Vocabulary knowledge and learning:

Individual differences in adult native speakers

ISBN: 978-90-76203-98-0

Cover picture: Drawing by Lisa Ihle Mainz.

Printed and bound by Ipskamp Drukkers b.v.

Vocabulary knowledge and learning: Individual differences in adult native speakers

Proefschrift
ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van de decanen,
in het openbaar te verdedigen op maandag 22 oktober 2018
om 12.30 uur precies

door
Nina Mainz
geboren op 1 januari 1991
te Aachen (Duitsland)

Promotor:

Prof. dr. A.S. Meyer

Copromotoren:

Dr. Z. Shao (MPI for Psycholinguistics, Nijmegen)

Dr. A.C. Smith (MPI for Psycholinguistics, Nijmegen)

Manuscriptcommissie:

Prof. dr. A.P.A. Roelofs

Prof. dr. J.M. McQueen

Prof. dr. M. Damian (University of Bristol, Verenigd Koninkrijk)

This research was supported by the The Max Planck Society for the Advancement of Science, Munich, Germany.

Table of Contents

1 | General introduction

Knowing the words of a language is an essential aspect of an individual's command of their language. The number of words that speakers of a language know is quite impressive. An average 20-year-old native speaker of American English is estimated to know 42,000 lemmas, i.e. uninflected word forms from which all inflections are derived (Brysbaert, Stevens, Mandera, & Keuleers, 2016b). However, word learning does not stop at some point in early adulthood. On the contrary, vocabulary size tends to improve with ageing and across the entire lifespan (e.g. Brysbaert et al., 2016b; Keuleers, Stevens, Mandera, & Brysbaert, 2015; Schroeder & Salthouse, 2004). Thus, over the years, humans are exposed to, continue to learn, and importantly also remember, an impressive number of words. It has been estimated that an average adult learns approximately 6,000 new lemmas between the ages of 20 and 60 years. That means in these 40 years, on average one new lemma is learned every two days (Brysbaert et al., 2016b; Keuleers et al., 2015).

Hence, a speaker's vocabulary is an ever-growing inventory of words to describe the world around her. Assuming that vocabulary is a collection of expressions to describe all kinds of experiences and information that are relevant to a specific speaker and her community, it might be expected that there is variation in which and how many words people know. Specialist vocabulary is a very intuitive example for the strong effects of experience on the types of words known by different individuals. The words *syncopation* or *lunge* are, for instance, known by individuals with certain occupations, interests, or hobbies, and completely opaque to others.[1] Another aspect of what we might very broadly refer to as *experience* or perhaps *skills* is education. It has been shown that educational level significantly affects vocabulary size across the entire age range, with higher educational levels being associated with considerably greater vocabulary knowledge.[2] This likely reflects effects of extensive reading and studying, which are

[1]In this case, *syncopation* is a word known by musicians or perhaps also dancers and it refers to a change to the rhythm of a piece of music where the stressed beats become unstressed and the unstressed ones become stressed. *Lunge* is a dance movement where one leg is in a bent position and the other leg is extended.

[2]I acknowledge that the terms *vocabulary size*, *vocabulary knowledge*, and *vocabulary* refer to different but closely related and conflated constructs. While *vocabulary size* refers to the number of different words in a *vocabulary*, with the latter referring to the set of words an individual knows, *vocabulary*

associated with formal education, on vocabulary growth (Brysbaert et al., 2016b; Keuleers et al., 2015).

Vocabulary size affects language processing

Hence, there are considerable individual differences in vocabulary size, important predictors of which are age and educational level. As the knowledge of words is assumed to be an essential aspect of an individual's language capacity, it is plausible to expect that individual differences therein affect language processing.

Previous research has indeed indicated that larger vocabularies are associated with more accurate and importantly also with faster language processing in both comprehension and production. In healthy older adults, for instance, accuracy of spoken word recognition (Janse & Jesse, 2014) and the use of predictive information in spoken contexts (Federmeier, McLennan, Ochoa, & Kutas, 2002) have been shown to benefit from greater vocabulary knowledge. Furthermore, large vocabulary size was associated with better listening comprehension in young adults (Andringa, Olsthoorn, van Beuningen, Schoonen, & Hulstijn, 2012). Speech recognition in suboptimal conditions was also found to be better for individuals with larger vocabularies (Bent, Baese-Berk, Borrie, & McKee, 2016). Furthermore, Yap, Balota, Sibley, and Ratcliff (2012) found higher vocabulary scores to be associated with more accurate and faster word recognition in lexical decision and speeded pronunciation tasks.

In addition, increased vocabulary knowledge was found to be associated with faster language production. In a study on healthy older adults, reaction times in various linguistic tasks, including verbal fluency and picture naming, were predicted by vocabulary knowledge. Individuals with larger vocabularies were overall faster at processing language (Rodriguez-Aranda & Jakobsen, 2011). In verbal fluency tasks participants are asked to produce as many words as possible from a given semantic category (category fluency) or starting with a given letter (letter fluency) within one minute. Unsworth, Spillers, and Brewer (2011) found better vocabulary knowledge to be related to more items generated in verbal fluency tasks. Similarly, Shao, Janse, Visser, and Meyer (2014) found that in letter as well as category fluency tasks, individuals with greater vocabularies were faster to initiate their response to the cue than individuals with weaker word knowledge.

To sum up, larger vocabularies have been associated with more accurate and faster language processing in various production and comprehension tasks. An open question is what the reason is for this vocabulary benefit. It might be somewhat counterintuitive that

knowledge refers to what is known about the words in a vocabulary. In the following I will use these terms interchangeably referring to what can be measured using vocabulary tests and what is assumed to be a combination of the underlying components size and knowledge.

retrieving lexical items from a larger vocabulary is faster than from a smaller vocabulary. As the lexicon becomes larger, more words might compete for selection, which would render lexical selection more difficult, hence slower and maybe less accurate (Diependaele, Lemhöfer, & Brysbaert, 2013).

Aside from faster reaction times in language processing tasks, previous research has found smaller word frequency effects for speakers and readers with larger vocabularies (Diependaele et al., 2013; Yap, Tse, & Balota, 2009) or more reading experience (Chateau & Jared, 2000; Kuperman & Van Dyke, 2013). Reaction times (RTs) typically decrease with increasing word frequency and this effect has been found to be smaller for individuals with larger vocabularies, meaning that the reaction time difference between low- and high-frequency words decreases with increasing vocabulary size. Brysbaert, Lagrou, and Stevens (2016a), for example, observed that individuals with greater vocabulary knowledge were not only faster to make lexical decisions but also showed smaller word frequency effects on their language processing speed (see also Diependaele et al., 2013). This frequency by skill interaction has been taken to indicate differences in the entrenchment of lexical representations in smaller as compared to larger vocabularies. Thus, the representations in individuals with greater vocabularies have been suggested to be more *robust* or *distinct*, enabling faster processing, as compared to individuals with smaller vocabularies (Diependaele et al., 2013). This interaction between word frequency and skill has been argued to result from differences in exposure to language (Brysbaert et al., 2016a; Monaghan, Chang, Welbourne, & Brysbaert, 2017). Increased exposure has been associated with an increase in efficiency of accessing lexical representations across the entire frequency range (Monaghan et al., 2017). As a result, the lexicon of individuals with limited language exposure and therefore weaker word knowledge is hypothesised to show a stronger difference in processing efficiency between low- and high-frequency words due to less entrenched representations (see also Yap et al., 2009).

These findings on the relationship between individual differences in vocabulary size and language processing raise a few questions. First of all, one might ask what the origin for such considerable variation in vocabulary size between the native speakers of a language is. Are there factors other than age, educational level, and exposure, which can be identified as being related to variation in vocabulary learning and size? Related to that, it is unclear whether individual differences in factors beyond exposure, such as cognitive skills, affect the interaction observed between vocabulary size and word frequency. Both of these questions and potential answers to them based on previous research are described later.

Measuring vocabulary size

Another issue central to any research on variation in vocabulary size concerns the measurement of knowledge of words. Importantly, the majority of studies that have indicated beneficial effects of greater vocabulary knowledge on language processing have not focused on the examination of vocabulary size specifically. This may be the reason why often only single vocabulary tests were employed to assess word knowledge. However, is a single test sufficient to assess vocabulary, especially if the focus is on individual differences in word knowledge?

Hence, the question arising is how to best assess a skill as complex and potentially multiply-determined as vocabulary size or knowledge. Bowles and Salthouse (2008), for example, argued that it is necessary to use a variety of measures of vocabulary size, especially in studies where vocabulary knowledge is in the focus of interest. The reason is that no vocabulary test is a pure measure of word knowledge but involves other cognitive abilities as well as world knowledge (Bowles & Salthouse, 2008). Consequently, the advice is to use a battery of tests of different types (antonym vs. synonym) and formats (multiple-choice vs. open) and calculate a composite score of performance on all tests, which is suggested to likely reflect vocabulary size.

Furthermore, a lot of the established and often-used vocabulary tests are multiple-choice tests and it has been claimed that this is not the ideal way of assessing word knowledge (Gyllstad, Vilkaite, & Schmitt, 2015). In multiple-choice tests, individuals can rely on elimination strategies or guessing to arrive at the correct response whereas open tests require knowledge of the target word. Following Gylstad and colleagues (2015) this might lead to an overestimation of participants' vocabulary size when using multiple-choice instead of open or interview-based measures of vocabulary. This in turn might be problematic in individual differences studies, where it is essential that the measures used to assess participants' cognitive abilities are able to elicit variation in test performance (Kidd, Donnelly, & Christiansen, 2018). Especially when testing highly able groups of participants, which often happens in psychological and psycholinguistic research where mainly university undergraduates are tested, the use of measures that are too easy due to the possibility of relying on strategies such as elimination or guessing might be highly problematic.

Furthermore, Henriksen (1999) proposed three dimensions along which lexical competence may develop and vary. These three dimensions are (a) partial to precise knowledge, (b) depth of knowledge, and (c) receptive to productive use ability. Assuming vocabulary development and knowledge to show variation along these different dimensions supports the assumption that different measures of vocabulary are needed. The 'partial to precise knowledge' dimension refers to a continuum from mere recognition of a lexical item to precise comprehension, describing the acquisition of form-meaning mappings. Multiple-choice tests might be seen as addressing this first dimension, requiring the recognition and knowledge of single words and their meanings. Secondly, development along the 'depth of knowledge' dimension is related to the creation of links or associations between words. Thus, while the first dimension is concerned with the process of acquiring single lexical items, this second dimension refers to the development of an internal structure of the lexicon acknowledging the relationships between words. Performance on different types of tests, namely antonym and synonym tests, might be thought to reflect knowledge on the depth continuum, more precisely knowledge of the complex antonymous or synonymous relationships between different words. The third and final 'receptive to productive ability' dimension reflects the observation that words are typically first available for reception and slowly become part of the productive vocabulary, which is a continuous process. Hence, only a combination of open and multiple-choice tests will be capable of accounting for the receptive to productive continuum (Henriksen, 1999).

Individual differences in vocabulary learning

While vocabulary appears to be difficult to assess and different ideas have been put forward as to how the knowledge of words is best measured, previous studies have demonstrated large variability in vocabulary size across native speakers of a language (Brysbaert et al., 2016b). One might ask what the origins are for such considerable variation in vocabulary size among native speakers of a language. Why do some people learn more words than others, thus end up having a larger vocabulary than others? It has been shown that differences in age as well as educational level are related to considerable individual variation in vocabulary size. Closely related to these effects is, as mentioned above, the impact of amount of exposure or input on word learning. Earlier experimental and modelling work has shown that greater amounts of exposure result in more words being learned, thus, in larger vocabularies (Hurtado, Marchman, & Fernald, 2008; Monaghan et al., 2017). Variation in the type and amount of exposure has, in particular, been shown to result in considerable individual differences in the number of words children acquire. Early in development the quantity of the input appears to be more beneficial for vocabulary learning, whereas later in development the diversity of the input, i.e. the number of different words, has been shown to have stronger advantageous effects on learning (Jones & Rowland, 2017).

In addition to being affected by the environmental factor exposure, vocabulary learning and size have been suggested to be influenced by individual differences in various cognitive abilities. Better phonological short-term memory has been associated with improved word learning and greater vocabulary size in children (Gathercole, 2006). Furthermore, individual differences in processing speed as well as in vocabulary size have been found to predict variation in vocabulary learning and size. Children with faster online processing speed at 25 months showed more accelerated word learning over the first two years of life (Fernald, Perfors, & Marchman, 2006), and children with better vocabulary knowledge at 2 years of age also showed greater word knowledge at 8 years of age (Marchman & Fernald, 2008). A question arising is what the origin is of the relationship between online processing speed and vocabulary learning. One possibility might be that greater amounts of exposure early in development result in some children having larger vocabularies, and this larger amount of training on linguistic input also leads to the emergence of advantages in language processing (Fernald et al., 2006). Alternatively, variation in general cognitive abilities might underlie differences in word learning and also differences in processing abilities, with the latter being present from the beginning of language development (Fernald et al., 2006). Related to this is the question of which factors give rise to variation in vocabulary size in the first place, which in turn has effects on subsequent vocabulary learning. Again, environmental factors, such as exposure, but also cognitive factors potentially cause variation in word learning and knowledge.

Summary and thesis outline

To sum up, earlier research has suggested that there are considerable individual differences in vocabulary size among the native speakers of a language, which impact language processing performance (Brysbaert et al., 2016a, 2016b; Kidd et al., 2018; Yap et al., 2012). However, not much research has focused on variation in vocabulary and its effects on lexical processing, which is why most studies have used only single tests to measure participants' vocabulary. It has been argued, though, that vocabulary can only be assessed using a battery of different measures varying in format and test type (Bowles & Salthouse, 2008). Furthermore, the majority of what we know about the relationship between individual differences in vocabulary and language processing is based on studies where a very small and homogeneous group of participants was tested, namely mostly undergraduate university students. This is the case for psychological and psycholinguistic research more generally but raises the question of whether our findings and theories are generalisable to individuals from more diverse backgrounds (Kidd et al., 2018).

These questions and gaps in the literature were addressed in Chapters 2 and 3 of this dissertation. Two groups of participants, namely university students (Chapter 2) and vocational college students (Chapter 3), were tested on a battery of different vocabulary measures. Two of the vocabulary measures were established tests of word knowledge, while five additional tests of different types (antonym, synonym, definition) and formats (multiple-choice, open) were newly developed. In addition, participants performed a lexical decision task to measure language processing performance. The aims were (a) to examine whether a battery of vocabulary tests is indeed necessary to assess vocabulary size or whether single measures are sufficiently representative of vocabulary test performance, and (b) to study the relationship between individual differences in language processing and variation in vocabulary knowledge in our typical and a more diverse group of participants.

In Chapter 4, I present data from a picture-word interference task, which was completed by the university students group in addition to the lexical decision task. In this vein, I was able to examine whether lexical production is influenced by variation in vocabulary size to a similar extent as word recognition. This also allowed us to examine the relationship between comprehension and production in the same individuals. The production task was not administered to the vocational college students due to constraints on the test setting and the length of the experiment.

The observation of considerable variation in vocabulary among the native speakers of a language raised the question of what causes these individual differences in the knowledge of words (e.g. Brysbaert et al., 2016b). Previous research has demonstrated that an important factor leading to variation in vocabulary learning and size is exposure (Hurtado et al., 2008; Jones & Rowland, 2017; Monaghan et al., 2017). Greater exposure has been associated with better word learning. Jones and Rowland (2017) have shown that early in development, input quantity is more important for vocabulary growth whereas later in development the diversity of the input is decisive. Furthermore, different cognitive skills have been suggested to influence lexical learning. Greater phonological short-term memory, higher processing speed, and larger vocabulary size have been associated with more successful word learning (Fernald et al., 2006; Gathercole, 2006; Marchman & Fernald, 2008; McMurray, Horst, & Samuelson, 2012). However, most of what is known about individual differences in word learning is based on developmental research. Hence, one open question in this context concerns the degree to which these factors continue to play a role in word learning in adulthood. Aside from that, not much is known about the relationships between individual differences in various cognitive abilities and their effects on language learning, especially in adult native speakers (Kidd et al., 2018). Instead of focusing on single cognitive abilities and examining the roles of, for instance, phonological short-term memory, processing speed, and vocabulary separately, it is considered necessary to look at different cognitive skills and their relationships between each other and with lexical learning. Finally, not much research has focused on the relationships between individual differences in internal sources for variation, i.e. cognitive skills, and differences in environmental factors, such as exposure (Kidd et al., 2018). What is the relationship, for instance, between variation in processing speed or vocabulary size and the amount of exposure an individual needs to learn a new word?

These open questions were addressed in Chapters 5 and 6 of the present dissertation. In the experiment detailed in Chapter 5 I examined novel word learning performance in adult native speakers and how individual differences therein relate to variation in different cognitive abilities, namely general processing speed, nonverbal intelligence, phonological short-term memory, and vocabulary. In addition, I was interested in the relationship between these cognitive factors and the factors amount of exposure and overnight consolidation, which have previously been found to affect word learning. This study was conducted with university students only due to limited availability of non-university students as participants. Additionally, constraints on the length of test sessions possible in a test setting at vocational colleges did not allow us to run this study with a more diverse group of participants. In the future, research on word learning should also be extended to include participants from more diverse educational backgrounds.

In Chapter 6 I present a computational modelling study that aimed at bringing together the research topics and questions from all previous chapters. A distributed connectionist model was used to examine causes and consequences of variation in vocabulary size. More precisely, I investigated the potential causal roles of variation in cognitive skills, namely processing speed and intelligence, and the environmental factors quantity and diversity of the input in determining variation in vocabulary size. In addition, I examined the effects of differences in vocabulary size as well as in these potential causes of variation in vocabulary size on language processing and novel word learning. Thus, the findings from Chapter 6 complement the observations from my behavioural work by providing insights into causal relationships between variation in cognitive as well as environmental factors and lexical learning and processing, and by shedding light on underlying mechanisms of the observed behavioural effects.

2 | Vocabulary knowledge predicts lexical processing: Evidence from lexical decision[1]

Abstract

With this study we pursued two goals; firstly the development and assessment of measures of vocabulary size in Dutch native speakers, and secondly the investigation of the relationship between individual differences in word knowledge and language processing. Five vocabulary tests were developed, including multiple-choice and open antonym and synonym tests and a definition test, and administered together with Andringa and colleagues' (2012) receptive multiple-choice test and the PPVT-III NL (Schlichting, 2005). Language processing performance was measured using a lexical decision task. We found the typical lexicality and word frequency effects in the lexical decision task. Importantly, RTs were predicted by vocabulary size, indicating that individuals with better vocabulary knowledge are better in language processing. Scores from six out of seven vocabulary tests were significantly related with speed of language processing. Implications of our findings concerning the assessment of vocabulary size in individual differences studies and concerning future research on the role of vocabulary in language processing are discussed.

[1]This and the following chapter are based on *Mainz, N., Shao, Z., Brysbaert, M., & Meyer, A.S. (2017). Vocabulary knowledge predicts lexical processing: Evidence from a group of participants with diverse educational backgrounds. Frontiers in Psychology, 8, 1164. doi: 10.3389/fpsyg.2017.01164*

Introduction

Knowing the words of the language is undeniably an important part of a speaker's command of their language. Due to differences in life-experience, interests, and skills, adults are likely to differ considerably in the structure and size of their native language vocabularies (Brysbaert et al., 2016b). Individual differences in vocabulary size are likely to affect language processing in adult native speakers, just as it has been shown for individual differences in general cognitive abilities, such as inhibitory control (Banks, Gowen, Munro, & Adank, 2015; Shao, Roelofs, Acheson, & Meyer, 2014; Shao, Roelofs, & Meyer, 2012), sustained attention (Jongman, Roelofs, & Meyer, 2015), or working memory (Hartsuiker & Barkhuysen, 2006). However, the role of adult speakers' vocabulary knowledge in language processing has not been investigated comprehensively. The present study aimed at investigating the relationship between individual differences in vocabulary size and language processing performance more closely. For this purpose we employed a battery of seven vocabulary tests and a lexical decision task.

Most studies that have considered the relationship between vocabulary size and language processing performance found beneficial effects of increased vocabulary size on language processing performance. Studies on healthy older adults showed, for instance, that better vocabulary knowledge is beneficial for accuracy of spoken word recognition (Janse & Jesse, 2014) and the use of predictive information in spoken contexts (Federmeier et al., 2002). In young adults, an increase in vocabulary size was associated with better listening comprehension (Andringa et al., 2012). Large vocabulary size was also found to be linked to better speech recognition in suboptimal conditions (Bent et al., 2016). Furthermore, Yap, Balota, Sibley, and Ratcliff (2012) found higher vocabulary scores to be associated with more accurate and faster word recognition in lexical decision and speeded pronunciation tasks. In addition, increased vocabulary knowledge was found to be associated with faster language production. In a study on healthy older adults, reaction times (RTs) in various linguistic tasks, including verbal fluency and picture naming, were predicted by vocabulary knowledge. Larger vocabularies were related to overall faster language processing (Rodriguez-Aranda & Jakobsen, 2011). Verbal fluency tasks require participants to produce as many words as possible within a semantic category (category fluency) or starting with a given letter (letter fluency) within one minute. Unsworth, Spillers, and Brewer (2011) found better vocabulary knowledge to be related to more items generated in verbal fluency tasks. Similarly, Shao, Janse, Visser, and Meyer (2014) studied the component processes that determine older adults' performance on verbal fluency tasks. In letter as well as category fluency tasks, individuals with greater word knowledge were faster to initiate their response to the cue than individuals with weaker vocabularies.

In a nutshell, better vocabulary knowledge has been associated with advantages in various language comprehension and production tasks. This is somewhat counterintuitive as one might expect that retrieving lexical items from a larger vocabulary would be slower than retrieval from a smaller lexicon, because more lexical items might compete for selection as the lexicon becomes larger or denser (Diependaele et al., 2013). Contrary to this expectation, individuals with better word knowledge appear to be able to access their knowledge fast and efficiently, perhaps even faster than individuals with smaller vocabularies.

In addition to faster RTs in language processing tasks, previous research has found smaller word frequency effects for speakers and readers with larger vocabularies (Diependaele et al., 2013; Yap et al., 2009) or more reading experience (Chateau & Jared, 2000; Kuperman & Van Dyke, 2013). Typically, RTs decrease with increasing word frequency. This word frequency effect has been reported to be smaller for individuals with better vocabulary knowledge, meaning that the RT difference between low- and high-frequency words decreases with increasing vocabulary size. Diependaele and colleagues (2013), for instance, reanalysed data from an earlier visual word recognition study (Lemhöfer et al., 2008) and found larger vocabularies to be associated with smaller effects of word frequency. Furthermore, Brysbaert, Lagrou, and Stevens (2016a) observed that individuals with higher vocabulary scores were not only faster to make lexical decisions but also showed smaller effects of word frequency on their RTs. In both studies, this frequency by skill interaction was taken to be indicative of differences in entrenchment between smaller and larger vocabularies. Thus, the representations in individuals with greater vocabularies are assumed to be more robust or distinct, enabling faster processing, as compared to individuals with smaller vocabularies. According to this lexical entrenchment hypothesis, the frequency by skill interaction is due to differences in exposure to language, especially to written language, which has a lower type-token ratio than spoken language (Brysbaert et al., 2016a). It is assumed that amount of exposure has a particularly strong impact on the representations of low frequency words (Kuperman & Van Dyke, 2013). As a result, the lexicon of individuals with limited language exposure and therefore weaker word knowledge is hypothesised to show a stronger frequency difference between low- and high-frequency words. Hence, the vocabulary of individuals with limited exposure to language has a steeper frequency curve than what would be observed in individuals with larger vocabularies (Brysbaert et al., 2016a). Consequently, lexical representations in low-vocabulary individuals are hypothesised to be weaker, less robust or distinct, especially for low-frequency words, and therefore slower to be processed.

A similar argument has been put forward by Yap, Tse, and Balota (2009). They also hypothesised that greater vocabulary knowledge led to overall increased precision and stability of lexical representations. This was based on their observation that vocabulary knowledge affects the joint effects of word frequency and associative priming. In a lexical decision task, participants with poorer vocabulary knowledge showed stronger associative priming effects for low-frequency than for high-frequency words, whereas individuals with better vocabulary scores exhibited equally strong priming effects for both types of words. This suggests that the lexical representations in readers with greater vocabulary knowledge do not differ much in quality or strength depending on word frequency, contrary to the representations in low-vocabulary individuals, which appear to show considerable differences in strength or robustness depending on word frequency (Yap et al., 2009).

To sum up, based on the observation that word frequency effects on word recognition are smaller in high-vocabulary than in low-vocabulary individuals, structural differences between the representations in vocabularies of varying sizes have been suggested. Different researchers have used different terms to refer to this idea; representations in individuals with better word knowledge or more experience with language have been proposed to be more *robust*, *entrenched*, *precise*, or higher in *lexical quality* making lexical access faster and less prone to effects of word frequency (Diependaele et al., 2013; Perfetti & Hart, 2001; Van Dyke, Johns, & Kukona, 2014; Yap et al., 2009).

While the aforementioned studies implicate a role of vocabulary size in comprehension and production tasks, most of them did not focus on vocabulary specifically. This may be the reason why usually only single vocabulary tests were employed to assess word knowledge. It has been argued, though, that a complex skill such as vocabulary cannot be measured using individual tests. Bowles and Salthouse (2008) claimed that it is necessary to use a variety of measures of vocabulary size, especially in studies where vocabulary knowledge is in the focus of interest. No vocabulary test is a pure measure of word knowledge but they involve other cognitive abilities (Bowles & Salthouse, 2008). Thus, using a battery of vocabulary tests is necessary in order to make sure that what is represented by a composite score of a participant's performance on all tests is in fact vocabulary size.

Additionally, the majority of established vocabulary tests are multiple- choice tests, but it is unclear whether this is the ideal way of assessing individuals' word knowledge. Gyllstad, Vilkaite, and Schmitt (2015) compared second language learners' performance on a multiple-choice vocabulary test with an open interview-based test of vocabulary knowledge. They found a mismatch between the multiple-choice and the open test scores with the former overestimating participants' vocabulary size as compared to the latter. This pattern was argued to arise because participants could use guessing or elimination strategies in the multiple-choice task (Gyllstad et al., 2015). Although that study concerned second language learners, the argument can be extended to the assessment of native speakers. This supports the idea that multiple tests and test formats are required to obtain a reliable indicator of vocabulary size. Such a comprehensive investigation of individual differences in vocabulary, assessed using various different types of vocabulary tests, and their relationship with language processing performance in healthy adult native speakers, is lacking so far.

With the present study we investigated the relationship between individual differences in vocabulary and word recognition more comprehensively. For the above-described reasons, it was considered necessary to employ not only one single measure of vocabulary size but a battery of different tests. Thus, the Dutch participants in the present study completed a set of seven vocabulary tests, two of which were the established Peabody Picture Vocabulary Test (PPVT-III NL; Schlichting, 2005) and Andringa and colleagues' (2012) receptive multiple-choice test. In addition to these multiple-choice tests, five tests of different types and formats were developed. These measures were a definition test, multiple-choice antonym and synonym tests, and open antonym and synonym tests. In this vein, we took into account that various test formats (i.e. multiple-choice and open tests), asking the participants to perform different tasks (such as the antonym or synonym of the target), are needed to reliably assess vocabulary size. The choice of the test types and formats was based on Henriksen's (1999) proposal that there are three dimensions of vocabulary development. The knowledge of words, which varies along a continuum from partial to precise, was addressed using the definition test and the various multiple-choice measures. Secondly, a deeper knowledge of the meaning of words and their relations to other words was assessed using the antonym and synonym tests. Finally, the distinction between receptive and productive vocabulary knowledge was taken into account by using open tests in addition to multiple-choice tests. Furthermore, we were inspired by the way vocabulary was assessed in earlier studies. The format of the multiple-choice synonym test was identical to widely used measures of vocabulary tests, such as the Shipley Vocabulary test (Shipley, 1946). The antonym test only differed from the synonym test in that participants were asked to select a word that had the opposite meaning to the target instead of the same meaning. The test items used in these measures covered a large range of word frequencies to make sure that they were able to measure sufficient variability in vocabulary.

In addition, participants completed a visual lexical decision task to assess their language processing performance. This task is a widely used measure of speed of word recognition (e.g., Balota et al., 2007; Brysbaert et al., 2016a; Keuleers, Lacey, Rastle, & Brysbaert, 2012). On each trial of the lexical decision task, participants are presented with a string of letters and are asked to decide whether or not it is an existing word in a given language. Two classic findings are the effects of lexicality and frequency. Responses for words are usually faster than responses for nonwords, and more frequent words elicit faster responses than less frequent words (e.g., Keuleers et al., 2015; Kuperman & Van Dyke, 2013; Whaley, 1978; Yap et al., 2012). Frequency manipulations were implemented in the present study such that the stimuli covered a large range of word frequencies.

We expected to find beneficial effects of increased vocabulary knowledge on language processing performance, with faster RTs and lower error rates for individuals with greater vocabulary knowledge. In addition, the lexical entrenchment account, which predicts an interaction between a participant's vocabulary score and the word frequency effect, was tested. Larger effects of word frequency were predicted for individuals with poorer vocabulary knowledge (e.g., Diependaele et al., 2013).

In addition, the nature of a potential effect of vocabulary on language processing was examined more closely, aiming at replicating earlier findings concerning the relationship between vocabulary size and diffusion model parameters (Brysbaert et al., 2016a; Yap et al., 2012). For this purpose, we used Ratcliff's (1978) diffusion model approach to analyse the lexical decision task data (see also Yap et al., 2012, for discussion). The model takes a participant's RTs for both correct and incorrect responses into account and decomposes them into a number of parameters representing the cognitive mechanisms underlying binary forced-choice tasks. In the context of a lexical decision task, the model assumes that the information necessary for a response, i.e. word versus nonword decision, is accumulated over time in a noisy process (Ratcliff, Gomez, & McKoon, 2004). It begins at a starting point until it reaches one of two response criteria, i.e. a word or nonword boundary (see Figure 2.1).

The standard version of the diffusion model estimates seven parameters by fitting the diffusion model to participants' lexical decision task data. An important parameter is the mean *drift rate*. It reflects the speed of information accumulation, which is dependent on task or stimulus difficulty. In addition, drift rate is hypothesised to fluctuate from one trial to the next. This is reflected in a parameter called the *variability in drift rate across trials*. As indicated above, the accumulation of information is not only characterised by its speed, i.e. the drift rate, but also by a particular *starting point*, which reflects a participant's bias towards word or nonword responses (see Figure 2.1). Just as for the drift rate, the diffusion model assumes *across-trial variability in starting point*. The information extracted from the stimulus accumulates with a certain drift rate beginning from a starting point towards one of two boundaries or response criteria. The parameter *boundary separation* reflects how far these boundaries are separated from each other and is typically considered to model the speed-accuracy tradeoff. If the boundaries are close together either of the response criteria is met quickly but chances of making a mistake are higher. When the boundaries are farther apart, decision making is slower but less error-prone. In addition to the decision process, the diffusion model assumes two other component processes, stimulus encoding and response execution (see Figure 2.1). The latter two are combined into the parameter *nondecisional components of processing* when the model is fit to data. Finally, the time needed for nondecision processes is assumed to fluctuate which is reflected in the parameter *across-trial variability in the nondecisional component of processing* (Brysbaert et al., 2016a; Ratcliff et al., 2004; Ratcliff & McKoon, 2008). We asked whether differences in vocabulary knowledge would be associated with specific parameters of the diffusion model.

We expected the drift rate to increase with increasing vocabulary size because individuals with greater vocabulary knowledge were assumed to build up lexical information more quickly as one of the reasons for their faster lexical decision RTs. The variability in drift rate was presumed to decrease with increasing vocabulary size. This would indicate more stable, smoother processing across different trials for individuals with larger vocabularies, for instance due to smaller effects of stimulus difficulty on drift rate. In addition, and based on previous findings, the starting point was assumed to be less word biased for high- than for low-vocabulary individuals (Brysbaert et al., 2016a). Furthermore, it was presumed that the time needed for nondecision components of processing may decrease with increasing vocabulary scores. This would indicate that individuals with improved word knowledge also exhibit advantages in their general speed of processing, for instance, reflected in faster response execution (Brysbaert et al., 2016a; Yap et al., 2012).

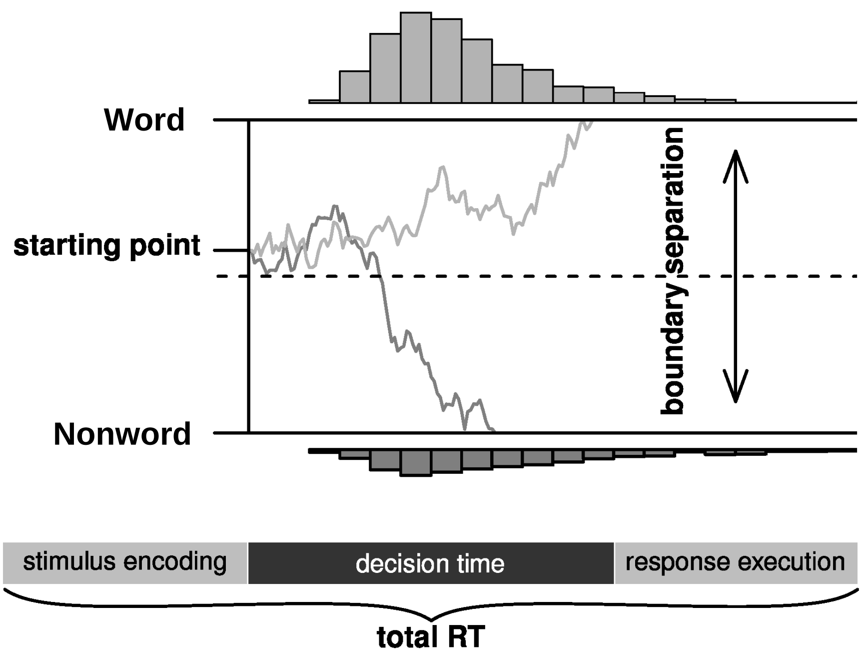

Figure 2.1: Ratcliff's (1978) diffusion model applied to the lexical decision task. Upon stimulus presentation, noisy information accumulates either towards a word or a nonword threshold. The figure shows the information accumulation for two distinct stimuli, one resulting in a word and the other in a nonword decision (Figure adapted from Dutilh et al., 2012).

Hence the present study aimed at investigating individual differences in vocabulary in young adult native speakers more closely by using a battery of measures to assess vocabulary size. Additionally, the origin of a potential vocabulary effect on language processing was examined using diffusion model analyses, accounting for the possibility that an overall effect of vocabulary on lexical decision task performance might originate in different component processes of lexical decision making.

Method

Participants

A total of 75 young adults (57 females) aged between 18 and 34 years (M = 21.9; SD = 3.7) gave informed consent to participate in this study.[2] All participants were completing their studies at the Radboud University Nijmegen or the Hogeschool van Arnhem en Nijmegen at the time of testing or had recently graduated. They were recruited using

[2]Five participants were 30 years or older. Inspection of their data did not suggest that they were to be treated as outliers.

the participant database of the Max Planck Institute for Psycholinguistics and were paid 12 Euros for their participation. Ethical approval was granted by the Faculty of Social Sciences of the Radboud University Nijmegen.

Materials and design

All participants completed a battery of seven vocabulary tests and a lexical decision task. Two of the vocabulary tests were established measures of vocabulary knowledge, namely Andringa et al.'s (2012) receptive multiple-choice test and the Peabody Picture Vocabulary Test (PPVT-III NL; Schlichting, 2005). The other five tests were newly developed.

Receptive multiple-choice test

This multiple-choice test was developed by Andringa and colleagues (2012). Participants were presented with target words, such as *mentaliteit* (mentality) or *tentatief* (tentative), embedded in different neutral carrier sentences. Each sentence was presented along with five answer options, one of which was a description of the target word and one being *Ik weet het echt niet* (I really don't know). For example, the target word *mentaliteit* (mentality) was presented with the answer options *tafel* (table), *persoon* (person), *manier van denken* (way of thinking), and *sfeer* (atmosphere; see Appendix A for all vocabulary tests). The target words covered a large range of word frequencies between 0 and 31.28 counts per million in the SUBTLEX corpus (M = 1.87; SD = 5.07; Keuleers, Brysbaert, & New, 2010).

The test consists of 60 target sentences. In the present study the first sentence was used as a practice item so that the test comprised a total of 59 questions. Both the original sequence of items and the positions of the correct responses were the same as in Andringa et al.'s (2012) test.

Peabody Picture Vocabulary Test (PPVT)

Every trial in the Dutch version of the PPVT-III (Schlichting, 2005) consists of a spoken target word and a set of four pictures. Participants were instructed to choose the picture that corresponded to the word they heard. The target word frequency ranged from 0 to 29.13 counts per million in the SUBTLEX corpus (M = 1.86; SD = 4.60; Keuleers et al., 2010).

The stimuli in the PPVT are organised in blocks of twelve words but the number and order of blocks varied depending on the participant's performance. Each participant started with the same first block of twelve items. Depending on the number of mistakes made, the following block was comprised of either easier or more difficult words. The same held true for all subsequent blocks of stimuli. Thus, individuals with poor word knowledge might complete five blocks of twelve words each, while others might complete the maximum of eight blocks of stimuli. All target words had been spoken by a female native speaker of Dutch and recorded.

Definition test

In this test, participants were presented with 20 definitions of words from four different semantic categories (animal, profession, body part, instrument/object; see Appendix A). The task was to name the word that corresponded to the definition. All definitions were taken from a definition naming experiment by La Heij and colleagues (1993). The frequencies of the correct responses ranged between 0.02 and 244.07 occurrences per million (M = 39.01, SD = 63.63) in the subtitle corpus SUBTLEX-NL (Keuleers, Brysbaert, & New, 2010). Furthermore, the correct responses displayed z-transformed prevalence values between 1.99 and 3.41 (M = 2.93, SD = 0.44). Prevalence is a measurement of how many people in a group of speakers of a language, in this case Dutch, know a given word (Keuleers et al., 2015). Following Keuleers and colleagues, words with a prevalence value of 1 or above are known by at least 85% of Dutch native speakers in the Netherlands. The order of items within this and the following tests was pseudo-randomised such that low- and high-frequency test items were well distributed and the order was fixed across participant.

Multiple-choice antonym test

The multiple-choice antonym test included 25 test items, which were presented without carrier sentences and along with five single-word answer alternatives (see Appendix A). Some of the target words were taken from the Toets Gesproken Nederlands (TGN), a Dutch language test used to assess language for immigration requirements (Kerkhoff, Poelmans, de Jong, & Lennig, 2005).

The test items in the multiple-choice antonym test represented a large frequency range with between 0 and 3838.54 counts per million (M = 200.16, SD = 764.93) in the subtitle corpus SUBTLEX-NL (Keuleers et al., 2010) and z-transformed prevalence values of between -1.73 and 3.37 (M = 2.37, SD = 1.22).

Open antonym test

Just as in the multiple-choice test, the open antonym test included 25 test items, which were presented individually (see Appendix A). Participants were instructed to write down an antonym for each word. Some of the target words were also taken from the TGN (Kerkhoff et al., 2005). The test items in the open antonym test represented a frequency range between 0 and 60.69 counts per million in the SUBTLEX-NL corpus (M = 9.09, SD = 13.14). The prevalence values of the target words ranged from 1.03 to 3.32 (M = 2.49, SD = 0.68).

Multiple-choice synonym test

The multiple-choice synonym test was structurally identical to the multiple-choice antonym test, the only difference being that participants were asked to select a word that has the same meaning as or is interchangeable with the given target. Hence, it consisted of 25 test items, which were presented along with five single-word answer alternatives (see Appendix A). The multiple-choice synonym test was based on a part of the Groningen Intelligence Test (Luteijn & van der Ploeg, 1983). This measurement consists of 20 test items, which are presented along with five answer options each. The majority of these words have very low frequencies in the SUBTLEX-NL corpus. In order to adapt the test for the present purposes and make the final test scores comparable to the 25-item antonym test, five new medium to high-frequency test words were added.

The word frequencies of the test items in the multiple-choice test ranged from 0 to 48.05 per million (M = 4.85, SD = 11.02) in the SUBTLEX-NL corpus. The words' prevalence values ranged between -0.64 to 3.35 (M = 1.77, SD = 1.11).

Open synonym test

The open synonym test was structurally identical to its antonym counterpart. It comprised 25 test items which were presented individually and without carrier sentences (see Appendix A). The construction of this open synonym test was inspired by the English version of the Mill Hill Vocabulary Scale (Raven, Court, & Raven, 1994). The target words had frequency values between 0 and 36.25 per million (M = 4.62, SD = 9.33). Furthermore, the test items in the open synonym test had prevalence values between -0.59 and 3.32 (M = 1.88, SD = 1.06). Importantly, each word appeared only once as target item in one of the vocabulary tests.

Lexical decision task

Ninety word and 90 nonword stimuli were included in the lexical decision task. The words covered a broad word frequency range from 0.02 to 89.92 (M = 9.69, SD = 16.38) occurrences per million in the SUBTLEX corpus (Keuleers et al., 2010). All words were uninflected and were not homophonous with other Dutch words. Furthermore, the words' prevalence values were not lower than 2.89, indicating that all words were known by at least 98% of Dutch speakers in the Netherlands (Keuleers et al., 2015).

The nonwords were created using the program Wuggy (Keuleers & Brysbaert, 2010), which generates nonwords based on real words. Each of the nonwords corresponded to one real word on the list. All nonwords differed from their real word counterparts by a letter, a sound, or an entire syllable, while being pronounceable but not homophonous or homonymous with an existing Dutch word. Using this set of stimuli, four lists including all 180 items were created. The order of stimuli within each of the lists was fixed. Not more than three consecutive trials belonged to the same experimental condition. The four stimuli lists were counterbalanced across participants.

Apparatus

All tasks were presented on a 17-inch screen (Iiyama LM704UT) using either Presentation software (version 16.5, www.neurobs.com), or the online questionnaire software LimeSurvey (www.limesurvey.org). The auditory stimuli in the PPVT were presented using HD 280 Sennheiser headphones.

Procedure

All participants were tested individually in experiment rooms at the Max Planck Institute for Psycholinguistics. Everyone completed the vocabulary tests in the same order, namely: Definition test, Andringa et al.'s (2012) test, multiple-choice antonym test, open antonym test, multiple-choice synonym test, and PPVT. The vocabulary tests were self-paced and participants were instructed to answer as accurately as possible without thinking about single test items for too long. Before each of the tests started, instructions were presented on the screen, along with example questions. In total, the vocabulary test battery took between 35 and 45 minutes. The lexical decision task was completed in a separate test session.

Receptive multiple-choice test

The target words were presented in neutral carrier sentences and were marked with two asterisks. The question and answer options were written in 25-point Arial font and stayed on the screen until the participant had selected an answer by pressing one of the number buttons 1 to 5 on the keyboard. Answering a question initiated the presentation of the following test item.

Peabody Picture Vocabulary Test

Participants were presented with a spoken stimulus and four pictures on the screen and were instructed to select one of the four pictures that corresponded to the word they heard by pressing one of the number buttons 1 to 4 on the keyboard. The target word was repeated until one of the four response buttons was pressed, which initiated the presentation of the next stimulus.

New vocabulary tests

In the definition and open antonym and synonym tests, participants were asked to type in the correct answer using the keyboard. For the definition test, it would be the animal, profession, body part, or object corresponding to the definition. In the open antonym and synonym tests, participants were required to type in a word that was an antonym or synonym of the test word. Participants could proceed to the next question without answering the previous one, if they did not know the answer.

In the multiple-choice antonym and synonym tests, participants were given five answer alternatives and instructed to select the word that had the opposite (antonym) or same (synonym) meaning as the target. Answering a question initiated the presentation of the following test item. Everything was written in 25-point Arial font.

Lexical decision task

The experiment was divided into two parts, consisting of 90 stimuli each. Between the two blocks, participants could take a short break. Each trial started with a fixation cross, which was shown in the center of the screen. After 500 ms, it was replaced by a word or nonword written in 24-point Arial font. The stimuli stayed on the screen for 3 seconds or until a response button was pressed. Half of the participants were instructed to press the "Z"-button on the keyboard if the string of letters on the screen was a word and "M" if it was a nonword; the other half did it the other way around. Participants were instructed to respond as quickly and accurately as possible. Before the test phase began, participants were familiarised with the task in four practice trials.

Analyses

Vocabulary tests

Peabody scores were calculated based on the total number of items participants responded to and the number of errors they made. These raw scores were then transformed into standardised scores, called woordbegripquotiënt (word comprehension score, WBQ), which was used for all further analyses (Schlichting, 2005). PPVT scores are adjusted for age. For our group of participants the maximum possible score varied between 139 for younger participants and 136 for the oldest participants.

One point was given for each correct answer in all other multiple-choice and open tests. Some exceptions applied to the definition as well as the open antonym and synonym tests. If a participant demonstrated knowledge of the word or concept without producing an actual antonym or synonym (e.g. writing *stil* (silent) instead of *stilte* (silence) as antonym for *lawaai* (noise)) or if the answer was misspelled, they received 0.5 points for that answer. Three native speakers of Dutch with backgrounds in linguistics or psycholinguistics independently categorised all answers. Some cases, which they did not agree on initially, were then discussed in the group. That always resulted in a judgment supported by all of them. It was possible that for one target several different responses were correct. For example, in the case of the definition test item *Iemand die werkt met patiënten* (Someone who works with patients), the responses *dokter* and *arts* (both doctor) as well as *verpleegster* (nurse) were considered correct. In the open antonym test, for instance, both *falen* and *mislukking* (both failure) were considered correct responses given the target *succes* (success).

The relationships between the vocabulary tests were analysed using bivariate Pearson correlation analyses in SPSS (version 20). In addition we conducted a Principal Component Analysis (PCA) in SPSS, which can be used to identify a small number of components that account for the variability in a larger number of original measures, in this case seven vocabulary tests. The goal of a PCA is data reduction in cases where the measures originally included in the study are to be reduced to a subset without losing information (DeCoster, 1998). We ran a PCA assuming two components for that exact reason, namely reduction of the number of vocabulary scores as predictors in the individual differences analyses to one measure or two measures (reflecting a distinction between multiple-choice and open tests).

Lexical decision task

In the lexical decision task, we measured RTs and accuracy. Responses were excluded from the analyses if they exceeded a participant's mean RT by more than three standard deviations or were shorter than 250ms.

Accuracy was investigated using mixed logit models employing the glmer function from the package lme4 (version 1.1.12; Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2016). The first model on words and nonwords included a fixed intercept, a fixed effect for lexicality (word vs. nonword) as well as random intercepts for both item and participant. Additionally, we modelled per-participant random slope adjustments to the lexicality effect. Secondly, a model on words alone was run including a fixed intercept, a fixed effect for frequency, and random intercepts for both items and participants. Per-participant random slope adjustments to the frequency effect were also included.

RTs were log-transformed and analysed in linear mixed-effects models using the lmer function of the lme4 package (version 1.1.12; Bates et al., 2015) in R (R Core Team, 2016). RTs for correct responses to words vs. nonwords and to words alone were analysed. The first mixed model on RTs for correct responses to words and nonwords included a fixed intercept, a fixed effect for lexicality, and by-participant and by-item adjustments to the fixed intercept (random intercepts). Additionally, by-participant random slope adjustments to the lexicality effect were included. The predictor lexicality was sum-to-zero contrast coded (nonword = 1; word = -1). The second model on correct responses to words only included a fixed intercept, and a fixed effect for the log-transformed continuous variable word frequency. Furthermore, by-participant and by-trial random adjustments to the fixed intercept (random intercepts) and by-participant random adjustments to the frequency slope were modelled (random slope). All possible correlations between the random effects were included. Hence, we followed Barr, Levy, Scheepers, and Tily (2013) using a maximal random effects structure. P-values were determined using the normal approximation, i.e. using the t-value as z-value.

In order to examine the effects of individual differences in vocabulary knowledge on lexical decision task performance, the models of RTs and accuracy for words were run with frequency and vocabulary score (see below for the definition) as an additional predictors. Both models included a fixed intercept, a fixed effect for the continuous variable word frequency, and by-participant and by-trial random adjustments to the fixed intercept (random intercepts). In addition, by-participant random adjustments to the frequency slope were modelled (random slope). The model for accuracy did not converge with this maximum random effects structure. We therefore had to remove the random slopes from the model.

In order to explore how well the scores from each of the vocabulary tests predicted speed and accuracy in the lexical decision tasks, we ran separate models for each score, yielding seven models predicting accuracy and seven models predicting speed. In addition, we used a composite measure of vocabulary described below. Based on the PCA on all vocabulary tests, which did not show a clear pattern distinguishing different types of tests from one another (see Results section below), we decided to use a component score of vocabulary reflecting each participant's performance on all seven measures of word knowledge. For that purpose, regression-based factor scores were calculated for each participant using the PCA method in SPSS (DiStefano, Zhu, & Mindrila, 2009). Assuming only one underlying factor, each individual's loading or score on that factor based on their seven vocabulary test scores was calculated and thus, the measurements of word knowledge were collapsed into one number with a mean of zero. This allowed us to compare the individual vocabulary measures with a composite measure reflecting performance on the entire battery of tests.

In addition, we ran a diffusion model analysis on the lexical decision task in order to investigate the origin of a potential vocabulary size effect. By fitting the model to each participant's RTs for both correct and incorrect responses, we obtained seven parameter estimates for each individual: (1) mean drift rate, i.e. the speed of information accumulation, (2) variability in drift rate across trials, (3) boundary separation, (4) mean starting point, i.e. participants' bias towards words or nonwords, (5) variability in starting point across trials, (6) non-decision component of processing, i.e. time needed for stimulus encoding and response execution, and (7) across-trial variability in time needed for non-decision component. We employed the fast-dm algorithm written by Voss and Voss (2007) to estimate the parameters of the diffusion model. These parameter estimates were then entered into regression analyses in R, with vocabulary score as predictor.

Results

Table 2.1: The distribution of test scores in all seven vocabulary tests. The maximum possible scores for each of the tests are provided in brackets in the column displaying participants' maximum scores.

Test	N	Minimum	Maximum	Mean	SD
Andringa	75	25.0	58.0 (59.0)	40.01	6.47
PPVT	75	56.0	125.0 (*)	102.61	12.5
Definition test	75	12.0	20.0 (20.0)	16.41	2.04
Antonym MC	75	14.0	25.0 (25.0)	23.0	1.60
Antonym open	75	14.0	24.0 (25.0)	19.29	2.37
Synonym MC	75	11.0	24.0 (25.0)	17.68	2.87
Synonym open	75	5.5	22.0 (25.0)	10.70	2.86

* Maximum possible score varied between 136 and 139.

Vocabulary test

Table 2.1 shows the vocabulary test scores averaged across participants and Table 2.2 displays the correlations between the test scores. There were moderate to strong correlations between all test scores, indicating that the vocabulary measures assessed, to some extent, a shared underlying ability. The multiple-choice antonym test, which was easier than the other tests, was least strongly correlated with the other measures.

The reliability measure Cronbach's α indicated that the test battery as a whole was highly reliable (α = .88). Dropping the multiple-choice antonym test would increase α (.89) while leaving out one of the other tests would lead to a lower α.

A PCA assuming two components was run on z-transformed vocabulary scores. This was based on the assumption that two components might distinguish between multiple-choice and open tests. The first component had an eigenvalue of 4.35, the other component had an eigenvalue below 1. This first component explained 62.13% of the total variance. As shown in Table 2.3, Factor 1 loaded on all tests with only a slightly smaller loading for the multiple-choice antonym test. No distinction between productive and receptive vocabulary tests was found.

Table 2.3: Results of the PCA assuming two components.

Vocabulary measure	**Component 1**	**2**
Andringa	.81	-.13
PPVT	.80	-.14
Definition	.86	.04
Antonym MC	.56	.82
Antonym open	.84	-.07
Synonym MC	.78	-.22
Synonym open	.82	-.06
Eigenvalue	4.35	.76
% Variance	62.13	10.89

Lexical decision task

Accuracy rates were overall high with 2.7% of all trials being false alarms and 1.6% misses. RTs were trimmed per participant according to the above-mentioned criteria. 1.7% of trials were excluded as outliers. As typically found in lexical decision tasks, accuracy was higher for words than for nonwords (z = -4.56; p < .001) and participants made fewer errors with increasing word frequency (z = 7.36; p < .001). In addition, RTs for words were significantly faster than for nonwords (t = 10.74; p < .001; see Appendix B for a table showing averaged lexical decision RTs for all conditions and a plot of the RT distribution). Finally, RTs for correct responses to words were significantly predicted by

Table 2.2: Bivariate correlations between the vocabulary scores.

	Definition	**Andringa**	**Antonym MC**	**Antonym open**	**Synonym MC**	**Synonym open**
Andringa	.62**					
Antonym MC	.46**	.36**				
Antonym open	.73**	.63**	.40**			
Synonym MC	.59**	.62**	.32**	.58**		
Synonym open	.68**	.61**	.38**	.63**	.56**	
PPVT	.63**	.57**	.35**	.67**	.58**	.59**

** $p < .01$.

word frequency (t = -15.17; p < .001), with faster RTs for high- than for low-frequency words.

Individual differences

The main interest of the present study was the relation between individual differences in vocabulary and language processing. Results of mixed-effects models on both accuracy and speed in the lexical decision task are reported below. For this individual differences investigation, we focused on responses to word trials.

Accuracy

Response accuracy for words was significantly predicted by word frequency (z = 7.33; p < .001) but not by the composite vocabulary score, which was based on all seven measures of vocabulary knowledge (z = 1.50; p = .13). Hence, error rates were lower for high-frequency words. The interaction between word frequency and the composite vocabulary score was not significant (z = .08; p = .93). The models using participants' vocabulary scores from the individual measures as predictors of accuracy showed overall the same results (see Table 2.4). Word frequency was a highly significant predictor of lexical decision accuracy in all seven models whereas vocabulary was insignificant in five of the models. Only the scores from Andringa et al.'s (2012) measure and the open antonym test turned out to be significant predictors of lexical decision accuracy.

Table 2.4: *P*- and *t*-values for the main effects in each of the models where scores of individual vocabulary measures were used as a predictor of lexical decision accuracy.

Model	**Variable**	***t***	***p***
Andringa	Vocabulary score	2.29	.02
	Word frequency	7.24	<.001
	Frequency x vocabulary	-.35	.73
PPVT	Vocabulary score	1.21	.23
	Word frequency	7.40	<.001
	Frequency x vocabulary	.83	.41
Definition test	Vocabulary score	.91	.36
	Word frequency	7.34	<.001
	Frequency x vocabulary	.38	.70
Multiple-choice antonym	Vocabulary score	.02	.99
	Word frequency	7.37	<.001
	Frequency x vocabulary	1.66	.1
Open antonym	Vocabulary score	2.43	.02
	Word frequency	7.10	<.001
	Frequency x vocabulary	-.39	.70
Multiple-choice synonym	Vocabulary score	.22	.83
	Word frequency	7.36	<.001
	Frequency x vocabulary	-1.13	.26
Open synonym	Vocabulary score	.64	.52
	Word frequency	7.36	<.001
	Frequency x vocabulary	-.25	.80

Reaction times

Lexical decision RTs on word trials were significantly predicted by log-transformed word frequency (t = -15.21; $p < .001$) and the composite vocabulary score (t = -3.12; p = .001). However, the interaction between the two main effects, word frequency and vocabulary score, was not significant (t = 1.23; p = .22). The seven distinct mixed-effects models each including the vocabulary score from one of the individual tests and word frequency as predictors confirmed this (see Table 2.5). Word frequency was a significant predictor of lexical decision RTs in all models. Only the multiple-choice antonym test did not significantly predict participants' RTs. All other models showed a significant main effect of vocabulary score (see Table 2.5). The interaction between vocabulary score and word frequency was significant for only one of the measures, namely the definition test (t = 2.02, p = .04). Participants with higher definition vocabulary test scores showed smaller effects of word frequency.

Table 2.5: *P*- and *t*-values for the main effects in each of the models where scores of individual vocabulary measures were used as a predictor of lexical decision RTs.

Model	**Variable**	***t***	***p***
Andringa	Vocabulary score	-2.30	.02
	Word frequency	-15.30	<.001
	Frequency x vocabulary	1.5	.14
PPVT	Vocabulary score	2.08	.038
	Word frequency	-15.08	<.001
	Frequency x vocabulary	.01	.99
Definition test	Vocabulary score	-2.75	<.01
	Word frequency	-15.26	<.001
	Frequency x vocabulary	2.0	.04
Multiple-choice antonym	Vocabulary score	-1.61	.12
	Word frequency	-15.12	<.001
	Frequency x vocabulary	.46	.64
Open antonym	Vocabulary score	-2.69	<.01
	Word frequency	-15.09	<.001
	Frequency x vocabulary	.98	.32
Multiple-choice synonym	Vocabulary score	-2.30	.02
	Word frequency	-15.14	<.001
	Frequency x vocabulary	.79	.43
Open synonym	Vocabulary score	-2.99	.004
	Word frequency	15.18	<.001
	Frequency x vocabulary	.97	.33

Thus, RTs decreased with increasing word frequency, as typically found in lexical decision tasks, and individuals with higher vocabulary scores responded faster than those with smaller vocabulary scores (see Appendix C).

Diffusion model analysis

The means and SDs of the seven parameter estimates are reported in Table 2.6. The mixed-effects models on all seven diffusion model parameters were only run with the composite measure of vocabulary as a predictor. Based on the above-reported results it was assumed that the composite score is representative of all individual vocabulary tests.

Table 2.6: Parameter estimates of the diffusion model.

Parameter	***Mean***	***SD***
Drift rate for words	3.74	0.99
Variability in drift rate	0.69	0.29
Boundary separation	1.40	0.44
Starting point	0.55	0.08
Variability in starting point	0.23	0.12
Non-decision components	0.46	0.05
Variability in non-decision components	0.12	0.06

Drift rate for words was predicted by vocabulary score (β = .47, *SE* = .10, t = 4.61, p < .001). As expected, higher vocabulary scores were associated with faster drift rates, i.e. faster information accumulation. The starting point was significantly closer to zero (β = -.02, *SE* = .01, t = -2.49, p =.02) or neutral between the word and nonword boundaries with increasing vocabulary knowledge, thus the word-bias was smaller. Individuals with higher vocabulary scores spent less time on stimulus encoding and response execution as indicated by the negative effect of vocabulary on the non-decision component of processing (β = -.01, *SE* = .01, t = -2.76, p = .007).

Discussion

A battery of seven vocabulary tests and a lexical decision task were used to examine the relationship between vocabulary size and word recognition speed in young adults. There was variation in vocabulary test performance although the group was quite homogeneous being comprised of university students only. The bivariate correlations were similarly strong for all vocabulary tests and indicate that all the tests measure largely the same ability. Performance on only one test, namely the multiple-choice antonym test, was less strongly correlated with all other tests. Furthermore, the reliability measure Cronbach's α was high for the test battery as a whole. This, together with a relatively high average score and low standard deviation on the multiple-choice antonym test, suggests that this specific test was easier for our participants than the other measures. We acknowledge this difference in difficulty. It is probably related to the higher word frequencies of the test items in this test as compared to the other tests.[3] If the entire battery of vocabulary tests were to be used again, one might want to adjust the frequency ranges so that they are more similar across all tests. In addition, the antonym tests seemed to be generally easier than their synonym test counterparts.[4]

To sum up, although the test battery was comprised of different types of vocabulary measures, i.e. open and multiple-choice tests, the correlational and reliability analyses indicated that all tests measured largely the same capacity. Only the multiple-choice antonym test did not match the other tests in difficulty. Finally, no distinction between multiple-choice and open tests was found. A PCA confirmed these conclusions. First, no distinction between multiple-choice and open tests in terms of two distinct components was found. Secondly, only the multiple-choice antonym test did not show a loading as high as the other tests on the first component and instead loaded highly on the second component. This is in line with the conclusions drawn from the descriptive, correlational, and reliability analyses indicating that the multiple-choice antonym test does not relate well to the other measures, presumably as it was much easier than the remaining tests. It has to be noted, however, that the open tests were not interview-based measures; they were not fully open as they still provided the participants with stimuli, i.e. single words, and specific tasks to perform on them (writing down the antonym or synonym of the given word). This might be the reason for not finding a distinction between multiple-choice and open tests as has been observed previously (Gyllstad et al., 2015).

