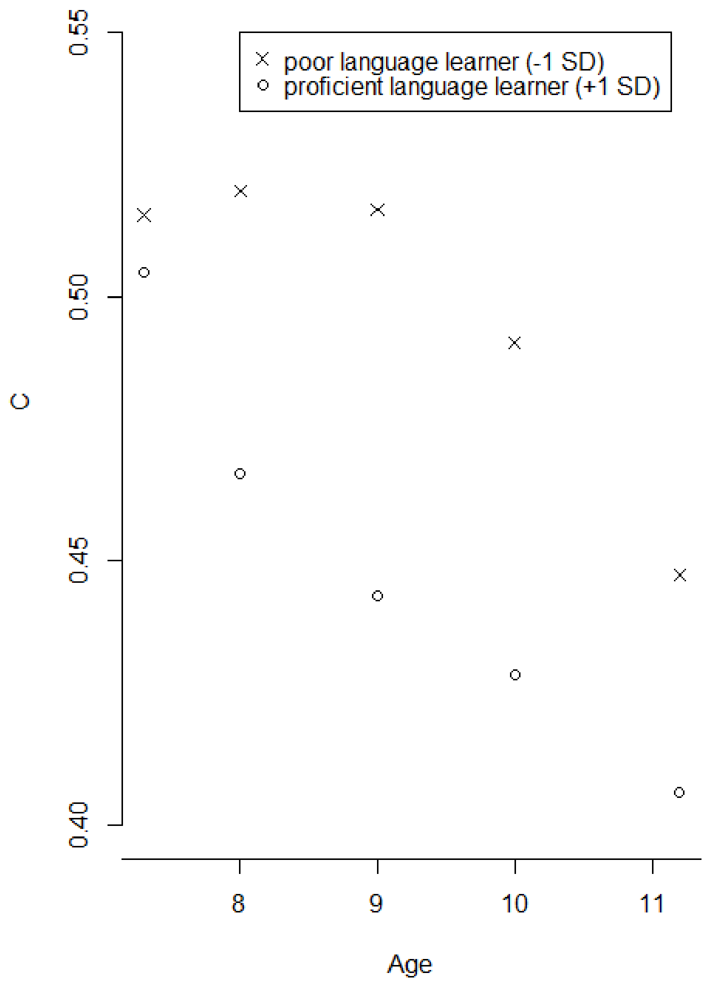

FIGURE 11.2. Clustering coefficient for a virtual poor (-1 *SD*) and a virtual proficient (+1*SD*) language learner.
Note: The network approach presented in Study 3 was applied to the average lexicon sizes – 1 SD and + 1 SD provided in Study 2.

The presented findings on interindividual differences lead to important further questions to be addressed in the future:

First, how do these great interindividual differences occur? Researchers have discussed several factors that might influence vocabulary learning. Among these are the home language environment (e.g. Hoff, 2006; Tamis-Lemonda, Schannon, Cabrera & Lamb, 2004) and school education (Beck & McKeown, 2007; Biemiller & Boote, 2006). Both factors were not examined within the present studies but should be addressed in future applications of the established methods to find causes for interindividual differences. Although the present research did not aim to explain these interindividual differences in the first place, the results on external validity of the vocabulary test from Study 1 provide a first hint: High correlations with measures of reading and writing skills with the vocabulary measure indicate that these skills are essential for lexical development.

This is in line with studies that argue that especially in late childhood, reading becomes an important source of vocabulary acquisition (e.g. Nagy, Anderson & Herman, 1987; Jenkins, Stein & Wysocki, 1984). However, the relation between vocabulary and reading ability is reciprocal as proposed by Perfetti's DVC triangle model (2010). The author claims that, on the one hand, a larger vocabulary enhances reading comprehension and, on the other hand, a higher reading comprehension allows the drawing of inferences of meanings from context resulting in a larger vocabulary. This reciprocal relation also strengthens the Matthew effect. One future aim of linguistic and educational research has thus to be to break this "vicious circle" to enable equal chances of education and academic achievement as vocabulary and reading are very important influential factors (Hoff, 2014; Walker, Greenwood, Hart & Carta, 1994; Graham, 1987).

This leads to a second central question regarding the present results: How do the interindividual differences in vocabulary size and structure affect other skills, e.g. language and literary skills or school achievement in general? Several authors have shown the effect of vocabulary on those skills (e.g. Rowe, Raudenbusch & Goldin-Meadow, 2012; Lee, 2011; Ouellette, 2006; Walley, Metsala & Garlock, 2003; Staehr, 2009) but their studies mostly include measures of semantic knowledge (e.g. through definition tasks) and mainly do not represent the total size or structure of vocabulary.

The newly established measure of vocabulary size presented in this thesis could additionally be used to inform and refine the relation between vocabulary size and other skills and vice versa. Future work could address, among others, the following substantial research questions: How does total lexicon size and content of an individual affect reading and listening comprehension of a certain sentence or text (see also Nation, 2006)? How many and which bi- and multimorphemic words are stored in an individual's lexicon and how does this influence morphological processing (see also Hasenäcker, 2016)? How much leisure time reading is necessary to reach a certain size of vocabulary (see also Cunningham & Stanovich, 1991)? With the presented approach, these questions could be addressed under a developmental approach informing theories of language and reading acquisition. In addition, the application of the present approach could inform the discussion on the distinction of vocabulary breadth and depth in that the measure of total lexicon size (breadth) could be compared to the assessment of deeper word knowledge (depth) to answer whether both measures address different concepts or not (see also Vermeer, 2001).

Besides lexicon size, another possible line of research derived from the findings of interindividual differences concerns the effect of individual network structure on language processing. As elaborated above, orthographic neighborhoods affect word processing but have been derived from corpora up to now (e.g. Andrews, 1992). Furthermore, results on the effect have been controversial with inhibitory as well as facilitative effects and mainly address the effect for adults (Andrews, 1997, see chapter 2.4) and only few studies include network measures (but see Chan & Vitevitch, 2009; 2010). The present thesis has shown that orthographic interconnectivity strongly changes in dependence of vocabulary size (see Figure 11.2). This holds for developmental differences as well as for differences between same-aged individuals, e.g. adults. Using the same corpus measure for neighborhoods or network characteristics may thus not be suitable to capture the effect of orthographic similarities on language processing. One central goal for future research could thus be to determine individual network characteristics and measure their effect on individual language processing mechanisms. Investigating changes of network effects on word processing during this period and attributing them to changes in the network itself could inform theories on the influence of the individual structure of the mental lexicon on language skills.

The findings on individual lexicon size, content and structure could in turn be used to refine and individualize models of language processing such as the Dual Route Cascaded Model of Visual Word (Coltheart et al., 2001) or the Multiple Read-Out Model (Grainger & Jacobs, 1996). These models aim to explain word retrieval processes and usually include an (orthographic) lexicon where known words are saved and retrieved from. This lexicon, however, is commonly based on corpus data such as the total content of the corpus and characteristics such as orthographic neighborhood size. The models thus assume that all words of the corpus and along with this, all possible orthographic neighbors of a word are known by every individual. However, the present results contradict this assumption. Determining the individual's lexicon size, content and structure thus would allow the adaptation of modelling on an individual level. This could in turn help to understand and explain processes of language performance as well as interindividual differences (see also Yap et al., 2012; Adelman, Sabatos-DeVito, Marquis & Estes, 2014).

Connected to the previous questions of how interindividual differences occur and affect other language skills is the question of how they can be prevented. Establishing effective reading training methods and promoting leisure time reading could be on promising approach here considering reading as an important source of vocabulary learning (Duff, Tomblin & Catts, 2015). Current approaches address peer tutoring (e.g. Kourea, Cartledge & Musti-Rao, 2007; Fuchs, Fuchs & Burish, 2000) or the application of computer software (e.g. Stodden, Roberts, Takahashi, Park & Stodden, 2012; Maracuso, Hook & McCabe, 2006). All these programs include the establishment of different reading strategies to improve reading abilities, mostly comprehension (for a review, see Gersten, Fuchs, Williams & Baker, 2001). Studies suggest that well-equipped libraries together with easy access to books and time to read are key factors in promoting reading habits (e.g. Rodrigo, Greenberg & Segal, 2014; Green, Peterson & Lewis, 2006; Morrow & Weinstein, 1986). Furthermore, since home language environment is an influential factor, parental trainings could be beneficial to promote reading habits and thereby vocabulary. Most studies and training methods address parents of preschool children and have shown that dialogic reading is an effective intervention to affect early language and literary skills and can be administered to parents for home use (e.g. Huebner & Meltzoff, 2005; Blom-Hoffman, O'Neil-Pirozzi, Volpe, Cutting & Bissinger, 2007). Early training of parents thus appears to be a promising approach since these early literacy experiences affect later reading habits (Baker, Scher & Mackler, 1997).