The fact that the vocabulary tests did not correlate perfectly, though, supports Bowles and Salthouse's (2008) hypothesis that each of them uniquely involves other cognitive abilities in addition to word knowledge. Hence, in individual differences studies focusing on vocabulary, the use of a composite measure based on participants' performance on different tests provides a more valid estimate of their vocabulary size than given by a single measure. Therefore, the mixed-effects model analyses on both lexical decision accuracy and RTs were run with i) participants' scores from all individual vocabulary tests, and ii) a composite score of vocabulary test performance as predictors. Thus, we tested whether using estimates of individuals' vocabulary size that

[3]There are a few items with extremely high frequency values, such as *aan* (on; 3838.54 counts per million), *achter* (behind; 473.8 counts per million), and *hetzelfde* (the same; 193.69 counts per million). Excluding these words leads to a drop in mean word frequency from 200.16 to 22.64 counts per million, which is still higher than in some of the other tests.

[4]Participants scored higher on the multiple-choice antonym (M = 23.0, SD = 1.60) than on the multiple-choice synonym test (M = 17.68, SD = 2.87), and higher on the open antonym (M = 19.29, SD = 2.37) than on the open synonym test (M = 10.70, SD = 2.86).

were based on individual measures of word knowledge or on a battery of tests provide a different picture of the relationship between individual differences in vocabulary and word recognition performance.

In the lexical decision task, the typical effects of lexicality and word frequency on accuracy and RTs were found, with more accurate and faster responses for words compared to nonwords, and for higher compared to lower frequency words. The focus of the present investigation, however, was the relationship between individual differences in vocabulary and language processing. Weaker vocabulary scores were associated with slower RTs in the lexical decision task. This finding is consistent with findings of several earlier studies (Brysbaert et al., 2016a; Diependaele et al., 2013; Unsworth et al., 2011; Yap et al., 2012). This was true not only for the composite measure of vocabulary knowledge but also for all of the individual vocabulary tests as predictors of lexical decision RTs, except for the multiple-choice antonym test.

Furthermore, the analyses suggest that the individual vocabulary tests (besides the multiple-choice antonym test) represent participants' vocabulary size as well as the composite score; all analyses showed the same pattern of the relationship between vocabulary and lexical decision accuracy and speed. Hence, the present study does not provide evidence for the assumption (Bowles & Salthouse, 2008) that a battery of different types of test has to be used to assess vocabulary knowledge reliably. Thus, a practical recommendation from this study is that for a broad assessment of Dutch university students' vocabulary, the use of one of the standard tests, e.g. Andringa's test or the PPVT, is adequate. Based on the current results no specific measure of vocabulary can be recommended as being superior over another. It should, of course, be kept in mind that we only used a single processing task. Thus, we cannot exclude that performance in other linguistic tasks may best be predicted by a composite score based on the results of several vocabulary tests.

Deviating from previous studies (Chateau & Jared, 2000; Diependaele et al., 2013; Kuperman & Van Dyke, 2013; Yap et al., 2009), the word frequency effect in our lexical decision experiment was independent of vocabulary size. A moderately significant interaction between word frequency and vocabulary was found solely for the definition test; none of the other individual test scores or the composite score showed an interaction with word frequency. Thus, it was concluded that overall the frequency x skill interaction can be considered absent in the present data. The fact that we obtained strong effects of word frequency on accuracy and RTs makes it rather unlikely that the frequency range covered by our materials was too small to elicit the interaction between word frequency and skill. However, it has to be noticed that the word frequency range in the materials used by Brysbaert and colleagues (2016a) was larger with a minimum

SUBTLEX frequency of 0.12 and a maximum of 501.33 (M= 18.73), as compared to our range of word frequencies (see Adelman et al., 2014, for the materials used in Brysbaert et al., 2016a). Hence, even though we found a main effect of word frequency on RTs, the word frequency range might still have been too small to obtain the frequency x skill interaction. In addition, it is possible that the group of participants we tested was too homogeneous and the variation in vocabulary size too small to elicit the frequency x skill interaction, which has previously been taken as evidence for the lexical entrenchment account.

The lexical entrenchment hypothesis assumes that the frequency x skill interaction reflects individual variation in language exposure, with more exposure leading to more distinct or robust lexical representations that are faster to be accessed (Brysbaert et al., 2016a; Diependaele et al., 2013). An increased amount of exposure has been argued to have particularly strong effects on the exposure to low-frequency words (Kuperman & Van Dyke, 2013). As a result, the frequency difference between low- and high-frequency words gets smaller and the frequency curve is less steep (Brysbaert et al., 2016a). Thus, individuals with increased exposure (and larger vocabularies) show a smaller word-frequency effect in tasks such as the lexical decision task (Brysbaert et al., 2016a; Diependaele et al., 2013). One might argue that the group of university students is probably rather homogeneous as to how much and which type of exposure they get to their native language by reading, attending lectures, and so forth. Hence, all participants were presumably highly proficient users of their native language. Although there was variation in vocabulary scores as well as in lexical decision RTs, maybe this was not strong enough to produce a frequency by skill interaction.

Alternatively, the present results might be taken to indicate that some individuals are simply fast processors whereas others are generally slower. Claiming that it is just processing speed that differs in relation to vocabulary size would be an argument against the lexical entrenchment account and the idea that there are representational differences between vocabularies of varying sizes. More research on language processing, the sensitivity to lexical characteristics, and vocabulary size is needed to get more insights into the applicability of the lexical entrenchment account. For this purpose it is crucial to not only test university students but a more varied participant group and, thus, get a more representative picture of individual differences in vocabulary and their relationship with language processing performance. If the frequency x skill interaction were found in a more heterogeneous participant sample using the same task as in the present study, this would support the idea that the group of university students exhibited too little variation in vocabulary and/or lexical decision performance to elicit the effect.

A diffusion model analysis was run to gain insights into the origin of the vocabulary effect on lexical decision times. We did not observe exactly the same but overlapping patterns with Brysbaert et al. (2016a) who found all parameters but boundary separation and the non-decision component of processing to be affected by vocabulary, and Yap et al. (2012) who found vocabulary to significantly predict all parameters but across-trial variability in starting point. The most interesting parameter is the drift rate for words, thus the speed of information accumulation. In line with previous studies, we found vocabulary score to predict drift rate for words, indicating that lexical information of word stimuli builds up faster in individuals with greater vocabulary knowledge (Brysbaert et al., 2016a; Yap et al., 2012). Hence, the high-vocabulary advantage has its origin at least to a certain extent in the fact that the information build-up is faster; perhaps because individuals with stronger word knowledge are faster in accessing their lexicon than individuals with weaker vocabulary knowledge.

Additionally, and in line with Brysbaert et al.'s (2016) observations, we found a significant effect of vocabulary score on starting point, showing that the word bias was less strong for individuals with higher vocabulary scores. Due to the fact that half of the lexical decision stimuli were words and the other half were nonwords, an ideal starting point for the lexical decision process would be right in the middle between the word vs. nonword response boundaries. This would then allow fast word as well as nonword decisions. Maybe participants with better vocabulary knowledge were more sensitive to the distribution of words vs. nonwords among the stimuli, resulting in a starting point that is neither strongly word nor strongly nonword biased, enabling efficient decision making in both directions.

Finally, we found non-decision processes to be significantly predicted by vocabulary score. Consistent with Yap et al. (2012), participants with higher vocabulary scores needed less time for non-decision processes. Hence, not only the build-up of lexical information (drift rate), i.e. word processing, was faster for individuals with greater word knowledge, but also stimulus encoding and response execution happened faster. It can be speculated that better vocabulary knowledge may be associated with an increase in general speed of processing, in addition to an increase in speed of language processing. Maybe greater vocabulary knowledge is associated with advantages in general cognitive ability and speed of processing so that individuals' vocabulary scores predict not only the time needed for language-related processing in the lexical decision task but also for non-decision processes. However, further research including measures of general cognitive abilities, such as general processing speed, is needed to make stronger claims about the relationship between language-specific and general cognitive abilities.

Conclusions

In this study, it was shown that greater vocabulary knowledge, as measured by a battery of vocabulary tests, is associated with faster word processing in a lexical decision task. Although the vocabulary tests differed in their degree of openness and in the task to be performed, the analyses showed the same patterns for all individual vocabulary tests and for the composite score, which reflects participants' performance on the test battery as a whole. Thus, our findings suggest that it is not necessary to use various measures to assess vocabulary. Furthermore, diffusion model analyses indicated that the high-vocabulary advantage in lexical decision speed originates in various component processes of lexical decision-making, namely drift rate, starting point, and time needed for non-decision processes.

The present study set out to provide a more comprehensive investigation of the relationship between individual differences in vocabulary and language processing. More research is needed examining potential consequences of individual differences in vocabulary size, such as structural or representation differences between vocabularies of varying sizes. Finally, despite considerable evidence for a relationship between individual differences in vocabulary size and language processing performance, to our knowledge no current model of language processing accounts for effects of variation in vocabulary size. This is, hence, an important issue for further experimental and computational studies.

Appendix A: Vocabulary tests

Andringa et al.'s (2012) receptive multiple-choice test

1. Deze schoenen *glanzen*.
    a) zijn heel erg nat
    b) glimmen
    c) zijn verschillend
    d) zijn kapot
    e) Ik weet het echt niet.
2. Hij *spaart* voor een auto.
    a) Hij stopt voor een auto langs de weg.
    b) Hij is onder een auto gekomen.
    c) Hij verzamelt geld om een auto te kopen.
    d) Hij verzamelt foto's van auto's.
    e) Ik weet het echt niet.
3. De leraar *prijst* de leerling.
    a) De leraar geeft de leerling een cadeau.
    b) De leraar zegt dat de leerling niet goed zijn best doet.
    c) De leraar geeft de leerling een klap.
    d) De leraar zegt dat de leerling goed werk levert.
    e) Ik weet het echt niet.
4. Ik vind honden *eng*.
    a) Ik vind honden vies.
    b) Ik ben bang voor honden.
    c) Ik vind honden leuk.
    d) Honden maken mij ziek.
    e) Ik weet het echt niet.
5. Hiermee *beëindigen* we de vergadering.
    a) beginnen

   b) besluiten
   c) vergeten
   d) verrassen
   e) Ik weet het echt niet.

6. Hij heeft pijn in zijn *hiel*.
   a) onderste deel van de rug
   b) achterkant van de voet
   c) bovenste deel van het hoofd
   d) zijkant van de knie
   e) Ik weet het echt niet.

7. Wat een vreemde *mentaliteit*!
   a) tafel
   b) persoon
   c) manier van denken
   d) sfeer
   e) Ik weet het echt niet.

8. Hij werkt bij een *uitgeverij*.
   a) bedrijf dat boeken laat drukken
   b) kantoor waar je geldzaken doet
   c) instelling die arme mensen helpt
   d) gebouw waar je dingen kunt kopen
   e) Ik weet het echt niet.

9. Die mensen hebben een *bok* in hun tuin.
   a) soort dier
   b) soort feest
   c) soort boom
   d) soort brievenbus
   e) Ik weet het echt niet.

10. Mag ik jouw *kam* even lenen?

a) ding waarmee je schrijft
b) ding waarmee je rekent
c) ding waarmee je het eten snijdt
d) ding waarmee je je haren netjes maakt
e) Ik weet het echt niet.

11. Hij kocht *onroerend goed*.
    a) stukken grond en gebouwen
    b) prachtige kleren
    c) mooie boeken
    d) beelden en schilderijen
    e) Ik weet het echt niet.
12. Het gaat *allengs* beter met haar.
    a) sinds lange tijd
    b) helemaal niet
    c) natuurlijk
    d) langzamerhand
    e) Ik weet het echt niet.
13. Hij is een *ordelijke* man.
    a) Hij is stil en eerlijk.
    b) Hij is regelmatig en netjes.
    c) Hij is rijk en gelukkig.
    d) Hij is gezellig en vrolijk.
    e) Ik weet het echt niet.
14. Ze gaan dat gebied *afbakenen*.
    a) Ze gaan de grenzen van dat gebied aangeven.
    b) Ze gaan dat gebied mooier maken.
    c) Ze gaan de bomen die op dat gebied staan weghalen.
    d) Ze gaan dat gebied veiliger maken.
    e) Ik weet het echt niet.

15. Heb jij geen last van *wroeging*?
    a) rugpijn
    b) een schuldgevoel
    c) nieuwsgierigheid
    d) boosheid
    e) Ik weet het echt niet.
16. Het *traject* is erg lang.
    a) de weg
    b) de trein
    c) de wachttijd
    d) de brug
    e) Ik weet het echt niet.
17. Ik heb in het museum een *harnas* gezien.
    a) postzegel van grote waarde die men vroeger verzamelde
    b) hoge koffer die men vroeger mee op reis nam
    c) ijzeren jas die vroeger tegen wapens beschermde
    d) groot mes dat men vroeger in de strijd gebruikte
    e) Ik weet het echt niet.
18. Die man heeft een *weerbarstig* karakter.
    a) Die man is meestal in een gezellige stemming.
    b) Die man begint snel te huilen.
    c) Die man voelt zich vaak onzeker.
    d) Die man is niet makkelijk te overtuigen.
    e) Ik weet het echt niet.
19. Ik durfde het in een moment van *overmoed*.
    a) Ik durfde het omdat ik ineens grote spanning voelde.
    b) Ik durfde het omdat ik me ineens heel sterk voelde.
    c) Ik durfde het omdat ik ineens een groot verlangen voelde.
    d) Ik durfde het omdat ik dacht dat het moest.

e) Ik weet het echt niet.

20. Zij bracht haar horloge naar de *lommerd*.
    a) iemand die klokken maakt
    b) iemand die de waarde van dure dingen beoordeelt
    c) plaats waar je geld kunt lenen als je een ding van jezelf achterlaat
    d) plaats waar je dure dingen kunt kopen
    e) Ik weet het echt niet.
21. Krijg jij ook een *toelage*?
    a) brief waarin staat dat je iets mag doen
    b) een goed cijfer voor de test
    c) woorden waarmee je iets duidelijk maakt
    d) geld om van te leven dat je regelmatig krijgt
    e) Ik weet het echt niet.
22. Er is aldoor *geharrewar* op mijn werk tegenwoordig.
    a) ruzie
    b) herhaling
    c) plezier
    d) drukte
    e) Ik weet het echt niet.
23. Hij zag *asgrauw*.
    a) Hij had grijs haar.
    b) Zijn gezicht had geen kleur.
    c) Hij kon niet goed zien.
    d) Hij zag het vuur uitgaan.
    e) Ik weet het echt niet.
24. Een *knoestige* boom.
    a) Een boom met een bepaald soort vruchten.
    b) Een boom van een bepaalde leeftijd.
    c) Een boom met een bepaalde vorm.

d) Een boom met een bepaald soort bladeren.

e) Ik weet het echt niet.

25. Dat zou ik anders *inschatten*.

a) beoordelen

b) aankondigen

c) opschrijven

d) doen

e) Ik weet het echt niet.

26. In dat land heerst *cholera*.

a) Het gaat in dat land slecht met de economie.

b) Het heeft in dat land allang niet geregend.

c) Veel mensen in dat land hebben een bepaalde ziekte.

d) Men voert al jaren oorlog in dat land.

e) Ik weet het echt niet.

27. Dat kunnen we niet langer *verhelen*.

a) mooier maken dan het is

b) verbergen, verzwijgen

c) stiekum verkopen

d) repareren

e) Ik weet het echt niet.

28. Zij voerden een *franke* discussie.

a) eerlijke

b) beleefde

c) felle

d) rustige

e) Ik weet het echt niet.

29. Zij kreeg *wanten* voor haar verjaardag.

a) dingen die zorgen dat je handen niet koud worden

b) dingen die zorgen dat je voeten niet koud worden

c) dingen die zorgen dat je oren niet koud worden
d) dingen die zorgen dat je knieën niet koud worden
e) Ik weet het echt niet.

30. Dat *frappeert* me.
    a) Dat maakt me aan het schrikken.
    b) Dat vind ik opvalllend.
    c) Dat doet me pijn.
    d) Dat verveelt me.
    e) Ik weet het echt niet.
31. Onze *primus* ging al na een paar dagen kapot.
    a) terras-verwarmer
    b) gaskachel
    c) gasbrander om op te koken
    d) olielamp
    e) Ik weet het echt niet.
32. Zij is een *bolleboos*.
    a) Zij drinkt erg veel.
    b) Zij doet erg veel aan sport.
    c) Zij kan erg goed leren.
    d) Zij vindt zichzelf erg mooi.
    e) Ik weet het echt niet.
33. Veel mensen vinden deze voetballer nogal *flegmatiek*.
    a) wisselvallig, onzeker
    b) agressief
    c) kalm, traag
    d) gevoelig
    e) Ik weet het echt niet.
34. We werden ermee *overvoerd*.
    a) We kregen er te veel van.

b) We konden onze gevoelens niet de baas blijven.

c) We werden tegen onze zin in meegenomen.

d) We kregen een ongeluk.

e) Ik weet het echt niet.

35. *Dat zit wel snor*.

a) Dat is goedkoop.

b) Dat staat je goed.

c) Dat is niet gelukt.

d) Dat is in orde.

e) Ik weet het echt niet.

36. Zij heeft iets moois *gewrocht*.

a) beleefd

b) gehoord

c) gemaakt

d) kapot gemaakt

e) Ik weet het echt niet.

37. De wereld van het internet is heel *libertijns*.

a) flexibel

b) zonder regels

c) vrij, vrijzinnig

d) makkelijk

e) Ik weet het echt niet.

38. Hij is lid van de *schutterij*.

a) vereniging die dieren beschermt

b) vereniging die de natuur beschermt

c) muziekvereniging

d) schietvereniging

e) Ik weet het echt niet.

39. Ik heb de *smoor* in.

a) Ik ben in een slechte stemming.
b) Ik kan geen lucht krijgen.
c) Ik voel me gelukkig.
d) Ik slaap slecht.
e) Ik weet het echt niet.

40. De zaak wordt *geseponeerd*.
    a) niet verder behandeld
    b) ergens anders voortgezet
    c) op een later tijdstip behandeld
    d) in tweeën gedeeld
    e) Ik weet het echt niet.
41. Die vraag wordt veel door *exegeten* onderzocht.
    a) iemand die de natuur bestudeert
    b) iemand die heilige teksten bestudeert
    c) iemand die geinteresseerd is in zwarte magie
    d) iemand die geinteresseerd is in het leven na de dood
    e) Ik weet het echt niet.
42. Er was een *oploop* op straat.
    a) Er was een voorstelling.
    b) De mensen liepen naar één punt.
    c) Er gebeurde een ongeluk.
    d) Er werd aan de straat gewerkt.
    e) Ik weet het echt niet.
43. Hij ligt onder de *hoogtezon*.
    a) de zon op het heetst van de dag
    b) ding waar je bruin van wordt
    c) ding om je tegen de zon te beschermen
    d) ding om je tegen de regen te beschermen
    e) Ik weet het echt niet.

44. Hij bracht een *jobstijding*.
    a) bericht over een nieuwe baan
    b) erg slecht bericht
    c) bericht over geld
    d) slecht weerbericht
    e) Ik weet het echt niet.
45. Ze zei het met een *aanminnig* lachje.
    a) mooi
    b) geheimzinnig
    c) gemeen, vals
    d) charmant
    e) Ik weet het echt niet.
46. Ik ga die boeken *kaften*.
    a) terug naar de bibliotheek brengen
    b) beschermend papier eromheen doen
    c) in een bepaalde orde in de boekenkast zetten
    d) bij de bibliotheek lenen
    e) Ik weet het echt niet.
47. Zijn familie vindt hem al weken lang *lethargisch*.
    a) afwezig
    b) vervelend
    c) slaperig, moe
    d) suf, inactief
    e) Ik weet het echt niet.
48. Voor een kind is zoiets *nefast*.
    a) slecht, schadelijk
    b) gezond, goed voor de groei
    c) onmisbaar
    d) fataal

e) Ik weet het echt niet.

49. Misschien moeten we de tafel laten *politoeren*.
    a) professioneel reinigen
    b) oppoetsen, glanzend maken
    c) restaureren, repareren
    d) taxeren
    e) Ik weet het echt niet.
50. Het is daar erg *gehorig*.
    a) Je hoort daar geluiden van de buren erg goed.
    b) Men luistert daar veel naar muziek.
    c) Er is erg veel lawaai daar.
    d) Er wordt veel daarover gepraat.
    e) Ik weet het echt niet.
51. De *vendutie* is waarschijnlijk volgende week.
    a) openbare verkoop, veiling
    b) uitspraak van de rechter
    c) opening
    d) rechtzaak
    e) Ik weet het echt niet.
52. Hij is een *slokop*.
    a) iemand die veel bier drinkt
    b) iemand die snel en veel eet
    c) iemand die nogal langzaam is
    d) iemand die te dik is
    e) Ik weet het echt niet.
53. De datum staat onder het *epitaaf*.
    a) tekst op een grafsteen
    b) versiering
    c) stempel

d) tekst voorin een boek

e) Ik weet het echt niet.

54. De *ondertiteling* valt weg.

a) goede naam

b) steun met geld

c) tekst onder televisiebeelden

d) uitbreiding van een titel

e) Ik weet het echt niet.

55. Je moet je geld tegen *ontwaarding* beschermen.

a) waardevermindering

b) diefstal

c) het niet onvangen van rente

d) rente-aftrek

e) Ik weet het echt niet.

56. Een *tentatieve* lijst is te vinden op internet.

a) verkeerde, foute

b) uitgebreide, volledige

c) voorlopige, tijdelijke

d) realistische

e) Ik weet het echt niet.

57. In deze *hagiografie* kun je daarover meer lezen.

a) beschrijving van de geschiedenis van een stad

b) beschrijving van de geschiedenis van een familie

c) beschrijving van een bepaalde periode

d) beschrijving van het leven van heiligen

e) Ik weet het echt niet.

58. Het taalgebruik op deze website is *affreus*.

a) informeel, niet zakelijk

b) slordig, niet netjes

c) walgelijk, schandelijk

d) onjuist, fout

e) Ik weet het echt niet.

59. Volgens mij is *chicaneren* haar hobby.

    a) je overal mee bemoeien

    b) kaarten, een speciaal kaartspel

    c) klagen, zeuren

    d) handwerken, een speciale haaktechniek

    e) Ik weet het echt niet.

60. Ik draag het liefst een *duffel*.

    a) lange regenjas

    b) grote, zware jas

    c) dikke trui

    d) pyjama

    e) Ik weet het echt niet.

Definition test

1. Een dier dat blaat. – Schaap.
2. Iemand die werkt met meel. – Bakker.
3. Vrouwtje van een kater. – Poes.
4. Iemand die werkt met eten. – Kok.
5. Vrouwtje van een reu. – Teef.
6. Lichaamsdeel om te ruiken. – Neus.
7. Iemand die werkt met patienten. – Arts.
8. Een dier dat blaft. – Hond.
9. Iemand die werkt met vlees. – Slager.
10. Vrouwtje van een hengst. – Merrie.

11. Voorwerp om mee te roeren. – Lepel.
12. Iemand die werkt met klokken. – Horlogemaker.
13. Lichaamsdeel om mee te proeven. – Tong.
14. Iemand die werkt met verf. – Schilderer.
15. Vrouwtje van een bok. – Geit.
16. Iemand die werkt met kip. – Poelier.
17. Lichaamsdeel om mee te zien. – Oog.
18. Iemand die werkt met eetgewoonten. – Dietist.
19. Een dier dat knort. – Varken.
20. Een dier dat hinnikt. – Paard.

Multiple-choice antonym test

1. kritiek
   a) ordening
   b) lof
   c) antwoord
   d) bejaarde
   e) theorie
2. aan
   a) terug
   b) mild
   c) door
   d) uit
   e) op
3. stijf
   a) lang
   b) vast
   c) gezellig
   d) flexibel
   e) selfzuchtig
4. gierig
   a) gul
   b) vrolijk
   c) handig
   d) grof
   e) koppig
5. bars
   a) groot
   b) vriendelijk
   c) roemorig
   d) aanwezig
   e) intelligent

6. vallen
   a) opstaan
   b) halen
   c) verdelen
   d) melden
   e) stappen
7. nauw
   a) kapot
   b) laag
   c) wijd
   d) tegen
   e) donker
8. hetzelfde
   a) zwaar
   b) verschillende
   c) blij
   d) iemand
   e) dergelijk
9. dwerg
   a) kleinkind
   b) kinky
   c) reus
   d) eender
   e) verdrietig
10. aanbod
    a) offer
    b) toekomst
    c) ongeluk
    d) bieding
    e) vraag
11. studiosus
    a) flat
    b) gesprek
    c) luiaard
    d) luisteraar
    e) leraar
12. minst
    a) nietig
    b) afwezig
    c) dezelfde
    d) licht
    e) meest
13. dapper
    a) helder
    b) nieuw
    c) krachtig
    d) bang
    e) permantig
14. mals
    a) hard
    b) straks
    c) soms
    d) netjes
    e) donzig
15. contra
    a) achter
    b) snel
    c) echt
    d) pro

e) alleen

16. bucolisch
    a) klein
    b) rustig
    c) urbaan
    d) duidelijk
    e) uitvoerig

17. prullerig
    a) altijd
    b) aardig
    c) slank
    d) kostbaar
    e) minuscuul

18. bedorven
    a) gekookt
    b) antiek
    c) anzienlijk
    d) voluit
    e) vers

19. interessant
    a) saai
    b) prettig
    c) zichtbaar
    d) juist
    e) ruim

20. moe
    a) soortgelijk
    b) samen
    c) raak
    d) zuur
    e) uitgerust

21. pessimisme
    a) humeur
    b) aankomst
    c) optimisme
    d) mogelijkheid
    e) plezier

22. aankomen
    a) stijgen
    b) vertrekken
    c) vergeten
    d) slapen
    e) rijden

23. ontwaken
    a) sluipen
    b) inslapen
    c) aankleden
    d) instappen
    e) brengen

24. achter
    a) onder
    b) hel
    c) slap
    d) beneden
    e) voor

25. winst
    a) verlies
    b) bol

c) dorst

d) hoop

e) moed

Open antonym test

1. Zaaien
   Antonym: Oogsten
2. Leugen
   Antonym: Waarheid
3. Nadeel
   Antonym: waarheid
4. Minimaal
   Antonym: Maximaal
5. Unanimiteit
   Antonym: onenigheid
6. Geforceerd
   Antonym: vrijwillig
7. Mager
   Antonym: Dik
8. Officieus
   Antonym: Officieel
9. Lawaai
   Antonym: Stilte
10. Traag
    Antonym: Sneel
11. Success
    Antonym: Mislukking
12. Bevestiging
    Antonym: Ontkenning
13. Negatief
    Antonym: Positief
14. Absent
    Antonym: Present
15. Goedkoop
    Antonym: Duur
16. Fluisteren
    Antonym: Schreeuwen
17. Passief
    Antonym: Actief
18. Vijand
    Antonym: Vriend
19. Overwinning
    Antonym: Nederlaag
20. Theoretisch
    Antonym: Praktisch
21. Ouderwets
    Antonym: Modern
22. Deficientie
    Antonym: Overdaad
23. Legaal
    Antonym: Illegaal
24. Tanen
    Antonym: Toenemen
25. Monochroom
    Antonym: Bont

Multiple-choice synonym test

1. vlug
   a) water
   b) snel
   c) hoekig
   d) aal
   e) hast
2. vlijt
   a) arm
   b) sterk
   c) ijver
   d) rijk
   e) vermoeid
3. staven
   a) lopen
   b) wanen
   c) bekrachtigen
   d) verrijken
   e) verwaalozen
4. amper
   a) stroom
   b) geneesmiddel
   c) groente
   d) nauwelijks
   e) netjes
5. schamel
   a) warboel
   b) kameel
   c) verbinding
   d) armoedig
   e) kleding
6. observeren
   a) luisteren
   b) bekijken
   c) optellen
   d) vermaken
   e) trouwen
7. karig
   a) smal
   b) kaal
   c) koud
   d) dun
   e) zuinig
8. ruiterlijk
   a) verzorgend
   b) openhartig
   c) snel
   d) branie
   e) durf
9. verleiding
   a) voorstel
   b) toeval
   c) achtergrond
   d) tegenvaller
   e) bekoring
10. vermetel

a) liniaal
b) gemeten
c) gewaagd
d) zwak
e) precies

11. blaam
a) smet
b) struik
c) blaasje
d) roddel
e) vrucht

12. vermoeiend
a) eerlijk
b) muf
c) hooghartig
d) inspannend
e) dramatisch

13. cursief
a) nauwsluitend
b) schuingedrukt
c) gauw
d) middel
e) leesbaar

14. miniem
a) aftrekken
b) verlies
c) verschil
d) mimiek
e) onbeduidend

15. abrupt
a) zwaar
b) apart
c) plotseling
d) vergeefs
e) onvoldoende

16. fameus
a) beroemd
b) waardig
c) bekend
d) aardig
e) beleefd

17. fragiel
a) vaartuig
b) lelie
c) waterloop
d) broos
e) licht

18. bolster
a) rond
b) planeet
c) buitenkant
d) goedheid
e) sloof

19. nestor
a) meester
b) eerwaarde
c) insect
d) heer
e) oudste

20. bombast
    a) vermomming
    b) schil
    c) granaat
    d) gezwollenheid
    e) boomsoort
21. frenetiek
    a) krachtig
    b) verlegen
    c) lichamelijk
    d) spiritueel
    e) verwoed
22. geschikt
    a) lastig
    b) samen
    c) passend
    d) soms
    e) duister
23. pendant
    a) tegenhanger
    b) klok
    c) verwaand
    d) groots
    e) ijdel
24. concies
    a) bondig
    b) beheer
    c) vasthoudend
    d) samenvatting
    e) aandacht
25. frugaal
    a) dubbelzinnig
    b) belemmerd
    c) sober
    d) breekbaar
    e) helder

Open synonym test

1. Woeden
   Synonym: Razen
2. Gekibbel
   Synonym: Ruzie
3. Verbinden
   Synonym: Schakelen
4. Overstijgen
   Synonym: Verwinnen
5. Koppig
   Synonym: Stug
6. Attitude
   Synonym: Houding
7. Latent
   Synonym: Onderliggend
8. Floreren
   Synonym: Bloeien, gedijen

9. Frequent
   Synonym: Vaak
10. Gewaand
    Synonym: vermeend
11. Permanent
    Synonym: Altijd
12. Mordicus
    Synonym: Hardnekkig
13. Morsig
    Synonym: Smerig
14. Tillen
    Synonym: Heffen
15. Obsederen
    Synonym: Bekijken
16. Beleefd
    Synonym: Hoofs
17. Vrolijk
    Synonym: Jolig
18. Emaneren
    Synonym: Uitstralen
19. Palliatief
    Synonym: Pijnstillend
20. Toezicht
    Synonym: Controle
21. Opulent
    Synonym: Copieus
22. Loyaal
    Synonym: Trouw
23. Caprice
    Synonym: Humeur
24. Wijs
    Synonym: Slim
25. Naarstig
    Synonym: Ijverig

Appendix B

Table 2.7: Reaction times (ms) for correct responses in the lexical decision task per condition. For illustration purposes, frequency is shown as a categorical variable although all models were run on frequency as a continuous variable.

Lexicality	**Frequency**	***Mean***	***SD***
	low	661	192
Words	medium	611	180
	high	591	174
	total	621	185
Nonwords	-	718	244

Note: Low-frequency words had frequency values of less than 1 count per million in the SUBTLEX corpus (M = 0.36; SD = 0.27), medium-frequency items between 1 and 10 counts per million (M = 3.47; SD = 2.34), and high-frequency words between 10 and 90 counts per million (M = 25.24; SD = 20.92).

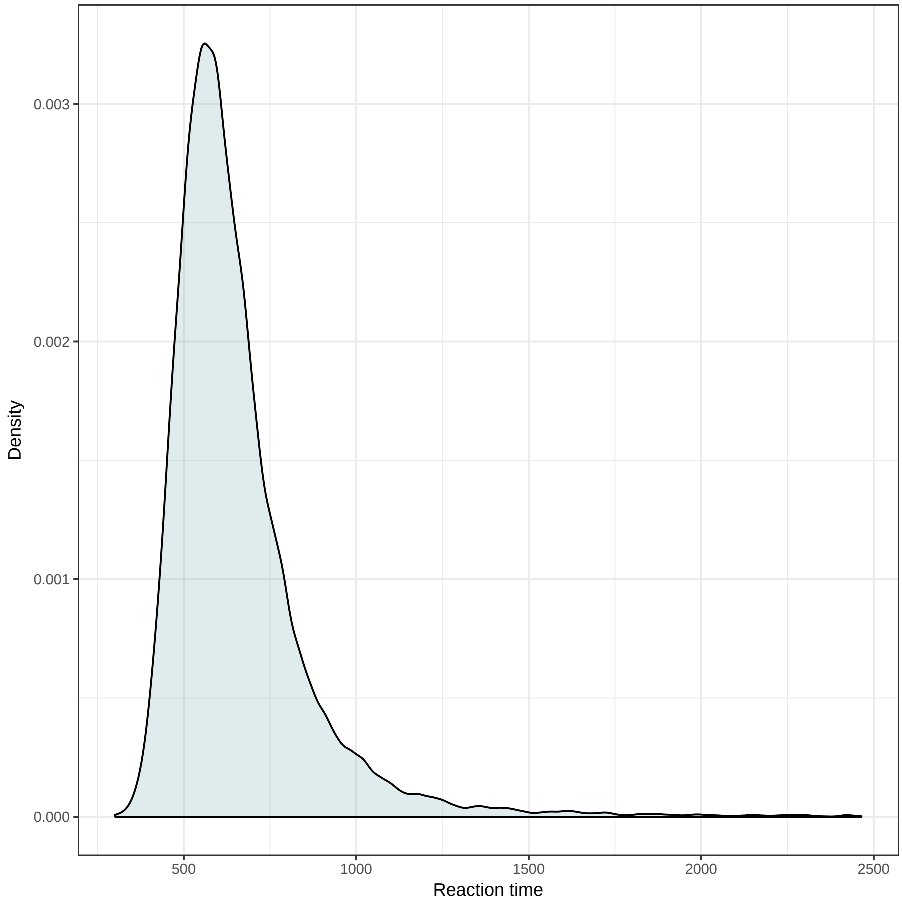

Figure 2.2: Distribution of the lexical decision RTs.

Appendix C

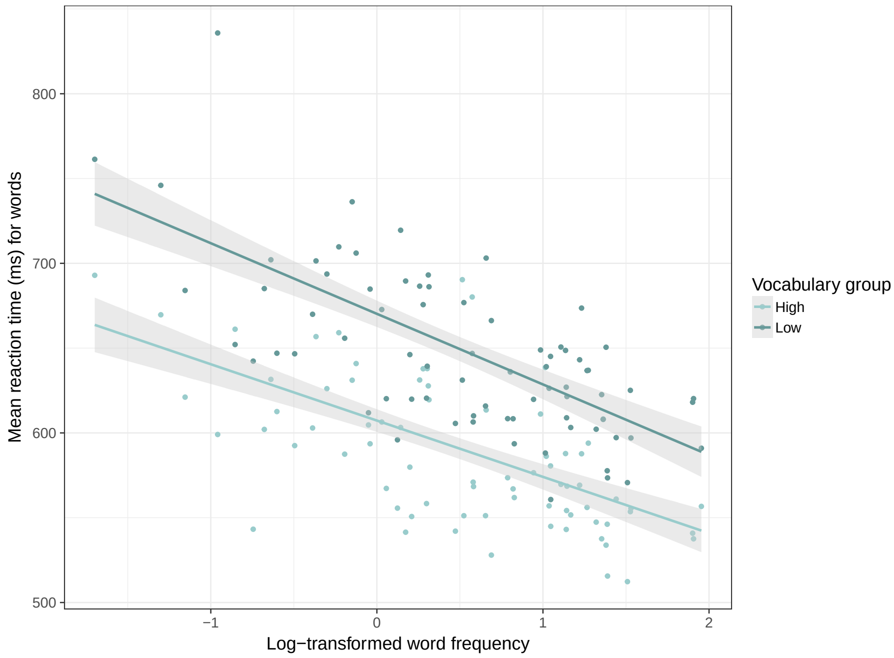

Figure 2.3: RT as a function of log-transformed word frequency for low- vs. high-vocabulary individuals. 95% confidence intervals are displayed in grey. Note that for illustration purposes, vocabulary score was transformed into a categorical variable with two levels performing a median split. The above-described model is run with vocabulary score as a continuous variable.

3 | Beyond the "typical" participants: Vocabulary and lexical processing in non-university students[1]

Abstract

Previous research has indicated an effect of individual differences in vocabulary on language processing. Greater word knowledge has been associated with faster and more accurate language production and comprehension. These findings are, however, mainly based on studies testing university students. It is questionable whether that is really representative of the population, hence, whether these studies provide a comprehensive picture of the relationship between vocabulary and language processing performance. The present study aimed at filling this gap. For this purpose, we administered a battery of six vocabulary tests and a visual lexical decision task to a group of young adult non-university students. Deviating from previous findings, the results show no main effect of word knowledge on lexical decision RTs; only a significant vocabulary effect on accuracy. Furthermore, in line with previous studies, individuals with greater vocabularies showed a smaller word frequency effect in the lexical decision task. Implications for measuring vocabulary size in native speakers and for future individual differences studies are discussed.

[1]This and the previous chapter are based on *Mainz, N., Shao, Z., Brysbaert, M., & Meyer, A.S. (2017). Vocabulary knowledge predicts lexical processing: Evidence from a group of participants with diverse educational backgrounds. Frontiers in Psychology, 8, 1164. doi: 10.3389/fpsyg.2017.01164*

Introduction

A growing body of research focuses on the effects of individual differences in various cognitive abilities on language processing performance (Banks et al., 2015; Hartsuiker & Barkhuysen, 2006; Jongman et al., 2015; Shao, Meyer, & Roelofs, 2013; Shao, Roelofs, et al., 2014; Shao et al., 2012). Vocabulary size is one of the variables that has been suggested to predict native speakers' language processing. As replicated in Chapter 2 of this dissertation, increased vocabulary size has been associated with better, i.e. faster and more accurate, performance on different linguistic tasks (see also, Brysbaert et al., 2016a; Janse & Jesse, 2014; Rodriguez-Aranda & Jakobsen, 2011).

However, just as in the vast majority of psychological investigations, these individual differences studies have examined undergraduate students. Hence, conclusions about the relationship between language processing performance and, for instance, vocabulary size are based on a rather small and presumably quite homogeneous group of people. It is questionable whether this provides a representative picture of i) the range of abilities present in the population, and ii) the role of vocabulary in language processing. This study aimed at filling this gap in previous research by investigating the effects of individual differences in vocabulary size on word recognition in non-university students.

In previous investigations, better vocabulary knowledge has, for instance, been found to have beneficial effects on older adults' use of predictive information in spoken contexts (Federmeier et al., 2002). Furthermore, an increase in vocabulary size has been related to more accurate spoken word recognition in older adults (Janse & Jesse, 2014) and better listening comprehension in young adults (Andringa et al., 2012). Additionally, higher vocabulary scores were found to be correlated with more accurate and faster word recognition in lexical decision and speeded pronunciation tasks (Yap et al., 2012). Better vocabulary knowledge has not only been associated with improved language comprehension but also with faster language production in various tasks, such as picture naming and verbal fluency (Rodriguez-Aranda & Jakobsen, 2011). In verbal fluency tasks participants are asked to produce as many words as possible within a semantic category (category fluency) or starting with a given letter (letter fluency) within one minute. Greater word knowledge was found to be related to more items generated in verbal fluency tasks (Unsworth et al., 2011). In addition, Shao, Janse, Visser, and Meyer (2014) report that in letter as well as category fluency tasks, individuals with larger vocabularies were faster to initiate their response to the cue than individuals with weaker vocabularies. In a nutshell, better vocabulary knowledge has been associated with advantages in various measures assessing language comprehension and production performance.

Besides faster RTs in language processing tasks, previous research has found smaller word frequency effects for speakers and readers with more print exposure (Chateau & Jared, 2000; Kuperman & Van Dyke, 2013) or larger vocabularies (Diependaele et al., 2013; Yap et al., 2009). With increasing word frequency, RTs typically decrease. This RT difference between low- and high-frequency words has been shown to decrease with increasing vocabulary size; hence, the word frequency effect appears to be smaller for individuals with better vocabulary knowledge. Diependaele and colleagues (2013), for instance, reanalysed data from an earlier visual word recognition study (Lemhöfer et al., 2008) and found that individuals with higher vocabulary scores showed smaller effects of word frequency. Furthermore, Brysbaert, Lagrou, and Stevens (2016a) observed that larger vocabularies were associated not only with faster lexical decision RTs but also with smaller effects of word frequency on the RTs. In both studies, this frequency by skill interaction was argued to support the *lexical entrenchment hypothesis*, which postulates differences in entrenchment between smaller and larger vocabularies. This means that the representations in individuals with greater word knowledge are hypothesised to be more robust or distinct, enabling faster processing, as compared to individuals with smaller vocabularies. According to the lexical entrenchment hypothesis, the frequency x skill interaction is a result of differences in language exposure. The amount of exposure is presumed to have a particularly strong impact on the exposure to, and therefore representations of, low-frequency words (Kuperman & Van Dyke, 2013). Thus, the lexicon of individuals with limited exposure to language and therefore weaker vocabulary knowledge is hypothesised to have a steeper frequency curve than individuals with larger vocabularies (Brysbaert et al., 2016a). As a result, the vocabularies of individuals with poorer word knowledge (due to limited language exposure) are assumed to show a stronger frequency difference between low- and high-frequency words. The lexical representations in low-vocabulary individuals are hypothesised to be weaker, less robust, entrenched, or distinct, especially for low-frequency words, and therefore slower to be processed.

Yap, Tse, and Balota (2009) put forward a similar argument, claiming that overall increased precision and stability of lexical representations is a result of greater vocabulary knowledge. This was based on their observation that vocabulary knowledge affects the joint effects of word frequency and associative priming. In a lexical decision task, participants with poorer vocabulary knowledge exhibited stronger associative priming effects for low-frequency than for high-frequency words, while individuals with better vocabulary scores showed equally strong priming effects for all words. Yap et al. (2009) have taken this to indicate that the lexical representations in readers with higher vocabulary scores do not differ much in quality or strength depending on word frequency. The representations in low-vocabulary individuals, by contrast, appear to show considerable differences in strength or robustness depending on word frequency.

To sum up, representational differences between vocabularies of varying sizes have been proposed based on the fact that low-vocabulary individuals showed stronger word frequency effects on word recognition than high-vocabulary individuals. This idea has been named with different terms: Representations in larger vocabularies have been hypothesised to be more *distinct*, *precise*, *robust*, *entrenched*, or higher in *lexical quality*, leading to increased lexical access speed and smaller word frequency effects (Diependaele et al., 2013; Van Dyke et al., 2014; Perfetti & Hart, 2001; Yap et al., 2009).

All of the aforementioned studies suggest a role of vocabulary size in comprehension and production. However, most of them did not focus on vocabulary in particular and therefore only used single vocabulary tests to assess word knowledge. Bowles and Salthouse (2008) have claimed, though, that it is necessary to use a variety of measures to assess a skill as complex as vocabulary knowledge. This was considered to be particularly important in studies focusing on individual differences in vocabulary size. The reason for this, as argued by Bowles and Salthouse (2008), is that vocabulary tests are never a pure measure of word knowledge but involve other cognitive abilities. Using a composite score of vocabulary reflecting participants' performance on a battery of different types of tests is a way to obtain a relatively reliable measure of vocabulary size. Therefore, the same vocabulary test battery that was used in the experiment presented in Chapter 2 of this dissertation was employed in this study.

Additionally, as indicated before, most of the insights into the role of vocabulary size reported on above are based on studies examining undergraduate university students. Hence, conclusions and theories about language processing and the lexicon are grounded in findings from a small, homogeneous, and in some sense special group of people. They are special to the extent that the amount and type of language exposure they get is very likely to be notably different from that of non-university students. In addition, most participants that are tested in psychological studies are very experienced participants, meaning that they have had quite a lot of practice with completing different kinds of experimental tasks. This may affect their performance, at least to a certain extent and in certain tasks. Investigations testing participants from a broader range of educational backgrounds is essential and long overdue. This is the case not only for reasons of societal impact but also to ensure that assumptions and theories about language capacities and processing are grounded in observations that are representative for the entire population, not just a very specific group of young adults.

With the present study we set out to address these issues by investigating the relationship between individual differences in vocabulary and word recognition in a group of young adults who are not university students but vocational college students. Vocational education in the Netherlands is subdivided into four different levels. Level four can be assumed to be closest to university education. The groups in levels one, two, and three are comprised of people who will continue their education by entering the next level while others finish after completing their current level of education. Thus, this group of participants is rather varied and should display a considerably broader range of vocabulary knowledge than university students.

The use of a battery of different vocabulary tests was considered particularly important in this study as the group of participants was assumed to be more heterogeneous and less experienced with psycholinguistic testing; hence, it was expected that they might display more variation in vocabulary test performance both across and possibly also within participants.

Thus, the participants in the present study completed almost the same set of vocabulary tests as had been used in the experiment reported on in Chapter 2. The test battery was comprised of six vocabulary tests, two of which were the established Peabody Picture Vocabulary Test (PPVT-III NL; Schlichting, 2005) and Andringa and colleagues' (2012) receptive multiple-choice test. In addition to these multiple-choice tests, four additional newly developed measures were administered: A definition test, multiple-choice antonym and synonym tests, and an open antonym test. The open synonym test was excluded from this study as it turned out to be extremely challenging for the university student population. We did not want to overtax and as a result demoralise the participants with an overly difficult test, and therefore decided to exclude it from the test battery.

In addition, participants' language processing performance was assessed using a visual lexical decision task, which is a widely used measure of speed of word recognition (e.g. Balota et al., 2007; Brysbaert et al., 2016a; Keuleers et al., 2012). On each trial of a lexical decision task, participants are presented with a string of letters and are asked to decide whether or not it is an existing word in a given language. Typical findings in the lexical decision task are effects of lexicality and frequency. Reaction times (RTs) for words are usually faster than RTs for nonwords, and more frequent words elicit faster responses than less frequent words (e.g. Keuleers et al., 2015; Kuperman & Van Dyke, 2013; Yap et al., 2012). In the present task, frequency manipulations were implemented such that the stimuli covered a large range of word frequencies. We used the same lexical decision task as in the experiment reported in Chapter 2.

We expected to find increased vocabulary knowledge to predict improved language processing performance, with faster RTs and lower error rates for participants with greater vocabulary knowledge. Furthermore, we tested the lexical entrenchment hypothesis, which predicts an interaction between a participant's vocabulary score and the word frequency effect. Larger effects of word frequency were predicted for individuals with poorer vocabulary scores (e.g. Diependaele et al., 2013).

In addition, the nature of a potential effect of vocabulary on language processing was examined more closely. For this purpose, we used Ratcliff's (1978) diffusion model approach to analyse the lexical decision task data (Ratcliff, 1978; Ratcliff et al., 2004; Yap et al., 2012, for a discussion). We looked at the same diffusion model parameters as described in Chapter 2 of this dissertation. A detailed description of the diffusion model approach and its application to lexical decision task data can be found in Chapter 2.

We expected the drift rate to increase with increasing vocabulary size, indicating faster build-up of lexical information in individuals with greater vocabulary knowledge. Furthermore, based on previous studies it was predicted that increased vocabulary knowledge would be associated with a reduced word bias in the starting point parameter (Brysbaert et al., 2016a; Yap et al., 2012, see also Chapter 2).

Hence the present study aimed at filling a gap in the research on vocabulary size and language processing by examining participants that are not usually tested in psychological studies, namely young adult vocational college students. For this purpose we employed a battery of vocabulary size measures and a visual lexical decision task to assess word recognition performance. Additionally, the potential vocabulary effect on language processing was examined more closely by using diffusion model analyses. In this vein we hoped to get insights into where, i.e. in which component processes of lexical decision making, an overall effect of vocabulary on lexical decision task performance might originate.

Method

Participants

A total of 231 young adults gave informed consent to participate in this study. The results obtained from fifty-seven individuals were excluded due to failure to perform one or several of the tasks correctly. Performance on the PPVT or the lexical decision task was in most cases the reason for excluding a participant. Hence, data of 174 participants (92 females) aged between 18 and 32 years (M = 20.3; SD = 2.7) were left for further analyses. All of them were students at vocational colleges in the Netherlands (ROC Nijmegen, ROC Tilburg, ROC Midden Nederland). The vocational training in the Netherlands is divided into four different levels, with level one being the lowest. Table 3.1 shows the distribution of participants across the different levels of vocational education. A more even spread across the levels would certainly have been desirable, but could not be attained due to timetabling constraints in the colleges.

All participants were recruited through the teachers at their vocational colleges. Participation was voluntary and not part of compulsory classes. In some of the cases, the schools were paid an expense allowance of 10 Euros per participant to spent on teaching materials; in other schools, participants were paid 10 Euros each for their participation. Ethical approval was granted by the Faculty of Social Sciences of the Radboud University Nijmegen.

Table 3.1: Numbers of female and male participants per level of vocational education.

Level	Female	Male	Total
1	3	-	3
2	14	-	14
3	24	27	51
4	50	56	106
total	92	83	174

Materials and design

The materials of the lexical decision task were the same as in the Experiment reported on in Chapter 2 of this dissertation, whereas the vocabulary test materials were slightly different. As mentioned before, the open synonym test was not administered because the scores achieved by the university students indicated that it would probably be very challenging and potentially frustrating for the vocational college students without providing valuable additional insights.[2] Furthermore, five additional high-frequency filler words were added to the multiple-choice antonym and synonym tests as well as the open antonym test, respectively. This was done in order to increase the number of relatively easy items and keep participants motivated throughout the test. These filler items were excluded from the final test score.

Apparatus

All tasks were administered using 14-inch HP laptops (Probook 640 G1) and Panasonic RP-HT030 headphones. All tests were implemented using Presentation software (version 16.5, www.neurobs.com).

[2]Participants in Experiment 1 obtained a relatively low mean score of 10.70 (SD = 2.86) as compared to the open antonym test (M = 19.29, SD = 2.37).

Procedure

Participants were tested in groups of between 9 and 30 students in their classrooms. They completed the tasks in the following fixed order: 1) Definition test, 2) Andringa et al.'s (2012) receptive multiple-choice test, 3) Multiple-choice antonym test, 4) Open antonym test, 5) Multiple-choice synonym test, 6) PPVT, 7) Lexical decision task. The procedure for the vocabulary tests was the same as explained in Chapter 2 of this dissertation.

In contrast to this, the procedure for the lexical decision task was slightly altered. First, a pilot study with twenty vocational college students showed that presenting the stimuli for 3 seconds was too short. Therefore, the presentation time was increased to 5 seconds. Secondly, the response buttons were kept constant, with "M" to be pressed for words and "Z" for nonwords. This was done to facilitate administering the task in a group setting.

Analyses

The vocabulary tests were scored and analysed in exactly the same way as in the study reported on in Chapter 2 of this dissertation. The only difference was that the five items added to the multiple-choice antonym and synonym tests and the open antonym test were excluded from the test scores, as mentioned before.

Furthermore, the same analyses were run on the vocabulary and lexical decision data from this study as in the experiment presented in Chapter 2. Hence, participants' vocabulary test scores were analysed using bivariate correlations and a Principal Component Analysis (PCA) in SPSS (version 20). Lexical decision accuracy for words as well as RTs on correct word trials were analysed in different mixed-effects models with either lexicality or word frequency as predictors. All models included random intercepts for both item and participant, and per-participant random slope adjustments to the lexicality or word frequency effect. All analyses were run in R (R Core Team, 2016) using the glmer function from the package lme4 (version 1.1.12, Bates et al., 2015).

The individual difference analyses were also the same as in Chapter 2. Thus, vocabulary score was added as a predictor to the above-described models. Different mixed models were run, each including participants' scores from one of the individual tests as a predictor. In addition, a composite score of vocabulary knowledge based on each participant's performance on the battery as a whole was calculated and entered as a predictor of lexical decision accuracy and RTs in a separate model.

The composite vocabulary score was obtained by calculating regression-based factor scores for each participant using the PCA method in SPSS (DiStefano et al., 2009). Assuming only one underlying factor, each individual's loading or score on that factor based on their six vocabulary test scores was calculated. By doing this, the measurements of word knowledge were collapsed into one number with a mean of zero. Hence, we were able to compare the individual vocabulary measures with a composite measure reflecting performance on the entire battery of tests.

Finally, we used the fast-dm algorithm written by Voss and Voss (2007) to estimate the diffusion model parameters. For each participant we obtained seven parameter estimates by fitting the model to each individual's RTs for both correct and incorrect responses: (1) mean drift rate, i.e. the speed of information accumulation, (2) variability in drift rate across trials, (3) boundary separation, (4) mean starting point, i.e. participants' bias towards words or nonwords, (5) variability in starting point across trials, (6) non-decision component of processing, i.e. time needed for stimulus encoding and response execution, (7) across-trial variability in time needed for non-decision component. These parameter estimates were then entered into regression analyses in R, with vocabulary score as a predictor.

Results

Vocabulary tests

The responses in the vocabulary tests were scored and analysed as described above. The mean vocabulary test scores per test are shown in Table 3.2.

Table 3.2: The distribution of test scores in all seven vocabulary tests. The maximum possible scores for each of the tests are provided in brackets in the column displaying participants' maximum scores.

Test	N	Minimum	Maximum	Mean	SD
Andringa	174	15.0	50.0 (59.0)	33.4	6.0
PPVT	174	55.0	115.0 (*)	87.9	10.0
Definition test	174	6.0	20.0 (20.0)	14.7	2.1
Antonym MC	174	8.0	25.0 (25.0)	20.8	2.6
Antonym open	174	7.0	21.5 (25.0)	14.9	2.4
Synonym MC	174	3.0	21.0 (25.0)	12.8	3.1

* Maximum possible score varied between 136 and 139.

Bivariate correlation coefficients between the vocabulary tests are displayed in Table 3.3. All measures are significantly moderately correlated with one another. All tests are correlated and, thus, appear to capture a shared underlying variable. The reliability measure Cronbach's α indicated that the test battery as a whole is highly reliable (α = .80). Dropping one of the tests would lead to a lower α, hence lower reliability of the vocabulary test battery.

A PCA assuming two components was run on z-transformed vocabulary scores. Only the first component had an eigenvalue greater than 1 and it explained 50.39% of the total variance. Factor 1 loaded on all vocabulary tests (see Table 3.4). Again, no clear picture of a distinction between productive and receptive vocabulary tests was obtained.

Table 3.4: Results of the PCA assuming two components.

Vocabulary measure	**Component 1**	**2**
Andringa	.80	-.17
PPVT	.65	.17
Definition	.63	-.64
Antonym MC	.77	.05
Antonym open	.64	.59
Synonym MC	.75	.01
Eigenvalue	3.02	.81
% Variance	50.39	13.56

Lexical decision task

Accuracy rates were lower than usually observed in lexical decision tasks and as in Chapter 2 of this dissertation, with 11.6% of all trials being false alarms and 2.3% misses. RTs were trimmed by excluding all responses that exceeded each participant's mean by 3 SD or were lower than 250 ms. Following these criteria, 3% of data were excluded as outliers. Accuracy was higher for words than for nonwords (z = -20.61; $p < .001$), and participants made fewer errors with increasing word frequency (z = 10.98; $p < .001$). RTs for words were slower than for nonwords (t = 19.18; $p < .001$; see Appendix A for a table showing averaged lexical decision RTs for all conditions and a plot of the RT distribution). Finally, RTs for correct responses to words increased with decreasing word frequency (t = -14.29; $p < .001$).

Individual differences

The relationship between individual differences in vocabulary size and word recognition performance was the focus of this study. Both lexical decision accuracy and speed were

Table 3.3: Bivariate correlations between the vocabulary scores.

	Definition	**Andringa**	**Antonym MC**	**Antonym open**	**Synonym MC**
Andringa	.46**				
Antonym MC	.39**	.57**			
Antonym open	.20**	.38**	.44**		
Synonym MC	.36**	.56**	.43**	.40**	
PPVT	.32**	.35**	.42**	.34**	.40**

** $p < .01$.

analysed using mixed-effects models. For this individual differences investigation, we focused on responses to word trials.

Accuracy

Response accuracy in the lexical decision task was predicted by word frequency (z = 8.41, p < .001) and composite vocabulary score (z = 4.45; p < .001), whereas the interaction between these main effects was not significant (z = -.55; p = .58). Similar results were obtained from the mixed models including participants' scores from the individual vocabulary tests as a predictor (see Table 3.5). Only the model using the PPVT scores as one of the predictors, along with word frequency, did not show a significant main effect of vocabulary score (z = 1.32; p = .19), but only of word frequency (z = 8.0; p < .001). The interaction was not significant (z = -.97; p =.33), just as in the model including the composite vocabulary score. Furthermore, the model including the multiple-choice synonym test scores as a predictor of lexical decision accuracy on word trials did not converge, even after excluding random slopes and the random intercept for item. All other vocabulary tests or models showed the same patterns of relationship between vocabulary score, word frequency, and lexical decision accuracy.

Table 3.5: *P*- and *t*-values for the main effects in each of the models where scores of individual vocabulary measures were used as predictors of lexical decision accuracy.[3]

Model	**Variable**	***t***	***p***
Definition test	Vocabulary score	2.35	.01
	Word frequency	8.13	<.001
	Frequency x vocabulary	-1.49	.14
Andringa	Vocabulary score	3.81	< .001
	Word frequency	7.42	<.001
	Frequency x vocabulary	-1.64	.10
Multiple-choice antonym	Vocabulary score	4.39	<.001
	Word frequency	8.79	<.001
	Frequency x vocabulary	.79	.03
Open antonym	Vocabulary score	2.23	.02
	Word frequency	7.77	<.001
	Frequency x vocabulary	-.09	.93
PPVT	Vocabulary score	1.32	.19
	Word frequency	8.0	<.001
	Frequency x vocabulary	-.97	.33

Reaction times

There was a significant main effect of word frequency (t = -14.99; p < .001) but not of the composite measure of vocabulary (t = .90; p = .37) on log-transformed RTs. Importantly, the interaction between word frequency and composite vocabulary score was significant (t = 4.10; p <.001), with a stronger word frequency effect for individuals with poorer word knowledge (see Appendix B). Most of the models including each of the individual vocabulary scores as predictors of lexical decision RTs yielded similar results (see Table 3.6). Only the open antonym and the multiple-choice synonym tests showed slightly different patterns. In the model with open antonym score and word frequency as predictors, the latter was highly significant (t = -12.22; p < .001); however, neither vocabulary (t = -1.40; p = .16) nor the interaction between word frequency and

[3]The model using the multiple-choice synonym test as a predictor of lexical decision accuracy did not converge, even when leaving out all random slopes and the random intercept for trial. Therefore, it is not reported on in the table.

vocabulary (t = 1.33; p = .18) were significant predictors of RT. The model with participants' multiple-choice synonym test scores as an independent variable showed significant main effects of word frequency (t = -12.69; $p < .001$) and vocabulary score (t = 2.48; p = .01), as well as a significant interaction between the two (t = 2.58; p = .01). Hence, individuals with higher synonym multiple-choice test scores were slower in making lexical decisions while the frequency effect was smaller for individuals, who scored highly in this test.

Table 3.6: *P*- and *t*-values for the main effects in each of the models where scores of individual vocabulary measures were used as predictors of lexical decision RTs.

Model	**Variable**	***t***	***p***
Definition test	Vocabulary score	.58	.56
	Word frequency	-13.60	<.001
	Frequency x vocabulary	3.56	<.001
Andringa	Vocabulary score	.93	.35
	Word frequency	-13.32	<.001
	Frequency x vocabulary	4.34	<.001
Multiple-choice antonym	Vocabulary score	-.26	.79
	Word frequency	-13.77	<.001
	Frequency x vocabulary	1.93	.05
Open antonym	Vocabulary score	-1.40	.16
	Word frequency	-12.22	<.001
	Frequency x vocabulary	1.33	.18
Multiple-choice synonym	Vocabulary score	2.48	.01
	Word frequency	-12.69	<.001
	Frequency x vocabulary	2.58	.01
PPVT	Vocabulary score	1.36	.17
	Word frequency	-12.49	<.001
	Frequency x vocabulary	3.42	<.001

Diffusion model analysis

The mean values and SDs of the seven parameter estimates as well as the t-values for the effect of vocabulary score are reported in Table 3.7. The drift rate for words, i.e. the speed with which information accumulates, was higher for individuals with higher vocabulary scores (β = .22, SE = .06, t = 3.50, $p < .001$). Furthermore, vocabulary score predicted the starting point (β = -.02, *SE* = .01, *t* = -2.50, *p* = .02), with increased vocabulary knowledge being associated with a starting point closer to zero, thus, a smaller word bias. The remaining parameters did not show effects of vocabulary size.

Table 3.7: Parameter estimates of the diffusion model.

Parameter	***Mean***	***SD***
Drift rate for words	2.27	0.83
Variability in drift rate	0.78	0.27
Boundary separation	1.96	0.45
Starting point	0.58	0.11
Variability in starting point	0.22	0.17
Non-decision components	0.42	0.09
Variability in non-decision components	0.16	0.12

Discussion

We used a battery of six vocabulary tests and a lexical decision task to investigate the relationship between individual differences in vocabulary and lexical processing performance. An important aspect of this study, distinguishing it crucially from earlier research, is the participant sample that was tested. Instead of undergraduate university students, the usual participants in psychological research, we recruited a large group of vocational college students as participants. This was done for the following reasons. The typical participants, i.e. university students, are assumed to form a rather homogeneous group presumably exhibiting less variation than can be found in the general population. With regards to language, this is due to the fact that they are presumed to get a similar and very specific amount and type of exposure to language (e.g. lectures, scientific reading). In addition, university students are trained to communicate thoughts and ideas in a certain way using a nuanced vocabulary. Overall, this rather specific group of young adults is assumed to not be representative of the general population with regards to their cognitive abilities; in this particular context, their vocabulary and linguistic abilities. Therefore, it was deemed necessary to extend the research on individual differences in vocabulary knowledge and their role in language processing to participants from a broader range of educational backgrounds. It has been argued before that examining university students only might, for instance, underestimate the strength of the relationship between vocabulary knowledge and word recognition performance (Yap et al., 2012).

Participants' scores on all vocabulary tests showed a high degree of variability. This was expected based on the fact that individuals from all levels of vocational education were tested, presumably covering a relatively large range of abilities. Scores from all vocabulary measures correlated similarly strongly with each other, but overall just moderately and not as strongly as has been shown in a group of university students tested previously with the same vocabulary measures (see Chapter 2). In other words, in the group of vocational college students the vocabulary tests captured less shared variance than in the group of university students. In line with this, the PCA showed a smaller percentage of variance explained by the first component as well as weaker loadings of all vocabulary measures on the first component. The reason might be that the vocational college students experienced some of the tests, such as the PPVT or the open antonym test, as much more difficult than others, leading to weaker correlations between participants' scores in the different tests. This illustrates the difficulty one faces when applying measures developed for testing a certain group (here university students) to a different population (here young adults from a different educational background). Finally, again no distinction between open and multiple-choice tests was found in the PCA. Cronbach's α ($\alpha = 0.80$) indicated that the test battery as a whole was reliable and similarly reliable to that observed in the experiment presented in Chapter 2 of this dissertation ($\alpha = .88$).

It has to be noted that all participants responded correctly to almost all of the easy items that had been added to the multiple-choice antonym and synonym and the open antonym tests to increase the number of targets which participants could presumably respond to. Thus, the fact that they gave a high proportion of correct responses to these items, shows that they made an effort to complete the tests and respond to as many items as possible.

Notably, just as in the experiment in Chapter 2, the mean vocabulary scores indicated that the multiple-choice antonym test was easier than the multiple-choice synonym and open antonym tests. We acknowledge this difference in difficulty, which was observed in both experiments. As noted, in future research one might want to adjust the frequency ranges so that they are more similar across all tests.

Finally, the fact that this group displays a considerably larger range of performance than has been observed in university students (see Chapter 2) supports the idea that it is necessary to extend psycholinguistic investigations to participants from more varied educational backgrounds; only in this way it is possible to cover a wider range of variability and potentially different relationships between various cognitive abilities. When comparing the findings from the experiments in Chapters 2 and 3, we note that the samples differed in size. Importantly, the more variable group in Chapter 3 is also the larger group. This supports the idea that the greater variability observed in this group (despite the larger sample size) is in fact characteristic of the population.

Analyses of the lexical decision task data showed the lexicality and word frequency effects that are typically found: Error rates and RTs were lower for words as opposed to nonwords and for high- as compared to low-frequency words (e.g., Keuleers et al., 2015; Kuperman & Van Dyke, 2013; Yap et al., 2012).

Our main interest was the relationship between individual differences in vocabulary size and word recognition performance. We found that both word frequency and vocabulary score had significant effects on response accuracy in the lexical decision task but did not interact. Hence, in line with Yap et al.'s (2012) findings, participants' accuracy rates increased with growing vocabulary. The effect was independent of word frequency. This was the case for the composite measure of vocabulary and almost all of the individual vocabulary test scores; only the model including the PPVT scores did not show a significant main effect of vocabulary. Thus, overall the vocabulary tests all exhibited a similar pattern of relationship with lexical decision accuracy.

Furthermore, there was overall no significant main effect of vocabulary on RTs in the lexical decision task; there was only a consistently significant main effect of word frequency with faster RTs with increasing word frequency, as was expected. However, participants' scores on the open antonym test did not interact with the frequency effect. In addition, the only test which we found to predict lexical decision speed was the synonym multiple-choice test and here the effect was in the opposite direction of what had been previously observed. Individuals with higher synonym multiple-choice test scores were slower in making lexical decisions. The lack of a vocabulary effect or an effect in the opposite direction is surprising because based on previous research, it was assumed that an advantage in lexical decision RTs for individuals with greater vocabulary knowledge would be a rather robust finding (Brysbaert et al., 2016a; Yap et al., 2012). The conclusions about the relationship between vocabulary and language processing might, hence, be slightly different depending on which measure of vocabulary size is used. Thus, for testing broader samples using several vocabulary tests and combining their scores is advisable. Some of the aforementioned differences between the various vocabulary tests might of course be random fluctuations and more research, especially in such varied groups of participants, is needed.

It seems that the participants in this group were overall rather slow in making lexical decisions. They had RTs to words around 170ms longer, RTs to nonwords around 300ms longer, and SDs twice as large as the participants in the experiment presented in Chapter 2. Maybe these longer RTs reflect strategic behaviour on the part of the participants, independent of their vocabulary size; if anything, high-vocabulary individuals show a stronger tendency towards longer RTs and potentially strategic behaviour. It might be possible that the participants in this experiment spent a rather long time thinking about the target words, delaying their button press until they were sure about what the correct response was. A reason for applying such a strategy may be that these participants are less confident about their knowledge of words and trust their intuition less than university students do. Hence, to prevent too many errors they may have re-read the targets several times or thought about their responses for a longer time. This might be the case especially for participants who were determined to do well on all tests, which may explain the relationship between higher vocabulary scores and slower RTs.

Another possibility would be that many of the participants in this experiment were overall rather poor readers. Thus, it might be that reading the strings of letters simply took them a longer time. Both explanations would be in line with the observation that participants' RTs in the lexical decision task are overall considerably slower than what would be expected based on previous word recognition studies. It is unlikely that the reason for this lies in the materials as all words had very high prevalence values, indicating that at least 98% of Dutch speakers in the Netherlands know all of them. Thus, each participant could be assumed to at least know the majority of words. The assumption that the participants might have been poor readers could be examined by replacing the visual lexical decision task by an auditory version. This would show whether the vocational college students are slower processors in general, thus, also slow in an auditory version of the task, or just poorer readers than university students.

The facts that testing a group of participants other than the usual undergraduate university students required changing the lexical decision task (displaying stimuli for 5 seconds instead of only 3 seconds) and yielded considerably different lexical decision RTs demonstrate the necessity of extending psycholinguistic research to participant groups from more varied backgrounds. The data collected from the group of university students is apparently not representative of the general population. Hence, grounding ideas and theories on language processing and cognitive abilities in research on university students only does not do justice to the variability in cognitive abilities that is present in the population.

Although there was only a significant main effect of word frequency but not of vocabulary, the interaction between word frequency and vocabulary was significant. RTs were generally faster for high-frequency words while this word frequency effect was stronger for individuals with poorer vocabulary knowledge. This is in line with previous studies using the lexical decision task, which found an interaction between word frequency and participants' vocabulary skills (e.g., Diependaele et al., 2013).

Additionally, our finding fits with the lexical entrenchment hypothesis (Brysbaert et al., 2016a; Diependaele et al., 2013) stating that the lexical representations in low-vocabulary individuals are not as robust or strong as the representations in high-vocabulary individuals (see also, Perfetti & Hart, 2001, 2002; Yap et al., 2009). As argued by Diependaele and colleagues (2013), this reduced representational strength in low-vocabulary individuals is mainly due to less exposure. Limited exposure does not affect the entire lexicon to the same extent. It is argued that a decrease in exposure has particularly strong effects on low-frequency words. Thus, the frequency curve is supposed to be steeper in individuals with smaller vocabularies than in those with larger vocabularies (Brysbaert et al., 2016a). As a result, word frequency has a smaller effect on language processing speed in high-vocabulary participants than it has in low-vocabulary participants. This hypothesis was confirmed because we found a frequency x skill interaction, as has been reported previously (e.g., Diependaele et al., 2013; Kuperman & Van Dyke, 2013). Participants with weaker vocabulary scores showed stronger effects of word frequency on lexical decision RTs than those with higher vocabulary scores.

In addition, the results support the assumption made in Chapter 2 that the lack of a frequency x skill interaction might be due to characteristics of the participant group tested there, namely a rather homogeneous group of university students. This is due to the fact that the same materials that failed to elicit the interaction between word frequency and vocabulary knowledge in that group, showed the interaction in the present experiment where we tested a more diverse group of young adults.

Moreover, the observation of a frequency x skill interaction is compatible with the explanations for the lack of a main effect of vocabulary. Even if participants apply the strategy of delaying their responses or re-read the target words several times, these processes might be faster for high- than for low-frequency words. Additionally, individuals with larger vocabularies might show a smaller effect of word frequency due to having more exposure to words from the entire range of frequencies, including low-frequency items, than individuals with weaker vocabulary knowledge.

Additionally, we ran diffusion model analyses and entered the resulting parameters in regression models to further examine the origin of the vocabulary effect on lexical decision times. As in Chapter 2 of this dissertation, we did not observe exactly the same patterns but overlapping results with Brysbaert et al. (2016a) and Yap et al. (2012). Two parameters of the diffusion model were predicted by vocabulary score, namely drift rate for words and starting point. As has been found previously, individuals with higher vocabulary scores showed faster build-up of lexical information for words than individuals who scored lower on the vocabulary tests (Yap et al., 2012, see also Chapter 2). Hence, the high-vocabulary advantage has at least to a certain extent its origin in the fact that the information build-up is faster; perhaps because individuals with stronger word knowledge are faster in accessing their lexicon as compared to individuals with weaker vocabulary knowledge. In addition, and just as in Chapter 2, higher vocabulary scores were associated with a smaller word bias. Hence, individuals with better vocabulary knowledge showed a starting point closer to neutral between the word vs. nonword boundaries. This is probably an efficient starting point given that the task included 50% words and 50% nonwords, and high-vocabulary individuals are apparently more likely to show a starting point that is neither strongly word nor strongly nonword biased. It has to be noted that in principle these observations and explanations are compatible with the observation that the participants in this experiment were considerably slower in making lexical decisions than the participants in the experiment presented in Chapter 2. However, if it is assumed that their slowness is due to strategically postponing their responses, it is questionable whether the diffusion model analysis provides valuable insights. The reason is that in this case the RTs would not really reflect participants' stimulus encoding, decision making, and button press speed but potentially also some other additional components. Testing these participants' general processing speed or using an auditory instead of a visual version of the lexical decision task might provide some insights into whether they were generally slow or whether their speed was affected by the specific task demands and corresponding strategic behaviour.

Conclusions

In this study, we demonstrated that greater vocabulary knowledge as measured by a battery of vocabulary tests was associated with more accurate but not faster word processing in a lexical decision task. Despite the absence of a main effect of word frequency on lexical decision RTs, we observed an interaction between frequency and vocabulary. In line with earlier research, the word frequency effect on lexical decision RTs decreased with increasing vocabulary knowledge. This was found in an experiment where, unlike in earlier research, non-university students were tested. Besides the theoretical insights on the relationship between individual differences in vocabulary and language processing, the findings demonstrate the importance of extending psycholinguistic research to include participants from a broader range of educational backgrounds. More research is needed to examine potential consequences of individual difference in vocabulary size, such as structural or representational differences between vocabularies of varying sizes, in a varied group of speakers.

Acknowledgements

We would like to thank the vocational colleges ROC Nijmegen, ROC Midden-Nederland, and ROC Tilburg for their cooperation and the possibility to conduct our experiment there. We would also like to thank Annelies van Wijngaarden for help in contacting the schools and scheduling the test sessions, and Carlien Alferink, Nikki van Gasteren, Hanna de Groot, Anne van Hoek, Moniek van de Kraats, and Ilse Wagemakers for their support in testing the participants.

Appendix A

Table 3.8: Reaction times (ms) for correct responses in the lexical decision task per condition. For illustration purposes, frequency is shown as a categorical variable although all models were run with frequency as a continuous variable.

Lexicality	**Frequency**	***Mean***	***SD***
Words	low	855	414
	medium	781	416
	high	743	388
	total	793	409
Nonwords	-	1035	476

Note: Low-frequency words had frequency values of less than 1 count per million in the SUBTLEX corpus (M = 0.36; SD = 0.27), medium-frequency words between 1 and 10 counts per million (M = 3.47; SD = 2.34), and high-frequency words between 10 and 90 counts per million (M = 25.24; SD = 20.92).

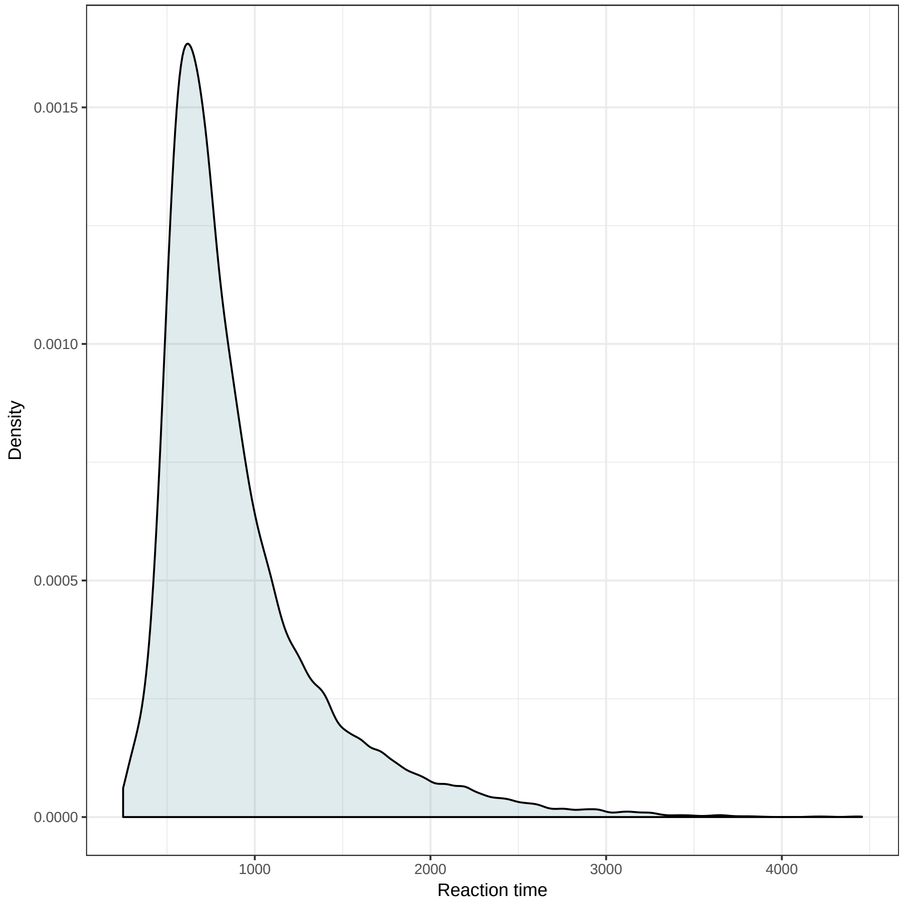

Figure 3.1: Distribution of the lexical decision RTs.

Appendix B

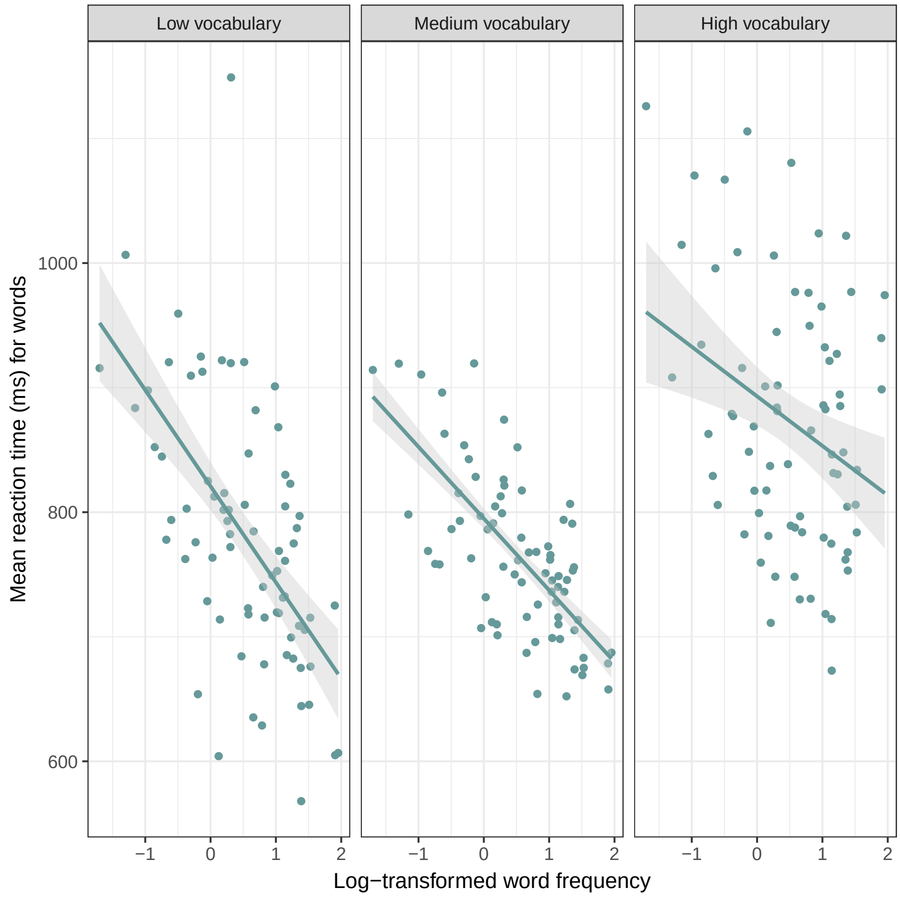

Figure 3.2: RT as a function of log-transformed word frequency for low- vs. medium- vs. high-vocabulary individuals. 95% confidence intervals are displayed in grey. Note that for illustration purposes, vocabulary score was transformed into a categorical variable with three levels. The above-described model was run with vocabulary score as a continuous variable.

4 | Vocabulary knowledge affects lexical production: Evidence from picture-word interference

Abstract

In this study we examined the relationship between vocabulary knowledge and language production in a picture-word interference task. Word knowledge was measured using the same battery of seven vocabulary tests, which was administered in the studies presented in Chapters 2 and 3. Hence, we used two established measures of word knowledge, namely the Peabody Picture Vocabulary Test (Schlichting, 2005) and Andringa et al.'s (2012) receptive multiple-choice test, and five newly developed ones. Language production was assessed in a picture-word interference (PWI) task. We found significant effects of semantic interference and distractor frequency. Slower RTs were obtained for semantically related than for unrelated distractors, and high-frequency distractor words induced less interference than low-frequency distractors. In addition, ex-Gaussian analyses showed a selective effect of semantic relatedness on the τ parameter and of distractor frequency on the μ parameter of the RT distribution. Importantly, RT in the PWI task were predicted by vocabulary size: Individuals with better word knowledge responded faster than those with smaller vocabularies.

Introduction

An average 20-year-old student has recently been estimated to know 42,000 lemmas and 4,200 multiword expressions, derived from 11,100 word families (Brysbaert et al., 2016b). While these numbers are impressive, the study also showed that there is considerable individual variation as a function of educational level. Likely underlying or related factors resulting in individual differences in vocabulary knowledge are, for instance, skills, interests, and life-experiences, including reading exposure (see Chapters 2 and 3; Brysbaert et al., 2016b). The knowledge of words is certainly an important aspect of a speaker's command of their language and it has been shown to affect language processing performance in various linguistic tasks (Banks et al., 2015; Rodriguez-Aranda & Jakobsen, 2011; Yap et al., 2009). Especially performance on comprehension tasks, such as lexical decision, spoken word recognition or speeded pronunciation, has been reported to vary as a function of individual differences in vocabulary size (e.g., Brysbaert et al., 2016a; Chateau & Jared, 2000; Yap et al., 2009). However, far less research has focused on language production. The present study therefore examined the relationship between individual differences in vocabulary size and spoken language production. We employed a battery of seven vocabulary tests and a picture-word interference (PWI) task.

As indicated before, previous studies of the role of vocabulary knowledge in language processing have often looked at word recognition and found beneficial effects of increased word knowledge on processing performance. Older adults' accuracy of spoken word recognition was found to be predicted by vocabulary knowledge with higher accuracy rates with increasing vocabulary size (Janse & Jesse, 2014). Better vocabulary knowledge was also found to be associated with improved speech recognition in suboptimal conditions (Bent et al., 2016). Furthermore, Yap and colleagues (2009) reported higher vocabulary scores to be linked to faster RTs and higher accuracy rates in both lexical decision and speeded pronunciation. In addition, smaller word frequency effects in the speeded pronunciation task were found for individuals with higher vocabulary scores. Similarly, Brysbaert, Lagrou, and Stevens (2016a) showed that higher vocabulary scores were associated with higher accuracy rates and faster RTs in a lexical decision task as well as reduced effects of word frequency on RTs. Hence, not only faster RTs were associated with higher vocabulary scores, but individuals with larger vocabularies (Diependaele et al., 2013; Yap et al., 2009) or increased reading experience (Chateau & Jared, 2000; Kuperman & Van Dyke, 2013) have also been observed to show smaller word frequency effects. The typical word frequency effect, namely faster RTs with increasing frequency, has been reported to be smaller for individuals with better vocabulary knowledge.

Representational differences between vocabularies of varying sizes have been suggested to be the origin of this frequency x skill interaction. High-vocabulary individuals are assumed to have more *robust*, *defined*, or *entrenched* representations as compared to low-vocabulary individuals leading to faster lexical access, which is less sensitive to effects of word frequency (Van Dyke et al., 2014; Perfetti & Hart, 2001; Yap et al., 2009). Diependaele and colleagues (2013) formulated the *lexical entrenchment hypothesis* to account for this pattern. Accordingly, the frequency x skill interaction is a result of differences in language exposure. Especially low-frequency words are assumed to be affected by variation in exposure, which leads to a stronger frequency difference between low- and high-frequency words in individuals with limited exposure to language, hence, smaller vocabularies (Kuperman & Van Dyke, 2013). Thus, the frequency curve in these individuals' vocabularies is steeper than in individuals with larger word knowledge (Brysbaert et al., 2016a). The lexical representations are assumed to be weaker or less robust in low- than in high-vocabulary individuals, in particular for low-frequency words, therefore slower to be processed.

Additional insights into the relationship between vocabulary knowledge and language production performance were gained from studies using verbal fluency tasks, where participants are required to produce as many words as possible starting with a given letter (letter fluency) or from a given semantic category (category fluency) within one minute. Unsworth and colleagues (2011) examined which component processes are involved in verbal fluency task performance. They found that individuals with greater vocabulary knowledge generated a higher number of items in total. Additionally, Shao et al. (2014) investigated the contributions of verbal ability and executive control to verbal fluency performance in older adults. In this study, participants with higher vocabulary scores were faster at giving the first response in different verbal fluency tasks than individuals with smaller vocabulary knowledge. Finally, Rodriguez-Aranda and Jakobsen (2011) found that vocabulary predicted RTs in letter as well as category fluency and a short picture naming task.

Hence, the insights on the relationship between individual differences in vocabulary and language processing are mainly based on studies using word recognition or other comprehension tasks (Brysbaert et al., 2016a; Diependaele et al., 2013; Yap et al., 2009, 2012), whereas there are considerably fewer studies focusing on language production. In addition, the latter mainly used verbal fluency tasks, which do not solely measure language processing performance but also executive control functions (Rodriguez-Aranda

& Jakobsen, 2011; Shao, Roelofs, et al., 2014; Unsworth et al., 2011), or a rather short picture naming task with only ten pictures (Rodriguez-Aranda & Jakobsen, 2011).

The present study filled this gap by relating language production performance to individual variation in vocabulary. A PWI task was administered to test for speed of lexical selection in spoken language production. In the classical PWI paradigm participants are instructed to name pictures in the presence of semantically related or unrelated distractor words. Additionally, we manipulated word frequency by pairing each stimulus with both low- and high-frequency distractor words. This was done to take into account the possibility that the effects of word frequency and vocabulary knowledge might interact with each other, just as was the case in the above-mentioned lexical decision experiments (e.g. Brysbaert et al., 2016a; Diependaele et al., 2013). It has to be noted that in addition to lexical selection for production, the PWI task involves word recognition processes (of the written distractor word), which the frequency manipulation in the present study is related to.

In general, mean naming RTs are longer for stimuli with semantically related than with unrelated distractor words (e.g., Damian & Martin, 1999; Miozzo & Caramazza, 2003; Schriefers, Meyer, & Levelt, 1990). Previous studies using PWI tasks including the distractor frequency manipulation have indicated that more interference was induced by low- as opposed to high-frequency distractor words (Dhooge & Hartsuiker, 2010, 2011; Hutson, Damian, & Spalek, 2013; Miozzo & Caramazza, 2003; Roelofs, Piai, & Schriefers, 2011; Scaltritti, Navarrete, & Peressotti, 2015).

The origins of the semantic interference and the distractor frequency effect are controversial. One explanation is based on the assumption of spreading activation from the target concept to related concepts. Hence, semantically related distractors receive bottom-up activation from the written word, and in addition top-down activation from the conceptually related target picture. This leads to stronger interference or competition for selection induced by semantically related as opposed to less strongly activated unrelated distractors. The effect is, hence, argued to be located early in the speech planning process (Roelofs, 1992, 2003). Within this framework the distractor frequency effect is hypothesised to be due to a reactive blocking of the distractor word, in order to give priority to the processing of the picture name. High-frequency words can be processed faster than low-frequency words, which leads to high-frequency distractors being blocked out more quickly as compared to their low-frequency counterparts. Picture naming latencies are, consequently, shorter for high-frequency distractors (Roelofs et al., 2011).

Others have located the semantic interference effect in an articulatory buffering stage, after lexical access has occurred. It is assumed that the distractor word activates an articulatory program, which is entered into an output buffer and has to be removed from it to allow articulation of the response to the target picture. This is hypothesised to take longer for semantically related than for unrelated distractor words (Dhooge & Hartsuiker, 2010, 2011; Scaltritti et al., 2015). Word frequency is presumed to affect the speed with which the distractor accesses the response buffer. High-frequency distractors enter the buffer earlier than low-frequency words and can thus be removed earlier; picture naming times for stimuli with high-frequency distractor words are consequently shorter than for those with low-frequency distractors (Dhooge & Hartsuiker, 2010; Scaltritti et al., 2015).

For the present purpose it is important to note that although there are differences between the two frameworks' explanations of the semantic interference and distractor frequency effects, they agree on important aspects. Both assume attentional control to be underlying the semantic interference effect, whereas the distractor frequency effect is driven by the speed of processing of the distractor word (Scaltritti et al., 2015).

A number of recent studies have investigated RT distributions using ex-Gaussian analyses besides looking at mean RTs (e.g., Jongman et al., 2015; Piai, Roelofs, & Schriefers, 2011, 2012; Shao et al., 2012). In contrast to analyses based on mean RTs, which assume a symmetric distribution around the mean, Ex-Gaussian analyses allow us to account for the typical positive skewness of RT distributions. This is due to the fact that the ex-Gaussian function is a convolution of a Gaussian, i.e. normal, and an exponential function, which fits RT distributions very well (Balota & Yap, 2011; Balota, Yap, Cortese, & Watson, 2008; Heathcote, Popiel, & Mewhort, 1991). An ex-Gaussian analysis provides two parameters describing RT distributions: μ which reflects the mean of the Gaussian part of the underlying RT distribution, and τ which describes the mean and standard deviation of the exponential part. Hence, using ex-Gaussian analyses RTs can be decomposed into two components, which characterise the normally distributed part (μ) and the slower tail (τ) of the underlying RT distribution (Roelofs, 2012).

Piai and colleagues (2011, 2012) offered pioneering insights into the semantic interference effect in the PWI paradigm by analysing not only mean RTs but the distributional features of the effect. Piai et al. (2011) report the semantic interference effect to be reflected solely in the μ parameter, thus, assuming the shape of a distributional shift. By contrast, Piai et al. (2012) found the semantic interference effect for clearly visible distractor words to be reflected on the τ part of the RT distribution. Thus, mixed results have been reported.

Scaltritti and colleagues (2015) went one step further by examining the hypothesised different underlying dynamics of the semantic interference and distractor frequency effects by analysing the distributional features of both effects. They found the semantic interference effect to selectively affect the slower tail (τ), whereas the distractor frequency effect was reflected in the normal part (μ) of the RT distribution. Hence, the latter effect was shown to result in a distributional shift such that RTs for stimuli with low-frequency distractors are overall slower than for those in the high-frequency condition. The semantic interference effect, by contrast, was particularly pronounced in the slow responses when individuals' attention efficiency is assumed to be reduced (Piai et al., 2012; Scaltritti et al., 2015, but see Roelofs & Piai; 2017).

In the present study, we intended to examine whether potential effects of vocabulary are mainly represented in the μ or τ part of the RT distribution. Earlier, it has been studied whether different cognitive abilities relate to different parameters of the RT distribution in various tasks. Shao et al. (2012), for instance, used ex-Gaussian analyses to examine the relationship between different aspects of executive control and participants' RTs in object and action naming. They found that updating ability was correlated with the exponential tail of both the action and object naming RT distributions. Hence, this suggests updating ability to be involved in only the slow trials in both tasks. By contrast, inhibition ability was correlated with the normal part (μ) of the RT distribution for action naming (which was the harder task), and with the slower tail (τ) for object naming. These observations indicate that in action naming, inhibiting ability is engage on most of the trials whereas in object naming inhibitory control is only involved in the very slow trials (Shao et al., 2012).

If increased vocabulary knowledge is associated with overall faster processing, the effect should selectively be reflected in the normal part of the RT distribution of the PWI task. In this case, the vocabulary effect would have the shape of a distributional shift with RTs being overall slower for individuals with smaller vocabulary scores. However, an effect of vocabulary knowledge only being exhibited in the slower tail of the RT distribution would indicate that the effect is mainly mediated by individuals' slow responses. Based on previous research on the RT distribution in PWI tasks, vocabulary effects on the slow tail of the RT distribution in this task would suggest some involvement of attentional capacity (Scaltritti et al., 2015).

Vocabulary knowledge was assessed using a battery of measures. Most earlier studies have used individual measures of vocabulary. However, it has been argued that the assessment of a skill as complex as word knowledge requires battery of vocabulary tests of different types and formats including multiple-choice and open tests (Bowles & Salthouse, 2008). The reason is that vocabulary test performance does not only depend on an individual's word knowledge but also on their world knowledge, attention, and guessing strategies. Especially for studies where vocabulary is the main interest, Bowles and Salthouse (2008) recommended the use of a battery of different vocabulary tests. We therefore employed two established measures of vocabulary size, namely the Peabody

Picture Vocabulary Test (PPVT; Schlichting, 2005) and Andringa et al.'s (2012) receptive multiple-choice test, in addition to five newly developed tests. These five tests were comprised of multiple-choice antonym and synonym tests, open antonym and synonym tests and a definition test, which all covered a large range of word frequencies.

Additionally, the various measures addressed the three dimensions of lexical competence proposed by Henriksen (1999). *Partial to precise knowledge of word meanings* was assessed in the multiple-choice tests as well as in the definition test. The *depth of word knowledge* dimension was addressed using the antonym and synonym tests as for those tests more complex relationships between words and their meanings was required. Finally, the distinction between *productive and receptive vocabulary knowledge* was taken into account by including open antonym and synonym tests in addition to the multiple-choice versions. Hence, this battery including various test types and formats was assumed to measure vocabulary knowledge comprehensively by addressing various aspects of it.

To sum up, we tested three hypotheses: (1) that better vocabulary knowledge is associated with faster naming latencies in a PWI task, (2) that individuals with larger vocabularies show weaker effects of distractor frequency than those with weaker vocabulary knowledge, and (3) that the semantic interference effect is smaller for participants with higher vocabulary scores. The reasoning behind the latter is that larger vocabularies have been associated with increased entrenchment or specificity of lexical representations, as indicated by smaller frequency effects in high- as compared to low-vocabulary individuals (Diependaele et al., 2013). As a result of these hypothesised differences in robustness or specificity of representations, the speed with which competition for selection between strongly activated lexical items can be resolved might vary between vocabularies of varying sizes.

For the purpose of analysing the relationship between RTs in the PWI task and vocabulary, we did not only analyse mean RTs but also conducted ex-Gaussian analyses. It was hypothesised that the distractor frequency effect would be reflected in the normal part of the RT distribution whereas the semantic interference effect was expected to affect the slower tail. We also examined whether vocabulary affected both parts of the distribution or one of them selectively.

Method

Participants

The PWI experiment was completed by the same 75 individuals who also participated in the lexical decision experiment presented in Chapter 2 of this dissertation. Data of eight participants had to be discarded prior to further data processing due to technical failure or failure to perform one of the tasks correctly.

Thus, data of 67 participants was left (51 females, 16 males) aged between 18 and 32 (M = 21.81 years; SD = 3.41 years). Most of them were students at the Radboud University Nijmegen or the Hogeschool van Arnhem en Nijmegen (N = 62), four participants were working, and one was unemployed. Most participants' highest level of education was a bachelor's degree (N = 25) or the gymnasium (N = 20), which is the highest variant of secondary education in the Netherlands. Others had completed a master's degree (N = 5), or other higher education certificates (N = 17).

All participants were recruited using the participant database of the Max Planck Institute for Psycholinguistics and gave informed consent to participate in this experiment. They were paid 12 Euros for their participation. Ethical approval was granted by the Faculty of Social Sciences of the Radboud University Nijmegen.

Materials and design

All participants completed seven vocabulary tests, five of which were newly developed and two were established measures of word knowledge, namely the Dutch version of the Peabody Picture Vocabulary Test (PPVT-III NL; Schlichting, 2005) and Andringa et al.'s (2012) receptive multiple-choice test. The test materials are described in Chapter 2 of this dissertation. Language production was assessed in a PWI task.

Picture-word interference task

The materials consisted of thirty line drawings with medium to high frequency names in the SUBTLEX-NL corpus (min = 1.72; max = 435.7; M = 65.94, SD = 88.88) were selected (Keuleers et al., 2010). The prevalence values of the stimuli (Keuleers et al., 2015) ranged from 2.27 to 3.38 (M = 2.89, SD = 0.33). According to Keuleers and colleagues, the values for the picture names included in the present picture-word interference experiment indicate that they are known by at least 98% of the native speakers of Dutch living in the Netherlands.

Two semantically related distractor words were assigned to each picture, one low-frequency (LF) and one high-frequency (HF) item, resulting in 60 distractor words in total. In addition, each picture was paired with two semantically unrelated distractors from the same list of 60 words, again a LF and a HF word. Hence, each picture appeared in four different distractor conditions: LF related, HF related, LF unrelated, HF unrelated. The picture *deur* (door) was, for instance, paired with the distractor words *raam* (window; HF related), *erker* (oriel; LF related), *huid* (skin; HF unrelated), and *penseel* (paintbrush; LF unrelated) (see Appendix A for all materials). Table 4.1 shows the frequency and prevalence values for LF and HF distractor words. The LF distractor words had significantly fewer counts per million in the SUBTLEX-NL corpus than the HF distractors ($t(29) = 3.78$, $p < .001$).

Word length of LF ($M = 6.47$, $SD = 1.48$) as compared to HF distractors ($M = 5.2$, $SD = 1.97$) was controlled for. LF words were on average 1.3 letters longer than the respective HF distractors of the same target picture ($SD = 2.1$). Word length and SUBTLEX frequency ($r = .45$, $p < .001$) were significantly correlated.

Table 4.1: Frequency and prevalence information for the low-frequency (LF) as compared to the high-frequency (HF) distractor words.

Distractor condition	**Measurement**	***N***	***Min***	***Max***	***M***	***SD***
LF words	SUBTLEX-NL*	30	0	9.67	1.03	1.97
	Prevalence	30	0	3.33	1.98	0.67
HF words	SUBTLEX-NL*	30	2.17	458	59.36	84.4
	Prevalence	30	2.46	2.97	2.97	0.27

* Frequency per million in the SUBTLEX-NL corpus.

Four pseudo-randomised stimuli lists including all 120 items each were created. Hence, each picture appeared in four experimental conditions in all four lists. The order of stimuli within each of the lists was fixed and pseudorandomised according to the following criteria. Two presentations of the same picture were separated by at least five trials with other pictures and not more than three consecutive trials belonged to the same experimental condition. Target pictures or distractor words from the same semantic category were always separated by items from another category. Finally, consecutive items, both picture names and distractor words, never had the same first phoneme.

To sum up, a total of 120 pictures were named by each participant. These included four instances of each picture, one from each experimental condition (LF related, HF related, LF unrelated, HF unrelated). Each distractor word occurred twice, once as a semantically related and once as an unrelated distractor word.

Procedure

All participants were tested individually in experiment rooms at the Max Planck Institute for Psycholinguistics. All tasks were presented on a 17-inch screen (Iiyama LM704UT) using either the Presentation software (version 16.5, www.neurobs.com) or as an online questionnaire using LimeSurvey (www.limesurvey.org). Headphones (HD 280 Sennheiser) were used to present the auditory stimuli in the PPVT.

The PWI task was completed on the same day as the lexical decision task, which is reported on in Chapter 2 of this dissertation. The PWI task was always followed by the lexical decision task, and the entire vocabulary test battery was completed on a different day, either before or after the experimental session. The vocabulary tests were self-paced and participants were instructed to respond as accurately as possible and without spending too much time thinking about single test items. Prior to the actual tests, participants saw instructions as well as example test items on the screen. The vocabulary test battery took in total about 35 to 45 minutes. As with the materials, a detailed description of the procedure of the vocabulary tests is provided in Chapter 2.

Picture-word interference task

Stimuli in the PWI task were presented using the Presentation software (version 16.5, www.neurobs.com). The experiment consisted of two parts. In the first part, participants were familiarised with the names of the pictures that were used in the experiment. During this familiarisation phase, all target pictures were shown one by one in the center of the screen together with the corresponding names written below them in Arial 47-point font. All drawings fitted into a virtual frame of approximately 12 cm by 12 cm and were shown on a white background. The presentation of the pictures and their names was self-paced. Pressing the enter key initiated the following picture. Participants were instructed to memorise the names assigned to the pictures so that they would be able to correctly name the pictures in the experiment.

After this familiarisation phase participants could take a short break. In the test phase, each trial started with a fixation cross for 500 ms, which appeared in the center of the screen and was replaced by a picture-distractor pair. The distractor was written in Arial 47-point font and was placed in the center, superimposed on the picture. The stimuli stayed on the screen for 3000 ms and were followed by the fixation cross and the next stimulus. Participants were instructed to respond as quickly and accurately as possible. Before the actual test phase started, the participants were shown four practice trials. The four pictures used to familiarise participants with the stimuli presentation were line drawings similar to the test items but not identical with any of them. Two practice pictures were paired with semantically related distractor words, two came with semantically unrelated distractors. The presentation of the four stimuli lists was counterbalanced across participants.

Analyses

Vocabulary tests

The vocabulary tests were scored and analysed in the same way as described in Chapter 2 of this dissertation. The only difference is that we did not repeat the Principal Component Analysis (PCA) and the reliability analysis for this is a subset of the data presented in Chapter 2.

Picture-word interference task

Responses were excluded from the analyses in cases of (a) incorrect naming, (b) verbal disfluencies (such as repairs, stuttering, or coughing), or (c) recording failures. Moreover, responses exceeding each participant's mean RT by three standard deviations were excluded from the analyses.

The picture naming RTs were log-transformed and analysed in R (R Core Team, 2016) employing a linear mixed-effects model approach using the lmer function of the lme4 package (version 1.1.12; Bates et al., 2015). The model on log-transformed RTs as dependent variable included an intercept as well as fixed effects for the factor semantic relatedness (related vs. unrelated) and the continuous variable distractor frequency. In addition, a fixed effect for the interaction between relatedness and distractor frequency was included. Furthermore, by-participant and by-item adjustments to the intercept (random intercepts) and to the relatedness and frequency slopes (random slopes) were modelled. All possible correlations between the random effects were included. Hence, we followed Barr, Levy, Scheepers, and Tily (2013) using a maximal random effects structure. P-values were determined using the normal approximation, i.e. using the t-value as z-value.

Ex-Gaussian analyses to estimate the parameters μ and τ of the RT distribution were run on raw RTs, i.e. untrimmed and not log-transformed RTs, from the PWI task. This was done using the program QMPE (Heathcote, Brown, & Cousineau, 2004). Ex-Gaussian parameters μ and τ were computed for all four conditions and then analysed in linear models using the lm function of the stats package (version 3.3.1; R Core Team, 2016) in R (R Core Team, 2016) with distractor frequency (LF vs. HF) and semantic relatedness (related vs. unrelated) as categorical predictors.

The main interest was the relationship between individual differences in vocabulary knowledge and PWI task performance. For this purpose, the above-described mixed-effects model on log-transformed RTs was run with vocabulary (see below for the definition) as additional predictor. Thus, the model included an intercept and fixed effects for semantic relatedness (related vs. unrelated) and the continuous predictors word frequency and vocabulary score. Fixed effects for the interactions between these effects were modelled. Finally, by-participant and by-item adjustments to the fixed intercept (random intercept) and random adjustments to the relatedness, frequency, and vocabulary score slopes (random slopes) were modelled.

We used a composite measure of vocabulary, which reflects each participant's performance on all seven vocabulary measures. We calculated regression-based factor scores for each participant using the PCA method in SPSS (DiStefano et al., 2009). As PCA analyses in a previous investigation have shown that the seven vocabulary tests can be reduced to only one component (see Chapter 2; Mainz, Shao, Brysbaert, & Meyer, 2017), we decided to calculate the factor scores assuming only one underlying factor. Thus we obtained each participant's score on that one factor based on their seven vocabulary scores, and in this vein the various measurements of vocabulary were collapsed into one number with a mean of zero and a standard deviation (SD) of 1.

Finally, the ex-Gaussian parameters μ and τ were correlated with each participant's composite measure of vocabulary and analysed using linear regression analyses with the composite measure of vocabulary as predictor.

Results

Vocabulary test

Table 4.2 shows the vocabulary test scores in all tests averaged across all participants, and Table 4.3 displays the bivariate correlations between all tests. The data are very similar to those presented in Chapter 2 because the 67 participants whose data are presented here are a subset of the 75 individuals who participated in the lexical decision experiment in Chapter 2. The tests correlated moderately to strongly with one another indicating that they assessed a shared underlying ability. Additionally, again the multiple-choice antonym test appears to be easier than the other measures and correlated least strongly with the other tests (see Chapters 2 and 3).

Table 4.2: Descriptive statistics of test scores in all seven vocabulary tests.The maximum possible scores for each test are provided in brackets in the column displaying participants' maximum scores.

Test	**N**	**Minimum**	**Maximum**	**Mean**	**SD**
Andringa	67	25.0	54.0 (59.0)	39.93	6.10
PPVT	67	56.0	125.0 (*)	103.34	11.80
Definition test	67	12.0	20.0 (20.0)	16.49	2.0
Antonym MC	67	14.0	25.0 (25.0)	23.06	1.60
Antonym open	67	14.0	24.0 (25.0)	19.37	2.35
Synonym MC	67	11.0	24.0 (25.0)	17.67	2.81
Synonym open	67	5.5	22.0 (25.0)	10.70	2.87

* Maximum possible score varied between 136 and 139.

Picture-word interference task

Following the above-mentioned criteria, 2.5% of the data were categorised as errors or misses and were excluded from further analyses. One stimulus, namely the picture *kwast* (paintbrush), was excluded as noticeably many naming errors were committed on this target picture (error rate = 17.8%). Finally, 2.2% of the remaining data were excluded as outliers. Table 4.4 presents the RTs for all four conditions averaged across 67 participants.[1]

[1]A bivariate correlation analysis of the PWI and lexical decision RTs (see Chapter 2) showed that participants' RTs in both tasks were positively correlated (r = .66, p < .001).

Table 4.3: Bivariate correlations between the vocabulary scores.

	Definition	**Andringa**	**Antonym MC**	**Antonym open**	**Synonym MC**	**Synonym open**
Andringa	.59**					
Antonym MC	.43**	.31**				
Antonym open	.73**	.60**	.35**			
Synonym MC	.57**	.61**	.36**	.57**		
Synonym open	.69**	.61**	.38**	.63**	.53**	
PPVT	.60**	.55**	.31**	.64**	.55**	.55**

** $p < .01$.

Table 4.4: Average reaction times (ms) and error rates and their standard deviations (SD) for pictures with related and unrelated distractor words in both frequency conditions.

Conditions		**Reaction times (ms)**		**Error rate (%)**	
		Mean	***SD***	***Mean***	***SD***
Related	low	804	167	3.5	4.1
	high	774	171	2.5	3.1
	total	789	170	3.2	3.5
Unrelated	low	791	166	2.8	3.1
	high	757	157	2.2	2.8
	total	774	152	2.4	2.6
Semantic interference effect	low	14	34		
	high	17	38		
	total	15	29		
Frequency effect	related	32	44		
	unrelated	35	34		
	total	33	33		

RTs in the PWI task were predicted by both semantic relatedness (t = 2.06, p = .04) and distractor frequency (t = -6.39, p < .001). The interaction between semantic relatedness and frequency was not significant (t = -0.28, p = .78). Thus, semantically related distractors induced interference effects leading to longer RTs as compared to unrelated distractors. RTs for targets with LF distractor words were slower than for targets with HF distractors (see Table 4.4).

Analyses of the ex-Gaussian parameters μ and τ using linear models showed significant effects of distractor frequency on μ (t = -2.84, p = .005), only, and of semantic relatedness on τ (t = 2.37, p = .02), only (see Table 4.5 for the raw values).

Table 4.5: Mean values of the ex-Gaussian parameters μ and τ per condition in the picture-word interference task.

Conditions		**Reaction times (ms)**	
		μ	τ
Related	low	674	151
	high	649	138
	total	661	145
Unrelated	low	673	132
	high	650	115
	total	662	123

Individual differences

Participants' RTs in the PWI task correlated negatively with their vocabulary scores (r = -.40, p = .001; see Figure 4.1). Higher vocabulary scores were related to faster RTs in the PWI task.

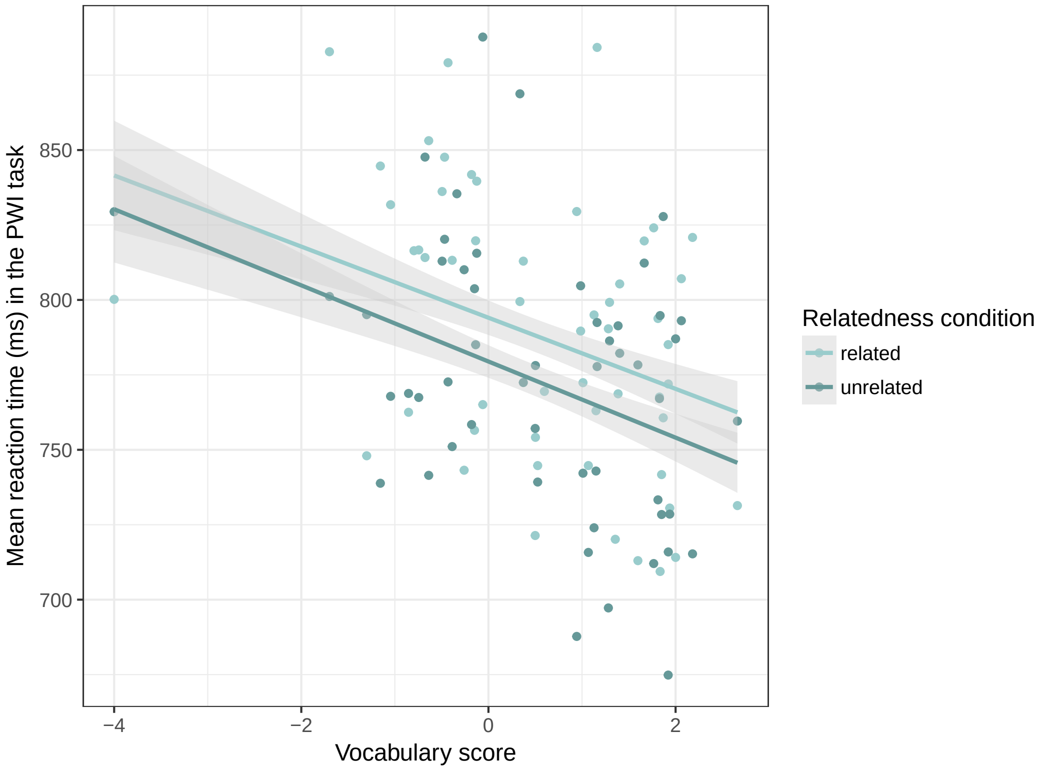

Figure 4.1: Correlation between participants' vocabulary test scores (factor scores) and their mean picture naming reaction times (in ms) for both conditions. Confidence intervals are displayed in shaded grey.

The mixed-effects analyses on the PWI task showed that distractor frequency (t = -6.4, p > .001), semantic relatedness (t = 2.1, p = .004), and vocabulary (factor) score (t = -3.3, p = .001) predicted individuals' RTs. Responses in the semantically related condition were slower than in the unrelated condition, and RTs were slower for stimuli in the low-frequency distractor condition than in the high-frequency distractor condition. In addition, poorer vocabulary scores were associated with slower RTs in the PWI task. These results confirm the findings of the correlational analyses.

The interaction between frequency and vocabulary score was significant (t = -2.2, p = .03), indicating that the distractor frequency effect was stronger for individuals with higher vocabulary scores.

Additionally, a significant three-way interaction between semantic relatedness, frequency, and vocabulary score was found (t = -2.0, p = .05). This interaction is plotted in Appendix B. As can be seen, higher distractor frequency was generally associated with faster responses. For unrelated distractors, the strength of the distractor frequency effect was independent of vocabulary size. For related distractors, the

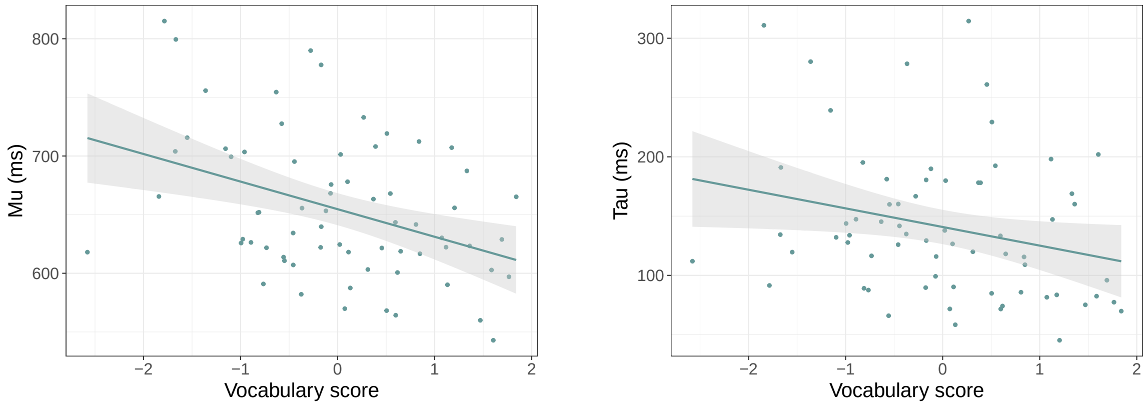

Figure 4.2: The relation between μ and τ of the reaction times in the picture-word interference task and individuals' vocabulary scores.

distractor frequency was weaker for participants with low vocabulary scores than for the high-vocabulary individuals.

Regression analyses on the ex-Gaussian parameters μ and τ showed that the composite measure of vocabulary predicted both μ (t = -3.41, p = .001) and τ (t = -2.15, p = .04). The correlations between vocabulary and μ (r = -.35, p = .004) as well as τ (r = -.29, p = .02) are illustrated in Figure 4.2.

Discussion

In this study, we used a battery of seven vocabulary tests and a PWI task to examine the relationship between individual differences in vocabulary size and word production. Although the group of participants was rather homogeneous, being comprised of university students only, we observed some variation in vocabulary test performance. The fact that all vocabulary tests correlated similarly strongly with one another indicates that they measure largely the same capacity, as was expected based on previous studies using the same vocabulary tests (Mainz et al., 2017, see also Chapters 2 & 3 in this dissertation). In line with these earlier observations, only one test appeared to be considerably easier than the others and thus correlated less strongly with the other measures of word knowledge, namely the multiple-choice antonym test.

In the PWI task, we replicated the classic semantic interference effect, with RTs in the semantically related condition being significantly longer than in the semantically unrelated condition. In addition, high-frequency distractor words induced less interference than low-frequency distractors, as was expected based on previous studies (e.g. Dhooge & Hartsuiker, 2010, 2011; Hutson et al., 2013). The results of the distributional analyses of the RTs in the PWI task are also congruent with previously reported findings (Scaltritti et al., 2015). The interference effect was mainly mediated by the slower tail of the RT distribution, i.e. the ex-Gaussian parameter τ, whereas the distractor frequency effect assumed the shape of a distributional shift, being reflected in the normal part of the RT distribution, i.e. μ. As stated before (e.g. Scaltritti et al., 2015), this indicates that different cognitive dynamics underlie the frequency as compared to the semantic interference effect. The latter effect is hypothesised to be at least partly due to fluctuating attentional efficiency. When the attentional system is highly efficient (fast responses), semantic interference is considerably reduced while it shows its full magnitude in slow responses when attention is considered to be less efficient (Roelofs, 2012; Scaltritti et al., 2015). In line with this, it has also been reported that selective inhibition ability affects the strength of the semantic interference effect (Shao, Roelofs, Martin, & Meyer, 2015).

Distractor frequency, by contrast, is assumed to affect the point in time when the distractor can be discarded and all subsequent operations leading to the articulation of the target word can be performed. The distractor frequency effect is, thus, reflected in an overall shift of the RT distribution (Scaltritti et al., 2015).

The main interest of the present study was the relationship between individual differences in vocabulary size and language processing speed. We showed that increased word knowledge was associated with faster RTs in the PWI task. Importantly, vocabulary was assessed in non-speeded tasks testing only for knowledge, whereas the experimental task was speeded. Hence, the observed relationships are not due to the tasks all testing for processing speed. These findings are in line with previous research indicating an advantage of high vocabulary individuals in language processing as assessed by various tasks (e.g. Andringa et al., 2012; Engelhardt, Nigg, & Ferreira, 2013; Janse & Adank, 2012; Salthouse, 1993). Having more words in one's mental lexicon does not seem to slow down or hinder language processing due to an increase in competition between a larger number of lexical items, as one might expect. Instead individuals with increased word knowledge are faster in accessing their vocabulary, not only in word recognition but also in production tasks, such as the PWI task in this study.

In the PWI task, a significant three-way interaction involving vocabulary size was obtained. For unrelated distractors the size of the distractor frequency effect did not depend on vocabulary size; but for related distractors, the distractor frequency effect was weaker for individuals with poorer word knowledge than for those with higher vocabulary scores. How this unexpected pattern arose is unclear and might warrant further investigation. At present we can only note that the picture naming data indicate that the strength of word frequency effects is not substantially moderated by vocabulary size. Hence, we only found weak evidence for a moderating effect of vocabulary size on the strength of the distractor frequency effect in the PWI task.

Previous studies using lexical decision tasks report this interaction, with high vocabulary individuals exhibiting smaller frequency effects than low vocabulary individuals (Chateau & Jared, 2000; Kuperman & Van Dyke, 2013; Yap et al., 2009). This has been taken as evidence for representational differences between vocabularies of varying sizes. Lexical representations in individuals with better word knowledge have, for example, been hypothesised to be more entrenched or robust, i.e. different in quality, from lexical representations in less skilled individuals (e.g. Diependaele et al., 2013; Yap et al., 2009).

The lack of a consistent interaction between word frequency and vocabulary is in line with findings from a lexical decision task completed by the same individuals (see Chapter 2). There, we also did not observe the frequency x skill interaction. We speculated that the group of participants might have been too homogeneous with regards to their knowledge of the words used in the experiment and potentially also with regard to their vocabulary test performance. This view was supported by the findings that the frequency x skill interaction could be elicited when administering the same lexical decision task and vocabulary tests to a group of participants from more diverse educational backgrounds (see Chapter 3).

Hence, further research, involving groups of participants with a wider range of linguistic abilities, is needed to determine how speakers with large or smaller vocabularies process distractors differing in frequency. In addition, future research into the structural characteristics of vocabularies of varying sizes might draw upon other techniques, in particular computational modelling.

In addition to analysing mean RTs, we conducted ex-Gaussian analyses and examined the relationship between the parameters μ and τ and individual differences on vocabulary knowledge. Participants' vocabulary scores correlated negatively with the parameters μ and τ of the RT distribution. Hence, the vocabulary effect on PWI task performance assumed the shape of a distributional shift with increased vocabulary resulting in overall shorter naming RTs, and was also reflected in the slower tail of the RT distribution. This indicates that individuals with poorer word knowledge gave more slow responses than those with higher vocabulary scores. Consequently, the vocabulary effect appears to be present both when attentional efficiency is assumed to be high, i.e. in the normal part of the RT distribution, and when individuals' attentional efficiency is argued to be reduced, i.e. in the slower tail of the RT distribution.