In addition, methods of vocabulary instructions in school could be evaluated and improved with the present approach. Several studies have shown that children benefit from direct instruction (e.g. Beck & McKeown, 2007; Biemiller, 2006). The presented findings can provide important implications for such training and teaching studies and practices. When wanting to teach new words one could investigate which words are unknown to the average learner of the target group and use them in word training studies as well as in class. With a pre- and post-test design, one could additionally use the newly introduced measure of lexicon size to evaluate the effectiveness of training or teaching.

The presented results additionally provide possibilities to handle interindividual differences in learning contexts: Knowing how many and which words an individual approximately knows (or does not know) can facilitate the adaptation of materials or texts for different groups of readers. For example, when writing a text for a certain age group with the goal that the average member of the audience can understand the text, one could elaborate which words are known by most of the individuals in that certain age group and include them into the text. Besides, when selecting given texts for certain individuals one could analyze their content in terms of which words are suitable for the target audience (see also Hsueh-Chao, Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010). For research purposes, the results on lexical content enable choosing items for tasks that fit the population of interest. For example, for conducting experiments on word recognition with children of different age groups, one could select words that are suitable for specific age groups. In this case, the definition of "suitable" could vary in that, for example, most participants of the target group know a certain word or that a word is very unfamiliar to a large part of the population of interest.

Taken together the thesis provides important implications for the course of lexical development from primary school onwards as well as future prospects for consecutive research. The findings highlight the relevance of late childhood for language development and point to the occurrence and increase of large interindividual differences in lexical development.

12 FINAL CONCLUSIONS

The present dissertation investigated the course of lexical development from primary school to adulthood. Its first aim was to develop a method to assess the mental lexicon in late childhood. The second aim included the description of lexical development in terms of lexicon size and lexical structure. One central outcome is the development of a sampling approach that can be adapted and extended to other contexts for future research on the mental lexicon. Results showed that the mental lexicon develops strongly from primary school onwards in terms of number of known words and connections between entries and thus emphasizes this time period as central for language acquisition. Supporting language development for this age group thus becomes more and more important, especially because my findings demonstrate the occurrence of extensive interindividual differences among children and adults. Since these differences might entail difficulties in other (language) skills and reduce chances of academic achievement, vocabulary training and instruction needs to be taken into account more intensely in primary school curricula. In conclusion, the present thesis can be regarded as a major contribution to basic research on lexical development referring to both methodological as well as substantial issues.

IV References

Aarnoutse, C., van Leeuwe, J., Voeten, M. & Oud, H. (2001). Development of decoding, reading comprehension, vocabulary and spelling during the elementary school years. *Reading and Writing: An Interdisiplinary Journal*, *14*, 61–89. doi: https://doi.org/10.1023/A:1008128417862

Abad, M. J., Noguera, C., & Ortells, J. J. (2003). Influence of prime–target relationship on semantic priming effects from words in a lexical-decision task. *Acta Psychologica*, *113*(3), 283-295.

Acha, J. & Perea, M. (2008). The effects of length and transposed-letter similarity in lexical decision: Evidence with beginning, intermediate, and adult readers. British Journal of Psychology, 99, 245-264. doi: 10.1348/000712607X224478

Adelman, J. S., Sabatos-DeVito, M. G., Marquis, S. J. & Estes, Z. (2014). Individual differences in reading aloud: A mega-study, item effects and some models. *Cognitive Psychology, 68,* 113-160. doi: 10.1016/j.cogpsych.2013.11.001

Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon* (4th ed.). Malden, Mass.: Wiley.

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching*. Newark: Interanational Reading Association.

Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. *Advances in Reading/Language Research*, *2*, 231–256.

Anderson, R. C., Wilson, P. T., & Fielding, L. G. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly*, *23*, 285–303.

Andrews, S. (1992). Frequency and Neighborhood Effects on Lexical Access: Lexical Simi-larity or Orthographic Redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(2), 234–254.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, *4*(4), 439–461.

Anglin, J. M., Miller, G. A. & Wakefield, P. C. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development, 58*, 1–186. doi: 10.2307/1166112

Anthony, J. L., Lonigan, C. J., Driscoll, K., Phillips, B. M., & Burgess, S. R. (2003). Phonological sensitivity: A quasi-parallel progression of word structure units and cognitive operations. *Reading Research Quarterly*, *38* (4), 470-487. doi: 10.1598/RRQ.38.4.3

Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science*, *14* (5), 255-259. doi: 10.1111/j.0963-7214.2005.00376.x

Anthony, J. L., & Lonigan, C. J. (2004). The nature of phonological awareness: converging evidence from four studies of preschool and early grade school children. *Journal of Educational Psychology*, *96* (1), 43 -55.

Arbesman, S., Strogatz, S. H., & Vitevitch, M. S. (2010). The Structure of Phonological Networks Across Multiple Languages. *International Journal of Bifurcation and Chaos*, *20*(3), 679–685.

Auer, M., Gruber, G., Mayringer, H. & Wimmer, H. (2005). *Salzburger Lese-Screening für die Klassenstufe 5-8*. Bern: Hans Huber.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (Release 2)* [CD-ROM]. Philadelphia, PA:Linguistic Data Consortium, University of Pennsylvania.

Baker, L., Scher, D. & Mackler, K. (1997). Home and family influences on motivations for reading. *Educational Psychologist, 32* (2), 69-82. doi: 10.1207/s15326985ep3020_2

Balota et al. (2007). The English Lexicon Project. *Behavior Research Methods, 39* (3), 445-459. doi: 10.3758/BF03193014

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286* (5439), 509-512. doi: 10.1126/science.286.5439.509

Bates, E., & Goodman, J. C. (1999). On the emergence of grammar from the lexicon. In B. MacWhinney (Ed.), *The emergence of language*. Mahwah, NJ: Lawrence Erlbaum.

Beck, I. L. & McKewon, M. G. (2007). Increasing Young Low-Income Children's Oral Vocabulary Repertoires through Rich and Focused Instruction. *The Elementary School Journal, 107* (3), 251-271.

Beckage, N., Smith, L. & Hills, T. (2011). Small Worlds and Semantic Network Growth in Typical and Late Talkers. *PLoS ONE 6* (5). doi: 10.1371/journal.pone.0019348

Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing*, *27*, 101–118. doi: 10.1177/0265532209340194

Bialystok, E., Craik, F. I. M. & Luk, G. (2008). Lexical access in bilinguals: Effects of vocabulary size and executive control. *Journal of Neurolinguistics*, *21*, 522–538. doi:10.1016/j.jneuroling.2007.07.001

Bialystok, E., Luk, G., Peets, K. F. & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*, 525–531. doi: 10.1017/S1366728909990423

Biemiller, A. (2003). Vocabulary: Needed if more children are to read well. *Reading Psychology*, *24*, 232–335. doi: 10.1080/02702710390227297

Biemiller, A. (2006). Vocabulary development and instruction: A prerequisite for school learning. In D. K. Dickinson & S. B. Neuman (Eds.), *Handbook of Early Literacy Research* (Vol. 2)*. 41-51.* New York: Guilford Press.

Biemiller, A. & Slonim, N. (2001). Estimating Root Word Vocabulary Growth in Normative and Advantaged Populations: Evidence for a Common Sequence of Vocabulary Acquisition. *Journal of Educational Psychology, 93* (3), 489-520. doi: 0.1037//0022-0663.93.3.498

Biemieller, A. & Boote, C. (2006). An Effective Method for Building Meaning Vocabulary in Primary Grades. *Journal of Educational Psychology, 98* (1), 44-62. doi: 10.1037/0022-0663.98.1.44

Bilson, S., Yoshida, H., Tran, C., Woods, E., & Hills, T. (2015). Semantic facilitation in bilingual first language acquisition. *Cognition, 140*, 122-134.