Previous studies on the role of selective inhibition ability in PWI task performance argued that this aspect of executive control is mainly reflected in the slower responses (e.g. Shao et al., 2015). Selective inhibition ability was correlated with the magnitude of the mean semantic interference effect. Thus, the RT distributions of individuals with better selective inhibition had smaller tails (Shao et al., 2015). In addition, Jongman et al. (2014) found that sustained attention ability correlated with the τ parameter of participants' RTs in picture description and picture naming tasks. Poorer sustained attention ability was associated with an increased number of slow responses in the language production tasks.

By contrast, the vocabulary effect is reflected in the entire RT distribution instead of being selectively mediated by the slower tail. Consequently, the effects of vocabulary and general cognitive abilities, such as selective inhibition and sustained attention, on language processing are different in nature.

Altogether, these results of the individual differences analyses suggest that enhanced speed of language processing – irrespective of the specific task to be performed – is associated with higher vocabulary scores. It is thus conceivable that individuals with larger vocabularies are faster processors or conversely, faster processors might be better word learners and therefore capable of accumulating greater vocabulary knowledge. Previous research on child language acquisition indicates that processing speed indeed plays a central role in word learning (Fernald et al., 2006; Marchman & Fernald, 2008; McMurray et al., 2012). Fernald and colleagues (2006), for instance, report that speed and accuracy in word recognition at 25 months predicted the learning rate for vocabulary across the 2nd year of age. Furthermore, Marchman and Fernald (2008) found that speed of word recognition and word knowledge in infancy predicted children's cognitive and linguistic performance at 8 years.

Conclusions

In this study we examined the relationship between vocabulary size and speed of lexical processing in production, measured using a PWI task. To sum up, we found strong evidence for an effect of vocabulary on lexical processing in production. An increase in vocabulary size was associated with overall faster RTs. This is in line with previous findings indicating that having a larger lexicon does not slow down lexical access but leads to faster word recognition (see Chapters 2 & 3) and production. Further investigations into the complex relationship between domain-general cognitive abilities (e.g. processing speed) and domain-specific ones (e.g. word knowledge), and their joint effects on both language learning and processing are necessary, ideally combining behavioural and computational techniques.

Appendix A: Materials used in the PWI task.

Table 4.6

Pictures	**Related**		**Unrelated**	
	HF	**LF**	**HF**	**LF**
1) deur (*door*)	raam (*door*)	erker (*oriel*)	huid (*skin*)	penseel (*paintbrush*)
2) jongen (*boy*)	baby (*baby*)	zuigeling (*baby*)	plant (*plant*)	zenuwarts (*neurologist*)
3) boot (*boat*)	schip (*ship*)	punter (*ship*)	verf (*paint*)	hoen (*hen*)
4) boom (*tree*)	plant (*plant*)	heester (*bush*)	auto (*car*)	schrift (*notebook*)
5) dokter (*doctor*)	tandarts (*dentist*)	zenuwarts (*neurologist*)	mes (*knife*)	skeelers (*skates*)
6) bot (*bone*)	huid (*skin*)	ligament (*ligament*)	ketting (*necklace*)	gilet (*waistcoat*)
7) envelop (*envelope*)	brief (*letter*)	telegram (*telegram*)	broek (*trousers*)	heester (*bush*)
8) fiets (*bike*)	auto (*car*)	skeelers (*skates*)	baby (*baby*)	coltrui (*turtleneck*)
9) gitaar (*guitar*)	piano (*piano*)	hobo (*oboe*)	beker (*mug*)	bokaal (*cup*)

Pictures	**Related**		**Unrelated**	
	HF	**LF**	**HF**	**LF**
10) regen (*rain*)	sneeuw (*snow*)	windhoos (*vortex*)	bus (*bus*)	pont (*ferry*)
11) hek (*hedge*)	poort (*gate*)	raster (*fence*)	tandarts (*dentist*)	knokkel (*knuckle*)
12) hond (*dog*)	paard (*horse*)	fret (*ferret*)	appel (*apple*)	kompres (*compress*)
13) jurk (*dress*)	broek (*trousers*)	gilet (*waistcoat*)	raam (*window*)	zuigeling (*baby*)
14) koe (*cow*)	haas (*rabbit*)	hoen (*hen*)	schip (*ship*)	buidel (*bag*)
15) jas (*jacket*)	schoen (*show*)	coltrui (*turtleneck*)	haas (*rabbit*)	raster (*fence*)
16) kwast (*paintbrush*)	verf (*paint*)	penseel (*paintbrush*)	poort (*gate*)	punter (*ship*)
17) maan (*moon*)	aarde (*earth*)	pluto (*pluto*)	tas (*bag*)	karaf (*carafe*)
18) druiven (*grapes*)	appel (*apple*)	vlierbes (*elderberry*)	verband (*bandage*)	raadhuis (*townhall*)
19) neus (*nose*)	oog (*eye*)	knokkel (*knuckle*)	paard (*horse*)	windhoos (*vortex*)
20) boek (*book*)	tijdschrift (*magazine*)	schrift (*notebook*)	knie (*knee*)	commode (*chest of drawers*)

Pictures	Related		Unrelated	
	HF	**LF**	**HF**	**LF**
21) ring (*ring*)	ketting (*necklace*)	collier (*necklace*)	piano (*piano*)	vlierbes (*elderberry*)
22) rugzak (*backpack*)	tas (*bag*)	buidel (*bag*)	oog (*eye*)	fret (*ferret*)
23) medaille (*medal*)	prijs (*prize*)	bokaal (*cup*)	bibliotheek (*library*)	telegram (*telegram*)
24) trein (*train*)	bus (*bus*)	pont (*ferry*)	aarde (*earth*)	collier (*necklace*)
25) voet (*foot*)	knie (*knee*)	scheen (*shin*)	tafel (*table*)	pluto (*pluto*)
26) kerk (*church*)	bibliotheek (*library*)	raadhuis (*townhall*)	tijdschrift (*magazine*)	ligament (*ligament*)
27) bijl (*axe*)	mes (*knife*)	houweel (*axe*)	sneeuw (*snow*)	hobo (*oboe*)
28) bank (*sofa*)	tafel (*table*)	commode (*chest of drawers*)	prijs (*prize*)	scheen (*shin*)
29) pleister (*plaster*)	verband (*bandage*)	kompres (*compress*)	brief (*letter*)	erker (*oriel*)
30) fles (*bottle*)	karaf (*carafe*)	beker (*mug*)	schoen (*show*)	houweel (*axe*)

Appendix B

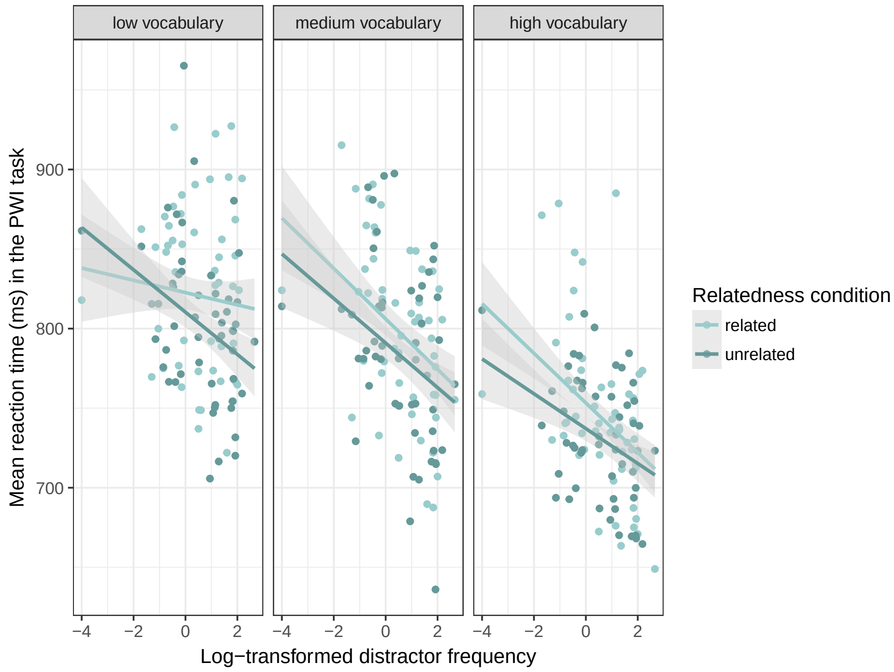

Figure 4.3: Interaction between the effects of distractor frequency and vocabulary (low vs. medium vs. high) on PWI reaction times (ms) in the unrelated vs. related condition. Confidence intervals are displayed in grey. Note that vocabulary was transformed into a categorical variable for illustrative purposes while all analyses were run on vocabulary as a continuous variable.

5 | Origins of individual differences in word learning: An exploration of cognitive and environmental effects in adult native speakers[1]

Abstract

In this study, we examined the relationship between individual differences in word learning and various cognitive abilities, as well as environmental effects in adult native speakers. We administered vocabulary measures, general processing speed tests, the Raven's Matrices (Raven, Raven, & Court, 1998), and a digit span test (Wechsler, 1997). In a word learning experiment, 111 participants learned 39 very low-frequency Dutch words and were tested on these words both right after training and after a delay of one week. In addition, we manipulated the number of exposures to the novel words so that each participant saw 13 low-, 13 medium-, and 13 high-exposure words.

Participants' test performance was better on Day 8 than on Day 1, indicating beneficial effects of overnight consolidation. Furthermore an increased number of exposures had positive effects on the retention of newly learned words. In addition, participants with greater vocabularies, higher processing speed, and higher Raven's scores showed advantages in the word learning task. Implications of our findings for future word learning studies are discussed.

[1]An adapted version of this has been submitted for publication as *Mainz, N., Smith, A.C., & Meyer, A.S. (submitted). An exploratory study of individual differences in adult novel word learning.*

Introduction

Previous research has shown that vocabulary knowledge varies considerably across native speakers of a language (Brysbaert et al., 2016b; Kidd et al., 2018). Individual differences in vocabulary have been associated with variation in language processing, with larger vocabularies being associated with more accurate and faster language comprehension (Brysbaert et al., 2016a; Yap et al., 2012, see also Chapters 2 and 3) and production (Rodriguez-Aranda & Jakobsen, 2011, see also Chapter 4). In addition, the effect of word frequency on lexical decision speed has been reported to be smaller for individuals with larger vocabularies. This has been claimed to indicate that the strength, robustness, or entrenchment of representations increases with the number of words in a vocabulary (Brysbaert et al., 2016a; Diependaele et al., 2013, see Chapter 2 for a more detailed description of this observation).

Thus, vocabulary size varies between individuals and this variation has an effect on individuals' language processing performance, which goes beyond simply knowing more words. A question which arises is what the origin is of this variation in vocabulary knowledge? Why do some people learn more words than others, thus ending up having a larger vocabulary? The origin of variation in vocabulary size is likely to be manifold and a number of factors, such as exposure and general intelligence, with presumably complex relationships between each other, are likely to be involved (Kidd et al., 2018). With the present study, we examined the relationship between individual differences in various verbal and nonverbal cognitive abilities and performance on a novel word learning task in adult native speakers.

Environmental factors and sleep

Previous studies on word learning have suggested a number of environmental as well as cognitive factors that might cause variation in word learning and vocabulary size (Kidd et al., 2018). Characteristics of the linguistic input that have been shown to affect word learning include the quantity and quality of the input (Hurtado et al., 2008; Jones & Rowland, 2017). Hurtado et al. (2008) observed that children who received more maternal input at 18 months of age showed better speech processing abilities and greater vocabularies at 24 months than children who had received less input early in development. In addition, the study demonstrated that children's real-time lexical processing abilities and vocabulary knowledge, both influenced by maternal input, are closely related and dependent on each other, and their relationship might be looked at in two different ways. Children who are exposed to more input are probably more practised at processing incoming speech, which makes their speech processing and therefore novel word learning more efficient and in turn leads to them acquiring a larger vocabulary. Alternatively, being exposed to more input, hence larger numbers of words, likely results in the child acquiring more words. This greater size and density of the child's lexicon might improve their language processing skills, simply due to a need for more fine-grained processing abilities (Hurtado et al., 2008).

Jones and Rowland (2017) examined the roles of amount of exposure to linguistic input (quantity) and diversity of linguistic input (quality) in children's language learning using both behavioural experiments and computational modelling. Input quantity was shown to be more important very early in the learning process, while lexical diversity turned out to have stronger positive effects on learning later in development. Furthermore, a model trained on lexically diverse input outperformed a model trained on quantitatively larger input on nonword repetition, sentence recall, and novel word learning. It is argued that the amount of a child's sublexical and lexical knowledge, which is dependent on the diversity of the linguistic input the child receives, affects novel word learning and consequently vocabulary size. Children with more sublexical and lexical knowledge are suggested to be able to use their existing knowledge to aid the processing of new incoming words. Consequently, they are faster and more efficient at processing novel input, which enables them to learn more and more quickly from their input, which eventually leads to greater vocabulary knowledge (Jones & Rowland, 2017).

Another characteristic of the learning situation that has been shown to influence word learning performance is whether or not novel phonological forms are presented along with meanings. Providing meaning, either in the format of a picture or a definition, had beneficial effects on the acquisition of new word forms (Hawkins, Astle, & Rastle, 2015; Savill, Ellis, & Jefferies, 2017; Takashima, Bakker, Van Hell, Janzen, & McQueen, 2014, 2017). This has been hypothesised to be due to the stronger involvement of both the semantic and episodic memory systems when learning meaningful words as compared to novel words without associated meaning (Takashima et al., 2017).

Furthermore, sleep has been reported to be an important factor affecting word learning. Lexical competition effects between novel and known words have, for example, often not been observed right after training on novel words but only in delayed tests, hence after intervening periods of sleep. It has been suggested that overnight consolidation is necessary for the integration of newly learned words or their representations within existing lexical knowledge (Bakker, Takashima, van Hell, Janzen, & McQueen, 2015; Brown & Gaskell, 2014; Weighall, Henderson, Barr, Cairney, & Gaskell, 2017).[2] Additionally, in an experiment by Kurdziel and Spencer (2016)

[2]But see Lindsay & Gaskell (2013) for a study showing that lexical integration of novel words might be possible without sleep.

participants learned novel words and were tested on these words right after training and after a delay either involving a period of sleep or not. Performance on the later test was significantly better for the individuals in the sleep group as compared to the awake group, indicating a beneficial effect of sleep on the retention of novel words. Similar observations have been made in other word learning studies, with generally better test performance after a delay involving sleep than right after training, i.e. without an intervening period of sleep (Takashima et al., 2014; Weighall et al., 2017).

Cognitive abilities

Besides these environmental effects and sleep, factors related to a learner's cognitive abilities have been shown to have an impact on word learning. There is a large body of research on the relationship between phonological short-term memory (PSTM) and vocabulary learning or size, with better PSTM being associated with greater vocabulary knowledge (Adams & Gathercole, 2000; Baddeley, Gathercole, & Papagno, 1998; Gathercole, 2006; Gupta & Tisdale, 2009a, 2009b). Children with relatively good PSTM abilities were, for instance, found to produce spontaneous speech comprising longer utterances, more diverse syntactic constructions, and a larger number of different words (Adams & Gathercole, 1995; Adams, 1996; Adams & Gathercole, 2000; Baddeley, 2003). Additionally, in a computational model Gupta and Tisdale (2009a) demonstrated that PSTM predicts vocabulary size and learning. More precisely, greater PSTM was found to be both cause and consequence of improved word learning. The fact that phonological word forms are serially ordered and extend over time has been argued to be the reason why some phonological storage ability is needed to process, reproduce and eventually learn novel words, which in turn trains and leads to improvement in PSTM (Gathercole, 2006; Gupta & Tisdale, 2009b).

Additionally, processing speed has been found to influence word learning. Fernald et al. (2006) and Marchman and Fernald (2008) have observed beneficial effects of faster processing speed on vocabulary learning in children. First of all, increased speed and accuracy of spoken word recognition at 25 months of age was related to faster and more accelerated growth in expressive vocabulary between 12 and 25 months (Fernald et al., 2006). Secondly, vocabulary size and speed of online word recognition at 25 months were shown to predict the same children's linguistic and cognitive skills, i.e. vocabulary, working memory, and intelligence, at eight years of age (Marchman & Fernald, 2008).[3] In addition, McMurray, Horst, and Samuelson (2012) demonstrated the beneficial effect of increased processing speed on word learning in a computational model. Faster processing

[3]See also Cristia, Seidl, Junge, Soderstrom, and Hagoort (2014), Ellis, Gonzales, and Deak (2014), and Weisleder and Fernald (2013).

networks showed higher accuracy rates on a multiple-choice test on novel words, which suggests processing speed to be one of the causes for more successful novel word learning (McMurray et al., 2012). Different proposals have been made as to which mechanisms or processes might underlie the word learning advantage that has been associated with increased processing speed. It has, for instance, been proposed that children who are faster at identifying words might have more time or resources that can be allocated to the processing of subsequent speech, which presents further opportunities for learning (Fernald et al., 2006). Furthermore, Marchman and Fernald (2008) suggested that as a consequence of being faster at processing a word form, these children might "have more resources available to process secondary aspects of the referential context that could support more richly instantiated lexical representations" (Marchman & Fernald, 2008, p. 14). Similarly, it may be that a word learner who is faster at processing the incoming auditory stimulus (word form) and its referent has more resources available to form not only a more detailed but also a more robust lexical representation on the first exposure to that stimulus, which can be strengthened and enriched during subsequent exposures to that word.

Finally, vocabulary size has been shown to predict word learning. In addition to observing a relationship between processing speed and word learning, Marchman and Fernald (2008) found that vocabulary at 25 months predicted children's vocabulary at 8 years. Hence, children with a larger vocabulary at first testing also learned more words in the following years (see also Feldman et al., 2005; Henderson & James, 2018). The question is why some children have a larger vocabulary than others in the first place. Measures of vocabulary perhaps capture a number of factors or skills involved in the acquisition of word knowledge. First of all, vocabulary size is likely an indicator of exposure to linguistic input. Maybe the high-vocabulary children had more exposure very early in development, and the vocabulary advantage they had thus gained was beneficial for further word learning. This increased exposure might be beneficial for further word learning because more exposure to language means more practice with linguistic input, and has been suggested to lead to an increase in sublexical and lexical knowledge and to benefit language processing speed, which in turn aid word learning (Fernald et al., 2006; Hurtado et al., 2008; Jones & Rowland, 2017; Weisleder & Fernald, 2013). Additionally, children who had more exposure to language in their first 25 months might also get more exposure in the following 6 years of life as they are presumably surrounded by the same speakers as in infancy. Hence, different effects related to the amount of exposure might underlie the vocabulary effect on word learning and thus might be capture by measuring vocabulary size. Another factor that might be captured in a measure of vocabulary is PSTM. As described above, a large number of studies have shown better PSTM to be associated with greater vocabulary knowledge. Hence, it could be that the underlying factor for some children having a larger vocabulary at 25 months and for their more successful vocabulary learning over the first 8 years of life is better PSTM.

Summary

Previous research has indicated that input quantity and quality, overnight consolidation, and the presence or absence of meaning affect novel word learning. Greater amounts of maternal speech input have been associated with greater vocabulary learning and faster language processing in children (Hurtado et al., 2008), and a computational model of word learning has built on and expanded these findings by demonstrating that input quantity is important for word learning very early in development, whereas the diversity of the input is more central to novel word learning at later stages in development (Jones & Rowland, 2017). What is unclear is whether and how individual differences in cognitive abilities are related to the effects of input quantity, especially in adult learners (Kidd et al., 2018). Does, for example, a learner who is a faster processor or has a larger vocabulary maybe need less input for learning? In addition, a considerable body of research has shown that overnight consolidation is beneficial for the retention of newly learned words (Kurdziel & Spencer, 2016; Takashima et al., 2014; Weighall et al., 2017). Hence, individual variation in consolidation might affect word learning performance, but again it is not known whether consolidation is related to or interacts with individual differences in verbal or non-verbal cognitive abilities. Individuals with larger vocabularies might, for instance, be better at consolidating lexical representations overnight, which would in turn aid novel word learning.

Aside from these factors, learners' cognitive abilities have been indicated to influence novel word learning. Better PSTM abilities and increased processing speed have been related to improved novel word learning and greater vocabulary size (Gathercole, 2006; Gupta & Tisdale, 2009b; Fernald et al., 2006; Marchman & Fernald, 2008). Both factors have mostly been studied in relation to word learning and vocabulary knowledge in childhood. In particular, PSTM has been argued to be especially important in early stages of word learning but supports novel word learning across the life span (Gathercole, 2006). It is unclear, though, how PSTM – just as general processing speed – is related to environmental factors and other cognitive factors in the context of word learning in adulthood (Kidd et al., 2018).

Finally, vocabulary size has been shown to predict word learning, with larger vocabularies being beneficial for learning (Henderson & James, 2018; Marchman & Fernald, 2008). Again, this has mostly been studied in children but might continue to affect word learning in adulthood. In addition to this, variation in vocabulary has been associated with variation in PSTM and processing speed, as indicated earlier, and is potentially also related to differences in nonverbal intelligence. Hence, it is not known how much variance in word learning is left for vocabulary to explain if factors that are arguably associated with variation in word knowledge, such as PSTM and processing speed, are taken into account.

Current study

The present study set out to address these various open questions on word learning. First of all, we studied word learning in adult native speakers. Most of what is known about individual differences in word learning is based on developmental studies, and little research has looked at variation in word learning in adults (Kidd et al., 2018). It is known, though, that word learning continues across the life span and extensive word learning takes place in adulthood (Hartshorne & Germine, 2015). Additionally, the acquisition of novel lexical items in adulthood presumably continues to be affected by the amount and type of input an adult receives, by the learning conditions, such as the possibility for overnight consolidation, and by individual variation in cognitive abilities. However, as stated by Kidd et al. (2018), little research has looked at role of individual differences in language learning across the life span.

Secondly, we examined the relationship between novel word learning performance, individual differences in various cognitive abilities, and environmental properties of the word learning process. As indicated in a recent review by Kidd and colleagues (2018), the question of how internal sources of variation interact with the environment is an important but open issue. We investigated the effects of variation in vocabulary, PSTM, general processing speed, and nonverbal intelligence on word learning, and looked at whether these factors interact with effects of input quantity and overnight consolidation. We assessed nonverbal intelligence in addition to the other cognitive abilities because it may be argued that nonverbal intelligence is the underlying factor driving variation in vocabulary and word learning as well as in processing speed and possibly also PSTM.

Due to the fact that not much is known about individual differences in word learning in adult native speakers and the relationships with the various cognitive abilities, this study was exploratory in nature. We aimed at providing first insights into potentially very complex relationships between adults' cognitive abilities and their word learning performance.

As stimuli in the word learning task, we used existing but very low-frequency and low-prevalence[4] Dutch words which could be depicted in photographs. Hence, participants learned the phonological forms of new words in their native language, which were associated with meaning in the format of one coloured photograph each. By using existing Dutch words we ensured that the task actually reflected first language learning in adulthood and not second language learning (Kurdziel & Spencer, 2016). We decided to use form-meaning pairings instead of novel phonological word forms only because it has been shown that meaning matters, with the presence of meaning being beneficial for word learning (Savill et al., 2017; Takashima et al., 2017). In addition, using not only word forms but also the associated meanings as stimuli was assumed to be more ecologically valid in the context of word learning in adult native speakers. Adults are presumably most of the time presented with not only novel word forms but with an associated meaning.

Participants were trained on these words in two tasks, similar to Takashima et al. (2017). First, they heard each word while seeing the corresponding picture on the screen and were asked to repeat the word aloud. Secondly, participants completed a four-alternative forced-choice (4AFC) picture-matching task where they were presented with the auditory form of a word and were asked to select one of four pictures on the screen that corresponded to the word. Feedback was provided after each trial to enable learning the correct form-picture pairings. In the second training task, the number of exposures of the novel words was varied. Words in the low-exposure condition were presented only twice, while words in the medium-exposure condition were presented eight times and those in the high-exposure condition 15 times. This manipulation was assumed to reflect effects of frequency or quantity of exposure in natural language use and learning (Brysbaert, Mandera, & Keuleers, 2018). Thus, the number of exposures was expected to affect word learning, with high numbers of exposure leading to more successful learning as indicated by higher accuracy rates at test than medium or low numbers of exposure during training. This manipulation enabled us to examine the relationship between individual differences, for example in vocabulary size or processing speed, and the effect of quantity of exposure. If individuals with larger vocabularies or those who are faster processors are indeed more efficient and faster at processing novel words allowing them to more quickly, i.e. after fewer exposures, form representations of the association between form and meaning (Hurtado et al., 2008; Marchman & Fernald, 2008), we might see an interaction between vocabulary score or processing speed and exposure condition in our experiment.

[4]Prevalence is a measure of how many speakers of Dutch in the Netherlands know a given word (Keuleers et al., 2015).

Participants' word learning performance was assessed using two different tasks, an associative and an explicit memory task. The first was a 4AFC picture-matching task, which was largely the same as the multiple-choice training task, the only difference being that no feedback was provided. The second task was a picture naming task, where participants were presented with the novel pictures one-by-one and were asked to produce the corresponding word. Both tests were administered right after training on the novel words and one week later. This allowed us to examine potential interactions between variation in cognitive abilities and the consolidation of novel words, as indicated before. Individual differences in the cognitive abilities of interest were measured in a battery of seven tasks. Three different tasks were administered to assess participants' general processing speed, namely a visual and an auditory simple reaction time (RT) task as well as a letter comparison task. In the simpler psychomotor speed tasks, participants had to press a button as quickly as possible upon presentation of a visual or auditory stimulus. Thus, no decision processes were involved. By contrast, in the letter comparison task, participants were asked to make a decision about whether two visually presented letter strings were the same or different and press a button to indicate their decision as quickly as possible upon presentation of the stimulus (Salthouse, 1996). Hence, the tasks differed in modality, i.e. visual vs. auditory, and in task complexity, i.e. psychomotor vs. perceptual speed (Albinet, Boucard, Bouquet, & Audiffren, 2012; Cepeda, Blackwell, & Munakata, 2013; Salthouse, 2000). It has been shown that more complex processing speed tasks are strongly related to executive control functions, thus, measuring not only processing speed but also executive functions. Simpler perceptual speed tasks, by contrast, have been found to be correlated with executive functions in children and older adults but not in young adults, indicating that these measures are purer measures of general processing speed in young adults than the perceptual speed tasks (Cepeda et al., 2013). We administered these different types of tasks and calculated composite scores representing participants' performance on all three tasks in order to yield a comprehensive measure of processing speed representing different aspects of speed of processing.

Vocabulary was assessed in two different tasks that have been used in previous studies, namely Andringa et al.'s (2012) receptive multiple-choice test and an open antonym test (Mainz et al., 2017). The measures of word knowledge, thus, included an open and a multiple-choice test both addressing participants' deeper knowledge of the semantic relationships of words by asking to select or generate words that are synonymous or antonymous to the targets (Henriksen, 1999). Previous studies have demonstrated that using a large battery of vocabulary tests might not be necessary, especially given time constraints in a study with several individual differences measures (Mainz et al., 2017).

Therefore, we decided to use two rather short tests which differed in test type and format (Bowles & Salthouse, 2008), and calculated a composite score of vocabulary based on both tests to be used in all analyses.

Non-verbal intelligence was assessed in a 20-minute version of the Raven's Advanced Progressive Matrices (Huettig & Janse, 2016; Raven et al., 1998). Finally, we used the forward digit span task taken from the Dutch version of the Wechsler Adult Intelligence Scale III (WAIS-III-NL Wechsler, 1997), which is a task typically employed to assess individuals' PSTM (Gathercole, 2006; Gupta & Tisdale, 2009a).

To sum up, the aim of the present study was the investigation of the relationship between individual differences in various cognitive abilities and word learning. The study's design involving a manipulation of the quantity of exposure per word and tests right after training and after a period allowing for overnight consolidation enabled us to explore potential relationships between sleep, environmental and cognitive factors affecting word learning in adults. The following hypotheses guided our study. (1) We expected the number of exposures to affect performance in the test phase. Specifically, we predicted that high-exposure words would elicit higher accuracy rates and faster RTs in the 4AFC test and higher accuracy rates and less severe naming errors in the naming task. (2) Participants were expected to perform better on Day 8 than on Day 1. (3) Faster general processing speed, higher vocabulary and intelligence scores, and better PSTM were hypothesised to be beneficial for word learning. (4) It was predicted that effects of number of exposures may interact with individual differences in these cognitive abilities. Hence, individuals with, for instance, higher nonverbal intelligence scores, better general processing speed abilities, or higher vocabulary scores might potentially show smaller effects of number of exposures on word learning performance. (5) Finally, we expected that we may observe interactions between the effect of Day of testing and any of the individual differences measures. This was based on the idea that variation in the cognitive abilities assessed in this study might affect the process of lexical consolidation and integration while not being predictive of the storage and retrieval of episodic representations of newly learned words. These expectations are speculative and we did not have strong predictions as to which cognitive abilities might affect short-term rather than long-term storage or the other way around.

Method

Participants

In total, 124 individuals gave informed written consent to participate in this study. Thirteen participants had to be excluded due to technical failure or failure to complete all experimental tasks. Hence, data of 111 individuals (89 females, 22 males) aged between 18 and 35 years (M = 22.81, SD = 2.81) was left for analyses. Most participants were students at Radboud University, Nijmegen (N = 83) or studied at other universities in the Netherlands (N = 3). A smaller group of participants were students at vocational universities, either at the Hogeschool van Arnhem en Nijmegen (N = 22) or other vocational universities (N = 3).

All participants were recruited using the participant database of the Max Planck Institute for Psycholinguistics and were paid 20 Euros for their participation. Ethical approval was granted by the Faculty of Social Sciences of the Radboud University Nijmegen.

Materials and design

Word learning experiment

Thirty-nine disyllabic and trisyllabic Dutch words, which were low in prevalence (min = -2.13; max = -1.32; M = -1.62; SD = 0.20) served as stimuli in the word learning experiment (see Appendix A for a list of all words). Following Keuleers and colleagues' (2015) prevalence norms, the words were known by less than 10% of the native speakers of Dutch in the Netherlands. In addition, none of the words appeared in the SUBTLEX corpus (Keuleers et al., 2010) and five native speakers of Dutch, with academic backgrounds in psycholinguistics and cognitive neuroscience indicated that they did not know any of the words. Furthermore, after the word learning experiment participants were asked to indicate whether they had come across any of the novel words before participating (see below).

The word form materials consisted of recordings of all 39 words, spoken by a female native speaker of Dutch. Each of the words was paired with one colour photograph depicting the referent. Each photograph appearing on the screen was at least 10 cm by 10 cm large but the exact size varied between pictures as some were vertical and some horizontal photographs. In some cases, red arrows and/or shading were added using Photoshop in order to clearly indicate the referent of the respective word. All photographs were downloaded from freely available images on the Internet (see Appendix A for examples).

Three sets of stimuli were created, each including all 39 words with equal numbers of words being assigned to the low-, medium-, and high-exposure conditions. Every word appeared in a different exposure condition in each of the sets. The word *wimberg*, for instance, was a low-exposure word in set 1, a medium-exposure word in set 2, and a high-exposure word in set 3. Thus, in the training phase, each participant was trained on 13 low-exposure words (3 exposures), 13 medium- exposure words (9 exposures), and 13 high-exposure words (16 exposures). Every participant completed two tasks in the training phase (task 1: Exposure, task 2: Four alternative forced-choice task). The manipulation of frequency of exposure was implemented in Training 2 (see below). The test phase also consisted of two tasks (task 1: Four alternative forced-choice task, task 2: Picture naming) used to assess associative as well as explicit memory of the novel words.

In the lists for the four alternative forced-choice tasks (4AFC) in Training 2 and Test 1, each target word was paired with three foil words. The foils also came from the respective set of trained novel words and the three foils in each display always consisted of one word from each exposure condition. Hence, every target picture was displayed along with one low-, one medium-, and one high-exposure foil. Each instance of a novel word was combined with a different combination of foils. In addition, the positions in which the target and foil images appeared on the display were counterbalanced within each list. Hence, the targets and the fillers appeared equally often in each of the four positions in the display.

The lists for all training and test phases were pseudorandomised. Two instances of the same target word were separated by at least five trials, and in the 4AFC tasks two consecutive trials differed in at least two images on the display. The order within the lists was fixed so that every participant assigned to the same set of novel words saw the same stimulus lists. Finally, assignment to one of the three stimulus sets was counterbalanced across participants.

Procedure

Participants carried out all tasks in a sound attenuated experiment room at the Max Planck Institute for Psycholinguistics. They were seated in front of a 17-inch screen (Iiyama LM704UT) and all tests were administered employing the software Presentation (version 16.5, www.neurobs.com). Auditory stimuli were presented using HD 280 Sennheiser headphones.

On Day 1, participants completed both training tasks of the word learning experiment. They were first exposed to the spoken form and corresponding pictures of all novel words (Training 1), followed by a 4AFC picture matching task including all novel words (Training 2; see Figure 5.1). Participants were instructed to memorise the novel word-picture pairs and they were informed that they would be tested on them later. Immediately after training, they performed the recognition test (Test 1), which was again a 4AFC picture matching task, and a picture naming test of explicit memory (Test 2). In addition, they filled in a questionnaire where they were asked to indicate their occupation, highest level of education, and information on whether they were fluent in a language other than their native language Dutch. One week later (Day 8), all participants returned to the lab and again performed both Tests 1 and 2 of the word learning experiment. In addition, participants completed a familiarity test where they were presented with the spoken form of all novel words and were asked to indicate whether they had known the word before participating in the experiment. If the answer was yes, they were asked to write down whether they knew the meaning and where they had come across that word, if they remembered. Finally, they were asked to complete the individual differences measures described above.

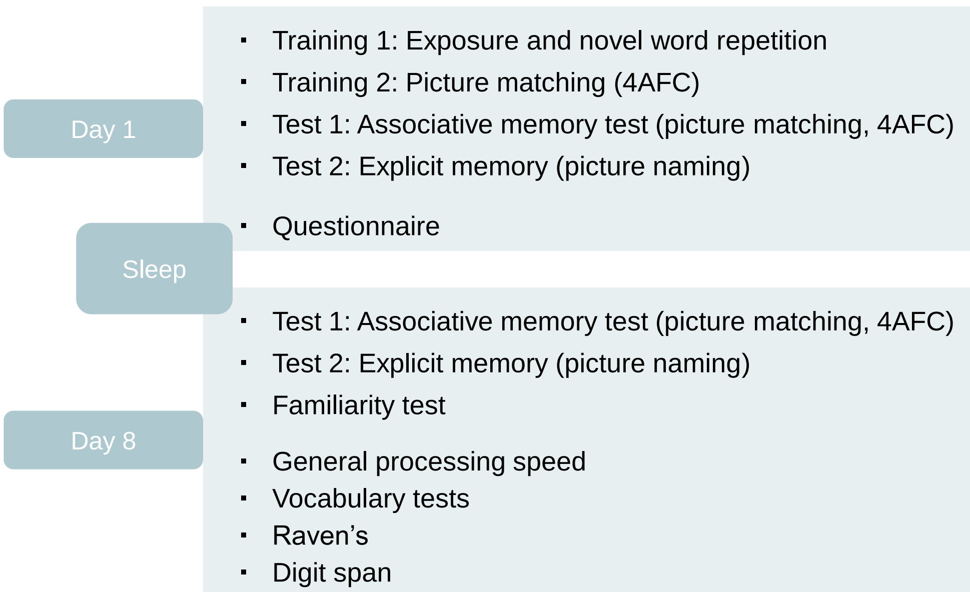

Figure 5.1: Schematic representation of the overall procedure.

Training 1: Exposure

Each trial of the exposure phase began with a fixation cross, which stayed on the screen for 500 ms. After that participants were simultaneously presented with the spoken form of one of the novel words and the corresponding picture on the screen. They were instructed to repeat the word upon hearing it. Training 1 was self-paced and the following stimulus was initiated by pressing the space bar. Each picture-word pair was presented once.

Training 2: Picture matching

On each trial in the picture matching task, participants first saw a fixation cross on the screen (500 ms) before they were presented with four pictures on the screen and a concurrent auditory stimulus. The task was to select the picture that corresponded to the word they heard by pressing one of the buttons on a button box. The four pictures were displayed until a response button was pressed. Each trial was followed by feedback. Thus, upon button press the auditory stimulus was repeated and only the correct picture was again presented in the centre of the screen until one of the buttons on the button box was pressed. Hence, the task was fully self-paced and participants were instructed to memorise the picture-word pairs.

As indicated before, each participant saw equal numbers of low-, medium-, and high-exposure words, while low-repetition words were presented twice, medium-repetition words 8 times, and high-repetition words 15 times throughout the entire Training 2 phase. Each participant was assigned to one of the three sets of words (see Materials section above); hence, which word appeared in which condition was counterbalanced across participants.

Training 2 was divided into three blocks. In the first block, participants were presented with the 13 high-frequency words only and saw each of them seven times (91 trials). Block 2 was comprised of 13 medium- and 13 high-frequency items, each of which was presented six times (156 trials). Block 3, finally, included all 39 novel words from all three frequency conditions and each picture-word pair was presented twice. In this way, we were able to control for the number of times that each novel word had been repeated before acting as foil for a low- vs. medium- vs. high-exposure target word. The blocks were ordered so that the number of items increased with each block because that appeared closer to a natural learning situation than decreasing the number of items within each block. Within each of the blocks, the order of the stimuli was pseudorandomised so that the same combination of visual stimuli was never repeated in two consecutive trials. Participants were instructed that they could take a short break between the blocks.

Test 1: Associative memory (picture matching)

The associative memory test was a 4AFC picture matching task, just as in Training 2. Each trial again started with a fixation cross (500 ms) followed by the display of four pictures on the screen and a concurrent auditory presentation of a trained novel word.

Participants were asked to select the picture that corresponded to the novel word they heard by again pressing one of the number buttons 1 to 4 on the keyboard. The pictures stayed on the screen until a button was pressed. No feedback was provided in this task and pressing one of the response buttons initiated the following trial. The test consisted of two blocks with each block containing each novel word once. Within each of the blocks, presentation of the stimuli was pseudorandomised so that words of the same repetition condition did not appear in more than three consecutive trials and two consecutive trials differed in at least two of the pictures on the display. All participants who had been trained on the same set of words, i.e. the words in the same exposure conditions, saw the same list, the order of which was fixed. Participants were instructed to respond as quickly and accurately as possible.

Test 2: Explicit memory (picture naming)

This explicit memory test of novel word learning was a picture naming task. Each trial started with a fixation cross, which stayed on the screen for 500 ms, followed by a picture corresponding to one of the trained novel words. Participants' task was to name the picture as quickly and accurately as possible and press the space bar to proceed to the following trial. The task was divided into two blocks each of which included all 39 novel words. Participants could take a short break between the blocks if they wished.

Individual differences measures

Visual processing speed

On each trial of the simple visual processing speed task, a fixation cross was presented in the centre of the screen followed by a black and white line drawing of a triangle. The stimulus onset varied randomly between 1 and 2 seconds. The triangle fitted into a virtual frame of 5 by 5 cm and was also presented in the centre of the screen. Participants were instructed to press a button on a response button box as quickly as possible upon presentation of the triangle. A button press initiated an intertrial interval of 1 second before the next trial started with the presentation of the fixation cross. Participants completed eight training trials before the actual experiment, which consisted of 20 trials.

In this task, RTs were measured and participants' mean speed across all trials was taken to indicate their processing speed performance.

Auditory processing speed

The auditory processing speed task was structurally identical with the visual task, the only difference being that the stimulus was a 550-Hz tone, which was 400ms long, instead of a triangle on the screen. Again, each trial started with a fixation cross, which was presented in the centre of the screen for a jittered time interval of between 1 and 2 seconds and was followed by a tone. The auditory stimulus was presented via headphones and participants were instructed to press a button on the button box as quickly as possible upon presentation of the tone. The experiment consisted of 20 trials in total and was preceded by eight practice trials.

Processing speed performance was operationalized as the mean RT across all trials in this task.

Letter comparison task (processing speed)

The materials of this test were based on a paper-and-pencil task assumed to measure processing speed as described in Earles and Salthouse (1995) and Salthouse (1996). Each trial started with a fixation cross, which was presented in the centre of the screen for 500 ms. After another 100 ms participants were presented with pairs of letter strings (all consonants) on a computer screen, one centered in the upper half of the screen and one in the lower half. These pairs were comprised of either the same letter string twice (same condition) or letter strings that differed by one letter (different condition). Participants were instructed to decide as quickly and accurately as possible whether the two strings of letters were the same or different by pressing response buttons on a button box. The button-press initiated a 1000 ms intertrial interval, followed by the next trial. All letters were presented in point 60 Arial font.

The task was divided into two blocks with a short break in between. The first block consisted of 24 three-letter strings and the second block of 24 six-letter strings. In both blocks, half of the letter strings belonged to the same condition while the other half came from the different condition.

RTs were measured and processing speed in this letter comparison task was operationalized as the mean RT across all correct trials.

Vocabulary: Receptive multiple-choice test

This test was developed by Andringa and colleagues (2012). Participants were presented with target words, such as *mentaliteit* (mentality) or *tentatief* (tentative), embedded in different neutral carrier sentences with the target word being marked with two asterisks. Each of the sentences was presented along with five answer options, one of which was a description of the target word and one being *Ik weet het echt niet* (I really don't know). For example, the target word *mentaliteit* (mentality) was presented with the answer options *tafel* (table), *persoon* (person), *manier van denken* (way of thinking), and *sfeer* (atmosphere; see Appendix A of Chapter 2 for the test materials).

Participants were asked to select the word that was synonymous with or explained the meaning of the target word, or indicate that they did not know the answer, by pressing one of the number buttons 1 to 5 on the keyboard. The questions stayed on the screen until the participant had selected an answer, which initiated the presentation of the following test item.

The original test consists of 60 target sentences. In the present study the first sentence was used as example question so that the test comprised a total of 59 questions. Both the original sequence of items and the positions of the correct responses were kept from Andringa et al's (2012) test. The vocabulary score on this test was calculated by counting the number of correct responses.

Vocabulary: Open antonym test

The open antonym test included 25 test items, which were presented individually and without carrier sentences. Some of the target words were taken from the Toets Gesproken Nederlands (TGN), a Dutch language test used to assess language for immigration requirements (Kerkhoff et al., 2005).

Participants were instructed to type in an antonym for each word using the keyboard (see Appendix A of Chapter 2 for the test materials). They were allowed to proceed to the next question without answering the previous one, if they did not know the answer. Pressing the enter key initiated the following trial. Prior to the test, instructions were presented on the screen including two examples.

The test items in the open antonym test represented a frequency range between 0 and 60.69 counts per million in the SUBTLEX-NL corpus (M = 9.09, SD = 13.14). The prevalence values of the target words ranged from 1.03 to 3.32 (M = 2.49, SD = 0.68).

One point was given for each correct response and 0.5 points were given for answers that demonstrated a participant's knowledge of the word or concept without being completely correct. This was true for misspelled responses or cases similar to the following one: when *rustig* (calm) as opposed to *rust* (calmness) was given as antonym for *lawaai* (noise).

Short-term memory: Forward digit span

The materials for the forward digit span task used to assess phonological short-term memory were taken from the Dutch version of the Wechsler Adult Intelligence Scale III (WAIS-III-NL Wechsler, 1997). We used a computerized version of the task where on each trial participants were presented with a string of digits in spoken form. All stimuli were recordings spoken by a native speaker of Dutch. During the auditory presentation of the strings of digits, a fixation cross was presented on the screen and was replaced by the words *Typ nu* (Type now) after the end of the recording. Participants were instructed to type in the series of digits they heard using the computer keyboard.

Participants were tested on strings of two to nine digits, with two trials for each sequence length and starting with the two-digit strings followed by three-digit strings and so forth. When participants made errors on two trials of the same length, the task stopped automatically. The number of digits of the longest sequence they recalled correctly was participants' short-term memory score.

Non-verbal intelligence: Raven's matrices

The Ravens Advanced Progressive Matrices were administered to assess non-verbal intelligence. We used a computerised version where on each trial participants were presented with a matrix of 9 geometrical patterns with one pattern in the bottom right corner absent. Participants were instructed to select one out of eight additional patterns that would complete the matrix by clicking on it using a computer mouse. The target matrices were presented as large pictures in the centre of the screen and the eight answer alternatives were presented in two rows of four pictures each below that. Participants could skip trials, which would be presented again at the end.

The materials consisted of 36 matrices in total and participants were given 20 minutes to complete the task. The remaining time was indicated in the top right corner of the screen. Participants' intelligence score was operationalized as the number of correct responses within the given 20 minutes.

Analyses and results

Individual differences measures

Prior to all analyses, the RTs from the three processing speed tasks were trimmed by excluding all RTs that were 2.5 SDs above or below each participant's mean. For each of the three speed tasks, less than 3.5% of all data was excluded from all further analyses.

For the purpose of the individual differences analyses, composite measures of general processing speed and vocabulary were calculated. Regression-based factor scores for both the vocabulary tests and processing speed tasks were calculated for each participant using the PCA method in SPSS (DiStefano et al., 2009). For both composite measures, only one underlying factor was assumed and each individual's loading on that factor based on their performance on the three processing speed tasks or two vocabulary test scores was calculated. Hence, the processing speed score reflects participant's performance on all three processing speed tasks and the vocabulary score reflects each participant's performance on the two vocabulary tests.

Table 5.1 displays participants' RTs in the processing speed task as well as their vocabulary, digit span, and Raven's scores. The scores in the two vocabulary tests, i.e. Andringa et al.'s (2012) receptive multiple choice test and the open antonym test, were similar to Experiment 1 reported in Mainz et al. (2017), where a similar population was tested on the same vocabulary tests (see Chapter 2).

Table 5.1: RTs or raw test scores for all individual differences measures.

Test	**Minimum**	**Maximum**	**Mean**	**SD**
Visual speed (ms)	178	340	225	26
Auditory speed (ms)	181	374	230	31
Letter comparison (ms)	492	1602	980	205
Antonym test	14.0	23.0	19.37	2.0
Andringa	27.0	56.0	39.72	5.74
Digit span	2.9	9.0	6.77	1.46
Raven's	7.0	36.0	22.38	5.97

Bivariate correlations between all individual differences measures as used in all subsequent analyses are displayed in Table 5.2. We did not find any significant correlations between the individual differences measures. [5]

[5]See Appendix B for a table with the bivariate correlations between all seven individual differences tasks. As expected, the two vocabulary measures significantly positively correlated with one another (r = .37, $p < .01$), as did the simple auditory and visual processing speed measures (r = .61, $p < .01$). The third more complex measure of processing speed (letter comparison) did not correlate significantly with the two others. Finally, we found a significant (mild) correlation between participants' performance on the digit span and the letter comparison tasks (r = -.22, p = .02). Individuals with higher digit span scores were faster on the letter comparison task.

Table 5.2: Correlation coefficients for the bivariate correlations between the individual differences measures. *P*-values are given in brackets.

	Vocabulary	**Speed**	**Digit span**
Speed	-.15 (.13)		
Digit span	.17 (.08)	-.01 (.90)	
Raven's	.14 (.13)	.16 (.09)	.007 (.94)

Word learning experiment

As indicated above, following the experiment on Day 8, participants were asked whether they had known any of the words from the experiment prior to their participation. Forty-one participants indicated that they had known at least one word before participating in the experiment; ten participants were familiar with three words, three participants with four words, and two participants indicated that they had been familiar with five or seven words. All other participants ($N = 26$) had been familiar with one or two words prior to coming to the lab for the word learning experiment. Note that several of the words that participants indicated to be familiar with were specialised vocabulary that they had known due to having specific interests or hobbies.[6] For each participant, we excluded the words that they had known before coming to the lab from all further analyses.

To establish participant's performance on the word learning task we first analysed performance on Training 2 (picture matching), Test 1 (picture matching), and Test 2 (picture naming) using models without individual differences measures as predictors. In this way, we could first get an impression of whether our task manipulations (Exposure Condition, Day of testing) affected word learning performance using simpler models with fewer predictors, before adding all four individual differences measures and respective interactions to the mixed-effects models.

All performance measures of interest were analysed in mixed-effects models using the lme4 package (version 1.1.15, Bates et al., 2015) in R (R Core Team, 2017). For the purpose of analysing the binary dependent variable accuracy, we used the glmer function, and for analysing all other continuous dependent variables we used the lmer function. All models included an intercept and fixed effects for the task manipulations. These were

[6] The word *firn*, for example, refers to a specific layer of ice; one participant knew that because of extensive experience with hiking.

Repetition in the model on accuracy in Training 2 (picture matching), and Day (Day 1 vs. Day 8), Block (1 vs. 2), and Exposure Condition (low vs. medium vs. high) as well as the interactions between these fixed effects in the models on the data from Test 1 (picture matching) and Test 2 (picture naming). In addition, all models included by-participant and by-item adjustments to the intercept (random intercepts). The categorical variables Day, Block, and Exposure condition were dummy coded. Day 1 and Block 1 were the reference levels for the binary variables. In the case of the Exposure Condition factor, which had three levels (low vs. medium vs. high), *medium* was the reference level.

Training 2: Picture matching

We found a significant effect of Repetition on accuracy ($\beta = 0.28$, $z = 15.26$, $p < .001$). Response accuracy became higher as the number of repetitions increased. It is important to note that accuracy rates in this task were close to ceiling with on average 95.3% correct ($SD = 3.03$).

Test 1: Associative memory (picture matching)

Two performance measures were taken in Test 1, namely accuracy and reaction time (RT). The accuracy model failed to converge with the above-mentioned structure. After excluding interactions between the fixed-effects the accuracy model converged. Accuracy in the picture matching test was found to be predicted by the medium vs. high contrast of Exposure Condition (medium vs. high: $\beta = 0.50$, $z = 2.41$, $p = .02$), Block ($\beta = 0.37$, $z = 4.17$, $p < .001$), and Day of testing ($\beta = -0.70$, $z = -4.32$, $p < .001$). Participants' accuracy rates were higher for high vs. medium exposure words. In addition, their accuracy rates were higher on Day 1 than on Day 8, and in Block 2 as compared to Block 1. However, it has to be noted that accuracy rates were overall close to ceiling with mean accuracy rates of 97.5% on Day 1 and 95.9% on Day 8 (see Table 5.3).

Table 5.3: Mean accuracy rates (%) in Test 1 (picture matching) for both days and all conditions. Standard deviations are given in brackets.

Exposure condition	**Accuracy rates (%)**		
	Day 1	**Day 8**	**Total**
Low	97.64 (4.18)	94.69 (6.0)	96.16 (5.37)
Medium	97.97 (3.05)	96.02 (4.97)	96.99 (4.23)
High	98.73 (2.55)	96.80 (4.19)	97.76 (3.59)

The model on RTs in Test 1 showed significant effects of one of the Exposure Condition contrasts, namely low vs. medium (low vs. medium: β = 0.04, t = 7.22, $p <$ .001), with faster RTs for medium vs. low exposure words (see Table 5.4). In addition, the effect of Day was significant (β = 0.14, t = 25.05, $p <$ .001), with faster RTs on Day 1 than on Day 8. Furthermore, the two-way interaction between the medium vs. high exposure contrast and Day was significant (medium vs. high: β = -0.03, t = -3.52, $p <$ .001), indicating that the RT difference between medium and high exposure words was stronger on Day 8 than on Day 1. Furthermore, the interaction between Block and Day was significant (β = -0.10, t = -12.71, $p <$ .001). The RT difference between the two blocks was larger on Day 8 than on Day 1. Finally, the interactions between Block and the two Exposure Condition contrasts were significant (medium vs. low: β = -0.02, t = -2.54, p = .01; medium vs. high: β = -0.02, t = -2.37, p = .02). The medium vs. low contrast was stronger in Block 2, whereas the medium vs. high contrast was weaker in Block 2.

Table 5.4: Mean reaction times (ms) in Test 1 (picture matching) for both days and all conditions. Standard deviations are given in brackets.

Exposure condition	**Reaction times (ms)**		
	Day 1	**Day 8**	**Total**
Low	1884.92 (430.74)	2468.50 (763.98)	2176.71 (684.39)
Medium	1738.19 (350.50)	2225.81 (588.97)	1982.0 (541.77)
High	1683.97 (330.77)	2121.32 (620.71)	1902.64 (542.46)

Test 2: Explicit memory (picture naming)

In Test 2, we also analysed two measures of word learning performance, namely the binary measure accuracy and the Levenshtein Distances (LD) between target and response. Participants' responses in the naming task were coded independently by two native speakers of Dutch. They transcribed each participant's actual responses and judged whether they were correct or incorrect. In most cases, the native speakers agreed on their judgements. The unclear cases were solved by a third person, who listened to the recordings again and selected one of the transcriptions. The vast majority of disagreements were merely based on different spellings and could easily be solved. In addition to the binary measure of response accuracy, we calculated the LDs between participants' responses and the targets. This was assumed to be a more fine-grained measure of performance on this task indicating participants' novel word knowledge.

The mixed-effects model analyses on picture naming accuracy showed significant effects of the medium vs. low Exposure Condition contrast (medium vs. low: β = -0.49, z = -7.14, p < .001) and Day (β = 0.80, z = 12.21, p < .001). Accuracy rates were higher for medium vs. low exposure words and participants responded more accurately on Day 8 than on Day 1 (see Table 5.5).

Table 5.5: Mean accuracy rates in Test 2 (picture naming) for both days and all conditions. Standard deviations are given in brackets.

Exposure condition	**Accuracy rates (%)**		
	Day 1	**Day 8**	**Total**
Low	25.33 (18.24)	36.31 (21.24)	30.79 (20.52)
Medium	33.51 (21.06)	47.08 (22.81)	40.27 (22.95)
High	33.76 (23.79)	49.64 (24.10)	41.66 (25.21)

The LDs between participants' responses and the target words were significantly predicted by the medium vs. low contrast of the Exposure Condition factor (medium vs. low: β = 0.16, z = 4.14, p < .001) and by Day (β = -0.20, z = -5.75, p < .001) (see Table 5.6). LDs were greater for low- as compared to medium-exposure words, and LDs were smaller on Day 8 than on Day 1. Hence, participants' responses were closer to the target words in the medium vs. low exposure condition, and on Day 8 than on Day 1.

Table 5.6: Mean Levenshtein Distances in Test 2 (picture naming) for both days and all conditions. Standard deviations are given in brackets.

Exposure condition	**Levenshtein Distances**		
	Day 1	**Day 8**	**Total**
Low	3.41 (2.70)	3.01 (2.81)	3.21 (2.76)
Medium	2.91 (2.72)	2.38 (2.75)	2.65 (2.75)
High	3.06 (2.79)	2.32 (2.81)	2.69 (2.83)

Individual differences in word learning

The main interest of the present study was the relationship between individual differences in general processing speed, vocabulary, PSTM (digit span), non-verbal intelligence (Raven's), and performance in the word learning task. All performance measures of interest were analysed in mixed-effects models using the same functions from the lme4 package (version 1.1.15, Bates et al., 2015) in R (R Core Team, 2017) as indicated before. Composite scores representing performance on the three processing speed tasks and the two vocabulary tests, respectively, were calculated and used in all analyses. Digit span and Raven's scores were z-transformed using the scale function in R prior to all analyses.

All analyses were exploratory in nature. We started with maximal models (Barr et al., 2013), which in all cases did not converge. We, therefore, reduced model complexity as much as it was deemed acceptable and reasonable in order to enable the models to converge. Unless otherwise stated, all models included an intercept and fixed effects for all individual differences measures (processing speed, vocabulary, digit span, Raven's) as well as per-participant and per-item adjustments to the intercept (random intercepts). In addition and just as described above, task-relevant manipulations were included as fixed effects. This was Repetition in the model on Training 2 (picture matching), and Day of testing (Day 1 vs. 8), Block (block 1 vs. 2) and Exposure Condition (low vs. medium vs. high) for both Test 1 (picture matching) and Test 2 (picture naming). Importantly, all models were limited to two-way interactions between the fixed effects and, unless stated otherwise, they did not include random slopes.

Training 2: Picture matching

Likely due to participants rapidly performing at ceiling on this task, mixed-effects models that included individual differences measures as predictors applied to participants' performance (accuracy) in Training 2 did not converge. We therefore ran a linear regression using the lm function in R. The linear regressions on accuracy in this task showed significant effects of Repetition (β= 0.22, z = 22.87, p < .001) and vocabulary (β = 0.10, z = 2.17, p = .03). Accuracy rates increased with an increasing number of repetitions, and individuals with higher vocabulary scores had higher accuracy rates. In addition, the interactions between digit span and Raven's score was significant (β= -0.11, z = -2.43, p = .02), with stronger effects of digit span score for individuals with weaker Raven's performance. Finally, we observed a significant interaction between Raven's and vocabulary scores (β= 0.15, z = 2.91, p = .004). Individuals with higher Raven's scores showed stronger beneficial effects of high vocabulary scores than individuals with lower Raven's scores (see Appendix C for plots).

Test 1: Associative memory (picture matching)

None of the mixed-effects models on accuracy including the individual differences measures as predictors converged. This is likely due to mean accuracy rates close to ceiling on both days, with 97.5% on Day 1 and 95.9% on Day 8. We therefore decided to focus on RTs as the measure of performance on this task.

RTs were log-transformed and in addition to the above-described structure, we modelled by-participant random slope adjustments to the effects of Block, Day, and Exposure Condition.

Participants' RTs in Test 1 were significantly predicted by Exposure Condition (medium vs. low: β = 0.03, t = 7.40, p < .001; medium vs. high: β = -0.009, t = -1.99, p = .05), Block (β = -0.01, t = -2.36, p = .02), and Day (β = 0.13, t = 29.10, p < .001). Hence, faster RTs were elicited by medium vs. low and by high vs. medium exposure words, and RTs were faster in Block 2 than in Block 1, and faster on Day 1 than on Day 8. In addition, digit span (β = -0.02, t = -2.82, p = .005) predicted participants' RTs in the picture matching task. Individuals with higher digit span scores showed faster RTs. Furthermore, the interaction between Day of testing and the medium vs. high exposure contrast was significant (medium vs. high: β = -0.01, t = -2.50, p = .01), indicating the effect of this Exposure Condition contrast on RTs was stronger on Day 8 than on Day 1. Additionally, we found a significant interaction between Day and Block (β = -0.08, t = -18.72, p < .001), with the effect of block being stronger on Day 8 than on Day 1. The significant interaction between the medium vs. high exposure contrast and Raven's scores (β = 0.006, t = 2.13, p = .03) indicates that for individuals with high Raven's scores, the medium vs. high exposure contrast has a stronger effect than for individuals with lower Raven's scores. Finally, the interactions between Day and Raven's (β = 0.009, t = 3.77, p < .001) and between Day and vocabulary were significant (β = 0.006, t = 2.54, p = .01), showing that the effects of both Raven's and vocabulary scores were weaker on Day 8 than on Day 1 (see Appendix D for a plot of all coefficients).

Test 2: Explicit memory (picture naming)

The mixed-effects model on accuracy in the naming task included Block as a predictor, but we did not model any two-way interactions between Block and the individual differences measures due to convergence issues. All other two-way interactions between the fixed effects for Exposure, Block, and Day, and the four individual differences measures were included.

The accuracy analysis showed significant effects of the medium vs. low exposure contrast (medium vs. low: β = -0.49, z = -7.08, p < .001), Day (β = 0.79, z = 12.11, p < .001), and vocabulary (β = 0.23, z = 2.22, p = .03). Accuracy rates were higher for medium vs. low exposure words and participants gave a larger number of correct responses on Day 8 than on Day 1. In addition, individuals with higher vocabulary scores had higher accuracy rates than those with lower scores on the vocabulary tests.

The interaction between Day and Raven's scores (β = 0.12, z = 2.89, p =.004) was a significant predictor of naming accuracy (see Appendix E for a plot). Hence, Raven's scores had a stronger relationship with naming accuracy on Day 8 than on Day 1. The interaction between Raven's score and the exposure contrast medium vs. high was also significant (β = -0.15, z = -3.15, p = .002), showing that for high-exposure words the effect of Raven's scores was weaker than for medium-exposure words. In addition, the interaction between the medium vs. high exposure contrast and processing speed was marginally significant (β = 0.08, z = 1.78, p = .07), suggesting that for fast processors the exposure difference between medium and high-exposure words was weaker than for individuals with reduced processing speed. Finally, the interaction between Raven's and vocabulary scores (β = 0.21, z = 2.0, p = .05) was significant, indicating that the effect of vocabulary on naming accuracy increased with increasing Raven's scores. Participants with higher Raven's scores showed stronger vocabulary effects on naming accuracy than individuals with weaker Raven's scores (see Appendix E for plots of the interactions and a plot of all coefficients from the model).

In addition to the model structure stated above, the model on LDs included by-participant random slope adjustments to the effects of Day, Block, and Exposure Condition. One Exposure Condition contrast predicted LDs in the naming task (medium vs. low: (β = 0.15, t = 4.38, p < .001), with larger LDs for low as compared to medium exposure words. In addition, participants showed smaller LDs on Day 8 than on Day 1 (β = -0.19, t = -6.65, p < .001). No individual differences were found to predict LDs, although the coefficient plot suggests that there might be some trends towards beneficial effects of Raven's scores and processing speed on LDs in the naming task (see Appendix E).[7]

Discussion

This study was the first to explore individual differences in novel word learning in adult native speakers, and the relationship between word learning and variation in various cognitive abilities, namely vocabulary, processing speed, PSTM, and nonverbal intelligence. In the following, our findings concerning the relationships between the

[7]For one set of analyses, we excluded participants' responses where the LD value was equal to the number of letters in the target word, hence, the responses where in many cases participants did not say anything or where they responded with a completely different word from the target. This excluded 34.15% of the data. The results from this analysis showed no Exposure Condition effects, potentially because many of the low-exposure words were excluded, and only Day (β = -0.20, t = -3.51, p < .001) and vocabulary (β = -0.11, t = -2.30, p =.02) were significant main effects. Hence, the analysis was very similar to the accuracy analysis and was therefore considered to provide no insights beyond what the above-described accuracy and LD analysis did (see Appendix E for a coefficient plot for this analysis).

various cognitive abilities tested in the present study are discussed. This is followed by a discussion of the insights gained into the relationship between individual differences in word learning and variation in cognitive as well as environmental factors.

First of all, correlation analyses between participants' Raven's scores, digit span scores, and the factor scores for vocabulary and processing speed showed no significant correlations between any of the individual differences measures. There were only trends towards weak correlations between digit span performance and vocabulary, and between Raven's score and processing speed. It might be surprising and somewhat counterintuitive that we did not observe any correlations between, for instance, Raven's scores and vocabulary or Raven's scores and general processing speed (Kail & Salthouse, 1994). It is important to note that the tests we used to assess each of the cognitive abilities were standard measures and the sample was relatively large with more than 100 participants. In addition, we tested participants from a population that is frequently used as a representative sample within our field. It is thus interesting that the cognitive factors do not correlate with each other, indicating that little variation in vocabulary size is explained by variation in non-verbal intelligence, general processing speed, or PSTM. This indicates that the skills assessed in these individual differences measures are distinct and independent constructs, at least in the group of participants tested in this study, namely young adult university students.

It has to be noted that in the present study we used the abbreviated version of the Raven's Matrices due to time constraints on the test sessions. In future research it might be worth further investigating the relationship between various cognitive abilities (and word learning) by using more than just one, or more comprehensive measure, to assess the presumably complex skill of nonverbal intelligence. In addition, testing participants with more varied educational backgrounds might also have an effect on the relationships observed between the various measures of cognitive ability. Increasing the diversity of participants' educational backgrounds would presumably lead to a larger range of vocabulary, digit span, and Raven's scores and result in more diverse general processing speed abilities being observed. If there are correlations between the various skills they might only show when the range of test performances is increased. In addition, the results would be more representative of the wider population if a more heterogeneous group was examined. As indicated above, this is a prevalent issue in psychological and psycholinguistic research and does not only apply to the present study (Kidd et al., 2018; see Chapter 3).

The main interest of the present study was the relationship between variation in word learning performance and individual differences in cognitive abilities (vocabulary, processing speed, digit span, and nonverbal intelligence). We analysed participants' performance on three different tasks from a word learning experiment, namely the 4AFC task in Training 2 (picture matching with feedback), the 4AFC task in Test 1 (picture matching), and the picture naming task in Test 2.

The analyses of participants' performance on 4AFC training task (Training 2) showed that accuracy rates during training increased with increasing numbers of repetition, as would be expected. In addition, our analyses showed that individuals with higher vocabulary scores showed overall higher accuracy rates in Training 2. This is in line with previous research showing that having greater vocabulary knowledge is beneficial for learning novel words (Henderson & James, 2018; Marchman & Fernald, 2008). Furthermore, participants with higher Raven's scores showed stronger benefits from having greater vocabulary knowledge, and weaker Raven's scores were associated with stronger digit span effects. However, it has to be noted that these findings are based on linear regression analyses only, thus not accounting for random variation between participants and items, because performance close to ceiling led to convergence issues with the mixed-effects models. Therefore, the relationship between task performance and individual differences measures is difficult to interpret and the discussion of the results focuses on the analyses of the two test tasks.

For Test 1, the 4AFC picture matching test, we analysed participants' accuracy rates as well as their RTs. Accuracy was close to ceiling on both days of testing, with around 97% on Day 1 and 95% on Day 8, indicating that this multiple-choice test with four answer alternatives was very easy for our participants to perform. Due to the overall very high accuracy and therefore rather limited variation, we considered RTs to be a more appropriate and more informative measure of performance on this task.

The RT analyses indicated that participants performed better in Block 2 than in Block 1, which is probably a practice effect. Both blocks comprised all 39 words so that participants responded to all words for the second time when completing Block 2. Similarly, the fact that participants showed faster RTs on Day 1 than on Day 8 is very likely a practice effect. On Day 1, participants completed a very extensive training task (Training 2) before doing this test, which was almost identical with Test 1 in that both were 4AFC picture matching tasks. On the contrary, on Day 8 participants came to the lab and started with the 4AFC test, without any preceding training on a similar task or the items. This has probably resulted in participants' RTs being considerably faster on Day 1 than on Day 8. Furthermore, the difference between blocks was stronger on Day 8 than on Day 1, suggesting that the practice effect of Block was stronger on Day 8. This is due to the fact that, as indicated before, the only practice that participants had was the completion of Block 1 prior to Block 2. Finally, our manipulation of number of exposures was reflected in how fast participants were at responding to the picture-matching task, both immediately after training and after a period allowing for overnight consolidation. Participants responded faster to words that they had seen more often during training; hence, responses to high-exposure words were significantly faster than to medium- and low-exposure words. For the medium- vs. high-exposure contrast, this effect was even stronger on Day 8 than on Day 1, despite the fact that RTs on Day 8 were overall longer. This suggests that overnight consolidation of the novel words strengthened the differences in speed of lexical access between words from the three exposure conditions, which were already observed right after training. Although on a very small scale, the repetition effect observed in lexical access to newly learned form-meaning pairings reflects word frequency effects observed in lexical access to known words (Brysbaert et al., 2018), and suggests a role of exposure frequency in producing word frequency effects during language processing.

Furthermore, we observed a significant interaction between Exposure Condition and Raven's scores, showing that individuals with higher Raven's scores were more sensitive to the medium vs. high exposure contrast. Their RTs were particularly enhanced by additional exposures. In addition, the interactions between Day and Raven's as well as between Day and vocabulary were significant, indicating that the beneficial effects of Raven's and vocabulary on picture matching RTs were stronger on Day 1 than on Day 8. This means that better vocabulary knowledge and higher Raven's scores were particularly beneficial for the formation of and access to short-term memories for the newly learned lexical items in this specific task and were not so beneficial for the consolidation of these representations.

The only individual differences measure that had a main effect on participants' RTs on the 4AFC picture-matching test was digit span score. Individuals with higher digit span scores, taken to indicate better PSTM abilities, showed significantly faster RTs on both days and in all conditions. Having better PSTM abilities might enable individuals to form stronger phonological representations or stronger associations between phonology and semantics, which might be faster to be accessed during the picture matching test. Thus, our findings may indicate a role of PSTM in novel word learning as has been reported in previous research (Gathercole, 2006; Gupta & Tisdale, 2009b). Alternatively, the relationship between PSTM and performance on the 4AFC picture matching task might be due to similarities between the two tasks. Both require the processing of an auditory stimulus and keeping this stimulus in short-term memory in order to perform a specific task on it. The fact that PSTM was only related to participants' RTs on the 4AFC test, and not to any measure of performance on the naming task, is taken to support this alternative idea. If PSTM is assumed to influence native language novel word learning in adults, the naming task especially would be expected to require PSTM and show effects of individual differences in PSTM. The reason is that the ability to store and reproduce the phonological form of a word, for which PSTM is presumably essential, is assumed to be central to performance on the naming task (Gathercole, 2006). In addition, it has to be noted that the RTs in this multiple-choice task might reflect word learning success as has been supported by the existence of exposure condition effects, but they may also influenced by task-specific behaviour. Due to the fact that it was a multiple-choice test, strategic behaviour such as answering by process of elimination might have influenced participants' RTs. In addition and as mentioned before, the interpretability of the relationship between the individual differences measures and performance on the 4AFC picture matching task is limited due to participants' high abilities, as indicated by their accuracy being close to ceiling. That is why the findings from this task should be interpreted more cautiously.

Performance on the picture naming test might be less biased by strategic effects, which is why we believe that naming performance is probably the most appropriate test measure of word learning performance. What is more, for the present population, which is characterised by overall high abilities, the more difficult naming task is more suitable for the elicitation of individual differences (see Kidd et al., 2018) and therefore seems to be a more appropriate measure to examine the questions of interest. Thus, relatively high abilities in the group of participants can be compensated for by using or focusing on a more difficult task, in this case the picture naming task.

In the picture naming task (Test 2) we measured and analysed naming accuracy and Levenshtein Distances (LDs) between participants' responses and the target words. We did not find any significant relationships between individual differences measures and LDs, although the coefficient plot (see Appendix E) suggests that there might be trends towards beneficial effects of Raven's scores and processing speed. This might be due to a lack of variation in the LDs as dependent measure, or due to LDs not being an appropriate measure for word knowledge in this task. Therefore, we focus on the findings from the accuracy analyses. We found significant effects of day of testing on naming accuracy, with improved performance on Day 8 as compared to Day 1. This is the typical effect of testing the knowledge of newly learned words after a delay allowing for overnight consolidation and this replicates studies that have shown beneficial effects of overnight consolidation on memory for newly learned words (Bakker et al., 2015; Brown & Gaskell, 2014; Weighall et al., 2017). It has been suggested that sleep aids the lexicalisation of newly learned words, thus consolidating the representations of these words and integrating them within existing lexical knowledge, which leads to increased accuracy when tested on novel words after a delay (Takashima et al., 2014; Weighall et al., 2017). The present study shows the effects, confirming that the participants in our study in fact learned the form-picture pairings as was expected. The fact that this typical effect of sleep was observed on naming accuracy (as opposed to the picture matching test) supports the impression that this was an appropriate and reliable measure of word learning success, which is – as indicated above – presumably less influenced by the effects of practice and strategic behaviour that have been argued to potentially play a role in the 4AFC picture matching (Test 1) task.

In addition, we observed a significant effect of the medium vs. low contrast of Exposure Condition. As was expected, higher numbers of exposure, hence increased training on the novel form-meaning pairings, resulted in improved test performance. Thus, just as in the analyses of individuals' performance on the 4AFC picture matching test, the manipulation of number of exposures had an effect on word learning performance. As indicated before, this might reflect word frequency effects observed in language processing of known words, albeit on a smaller scale (Brysbaert et al., 2018). The fact that here only the medium vs. low contrast was significant indicates that the difference in amount of exposure between the medium (9 exposures) and high condition (16 exposures) was not big enough to elicit an effect of Exposure Condition on picture naming performance.

Furthermore, we observed some relationships between the individual differences measures and performance on the naming task. We found a significant effect of participants' Raven's scores on naming performance on Day 8, but not on Day 1. After a period allowing for overnight consolidation, higher scores on the Raven's Advanced Progressive Matrices were associated with higher accuracy scores in the naming task. Thus, immediately after training, i.e. on the formation of episodic memory representations for the novel words, nonverbal intelligence as measured by this test did not show any effect, but it is indicated to be beneficial for the consolidation of novel word representations. Higher nonverbal intelligence is suggested to benefit the consolidation of novel word representations and potentially their integration into the lexicon. The relationship between a cognitive ability and consolidation warrants further investigation to determine the underlying mechanisms.

Additionally, the significant interaction between the medium vs. high exposure contrast and Raven's scores indicated that for high-exposure words the effect of Raven's score was weaker than for medium-exposure words. This suggests that higher nonverbal intelligence, as measured by the Raven's, is less beneficial for easy words that are presented very often than for slightly more difficult items.

Additionally, our findings suggest smaller effects of quantity of exposure (medium vs. high contrast) for individuals with higher general processing abilities than for those who were slower on the general processing speed tasks. It has to be noted that this effect was only marginally significant. Nevertheless, it appears that there is a tendency that high processing speed individuals are less sensitive to constraints on the number of exposures to novel words during learning. Hence, higher processing speed might be beneficial in that it might enable individuals to form relatively strong memory representations of novel words already after only one or two exposures to that word, and these representations may be accessed equally well as representations of words with higher frequencies of exposure. A similar relationship between online processing speed and vocabulary has been observed in children where processing speed at 25 months has been shown to predict language and cognitive skills at 8 years old (Fernald et al., 2006; Marchman & Fernald, 2008; McMurray et al., 2012). It has been suggested that children who are faster processors might have more time or resources to process subsequent speech input allowing for more opportunities for learning. Alternatively, children with increased processing speed may have more resources or time to process secondary aspects of the input which may help them to build richer or more robust representations at first exposure. From our data, we cannot conclude what the mechanism behind a potential processing speed advantage in word learning might be. It is conceivable though that having more resources or time for detailed processing of the visual input (i.e. the meaning) in relation to the auditory forms of the novel words, thus being able to form richer representations at first exposure, may have been the origin of the speed advantage in the present experiment. However, more research on this relationship is clearly needed, potentially testing individuals from more diverse educational backgrounds to further examine the relationship between processing speed and quantity of exposure in word learning, which in our study is suggested by a marginally significant interaction between Exposure Condition and processing speed.

Aside from this, vocabulary was the only individual differences measure that had a main effect on naming accuracy on both days of testing and in all conditions. Individuals with greater vocabulary knowledge showed higher accuracy rates in the naming test. Hence, having more word knowledge appears to be beneficial overall for word learning, i.e. for the formation of and access to episodic representations of novel words immediately after training and also for the consolidation and retrieval of lexical items at a later point in time. This suggests that individuals with larger vocabularies are better word learners due to mechanisms that are distinct from processing speed, PSTM, and nonverbal intelligence. Our finding is in line with previous findings on the role of vocabulary knowledge in word learning in children (Fernald et al., 2006; Henderson & James, 2018; Marchman & Fernald, 2008; Weisleder & Fernald, 2013). Hence, having greater word knowledge appears to be beneficial for word learning, not only in children but also in adults. What is more, this vocabulary benefit was even stronger for individuals with higher Raven's scores. This indicates that those whose performance on the Raven's Matrices was better were able to use their existing knowledge more efficiently to aid novel word learning compared to individuals with lower Raven's scores.

Therefore, next to nonverbal intelligence as measured by the Raven's Matrices, vocabulary appears to be a good predictor of word learning performance as measured by the picture naming task. Thus, despite the fact that PSTM, general processing speed, and performance on the Raven's Matrices were taken into account as perhaps affecting word learning performance and potentially being captured by measures of vocabulary size, vocabulary turned out to be an important and stronger predictor of word learning. One may argue that it is not very surprising that individuals with larger vocabularies show advantages in this word learning task. These individuals are probably better word learners, which enabled them to acquire more word knowledge and outperform others on the vocabulary tests. The question arising is what the mechanism is through which greater knowledge of words benefits novel word learning. As indicated before, something must have led to some individuals having a larger vocabulary in the first place, which is then beneficial for all subsequent word learning. It might be argued that those children who get quantitatively and qualitatively more exposure early in development learn more words, thus getting more practice with speech input allowing them to gain more sublexical and lexical knowledge (Jones & Rowland, 2017). This advantage for novel word learning may remain across the lifespan. However, this explanation leaves the question about the underlying mechanism open.

One possibility may be that individuals with greater vocabularies are generally better learners. If this was the case, the vocabulary benefit should also apply to non-linguistic learning tasks. Alternatively, the participants with higher vocabulary scores may be better language learners due to mechanisms other than processing speed, PSTM, or nonverbal intelligence, which were assessed in the present study. As indicated above, a measure of vocabulary size likely captures differences in the amount of exposure, hence prior experience with language, which in turn might be beneficial for word learning for various reasons. One might be that greater input or experience with language improves certain language learning mechanisms. It is, however, unclear what those might be. Individuals with more experience (and therefore greater vocabularies) may be better at encoding novel word forms or the mappings between form and meaning, possibly due to more lexical and sublexical knowledge. Alternatively, there might be structural or representational differences between vocabularies of varying sizes leading to greater vocabulary knowledge being beneficial for novel word learning. Previous research has indicated that characteristics of a learner's vocabulary, for instance phonological or semantic knowledge, affect the ease with which novel words are learned. Words from denser phonological neighbourhoods have been shown to be easier to learn than words from less dense neighbourhoods (Storkel, Bontempo, & Pak, 2014). In addition, it has been demonstrated that it is easier to integrate novel semantic representations within denser semantic networks than within less dense networks (Sailor, 2013). Learning new words that possess similar semantic concepts as known words is easier. Hence, differences in the density of the phonological or semantic networks that potentially result from differences in vocabulary size might be a cause of the word learning advantage associated with greater vocabulary knowledge.

Future research might use computational modelling to examine the underlying mechanisms associated with the vocabulary benefit in novel word learning. In such a model, it would be possible to investigate the contributions of factors such as the quantity and quality of exposure, general intelligence, or processing speed while keeping vocabulary size constant. In this way, the effects of vocabulary size on word learning and closely related factors such as variation in exposure, processing speed, or capacity could be disentangled. In addition, computational modelling would allow us to study both developing and mature systems and examine whether different factors are important at different points during development and learning.

Furthermore, future studies should examine the relationship between word learning and individual differences in various cognitive abilities in more diverse groups of participants. As mentioned before, in a sample of undergraduate university students the range of test scores on the measures of cognitive ability is presumably limited as compared to the general population (see Chapter 3). A change in the range of abilities in the sample tested might result in different conclusions about both the relationships between the cognitive abilities and between these abilities and word learning.

Conclusions

This exploratory study of novel word learning was the first to examine the role of individual differences in various cognitive abilities and word learning in adult native speakers. The results indicate that nonverbal intelligence as measured by the Raven's Advanced Progressive Matrices is beneficial for the consolidation of novel words. Furthermore, it is suggested that participants with higher Raven's scores were able to use their existing vocabulary knowledge more efficiently to aid novel word learning than individuals with weaker Raven's scores. Finally, vocabulary knowledge has been shown to be an important predictor of novel word learning ability. Knowing larger numbers of words provided a significant advantage when being faced with the task of learning new words. Our study provides first insights into the complex relationships between quantity of exposure, delay of testing (here confounded with presence/absence of sleep), and internal sources for variation in adults. Future research using, for instance, computational modelling techniques should further examine what the underlying mechanisms are that lead to the beneficial effects of larger vocabularies on novel word learning.

In addition, some methodological implications can be derived from the current study. The high accuracy rates on the picture matching test as compared to the picture naming test and the differences in the relationship between the respective task performance measures and participants' cognitive abilities demonstrate that for the presumably highly skilled population tested in this study more difficult tests eliciting greater variation in performance are more suitable. Furthermore, the fact that both word learning and cognitive skills are probably rather high in these young adults supports the assumption that future research should strive for testing individuals from more diverse educational backgrounds.

Acknowledgements

We would like to thank Sanne van Eck and Monique de Nijs for testing part of the participants, and Mirte Dikmans, Sanne van Eck, and Annelies van Wijngaarden for coding the data.

Appendix A

Words used in the novel word learning experiment

1. Spalier
2. Oribi
3. Wimberg
4. Firn
5. Ojief
6. Amsoi
7. Gerenoek
8. Gnomon
9. Napi
10. Blaffetuur
11. Golok
12. Pinang
13. Knippa
14. Liaspen
15. Deuvik
16. Margay
17. Gamander
18. Tsuba
19. Alikas
20. Tussor
21. Amict
22. Bendir
23. Genoffel
24. Fijfel
25. Paletot
26. Putto
27. Vair
28. Winket
29. Blimbing
30. Averuit
31. Euzie
32. Twatwa
33. Sifaka
34. Zurkel
35. Dolik
36. Kossem
37. Zaathout
38. Zwad
39. Angisa

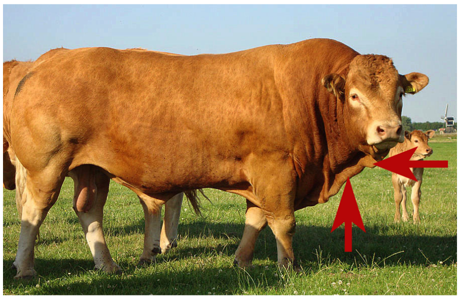

Figure 5.2: Kossem

Figure 5.4: Oribi

Figure 5.3: Winket

Table 5.7: **Appendix B:** Correlation table displaying the bivariate correlation coefficients and p-values for all individual differences measures.

	Visual speed	**Auditory speed**	**Letter comparison**	**Antonym test**	**Andringa**	**Digit span**
Auditory speed	.61**					
Letter comparison	.15	-.05				
Antonym test	-.02	-.10	.03			
Andringa	-.13	-.17 †	-.15	.37**		
Digit span	.02	-.002	-.22*	.13	.15	
Raven's	.14	.18	-.16	.07	.17 †	.01

† $p = .08$
$p < .05$
** $p < .001$

Appendix C

Training 2 (picture matching)

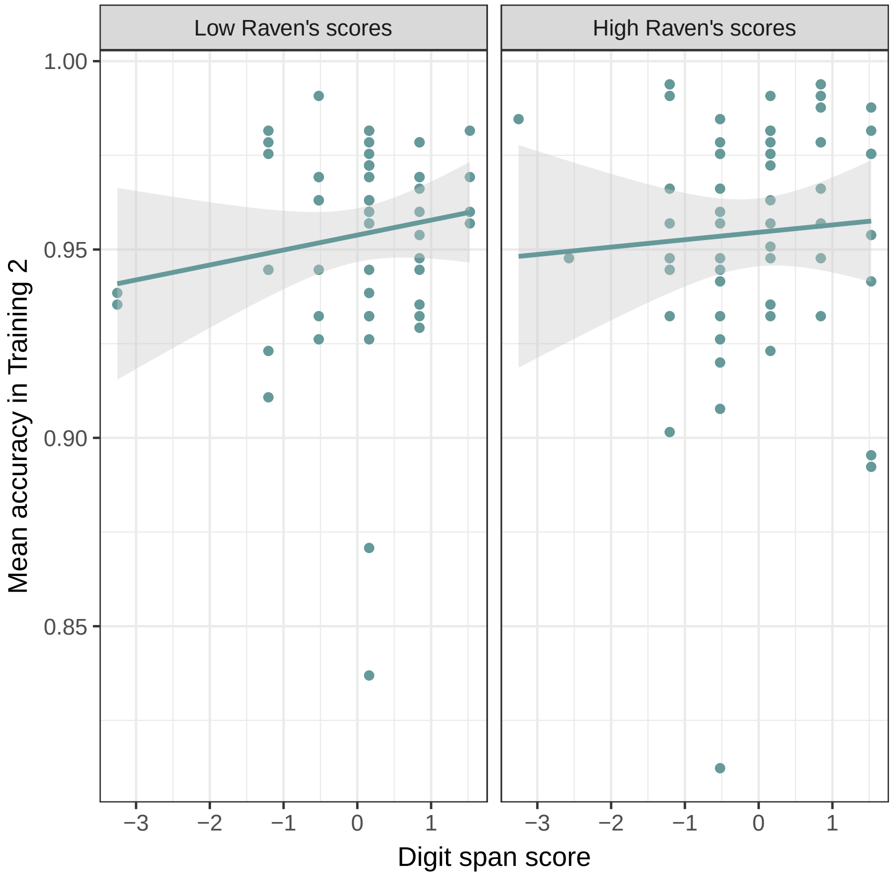

Figure 5.5: Correlation between accuracy rates in Training 2 (4AFC) and digit span scores for individuals with low Raven's scores vs. individuals with high Raven's scores. Raven's scores were z-transformed and scores below 0 were grouped together as low scores, scores above 0 as high Raven's scores.

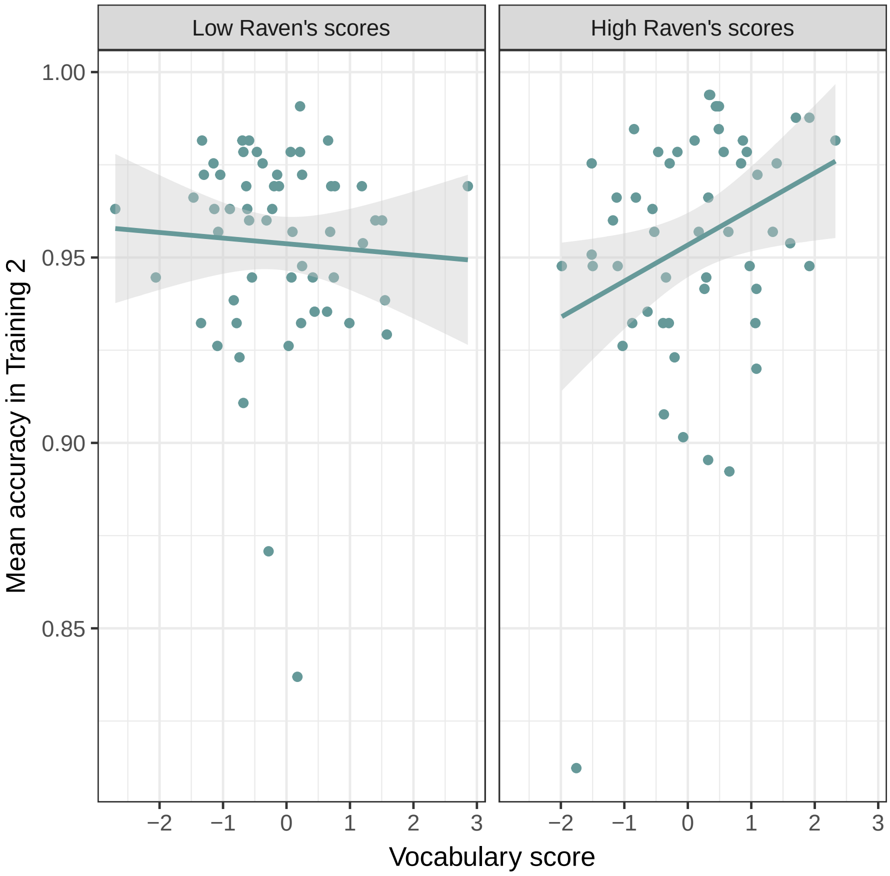

Figure 5.6: Correlation between accuracy rates in Training 2 (4AFC) and vocabulary scores for individuals with low Raven's scores vs. individuals with high Raven's scores. Again, z-transformed Raven's scores below 0 are low scores and those above 0 are high scores.

Appendix D

Associative memory Test 1 (picture matching)

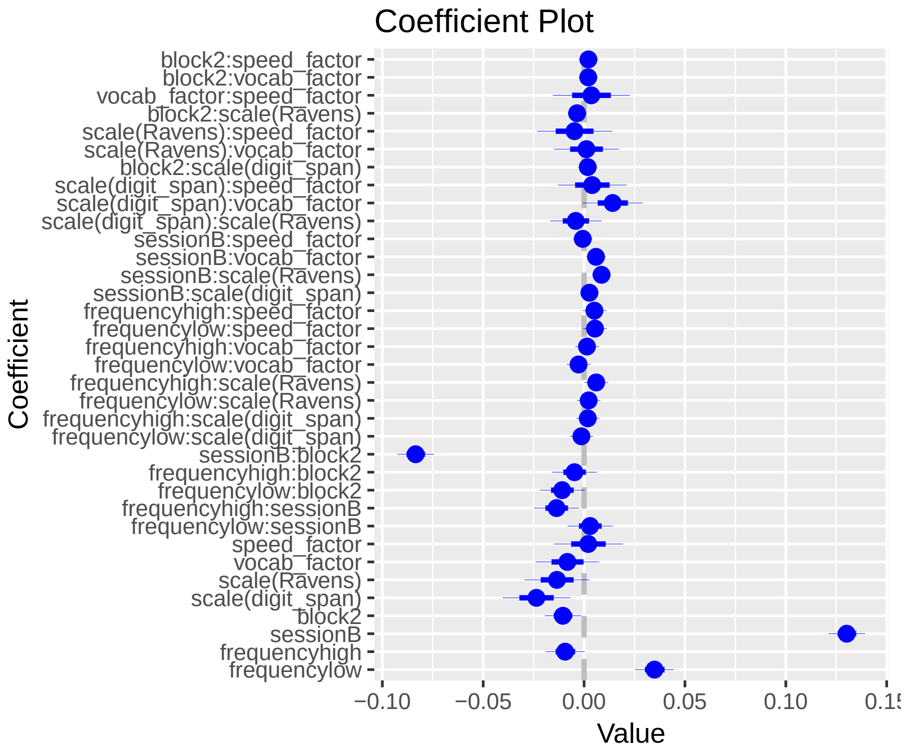

Figure 5.7: Coefficient plot for the mixed-effects model analysing reaction times in the associative memory Test 1 (picture matching). The estimate for the intercept in this model was 3.29. The coefficient plot does not include the intercept because including this large estimate (much larger than the remaining estimates) would have resulted in the differences between all other estimates being invisible.

Appendix E

Explicit memory Test 2 (picture naming)

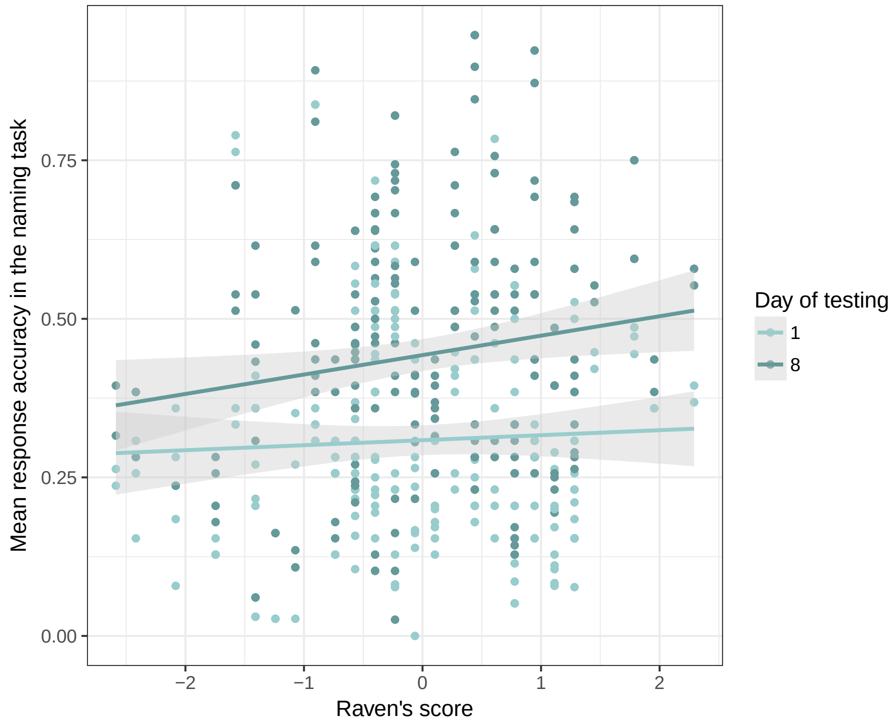

Figure 5.8: Correlation between accuracy rates the naming task and z-transformed Raven's scores for the test on Day 1 vs. Day 8.

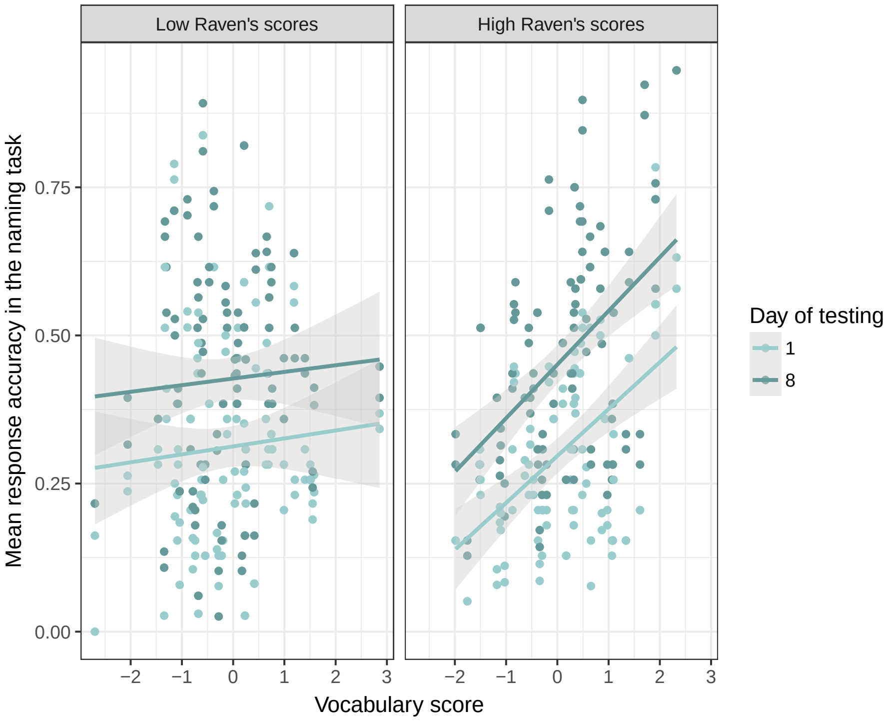

Figure 5.9: Correlation between accuracy rates in the naming test (both Day 1 and Day 8) and the factor score for vocabulary, for individuals with low Raven's scores vs. individuals with high Raven's scores. Raven's scores were z-transformed and scores below 0 were grouped together as low scores, scores above 0 as high scores.

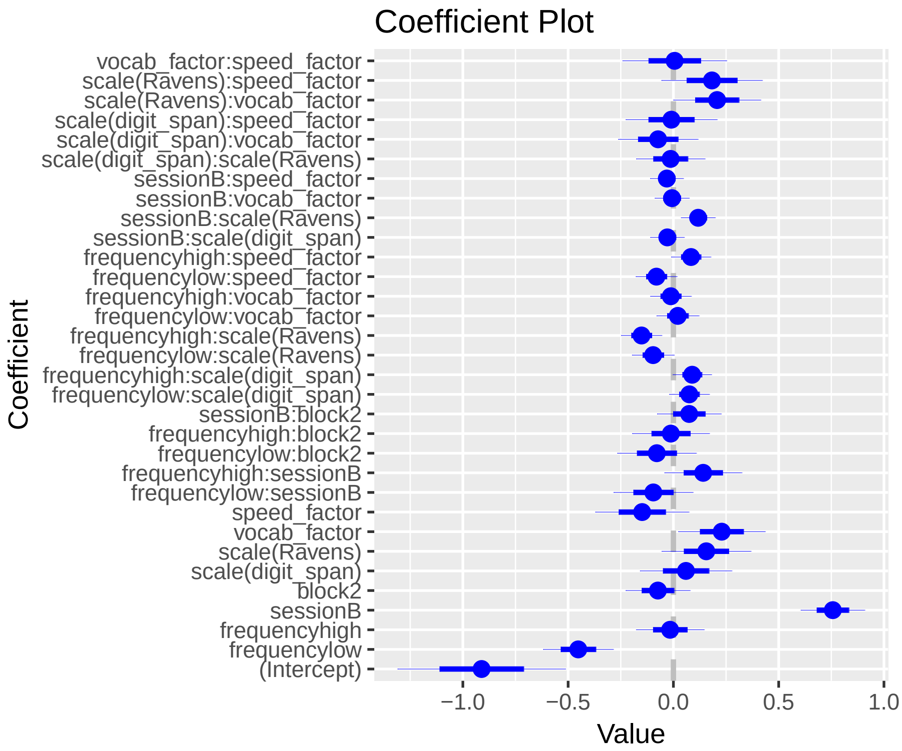

Figure 5.10: Coefficient plot for the mixed-effects model analysing response accuracy in the explicit memory Test 2 (picture naming).

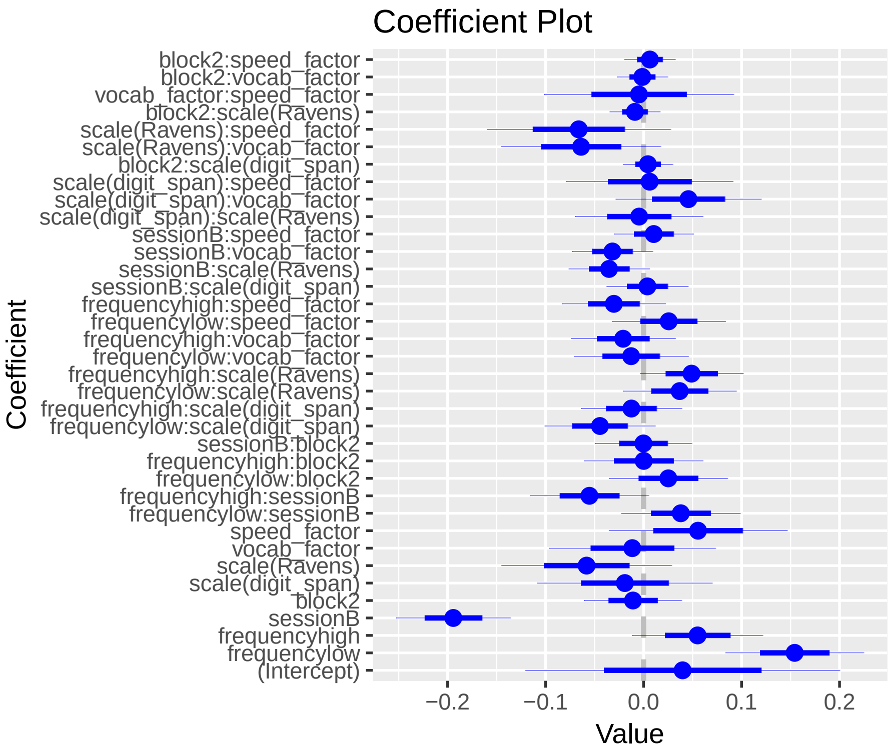

Figure 5.11: Coefficient plot for the mixed-effects model analysing Levenshtein Distances in the explicit memory Test 2 (picture naming).

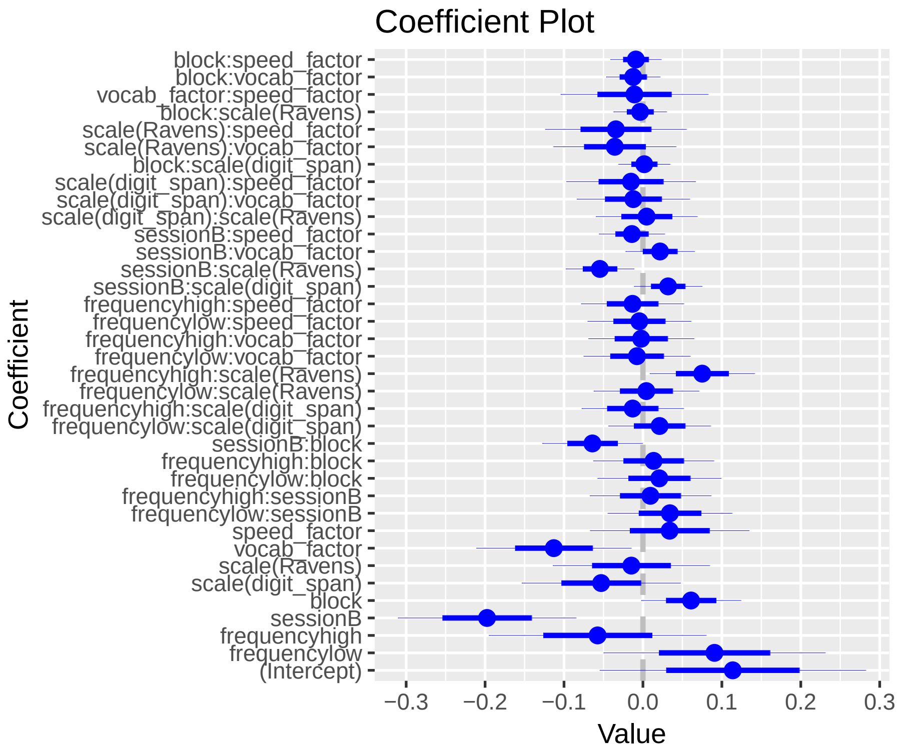

Figure 5.12: Coefficient plot for the mixed-effects model analysing Levenshtein Distances only for the words where the distance was smaller than the number of letters in the target.

6 | The causes and consequences of variation in vocabulary size: A computational model

Abstract

Individuals vary considerably in their vocabulary size. These individual differences have been shown to predict performance on a broad range of language related tasks, with larger vocabularies being associated with beneficial effects on language processing. Furthermore, previous research has suggested a number of potential environmental and cognitive causes for observed differences in vocabulary learning and resulting vocabulary size, which in turn likely have additional consequences for language processing. The present study used a connectionist computational model that learned phonological and semantic representations of English words and the mappings between these representations (Harm & Seidenberg, 2004) in order to simulate the learning and processing of spoken English words. We manipulated both internal properties of the model's processing (processing speed, processing resources) as well as external properties of the model's linguistic environment (quality and quantity of language exposure) in order to establish first their effects on vocabulary development, and secondly their consequences for language processing. Our simulations demonstrate that a complex interaction of cognitive and environmental factors is likely to cause variation in vocabulary size. In addition, these causes for variation in vocabulary size, i.e. processing speed and resources as well as differences in the quantity and quality of exposure, all had similar positive consequences for the model's language processing and novel word learning performance. Our analyses do not provide strong evidence for added beneficial effects of greater vocabulary size which go beyond those that can be attributed to underlying cognitive and environmental factors. Furthermore, we use the model to describe explicitly the mechanisms that likely underlie the observed beneficial effects.

Introduction

Previous research has shown that the speakers of a language vary considerably in their vocabularies. Age as well as educational background have been found to be important predictors of the number of words an individual knows (Brysbaert et al., 2016b, see Chapters 2 and 3). Furthermore, individual differences in vocabulary size have been associated with variation in language processing performance. Individuals with larger vocabularies have been reported to show more accurate and faster language comprehension (Brysbaert et al., 2016a; Diependaele et al., 2013; Yap et al., 2012, see Chapters 2 and 3) and production (Rodriguez-Aranda & Jakobsen, 2011, see Chapter 4). In addition, speakers with better vocabulary knowledge show reduced effects of word frequency as compared to individuals with smaller vocabularies (Brysbaert et al., 2016a; Diependaele et al., 2013; Monaghan et al., 2017).[1] High-vocabulary individuals' word recognition speed is less influenced by the frequency of the target words, thus, lexical processing is similarly fast for low- and high-frequency words. This has been taken to indicate that the lexical representations in individuals with greater vocabulary knowledge are more entrenched or more robust, and therefore faster to be accessed and less sensitive to effects of lexical characteristics (*lexical entrenchment hypothesis;* Diependaele et al., 2013). Monaghan et al. (2017) have examined the effect of variation in the amount of exposure on the size of the word frequency effect and on vocabulary size in a computational model. Greater exposure has been demonstrated to cause a reduction in frequency effects on processing as well as to cause greater vocabulary size. The model was also trained to learn a second language, which replicated effects of second language proficiency on frequency effects in first and second languages. Increased second language proficiency was associated with reduced frequency effects in the second language, whereas the frequency effects in the first language increased. It has been argued that variation in exposure alone therefore can explain differences in vocabulary size and in the sensitivity to word frequency effects, in both first and second languages (Monaghan et al., 2017).

However, is variation in the amount of exposure the only cause for variation in vocabulary size? Previous research has provided some insights into other factors that might play a role in determining an individual's word learning performance and, consequently, vocabulary size. Mostly developmental studies have indicated that individual differences in both environmental and cognitive factors drive variation in word learning.[2] Behavioural and modelling investigations have demonstrated that input quantity, i.e. the number of words in the input, benefits word learning very early in

[1] See Chapters 2 and 3 for a more detailed literature review.

[2] See Chapter 5 for a more detailed review and discussion of the literature.

development, while the quality, i.e. diversity or number of different words, within the input plays a role in word learning later in development (Jones & Rowland, 2017, see also Hurtado et al., 2008). In addition, a considerable body of research has focused on the relationship between phonological short-term memory (PSTM) and vocabulary learning and size, indicating that better PSTM abilities are associated with improved word learning and greater word knowledge (Gathercole, 2006; Gupta & Tisdale, 2009b). Another cognitive ability that has been associated with variation in word learning is processing speed. Both behavioural and computational studies have demonstrated that higher (language) processing speed in children predicts better word learning and greater vocabulary size (Fernald et al., 2006; Marchman & Fernald, 2008). Two potential reasons for this close relationship between vocabulary size and (language) processing speed have been suggested. Firstly, richer language input in early development might benefit word learning and vocabulary size in some children. As a result, their practice with linguistic input is greater as compared to children with quantitatively and qualitatively less input, which in turn might lead to more efficient and thus faster speech processing (Fernald et al., 2006). Alternatively, Fernald and colleagues (2006) proposed that initial individual differences in general cognitive skills, more precisely general processing skills, might be the origin of variation in language processing and word learning performance. Both suggestions are compatible with the findings from Fernald et al. (2006) and importantly, they are not mutually exclusive. It is conceivable that rather complex interactions between cognitive abilities and external factors lead to variation in word learning and processing.

Finally, another factor that has been indicated to predict word learning is vocabulary size. Children with larger vocabularies at 2 years of age showed a greater learning rate over the following 6 years of life, thus had larger vocabularies at 8 years of age, than children with less word knowledge earlier in development (Marchman & Fernald, 2008). It is possible that such a relationship between vocabulary size and learning rate is driven by shared underlying causes; it is also possible however that possessing a larger vocabulary offers distinct advantages that generate the observed increase in rate of word learning. Evidence supporting such a mechanism can be found in Steyvers and Tenenbaum (2005) whose preferential attachment theory proposes that the probability of a new item being acquired is positively related to the number of connections already within the network. Thus, a network that already has many densely connected nodes is more likely to acquire new nodes. Network properties generated by such a mechanism have been shown to mirror the properties observed within the natural emergent structure of human semantic networks (Sailor, 2013).

The study presented in Chapter 5 of this dissertation built upon this earlier work and extended it by examining young adult native speakers of Dutch instead of children. In addition, we were interested in the relationship between different cognitive abilities and the relationship between cognitive and environmental factors in affecting word learning performance. Knowing of the existence of considerable individual variation in cognitive skills and their effects on language processing, we assumed that there are not only external, exposure-related sources of variation, i.e. environmental factors, but also internal sources for variation in word learning and knowledge (Kidd et al., 2018). Most importantly, the experiment showed that vocabulary size predicts novel word learning performance in addition to the factors amount of exposure and sleep and associated overnight consolidation (see Chapter 5). Furthermore, nonverbal intelligence and general processing speed were suggested to potentially also influence word learning ability, though not as strongly as vocabulary.

To sum up, larger vocabularies have been associated with more accurate and faster language processing. Furthermore, it has been shown that individuals with larger vocabularies are less sensitive to word frequency effects, which has been taken to indicate that their representations are more entrenched or robust than those of individuals with less word knowledge. Monaghan et al. (2017) have demonstrated that variation in exposure alone offers an explanation for differences in the sensitivity to frequency effects as well as differences in vocabulary size, without taking individual variation in cognitive skills into account (Monaghan et al., 2017). However, we know that there are considerable individual differences in cognitive skills and that they affect not only language processing but also learning (Kidd et al., 2018). Aside from differences in the quantity and quality of input (Jones & Rowland, 2017), variation in PSTM, processing speed, general intelligence and vocabulary size have been associated with individual differences in word learning (Fernald et al., 2006; Gupta & Tisdale, 2009b; Kidd et al., 2018; Marchman & Fernald, 2008; McMurray et al., 2012). In addition, the experiment presented in Chapter 5 was in line with these findings from earlier research, indicating beneficial effects of increased vocabulary size on novel word learning. Furthermore, nonverbal intelligence was suggested to potentially also play a role in word learning.

A model that describes variation in lexical processing and learning, therefore, needs to account for such individual differences. The present computational modelling study aims to do so by simulating variation in cognitive abilities, namely processing speed and intelligence, as well as quality and quantity of exposure to linguistic input. We examined the potentially interactive effects of these factors on vocabulary size as well as processing. Underlying this is the assumption that learners who received the same amount of exposure might show differences in vocabulary size, lexical processing, and the sensitivity to word frequency effects due to internal or cognitive sources of variation in learning and processing.

Another important and related question that we addressed with the present study is why vocabulary size is such a good predictor of word learning (see e.g. Fernald et al., 2006, Chapter 5). Individuals with large vocabularies are probably very proficient word learners and this is likely to be reflected in a novel word learning study or across development. This might seem trivial at first but the interesting question is what mechanisms underlie such variation in performance, what makes some individuals good word learners in the first place. Is it just early exposure and variation therein that leads to differences in vocabulary size, which in turn affect all subsequent word learning? Alternatively, and more likely, more complex interactions between cognitive skills and environmental factors are at play, causing variation in word learning and vocabulary size, which persists beyond childhood and thus also affects language processing skills across adulthood. The latter would imply that vocabulary size is such a good predictor of word learning because it captures a variety of underlying factors, both environmental and cognitive. Further, computational and behavioural studies suggest that additional distinct benefits may result from increased vocabulary size (Steyvers & Tenenbaum, 2005; Sailor, 2013). Thus in addition to cognitive and environmental causal factors, mechanisms dependent on increased vocabulary size might have beneficial consequences for language learning and processing.

The present computational modelling study aimed to examine the causes and consequences of variation in vocabulary learning and size. In a parallel distributed processing model that learns semantic and phonological representations of spoken English words and mappings between these representations, we simulated individual differences in general processing speed and processing resources as well as in quality and quantity of exposure to linguistic input. Within this model, we tested the role of each factor independently and their interactions in causing variation in vocabulary size. We then examined their broader consequences for language processing, specifically effects on the structure of the emergent lexicon (profile of words acquired and frequency effects), speed of lexical processing, and rates of novel word learning. The purpose of the model was to explicitly describe the mechanisms that may underlie the factors suggested to cause variation in vocabulary size and their consequences for language processing. We believe it is only through such modelling exercises, in combination with behavioural studies, that we can isolate the effects of individual factors and gain insights into the complex interactions between them.

The computational model used in this study was developed from the Parallel Distributed Processing (PDP) framework (McClelland, Rumelhart, & the PDP

Research Group, 1986; Rogers & McClelland, 2014). Such models have proved successful in modelling a broad range of cognitive and developmental phenomena (see Rogers & McClelland, 2014 for review). These models, sometimes referred to as connectionist neural networks, consist of multiple layers of non-linear processing units each connected via weighted connections, which together are able to learn complex mappings between distributed representations. Such networks are able to learn through applying error-based learning algorithms (e.g. recurrent backpropagation, Rumelhart, Hinton, & Williams, 1986) that incrementally make small adjustments to the connection weights within the network in order for the network's behaviour to adapt to dynamic constraints imposed by the learning environment. This leads to the network's representational structure, processing, and behaviour being emergent properties of the interaction between environmental constraints and the network's computational architecture. Such a framework is thus well suited to the goals of the current project. It allows the implementation of variation in initial internal processing constraints (thus offering a means to model differences in initial cognitive capacities) in addition to variation in properties of the learning environment (e.g. quality and quantity of linguistic input). Using distributed phonological and semantic representations, the model is also able to capture the complex relationships that exist within and between representational domains that interact to influence learning and processing. We are thus able to implement limited assumptions regarding the structure of information available in the learning environment, the incremental learning mechanism that generates sensitivity to such structure, and the architecture that supports lexical learning and processing, and then observe the emergent consequences. As a computational model it also allows for controlled and independent manipulation of both environmental and cognitive constraints, the emergent consequences of which can then be analysed (e.g. speed of processing, representational structure, rate of learning) and thus isolated. The degree of control possible in computational modelling is unique to this technique and therefore complements any findings from behavioural studies, where the amount of control over or means of measuring, for example amount of language exposure, are rather restricted. Computational modelling is an excellent method for the investigation of a process as complex as word learning, variation in which is likely to be multiply-determined. In addition, it provides an excellent means of getting a clearer picture of the relationships between the various factors and skills involved and their effects on learning and processing.

Within this study we model the learning and processing of spoken English words. The model architecture supports learning and processing of phonological and semantic representations of words and the mappings between such representations. It is derived from Harm and Seidenberg's (2004) implementation of the triangle model of reading, used previously to simulate word learning and processing as an interaction between phonology, semantics, and orthography. They demonstrated the model's ability to reflect word frequency effects on learning and processing. In addition, Monaghan and colleagues (2017) have reported that the same model shows effects of exposure on vocabulary size and the size of the frequency x skill interaction, as indicated above. Thus, an architecture similar to that used in this study has been applied to simulate a variety of behavioural phenomena including variation in word learning and processing, and importantly also the size of the word frequency effect. It was therefore deemed well suited for the endeavour of the present study, namely the modelling of individual differences in vocabulary learning and size as a function of variation in internal (cognitive) and external (environmental) factors.

We decided to adopt the pre-literate version of the model by Harm and Seidenberg (2004; see also Monaghan et al., 2017), focusing on the mapping between phonology and semantics and, thus, simulating spoken word comprehension and production. The reason was first of all that we intended to complement the word learning study presented in Chapter 5, which did not include the orthographic but only spoken forms of the novel words. Secondly, we intended to keep this first attempt at modelling individual variation in word learning as a function of differences in cognitive abilities and input as simple as possible, thus reducing complexity and the time needed for the model to be trained. Additionally, we did not have a reason to believe that the conclusions from this pre-literate version of the model would differ from those based on a model that, in addition to the mappings between phonology and semantics, learns to map onto orthographic representations.

In summary, this study describes simulations in a parallel distributed processing model of spoken English, which learns mappings between phonological and semantic representations. Within the neural network model we manipulated the processing speed and processing resources of the network, and in addition the quality (number of types) and quantity (number of tokens) of linguistic input to which it is exposed in the learning environment. Networks were analysed to examine the consequences of these manipulations (including interactions) for the size of vocabulary acquired, the rate of vocabulary development, the structure of the emergent lexicon (frequency, semantic density, and phonological density of acquired words), the speed of language processing (including effects of frequency, phonological density, and semantic density), and novel word learning.

This work thus aimed to offer an explicit description of mechanisms likely to cause variation in vocabulary size, and an explicit description of how such mechanisms generate further consequences for language processing and learning. Are, for example, multiple mechanisms – in addition to exposure – responsible for generating the observed frequency by skill interactions?

Further, we wished to use such simulations to explore what has been suggested to be distinct consequences of variation in vocabulary size. As described above, variation in language processing performance and in the sensitivity to word frequency effects have been related to differences in vocabulary size. Modelling word learning as a function of variation in different environmental and cognitive factors allows us to examine the consequences of variation in vocabulary size on processing while controlling for the factors potentially underlying this variation in vocabulary size. In this way, we built on studies as the ones presented in Chapters 2 to 4 in this dissertation, aiming at providing insights into the relationship between factors leading to variation in vocabulary size and language processing performance (see also Fernald et al., 2006). An important question was whether increased vocabulary size has beneficial effects on processing, which go beyond effects that can be attributed to potential causes for this greater vocabulary size. Our investigation examines the relationships between vocabulary size and language processing, vocabulary size and rate of vocabulary development, and vocabulary size and novel word learning while taking into account environmental and cognitive factors and underlying mechanisms that cause variation in vocabulary size.

In a nutshell, two questions guided the present modelling investigation: (1) What are the causes for variation in vocabulary size?, and (2) What are the consequences of variation in vocabulary size? In our model, we manipulated processing speed, processing resources, amount of exposure, and diversity of the input (Corpus Size), and examined how they relate to variation in the number of words learned by the network (causes). In addition, we investigated the effects of these manipulations and of emergent differences in vocabulary size on language processing performance, i.e. speed of processing and the frequency x skill interaction (consequences). Finally, the model was tested on a short novel word learning task, and again the effects of the manipulated environmental and cognitive factors as well as of variation in emergent vocabulary size were examined.

Methods

Architecture

The architecture of the model was based on the parallel distributed processing (PDP) framework (see McClelland et al., 1986; Rogers & McClelland, 2014), with the network consisting of multiple layers of nonlinear processing units connected via weighted connections. The architecture of our model replicates that of the pre-literate network used within Harm and Seidenberg (2004), consisting of connected phonological and semantic layers to simulate spoken language processing.

The network comprised two visible layers; a phonological layer consisting of 200 units, which simulated processing of phonological information, and a semantic layer consisting of 2446 units, which simulated processing of semantic information (see Figure 6.1). A set of 50 clean-up units was fully connected to and from the phonological layer, while a further set of 50 clean-up units was fully connected to and from the semantic layer. Hence, the model was able to develop stable phonological and semantic representations for words. Activation could pass between semantic and phonological layers via two distinct routes. The ‘production’ route allowed activation to pass from semantics to phonology, in which the semantic layer was fully connected in a forward direction to a hidden layer consisting of 100 or 300 units (see Simulations section below), which in turn was fully connected in a forward direction to the phonological layer. A separate ‘comprehension’ route allowed activation to pass in the opposite direction from the phonological layer to the semantic layer, via a hidden layer consisting of 100 or 300 units. Phonological layer units were fully connected in a forward direction to units in the hidden layer, which in turn were fully connected in a forward direction to units in the semantic layer. The numbers of units in the hidden and clean-up unit layers were determined by pilot studies with the aim of identifying the minimum number of units needed to form semantics-phonology mappings for approximately 5000 words.

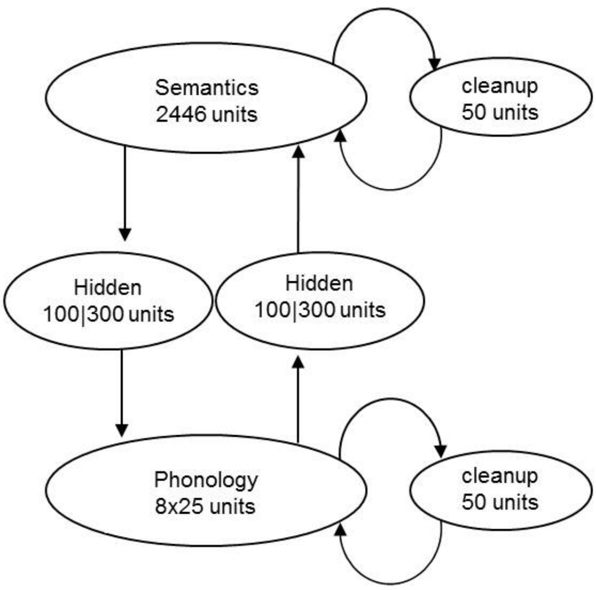

Figure 6.1: Architecture of the model used in the current simulations.