Blom-Hoffman, J., O'Neil-Pirozzi, T., Volpe, R., Cutting, J. & Bissinger, E. (2007). Instructing Parents to Use Dialogic Reading Strategies with Preschool Children. *Journal of Applied Psychology, 23* (1), 117-131. doi: 10.1300/J370v23n01_06

Bock, R. D. & Zimowski, M. F. (1997). Multiple group IRT. In van der Linden, W. J. & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 433–448). New York: Springer.

Bonin, P., Barry, C., Méot, A., Chalard, M. (2004). The influence of age of acquisition in word reading and other tasks: A never ending story? *Journal of Memory and Language, 50,* 456-476. doi: 10.1016/j.jml.2004.02.001

Borgatti, S. P. (2006). Identifying sets of key players in a social network. *Computational and Mathematical Organization Theory*, *12*(1), 21–34. doi: 10.1007/s10588-006-7084-x

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science*, *323*(5916), 892-895. doi: 10.1126/science.1165821

Braze, D., Tabour, W., Schankweiler, D. P., & Mencl, W. E. (2007). Speaking up for vocabulary: Reading skill differences in young adults. *Journal of Learning Disabilities*, *40*(3), 226–243.

Buchanan, L., Westbury, C. & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review, 8* (3), 531-544.

Burke, D. M., & Shafto, M. A. (2004). Aging and language production. *Current directions in psychological science*, *13* (1), 21-24. doi: 10.1111/j.0963-7214.2004.01301006.x

Burgess, C., & Lund, K. (2000). The dynamics of meaning in memory. In E. Dietrich & A. B. Markman (Eds.), *Cognitive dynamics: Conceptual and representational change in humans and machines* (pp. 117-156). Mahwah, NJ: Erlbaum.

Carey, S. (1978). The child as a word learner. In J. Bresnan, G. Miller, & M. Halle (Eds.), *Linguistic Theory and Psychological Reality* (pp. 264–293). Cambridge: MIT Press.

Carey, S. & Bartlett, E. (1978). Acquiring a Single New Word. *Papers and Reports on Child Language Development, 15,* 17-29.

Castles, A., Davis, C. & Letcher, T. (1999). Neighbourhood Effects on Masked Form Priming in Developing Readers. *Language and Cognitive Processes*, *14*(2), 201–224. doi: 10.1080/016909699386347

Castles, A., Davis, C., Cavalot, P. & Forster, K. (2007). Tracking the acquisition of orthographic skills in developing readers: Masked priming effects. *Journal of Experimental Child Psychology, 97,* 165-182, doi: 10.1016/j.jecp.2007.01.006

Cattell, R. B., Weiß, R. H. & Osterland, J. (1997). *Grundintelligenztest Skala 1*. Göttingen: Hogrefe.

Chan, K. Y., & Vitevitch, M. S. (2009). The influence of the phonological neighborhood clustering coefficient on spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1934–1949. doi: 10.1037/a0016902

Chan, K. Y., & Vitevitch, M. S. (2010). Network structure influences speech production. *Cognitive Science*, *34*(4), 685–697. doi:10.1111/j.1551-6709.2010.01100.x

Charles-Luce, J. & Luce, P. A. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language, 17*, 205-215. doi: 10.1017/S0305000900013180

Chateau, D., & Jared, D. (2000). Exposure to print and word recognition processes. *Memory & Cognition*, *28*(1), 143-153.

Clark, R., Hutcheson, S., & van Buren, P. (1974). Comprehension and Production in Language Acquisition. *Journal of Linguistics*, *10*(1), 39–54.

Clark, E. V. (1993). *The lexicon in acquisition.* Cambridge: Cambridge University Press.

Colé, P., Cavalli, E., Duncan, L. G., Theurel, A., Gentaz, E., Sprenger-Charolles, L. & El-Ahmadi, A. (2018). What is the influence of morphological knowledge in the early stages of reading acquisition among low SES children? A graphical modelling approach. Frontiers in Psychology, 9, 1-15. Doi: 10.3389/fpsyg.2018.000547

Coltheart, M.; Davelaar, E.; Jonasson, J. T & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp.535-555). Hillsdale, NJ: Erlbaum.

Coltheart, M., Rastle, K., Perry, C., Langdon, R. & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological review*, *108*, 204–256.

Csardi, G. & Nepusz, T. (2006). The igraph software package for complex network research. Retrieved from http://igraph.org

Cunningham, A. E., & Stanovich, K. E. (1991). Tracking the Unique Effects of Print Exposure in Children: Associations with Vocabulary, General Knowledge, and Spelling. *Journal of Educational Psychology*, *83*(2), 264–274.

D'Anna, C., Zechmeister, E., & Hall, J. (1991). Toward a meaningful definition of vocabulary size. *Journal of Reading Behavior*, *23*(1), 109–122. doi: 10.1080/10862969109547729

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first-and second-language word recognition: A lexical entrenchment account. *Quarterly Journal of Experimental Psychology*, *66*(5), 843-863. doi: 10.1080/17470218.2012.720994

Dollaghan, C. (1985). Child meets word: Fast mapping in preschool children. *Journal of Speech, Language, and Hearing Research*, *28* (3), 449-454.

Donlan, C., Cowan, R., Newton, E. J. & Lloyd, D. (2007). The role of language in mathematical development: Evidence from children with specific language impairments. *Cognition*, *103*, 23–33. doi:10.1016/j.cognition.2006.02.007

Duff, D., Tomblin, J. B. & Catts, H. (2015). The Influence of Reading on Vocabulary Growth: A Case for a Matthew Effect. *Journal of Speech, Language and Hearing Research, 58,* 853-864. doi: 10.1044/2015_JSLHR-L-13-0310

Dunn, D. M., & Dunn, L. M. (2007). *Peabody picture vocabulary test*. London: Pearson.

Embretson, S. E. & Reise, S. P. (2000). *Item Response Theory for Psychologists*. New Jersey: Lawrence Erlbaum.

Durda, K., Buchanan, L., & Caron, R. (2006). WordMine2 [Online]. Available at www.wordmine2.org.

Eyckmans, J. (2004). *Measuring receptive vocabulary size: Reliability and validity of the Yes/No Vocabulary Test for french-speaking learners of dutch*. Utrecht: LOT.

Fengxiang, Fan. (2010). An asymptotic model for the English hapax/vocabulary ratio. *Computational Linguistics, 36* (4), 631-637.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D., & Pethik, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., . . . Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual*. Baltimore: Paul H. Brokes Publishing Co.

Ferrer i Chancho, R. & Solé, R. V. (2001). The small world of human language. *Proceedings of the Royal Society B., 268*, 2261-2265, doi: 10.1098/rspb.2001.1800

Fleischer, W., Barz, I., & Schröder, M. (2012). *Wortbildung der deutschen Gegenwartssprache* (4. Edition.). *De Gruyter Studium*. Berlin: De Gruyter.

Frauenfelder, U. H., Baayen, R. H., & Hellwig, F. M. (1993). Neighborhood density and frequency across languages and modalities. *Journal of Memory and Language*, *32*(6), 781-804.

Fuchs, D. , Fuchs, L. S. & Burish, P. (2000). Peer-Assisted Learning Strategies: An Evidence-Based Practice to Promote Reading Achievement. *Learning Disabilities Research & Practice, 15* (2), 85-91. doi: 10.1207/SLDRP1502_4

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., et al. (2006). The cognitive correlates of third grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology*, *98*, 29–43. doi:0.1037/0022-0663.98.1.29

Gathercole, S., & Baddeley, A. D. (1989). Evaluation of the Role of Phonological STM in the Development of Vocabulary in Children: A Longitudinal Study. *Journal of Memory and Language*, *28*, 1989.

Gersten, R., Fuchs, L. S., Williams, J. P. & Baker, S. (2001). Teaching Reading Comprehension Strategies to Students With Learning Disabilities: A Review of Research. *Review of Educational Research, 71* (2), 279-320.

Geyken, A. (2007). The DWDS corpus: A reference corpus for the German language of the 20th century. In C. Fellbaum (Ed.). *Collocations and idioms: Linguistic, lexicographic, and computational aspects* (pp. 23–41). London, UK: Continuum.

Geyken, A., & Hanneforth, T. (2006). TAGH: A complete morphology for German based on weighted finite state automata. In A. Yli-Jyrä, L. Karttunen, & J. Karhumäki (Eds.), *Finite state methods and natural language processing* (pp. 55–66). Berlin, Germany: Springer. doi: 10.1007/11780885_7

Glas, C. A. W. & Verhelst, N. D. (1995). Testing the Rasch model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models* (pp. 69–95). New York: Springer.