As in Harm and Seidenberg (2004) and Monaghan et al. (2017), the model utilizes time averaged input units (Plaut, McClelland, Seidenberg, & Patterson, 1996). In contrast to discrete time units, where a unit's activation from input is computed instantaneously (i.e. no state transitions), the activation of continuous time units ramps up over time (i.e. smooth/dynamic state transitions), with the speed of this process being determined by an integration constant (σ: see Equation 1 below). Hence, activation passes gradually from one layer to the next. In addition, using time averaged input units implements the assumption that the network is under pressure to produce the correct output rapidly and that greater activation leads to faster responses (Harm & Seidenberg, 2004; Plaut et al., 1996).

Representations

Phonological forms of monosyllabic words were presented at the phonological layer. Just as in Monaghan et al.'s (2017) work, the phonological layer was composed of 8 phoneme slots and each slot comprised a set of 25 units each representing distinct phonetic features (see also Harm & Seidenberg, 2004). A unique binary phonetic feature vector represented each of 40 distinct phonemes. Thus, the phonological representations were comprised of 200 unit binary feature vectors. We used slot-based representations, with three slots for the onset, one slot for the vowel, and four slots for the coda of a syllable. In the resulting CCCVCCCC syllable template, the word pen would be presented as follows: _ _ p e n _ _ _. Hence, onset and coda consonants were presented in slots adjacent to the vowel. The vowel slot comprised a set of features to represent both short and long vowels as well as diphthongs, with each unit representing the presence or absence of a single phonetic feature.

The semantic layer was composed of 2446 units. We used the same binary semantic representations as those described in Harm and Seidenberg (2004) and Monaghan et al. (2017), which were retrieved from Wordnet (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990). Hence, each word's semantic representation was an activated subset of 2446 semantic features, where the presence of a feature was indicated by activity 1.

The corpus comprised a total of 5641 monosyllabic words in English. No homophones were included in the training corpus and word frequencies were retrieved from the CELEX database (R. Baayen, Piepenbrock, & Van Rijn, 1993) and log-transformed. After a number of pilot studies, we decided to use a word frequency range of 0.1 to 1.0. Thus, all words with a log-transformed word frequency between 0.05 and 0.1 were assigned a word frequency of 0.1. In this way, the model was able to learn all words within a reasonable number of training trials per word conditioned by word frequency.

Training

For all manipulations (see below), three different versions of the model with distinct random seeds, thus with different random orderings of training patterns and different initial random weight matrices, were trained on four different corpora.

The networks were trained to map between phonological and semantic representations for all words in the training corpus. Four different tasks were used for that purpose, (1) phonological and (2) semantic retention tasks, used to train the network on maintaining phonological/semantic representations over time, and (3) phonology to semantics (pho → sem) and (4) semantics to phonology (sem → pho) mapping tasks, simulating spoken word comprehension (pho → sem) and production (sem → pho). Each trial lasted a total of 12 time steps with time in the model represented by the flow of information between units in the network. For the phonological retention task, a phonological representation was clamped on the phonological units from time step 0-7. After time step 7 activity was free to cycle through the network. The target's phonological representation was presented to the phonological layer in time steps 9 to 12, and error was back propagated. For the semantic retention task, the semantic representation of a word was clamped on the semantic layer for time steps 0-7 and activity in the model was allowed to cycle. In time steps 9-12 the target's semantic representation was provided and error was back propagated.

For the phonology to semantics mapping task (comprehension), the phonological representation was clamped to the phonological layer for all 12 time steps and for time steps 9-12 the word's semantic representation was presented to the semantic layer and error was back propagated. Finally, in the semantics to phonology mapping task (speech production), the semantic representation of a word was clamped to the semantic layer of the model for all 12 time steps. The word's phonological representation was provided at time steps 9-12 error was back propagated. As indicated before, the model was trained using recurrent backpropagation (Rumelhart et al., 1986). For each word, cross-entropy error was computed between the model's output and the target representation. Cross-entropy error [3] is a commonly used cost function for training PDP neural network models, where the error increases exponentially as the distance between target and model output increases. A learning rate of 0.2 was used for the entirety of training, which was comprised of two stages. In training stage 1, the model was trained on a corpus of 2500 or 5000

[3] For the binary target vectors used in our simulations, cross entropy error is defined as follows:

$$-\sum_{i} \log q(x)$$

Where i represents each unit in the ouput layer. When the target for unit i=1, $q(x)$ is the actual output for unit i. While in cases where the target for unit i=0, $q(x)$ = 1-actual output for unit i.

words while in training stage 2, the model was only trained on an additional 250 novel words. Training stage 2 was not interleaved with training on words from stage 1 in order to control for language exposure and to limit effects of capacity limitations.[4]

In both training stages, words were selected pseudo randomly according to their frequency and then assigned to one of the training tasks. Due to differences in task complexity in both training stages, the phonological retention task made up 10% of the training trials, semantic retention task was completed on 10% the trials, on 40% of trials the model was trained on the mapping between phonology and semantics, and 40% of trials were the semantics to phonology mapping task (see Harm & Seidenberg, 2004; Monaghan et al., 2017). Finally, small amounts of noise were applied to input clamps for the phonological and semantic layer units, simulating noisy visual and auditory input. The levels of noise applied was the same across all tasks and trials.

Test measures

The model was tested on all four tasks: The (1) phonological and (2) semantic retention tasks, where the network had to maintain phonological/semantic representations over time, and the mapping tasks (3) from phonology to semantics (comprehension) and (4) from semantics to phonology (production). For the retention as well as the mapping tasks, the phonological or semantic representation of each word was presented to model. At the end of the 12 time steps of activation, the model's production at the phonological or semantic layer was recorded.

As there was less variation in performance on phonological and semantic mapping tasks (most networks performed at ceiling on this task) our analysis focused on performance on the comprehension and production tasks. For both of these tasks, we recorded cross-entropy error (calculated between output layer activity at the end of training trials and the target representation), accuracy and reaction time (RT). The network was considered to have accurately generated the correct word if the cosine distance between the output layer and target representation was lower than the distance between the output and all other representations in the training corpus. Overall task accuracy was measured as the proportion of words for which the network 'accurately' generated the target representation. From this we derived the number of words known, i.e. a network's vocabulary size, at a particular point in time during training (accuracy

[4]We accept that given an absence of further training on stage 1 words, training on novel words will ultimately lead to catastrophic interference and thus knowledge of stage 1 words will diminish. This could be resolved within the current model by interleaving training on both stage 1 and stage 2 words. However, as this introduces additional variation across manipulations in total stage 2 language exposure and increases effects of capacity limitations, we avoided such complexity and in the interest of time, training on stage 2 words only was implemented.

x corpus size). The model's RT was measured as the mean time step at which the cosine distance between output layer activation and target representation first became lower than the distance to any other representation in the training corpus.

Simulations

Environmental factors

Quality of language exposure

The number of words in the training corpus was varied in order to investigate the influence of differences in quality of exposure (number of types) on the model's word learning performance. We generated corpora of two different sizes, comprising 2500 or 5000 words, by randomly sampling words from a larger corpus of 5461 words. As indicated above, a total of four corpora per level were generated and three versions of the model (with different seeds) were trained on each corpus. This yielded a total of 24 simulations, 12 per corpus size.

Quantity of language exposure

The total amount of language exposure was the same for all networks, with all of them being trained on a total of 3 million training trials in training stage 1. Hence, the 2500 word corpus was trained for 1200 training trials per word (2500 words x 1200 trials = 3 million training trials in total). The 5000 words corpus completed 600 training trials per word (5000 words x 600 trials = 3 million training trials in total). Consequently, at the end of training the number of trials per word (number of tokens per word) differed between the two levels of corpus size while the total language exposure, i.e. 3 million training trials, was constant across the corpus sizes. The effect of quantity of language exposure (number of tokens) on each network's learning performance was examined by testing their performance on the words at various time points during training, namely at intervals of 50 trials per word. In this way, it was possible to examine the networks' performances at equal levels of exposure per word and at equal levels of overall language exposure (trials per word x corpus size).

In training stage 2 all networks were trained on the same number for words, i.e. 250 novel words, for 10 trials per word, independent of the corpus size of training stage 1. This enabled us to examine the effects of previous knowledge or exposure on novel word learning.

Cognitive factors

For each level of each cognitive factor manipulated, 3 networks each with different random seeds were trained on each of two levels of corpus size (2500 vs. 5000). As indicated before, four different corpora were generated for each of the three corpus size levels. This resulted in a total of 24 networks per cognitive factor level (3 networks x 2 corpus sizes x 4 corpora).

Processing speed

To simulate variation in processing speed we varied the integration constant (σ). This parameter determines the rate at which activation of a processing unit within the network ramps up over time as a function of the unit's input (see equation 6.1; equation 14 in Plaut et al., 1996). Within equation 1: x_t is the input to the unit at time t, (σ) is the integration constant, $x_{t-1}$is the input to the unit at time t-1 and x_i is the summed external input at time t-1 (i.e. the true input received by the unit at time t-1).

$$x_t = (1 - \sigma)x_{t-1} + \sigma x_i \tag{6.1}$$

Figure 6.2, displays differences in the temporal dynamics of a high processing speed unit (integration constant = 4 / [number of time steps, i.e. 12] = 0.333) and low processing speed unit (integration constant = 2 / [number of time steps, i.e. 12] = 0.167). As indicated by the figure the activation of a unit with a high integration constant is able to adapt more rapidly to a given input (see Figure 6.2), such a unit is therefore able to propagate more information through the network within a given period of time, thus process information at a faster rate. It therefore follows that it is feasible for networks with increased integration constants to learn more within a fixed period of time.

Relative to a number of previous implementations, which aim to vary processing speed within PDP networks (e.g. input gain: Kello & Plaut, 2003, multiple hyper-parameters: McMurray et al., 2012), varying the integration constant has limited consequences for the computational properties of a network (see Plaut et al., 1996) and thus offers a more tractable means of isolating the effects of variation in speed of processing.

General intelligence

In the present model, general intelligence was operationalised by manipulating the number of hidden units in both the comprehension and production pathways. Baseline networks possessed 300 units in each hidden layer, while networks with reduced processing resources contained 100 units per hidden layer. We thus simulate variation in general intelligence as differences in the processing resources available at processing

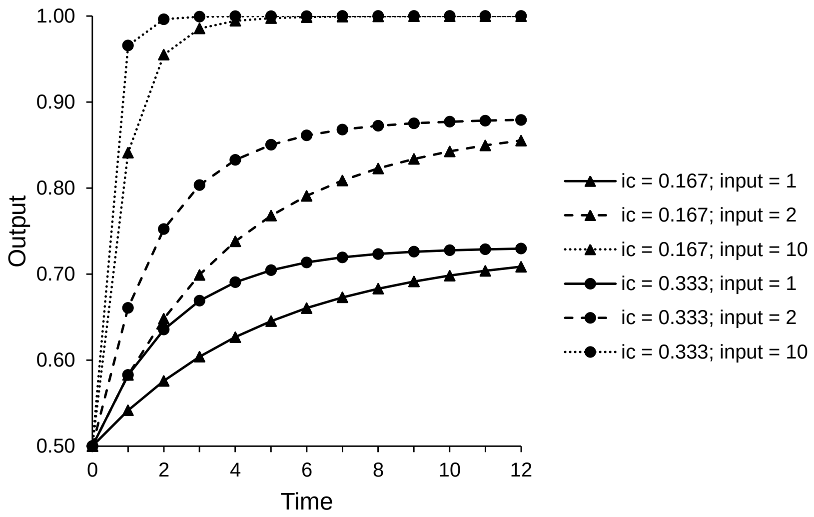

Figure 6.2: The activation over time of a continuous time averaged input unit when presented with a fixed input (1, 2, or 10) and assigned a high (0.333) or low (0.167) integration constant (ic).

bottlenecks within the cognitive system. This manipulation within the model aims to simulate a plausible mechanism that may underlie variation in performance on tasks designed to capture variation in general intelligence. We do not argue that this is the only plausible mechanism for simulating variation in general intelligence. Instead we believe that variation in performance on tasks that aim to measure this cognitive component are likely driven by variation in multiple underlying cognitive mechanisms, and thus we argue that our implementation represents one such mechanism.[5]

The number of units within a hidden layer ultimately determines the complexity of the mappings a given network is able to learn. However, it is not our assertion that variation in vocabulary size as described in the literature review above results from ultimate capacity limitations of individuals' cognitive architectures. Instead, care has been taken to ensure that at all networks with reduced processing capacity reported within this study possess the capacity to learn all mappings learnt by the networks with increased processing capacity to which they are compared. Instead as previously stated this manipulation aims to simulate differences in processing, representation, and behaviour that may result from reduced computational resources at processing

[5]In pilot simulations, as an alternative mechanism for modelling variation in general intelligence, we varied the efficiency of information processing by varying the level of noise added to activation as it passed along connections in the network (similar to Smith, Monaghan, & Huettig, 2014). Pilot simulations were unable to identify an appropriate level of noise that effected rate of learning without having catastrophic effects on the total number of mappings the network was able to learn. We do not, however, rule out such a mechanism as an alternative viable cause of variation in vocabulary size.

bottlenecks within the cognitive system. We believe variation in the number of hidden units is appropriate for this role as it generates variation in network properties such as the size of the representational space into which representations can be projected, thus the richness of representations, the confusability of items, and sensitivity to systematic relationships at higher levels of abstraction; all of which have additional consequences for processing and learning.

Summary

In summary, different networks were trained on corpora of either 2500 or 5000 words. In addition to variation in corpus size, the networks varied with regards to the number of hidden units (100 vs. 300), processing speed (2 vs. 4, with 4 being faster), with the baseline parameter settings being 300 hidden units and a speed of 4. In this way, we obtained three different types of networks, namely baseline networks, those with decreased hidden units, and those with decreased processing speed (see Table 6.1).[6]

Table 6.1: Summary of networks and manipulations.

Type/ manipulation	**Corpus size**	**Hidden units**	**Processing speed**
Baseline	2500	300	4
Hidden units	2500	100	4
Speed	2500	300	2
Baseline	5000	300	4
Hidden units	5000	100	4
Speed	5000	300	2

Analyses and Results

Before turning to the analyses we ran to address the research questions underlying the present study, the product of the above-described training over a total of 3 million training trials is presented. Figure 6.3 displays the development of cross entropy error for the production and comprehension tasks across all training trials. As indicated by the figure, production and comprehension error reduces over time, reflecting increasing of knowledge

[6]Networks either diverged in the number of hidden units or processing speed relative to baseline networks, we did not manipulate both cognitive factors simultaneously, hence we do not report interactions between cognitive factors.

of the cross-modal mappings over time, with all networks converging on similar levels of performance at the end of training.

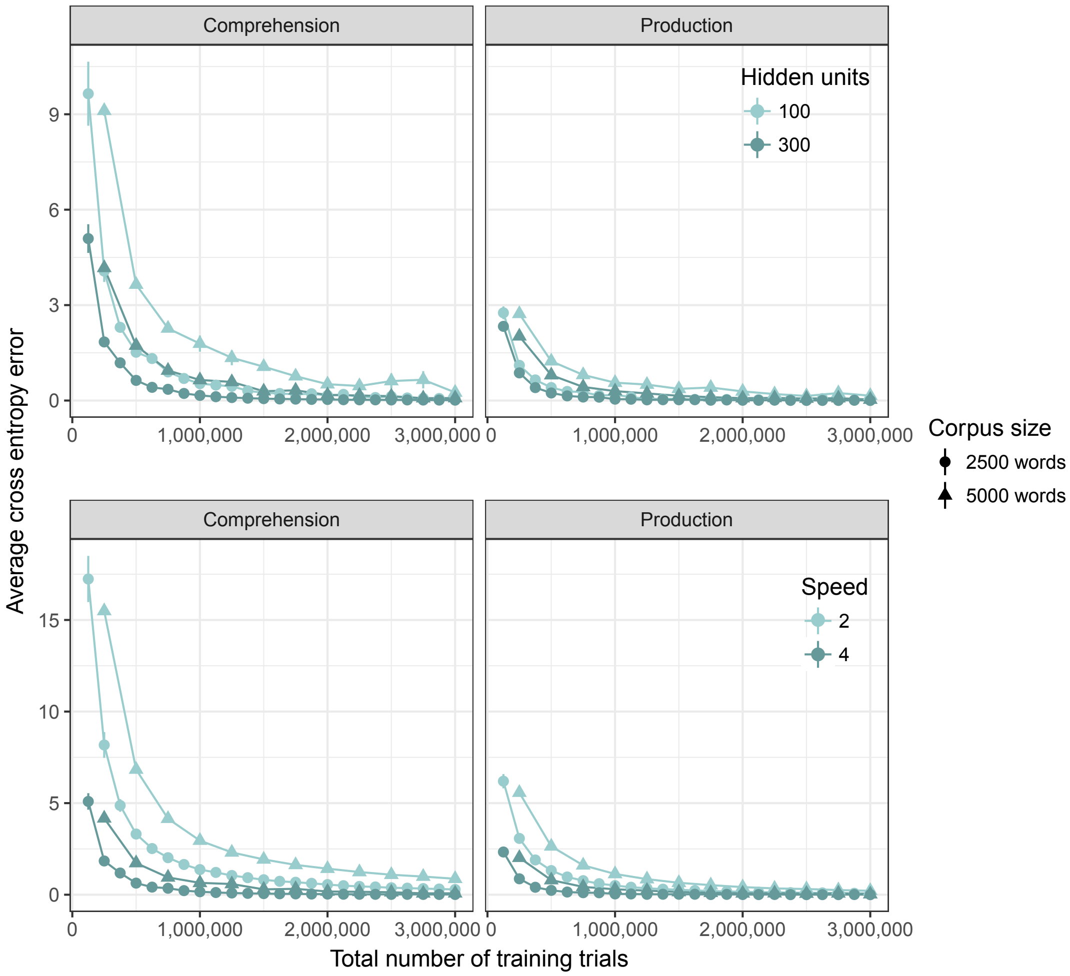

Figure 6.3: The development of the mean cross entropy error for the comprehension and production tasks over the total number of training trials as a function of corpus size and number of hidden units (top panels) as well as corpus size and processing speed (bottom panels). Standard errors are indicated in the plot.

Figure 6.4 displays the development of the number of known words, i.e. vocabulary size, of the networks exposed to 2500 and 5000 words over the total number of training trials for both comprehension and production. The learning curve is asymptotic with the number of words in the corpus determining the maximum number of words that can be acquired by each of the networks.

As indicated by Figures 6.3 and 6.4 the rate of increase in vocabulary size and decrease in task error for production exceeds that of comprehension. This is due to differences in the complexity of the mapping tasks; it is more difficult for networks to generate the sparsely distributed semantic patterns than it is to generate the componential

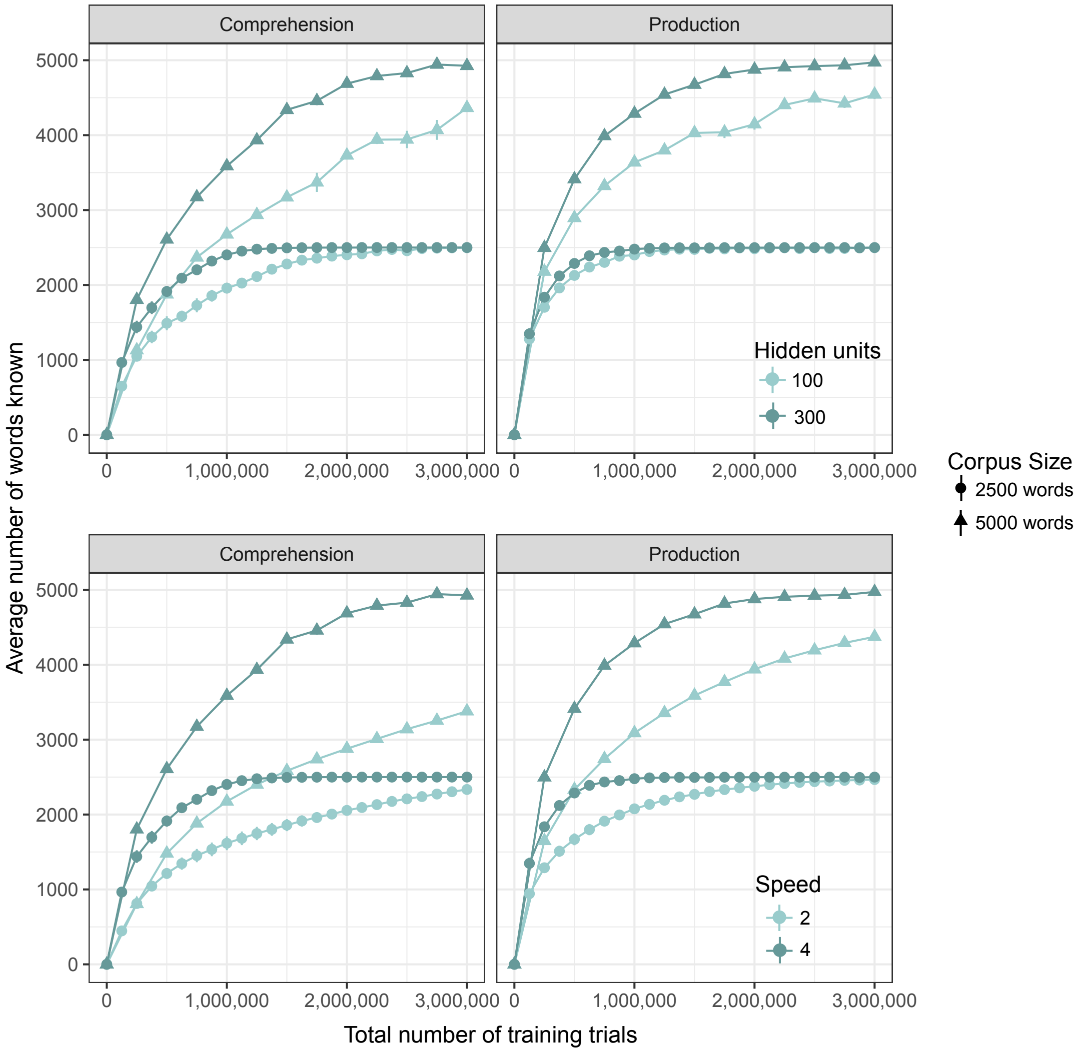

Figure 6.4: The development of the mean comprehension and production vocabulary size (and standard error) over the total number of training trials as a function of corpus size and number of hidden units (top panels), and corpus size and processing speed (bottom panels).

phonological patterns.[7] Our analysis focuses on behaviour common to both production and comprehension measures. Within this study we do not aim to draw conclusions generated by differences in task complexity between production and comprehension.

All further analyses were run on two time points during training namely after 250,000 and 500,000 training trials. This period was selected as it occurs before the

[7]Behavioural data suggest the opposite, namely that production lags behind comprehension in vocabulary learning and size. It might be possible that the model is doing a more difficult task by being required to produce the full semantic code, which is unlikely to be necessary in an average comprehension task. Alternatively, production may be better than comprehension performance because the model might be doing an easier production task than humans as the full semantic code is presented as input and does not need to be generated by other components of cognition.

networks' learning performance had reached asymptote and before capacity limitations greatly influence behaviour (Figure 6.4 demonstrates that vocabulary size was able to increase substantially beyond levels recorded at 500k training trials given further training). Further, selecting this period ensured that we analysed time points during training, at which the networks had already acquired substantial knowledge of their linguistic environment, and hence results do not reflect behaviour specific to early stages of language development. Thus, we examine networks at a stage that is likely to reflect the behaviour of adults, which has been the focus of the experiments presented in the previous chapters of this dissertation.

Two broad questions guided this study and the analyses of the data generated by the computational model. First of all, we were interested in the causes of variation in vocabulary size and secondly, we aimed at examining consequences of variation in vocabulary size. The presentation of the analyses and results reflects these questions and is structured accordingly, first addressing the causes of variation in vocabulary size followed by sections examining potential consequences.

All analyses were run on a total of 72 networks that had been exposed to corpora of 2500 or 5000 words in total and we looked at two different time points (250k and 500k training trials). Hence, the total amount of exposure (i.e. training trials) was equated across networks with different corpus sizes while that meant that the number of exposures per word differed depending on the corpus size. The networks that saw 5000 words in total had on average 50 exposures per word in 250k training trials and 100 exposures per word in 500k training trials, whereas the networks that were exposed to 2500 words in total had seen each word on average 100 times in 250k training trials and 200 times in 500k training trials. As indicated before, in addition to variation in corpus size, the networks varied with regards to the number of hidden units (100 vs. 300) processing speed (2 vs. 4, with 4 being faster), with the baseline parameter settings being 300 hidden units and a speed of 4.

All analyses were run using the lmer function from the lme4 package (Bates et al., 2015) in R (version 3.4.1; R Core Team, 2017). All models included an intercept and fixed effects for the factors Training Trials per word, Corpus Size, Hidden Units, and Speed. In addition, we modelled the two-way interactions between Training Trials per Word and all other factors, and between Corpus Size and all other variables. All models included random intercepts for each network and for the four different corpora per corpus size. All categorical predictors (Corpus Size, Hidden Units, Speed) were dummy coded and the lower level was in each case used as the reference level (2500 words corpus, 100 hidden units, 2 for processing speed). The continuous predictors, i.e. vocabulary size, density, and word frequency, used in some of the analyses, were all z-transformed using the scale function in R.

Causes of variation in vocabulary learning and size

As indicated before, the model's performance error was measured as the cosine difference between the output layer and all phonological or semantic representations in the corpus. Overall accuracy was measured as the proportion of words for which the cosine distance between the target and the model's actual production was closest relative to all other words in the corpus. The measure of interest for the present analyses was the number of words acquired by the model, i.e. its vocabulary size, which was derived from accuracy (corpus size x accuracy rate).

For the purpose of examining the causes for individual differences in vocabulary, we ran network level analyses examining the number of words known for comprehension as well as for production at the two different points in time. The analyses on comprehension vocabulary size showed significant effects of Trials per Word ($\beta = 0.20$, $t = 9.59$, $p < .001$), Corpus Size ($\beta = 0.90$, $t = 17.14$, $p < .001$), Hidden Units ($\beta = 0.68$, t = 25.89, p < .001), and Speed ($\beta = 1.08$, $t = 41.21$, $p < .001$). Hence, more exposures per word were associated with larger vocabularies and networks that had been exposed to the larger corpus or those with more processing capacities or higher processing speed knew more words after 250k or 500k training trials.

In addition, all two-way interactions were significant. The interaction between Training Trials per Word and Corpus Size ($\beta = 1.15$, $t = 67.83$, $p < .001$) indicated stronger effects of the number of exposures per word for the larger corpus size. With more exposures per word, the networks exposed to 5000 words learned more words than those exposed to 2500 words. Thus, networks with an increased input quality (more diverse input) but not increased input quantity (not more exposure in total) benefitted more from getting more exposures per word. The learning rate was higher for networks that were exposed to 5000 vs. 2500 words. Potential reasons for this are further examined below. Furthermore, Training Trials per Word showed significant interactions with Hidden Units ($\beta = 0.15$, $t = 8.81$, p < .001) and Speed ($\beta = 0.20$, $t = 11.82$, $p < .001$), showing that the effects of both manipulations were stronger with increasing training. In this early phase of learning, the benefits of having greater processing capacities and higher processing speed increased with increased exposures per words while Figure 6.4 shows that the effects of processing capacity and speed on vocabulary size reduce towards the end of training. Hence, networks with greater processing capacity and higher processing speed showed higher learning rates but no increase in overall capacity. Finally, Corpus Size interacted significantly with both Hidden Units (β

= 0.80, t = 19.91, p < .001) and Speed (β = 1.11, t = 27.45, p < .001). This suggests that the effects of Hidden Units and Speed were stronger for the networks exposed to 5000 words as compared to those that are faced with the task of learning a smaller number of words (2500 words). The benefits of having greater processing capacity (more hidden units) and higher processing speed were greater for networks that were exposed to the larger corpus (see Appendix A).

The analyses on the networks' production vocabulary showed exactly the same effects (see Table 6.2 for estimates, t-, and p-values).

Table 6.2: The estimates, t-, and p-values for the fixed effects in the mixed-effects analyses on production vocabulary size after 250k and 500k training trials. For the dummy coded binary variables, the reference level was in all cases the smaller value and the comparison level is indicated in brackets.

Fixed effect	**Estimate**	***t*-value**	***p*-value**
Intercept	-1.64	-51.22	<.001
Corpus (5000)	1.56	30.26	<.001
Hidden Units (300)	0.19	7.79	<.001
Speed (4)	0.91	37.32	<.001
Training Trials per Word	0.19	7.61	<.001
Training Trials:Corpus	1.0	49.37	<.001
Training Trials:Hidden Units	0.12	6.02	<.001
Training Trials:Speed	0.21	10.41	<.001
Corpus:Hidden Units	0.58	14.83	<.001
Corpus:Speed	0.83	20.96	<.001

Characteristics of the known words[8]

Corpus measures that could be extracted to describe the words learned by the model were average word frequency of the words known by each of the networks as well as phonological and semantic density. Phonological and semantic density of each word was measured as the number of words within a hamming distance of 10 from the target. Hence, local phonological or semantic density was indicated by the number of phonological/semantic representations that differed in up to 10 units from the target representation.

[8]We initially intended to run mixed-effects models with error as dependent variable, predicted by the different cognitive and environmental factors as well as the lexical characteristics word frequency and semantic as well as phonological density. These logistic models failed to converge. Therefore, we decided that for the present purposes it is sufficient to provide some insights into the characteristics of the networks' lexica by examining the mean word frequency and phonological and semantic densities of the known words.

Frequency

We further examined potential reasons for why networks that are exposed to a larger corpus show an increased learning rate, hence, why they benefit more from an increase in the number of exposures per word. For this purpose, we looked at the average word frequency of the words known by each network. The idea was that the larger corpus might contain a larger number of easier-to-learn high-frequency words, which might initially increase the learning rate of those networks that are exposed to the larger corpus. The mean frequency values for both corpus sizes confirmed this.

At same numbers of total training trials, the mean frequencies were higher for the networks exposed to 5000 words than for those exposed to 2500 words (see Table 6.3 below and Appendix A for Figures). Thus, the networks that were exposed to a larger corpus were probably exposed to a larger number of high-frequency words, which were easier to learn. Therefore, getting more exposures per word was particularly beneficial for the networks exposed to 5000 words as they were, hence, able to learn the larger number of high-frequency words whereas the networks exposed to 2500 words had fewer of those easy words in their input. T-tests confirmed a significant difference in mean word frequencies between networks that had been exposed to 2500 vs. 5000 words during training for both time points of interest and for both comprehension (250k trials: t(32373) = -35.62, $p <$ .001; 500k trials: t(45541) = -35.94, $p < .001$) and production (250k trials: t(44370) = -34.86, $p < .001$; 500k trials: t(58725) = -30.09, $p < .001$; see Appendix A for Figures).

Furthermore, the differences in mean word frequency of the known words between 50 and 100 trials per word for the networks trained on the larger corpus (comprehension: t(47308) = 44.74; $p < .001$; production: t(64048) = 40.85, $p < .001$) and between 100 and 200 trials per word for the networks trained on the 2500 words corpus (comprehension: t(33541) = 32.46, $p < .001$; production: t(43729, $p < .001$) were statistically significant. Hence, overall mean word frequency reduced significantly over training confirming the impression that high-frequency words are learned first and low-frequency words later in training.

Table 6.3: Mean word frequency for comprehension and production as a function of corpus size and number of training trials per word.

Trials per Word	Corpus	Comprehension		Production	
		Mean	SD	Mean	SD
50	5000	0.64	0.29	0.52	0.31
100	2500	0.52	0.30	0.42	0.31
100	5000	0.52	0.31	0.43	0.31
200	2500	0.42	0.31	0.35	0.30

In addition, we also examined whether the mean word frequencies (and mean semantic and phonological densities) of the words known by each network reveal anything about the origins of the advantages for networks with higher processing capacities or speed.

Interestingly, for both comprehension and production we observed that the mean word frequency of the known words was lower for the networks with greater processing capacities (300 hidden units) and for those with higher processing speed (see Tables 6.4 and 6.5). This suggests that the networks that were intended to simulate more intelligent or faster processing word learners learned a greater number of low-frequency words than the networks with reduced processing capacities or speed. Note that these low-frequency words are assumed to be more difficult to learn, only because they are low in frequency; no other lexical characteristics were manipulated explicitly to render these words more difficult to learn.

T-tests confirmed significant differences between the mean word frequencies for networks with 300 vs. 100 hidden units (comprehension: $t(138220) = 63.21$, $p < .001$; production: $t(218160) = 26.25$, $p < .001$) and for those with higher vs. lower processing speed (comprehension: $t(109870) = 99.74$, $p < .001$; production: $t(171850) = 81.26$, $p < .001$).[9]

[9] For these analyses, data was collapsed across both corpus sizes and both time points of interest. Analysing data for the networks exposed to 2500 vs. 5000 words and for time points 250k vs. 500k trials separately yields the same results, with all comparisons between 100 vs. 300 hidden units and speed 2 vs. 4 being significant.

Table 6.4: Mean word frequency for comprehension and production as a function of number of hidden units and number of training trials per word.

Corpus	Trials per Word	Hidden Units	Comprehension Mean	Comprehension SD	Production Mean	Production SD
2500	100	100	0.62	0.29	0.45	0.31
2500	100	300	0.52	0.30	0.42	0.31
2500	200	100	0.51	0.31	0.37	0.31
2500	200	300	0.42	0.31	0.35	0.30
5000	50	100	0.75	0.27	0.56	0.31
5000	50	300	0.64	0.29	0.52	0.31
5000	100	100	0.62	0.30	0.47	0.32
5000	100	300	0.52	0.31	0.43	0.31

Table 6.5: Mean word frequency for comprehension and production as a function of processing speed and number of training trials per word.

Corpus	Trials per Word	Speed	Comprehension Mean	Comprehension SD	Production Mean	Production SD
2500	100	2	0.70	0.27	0.55	0.31
2500	100	4	0.52	0.30	0.42	0.31
2500	200	2	0.58	0.29	0.46	0.31
2500	200	4	0.42	0.31	0.35	0.30
5000	50	2	0.81	0.24	0.64	0.30
5000	50	4	0.64	0.29	0.52	0.31
5000	100	2	0.70	0.28	0.55	0.31
5000	100	4	0.52	0.31	0.43	0.31

Phonological density

Comprehension. At the same stages of training (i.e. 250k trials and 500k trials), the phonological density as measured by the mean number of words within a hamming distance of 10 from each word, was larger in the networks exposed to 5000 as compared to those exposed to 2500 words. As would be expected, more diverse input, i.e. a greater number of different words, leads to denser phonological networks in the vocabularies acquired by the model. This was confirmed by a t-test comparing the mean phonological density between the two corpus sizes (t (81000) = -178.84, p < .001). Furthermore, mean phonological density reduced over training. Hence, with increasing exposure (i.e. training trials per word), the words learned by the networks exposed to both corpus sizes had on average significantly sparser phonological neighbourhoods (2500 words: t(32738) = 5.84, p < .001; 5000 words: t(45139) = 7.35, p < .001). Just as for word frequency, this confirms the idea that words from denser neighbourhoods are easier to learn and are, thus, learned first before words from less dense neighbourhoods are acquired.

The phonological density of the words known by the networks at the two different time points, also differed as a function of number of hidden units (t (133920) = 7.38, p < .001) and processing speed (t (102910) = -2.86, p = .004). The vocabularies of networks with greater numbers of hidden units and higher processing speed were phonologically less dense; hence, the number of words within a hamming distance of 10 from each known word was overall smaller for networks with 300 vs. 100 hidden units and for faster vs. slower networks.

Production. The relationship between corpus size and phonological density was the same for production as for comprehension, with the vocabularies known by the networks that received greater input (5000 words corpus) being phonologically more dense than the words known by the networks exposed to 2500 words (t (109690) = -208.15, p < .001). Again, just as for comprehension, phonological neighbourhood density reduced significantly over training (2500 words: t(43992) = 3.92, p < .001; 5000 words: t(62231), p < .001).

The number of hidden units did not have an effect on phonological density in the production vocabularies, although networks with 300 hidden units show numerically smaller phonological density (t (218300) = -0.07, p = .94). Finally, the phonological density of the vocabularies known by the faster networks was smaller than the density of the words known by the slower networks. Hence, networks with faster processing speed learned more words with more sparse neighbourhoods, i.e. with fewer words within a hamming distance of 10 (t (168510) = 9.28, p < .001).

Table 6.6: Mean number of words within a hamming distance of 10 of the phonological representations. Mean phonological density of all known words depending on Corpus Size.

Trials per Word	**Corpus**	**Comprehension**		**Production**	
		Mean	**SD**	**Mean**	**SD**
50	5000	752.57	414.70	738.35	403.68
100	2500	372.12	204.40	365.32	199.74
100	5000	725.67	405.74	718.98	396.52
200	2500	359.60	200.27	358.13	196.42

Table 6.7: Mean number of words within a hamming distance of 10 of the phonological representations. Mean phonological density of all known words depending on number of Hidden Units.

Corpus	**Trials per Word**	**Hidden Units**	**Comprehension Mean**	**Comprehension SD**	**Production Mean**	**Production SD**
2500	100	100	388.73	207.76	369.48	200.33
2500	100	300	372.12	204.40	365.32	199.74
2500	200	100	374.85	204.09	360.02	197.54
2500	200	300	359.60	200.27	358.13	196.42
5000	50	100	769.58	414.50	746.19	407.18
5000	50	300	752.57	414.70	738.35	403.86
5000	100	100	762.53	415.87	726.76	400.13
5000	100	300	725.67	405.74	718.98	396.52

Table 6.8: Mean number of words within a hamming distance of 10 of the phonological representations. Mean phonological density of all known words depending on Processing Speed.

Corpus	Trials per Word	Speed	Comprehension		Production	
			Mean	SD	Mean	SD
2500	100	2	383.05	206.83	379.15	203.87
2500	100	4	372.12	204.40	365.32	199.74
2500	200	2	375.27	206.10	396.35	201.01
2500	200	4	359.60	200.27	358.13	196.42
5000	50	2	723.61	402.58	756.79	409.01
5000	50	4	752.57	414.70	738.35	403.86
5000	100	2	743.82	413.30	741.17	405.05
5000	100	4	725.67	405.74	718.98	396.52

Semantic density

Comprehension. Just as for phonological density, the semantic density of the vocabularies learned by the different networks varied as a function of corpus size. Just as in the case of phonological density, a word's semantic density was measured as the mean number of words within a hamming distance of 10. The networks exposed to 5000 words acquired semantically denser vocabularies than those exposed to 2500 words during training (t (81028) = -123.56, $p < .001$). In the vocabularies of the networks exposed to 5000 words, semantic density did not differ significantly as a function of Training Trials per Word (t(45018) = 0.14, $p = .89$), whereas the semantic density of the words known by networks exposed to the smaller corpus reduced significantly with increasing training per word (t(32758) = 3.95, $p < .001$). Hence, only for the small corpus, higher number of trials per word were associated with reduced semantic density, again suggesting that words with semantically denser neighbourhoods are learned first, supposedly because they are easier to learn, followed by words from semantically less dense neighbourhoods.

The mean number of words known within a hamming distance of 10 of each target word was larger for networks with 300 vs. 100 hidden units (t (134890) = -5.71, $p < .001$). Hence, the semantic density of the words known was larger for networks with greater processing capacities. Finally, the semantic density of the comprehension vocabulary was overall smaller for faster processors (t (100970) = -4.21, $p < .001$).

Production. For production, the relationship between semantic density and corpus size was the same as for comprehension, with a larger corpus size being associated with semantically denser vocabularies learned by the model (t (110070) = -147.66, $p < .001$). In addition, the mean semantic density of the known words was lower with increasing training trials per word; but this was significant only for networks exposed to the 5000 words corpus (t(63414) = 3.01, $p = .003$), and not for those exposed to the smaller corpus (t(44266) = 1.21, $p = .23$). As opposed to comprehension, the production vocabularies showed semantically less dense vocabularies for networks with 300 vs. 100 hidden units. The difference was statistically significant for the networks trained on the 5000 words corpus (t (127640) = 4.32, $p < .001$), and marginally significant for the networks trained on the smaller corpus (t (90250) = 1.94, $p = .06$).

Finally, the number of words within a hamming distance of 10 from each target did not differ between faster and slower processors for networks trained on 5000 words (t (99757) = -1.23, $p = .22$) and those trained on 2500 words (t (68370) = -0.01, $p = .99$).

Table 6.9: Mean word frequency for comprehension and production as a function of corpus size and number of training trials per word.

Trials per Word	**Corpus**	**Comprehension**		**Production**	
		Mean	**SD**	**Mean**	**SD**
50	5000	1183.85	965.13	1179.44	927.41
100	2500	598.35	474.55	571.13	460.70
100	5000	1182.67	940.51	1158.25	915.0
200	2500	578.66	465.44	565.99	458.63

Table 6.10: Mean word frequency for comprehension and production as a function of number of hidden units and number of training trials per word.

Corpus	Trials per Word	Hidden Units	Comprehension		Production	
			Mean	SD	Mean	SD
2500	100	100	591.90	484.58	579.35	463.45
2500	100	300	598.35	474.55	571.13	460.70
2500	200	100	591.78	474.0	570.15	459.87
2500	200	300	578.66	465.44	565.99	458.63
5000	50	100	1133.69	977.29	1198.81	935.51
5000	50	300	1183.85	965.13	1179.44	927.41
5000	100	100	1172.74	960.76	1182.12	922.42
5000	100	300	1182.67	940.51	1158.25	915.0

Table 6.11: Mean word frequency for comprehension and production as a function of processing speed and number of training trials per word.

Corpus	Trials per Word	Speed	Comprehension		Production	
			Mean	SD	Mean	SD
2500	100	2	602.69	490.63	572.41	470.19
2500	100	4	598.35	474.55	571.13	460.70
2500	200	2	602.60	479.33	564.94	463.94
2500	200	4	578.66	465.44	565.99	458.63
5000	50	2	1106.69	979.04	1161.34	956.51
5000	50	4	1183.85	965.13	1179.44	927.41
5000	100	2	1204.35	976.44	1159.57	935.16
5000	100	4	1182.67	940.51	1158.25	915.0

Consequences of variation in vocabulary learning and size

To get insights into the consequences of variation in vocabulary size, we conducted two different analyses. First of all, we examined the same time points as described above and looked at the networks' comprehension and production RTs and the interactions between word frequency effects and skill, i.e. Hidden Units, Speed, and Vocabulary Size. Secondly, we examined novel word learning performance of the networks described above. All results are briefly summarised in two tables at the end of the Results section.

Reaction times

Reaction times (RTs) are assumed to be an adequate measure of the knowledge of words and the strength of the representations. As indicated before, the calculation of RTs was closely related to task error (see Figure 6.3). The RT measure was derived by first calculating at each time step within the test trial (comprehension or production trial) the cosine distance between the output layer (semantic layer for comprehension trials, phonological layer for production trials) and all other word's representations in the corpus. The production or comprehension RT was the time step at which the distance between the model's output and the target was lowest relative to all other representations in the corpus.

The interaction between word frequency effects and skill were of particular interest. Previous studies have shown weaker frequency effects for individuals or networks with larger vocabularies (i.e. greater linguistic skills) and have proposed that this is a result of stronger or more entrenched representations in high-vocabulary individuals, which are less affected by lexical characteristics such as word frequency (Brysbaert et al., 2016a; Diependaele et al., 2013).

For the purpose of further examining this observation and potential underlying mechanisms, we analysed the words that all networks had learned by the end of 250k training trials (this ensured the complexity of the mapping tasks was equated across manipulations within the model). For these items, we analysed all networks' RTs at 250k training trials and at 500k training trials, again separately for comprehension and production. The first mixed-effects models included the predictors Corpus Size, Hidden Units, and Speed and their interactions, just as before, and in addition to that Word Frequency, Semantic Density, and Phonological Density. The latter were z-transformed using the scale function in R, and all two-way interactions between these word-level characteristics and the network features Corpus Size, Hidden Units, and Speed were included. In addition to the random intercepts for Network and Corpus, we included a random intercept for Word. Thus, in the first analyses we included those factors as predictors, which have been shown to cause variation in vocabulary size (see previous section). In a second set of analyses, we included Vocabulary Size as continuous, z-transformed predictor in addition to those other factors to examine whether Vocabulary Size captures any variance in RTs even when closely related factors (such as Hidden Units and Speed) are included.

Comprehension. The first analyses showed significant effects of Corpus Size ($\beta = 0.84$, $t = 9.48$, $p < .001$), Training Trials per Word ($\beta =$ -0.31, $t =$ -10.19, $p < .001$), Hidden Units ($\beta =$ -0.14, $t =$ -3.49, $p < .001$), and Speed ($\beta =$ -0.74, $t =$ -17.95, $p < .001$) on comprehension RTs. Networks trained on 5000 words had slower RTs than those trained on 2500 words. Furthermore, increased training per word, greater processing capacity and higher processing speed were associated with faster RTs. In addition, networks were faster at comprehending high-frequency words than words with lower frequencies (β = -0.26, t = -4.90, $p < .001$), and showed faster RTs with increasing phonological density (β = -0.39, t = -7.35, $p < .001$) and semantic density (β = -0.12, t = -2.19, p = .03).

Additionally, some interactions were significant predictors of comprehension RTs. The interactions between Corpus Size and Hidden Units (β = -0.25, t = -3.94, $p < .001$) and between Corpus Size and Speed (β = -0.44, t = -6.93, $p < .001$) indicated that the effects of Hidden Units and Speed differed were stronger for the networks trained on the 5000 vs. 2500 words corpus. Furthermore, the Trials per Word effect was stronger for the 5000 vs. 2500 words corpus (β = -0.51, t = -10.37, $p < .001$). Training Trials per Word also showed significant interactions with Speed (β = 0.13, t = 5.95, $p < .001$), Semantic (β = 0.13, t = 5.95, $p < .001$), and Phonological Density (β = 0.13, t = 5.95, $p < .001$). The Trials per Word effect was stronger for the slower than for the faster networks, indicating that increased training was generally beneficial but even more so for networks with reduced processing speed. Besides that, the Training Trials per Word effect was stronger for the words with lower Phonological and Semantic densities. Hence, for the words that are presumably more difficult to learn (those from more sparse neighbourhoods), additional training was particularly beneficial for the creation of strong representations that are fast to be accessed. Additionally, significant interactions were found between Semantic Density and Speed (β = 0.27, t = 15.31, $p < .001$), Semantic Density and Hidden Units (β = 0.06, t = 3.30, $p < .001$), and between Phonological Density and Speed (β = 0.12, t = 6.40, $p < .001$). The beneficial effects of high semantic density on comprehension RTs were stronger for networks with fewer hidden units and lower processing speed. Hence, networks with processing advantages due to higher processing capacity or speed benefitted less from high semantic density. The same relationship holds for Phonological Density and Speed.

The interactions we were most interested in are the interactions between Word Frequency and different parameters determining what might be called skill, namely the causes for variation in vocabulary size Corpus Size, Training Trials per word, Hidden Units and Speed.[10] The Word Frequency effect was weaker for the networks trained on the 5000 words corpus (β = 0.13, t = 5.95, $p < .001$), as well as for those with higher numbers of Hidden Units (β = 0.04, t = 2.15, p = .03) and higher Speed (β = 0.05, t = 2.90, p = .004). In addition, Word Frequency effects decreased with an increase in the

[10] See Appendix C for effects plots of all Frequency x Skill interactions for this analysis of comprehension RTs. All plots were created based on data extracted using the effects package (version 4.0.0; Fox et al., 2003) in R.

number of trials per word (β = 0.07, t = 6.31, p < .001). This suggests smaller frequency effects on RTs for networks with higher skills, replicating the frequency x skill interaction, which has previously been reported as an interaction between vocabulary size and word frequency effects.

As indicated before, we ran the same mixed-effects model but this time added vocabulary size as an additional predictor. This changed some of the effects but not all of them. Corpus Size significantly predicted comprehension RTs (β = 3.02, t = 6.08, p < .001) just as before, with the larger corpus being associated with longer RTs. Also the effect of Word Frequency stayed the same (β = -0.23, t = -3.06, p = .002), indicating that high-frequency words elicited faster RTs than low-frequency words. The effects of Training Trials per Word (β = 0.80, t = 2.63, p = .003) and Speed (β = 1.03, t = 3.20, p = .001) were positive, now that vocabulary size was included as predictor. Hence, other than before increased training and higher processing speed were in this model associated with longer RTs. Hidden Units as well as Phonological and Semantic Density did not show any significant effects on comprehension RTs when Vocabulary Size was included as predictor. Vocabulary Size showed a negative effect (β = -1.48, t = -5.96, p < .001); thus, networks with larger vocabularies had faster RTs.

In addition, the results indicated stronger effects of Speed associated with the smaller Corpus Size (β = -1.69, t = -4.28, p < .001), and stronger Phonological (β = -0.15, t = -3.81, p < .001) and Semantic Density (β = -0.24, t = -6.48, p < .001) effects for the networks trained on the 5000 vs. 2500 words corpus. The first interaction is in the opposite direction as compared to the model without Vocabulary size, and the others were not significant in the previous smaller model. Furthermore, the Training Trials per Word effect was again stronger for the slower networks (β = -0.50, t = -2.65, p = .008), although this time the Training per Word effect is a positive one (increasing RTs with increasing number of trials per word). Just as before, the Training per Word effect was stronger for words from semantically less dense neighbourhoods (β = -0.05, t = -3.20, p = .001). Furthermore, the interactions between Hidden Units and Semantic Density (β = -0.09, t = -3.47, p < .001) and Hidden Units and Phonological Density (β = -0.12, t = -3.81, p < .001) showed even stronger beneficial effects of high density neighbourhoods on comprehension RTs for networks with 300 vs. 100 hidden units. Vocabulary Size also interacted with both Semantic (β = 0.14, t = 7.57, p < .001) and Phonological Density (β = 0.07, t = 3.40, p < .001) indicating opposite effects of density depending on the network's vocabulary size. Networks with larger vocabularies showed negative effects of density, i.e. high-density words were associated with slower RTs for these networks, whereas low-vocabulary networks showed beneficial effects of higher neighbourhood density on comprehension RTs. Furthermore, Vocabulary Size interacted with Speed ($\beta = 1.0$, $t = 5.43$, $p < .001$). Networks with lower processing speed showed stronger benefits of having a large vocabulary. Or in other words, networks with processing advantages due to high processing speed showed reduced benefits of having large vocabularies.

The most important interactions were again those between skill or factors determining a network's skill (Corpus Size, Training Trials per word, Hidden Units, Speed, Vocabulary Size) and Word Frequency. The interactions between Corpus Size and Frequency ($\beta = 0.10$, $t = 2.64$, $p = .008$) and between Training Trials per Word and Frequency ($\beta = 0.06$, $t = 3.52$, $p < .001$) were significant. Just as in the model without Vocabulary Size, the word frequency effects were stronger for networks trained on the smaller corpus, and they reduced over training; hence, with increasing trials per word, the frequency effect became weaker. The interactions between Frequency and Hidden Units ($\beta = 0.03$, $t = 0.95$, $p = .34$) and Frequency and Speed ($\beta = 0.03$, $t = 0.91$, $p = .36$) were not significant but when plotting the effects, a tendency towards smaller frequency effects for more skilled networks (as has been shown before) can be seen. The interaction between Vocabulary and Frequency was not significant. Hence, adding Vocabulary Size as predictor, which probably captures a lot of variance that had in the previous model been captured by Hidden Units and Speed, influences not only their effects on comprehension RTs but also the frequency x skill interactions.

When running a model with only Vocabulary Size as predictor in addition to the lexical characteristics Frequency, Phonological and Semantic Density, we observed the same beneficial effects of high Frequency and neighbourhood density as previously (see Appendix C for a table with the model output). Importantly, Vocabulary Size now showed a strong negative effect indicating larger vocabularies being associated with faster RTs ($\beta = -0.25$, $t = -20.60$, $p < .001$), and a frequency x skill interaction ($\beta = 0.05$, $t = 6.08$, $p < .001$) that matched the one observed between Frequency and each Corpus Size, Training Trials per word, Hidden Units, and Speed when Vocabulary Size was not included. Hence, this confirms the impression that Vocabulary Size captures the variance associated with these factors that determine skill. Furthermore, this is in line with what has been shown in the analyses on the causes for variation in vocabulary size, which indicated causal relationships between vocabulary and all of the above-mentioned factors.

Production. Most of the effects on RTs in the production task corresponded to those observed in the analyses on comprehension RTs. All estimates, t-, and p-values for are displayed in a Table in Appendix D, and a short summary of the main effects of all models is provided in the form of a Table in the end of the Results section. Here we only report on the effects that deviate from those observed on comprehension.

In addition to the interactions reported on for comprehension, production RTs were predicted by interactions between Corpus Size and Phonological Density (β = -0.03, t = -3.17, p = .002) and Corpus Size and Semantic Density (β = -0.03, t = -3.87, p < .001). The effects of phonological and semantic density were stronger for networks trained on the 5000 as compared to those trained on the 2500 words corpus. Furthermore, the interaction between Training Trials per Word and Phonological Density observed for comprehension was not present in the analyses on production RTs.

Just as for comprehension, we ran the same analyses with Vocabulary Size as additional predictor to examine whether vocabulary would explain any variance in comprehension RTs if factors closely related to or even causing variation in vocabulary size (see section 3.1), such as Corpus Size, Training Trials per word, Hidden Units, and Speed, are also included in the model. The patterns were again very similar to those observed in the analyses on comprehension RTs. While the effects of Corpus Size (β = 1.87, t = 7.94, p < .001) and Word Frequency (β = -0.24, t = -8.0, p < .001) stayed the same when including Vocabulary Size as an additional predictor, the effects of Training Trials per Word (β = 0.50, t = 3.48, p < .001) and Speed (β = 0.51, t = 4.06, p < .001) changed direction. Again just as for comprehension, Vocabulary Size showed a significant negative effect on production RTs (β = -0.77, t = -5.56, p < .001). Hence, larger vocabularies were associated with faster RTs. Furthermore, just as in the mixed-effects model without Vocabulary Size as predictor, significant effects of Phonological (β = -0.18, t = -5.90, p < .001) and Semantic Density (β = -0.25, t = -8.40, p < .001) indicated that faster RTs were elicited by words from dense neighbourhoods. We observed significant interactions between Corpus Size and each Speed (β = -1.41, t = -10.63, p < .001), Training Trials per Word (β = -0.37, t = -2.14, p = .03), Frequency (β = -0.09, t = -4.18, p < .001), and Phonological (β = -0.08, t = -3.67, p < .001) and Semantic Density (β = -0.09, t = -4.42, p p < .001). Stronger effects of Speed, Frequency, and Phonological and Semantic Density were associated with the larger corpus. The effect of Training Trials per Word was positive for the small and negative for the large corpus; hence, RTs of the networks trained on 5000 words showed a reduction of RTs with an increasing number of trials per word, whereas networks trained on 2500 words showed increasing RTs with training. Training per Word also showed significant interactions with Speed (β = -0.37, t = -4.10, p < .001), Phonological Density (β = -0.02, t = -2.41, p < .02), and Vocabulary Size (β = 0.26, t = 2.61, p = .009). Increasing numbers of trials per word resulted in faster RTs for the networks with higher speed, and in slower RTs for those with low processing speed. In addition, networks with greater vocabularies got slower over training whereas networks with smaller vocabularies showed an increase in speed over training. Finally, the effect of Phonological Density increased over training. Semantic Density ($\beta = 0.09$, $t = 6.89$, $p < .001$) and Phonological Density ($\beta = 0.14$, t = 10.13, $p < .001$) also both interacted with Speed, showing that the beneficial effect of high semantic density on production RTs was stronger for slower networks. In addition, both Phonological ($\beta = 0.02$, $t = 2.62$, $p < .009$) and Semantic Density ($\beta = 0.03$, $t = 3.14$, $p < .002$) showed significant interactions with Vocabulary Size. The beneficial effect of high density neighbourhoods was stronger for larger vocabularies. Finally, Speed and Vocabulary Size interacted ($\beta = 0.49$, $t = 7.44$, $p < .001$), with stronger vocabulary size effects for slower networks. Just as for comprehension, networks with processing advantages due to high processing speed showed reduced benefits of having large vocabularies.

Again, we were most interested in the interactions between Word Frequency and skill (i.e. Corpus Size, Training Trials per word, Hidden Units, Speed, Vocabulary Size). As opposed to the analyses without Vocabulary Size (and to the analyses on comprehension RTs), the effect of Word Frequency was stronger for the 5000 vs. the 2500 words corpus ($\beta = -0.09$, $t = -4.18$, $p < .001$). However, just as in the previous analysis on production RT, the word frequency effect was weaker for faster networks ($\beta = 0.10$, $t = 7.17$, $p < .001$) and reduced with increasing numbers of trials per word ($\beta = 0.03$, $t = 4.06$, $p < .001$). The interaction between Word Frequency and Hidden Units was not significant. Finally, Vocabulary Size and Word Frequency showed a significant interaction ($\beta = 0.07$, $t = 8.15$, $p < .001$), with larger vocabularies being associated with smaller frequency effects. Hence, the typical frequency x skill interaction is present as an interaction between Frequency and different factors determining skill. Again, adding Vocabulary Size as predictor likely affects different variables and interactions because of the close or even causal relationship between the skill-related factors Corpus Size, Training Trials per word, Hidden Units, and Speed and Vocabulary Size.

Just as for comprehension, we ran a mixed-effects model with only Vocabulary Size as predictor in addition to the lexical characteristics Frequency, Phonological and Semantic Density. We again found a significant negative effect of vocabulary size ($\beta = -0.58$, $t = -19.31$, $p < .001$), with larger vocabularies being associated with significantly shorter RTs. In addition, the frequency x skill interaction was again significant ($\beta = 0.07$, $t = 22.41$, $p < .001$), with reduced frequency effects for networks with larger vocabularies. For the estimates, t-, and p-values of all other coefficients, see Appendix D.

Novel word learning

We examined the networks' novel word learning performance, and how the factors causing variation in vocabulary size (Corpus Size, Hidden Units, Speed, prior number of training trials) relate to novel word learning. All above-described and analysed networks performed a novel word learning task. After 250k and 500k training trials all networks were exposed to the same set of 250 words that they had not seen before. This novel word learning task was run in a total of 2500 training trials, words were selected randomly from the novel set with each word presented on average 10 times. The networks had to learn these new form-meaning mappings and were tested on each of them after 2, 6, and 10 trials per novel word. For the purpose of examining novel word learning performance, we analysed networks' vocabulary size at these three points during Training 2. Vocabulary size was again derived from error, i.e. the cosine difference between the model output and all other representations in both the stage 1 training corpus and stage 2 (novel word) training corpus. The model was considered accurate if the error was lowest for the target representation and accuracy was determined as the proportion of words for which the cosine distance between the target and the model's actual output was closest relative to all other words in the corpus. From this we derived the number of words known, i.e. a network's vocabulary size, at a particular point in time during stage 2 training (accuracy x cor).

Mixed-effects models were run on Vocabulary Size as predicted by Training 1 Corpus Size (2500 vs. 5000), Hidden Units (100 vs. 300), Speed (2 vs. 4), Trials per Word during training 1 after which the novel word learning started, and Training 2 Trials per Word (2 vs. 6 vs. 10). As before, all analyses were run on comprehension and production separately. The continuous predictors Training1 and Training 2 Trials per Word were z-transformed using the scale function in R and all other predictors were dummy coded as described before.

We ran additional analyses where we added Vocabulary Size at the end of Training 1 as a continuous (z-transformed) predictor. In this vein, we were able to determine whether having a larger vocabulary would show benefits for novel word learning that are not already captured by the factors Corpus Size, Training Trials per word, Hidden Units, and Speed.

Again, the results of all models are briefly summarised in two Tables in the end of the Results section.

Comprehension. Novel comprehension vocabulary size was predicted by Hidden Units ($\beta = 40.33$, t = 31.51, $p < .001$) and Speed ($\beta = 40.04$, $t = 31.29$, $p < .001$). Networks with greater processing capacities, i.e. 300 as compared to 100 hidden units and higher processing speed learned more words during the novel word learning task. In addition, Training 2 Trials per Word predicted the number of novel words known ($\beta = 17.59$, $t = 312.26$, $p < .001$); more training was associated with larger novel vocabularies. Training 1 Trials per Word at which novel word learning started was significant ($\beta = 6.23$, $t = 83.38$, $p < .001$), with larger novel vocabularies being learned when the networks had more stage 1 training completed before. Finally, networks that had been exposed to the larger corpus in Training 1, learned more novel words in Training 2 ($\beta = 16.17$, $t = 4.17$, $p < .001$).

The significant interaction between Corpus Size and Training 1 Trials per Word ($\beta = 16.53$, $t = 269.26$, $p < .001$) reflects the stronger effect of Trials per Word on the networks exposed to the larger corpus which was reported on before. Hence, due to the fact that Trials per Word has a stronger effect on the learning for the networks exposed to the larger corpus, the effect of Training 1 Trials per Word when the novel word learning is initiated is also stronger on the networks exposed to the 5000 words corpus than on those exposed to 2500 words in Training 1. In addition, the interaction between Corpus Size and Training 2 Trials per Word (β = -2.14, t = -59.99, $p < .001$) indicates weaker effects of Corpus Size from Training 1 with increasing Training 2 Trials per Word. Hence, with increasing training on the novel words, the effect of Training 1 Corpus Size diminished. Furthermore, the interactions between Corpus Size and Hidden Units ($\beta = 3.86$, $t = 2.13$, $p = .03$) and between Corpus Size and Speed ($\beta = 4.44$, $t = 2.45$, $p = .01$) were significant, indicating the beneficial effects of increased processing capacity and speed were even more beneficial for networks exposed to 5000 vs. 2500 words in Training 1. Speed and Training 2 Trials per Word ($\beta = 10.34$, $t = 236.97$, $p < .001$) and Hidden Units and Training 2 Trials per Word ($\beta = 9.69$, t = 221.96, $p < .001$) also showed significant interactions, indicating stronger effects of Training 2 Trials per Word for networks with higher processing speed and greater processing capacity. Hence, the learning rate of faster networks and those with more processing capacity was higher as compared to networks with reduced speed and processing capacity. The interaction between Speed and Training 1 Trials per Word (β = -3.19, t = -53.03, $p < .001$) suggested that for faster networks the time point at which the novel word learning is initiated has a weaker effect, i.e. leads to less of a difference in the number of novel words learned, than for slower networks. Hence, for faster networks, additional exposure to Training 1 items has weaker effects on the number of words learned in Training 2 than for slower networks. Finally, the interaction between Hidden Units and Training 1 Trials per Word ($\beta = 1.66$, $t = 27.67$, $p < .001$) turned out significant, showing that the starting point of Training 2 has stronger effects on the number of words learned for networks with 300 vs. 100 hidden units.

As indicated before, we repeated the same analyses, the only difference being that we added Vocabulary Size after Training 1 as an additional predictor. The results show the same positive effects of Training 1 Trials per Word ($\beta = 0.52$, $t = 2.43$, $p = .02$) and Training 2 Trials per Word ($\beta = 20.57$, $t = 246.66$, $p < .001$), Hidden Units ($\beta = 29.57$, t = 20.06, $p < .001$), and Speed ($\beta = 19.08$, $t = 11.07$, $p < .001$) as the analyses without Vocabulary Size as predictor. Furthermore, all interactions between these factors were the same as in the previously reported analyses (full table in Appendix E). In addition, Vocabulary Size predicted novel word learning (β = 17.81, t = 18.52, $p < .001$), with larger vocabularies being associated with better novel word learning performance.

However, other than in the analyses reported on before, the effect of Corpus Size on number of words learned in Training 2 turned out negative (β = -20.62, t = -4.98, $p < .001$), indicating that networks exposed to a larger corpus in Training 1 learn fewer words in Training 2. Additionally, Corpus Size interacted with Vocabulary Size (β = -21.77, t = -17.04, $p < .001$), showing that for the networks exposed to 5000 words in Training 1 the vocabulary size effect is much weaker (or almost reversed, hence negative) as compared to networks exposed to 2500 words in Training 1. Hence, the unexpected effect of Corpus Size appears to be due to the additional predictor Vocabulary Size, and its interaction with Corpus Size. The interaction between Vocabulary Size and Speed (β = -2.73, t = -10.61, $p < .001$) indicated stronger beneficial effects of Vocabulary on novel word learning for networks with reduced speed and processing capacities. Hence, networks that have processing advantages due to increased speed show less additional benefits due to having larger vocabularies. Furthermore, Vocabulary Size and Training 2 Trials per Word interacted (β = 1.34, t = 47.49, $p < .001$) showing increasing effects of Vocabulary Size with increasing training on the novel words. All other interactions stay the same as in the analyses without Vocabulary Size.

To sum up, adding Vocabulary Size as a predictor only affects the effect Corpus Size, probably due to an interaction between Corpus Size and Vocabulary Size. All other effects stay the same. Hence, despite the fact that factors causing variation in Vocabulary Size (Hidden Units, Speed) are included in the analyses, Vocabulary Size appears to have additional beneficial effects on novel word learning, independent of these underlying factors.

Production. The results of the mixed-effects mode on novel production vocabulary size without vocabulary size as predictor were exactly the same as those of the analyses on novel comprehension vocabulary size. Therefore the estimates, t-, and p-values of all predictors can be found in Appendix F. A summary of the main effects can be found in Table 6.13 in the end of the Results section.

A second mixed-effects model was run including Vocabulary Size as additional predictor. All effects stayed exactly the same as in the analysis without Vocabulary Size as predictor (see Appendix F for a full table). Vocabulary Size significantly predicted novel word learning (β = -3.31, t = -3.10, p = .002) with larger Training 1 vocabularies being associated with fewer words learned during Training 2. Thus, other than for comprehension vocabulary knowledge does not seem to benefit the learning of novel production vocabulary.[11] In addition, Vocabulary Size showed a significant interaction with Speed (β = -4.50, t = -12.79, $p < .001$), indicating stronger negative effects of Vocabulary Size on networks with higher processing speed. Finally, Vocabulary Size and Training 2 Trials per Word interacted ($\beta = 1.18$, $t = 32.57$, $p < .001$), indicating that the negative effects of vocabulary size reduce with increasing training on the novel words.

[11] Just as in the RT analyses, we ran an additional mixed-effects model with only Training 1 Vocabulary Size and Training 2 Trials per Word as predictors. Hence, we did not include the factors that had been identified as causes for variation in vocabulary size to examine whether the inclusion of different closely related effects might be the origin of the unexpected negative vocabulary effect. This analysis showed strong positive effects of both Training 2 Trials per Word ($\beta = 36.06$, $t = 1524.9$, $p < .001$) and Vocabulary Size ($\beta = 12.81$, $t = 287.85$, $p < .001$) on the number of words learned in the novel word learning task, as well as a significant interaction between the two effects ($\beta = 0.78$, $t = 33.07$, $p < .001$), similar to the effects observed for comprehension. Hence, when only including Vocabulary Size as factor determining skill, the effect of Vocabulary is similar to that of processing speed and capacity in the models that did not include Vocabulary Size. Including all factors that determine skill and likely share considerable variance in novel word learning performance, appears to affect the outcome. Overall, more skilled networks (whether skill is determined based on our manipulations or the number of known words) show stronger novel word learning performance.

Table 6.12: The effects of all manipulations on the three dependent measures. All effects are the same for comprehension and production, and based on models without vocabulary size as a predictor.

Factor	Dependent measures		
	Vocabulary size	**Reaction times**	**Novel word learning**
Corpus size (input quality)	Larger corpus → larger vocabulary → higher learning rate	Larger corpus → slower RTs	Larger corpus → better novel word learning
Trials per word in stage 1 (input quantity)	More exposures → larger vocabulary → higher learning rate	More exposures → faster RTs	More training 1 exposure → better novel word learning
Trials per word in stage 2			More exposures → better novel word learning
Hidden units (resources)	More resources → larger vocabulary → higher learning rate	More resources → faster RTs	More resources → better novel word learning
Speed	Higher speed → larger vocabulary → higher learning rate	Higher speed → faster RTs	Higher speed → better novel word learning
Word frequency		High-frequency → faster RTs	
Phonological density		High density → faster RTs	
Semantic density		High density → faster RTs	

Table 6.13: The effects of the manipulations on RTs and novel word learning performance, when including Vocabulary Size as a predictor in the models. Unless stated otherwise, all effects are the same for comprehension and production.

Factor	Dependent measures	
	Reaction times	**Novel word learning**
Corpus size (input quality)	Larger corpus → slower RTs	Larger corpus → decreased novel word learning
Trials per word in stage 1 (input quantity)	More exposures → slower RTs	More training 1 exposure → better novel word learning
Trials per word in stage 2		More exposures → better novel word learning
Hidden units (resources)		More resources → better novel word learning
Speed	Higher speed → slower RTs	Higher speed → better novel word learning
Vocabulary size	Larger vocabulary → faster RTs	Larger vocabulary → Comprehension: better novel word learning → Production: decreased novel word learning
Word frequency	High-frequency → faster RTs	
Phonological density	High density → faster production RTs	
Semantic density	High density → faster production RTs	

Discussion

The present study was the first to model explicitly the relationships between four distinct environmental and cognitive factors suggested to cause variation in vocabulary learning and size. We examined the roles of two environmental factors, namely (a) amount or quantity of exposure, i.e. the number of training trials per word, and (b) the diversity of the input, i.e. the number of types (2500 vs. 5000). In addition, we manipulated the cognitive factors (c) processing speed and (d) computational resources, with the latter being assumed to simulate variation in general intelligence. We demonstrated that each of these factors influenced vocabulary size and rate of word learning, with increased exposure and diversity of the input as well as higher processing speed and intelligence being associated with larger vocabularies and higher learning rates (see Figures in Analyses and Results section and Appendix A).

Notably, we examined two time points during training, namely after 250k training trials and after 500k training trials. We analysed the network's performance after it had acquired substantial knowledge of its linguistic environment, therefore avoiding behaviour limited to early stages of development. The selected time points also occurred before learning reached asymptote, hence before the number of words they knew was limited by either the number of words in the training corpus or capacity limitations. Hence, choosing the two time points for the analyses limited asymptotic effects on vocabulary size and learning rate and with that effects of over-training and capacity limitations. We thus simulated adult performance. While asymptotic effects cannot be excluded completely, it was assumed that small effects are valid because human learners might also be assumed to gain less from additional exposure once a certain amount of training is reached (for example in the case of high-frequency words). Finally, although developmental effects could be analysed using this model, this is beyond scope of current study. We believe, however, that many of our conclusions extend to early stages of development. One exception is likely to be the effect of corpus size, where at very early stages of training a limited set may be advantageous (e.g. Jones & Rowland, 2017).

Variation in vocabulary size

The first set of analyses was run to examine the factors causing variation in the number of words learned by the model. We found all factors that we manipulated affected the network's vocabulary size. An increase in exposure, i.e. training trials per word, was associated with larger vocabularies, as has been shown in earlier behavioural and computational studies (Hurtado et al., 2008; Monaghan et al., 2017, see exposure manipulation in Chapter 5). Furthermore, networks that were exposed to the larger 5000 words corpus, hence those that received more diverse input, ended up with larger vocabularies than networks exposed to 2500 words. In addition, the two environmental factors amount of exposure (i.e. input quantity) and diversity (i.e. input quality) showed an interaction, indicating that the diversity of the input affects learning rate. Networks exposed to the large corpus, i.e. more types and therefore more diverse input, showed higher learning rates than those exposed to the small corpus. This is in line with previous research by Jones and Rowland (2017) indicating that increased diversity of the input has beneficial effects on vocabulary learning later in development although mechanisms underlying this shared property of the two models differ (see later discussion on the diversity of input mechanisms).

As indicated before, not only the environmental but also the two cognitive factors we manipulated in the model were shown to predict vocabulary size. Networks with higher processing speed and more processing resources, i.e. higher intelligence, were found to develop larger vocabularies than those with reduced speed and resources. Furthermore, both processing speed and resources, i.e. general intelligence, showed interactions with the environmental factor exposure, i.e. training trials per word. Hence, both cognitive factors generated not only greater vocabularies but also an increased word learning rate. This corresponds to earlier findings from behavioural as well as computational investigations on the relationship between processing speed and word learning. Higher (language) processing speed has been associated with improved vocabulary learning and size in children (Fernald et al., 2006; Marchman & Fernald, 2008). Consistent with previous modelling of effects of variation in processing speed (McMurray et al., 2012) our simulations suggest that variation in general processing speed might indeed be a cause of individual differences in word learning and knowledge, although the mechanisms used to generate such effects were distinct (see later discussion of cognitive mechanisms).

Furthermore, processing speed and resources also showed interactions with the second environmental factor, i.e. input diversity. The effects of the two cognitive factors were stronger for networks trained on more types, i.e. more diverse input, than for those trained on the smaller corpus. This is a novel finding concerning the relationship between cognitive and environmental factors causing variation in vocabulary learning and size.

Properties determining ease of word learning

Analyses performed on the properties of words known by each network have shown that certain lexical characteristics make some words easier to learn for the network than others. The results of these analyses are supported by the RT analyses, which are discussed below. As would be expected, high-frequency words are learned more easily than low-frequency words, due to increased training on such items (more tokens). This replicates findings from earlier behavioural as well as computational modelling studies, indicating that words which a learner is exposed to more often are learned more easily and faster (Harm & Seidenberg, 2004; Monaghan et al., 2017, see effect of exposure in Chapter 5). This is also in line with the observations made in the behavioural experiment present in Chapter 5, where words that participants had seen more often during training were learned better than low-exposure words.

Furthermore, the network learned words with phonologically dense neighbourhoods more easily than those with sparse neighbourhoods, which also replicates earlier findings. Storkel and Adlof (2009), for example, found that children learn novel words from dense phonological neighbourhoods more easily than words from sparse phonological neighbourhoods (see also Storkel, 2009). Similar to the observations concerning phonological neighbourhood density, words with dense semantic neighbourhoods were learned more easily by the network than words with semantically less dense neighbourhoods. There are a number of studies showing this beneficial effect of high semantic density on lexical learning (Hills, Maouene, Maouene, Sheya, & Smith, 2009; Sailor, 2013; Steyvers & Tenenbaum, 2005), while the picture is more mixed than in the case of phonological density (e.g. Sahni, 2010).

Although, as mentioned above there is behavioural evidence that mirrors the model's behaviour that increased neighbourhood density is beneficial for learning, this analysis of the properties of 'known' words serves a further purpose of establishing which characteristics make words easy to learn within the model used in this study.

Aside from word frequency, which has been shown to affect word learning and processing in emergent connectionist models of the type used in this study (e.g. Harm & Seidenberg, 2004; Monaghan et al., 2017), high phonological and semantic neighbourhood density were shown to be advantageous for word learning. The reason is that representations with dense phonological/semantic neighbourhoods, i.e. those that are similar to many other words in the input, require smaller changes to the representational space in order to be accommodated, i.e. learned, by the network. This makes learning phonological and semantic representations with many neighbours easier than learning items from sparse neighbourhoods, which thus require larger changes to the weight structure to be represented and learned. Although this mechanism is valid for generating the density relationships observed, it is important to acknowledge that this belies the complexity of the representational factors affecting word learning ease and success in the current model. As a result of the rich representational structure implemented in the current model, complex interactions between the representational structure both within phonological and semantic domains, and across modalities influence learning. Thus, further analyses are required in order to offer a comprehensive description of how mechanisms dependent on neighbourhood density operate within the current model.

Reaction times

Reaction times derived from the model record the time step at which the cosine distance between the model output and the target is smallest relative to all other representations in the corpus. Within the model, RT is dependent on the fidelity of a network's representations, as with richer representations the network is quicker to generate properties of the targets representation that distinguish it from competitors. Thus, RTs reflect the model's knowledge of a given word and are assumed to display a similar non-linear decreasing function with respect to training as observed for task error in Figure 6.3.

Increased training per word, greater processing speed, and increased computational resources generated faster RTs. The same factors were also shown to cause variation in vocabulary size within the model. This is in line with previous behavioural research indicating a close relationship between language processing speed and vocabulary size (Fernald et al., 2006).

Our results suggest that the mediating or circular relationship between vocabulary size and language processing speed, as has been observed by Fernald and colleagues (2006) and simulated by McMurray and colleagues (2012), might at least in part be due to individual differences in underlying cognitive and environmental factors that benefit both vocabulary size and language processing speed. Fernald and colleagues (2006; see also Marchman & Fernald, 2008) have suggested two explanations for the relationship observed between word knowledge and language processing, one being that differences in exposure very early in development lead to variation in early vocabulary size and that this benefits language processing and therewith subsequent word learning. This mechanism is captured by our modelling results, where increased exposure (number of tokens) leads to both increased vocabulary learning (size of vocabulary) and language processing speed (RTs). The other proposal was that variation in cognitive skills, i.e. factors independent of input, is behind variation observed both in language processing speed and word learning but is not limited to the language domain (Fernald et al., 2006). Again our simulations capture such a relationship both increased general processing speed and processing resources lead to both increased vocabulary size and faster language processing (RTs). Hence, our simulations indicate that the two hypotheses (cognitive and environmental) about the origins of the relationship between vocabulary size and processing speed put forward by Fernald et al. (2006) are not mutually exclusive but that likely a complex relationship between external and internal sources of variation results in differences in vocabulary size and speed of lexical processing.

Furthermore, networks were faster at processing high-frequency words and words with phonologically and semantically dense neighbourhoods. Hence, the lexical characteristics that have been shown to have beneficial effects on word learning (see above) were also found to be advantageous for lexical processing. These properties of our simulations replicate the typical effects of word frequency (Brysbaert et al., 2018) as well as phonological (Vitevitch, 2002; Vitevitch & Sommers, 2003; Yates, Locker, & Simpson, 2004) and semantic density (Yap, Lim, & Pexman, 2015; Yap, Tan, Pexman, & Hargreaves, 2011) on RTs, with faster RTs for high-frequency words and those with denser neighbourhoods. Thus, words with higher frequencies and denser phonological and semantic neighbourhoods were easier to learn and also easier (i.e. faster) to process for our model. This makes sense given the close relationship between fidelity of representations and RTs; properties that allow information about a word to be learnt more rapidly will result in richer representation and hence faster RTs.

Greater input diversity, by contrast, had negative effects on the model's lexical processing speed. Networks that had been exposed to the larger corpus, i.e. a larger number of distinct types, showed slower comprehension and production RTs. One factor underlying this behaviour is that RTs are calculated relative to all words in the corpus, namely as the point in time when the distance between model's production and the target is smaller relative to all other words in the corpus. Due to the fact that the larger 5000 words corpus contains more words that are closer to the given target, there are more words between which the network needs to distinguish and therefore RTs are likely higher. While this is an inherent property of the larger corpus and RTs associated with it within the model, it is a potentially valid distinction, for if you know more words the target must be distinguished from larger numbers of other known words.

It has been shown that larger vocabularies are associated with faster lexical decision RTs (see Chapters 2 and 3), which may appear at odds to predictions based on the observation that larger corpora generate slower RTs. However, this data does not necessarily contradict the model for as the model demonstrates many underlying factors determine vocabulary size. In addition, different causes underlying variation in vocabulary size might be responsible for faster RTs being associated with larger vocabularies instead of vocabulary size itself driving this effect. See below for a more detailed discussion of that possibility.

We were particularly interested in the frequency x skill interaction, which has been observed in previous studies. Word frequency effects have been reported to be smaller for individuals with larger vocabularies (Brysbaert et al., 2016a; Diependaele et al., 2013; Yap et al., 2009). As stated in the lexical entrenchment hypothesis, this has been suggested to be due to more entrenched or robust representations in individuals with larger vocabularies, which are faster to be accessed than those of individuals with smaller vocabularies. Differences in exposure have been argued to be the main reason for these differences in sensitivity to word frequency between speakers with smaller vs. larger vocabularies (Kuperman & Van Dyke, 2013; Monaghan et al., 2017). Monaghan and colleagues (2017) demonstrated that in their model increased exposure is generates an increase in the efficiency of mappings. They demonstrate that the non-linear relationship between increased training (exposure) and increased knowledge of a given item (task error) means that additional training on an item leads to increasingly weaker improvements to stored representations (less additional knowledge or smaller change in task error). It is such a mechanism that Monaghan et al. (2017) demonstrate to generate the frequently observed interaction between frequency and vocabulary size. As previously discussed they argue that increased training (exposure/number of tokens) is the underlying factor generating both increased vocabulary size and weaker effect of frequency on processing.

Interestingly, we found two-way interactions between word frequency and all environmental and cognitive factors that we previously demonstrate to cause variation in vocabulary size, namely number of of training trials per word (number of tokens), diversity of input (number of types), processing speed, and processing resources (i.e. general intelligence). Thus, all of the factors that have been shown to determine the networks' skill, i.e. vocabulary size, generated the frequency x skill interaction, with higher skill networks displaying a reduced word frequency effect.

Furthermore, we observed similar interactions with the two density measures, namely phonological and semantic neighbourhood density. Thus, what has previously found to be an interaction between skill, more precisely vocabulary size, and the word frequency effect might – based on the present model – be extended to interactions between other lexical characteristics and other cognitive and environmental factors that influence learning and therefore 'skill'. Ultimately, our simulations predict that internal properties of the network or properties of the learning environment that enable a network to learn faster interact with lexical properties that make certain words easier to learn.

In each of these cases, the relationship between increased skill (i.e. processing speed, resources, and exposure [number of training trials per word]) and weaker effects of the lexical characteristics of words (i.e. frequency and density) on processing is generated by non-linear relationships similar to those between learning and training reported in Monaghan et al. (2017). As training increases the rate of learning, i.e. the rate of error reduction, decreases (see Figure 6.3 in the Analyses and Results section), the fidelity of representations plateaus. Thus, any property of networks that affects the rate of learning, i.e. the rate of decent along this function, will interact with properties of words that affect learning. This is because the difference in fidelity of representations (affected by a given word level characteristic) for networks ahead in their training, (i.e. those with higher learning rates therefore further down on the learning curve) will thus be smaller. Therefore, as shown in our simulations, such networks will display smaller differences in performance on sets of word that differ in such word level characteristics.

As indicated before, the frequency x skill interaction has been taken to suggest that the representations in individuals with larger vocabularies are more robust or more entrenched than those in individuals with smaller vocabularies, leading to reduced sensitivity to lexical characteristics (Brysbaert et al., 2016a; Diependaele et al., 2013). The findings of the present study confirm this. A weaker effect of lexical characteristics indicates, as described above, that the network is further down in the error curve, thus, has more developed representations. A further examination of the model may offer a means of describing explicitly in what way representations differ between high-vocabulary vs. low-vocabulary networks, in what respect are representations more entrenched (Diependaele et al., 2013) or more robust (Yap et al., 2009), and whether there are distinctions in such properties due to differing causes of variation in vocabulary size.

The interaction we observed between Corpus Size, i.e. input diversity, and word frequency potentially suggests a different origin of the effect. The RTs were found to be overall slower for networks exposed to more diverse input however the frequency x skill interaction was still present. As previously described, slower RTs likely result from RTs being dependent on a network's ability to differentiate a target from all other words in the corpus, i.e. all other words to which it is exposed. This leads to an overall increase in RTs for networks exposed to more diverse input. It is possible this also has consequences for the frequency by corpus size interaction. However, an alternative explanation is that the representation of low-frequency words, may gain additional benefit from knowledge of more tokens relative to high-frequency words. One possible mechanism that may generate such beneficial effects for larger corpora would be that additional training on similar mappings or representations leads to faster activation of a given low-frequency mapping. Further analyses of the corpora of different sizes and the emergent representations developed within the model are required to understand this intriguing interaction.

In summary, we replicated the frequency x skill interaction that has previously been found in behavioural and computational modelling studies and extend them to other lexical characteristics that affect learning (Brysbaert et al., 2016a; Monaghan et al., 2017). Additionally, our findings confirm suggestions made earlier that the frequency x skill interaction indicates differences in how far developed, i.e. robust (Yap et al., 2009) or entrenched (Diependaele et al., 2013) the network's representations are. Furthermore, our findings extend the conclusions from earlier investigations because we did not only show that vocabulary size is multiply determined by both environmental and cognitive factors, but also that all of these factors show an interaction with word frequency effects. This adds further support to arguments that the factor vocabulary size captures individual variation in multiple underlying environmental and cognitive components. The fact that processing speed and resources as well as exposure interact with frequency, in turn, adds weight to their potential role as underlying variation in vocabulary size (see above). Hence, it is indicated that not only variation in exposure leads to individual differences in vocabulary size and differences in the sensitivity to effects of word frequency (see Monaghan et al., 2017), but that complex relationships between individual differences in cognitive and environmental effects results in the observed individual differences in vocabulary size and the frequency x skill interaction. We acknowledge that, as demonstrated in Monaghan et al. (2017), differences in exposure alone potentially explain the observed interaction between frequency and skill. However, given the extensive variation in learners not only in environmental factors but also in cognitive skills, it is important to aim at explaining observed effects in their full complexity. Part of this complexity is to take individual variation in cognitive abilities into account as potentially interacting with effects of exposure and affecting the learning of words with different degrees of difficulty.

Novel word learning

Results of novel word learning simulations replicate many of the relationships described above relating to environmental and cognitive effects on learning. We found that networks with higher processing speed, greater processing resources, increased quantity of exposure (number of tokens) and diversity of exposure (number of types) learned more words over the course of Training 2. Furthermore, networks with higher speed and greater resources learned more words at higher learning rates than their less skilled counterparts. Our novel word learning simulations aimed to simulate learning of novel words in adulthood, hence simulations occurred once networks had acquired substantial knowledge of their linguistic environment, not at early stages of language development. The simulations thus indicate that cognitive and environmental factors that affect vocabulary size are likely to continue to influence word learning throughout the lifespan. This is in line with findings from behavioural research conducted in adult populations (see Chapter 5), which has indicated positive relationships between novel word learning and increased vocabulary size. The reason is that our earlier simulations suggested measures of vocabulary size to capture various underlying factors (environmental and cognitive). Hence, the beneficial effects of higher processing speed, greater resources (i.e. intelligence), and increased exposure (number of tokens and number of types) on novel word learning in mature networks displayed by the model are consistent with and might underlie the vocabulary size advantages observed in experimental research.

Novel word simulations were also run to provide further insight into the mechanisms that may underlie relationships between each of the factors manipulated and vocabulary size by allowing a separation between learning environments. We discuss such insights in the following sections of the discussion.

Cognitive factors

Our analyses of networks varying in different cognitive factors demonstrated that higher processing speed and greater processing resources accelerate learning. Hence, faster networks and those with more processing resources showed larger vocabularies as well as faster learning rates, as indicated above. This was true for Training 1 on corpora of 2500 or 5000 words, as well as for Training 2, where all networks were exposed to an additional set of 250 novel words. In addition, both cognitive factors interacted with the diversity of the input, indicating that the advantages associated with improved cognitive skills were stronger for those networks trained on the larger corpus (see above). One reason behind this is that the networks exposed to more diverse input, hence to the larger corpus, are presented with more words to learn from. Therefore, processing advantages due to higher processing speed or more resources leads to faster learning, such that more words breach the 'knowledge' criteria threshold sooner. Thus, additional potential benefits of learning from more diverse input have an impact on learning more rapidly in skilled than in less skilled networks (such potential benefits are discussed in the diversity of input section of the discussion). Overall, many of the effects of improved cognitive skills mirrored those of increased exposure, thus additional learning.

Analyses of the words known as well as the RTs for known words at 250k and 500k training trials indicated that more skilled networks learned larger numbers of more difficult words than their less skilled counterparts. More precisely, the words known by networks with increased speed and processing resources showed lower mean word frequencies and overall lower phonological and semantic neighbourhood densities. These observations provided some insights into what the mechanisms underlying the cognitive advantages might be. These properties of the acquired words demonstrate that our implemented mechanisms for varying the speed of information processing and the processing resources available to a learner lead to a lowering of the threshold on the amount of exposure required to learn a word of given complexity (e.g. semantic or phonological).

The model's general speed of processing information was manipulated through changing the integration constant within each time averaged input processing unit of the network. This was assumed to be an appropriate means of simulating differences in general processing speed as the integration constant controls the speed with which input activation in a given unit ramps up, and thus the rate at which information can propagate through the network. Such an implementation thus manipulates speed directly, with weaker effects on the computational properties of the network than, for example, a manipulation of input gain would have (Plaut et al., 1996) or a range of hyperparameters (temperature, inhibition, learning rate: McMurray et al. 2012) that likely have many additional consequences for network behaviour. Our manipulation of speed in relation to effects on vocabulary development was in this way distinct from previous studies (e.g. hebbian-learning algorithm: McMurray et al. 2012) yet replicated the same behavioural effects, i.e. relationships between RTs, vocabulary size, rate of vocabulary development. As indicated by Figure 6.2 (Methods section) a larger integration constant means that a network can react quicker to information available in the learning environment. This implementation of the mechanism underlying relationships between general processing speed and vocabulary size proposes that it is thus possible for faster processors to extract more information from a given period of exposure, thus extracting more information from which they are able to learn. This is in line with the behavioural word learning study presented in Chapter 5, which implied that – although effects may be small – higher processing speed is associated with reduced sensitivity to differences in the number of exposures to novel words. Further, online language processing speed, as it was measured in the behavioural studies of Fernald et al. (2006; see also Marchman & Fernald, 2008), might show a mutual relationship with vocabulary growth because online language processing skill might improve with increased training on linguistic input , hence with increased vocabulary size (as seen in the relationship between iteration, vocabulary size and RTs within the current model). However, our simulations like those of McMurray et al. (2012) also demonstrate that such a relationship may exist due to underlying initial differences in general processing speed, that generate increased vocabulary size, increased language speed of processing and increased rates of word learning.

The second cognitive factor simulated in this study that, as our results show, also offers explanation for such relationships observed between vocabulary size, speed of language processing and rate of word learning, is variation in processing resources. Although this mechanisms is distinct from that of processing speed, it generates many similar consequences i.e. faster RTs, frequency x skill interactions, faster rate of word learning, larger vocabulary size. Increasing the number of hidden units (processing resources) within a network affects learning and processing as it alters the size of the representational space into which representations can be projected and thus the richness of internal representations and the level of abstraction at which networks able to develop sensitivity.

This mechanism, implemented within our model, aims to represent variation in computational resources at bottlenecks within cognitive system. We argue that such a mechanism is likely to be one of many that underlie variation in performance on tasks that aim to measure variation in general intelligence, and thus generate relationships between vocabulary size and general intelligence.

There may of course be alternative ways of implementing differences in intelligence, one being the introduction of different degrees of noise across all connections in the network, which would affect variation in general processing efficiency. However, pilot studies showed that even very small amounts of noise introduced across all connections in the network had detrimental effects on the model's learning performance, leading not only to changes in learning rate but to a significant limitation of overall learning ability. It is possible however that an appropriate parameter range could be found that would not lead to such catastrophic effects on performance.

Variation in general intelligence has, to our knowledge, not previously been implemented to model variation in vocabulary size explicitly but there are potentially many ways of doing that. Our results support the argument that variation in processing resources may be a viable mechanism underlying this relationship. Future research is needed to further examine whether and how manipulations of connection noise or alternative parameters can be used to simulate variation in general intelligence, and whether these different implementations have distinct effects on learning and processing.

Finally, the only difference observed between the two cognitive factors manipulated concerned their interactions with Training 1 Trials per Word in the novel word learning task. Hence, the point in time when Training 2 started showed different interactive effects with processing speed vs. resources. The effect of increased Training 1 showed weaker effects for faster than for slower networks, but stronger effects for networks with more vs. less processing resources. This might suggest that additional resources do not simply accelerate learning as in the case of increased processing speed, but also allow for greater learning capacity. A further examination of differences in the emergent representational structure of networks with higher processing speed vs. greater resources is required to further understand the mechanism underlying this observed difference in model behaviour.

In summary, an increase in both the amount of computational resources, i.e. intelligence, and general processing speed are associated with greater vocabulary size and faster learning rates. Both cognitive factors showed similar effects on learning in Training 1 as well as in the novel word learning task. The model presented offers an explicit description of multiple distinct mechanisms that potentially underlie the beneficial effects of improved cognitive abilities on lexical learning and language processing. As indicated before, this is in line with previous research showing beneficial effects of processing speed on vocabulary learning and size (Fernald et al., 2006; McMurray et al., 2012). Further, given our simulations aim to simulate effects present within adult populations, results of our simulations are consistent with both evidence for the effects of individual differences in cognitive abilities in children as well as across the lifespan (Kidd et al., 2018). This indicates that it is not only exposure that leads to such variation in vocabulary learning, knowledge, and processing but that both environmental and internal sources for variation interact in determining learning and processing performance.

Diversity of the input

Our analyses shed light on multiple mechanisms potentially underlying the beneficial effects of increased input diversity, i.e. larger number of types, on word learning. First, one consequence of networks being exposed to the larger corpus, is that they were presented with a higher number of easy to learn words than were present in the smaller corpus. More precisely, the number of high-frequency words and words with high phonological or semantic neighbourhood densities was larger in the corpus consisting of a higher number of distinct types. This was a necessary consequence of randomly sampling words to form corpora of different sizes and is reflected in the analyses of the words known by networks trained on either small or large corpora performed at 250k and 500k training trials. These properties of the input highlight three distinct mechanisms that generate beneficial effects of greater input diversity: word frequency, phonological neighbourhood density, and semantic neighbourhood density. More precisely greater diversity of input leads to exposure to a higher number of high-frequency and -density words. Instead of a change to the learning threshold (number of training trials network needs to learn a given word), which was associated with improved cognitive skills, the mechanism in this case is that there are a larger number of easier to learn words (i.e. high-frequency and high-density words that require fewer training trials to learn) in the more diverse input. Thus, the larger, i.e. more diverse, corpus comprised more words that require less training to breach the learning ('known' word) threshold.

Further, such advantages are likely to extend to other properties of the representational structure both within and across modalities that affect learning. For example, larger corpora are also likely to contain more words with easier to learn cross-modal mappings, i.e. mappings that follow systematic relationships between phonology and semantics. However, such relationships are yet to be analysed and thus demonstrated within the current model.

The observation that larger numbers of easier to learn words exist in the larger corpus is closely related to basic assumptions about the characteristics of corpora of different sizes.

Within the current study, to limit implementation of assumptions regarding the factors that generate variation in the diversity of input, we simply randomly sampled from a larger corpus (5641 words) to generate a smaller sets of different sizes (2500 and 5000 words). It is however likely that many complex factors play a role in determining differences in the distribution of word properties (i.e. frequency, semantic density, phonological density) between populations that differ in the diversity of their linguistic input.[12] Investigating within this model the effects of such factors on the distribution of properties that affect word learning and their consequences for both vocabulary size and broader aspects of word processing is beyond the scope of the current study.

Irrespective of sampling method or assumptions about different distributions in corpora of varying sizes, it is safe to assume that increasing the number of types will increase the number 'easy to learn' words, i.e. words with sufficiently high frequency and/or density to breach the learning threshold. Our modelling highlights that such mechanisms may potentially play a major role in generating variation in vocabulary size. More research is needed, though, to describe how the distributions of lexical properties that affect word learning differ between populations that differ in their exposure to language. One important factor to be taken into account in this context is for instance literacy, i.e. reading and exposure to written text, which has been shown to both influence the lexical characteristics of the input (Cunningham, 2005) and accelerate language learning (Nagy, Herman, & Anderson, 1985; Sternberg, 1987). Effects related to literacy could be modelled in the current framework in order to assess the impact on word learning, vocabulary size, and broader aspects of language processing.

Novel word learning simulations also show a beneficial effect of increased diversity of exposure on word learning. In stage 2 (novel word) training all networks are trained on the same set of 250 words, therefore the above mechanisms are unable to explain this advantage. Instead a further mechanism must be at work. Knowledge the network has already acquired from diverse exposure in stage 1 training affects the ease of learning new words, thus increased diversity of input generates network characteristics that benefit

[12]Notably, it has been shown that increasing sample size (when sampling from a larger corpus as we did to create the training corpora) has particularly strong effects on the percentage of low-frequency words in a sample and less strong effects on the percentage of high-frequency words (Baayen, 2001; Kuperman & Van Dyke, 2013). Thus, increasing the sample size has an impact in particular on the relative number of low-frequency words, which leads to an increasing sample size being associated with decreasing ratios of relative frequencies between low- and high-frequency words (Kuperman & Van Dyke, 2013). Thus, we acknowledge that differences in the relative frequencies might also exist within the corpora of different sizes in our study, though are probably smaller in size due to the smaller corpus size difference than the one described in Kuperman & van Dyke (2013). The observed advantages of the larger corpus size are, however, likely to be due to the absolute number of high-frequency words, which is higher in the 5000 than the 2500 words corpus, affecting the number of words that are learned more easily. We cannot make claims about consequences of differences in the percentage of low- vs. high-frequency words in the corpora of different size.

novel word learning. Our proposal for this mechanism is as follows. In the current model, all words to which the network is exposed have an impact on the representational space. Hence, the more diverse the input to which a network is exposed, the more likely it is that the network already possesses a representation or a mapping that is similar to that of the novel word. Thus, smaller changes are required to the weight structure in order to accommodate the novel word (this is similar to the mechanism described above that makes high density words easier to learn). This is potentially driving the beneficial effects observed in the novel word learning simulations, where the greatest benefit of greater input diversity during Training 1 was found at early stages of Training 2, hence when the structure of the Training 1 representations were able to exert the greatest influence on novel word learning. This mechanism overlaps with others in the literature that propose beneficial effects of increased network size (Steyvers & Tenenbaum, 2005; Sailor, 2013). This might therefore also be an explanation for why including vocabulary size in the analyses on novel word learning performance appears to absorb much of the variance that was previously, i.e. without vocabulary size as predictor, captured by differences in corpus size, i.e. diversity of the input. Further, analysis of the model is required to test this hypothesis.

In summary, our simulations of the effects of quality of exposure (number of types) on vocabulary size and language processing more generally replicate a number of observed relationships between vocabulary size and language processing (RTs, frequency effects, novel word learning). As discussed above the model offers an explicit description of multiple mechanisms driven by increased diversity (number of types) that potentially explain such observed effects. Ours is not the first computational modelling investigation of such relationships, Jones and Rowland (2017) also demonstrate that advantages of increased lexical diversity emerge when a simple (chunk-based) learning mechanism is applied to learn phonological representations. Our simulations, however, build on this study, by offering many additional mechanisms that arise from a more detailed description of the processes involved. More specifically, our model includes both semantic and phonological components of this task. In doing so we are able to demonstrate that increased diversity increases the number of words with favourable characteristics (high frequency, high semantic density, high phonological density) to which a system is exposed. Further, exposure to more diverse input affects the internal emergent structure of the network (in phonological, semantic and cross-modal dimensions) in a manner that is beneficial to novel word learning.

Vocabulary size

Our simulations demonstrate that many underlying causes of variation in vocabulary size share beneficial consequences for both language learning and processing (e.g. speed of processing, rate of word learning). Therefore it is possible to explain positive relationships between increased vocabulary size and language learning and processing as consequences of shared underlying factors. It is also possible however, that variance explained in language learning and processing is attributable to mechanisms that are driven purely by variation in vocabulary size, i.e. there exist mechanisms that generate beneficial effects (e.g. increased speed of processing, or rate of word learning) attributable purely to properties of possessing a larger vocabulary.

Novel word learning analyses that included Training 1 vocabulary size (especially the models with only vocabulary size as predictor) as a predictor indicated that networks with larger vocabularies learned more words and had a faster learning rate in the novel word learning task. Furthermore, as indicated before, networks which had been exposed to the larger corpus (i.e. more types) in Training 1 showed better novel word learning in Training 2. This suggests an advantage for networks that know larger numbers of distinct words, i.e. networks with larger vocabularies. This is consistent with evidence from earlier modelling and experimental studies and preferential attachment theory (Sailor, 2013; Steyvers & Tenenbaum, 2005) that argues that new items are more likely to be acquired should the network already possess a similar item that is densely connected. Although, in in our analysis of the model a similar mechanism is attributed to consequences of variation in the diversity on input, it is also possible that such a mechanism is at work across all networks that have greater language knowledge (larger vocabulary size). Further, examination of the behaviour and internal processing of networks before and during stage 2 training is required to tease apart cause and effect.

However, analyses of RTs and novel word learning performance where vocabulary size was included as predictor while also accounting for the various environmental and cognitive causes for variation in vocabulary size generated mixed results, providing little evidence for beneficial effects resulting from mechanisms driven by increased vocabulary size.

Further, we acknowledge that we are unable to rule out effects of mechanisms that are not captured by our model which may generate distinct advantages as a result of increased vocabulary size. One such mechanism may result from interactions between fast mapping, mutual exclusivity, and vocabulary size as, for example, with more words known it is likely easier to identify novel items and attribute novel labels (McMurray et al., 2012). Another set of potentially beneficial mechanisms beyond the scope of the current model may exist in explicit (rather than implicit) learning strategies or heuristics which potentially play more significant roles in adulthood. For example, with greater language knowledge it may become easier to identify explicitly systematic relationships within and between semantic and phonological representations that then aid learning.

Our examination of the emergent behaviour of a system in which an error based learning algorithm is applied to learn phonological and semantic representations and the mappings between such representations, offers little evidence for distinct benefits purely attributable to increased vocabulary size. Instead, the results of simulations conducted within such a model suggest that measures of vocabulary size capture variance generated by a complex interaction of environmental and cognitive causal factors, many of which share similar consequences for language learning and processing (e.g. increased vocabulary size, faster processing, faster rates of word learning, frequency x skill interactions). Thus, if relationships are analysed between vocabulary size and measures of language processing or learning without a comprehensive representation of such causal factors, then variance attributable to such factors is likely to load on vocabulary size. Further, given the difficulty in obtaining measures that accurately capture variance in such underlying causal factors, such circumstances are likely to often occur, and therefore exaggerate the distinct effects of variation in vocabulary size. Should such measures be obtainable, it remains to be seen how much variance can be explained by vocabulary size. However, our simulations suggest that a substantial part of the frequently observed relationships between vocabulary size and language processing and learning might be generated by factors underlying differences in vocabulary size instead of by mechanisms driven purely by variation in vocabulary size. Given such relationships between the factors that generate variation in vocabulary size and the difficulty of obtaining measures that accurately capture variation in such underlying causes, we believe it is only through further computational modelling that it may be possible to isolate properties of behaviour that can distinguish between distinct causes and distinct consequences of variation in vocabulary size.

Conclusions

The present study highlights the complexity of the interaction between environmental and cognitive factors that give rise to variation in vocabulary size. Further, our model provides an explicit description of multiple mechanisms that cause variation in vocabulary size and their behavioural consequences, namely differences in lexical processing speed, learning rate, novel word learning performance, and in the structure of the emergent lexicon. Our simulations indicate that advantages in language processing observed for individuals with larger vocabularies are likely to largely result from interactions between underlying environmental and cognitive factors that generate both improved language processing ability and increased vocabulary size (see e.g. Fernald et al., 2006). Further, our simulations suggest that any additional advantage that results purely from having a larger vocabulary is likely very small relative to the effects of such causal factors, i.e. input diversity, exposure, intelligence, processing speed.

Appendix A

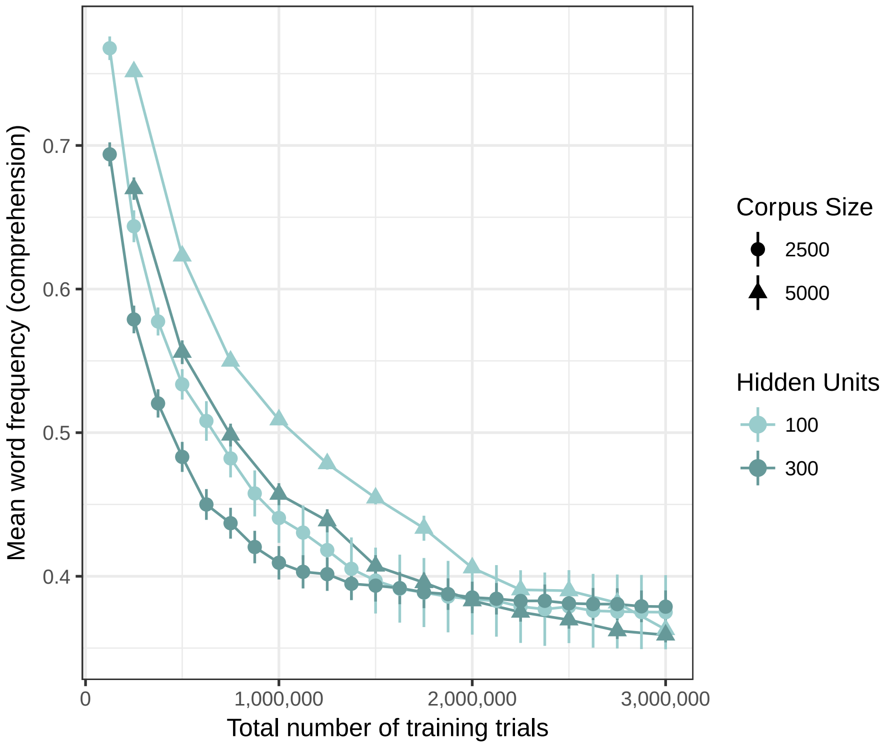

Figure 6.5: The development of average word frequency for the model's comprehension vocabulary across training as a function of Hidden Units, i.e. general intelligence. Standard errors are indicated in the plot.

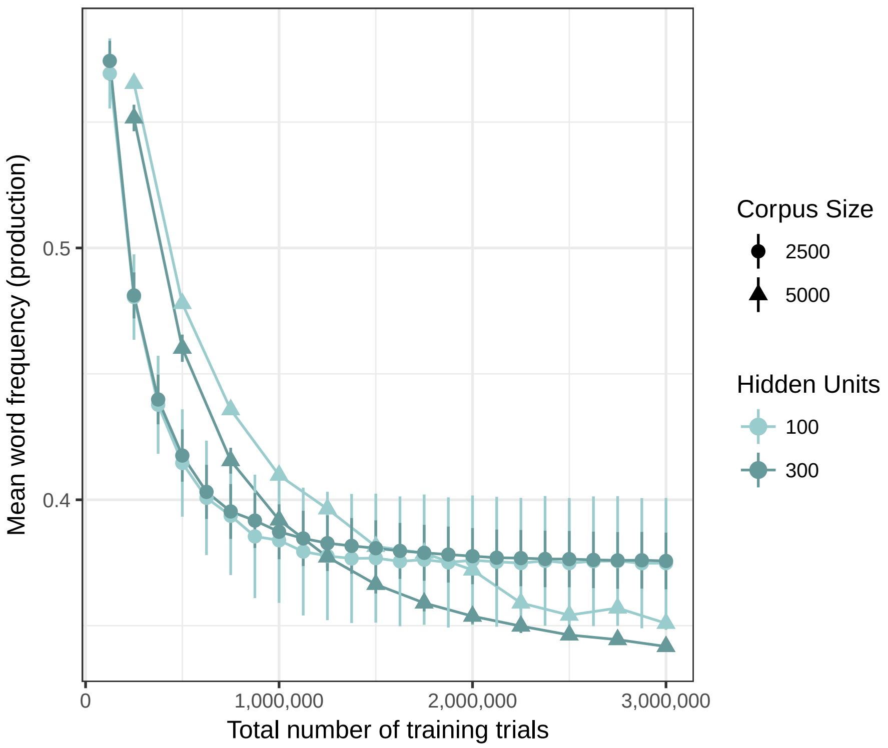

Figure 6.6: The development of average word frequency for the model's production vocabulary across training as a function of Hidden Units, i.e. general intelligence. Standard errors are indicated in the plot.

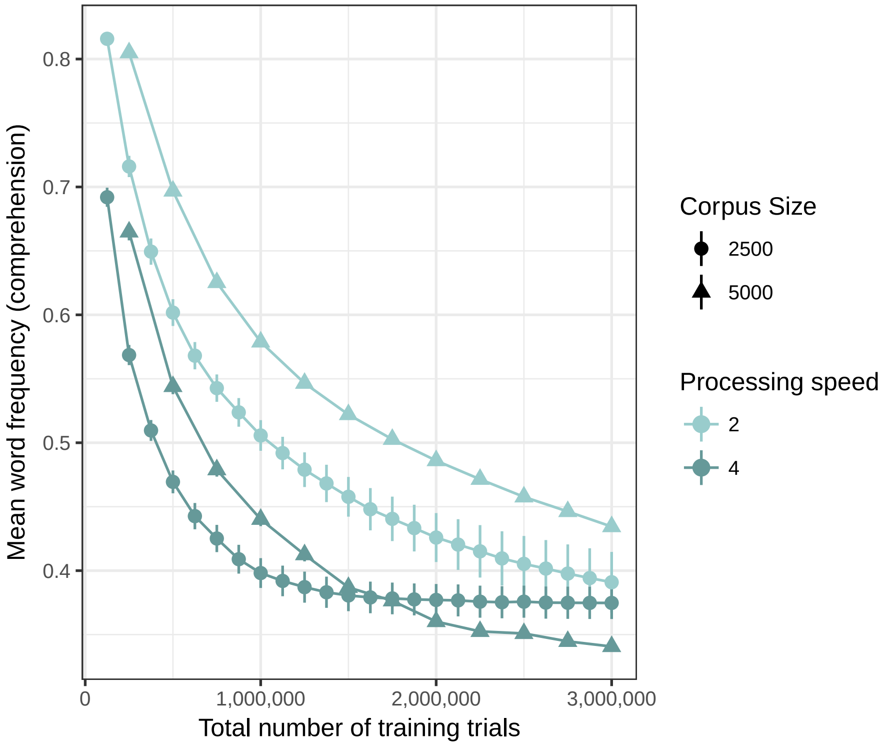

Figure 6.7: The development of average word frequency for the model's comprehension vocabulary across training as a function of Processing Speed. Standard errors are indicated in the plot.

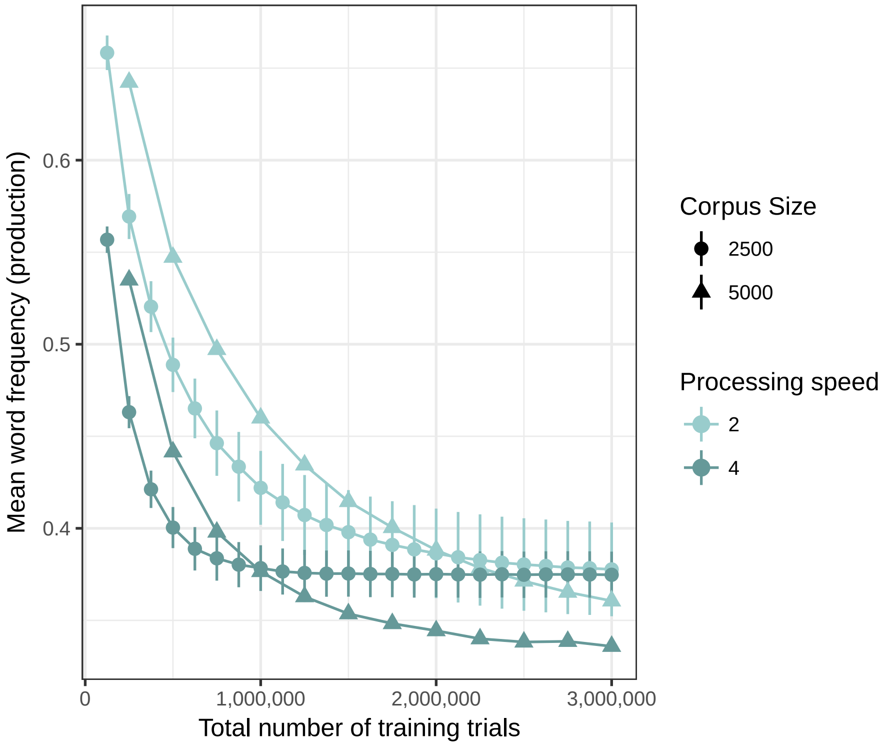

Figure 6.8: The development of average word frequency for the model's production vocabulary across training as a function of Processing Speed. Standard errors are indicated in the plot.

Appendix B

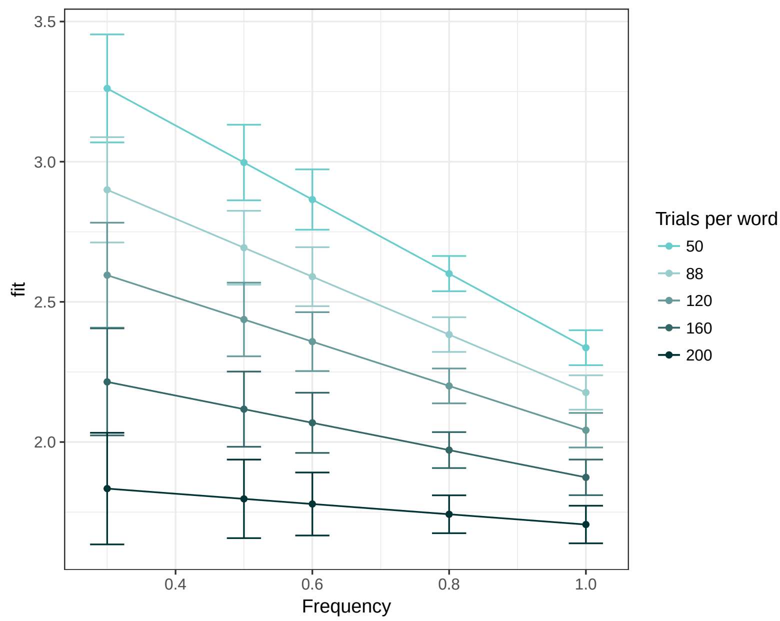

Figure 6.9: The Interaction between the word frequency effect and the effect of Trials per Word. Error bars indicate the standard error.

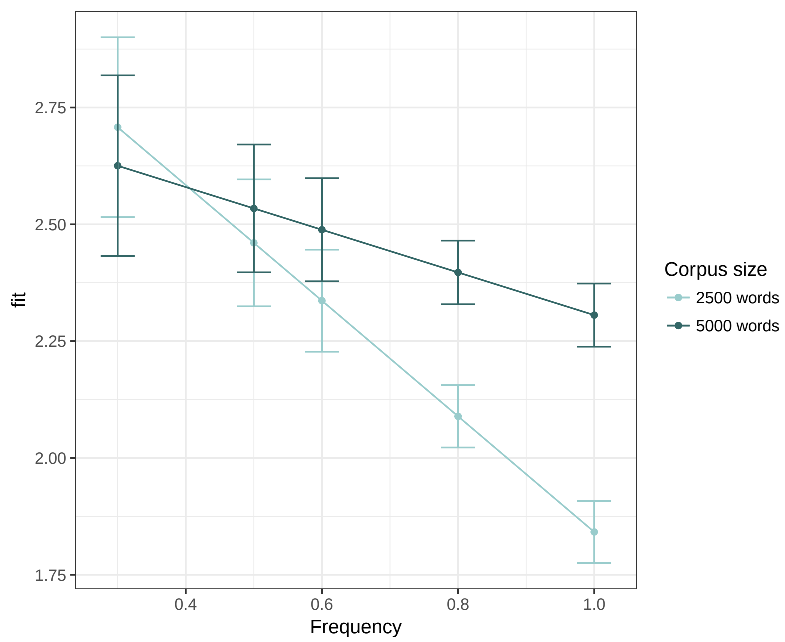

Figure 6.10: The Interaction between the word frequency effect and the effect of Corpus Size. Error bars indicate the standard error.

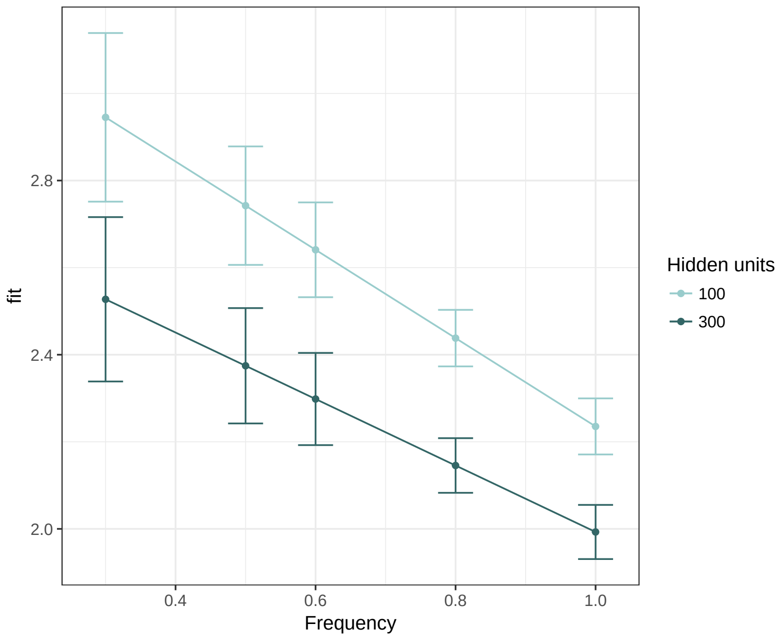

Figure 6.11: The Interaction between the word frequency effect and the effect of Hidden Units, i.e. processing capacity, which was assumed to simulate variation in general intelligence. Error bars indicate the standard error.

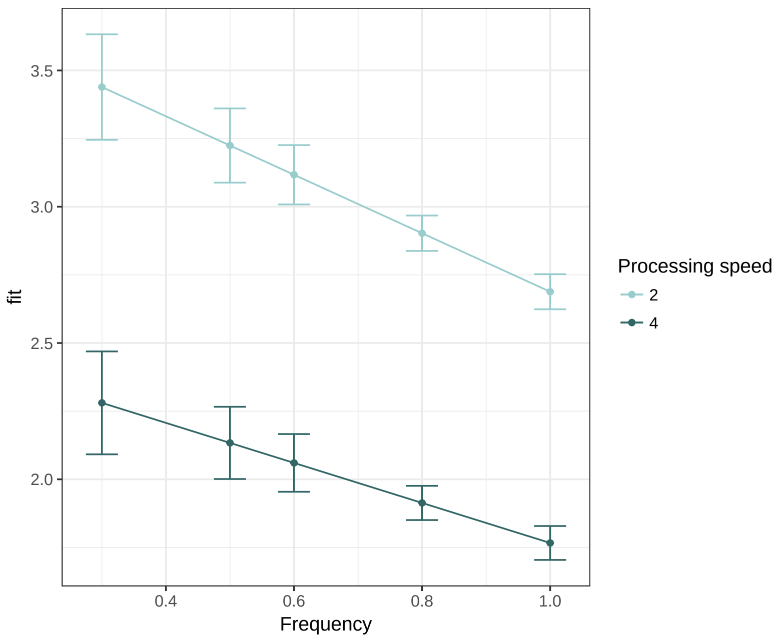

Figure 6.12: The Interaction between the word frequency effect and the effect of Processing Speed, variation in which was implemented by varying the integration constant. Error bars indicate the standard error.