Glück, C. W. (2011). *Wortschatz- und Wortfindungstest für 6- bis 10-Jährige*. Amsterdam: Elsevier.

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, *35*(3), 515–531. doi: 10.1017/S0305000907008641

Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, *11*(4), 341–363.

Graham, J. G. (1987). English language proficiency and the prediction of academic success. *TESOL quarterly*, *21* (3), 505-521.

Grainger, J. (1990). Word Frequency and Neighborhood Frequency Effects in Lexical Decision and Naming, *Journal of Memory and Language, 29*, 228-244.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: a multiple read-out model. *Psychological review*, *103*(3), 518.

Grainger, J., Muneaux, M., Farioli, F., & Ziegler, J. C. (2005). Effects of phonological and orthographic neighbourhood density interact in visual word recognition. *The Quarterly Journal of Experimental Psychology Section A*, *58*(6), 981-998. doi: 10.1080/02724980443000386

Grainger, J., & Ziegler, J. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology*, *2*, 54, 1-13. doi: 10.3389/fpsyg.2011.00054.

Green, S. D., Peterson, R. & Lewis, J. R. (2006). Language and Literacy Promotion in Early Childhood Settings: A Survey of Center-Based Practices. *Early Childhood Research & Practice, 8* (1).

Grimm, H. & Doil, H. (2005). *Elternfragebögen für die Früherkennung von Risikokindern (ELFRA)* (2. Edition). Göttingen: Hogrefe.

Gutierrez-Cleflen, V. F. & DeCurtis, L. (1999). Word definition skills in Spanish-speaking children with language impairment. *Communication Disorders Quarterly*, *21*, 23-31. doi: 10.1177/152574019902100104

Hammill, D. D., & Larsen, S. C. (2009). *Test of written language: TOWL4*. Pro-ed.

Hart, L., & Perfetti, C. A. (2008). Learning words in Zekkish: Implications for understanding lexical representation. In Grigorenk, E. L. & Naples, A. J. (Eds.) *Single word reading: Behavioral and biological perspectives*, New York: Taylor & Francis, 107-128.

Hasenäcker, J. (2016). *Learning to read complex words: Morphological processing in reading acquisition.* Dissertation, Freie Universität Berlin.

Hasenäcker, J., & Schroeder, S. (2017). Syllables and morphemes in German reading development: Evidence from second graders, fourth graders, and adults. *Applied Psycholinguistics*, *38*(3), 733-753. doi: 10.1017/S0142716416000412

Hawkins, K. A., & Bender, S. (2002). Norms and relationship of the Boston Naming Test performance to vocabulary and education: A review. *Aphasiology, 16*, 1143–1153

Hayes, D. P., & Ahrens, M. G. (1988). Vocabulary simplification for children: A special case of 'motherese'? *Journal of Child Language*, *15*(02), 395. doi: 10.1017/S0305000900012411

Heibeck, T. H., & Markman, E. M. (1987). Word learning in children: An examination of fast mapping. *Child Development, 58* (4), 1021-1034.

Heller, K. & Geisler, H. J. (1983). *Kognitiver Fähigkeitstest für 1. bis 3. Klassen*. Weinheim: Beltz.

Heller, K. A. & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision*. Weinheim: Beltz.

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Longitudinal analysis of early semantic networks: preferential attachment or preferential acquisition? *Psychological Science*, *20*(6), 729–739. doi:10.1111/j.1467-9280.2009.02365.x

Hills, T., Maouene, J., Riordan, B., & Smith, L. (2010). The associative structure of language and contextual diversity in early language acquisition. *Journal of Memory and Language, 63*, 259-273.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review, 26,* 55-88. doi: 0.1016/j.dr.2005.11.002

Hoff, E. (2014). *Language development* (5. ed.). S.l.: Wadsworth Cengage Learning.

Hoff, E. & Naigles, L. (2002). How Children Use Input to Acquire a Lexicon. *Child Development, 73* (2), 418-433. doi: 10.1111/1467-8624.00415

Holderbaum, C. S. & Fumagalli de Salles, J. (2011). Semantic Priming Effects in a Lexical Decision Task: Comparing third Graders and College Students in two Different Stimulus Onset Asynchronies. The Spanish Journal of Psychology, 14 (2), 589-599. doi: 10.5209/rev_SJOP.2011.v14.n2.7

Hsueh-Chao, M. H. & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a foreign language*, *13*(1), 403-430.

Huebner, C. E. & Meltzoff, A. N. (2005). Intervention to change parent-child reading style: A comparison of instructional methods. *Journal of Applied Developmental Psychology, 26* (3), 296-313. doi: 10.1016/j.appdev.2005.02.006

Huibregtse, I., Admiraal, W. & Merea, P. (2002). Scores on a yes-no vocabulary test: Correction for guessing and response style. *Language Testing*, *19*, 227–245. doi: 10.1191/0265532202lt229oa

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M. & Lyons, T. (1991). Early Vocabulary Growth: Relation to Language Input and Gender. *Developmental Psychology, 27* (2), 236–248.

Jenkins, J. R., Stein, M. L., & Wysocki, K. (1984). Learning Vocabulary Through Reading. *American Educational Research Journal*, *21*(4), 767–787. doi: 10.3102/00028312021004767

Johnson, C. J. & Anglin, J. M. (1995). Qualitative developments in the content and form of children's definitions. *Journal of Speech, Language, and Hearing Research*, *38* (3), 612-629.

Joshi, R. M. (2005) Vocabulary: A Critical Component of Comprehension, *Reading & Writing Quarterly: Overcoming Learning Difficulties, 21* (3), 209-219. doi: 10.1080/10573560590949278

Jurish, B. (2003). *Part-of-speech tagging with finite state morphology*. Berlin, Germany: Poster presented at the conference Collocations and Idioms: Linguistic, Computational, and Psycholinguistic Perspectives.

Jurish, B. & Würzner, K.-M. (2013): Word and Sentence Tokenization with Hidden Markov Models. *JLCL*, *28*(2), 61-83.

Kauschke, C. (1999). Früher Wortschatzerwerb im Deutschen: eine empirische Studie zum Entwicklungsverlauf und zur Komposition des kindlichen Lexikons. In: Meibauer, J. & Rothweiler, M. (Eds.) *Das Lexikon im Spracherwerb,* Stuttgart, UTB.

Kauschke, C. & Hofmeister, C. (2002). Early lexical development in German. A study on vocabulary growth and vocabulary composition during the second and third year of life. *Journal of Child Language, 29* (4), 735–757.

Kauschke, C. & Siegmüller, J. (2009). *Patholinguistische Diagnostik bei Sprachentwicklungsstörungen* (2. Ed.). Amsterdam: Elsevier.

Kemper, S., Greiner, L., Marquis, J. G., Prenovost, K. & Mitzner, T. L. (2001). Language Decline Across the Life Span: Findings From the Nun Study. *Psychology and Aging, 16* (2), 227-239. doi: 10.1037//0882-7974.16.2.227

Keuleers, E. (2013). vwr: Useful functions for visual word recognition research. Retrieved from https://CRAN.R-project.org/package=vwr

Kidd, E., Donnelly, S. & Christiansen, M. H. (2018). Individual Differences in Language Acquisition and Processing. *Trends in Cognitive Science, 22* (2), 154-169. doi: 10.1016/j.tics2017.11.006

Kiefer, T., Robitzsch, A. & Wu, M. (2016). *TAM: Test Analysis Modules* [Computer Software].

Kiese-Himmel, C. (2005). *AWST-R. Aktiver Wortschatztest für 3- bis 5-jährige Kinder – Revision.* Göttingen: Hogrefe.

Kolen, M. J. & Brennan, R. L. (2004). *Test Equating, Scaling and Linking. Methods and Practices.* New York: Springer.

Kornai, A. (2002). How many words are there? *Glottometrics, 4*, 61-86.

Kourea, L., Cartledge, G. & Musti-Rao, S. (2007). Improving the Reading Skills of Urban Elementary Students Through Total Class Peer Tutoring. *Remedial and Special Education, 28* (2), 95-107. doi: 10.1177/07419352070280020801

Kučera, H., & Francis, W. (1967). *Computational analysis of presentday American English*. Providence, RI: Brown University Press.

Kuperman, V., Stadthagen-Gonzales, H. & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44,* 978-990. doi: 10.3758/s13428-012-0210-4

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 802-823. doi: 10.1037/a0030859

Kyte, C. S., & Johnson, C. J. (2006). The role of phonological recoding in orthographic learning. *Journal of Experimental Child Psychology*, *93*(2), 166–185. doi:10.1016/j.jecp.2005.09.003

Landauer, T. K., & Streeter, L. A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior*, *12*(2), 119-131.

Landi, N. (2010). An examination of the relationship between reading comprehension, higher-level and lower-level reading sub-skills in adults. *Reading and Writing*, *23*(6), 701–717. doi: 10.1007/s11145-009-9180-z

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical Threshold Revisited: Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension. *Reading in a Foreign Language*, *22*(1), 15-30.

Lee, J. (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics, 32,* 69-92. doi: 10.1017/S0142716410000299

Lemhöfer, K. & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325–343. doi: 10.3758/s13428-011-0146-0

Lenhard, A., Lenhard, W., Segerer, R. & Suggate, S. (2015). *Peabody Picture Vocabulary Test – 4. Ausgabe: Deutsche Fassung*. Frankfurt am Main: Pearson Assessment.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet physics doklady*, *10*, 707.

Locker, L. Jr., Simpson, G. B. & Yates, M. (2003). Semantic neighborhood effects on the recognition of ambiguous words. *Memory & Cognition, 31* (4), 505-515. doi: 10.3758/BF03196092

Lorge, I., & Chall, J. (1963). Estimating the Size of Vocabularies of Children and Adults: An Analysis of Methodological Issues. *The Journal of Experimental Education*, *32*(2), 147–157.

Macaruso, P., Hook, P.E. & McCabe, R. (2006). The efficacy of computer-based supplementary phonics programs for advancing reading skills in at-risk elementary students. *Journal of Research in Reading, 29* (2), 162-172. doi: 10.1111/j.1467-9817.2006.00282.x

Marinellie, S. A., & Chan, Y. L. (2006). The effect of word frequency on noun and verb definitions: A developmental study. *Journal of Speech, Language, and Hearing Research*, *49* (5), 1001-1021. doi: 10.1044/1092-4388(2006/072)

Mason, O., & Verwoerd, M. (2007). Graph theory and networks in biology. *IET Systems Biology*, *1* (2), 89-119.

May, P. (2002). *Hamburger Schreibprobe 1-10*. Stuttgart: Ernst Klett Verlag.

May, P. (2012). *Hamburger Schreibprobe 1-10*. Stuttgart: Ernst Klett Verlag.

Mayer, A. (2016). TEPHOBE. Test zur Erfassung der phonologischen Bewusstheit und der Benennungsgeschwindigkeit. München: Reinhardt.

Mayne, A. M., Yoshinaga-Itano, C., Sedey, A. L., & Carey, A. (1998). Expressive vocabulary development of infants and toddlers who are deaf or hard of hearing. *The Volta Review*, *100* (5), 1-28.

Mayringer, H. & Wimmer, H. (2003). *Salzburger Lese-Screening für die Klassenstufe 1-4*. Bern: Hans Huber.

McGregor, K. K., Oleson, J., Bahnsen, A. & Duff, D. (2013). Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. *International Journal of Language & Communication Disorders, 48* (3), 307-319. doi: 10.1111/1460-6984.12008

McLean, S., Kramer, B. & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research*, *19*, 741–760. doi: 10.1177/1362168814567889

McNamara, T. (1996). *Measuring second language performance*. Harlow: Addison Wesley Longman.

Merea, P. & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, *4*, 142–154.

Meara, P. & Wolter, B. (2004). V_Links: beyond vocabulary depth. *Angles on the English-speaking World, 4,* 85-96.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*, 741–749.

Metsala, J. L. (1997). An examination of word frequency and neighborhood density in the development of spoken word recognition. *Memory & Cognition, 25 (1)*, 47-56.

Miestamo, M. (2017). Linguistic Diversity and Complexity. *Lingue e Linguaggio, 106* (2), 227-253.

Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, *38* (11), 39-41. http://web2.uwindsor.ca/wordmine

Miller, G. A. (1999). On knowing a word. *Annual Review of Psychology, 50*, 1-19.

Mislevy, R. J. & Stocking, M. L. (1989). A consumer's guide to LOGIST and BILOG. *Applied Psychological Measurement*, *13*, 57–75.

Mochida, K. & Harrington, M. (2006). The yes/no test as a measure of receptive vocabulary knowledge. *Language Testing*, *23*, 73–98. doi: 10.1191/0265532206lt321oa

Mol, S. E. & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin, 137* (2), 267-296. doi: 10.1037/a0021890

Monsell, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner & G. W. Humphreys (Eds.)*, Basic processes in reading: Visual word recognition*. Hillsdale, NJ: Erlbaum.

Morrow, L. M. & Weinstein, C. S. (1986). Encouraging Voluntary Reading: The Impact of a Literature Program on Children's Use of Library Centers. *Reading Research Quarterly, 21* (3), 330-346. doi: 10.2307/747713

Mulatti, C., Reynolds, M. G. & Besner, D. (2006). Neighborhood Effects in Reading Aloud: New Findings and New Challenges for Computational Models. Journal of Experimental Psychology: *Human Perception and Performance, 32* (4), 799-810, doi: 10.1037/0096-1523.32.4.799

Muter, V., Hulme, C., Snowling, M. J. & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology*, *40*, 665–681. doi:10.1037/0012-1649.40.5.665

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning Words from Context. *Reading Research Quarterly*, *20*(2), 233–253.

Nagy, W. E., Anderson, R. C. & Herman, P. A. (1987). Learning Word Meanings from Context during normal Reading. *American Educational Research Journal, 24* (2), 237-270.

Naigles, L. R., & Hoff-Ginsberg, E. (1998). Why are some verbs learned before other verbs? Effect of input frequency and structure on children's early verb use. *Journal of Child Language*, *25*, 95–120.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(1), 9-13.

Nation, P. (1993a). Using dictionaries to estimate vocabulary size: Essential, but rarely followed, procedures. *Language Testing*, *10*(1), 27–40. doi: 10.1177/026553229301000102

Nation, P. (1993b). Vocabulary Size, Growth, and Use. In R. Schreuder & B. Weltens (Eds.), *The Bilingual Lexicon* (pp. 115–135). Amsterdam: John Benjamins Publishing Co.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63* (1), 59-82.

Nation, P. (2012). Measuring vocabulary size in an uncommonly taught language. *International Conference on Language Proficiency Testing in the Less Commonly Taught Languages, Bangkok, Thailand*,

Neely, J. H. (2012). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In *Basic processes in reading* (pp. 272-344). Routledge.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1999). *The University of South Florida word association norms.*

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, *98*(3), 554–566. doi:10.1037/0022-0663.98.3.554

Ouellette, G. & Beers, A. (2010). A not-so-simple view of reading: How oral vocabulary and visual-word recognition complicate the story. *Reading and Writing*, *23*, 189–208. doi: 10.1007/s11145-008-9159-1

Pearson, P. D., Hiebert, E. H. & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly*, *42*, 282–298. doi: 10.1598/RRQ.42.2.4

Pellicer-Sánchez, A. & Schmitt, N. (2012). Scoring yes-no vocabulary tests: Reaction time vs. nonword approaches. *Language Testing*, *29*, 489–509. doi: 10.1177/0265532212438053

Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific Studies of Reading, 11*(4), 357–383. doi: 10.1080/10888430701530730

Perfetti, C. A. & Hart, L. A. (2002). The lexical quality hypothesis. *Precursors of Functional Literacy*, *11*, 67–86.

Perfetti, C. A., Wlotko, E. W. & Hart, L. A. (2005). Word Learning and Individual Differences in Word Learning Reflected in Event-Related Potentials. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31 (6). 1281-1292. doi: 10.1037/0278-7393.31.6.1281

Perfetti, C. A. & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*, 22–37. doi:10.1080/10888438.2013.827687

Pfost, M., Hattie, J., Dorfler, T., & Artelt, C. (2014). Individual Differences in Reading Development: A Review of 25 Years of Empirical Research on Matthew Effects in Reading. *Review of Educational Research*, *84*(2), 203–244. doi: 10.3102/0034654313509492

Pollatsek, A., Perea, M., & Binder, K. S. (1999). The effects of "neighborhood size" in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(4), 1142–1158. doi:10.1037//0096-1523.25.4.1142

Pregel, D., & Rickheit, G. (1987). *Der Wortschatz im Grundschulalter: Häufigkeitswörterbuch zum verbalen, substantivischen und adjektivischen Wortgebrauch*. Hildesheim: Georg Olms Verlag.

Purpura, D. J., Hume, L. E., Sims, D. M. & Lonigan, C. J. (2011). Early literacy and early numeracy: The value of including early literacy skills in the prediction of numeracy development. *Journal of Experimental Child Psychology, 110*, 641-658. doi:10.1016/j.jecp.2011.07.004

R Core Team. (2015). R: A Language and Environment for Statistical Computing: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Regier, T. (2005). The Emergence of Words: Attentional Learning in Form and Meaning. *Cognitive Science, 29*, 819-865.

Reif, M. (2014). PP. Retrieved from https://github.com/manuelreif/PP

Rescorla, L., Mirak, J. & Singh, L. (2000). Vocabulary Growth in Late Talkers: Lexical Development from 2;0 to 3;0. *Journal of Child Language, 27* (2), 293-311, doi: 10.1017/S030500090000413X

Richter, T., Isberner, M.-B., Naumann, J., & Neeb, Y. (2013). Lexical Quality and Reading Comprehension in Primary School Children. *Scientific Studies of Reading, 17*(6), 415–434. doi:10.1080/10888438.2013.764879

Ricketts, J., Nation, K. & Bishop, D. V. M. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading*, *11*, 235–257. doi: 10.1080/10888430701344306

Rizopoulos, D. (2006). ltm: An R package for Latent Variable Modelling and Item Response Theory Analyses. *Journal of Statistical Software, 17* (5), 1-25.

Rodrigo, V., Greenberg, D. & Segal, D. (2014). Changes in Reading Habits by Low Literate Adults through Extensive Reading. Reading in a Foreign Language, 26 (1), 73-91.

Roedder John, D., & Sujan, M. (1990). Age differences in product categorization. *Journal of Consumer Research*, *16*(4), 452-460.

Roget, P. M. (1911). *Roget's Thesaurus of English Words and Phrases.* TY Crowell Company.

Rost, J. (1999). Was ist aus dem Rasch-Modell geworden? *Psychologische Rundschau, 50*, 140-156. doi: 10.1026//0033-3042.50.3.140

Rott, S. (1999). The Effect of Exposure Frequency on Intermediate Language Learners' Incidental Vocabulary Acquisition and Retention through Reading. *Studies in Second Language Acquisition, 21*(4), 589-619.

Rowe, D. C., Jacobson, K. C., & Van den Oord, E. J. (1999). Genetic and environmental influences on vocabulary IQ: Parental education level as moderator. *Child Development*, *70* (5), 1151-1162.

Rowe, M., Raudenbush, S. W., Goldin-Meadow, S. (2012). The Pace of Vocabulary Growth helps predict Later Vocabulary Skill. *Child Development, 83* (2), 508-525. doi: 10.1111/j.1467-8624.2011.01710.x

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, *52*(3), 1059-1069.

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, *52*(3), 1059-1069.

Ruoff, A. (1981). *Häufigkeitswörterbuch gesprochener Sprache: Gesondert nach Wortarten, alphabethisch, rückläufig alphabetisch und nach Häufigkeit geordnet*. Berlin: Walter de Gruyter.

Sánchez-Casas, R., Ferré, P., García-Albea, J. & Guasch, M. (2006) The nature of semantic priming: Effects of the degree of semantic similarity between primes and targets in *Spanish, European Journal of Cognitive Psychology, 18* (2), 161-184, doi: 10.1080/09541440500183830

Schmid, H., Fitschen, A. & Heid, U. (2004). SMOR: A German Computational Morphology Covering Derivation, Composition, and Inflection. *Proceedings of the IVth International Conference on Language Resources and Evaluation (LREC 2004)*, 1263-1266.

Schmitt, N. (2014). Size and Depth of Vocabulary Knowledge: What the Research Shows. *Language Learning, 64* (4), 913-951. doi: 10.1111/lang.12077

Schroeder, S., Würzner, K.-M., Heister, J., Geyken, A. & Kliegl, R. (2015a). childLex: A lexical database of German read by children. *Behavior Research Methods*, *47*, 1085–1094. doi: 10.3758/s13428-014-0528-1

Schroeder, S., Würzner, K.-M., Heister, J., Geyken, A. & Kliegl, R. (2015b). childLex: Eine lexikalische Datenbank zur Schriftsprache für Kinder im Deutschen. *Psychologische Rundschau, 66,* 155-165. doi: 10.1026./0033-3042/a000275

Schroeder, S., Segbers, J., & Schröter, P. (2016). Der Kinder-Titelrekognitionstest (K-TRT). *Diagnostica, 62,* 16-30. doi: 10.1026/0012-1924/a000131

Schröter, P. & Schroeder, S. (2017). The developmental lexicon project: A behavioral database to investigate visual word recognition across the lifespan. *Behavior Research Methods, 47*, 2183–2203. doi: 10.3758/s13428-016-0851-9

Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(4), 876–900. doi:10.1037//0096-1523.21.4.876

Seashore, R. H., & Eckerson, L. D. (1940). The measurement of individual differences in general english vocabularies. *The Journal of Educational Psychology*, *31*(1), 14–38.

Segbers, J. & Schroeder, S. (2017). How many words do children know? A corpus-based estimation of children's total vocabulary size. *Language Testing, 34*, 297–320. doi: 10.1177/0265532216641152

Shillaw, J. (1996). The application of Rasch modelling to yes/no vocabulary tests. *Vocabulary Acquisition Research Group*, University of Wales Swansea.

Siew, C. S. (2018). The orthographic similarity structure of English words: Insights from network science. *Applied Network Science*, *3*(1), 13.

Siew, C. S. Q. (2013). Community structure in the phonological network. *Frontiers in psychology, 4*, 1-17. doi: 10.3389/fpsyg.2013.00553

Siew, C. S., & Vitevitch, M. S. (2016). Spoken word recognition and serial recall of words from components in the phonological network. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(3), 394-411. doi: 10.1037/xlm0000139

Smith, M. K. (1941). Measurement of the size of general English vocabulary through the elementary grades and high school. *Genetic Psychology Monographs*, *24*, 311–345.

Staehr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisiton, 31*, 577-605. Doi: 10.1017/S0272263109990039

Stamer, M. K., & Vitevitch, M. S. (2012). Phonological similarity influences word learning in adults learning Spanish as a foreign language. *Bilingualism*, *15*(3), 490–502. doi: 10.1017/S1366728911000216

Stanovich, K. E. (2009). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Journal of Education*, *189*(1-2), 23-55.

Stanovich, K. E. & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Reasearch Quarterly, 24* (4), 402-433.

Stella, M., Beckage, N. M., Brede, M., & De Domenico, M. (2018). Multiplex model of mental lexicon reveals explosive learning in humans. *Scientific reports*, *8*(1), 2259.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cognitive Science*, *29*(1), 41–78. doi: 10.1207/s15516709cog2901_3

Stodden, R. A., Roberts, K. D., Takahashi, K., Park, H. J., Stodden, N. J. (2012). Use of Text-to speech Software to Improve Reading Skills of High School Struggling Readers. *Procedia Computer Science, 14*, 359-362. doi: 10.1016/j.procs.2012.10.041

Storkel, H. L. (2004). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics*, *25*, 201–221.

Szagun, G., Steinbrink, C., Franik, M., & Stumper, B. (2006). Development of vocabulary and grammar in young German-speaking children assessed with a German language development inventory. *First Language, 26*(3), 259–280. doi: 10.1177/0142723706056475

Tamis-LeMonda, C. S., Shannon, J. D., Cabrera, N. J., & Lamb, M. E. (2004). Fathers and mothers at play with their 2-and 3-year-olds: Contributions to language and cognitive development. *Child development*, *75* (6), 1806-1820.

Tannenbaum, K. R., Torgesen, J. K., & Wagner, R. K. (2006). Relationships Between Word Knowledge and Reading Comprehension in Third-Grade Children. *Scientific Studies of Reading*, *10*(4), 381–398. doi: 10.1207/s1532799xssr1004_3

Torgesen, J. K., & Bryant, B. R. (2004). *Test of phonemic awareness.* Pro-Ed.

Trautwein, J. & Schroeder, S. (2018). WOR-TE: Ein Ja/Nein Wortschatztest für Kinder verschiedener Altersgruppen. Entwicklung und Validierung basierend auf dem Rasch-Modell. doi: 10.1026/0012-1924/a000212

Trautwein, J. & Schroeder, S. (2018). Orthographic Similarities in the Developing Mental Lexicon. Insights from Graph Theory and Implications for Orthographic Development. *Frontiers in Psychology*, 9, 2552. doi: 10.3389/fpsyg.2018.02252

Topping, K. (2015). *What Kids Are Reading: The Book-Reading Habits of Students in British Schools 2015*. London: Renaissance Learning.

Van Wijk, B. C., Stam, C. J., & Daffertshofer, A. (2010). Comparing brain networks of different size and connectivity density using graph theory. *PloS one*, *5*(10), e13701.

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics, 22*, 217-234.

Vicentini, A. (2003). The economy principle in language. *Notes and Observations from early modern english grammars. Mots, Palabras, Words*, *3*, 37-57.

Vitevitch, M. S. (2008). What Can Graph Theory Tell Us About Word Learning and Lexical Retrieval? *Journal of Speech, Language, and Hearing Research*, *51*(2), 408–422. doi: 10.1044/1092-4388(2008/030)

Vitevitch, M. S., & Storkel, H. L. (2012). Examining the Acquisition of Phonological Word Forms with Computational Experiments. *Language and Speech*, *56*(4), 493–527. doi: 10.1177/0023830912460513

Vitevitch, M. S., & Goldstein, R. (2014). Keywords in the mental lexicon. *Journal of Memory and Language*, *73*, 131–147. doi:10.1016/j.jml.2014.03.005

Walker, D., Greenwood, C., Hart, B., & Carta, J. (1994). Prediction of school outcomes based on early language production and socioeconomic factors. *Child development*, *65* (2), 606-621.

Walley, A. C., Metsala, J. L. & Garlock, V. M. (2003). Spoken vocabulary growth: Its role in the development of phoneme awareness and early reading ability. *Reading and Writing: An Interdisciplinary Journal, 16*, 5-20.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *Nature*, *393*(6684), 440–442.

Waxman, S., & Gelman, R. (1986). Preschoolers' use of superordinate relations in classification and language. *Cognitive Development*, *1* (2), 139-156.

Weiß, R. H. (2006). *Grundintelligenztest Skala 2 - Revision mit Wortschatztest und Zahlenfolgentest - Revision*. Göttingen: Hogrefe.

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L. & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology, 102*, 43–53. doi: 10.1037/a0016738

Williams, K. T. (2007). *Expressive Vocabulary Test, Second Edition: EVT-2*. London: Pearson.

Yap, M. J., Balota, D. A., Sibley, D. E. & Ratcliff, R. (2012). Individual Differences in Visual Word Recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance, 38* (1), 53-79. doi: 10.1037/a0024177

Yarkoni, T., Balota, D. & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15* (5), 971-979. Doi:10.3758/PBR.15.5.971

Yates, M. (2005). Phonological neighbors speed visual word processing: evidence from multiple tasks. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(6), 1385–1397. doi:10.1037/0278-7393.31.6.1385

Yen, W. M. (1984). Effect of local item dependence on fit and equating performance of the Three-Parameter Logistic Model. *Applied Psychological Measurement, 8*, 125–145.

Ziegler, J. C. & Goswami, U. (2005). Reading Acquisition, Developmental Dyslexia, and Skilled Reading Across Languages: A Psycholinguistic Grain Size Theory. *Psychological Bulletin, 131* (1), 3-29. doi: 10.1037/0033-2909.131.1.3

Ziegler, J. C., Muneaux, M., & Grainger, J. (2003). Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language*, *48* (4), 779-793. doi: 10.1016/S0749-596X(03)00006-8

Ziegler, J. C. , & Muneaux, M. (2007). Orthographic facilitation and phonological inhibition in spoken word recognition: A developmental study. *Psychonomic Bulletin & Review, 14*(1), 75–80.

Zortea, M. & Fumagalli de Salles, J. (2012). Semantic word association: comparative data for Brazilian children and adults. *Psychology & Neuroscience, 5*, (1), 77-81. doi: 10.3922/j.psns.2012.1.10

Zortea, M., Menegola, B., Villavicencio, A., & Salles, J. F. d. (2014). Graph Analysis of Semantic Word Association among Children, Adults, and the Elderly. *Psicologia: Reflexao e Critica*, *27*(1), 90–99.

V APPENDIX

APPENDIX A1: Items of the yes/no vocabulary test in Study 1

Items. Frequenzen (Lemma Log, childLex 0.15.01) sind in Klammern angegeben; kursiv markiert sind die Wörter, die laut Analyse vom Test ausgeschlossen werden sollten.

Ankeritems (*N* = 20):

taub [1.04], rümpfen [1.02], polieren [1.01], schroff [1.00], beichten [0.40], Reparatur [0.40], originell [0.39], rotieren [0.39], Statist [-0.04], *wähnen* [-0.04], Prozession [-0.09], Schlot [-0.09], *konstatieren* [-0.39], porös [-0.52], reffen [-0.52], *Reuse* [-0.39], ignorant [-0.69], *Tresse* [-0.69], ostentativ [-0.99], sedieren [-0.99]

Version 1./2. Klasse (*N* = 70):

nicken [2.70], *starren* [2.51], *passieren* [2.46], Sorge [2.15], deuten [2.10], zufrieden [2.04], heftig [2.03], knurren [1.99], Atem [1.97], Insel [1.96], sinken [1.95], nervös [1.92], Pfote [1.89], keuchen [1.89], eng [1.87], furchtbar [1.85], ähnlich [1.82], Mühe [1.80], Kerl [1.79], eilig [1.76], finster [1.75], wirbeln [1.71], *Gegend* [1.67], flattern [1.63], Stute [1.62], *reagieren* [1.58], recken [1.53], *Planet* [1.51], *Stapel* [1.48], eisig [1.48], *Museum* [1.48], *Schüssel* [1.46], falten [1.46], konzentrieren [1.45], beeindrucken [1.44], Knall [1.43], bequem [1.43], Gehirn [1.42], behutsam [1.41], spöttisch [1.40], empor [1.39], *Backe* [1.39], dröhnen [1.39], kompliziert [1.37], Nerv [1.37], fies [1.35], beben [1.34], Gerät [1.33], Bucht [1.33], Notiz [1.32], schielen [1.30], kahl [1.30], schmunzeln [1.30], *Statue* [1.29], grob [1.28], genial [1.28], *Strahl* [1.28], *ignorieren* [1.25], köstlich [1.23], zappeln [1.23], Idiot [1.23], knipsen [1.21], Kelch [1.20], *grell* [1.19], japsen [1.19], Amulett [1.18], flehen [1.18], zünden [1.16], Nüstern [1.16], drängeln [1.15]

Überschneidung der Versionen 1./2. Klasse und 3./4. Klasse (*N* = 10):

glotzen [1.15], *knirschen* [1.14], Kurs [1.12], scheußlich [1.11], unterirdisch [1.11], Ader [1.09], hüten [1.09], eignen [1.07], skeptisch [1.06], hämisch [1.06]

Version 3./4. Klasse (*N* = 60):

Reling [0.86], enorm [0.86], Kittel [0.86], Lektion [0.85], Erbe [.84], nüchtern 0.82], Sohle [0.81], Sirene [0.81], hager [0.80], dulden [0.79], empören [0.79], Vernunft [0.79], bange [0.78], *artig* [0.77], ruinieren [0.77], sympathisch [0.77], aktivieren [0.75], Salbe [0.73], kritisch [0.71], brodeln [0.70], Hieb [0.70], listig [0.70], Wildnis [0.70], Humor [0.69], Rumpf

[0.68], Schwur [0.68], Gurgel [0.67], Existenz [0.66], keifen [0.66], tückisch [0.65], raunzen [0.64], spreizen [0.64], Klinik [0.63], morsch [0.62], schuften [0.62], rumoren [0.62], Wade [0.62], zieren [0.61], lotsen [0.61], Prärie [0.61], bugsieren [0.60], karg [0.60], anonym [0.60], inspizieren [0.57], hecheln [0.57], leugnen [0.56], wabern [0.56], brenzlig [0.56], büßen [0.55], Charakter [0.55], Instinkt [0.55], Auktion [0.54], fahrig [0.54], Attacke [0.53], dreist [0.53], kombinieren [0.52], relativ [0.52], Ahne [0.50], Dialog [0.50], irritieren [0.47]

Überschneidung der Versionen 3./4. Klasse und 5./6. Klasse (*N* = 10):

berüchtigt [0.48], flankieren [0.48], gütig [0.47], streben [0.47], schlaksig [0.45], welk [0.45], perplex [0.44], Zeremonie [0.44], Identität [0.42], salutieren [0.42]

Version 5./6. Klasse (*N* = 70):

montieren [0.37], Wams [0.35], grotesk [0.33], naiv [0.31], Semester [0.29], *äsen* [0.26], *Spind* [0.26], analysieren [0.24], drapieren [0.24], *brüsk* [0.21], Reklame [0.18], Hektik [0.15], imponieren [0.12], Antlitz [0.09], kurieren [0.09], elastisch [0.05], Pose [0.05], quittieren [0.05], weilen [0.05], appellieren [0.01], Euphorie [0.01], Wabe [0.01], zentral [0.01], akut [-0.04], Anstalt [-0.04], Apparatur [-0.04], argumentieren [-0.04], assistieren [-0.04], Dimension [-0.04], *Fanfare* [-0.04], *Galosche* [-0.04], Hobel [-0.04], konfus [-0.04], meucheln [-0.04], Monokel [-0.04], Pflug [-0.04], rebellieren [-0.04], Schliere [-0.04], Schote [-0.04], *schwelen* [-0.04], *süffisant* [-0.04], triezen [-0.04], Amnestie [-0.09], effektiv [-0.09], Flaum [-0.09], *Häme [-0.09]*, Insasse [-0.09], Kandis [-0.09], *schartig* [-0.09], souverän [-0.09], akribisch [-0.15], deichseln [-0.15], Devise [-0.15], effizient [-0.15], exzentrisch [-0.15], abstrus [-0.22]], Inbrunst [-0.22], mental [-0.22], heroisch [-0.22], beseelen [-0.04], famos [-0.04], inszenieren [-0.04], Moral [-0.04], scheckig [-0.04], demoliert [- 0.09], galant [-0.09], *graziös* [-0.09], *imitieren* [-0.09], kapitulieren [-0.15], *sengen* [-0.22]

Pseudowörter

verglauben, luben, Ragel, nickisch, Werbel, Runte, schwich, neidig, Fristerei, arben, Bicher, beräkeln, blumisch, Sift, klogen, Führtum, plestisch, Elge, hasteln, friedhaft, weppen, dinn, leufern, hörisch

APPENDIX A.2:

TABLE 8.2:

Lexicon sizes and network measures in different age groups for lemmas from Study 3.

	M Lexicon Size	Network Measures *M (SD)* **				
Grade	Lemmas	*n*	*<k>*	*L*	*D*	*C*
1	5925	1922	4.79	10.81	32.00	.61
		(25)	(0.14)	(1.09)	(4.18)	(.04)
2	6097	1980	4.81	10.79	31.84	.61
		(26)	(0.14)	(0.98)	(3.41)	(.05)
3	11182	3758	5.50	9.94	30.02	.64
		(41)	(0.13)	(0.35)	(2.71)	(.04)
4	14819	5021	5.87	9.57	30.28	.65
		(48)	(0.11)	(0.22)	(2.62)	(.03)
5	18812	6378	6.21	9.39	31.68	.66
		(52)	(0.13)	(0.20)	(3.11)	(.03)
6	25694	8651	6.67	9.16	33.58	.66
		(62)	(0.10)	(0.20)	(4.74)	(.02)
8	38029	12589	7.12	9.09	38.02	.63
		(61)	(0.06)	(0.27)	(6.02)	(.01)

APPENDIX A.3: Erklärungen der Promovendin zum eigenen Anteil an den vorgelegten wissenschaftlichen Abhandlungen

WISSENSCHAFTLICHE ABHANDLUNG 1

TITEL	WOR-TE: Ein Ja/Nein-Wortschatztest für Kinder verschiedener Altersgruppen. Entwicklung und Validierung basierend auf dem Rasch Modell
AUTOREN	Jutta Trautwein, Sascha Schroeder
JOURNAL	Diagnostica
	PUBLIKATIONSSTATUS
( )	Nicht eingereicht
( )	Eingereicht
( )	In Begutachtung
( )	Angenommen
(x)	Veröffentlicht / Publikationsjahr: 2018 (online first)

Zur Entwicklung dieses Artikels habe ich in folgender Art und Weise beigetragen:

- Mitwirkung an der Konzeption und Fragestellung
- Klärung der Umsetzung und des statistischen Modells
- Eigenständige Rekrutierung und Akquise von Probanden
- Eigenständige Durchführung der Untersuchung
- Eigenständige Aufbereitung der Daten
- Analyse der Daten und Interpretation der Ergebnisse in Absprache
- Niederschrift und Überarbeitung des Artikels in Absprache
- Revision des Manuskripts und Beantwortung der Gutachterkommentare in Absprache

____________________________ ____________________________

Jutta Trautwein Sascha Schroeder

WISSENSCHAFTLICHE ABHANDLUNG 2

TITEL	How many words do children know? A corpus-based estimation of children's lexicon size.
AUTOREN	Jutta Segbers (now Trautwein), Sascha Schroeder
JOURNAL	Language Testing 34 (3)
	PUBLIKATIONSSTATUS
( )	Nicht eingereicht
( )	Eingereicht
( )	In Begutachtung
( )	Angenommen
(x)	Veröffentlicht / Publikationsjahr: 2017

Zur Entwicklung dieses Artikels habe ich in folgender Art und Weise beigetragen:

- Mitwirkung an der Konzeption und Fragestellung
- Klärung der Umsetzung und des statistischen Modells
- Eigenständige Rekrutierung und Akquise von Probanden
- Eigenständige Durchführung der Untersuchung
- Eigenständige Aufbereitung der Daten
- Analyse der Daten und Interpretation der Ergebnisse in Absprache
- Niederschrift und Überarbeitung des Artikels in Absprache
- Revision des Manuskripts und Beantwortung der Gutachterkommentare in Absprache

____________________________ ____________________________

Jutta Trautwein Sascha Schroeder

WISSENSCHAFTLICHE ABHANDLUNG 3

TITEL	Orthographic Similarities in the Developing Mental Lexicon. Insights from Graph Theory and Implications for Orthographic Development.
AUTOREN	Jutta Trautwein, Sascha Schroeder
JOURNAL	Frontiers in Psychology, 2552 (9)
	PUBLIKATIONSSTATUS
( )	Nicht eingereicht
( )	Eingereicht
( )	In Begutachtung
()	Angenommen
(x)	Veröffentlicht / Publikationsjahr: 2018

Zur Entwicklung dieses Artikels habe ich in folgender Art und Weise beigetragen:

- Mitwirkung an der Konzeption und Fragestellung
- Klärung der Umsetzung und des statistischen Modells
- Eigenständige Rekrutierung und Akquise von Probanden
- Eigenständige Durchführung der Untersuchung
- Eigenständige Aufbereitung der Daten
- Analyse der Daten und Interpretation der Ergebnisse in Absprache
- Niederschrift und Überarbeitung des Artikels in Absprache
- Revision des Manuskripts und Beantwortung der Gutachterkommentare in Absprache

____________________________ ____________________________

Jutta Trautwein Sascha Schroeder

APPENDIX A.4: Allgemeine Erklärung

1. Ich erkläre, an keiner anderen Hochschule ein Promotionsverfahren eröffnet zu haben.
2. Ich erkläre, dass die Dissertation in der gegenwärtigen Fassung keiner anderen Hochschule zur Begutachtung vorgelegen hat oder vorliegt.
3. Ich erkläre, dass die Arbeit selbstständig und ohne Hilfe Dritter verfasst wurde und bei der Abfassung alle Regelungen guter wissenschaftlicher Standards eingehalten wurden.

__________________________ _______________________________________

Datum Unterschrift

APPENDIX A.5: Curriculum Vitae

Der Lebenslauf ist in der Onlineversion aus Gründen des Datenschutzes nicht enthalten.