Appendix C

Table 6.14: The estimates, *t*-, and *p*-values for all coefficients in the mixed-effects model on comprehension RT with only Vocabulary Size as skill-related predictor.

Fixed effect	**Estimate**	***t*-value**	***p*-value**
Intercept	2.33	12.32	<.001
Frequency	-0.14	-2.84	<.001
Semantic density	-0.17	-3.60	<.001
Phonological density	-0.08	-1.66	.10
Vocabulary size (comp)	-0.25	-20.60	<.001
Frequency:Sem density	-0.08	-1.93	.05
Frequency:Pho density	0.03	0.64	.52
Frequency:Vocabulary	0.05	6.08	<.001
Sem density:Pho density	-0.03	-0.57	.57
Sem density:Vocabulary	0.09	12.66	<.001
Pho density:Vocabulary	0.03	4.19	<.001

Appendix D

Table 6.15: The estimates, t-, and p-values for all coefficients in the mixed-effects model on production RT (without vocabulary size as predictor).

Fixed effect	**Estimate**	***t*-value**	***p*-value**
Intercept	2.36	63.44	<.001
Corpus (5000)	0.47	10.31	<.001
Speed (4)	-0.62	-27.48	<.001
Hidden Units (300)	-0.08	-3.61	<.001
Frequency	-0.41	-18.24	<.001
Trials per Word	-0.14	-11.94	<.001
Semantic density	-0.31	-14.14	<.001
Phonological density	-0.23	-10.28	<.001
Corpus:Speed	-0.17	-5.21	<.001
Corpus:Hidden Units	-0.08	-2.47	.01
Corpus:Frequency	0.07	8.37	<.001
Corpus:Trials per Word	-0.24	-25.03	<.001
Corpus:Sem density	-0.03	-3.87	<.001
Corpus:Pho density	-0.03	-3.17	.002
Speed:Frequency	0.20	27.34	<.001
Hidden Units:Frequency	0.04	5.55	<.001
Frequency:Trials	0.07	17.67	<.001
Frequency:Sem density	0.01	0.49	.62
Frequency:Pho density	0.02	1.09	.27
Speed:Trials per Word	0.15	15.64	<.001
Hidden Units:Trials	0.0009	0.10	.92
Trials:Sem density	0.01	3.79	<.001
Trials:Pho density	-0.002	-0.46	.65
Speed:Sem density	0.13	18.50	<.001
Hidden Units:Sem density	0.02	3.05	.002
Sem density:Pho density	0.02	0.80	.43
Speed:Pho density	0.17	23.92	<.001
Hidden Units:Pho density	0.003	0.41	.68

Table 6.16: The estimates, t-, and p-values for all coefficients in the mixed-effects model on production RT with only Vocabulary Size as skill-related predictor.

Fixed effect	**Estimate**	***t*-value**	***p*-value**
Intercept	4.45	5.40	<.001
Frequency	-0.36	-1.84	.07
Semantic density	-0.08	-1.46	.14
Phonological density	-0.10	-1.74	.08
Vocabulary size (prod)	-0.58	-19.31	<.001
Frequency:Sem density	0.001	0.25	.80
Frequency:Pho density	0.009	0.73	.57
Frequency:Vocabulary	0.07	22.41	<.001
Sem density:Pho density	0.001	0.30	.76
Sem density:Vocabulary	0.006	7.56	<.001
Pho density:Vocabulary	0.02	12.09	<.001

Appendix D

Table 6.17: The estimates, t-, and p-values for all coefficients in the mixed-effects model on comprehension vocabulary size in the novel word learning task including Vocabulary Size as predictor.

Fixed effect	**Estimate**	***t*-value**	***p*-value**
Intercept	58.76	17.57	<.001
Corpus (5000)	-20.62	-4.98	<.001
Speed (4)	19.08	11.07	<.001
Hidden Units (300)	19.57	20.06	<.001
Trials per Word (2)	20.57	246.66	<.001
Trials per Word (1)	0.52	2.43	.02
Vocabulary size (comp)	17.81	18.52	<.001
Trials_2:Vocabulary	1.34	47.49	<.001
Speed:Corpus	37.49	17.01	.01
Hidden Units:Corpus	22.69	11.59	<.001
Trials_2:Corpus	-3.22	-77.39	<.001
Trials_1:Corpus	32.44	32.84	<.001
Vocabulary:Corpus	-21.77	-17.04	.002
Speed:Vocabulary	-2.73	-10.62	<.001
Speed:Trials_2	8.14	129.36	<.001
Speed:Trials_1	-3.31	-11.13	<.001
Hidden Units:Vocabulary	0.09	0.47	.64
Hidden Units:Trials_2	8.23	156.86	<.001
Hidden Units:Trials_1	0.17	0.68	.50

Appendix D

Table 6.18: The estimates, t-, and p-values for all coefficients in the mixed-effects model on novel production vocabulary size (without vocabulary size as predictor).

Fixed effect	**Estimate**	***t*-value**	***p*-value**
Intercept	60.94	18.13	<.001
Corpus (5000)	14.83	3.12	.002
Speed (4)	32.51	22.85	<.001
Hidden Units (300)	22.53	15.84	<.001
Trials per Word (2)	31.14	490.76	<.001
Trials per Word (1)	4.99	59.31	<.001
Speed:Corpus	-1.23	-0.61	.54
Hidden Units:Corpus	11.62	5.77	<.001
Trials_2:Corpus	-3.05	-76.01	<.001
Trials_1:Corpus	16.49	238.51	<.001
Speed:Trials_2	6.34	129.09	<.001
Speed:Trials_1	-4.17	-61.48	<.001
Hidden Units:Trials_2	3.32	67.49	<.001
Hidden Units:Trials_1	2.61	38.69	<.001

Table 6.19: The estimates, t-, and p-values for all coefficients in the mixed-effects model on novel production vocabulary size (including vocabulary size as predictor).

Fixed effect	**Estimate**	***t*-value**	***p*-value**
Intercept	58.01	15.21	<.001
Corpus (5000)	20.74	4.09	<.001
Speed (4)	29.13	16.80	<.001
Hidden Units (300)	26.59	17.97	<.001
Trials per Word (2)	33.32	363.78	<.001
Trials per Word (1)	5.84	17.72	.02
Vocabulary size (prod)	-3.31	-3.10	.002
Trials_2:Vocabulary	1.18	32.57	<.001
Speed:Corpus	8.44	3.76	<.001
Hidden Units:Corpus	5.94	2.86	.004
Trials_2:Corpus	-4.55	-74.96	<.001
Trials_1:Corpus	19.02	18.62	<.001
Vocabulary:Corpus	0.91	0.70	.48
Speed:Vocabulary	-4.50	-12.79	<.001
Speed:Trials_2	4.78	69.92	<.001
Speed:Trials_1	-1.13	-4.15	<.001
Hidden Units:Vocabulary	3.71	12.84	<.001
Hidden Units:Trials_2	2.75	53.25	<.001
Hidden Units:Trials_1	0.84	4.72	<.001

7 | Summary and discussion

The present dissertation aimed at answering various questions all centering around the observation that native speakers of a language vary considerably in their vocabulary. What is the relationship between vocabulary size and language processing? Is a battery of vocabulary tests really necessary to assess the knowledge of words in native speakers? And do findings from studies on the typical participants in psychological research, namely undergraduate university students, apply to individuals from more diverse backgrounds? These questions were addressed in Chapters 2 to 4 of this dissertation. One central finding was that there are big differences between native speakers' knowledge of words.

Having observed such considerable variation in vocabulary size among the native speakers of a language, I asked what the origin is for some people being better word learners than others and, therefore, ending up having larger vocabularies than others. This was the main focus of the remaining two chapters of this dissertation, Chapters 5 and 6. In addition, the computational modelling study presented in Chapter 6 examined the relationship between causes for variation in vocabulary size and consequences thereof for language processing and novel word learning performance.

In the following, I provide a summary of the main findings from all chapters. Subsequently, I discuss them in light of earlier research on lexical processing, vocabulary size and learning. In addition, implications for future research in general and more specifically individual differences studies are discussed.

In the experiment detailed in Chapter 2 I investigated individual differences in vocabulary size – as measured by a battery of seven vocabulary tests – and their relationship with variation in lexical processing – more precisely word recognition. Performance on a lexical decision task (LDT) was found to be predicted by word frequency and vocabulary, with high-frequency words and higher vocabulary scores being related to faster responses on the LDT. Hence, greater vocabulary knowledge was indicated to be beneficial for lexical processing, which is in line with the findings from earlier research (Brysbaert et al., 2016a; Diependaele et al., 2013; Yap et al., 2009). Furthermore, this relationship between higher vocabulary scores and faster reaction times (RTs) was not only found for the composite score of vocabulary representing individuals' performances on all seven vocabulary tests, but also for the majority of individual vocabulary tests. This leads to the second important implication of the study presented in Chapter 2, namely that based on these findings, six out of the seven vocabulary tests were each considered to be as representative of participants' vocabulary size as the composite score. The PPVT (Schlichting, 2005) as well as Andringa et al.'s (2012) receptive multiple-choice test, and the newly developed open antonym and synonym, multiple-choice synonym, and definition tests all showed the same pattern of the relationship between vocabulary and word recognition accuracy and speed. Solely the multiple-choice antonym test appeared to be too easy for the present group of participants; an impression which was supported by the considerably higher test scores on this test, the lower correlations between this and the other tests, and the Principal Components Analysis, which showed the smallest loading of the first component on the multiple-choice antonym test. Consequently, the experiment reported on in Chapter 2 did not provide any evidence for the supposed need to use a battery of different vocabulary tests to assess word knowledge in the group of participants typically tested in psychological research, namely undergraduate university students. This is true at least if the vocabulary test used is capable of eliciting individual differences in test performance, which was apparently not true for the multiple-choice antonym test.

The experiment reported on in Chapter 3 was intended to complement Chapter 2, in that the same tasks were used to test participants from more diverse educational backgrounds. It has to be noted that the materials used in this study were in fact not exactly the same as in the previous experiment. One vocabulary test, namely the open synonym test, had been indicated to be very challenging for the university students, which is why I did not include it in the testing of the vocational college students. Furthermore, an additional five high-frequency filler items were added to the multiple-choice antonym and synonym tests and to the open antonym test, respectively. This was done to increase the number of relatively easy items and keep participants motivated throughout the tests. In addition, pilot studies showed that presenting the LDT stimuli for 3 seconds was too short for the vocational college students. The length of the stimulus display was, therefore, increased to 5 seconds. Thus, even without looking at the results, the fact that these changes were necessary to make the tasks suitable for this more diverse group of participants than is typically tested tells us something about how homogeneous and unrepresentative of variation present in the population the group of "typical" participants, i.e. university students, appears to be.

As might be expected, the vocabulary scores as well as the LDT accuracy rates of the participants in this experiment were considerably lower than those of the university students, and the RTs were slower. These performance differences support what has been indicated before, namely that conclusions based on the "typical" participants cannot necessarily be extended to participants from more diverse backgrounds, as the skills present in the latter group are also much more diverse. It is, hence, questionable whether and in how far insights gained from testing undergraduate university students are representative of the variation present in the population, especially when it comes to studies focusing on individual differences.

With regards to the relationship between lexical processing and vocabulary, we observed that LDT accuracy but notably not RTs were affected by vocabulary in this experiment. Hence, higher vocabulary scores were in this group of participants associated with higher response accuracy on the LDT but not with faster responses. Besides that, word frequency predicted both accuracy and RTs on the LDT, with higher-frequency words eliciting higher accuracy rates and faster RTs. In addition, the word frequency effect on RTs was found to vary as a function of vocabulary score, with individuals with higher vocabulary scores showing a reduced word frequency effect. This frequency by skill interaction has previously been reported by studies on word recognition and has been taken to indicate representational differences between vocabularies of varying sizes (Diependaele et al., 2013). The lexical representations in high-vocabulary individuals have been assumed to be more robust or more entrenched and, therefore, not only faster to be accessed but also less sensitive to effects of word frequency (Brysbaert et al., 2016a; Diependaele et al., 2013). Monaghan et al. (2017) used a computational model to demonstrate that differences in the amount of exposure to linguistic input can explain this frequency by skill interaction, with increased exposure leading to reduced word frequency effects on processing.

Just as with the results from the experiment in Chapter 2, the findings in this study indicate beneficial effects of greater vocabulary knowledge on language processing. However, differently to that reported in the previous chapter, the positive effect of increased vocabulary size was not found on lexical processing speed but on response accuracy and participants' sensitivity to the word frequency effect. Hence, it is suggested that in this group of participants, increased vocabulary knowledge might be associated in particular with differences in the entrenchment of lexical representations.

Why the expected relationship between overall RTs and vocabulary did not show is unclear. If anything, it seems that there was a tendency of high-vocabulary individuals to show slower RTs. Perhaps participants in this group relied on different strategies when responding to the lexical decision task. Thus, it might be that participants spent a rather long time thinking about the target items, delaying their button press until they were sure enough about what the correct response is. A potential reason for this kind of strategy might be less confidence in their knowledge relative to, for instance, university students, leading to these participants' responses being less intuitive and quick. This might be the case in particular for individuals who were determined to do well on this task, which may explain the relationship between slower RTs and higher vocabulary scores. Alternatively, performance on this task may be affected by these young adults being less proficient readers than university students, who receive more extensive exposure to written language. This would be in line with the observation of considerably longer RTs in this experiment than in the one presented in Chapter 2, and with the need to increase the length of the stimulus display. The role of reading proficiency in affecting LDT performance could in future research be tested by using an auditory version of the task.

Finally, vocabulary test performance in this group of participants was rather varied both within and across individuals, as indicated by the significant but only moderate correlations between all vocabulary test scores. Hence, based on the research presented in Chapter 3, we would recommend using at least two vocabulary tests to appropriately assess word knowledge when testing a similarly diverse group of participants.

In the experiment presented in Chapter 4, the same university students were tested as in the experiment in Chapter 2. In addition to completing a LDT and the battery of seven vocabulary tests, they performed a picture-word interference task. Similarly to what has been found in the previous chapters, higher vocabulary scores were associated with faster lexical processing. Hence, individuals with greater vocabulary knowledge were faster at producing the target words in this task than individuals with smaller vocabularies.

After having observed such considerable variation in vocabulary size among the native speakers of a language, the question of what the origins of these individual differences in vocabulary learning and size are arose. Chapter 5 presents the findings from an experiment on novel word learning in adult native speakers, which aimed at providing first insights into the relationship between variation in cognitive and environmental factors and their effects on word learning. Not only amount of exposure and sleep were found to influence novel word learning performance as measured by a picture naming test, but also nonverbal intelligence and vocabulary. Participants were tested on the novel word-picture pairings right after training and after a delay of one week. Their performance was overall better after a delay involving periods of sleep, thus allowing for overnight consolidation. In addition, test performance was better for words that had been presented more often during training than for low-exposure words, and these differences depending on exposure frequency became stronger after a delay of one week. This supported the impression that lexical consolidation might have taken place between training and the test after a week. Furthermore, participants with higher vocabulary scores performed better on the novel word learning task, as indicated by more accurate picture naming performance. In addition, individuals with higher Raven's scores, which were taken to be an indication of nonverbal intelligence, showed stronger beneficial effects of higher vocabulary scores on novel word learning than individuals with lower Raven's scores. Finally, we observed a tendency towards higher general processing speed being associated with reduced sensitivity to constraints on the number of exposures to the novel words. Importantly, overall not only environmental but also cognitive factors, and potentially complex relationships between them, were implied to play a role in determining word learning ability.

Especially interesting was the strong effect that vocabulary size was shown to have on novel word learning performance. It might seem trivial that individuals with higher vocabulary scores, thus probably better word learning performance, outperform those with lower vocabulary scores on a task that assesses word learning performance. However, the question arises: what are the underlying mechanisms for the beneficial effect of greater vocabulary knowledge on word learning? In addition, one might ask why some individuals have larger vocabularies in the first place, which appear to benefit subsequent word learning. Are differences in exposure to linguistic input early in development the reason for some children acquiring larger vocabularies than others, and do these advantages then persist throughout the lifespan? Or do individual differences in general cognitive skills cause the emergence of variation in vocabulary size and do these differences in cognitive skills continue to influence word learning later in life?

The computational modelling study presented in Chapter 6 of this dissertation was meant to provide answers to these questions. A distributed connectionist model was used to investigate potential environmental and cognitive causes for variation in vocabulary size, and the interactions between them. In addition, the relationship between variation in vocabulary size and differences in language processing and novel word learning were examined. In this way, this last chapter combined the research questions and topics from all previous chapters to provide a bigger picture of the relationships between causes for variation in vocabulary learning and differences in processing, and underlying mechanisms of these effects. The environmental factors quantity and quality of exposure, as well as the cognitive factors general processing speed and processing resources, and interactions between them were found to influence vocabulary learning and size. Greater input quantity and quality were associated with larger vocabularies. In addition, higher processing speed and increased resources were found to cause higher learning rates as well as greater word knowledge. A mechanism underlying the latter effects appeared to be that more skilled networks, i.e. those with greater processing speed and resources, are capable of learning larger numbers of more difficult to learn words. Greater input diversity, by contrast, was indicated to be beneficial for word learning due to larger numbers of easy to learn words being included in larger corpora. Importantly, these causes for variation in vocabulary size were observed to all have very similar consequences for language processing and novel word learning. Origins for improved word learning and vocabulary size appeared to also cause higher language processing speed, reduced word frequency effects, and better novel word learning performance. Hence, the computational modelling investigation suggested that the beneficial effects on language processing and novel word learning that were earlier attributed to larger vocabularies might be driven by variation in underlying cognitive and environmental factors. We are, however, unable to rule out that increased vocabulary size is associated with advantages for word learning that are distinct from those of underlying cognitive and environmental factors. Mechanisms underlying such additional benefits might be related to structural or representational characteristics of greater vocabularies (Sailor, 2013; Steyvers & Tenenbaum, 2005).

Measuring vocabulary size

One of the questions underlying the work presented in the first chapters of this dissertation was whether a battery of measures is necessary to assess vocabulary size. It has been argued that no vocabulary test is a pure measure of the knowledge of words but involves other general cognitive abilities (Bowles & Salthouse, 2008). Therefore, Bowles and Salthouse (2008) advised researchers to employ a battery of different measures of vocabulary size, especially in studies where vocabulary knowledge is the center of interest. Furthermore, Henriksen (1999) has proposed three dimensions along which vocabulary learning and knowledge varies, namely along continua from partial to precise knowledge, from shallow to deep knowledge, and from receptive to productive use ability. Assuming that variation in vocabulary knowledge can be assessed along these three dimensions also supports the claim that vocabulary tests of different types (antonym vs. synonym vs. definition) and formats (multiple-choice vs. open) are necessary to appropriately and comprehensively assess vocabulary size.

This is what I applied in the experiments presented in the first three experimental chapters of this dissertation. As indicated above, my findings from Chapter 2 (and Chapter 4) suggest that a single vocabulary test would have been sufficient to assess vocabulary. All measures except for the multiple-choice antonym test, which appeared to be too easy for the participants in this experiment, correlated strongly with each other and showed the same relationship with lexical processing performance. Hence, this experiment does not provide evidence for the need of multiple measures of vocabulary, as long as the test that is used is capable of eliciting individual differences in the target group.

However, the experiment detailed in Chapter 3 where individuals from more diverse backgrounds were tested provides a slightly different picture and, hence, a different answer to the question of whether single vocabulary tests are sufficient to assess word knowledge. Vocabulary test performance in this more diverse group of participants was more varied both within and across individuals. It appeared that the relative difference in difficulty between the vocabulary measures was larger in this than in the university students group, as suggested for example by the smaller correlations between the tests. In addition, the predictions about the relationship between vocabulary and lexical processing differed more in this than in the previous group depending on which vocabulary test was used in the analyses. Hence, overall my findings indicate that when testing a more diverse group, which is presumably characterised by a larger range of skills, the use of at least two different measures of vocabulary is advisable. In this way, the composite score calculated based on two different test types and/or formats is probably more representative of participants' vocabulary than relying on only one test; especially given the greater performance variability between tests that was observed in the study presented in Chapter 3.

Testing more diverse groups of participants

In addition to greater variability in vocabulary test performance within the group of participants in the second experiment (Chapter 3) than in the first experiment (Chapter 2), the average test scores reflected the effect of educational background, which has been previously reported (see e.g. Brysbaert et al., 2016b). The vocabulary scores achieved by the participants tested in the first experiment, who were mainly university undergraduates, were not only more homogeneous but also considerably higher than those from the participants with more diverse backgrounds tested in the second experiment. These observations mirror my findings in regards to lexical processing performance. The latter was also found to be weaker and more varied in the second, more diverse group of participants. Importantly, as indicated above, the materials of both the vocabulary tests and the LDT had to be modified to make them suitable for testing young adults at vocational colleges.

All of these observations point in the same direction, namely that the group of participants typically tested in psychological and psycholinguistic research, i.e. undergraduate university students, are not representative of the wider population. An important implication of my work for future research is the necessity to test participants from more diverse educational backgrounds (see also Kidd et al., 2018). Theories about the relationship between cognitive abilities and language processing as well as individual differences therein should be informed not only by findings from a small, homogeneous, and likely highly skilled group of individuals but by studies that account for and represent the full range of skills to be found in the population. Importantly, the present work has also demonstrated some of the challenges associated with applying typical psycholinguistic tasks to a more diverse group of participants. Some measures might need to be modified to make them suitable for individuals outside the typical group of participants. This, however, only supports the importance of extending our research to include participants from more diverse backgrounds.

Vocabulary size and language processing

Besides these methodological implications, the findings from Chapters 2 and 3, as well as from Chapter 4, provided insight into the relationship between vocabulary knowledge and language processing. My observations were in line with previous research indicating that greater vocabularies are associated with beneficial effects on lexical processing in both comprehension and production. Overall, individuals with higher vocabulary scores showed higher accuracy rates (Chapter 3) and faster RTs (Chapters 2 and 4) in word recognition and production. In addition, the frequency by skill interaction, which has been observed in earlier studies (Brysbaert et al., 2016a; Diependaele et al., 2013; Monaghan et al., 2017), was replicated in the experiment presented in Chapter 3. Hence, in the more diverse group of participants higher vocabulary scores were associated with reduced word frequency effects on lexical decision RTs. It is unclear why the experiment where the typical group of participants, namely university students, was tested did not show this effect. One possibility may be that the range of vocabulary scores and LDT performance in this sample was too small to elicit the frequency by skill interaction.

This interaction between the word frequency effect and skill has been taken to indicate differences in the degree of robustness or entrenchment of lexical representations in individuals with smaller vs. larger vocabularies (Brysbaert et al., 2016a; Diependaele et al., 2013). The typical frequency effect, with faster RTs for high- as compared to low-frequency words, has been suggested to result from differences in the efficiency of processing words varying in frequency of occurrence. Especially, increased exposure, which can be measured as greater vocabulary size, has been argued to cause a reduction in this difference in processing efficiency between low- and high-frequency words, which in turn is indicated by reduced frequency effects on lexical processing (Brysbaert et al., 2016a; Monaghan et al., 2017). Hence, the findings from the experiment detailed in Chapter 3 are in line with the assumption that higher vocabulary scores are associated with more entrenched or robust lexical representations, as indicated by a reduced effect of word frequency on lexical processing.

A question arising in light of these observations is why having a larger vocabulary is beneficial for lexical processing. It may be assumed that having more words in one's lexicon might render processing more difficult or slower as more words might compete for selection (Diependaele et al., 2013). The computational modelling study reported on in Chapter 6 of this dissertation provided some potential answers to this question. First of all, I replicated the beneficial effects of increased vocabulary size on language processing speed. Just as observed in earlier behavioural research, networks with larger vocabularies were faster to comprehend and produce words. Interestingly, however, the model showed that once factors causing variation in vocabulary size are controlled for, there is little evidence for additional beneficial effects of vocabulary on processing. The same was true for the frequency by skill interaction, which was elicited by various skill-related factors, i.e. factors that were shown to cause variation in vocabulary size. Hence, the factors quantity and quality of input, processing speed, and processing resources (with the latter taken to represent general intelligence), influenced vocabulary learning and size, as well as lexical processing, and showed interactions with the word frequency effect. The computational modelling study, thus, suggests that at least part of the mechanism behind the vocabulary advantage in processing might be variation in environmental and cognitive factors that underlie the observed differences in vocabulary size.

Importantly, the frequency by skill interaction appears to not only be a result of variation in exposure, which has previously been argued to be the case (Monaghan et al., 2017). I demonstrated that complex interactions between cognitive and environmental factors likely cause variation in vocabulary learning rate as well as size and also in lexical processing efficiency and the sensitivity to word frequency effects. Furthermore, it was demonstrated that the frequency by skill interaction might be extended to any factor (environmental and cognitive) determining the learning rate and success of a network, but also to any characteristic that affects the difficulty of learning and processing a given lexical item. Aside from word frequency, the lexical characteristics phonological and semantic density, which in the present model were shown to affect degree of difficulty of learning and processing lexical representations, interacted with skill. This finding certainly needs to be examined further in future behavioural and computational investigations.

The present computational model contributed new insights to findings based on previous studies by taking into account not only variation in exposure but also the extensive variation in cognitive skills that has been observed across a large range of individual differences studies (Kidd et al., 2018).

Future research, however, needs to more closely examine the effects of variation in vocabulary size on structural and representational characteristics of the lexicon. The present findings cannot rule out that greater vocabulary knowledge has additional beneficial effects on language processing due to internal, structural properties of the lexicon, i.e. beneficial effects that go beyond those that can be attributed to underlying environmental and cognitive factors (Sailor, 2013; Steyvers & Tenenbaum, 2005).

Individual differences in vocabulary learning

This leaves the question of what causes the considerable variation in vocabulary knowledge or size observed in the experimental work presented in Chapters 2 to 4. In the novel word learning experiment reported in Chapter 5, I explored the roles of individual differences in different environmental and cognitive factors in affecting word learning performance. Both sleep and higher numbers of exposure were associated with beneficial effects on learning, which is in line with findings from previous research (Hurtado et al., 2008; Monaghan et al., 2017; Weighall et al., 2017). Furthermore, both vocabulary and nonverbal intelligence as measured by the Raven's Matrices (Raven et al., 1998) were indicated to influence novel word learning. Individuals with greater vocabulary knowledge outperformed those with lower vocabulary scores, which replicated findings from developmental research showing an association between larger vocabularies and improved word learning performance (Henderson & James, 2018; Marchman & Fernald, 2008). Higher Raven's scores were associated with even stronger beneficial effects of increased vocabulary knowledge and with stronger effects of overnight consolidation. In addition, there was a tendency of higher general processing speed to be related to reduced effects of exposure frequency on learning. In a nutshell, the findings from this word learning study support the assumption that not only exposure alone but a complex interaction between environmental and cognitive factors likely causes variation in word learning.

More research on the role of differences in various cognitive skills and their interactions with properties of the learning environment in adult learners is, however, necessary. While in this study only university students were tested due to limited availability of more diverse participant groups and constraints on the test setting, future investigations into word learning need to be run with participants from more diverse backgrounds, as has been argued above.

The computational model detailed in Chapter 6 provided valuable insights into potential causes for variation in vocabulary learning and size. As indicated above the factors quantity and quality of the input (environmental) and processing speed and resources (cognitive), with the latter being assumed to simulate variation in intelligence, were observed to cause variation in vocabulary learning and size. What is more, once these causes for individual differences in vocabulary size were controlled for, there was not much evidence for additional beneficial effects of greater vocabulary knowledge on novel word learning. Hence, the advantages in novel word learning, which have previously been attributed to vocabulary size itself, appear in this model to be driven by factors underlying variation in vocabulary size. In line with suggestions made in the literature, differences in cognitive skills might cause the emergence of differences in vocabulary size and at the same time underlie variation in language processing speed, as described above (Fernald et al., 2006).

In addition to identifying factors likely causing variation in vocabulary size, the computational model helped shed light on mechanisms potentially underlying the beneficial effects of increased input quality as well as greater processing speed and resources. Greater input quality, i.e. greater input corpora, were indicated to aid word learning due to being comprised of a larger number of easy to learn words as compared to smaller corpora. Greater skill, i.e. higher speed or greater resources, by contrast appeared to be advantageous for vocabulary learning due to being associated with the ability to learn more difficult to learn items. Thus, while larger corpora, i.e. more diverse input, appeared to be beneficial due to including more words that breach the learning threshold of the network, increased processing speed and resources (i.e. intelligence) were associated with a lowering of the learning threshold. These mechanisms need to be examined more closely in future research; the effects of differing numbers of types in the input on the distribution of certain lexical characteristics determining the difficulty of learning and processing a given word, but also potentially different mechanisms underlying the beneficial effects of different cognitive skills on word learning require further experimental and computational investigation.

Conclusions

As speakers of a language we know a huge number of words already by early adulthood, and we continue to be exposed to and learn new words across the entire lifespan. I have shown that individuals with larger vocabularies are faster at processing language. Thus, despite knowing more words, which may be expected to render the selection of a single lexical item more difficult and slower, speakers with larger vocabularies are faster at making lexical decisions and naming pictures. Furthermore, I have demonstrated that variation in word learning and vocabulary size is likely caused by complex interactions between environmental and cognitive factors, such as input quantity and quality, processing speed, and intelligence. These causes for variation in vocabulary size seem to a great extent to also underlie variation in language processing performance and vocabulary learning. Hence, individual differences in lexical processing speed and word learning that have previously been associated directly with variation in vocabulary size might, in fact, be driven by factors that cause the variation in vocabulary size.

References

Adams, A.-M. (1996). Phonological working memory and spoken language development in young children. *The Quarterly Journal of Experimental Psychology: Section A*, *49*(1), 216–233. doi: 10.1080/713755610

Adams, A.-M., & Gathercole, S. E. (1995). Phonological working memory and speech production in preschool children. *Journal of Speech, Language, and Hearing Research*, *38*(2), 403–414. doi: 10.1044/jshr.3802.403

Adams, A.-M., & Gathercole, S. E. (2000). Limitations in working memory: Implications for language development. *International Journal of Language & Communication Disorders*, *35*(1), 95–116. doi: 10.1080/136828200247278

Adelman, J. S., Johnson, R. L., McCormick, S. F., McKague, M., Kinoshita, S., Bowers, J. S., … Davis, C. J. (2014). A behavioral database for masked form priming. *Behavior Research Methods*, *46*(4), 1052–1067. doi: 10.3758/s13428-013-0442-y

Albinet, C. T., Boucard, G., Bouquet, C. A., & Audiffren, M. (2012). Processing speed and executive functions in cognitive aging: how to disentangle their mutual relationship? *Brain and cognition*, *79*(1), 1–11. doi: 10.1016/j.bandc.2012.02.001

Andringa, S., Olsthoorn, N., van Beuningen, C., Schoonen, R., & Hulstijn, J. (2012). Determinants of success in native and non-native listening comprehension: An individual differences approach. *Language Learning*, *62*(s2), 49–78. doi: 10.1111/j.1467-9922.2012.00706.x

Baayen. (2001). *Word frequency distributions* (Vol. 18). Springer Science & Business Media.

Baayen, R., Piepenbrock, R., & Van Rijn, H. (1993). The celex lexical database [cd-rom] philadelphia: University of pennsylvania. *Linguistic Data Consortium*.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of communication disorders*, *36*(3), 189–208. doi: 10.1016/S0021-9924(03)00019-4

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological review*, *105*(1), 158. doi: 10.1037/0033-295X.105.1.158

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2015). Tracking lexical consolidation with erps: Lexical and semantic-priming effects on n400 and lpc responses to newly-learned words. *Neuropsychologia*, *79*, 33–41. doi: 10.1016/j.neuropsychologia.2015.10.020

Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry the power of response time distributional analyses. *Current Directions in Psychological Science*, *20*(3), 160–166. doi: 10.1177/0963721411408885

Balota, D. A., Yap, M. J., Cortese, M. J., & Watson, J. M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory and Language*, *59*(4), 495–523. doi: 10.1016/j.jml.2007.10.004

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... Treiman, R. (2007). The English lexicon project. *Behavior research methods*, *39*(3), 445–459. doi: 10.3758/BF03193014

Banks, B., Gowen, E., Munro, K. J., & Adank, P. (2015). Cognitive predictors of perceptual adaptation to accented speech. *Journal of the Acoustical Society of America*, *137*(4). doi: 10.1121/1.4916265

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255–278.

Bates, D., Mächler, M. M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48. doi: 10.18637/jss.v067.i01

Bent, T., Baese-Berk, M., Borrie, S. A., & McKee, M. (2016). Individual differences in the perception of regional, nonnative, and disordered speech varieties. *The Journal of the Acoustical Society of America*, *140*(5), 3775–3786. doi: 10.1121/1.4966677

Bowles, R. P., & Salthouse, T. A. (2008). Vocabulary test format and differential relations to age. *Psychology and Aging*, *23*(2), 366–376. doi: 10.1037/0882-7974.23.2.366

Brown, H., & Gaskell, M. G. (2014). The time-course of talker-specificity and lexical competition effects during word learning. *Language, Cognition and Neuroscience*, *29*(9), 1163–1179. doi: 10.1080/23273798.2014.916409

Brysbaert, M., Lagrou, E., & Stevens, M. (2016a). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Bilingualism: Language and Cognition*, *20*(3), 1–19. doi: 10.1017/S1366728916000353

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, *27*(1), 45–50. doi: 10.1177/0963721417727521

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016b). How many words do we know? practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology*, *7*, 1116. doi: 10.3389/fpsyg.2016.01116

Cepeda, N. J., Blackwell, K. A., & Munakata, Y. (2013). Speed isn't everything: Complex processing speed measures mask individual differences and developmental changes in executive control. *Developmental Science*, *16*(2), 269–286. doi: 10.1111/desc.12024

Chateau, D., & Jared, D. (2000). Exposure to print and word recognition processes. *Memory & Cognition*, *28*(1), 143–153. doi: 10.3758/BF03211582

Cristia, A., Seidl, A., Junge, C., Soderstrom, M., & Hagoort, P. (2014). Predicting individual variation in language from infant speech perception measures. *Child development*, *85*(4), 1330–1345. doi: 10.1111/cdev.12193

Cunningham, A. E. (2005). Vocabulary growth through independent reading and reading aloud to children. In M. L. Kamil & E. H. Hiebert (Eds.), *Teaching and learning new vocabulary: Bringing research to practice* (pp. 45–68). Mahwah, NJ: Erlbaum.

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 345. doi: 10.1037/0278-7393.25.2.345

DeCoster, J. (1998). *Overview of factor analysis.* Retrieved 21 January 2016, from `http://www.stat-help.com/notes.html`

Dhooge, E., & Hartsuiker, R. J. (2010). The distractor frequency effect in picture–word interference: Evidence for response exclusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(4), 878. doi: 10.1037 /a0019128

Dhooge, E., & Hartsuiker, R. J. (2011). The distractor frequency effect in a delayed picture-word interference task: Further evidence for a late locus of distractor exclusion. *Psychonomic Bulletin & Review*, *18*(1), 116–122. doi: 10.3758/s13423-010-0026-0

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

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, *14*(20), 1–11.

Dutilh, G., van Ravenzwaaij, D., Nieuwenhuis, S., van der Maas, H. L., Forstmann, B. U., & Wagenmakers, E.-J. (2012). How to measure post-error slowing: A confound and a simple solution. *Journal of Mathematical Psychology*, *56*(3), 208–216. doi: 10.1016/j.jmp.2012.04.001

Earles, J. L., & Salthouse, T. A. (1995). Interrelations of age, health, and speed. *The*

Journals of Gerontology Series B: Psychological Sciences and Social Sciences, *50*(1), P33–P41. doi: 10.1093/geronb/50B.1.P33

Ellis, E. M., Gonzalez, M. R., & Deak, G. O. (2014). Visual prediction in infancy: what is the association with later vocabulary? *Language Learning and Development*, *10*(1), 36–50.

Engelhardt, P. E., Nigg, J. T., & Ferreira, F. (2013). Is the fluency of language outputs related to individual differences in intelligence and executive function? *Acta psychologica*, *144*(2), 424–432. doi: 10.1016/j.actpsy.2013.08.002

Federmeier, K. D., McLennan, D. B., Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*(2), 133–146. doi: 10.1017.S004857720139203X

Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Rockette, H. E., & Paradise, J. L. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child development*, *76*(4), 856–868. doi: 10.1111/j.1467-8624.2005.00882.x

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental psychology*, *42*(1), 98. doi: 10.1037/0012-1649.42.1.98

Fox, J., et al. (2003). Effect displays in r for generalised linear models. *Journal of statistical software*, *8*(15), 1–27.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*(4), 513–543. doi: 10.1017/S0142716406060383

Gupta, P., & Tisdale, J. (2009a). Does phonological short-term memory causally determine vocabulary learning? toward a computational resolution of the debate. *Journal of memory and language*, *61*(4), 481–502. doi: 10.1016/j.jml.2009.08.001

Gupta, P., & Tisdale, J. (2009b). Word learning, phonological short-term memory, phonotactic probability and long-term memory: towards an integrated framework. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3755–3771. doi: 10.1098/rstb.2009.0132

Gyllstad, H., Vilkaite, L., & Schmitt, N. (2015). Assessing vocabulary size through multiple-choice formats: Issues with guessing and sampling rates. *ITL - International Journal of Applied Linguistics*, *166*(2), 278–306. doi: 10.1075/itl.166.2.04gyl

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes.

Psychological review, *111*(3), 662. doi: 10.1037/0033-295X.111.3.662
Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? the asynchronous rise and fall of different cognitive abilities across the life span. *Psychological science*, *26*(4), 433–443. doi: 10.1177/0956797614567339
Hartsuiker, R. J., & Barkhuysen, P. N. (2006). Language production and working memory: The case of subject-verb agreement. *Language and Cognitive Processes*, *21*(1-3), 181–204. doi: 10.1080/01690960400002117
Hawkins, E., Astle, D. E., & Rastle, K. (2015). Semantic advantage for learning new phonological form representations. *Journal of cognitive neuroscience*, *27*(4), 775–786. doi: 10.1162/jocn$_{a0}$0730
Heathcote, A., Brown, S., & Cousineau, D. (2004). Qmpe: Estimating lognormal, wald, and weibull rt distributions with a parameter-dependent lower bound. *Behavior Research Methods*, *36*(2), 277–290. doi: 10.3758/BF03195574
Heathcote, A., Popiel, S. J., & Mewhort, D. (1991). Analysis of response time distributions: An example using the stroop task. *Psychological Bulletin*, *109*(2), 340. doi: 10.1037/0033-2909.109.2.340
Henderson, L.-M., & James, E. (2018). Consolidating new words from repetitive versus multiple stories: Prior knowledge matters. *Journal of experimental child psychology*, *166*, 465–484. doi: 10.1016/j.jecp.2017.09.017
Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in second language acquisition*, *21*(02), 303–317. doi: 10.1017/S0272263199002089
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
Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, *31*(1), 80–93. doi: 10.1080/23273798.2015.1047459
Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? links between maternal talk, processing speed and vocabulary size in spanish-learning children. *Developmental science*, *11*(6). doi: 10.1111/j.1467-7687.2008.00768.x
Hutson, J., Damian, M. F., & Spalek, K. (2013). Distractor frequency effects in picture–word interference tasks with vocal and manual responses. *Language and Cognitive Processes*, *28*(5), 615–632. doi: 10.1080/01690965.2011.605599
Janse, E., & Adank, P. (2012). Predicting foreign-accent adaptation in older adults. *The Quarterly Journal of Experimental Psychology*, *65*(8), 1563–1585. doi: 10.1080/17470218.2012.658822

Janse, E., & Jesse, A. (2014). Working memory affects older adults' use of context in spoken-word recognition. *The Quarterly Journal of Experimental Psychology*, *67*(9), 1842–1862. doi: 10.1080/17470218.2013.879391

Jones, G., & Rowland, C. F. (2017). Diversity not quantity in caregiver speech: Using computational modeling to isolate the effects of the quantity and the diversity of the input on vocabulary growth. *Cognitive psychology*, *98*, 1–21. doi: 10.1016/j.cogpsych.2017.07.002

Jongman, S. R., Roelofs, A., & Meyer, A. S. (2015). Sustained attention in language production: An individual differences investigation. *The Quarterly Journal of Experimental Psychology*, *68*(4), 710–730. doi: 10.1080/17470218.2014.964736

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta psychologica*, *86*(2-3), 199–225. doi: 10.1016/0001-6918(94)90003-5

Kello, C. T., & Plaut, D. C. (2003). Strategic control over rate of processing in word reading: A computational investigation. *Journal of Memory and Language*, *48*(1), 207–232. doi: 10.1016/S0749-596X(02)00512-0

Kerkhoff, A., Poelmans, P., de Jong, J. H., & Lennig, M. (2005). *Verantwoording toets gesproken nederlands. ontwikkeld in opdracht van het ministerie van justitie van het koninkrijk der nederlanden.* 's-Hertogenbosch: Cinop.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633. doi: 10.3758/BRM.42.3.627

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*(3), 643–650. doi: 10.3758/BRM.42.3.643

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*(1), 287–304. doi: 10.3758/s13428-011-0118-4

Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *The Quarterly Journal of Experimental Psychology*, *68*(8), 1665–1692. doi: 10.1080/17470218.2015.1022560

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, *22*(2), 154 - 169. doi: 10.1016/j.tics.2017.11.006

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

Kurdziel, L. B., & Spencer, R. M. (2016). Consolidation of novel word learning in native english-speaking adults. *Memory*, *24*(4), 471–481. doi: 10.1080/09658211.2015.1019889

La Heij, W., Starreveld, P. A., & Steehouwer, L. C. (1993). Semantic interference and orthographic facilitation in definition naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 352.

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 12–31. doi: 10.1037/0278-7393.34.1.12

Lindsay, S., & Gaskell, M. G. (2013). Lexical integration of novel words without sleep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 608–622. doi: 10.1037/a0029243

Luteijn, F., & van der Ploeg, F. (1983). *Handleiding groninger intelligentietest (manual groningen intelligence test).* Lisse, The Netherlands: Swets and Zeitlinger.

Mainz, N., Shao, Z., Brysbaert, M., & Meyer, A. S. (2017). Vocabulary knowledge predicts lexical processing: Evidence from a group of participants with diverse educational backgrounds. *Frontiers in Psychology*, *8*, 1164. Retrieved from `http://journal.frontiersin.org/article/10.3389/fpsyg.2017.01164` doi: 10.3389/fpsyg.2017.01164

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental science*, *11*(3), F9–F16. doi: 10.1111/j.1467-7687.2008.00671.x

McClelland, J. L., Rumelhart, D., & the PDP Research Group. (1986). Parallel distributed processing: Explorations in the microstructure of cognition. vol. i: Foundations, and vol. ii: Psychological and biological models. *Cambridge, MA: MIT Press*.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological review*, *119*(4), 831. doi: 10.1037/a0029872

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to wordnet: An on-line lexical database. *International journal of lexicography*, *3*(4), 235–244.

Miozzo, M., & Caramazza, A. (2003). When more is less: a counterintuitive effect of distractor frequency in the picture-word interference paradigm. *Journal of Experimental Psychology: General*, *132*(2), 228. doi: 10.1037/0096-3445.132.2.228

Monaghan, P., Chang, Y.-N., Welbourne, S., & Brysbaert, M. (2017). Exploring the relations between word frequency, language exposure, and bilingualism in a computational model of reading. *Journal of Memory and Language*, *93*, 1–21. doi: 10.1016/j.jml.2016.08.003

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading research quarterly*, 233–253. doi: 10.2307/747758

Perfetti, C. A., & Hart, L. (2001). The lexical basis of comprehension skill. In D. S. Gorfien (Ed.), *On the consequences of meaning selection* (p. 67-86). Washington, DC: American Psychological Association.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven (Ed.), *Precursors of functional literacy* (p. 189-213). Philadelphia: John Benjamins.

Piai, V., Roelofs, A., & Schriefers, H. (2011). Semantic interference in immediate and delayed naming and reading: Attention and task decisions. *Journal of Memory and Language*, *64*(4), 404–423. doi: 10.1016/j.jml.2011.01.004

Piai, V., Roelofs, A., & Schriefers, H. (2012). Distractor strength and selective attention in picture-naming performance. *Memory & Cognition*, *40*(4), 614–627. doi: 10.3758/s13421-011-0171-3

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological review*, *103*(1), 56. doi: 10.1037/0033-295X.103.1.56

R Core Team. (2016). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review*, *85*(2), 59.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A Diffusion Model Account of the Lexical Decision Task. *Psychological Review*, *111*(1), 159–182. doi: 10.1037/0033-295X.111.1.159

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, *20*(4), 873–922. doi: 10.1162/neco.2008.12-06-420

Raven, J., Raven, J., & Court, J. (1998). *Manual for raven's progressive matrices and vocabulary scales. section 5: The mill hill vocabulary scale.* Oxford, England: : Oxford Psychologists Press/San Antonio, TX: The Psychological Corporation.

Rodriguez-Aranda, C., & Jakobsen, M. (2011). Differential contribution of cognitive and psychomotor functions to the age-related slowing of speech production. *Journal of the International Neuropsychological Society*, *17*(05), 807–821. doi: 10.1017/S1355617711000828

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking.

Cognition, *42*(1-3), 107–142. doi: 10.1016/0010-0277(92)90041-F

Roelofs, A. (2003). Goal-referenced selection of verbal action: modeling attentional control in the stroop task. *Psychological review*, *110*(1), 88. doi: 10.1037/0033-295X.110.1.88

Roelofs, A. (2012). Attention, spatial integration, and the tail of response time distributions in stroop task performance. *The Quarterly Journal of Experimental Psychology*, *65*(1), 135–150. doi: 10.1080/17470218.2011.605152

Roelofs, A., & Piai, V. (2017). Distributional analysis of semantic interference in picture naming. *The Quarterly Journal of Experimental Psychology*, *70*(4), 782–792. doi: 10.1080/17470218.2016.1165264

Roelofs, A., Piai, V., & Schriefers, H. (2011). Selective attention and distractor frequency in naming performance: comment on dhooge and hartsuiker. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4). doi: 10.1037/a0023328

Rogers, T. T., & McClelland, J. L. (2014). Parallel distributed processing at 25: Further explorations in the microstructure of cognition. *Cognitive science*, *38*(6), 1024–1077. doi: 10.1111/cogs.12148

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, *323*(6088), 533.

Sahni, S. (2010). Who's afraid of similarity? effects of phonological and semantic similarity on lexical acquisition. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 32).

Sailor, K. M. (2013). Is vocabulary growth influenced by the relations among words in a language learner's vocabulary? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1657. doi: 10.1037/a0032993

Salthouse, T. A. (1993). Speed and knowledge as determinants of adult age differences in verbal tasks. *Journal of Gerontology*, *48*(1), P29–P36.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological review*, *103*(3), 403. doi: 10.1037/0033-295X.103.3.403

Salthouse, T. A. (2000). Aging and measures of processing speed. *Biological Psychology*, *54*(1), 35 - 54. doi: 10.1016/S0301-0511(00)00052-1

Savill, N., Ellis, A. W., & Jefferies, E. (2017). Newly-acquired words are more phonologically robust in verbal short-term memory when they have associated semantic representations. *Neuropsychologia*, *98*, 85–97. doi: 10.1016/j.neuropsychologia.2016.03.006

Scaltritti, M., Navarrete, E., & Peressotti, F. (2015). Distributional analyses in the picture–word interference paradigm: Exploring the semantic interference and the distractor frequency effects. *The Quarterly Journal of Experimental Psychology*, *68*(7), 1348–1369. doi: 10.1080/17470218.2014.981196

Schlichting, L. (2005). *Peabody picture vocabulary test, dutch version.* Amsterdam: Pearson Assessment and Information BV.

Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of memory and language*, *29*(1), 86–102. doi: 10.1016/0749-596X(90)90011-N

Schroeder, D. H., & Salthouse, T. A. (2004). Age-related effects on cognition between 20 and 50 years of age. *Personality and individual differences*, *36*(2), 393–404. doi: 10.1016/S0191-8869(03)00104-1

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*. doi: 10.3389/fpsyg.2014.00772

Shao, Z., Meyer, A. S., & Roelofs, A. (2013). Selective and nonselective inhibition of competitors in picture naming. *Memory & Cognition*, *41*(8), 1200–1211. doi: 10.3758/s13421-013-0332-7

Shao, Z., Roelofs, A., Acheson, D. J., & Meyer, A. S. (2014). Electrophysiological evidence that inhibition supports lexical selection in picture naming. *Brain Research*, *1586*, 130–142. doi: 10.1016/j.brainres.2014.07.009

Shao, Z., Roelofs, A., Martin, R. C., & Meyer, A. S. (2015). Selective inhibition and naming performance in semantic blocking, picture-word interference, and color–word stroop tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1806. doi: 10.1037/a0039363

Shao, Z., Roelofs, A., & Meyer, A. S. (2012). Sources of individual differences in the speed of naming objects and actions: The contribution of executive control. *The Quarterly Journal of Experimental Psychology*, *65*(10), 1927–1944. doi: 10.1080/17470218.2012.670252

Shipley, W. (1946). *The shipley institute of living scale. los angeles: Western psychological services.* Los Angeles: Western Psychological Services.

Smith, A. C., Monaghan, P., & Huettig, F. (2014). Literacy effects on language and vision: Emergent effects from an amodal shared resource (asr) computational model. *Cognitive Psychology*, *75*, 28–54. doi: 10.1016/j.cogpsych.2014.07.002

Sternberg, R. J. (1987). Most vocabulary is learned from context. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 89–106). Hillsdale, NJ: Erlbaum.

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$_3$

Storkel, H. L. (2009). Developmental differences in the effects of phonological, lexical and semantic variables on word learning by infants. *Journal of child language*, *36*(2), 291–321. doi: 10.1017/S030500090800891X

Storkel, H. L., & Adlof, S. M. (2009). The effect of semantic set size on word learning by preschool children. *Journal of Speech, Language, and Hearing Research*, *52*(2), 306–320.

Storkel, H. L., Bontempo, D. E., & Pak, N. S. (2014). Online learning from input versus offline memory evolution in adult word learning: Effects of neighborhood density and phonologically related practice. *Journal of speech, language, and hearing research*, *57*(5), 1708–1721. doi: 10.1044/2014$_JSLHR-L-13-0150$

Takashima, A., Bakker, I., Van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Richness of information about novel words influences how episodic and semantic memory networks interact during lexicalization. *NeuroImage*, *84*, 265–278. doi: 10.1016/j.neuroimage.2013.08.023

Takashima, A., Bakker, I., Van Hell, J. G., Janzen, G., & McQueen, J. M. (2017). Interaction between episodic and semantic memory networks in the acquisition and consolidation of novel spoken words. *Brain and language*, *167*, 44–60. doi: 10.1016/j.bandl.2016.05.009

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2011). Variation in verbal fluency: A latent variable analysis of clustering, switching, and overall performance. *The Quarterly Journal of Experimental Psychology*, *64*(3), 447–466. doi: 10.1080/17470218.2010.505292

Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition*, *131*(3), 373–403. doi: 10.1016/j.cognition.2014.01.007

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 735. doi: 10.1037/0278-7393.28.4.735

Vitevitch, M. S., & Sommers, M. S. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory & Cognition*, *31*(4), 491–504. doi: 10.3758/BF03196091

Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods*, *39*(4), 767–775. doi: 10.3758/BF03192967

Wechsler, D. (1997). *Wechsler adult intelligence scale–third edition (wais–iii).*

Weighall, A., Henderson, L.-M., Barr, D., Cairney, S. A., & Gaskell, M. G. (2017). Eye-tracking the time-course of novel word learning and lexical competition in adults and children. *Brain and language*, *167*, 13–27. doi: 10.1016/j.bandl.2016.07.010

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological science*, *24*(11), 2143–2152. doi: 10.1177/0956797613488145

Whaley, C. P. (1978). Word — nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, *17*(2), 143–154. doi: 10.1016/S0022-5371(78)90110-X

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

Yap, M. J., Lim, G. Y., & Pexman, P. M. (2015). Semantic richness effects in lexical decision: the role of feedback. *Memory & cognition*, *43*(8), 1148–1167. doi: 10.3758/s13421-015-0536-0

Yap, M. J., Tan, S. E., Pexman, P. M., & Hargreaves, I. S. (2011). Is more always better? effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychonomic Bulletin & Review*, *18*(4), 742–750. doi: 10.3758/s13423-011-0092-y

Yap, M. J., Tse, C.-S., & Balota, D. A. (2009). Individual differences in the joint effects of semantic priming and word frequency revealed by rt distributional analyses: The role of lexical integrity. *Journal of Memory and Language*, *61*(3), 303–325. doi: 10.1016/j.jml.2009.07.001

Yates, M., Locker, L., & Simpson, G. B. (2004). The influence of phonological neighborhood on visual word perception. *Psychonomic Bulletin & Review*, *11*(3), 452–457. doi: 10.3758/BF03196594

Nederlandse samenvatting

De kennis van woorden is een belangrijk onderdeel van onze taalbeheersing. Zowel onderwijsgerelateerde als beroepsmatige prestaties zijn in grote mate afhankelijk van in hoeverre we geschreven en gesproken instructies kunnen begrijpen, alsook van of we onze gedachten en meningen duidelijk kunnen communiceren.

Het aantal woorden dat we kennen is enorm. Een gemiddelde twintigjarige moedertaalspreker van het Amerikaans-Engels kent naar schatting 42.000 lemma's, oftewel de niet-verbogen woordvormen waaruit verbuigingen zijn afgeleid (Brysbaert et al., 2016b). Echter, ook volwassenen blijven nieuwe woorden bijleren. Sterker nog, dikwijls verbetert onze woordenschat naarmate we steeds ouder worden (Brysbaert et al., 2016b; Schroeder & Salthouse, 2004). In de loop der jaren worden we dus blootgesteld aan een opvallend groot aantal woorden, die we vervolgens leren en, belangrijker nog, ook onthouden. Naar schatting leert de gemiddelde volwassene er ongeveer 6000 nieuwe woorden bij tussen zijn twintigste en zestigste. Dat betekent dat moedertaalsprekers gedurende deze veertig jaar gemiddeld om de dag een nieuw lemma leren (Brysbaert et al., 2016b; Keuleers et al., 2015).

Daarom is onze woordenschat een voortdurend toenemende inventaris van woorden om de wereld om ons heen te omschrijven, een verzameling uitdrukkingen voor allerlei voor ons relevante ervaringen en informatie. Als zodanig ligt het voor de hand om aan te nemen dat er variatie is in hoeveel en welke woorden mensen kennen. Vaktaal is een goed voorbeeld van de sterke effecten die ervaring kan hebben op het type woorden dat iemand kent. De woorden *syncope* of *lunge* zijn bijvoorbeeld woorden die mensen met bepaalde beroepen, interesses of hobby's kennen, maar ze zijn volledig ondoorzichtig voor anderen.[1] Een andere factor die beïnvloedt hoeveel en welke soort woorden we kennen is onderwijsachtergrond. Het is aangetoond dat opleidingsniveau een significante invloed heeft op de woordenschat van mensen over de gehele leeftijdsspanne. Hogere opleidingsniveaus worden geassocieerd met een grotere woordenschat. Dit is mogelijk toe te schrijven aan de gevolgen van veel lezen en studeren – aspecten die worden geassocieerd

[1] In dit geval is syncope een woord dat muzikanten en wellicht ook dansers kennen en het duidt op een verandering van het ritme in een muziekstuk, waarbij de zware tellen licht worden en de lichte tellen zwaar. Lunge is een dansbeweging waarbij één been gebogen is en het andere been uitgestrekt.

met formeel onderwijs – op het groeien van de woordenschat (Brysbaert et al., 2016b; Keuleers et al., 2015).

Woordenschat

Eerder onderzoek heeft aangetoond dat variatie in het aantal woorden dat mensen kennen niet alleen invloed heeft op welke ongebruikelijke woorden iemand kent en begrijpt. Ook de verwerking van veel voorkomende woorden die bijna alle volwassenen kennen wordt beïnvloed door het aantal woorden dat iemand kent. Van mensen met een betere woordenschat is vaak aangetoond dat ze ook beter zijn in het verwerken van taal (Bent et al., 2016; Janse & Jesse, 2014; Rodriguez-Aranda & Jakobsen, 2011; Yap et al., 2012). Dat betekent dat ze sneller en/of nauwkeuriger zijn in het uitvoeren van verschillende taken die taalproductie en taalbegrip testen. Het eerste deel van dit proefschrift (Hoofdstukken 2, 3 en 4) bestudeert de relatie tussen woordenschat en taalverwerking nader.

Een belangrijke vraag die het bestuderen van woordenschat en taalverwerking oproept is: hoe kan woordkennis het best worden gemeten? De meeste eerdere studies hebben slechts één woordenschattoets gebruikt. Er is echter betoogd dat verscheidene toetsen noodzakelijk zijn om een complexe vaardigheid als woordkennis te kunnen kwantificeren, met bijvoorbeeld zowel meerkeuze- als open vragen. Daarom is hier gebruikgemaakt van niet één, maar zeven verschillende woordenschattoetsen, waarvan er vijf werden ontwikkeld voor de eerste experimenten in dit proefschrift.

In Hoofdstuk 2 en 3 werd de relatie tussen variatie in woordenschat en taalbegrip onderzocht. Taalbegrip werd gemeten aan de hand van een gangbare taak in de psycholinguïstiek, *lexicale decisie* genoemd. In deze taak zien proefpersonen letterreeksen op een computerscherm en vervolgens moeten ze middels knoppen aangeven of de reeks op het scherm een bestaand al dan niet een Nederlands woord vormt. Al met al zijn de bevindingen van dit experiment in overeenstemming met eerder onderzoek: individuen met een betere woordenschat presteren ook beter op deze taak.

Een belangrijk aspect waarin deze experimenten eerdere studies aanvullen is dat hier niet alleen universitaire studenten werden getest (Hoofdstuk 2), maar ook jongvolwassenen met andere achtergronden (Hoofdstuk 3). Het meeste wat we weten over taalverwerking is gebaseerd op onderzoek met universitaire bachelorstudenten als proefpersonen. Men kan zich afvragen of de bevindingen van dat onderzoek gegeneraliseerd kunnen worden naar de gemiddelde jongvolwassene. De reden hiervoor is dat universitaire studenten bijvoorbeeld vaker worden blootgesteld aan geschreven taal dan de gemiddelde jongvolwassene, hetgeen waarschijnlijk weer hun taalvaardigheid beïnvloedt.

Het experiment in Hoofdstuk 3 liet zien dat de scores op woordenschat en lexicale decisie (oftewel de nauwkeurigheid en snelheid van de gemaakte beslissingen) veel grotere variatie vertoonden in de groep van niet-universitaire leerlingen. Dat betekent dat er een grote hoeveelheid variatie in de vaardigheden van mensen onopgemerkt blijft wanneer alleen universitaire studenten worden getest. Daarnaast was het verband tussen de woordenschatgrootte en taalverwerking (grotere woordenschat betekent betere taalverwerking) tussen niet-universitaire leerlingen en universitaire studenten vergelijkbaar, maar niet hetzelfde. De boodschap hier is dat het noodzakelijk is niet alleen de "gangbare" groep proefpersonen (universitaire studenten) te toetsen, maar te streven naar psychologische en psycholinguïstische studies waarin proefpersonen met uiteenlopender achtergronden deelnemen.

In Hoofdstuk 4 werd getest of het hebben van een grotere woordenschat ook gunstig is voor taalproductie. Ook hier werden proefpersonen getoetst op de bovengenoemde zeven woordenschattoetsen. Daarnaast deden ze een taak waarbij ze plaatjes benoemden en getest werd hoe snel ze dat deden. De resultaten van dit experiment toonden aan dat mensen met betere woordenschatscores ook sneller waren in het benoemen van plaatjes. Daarmee is niet alleen taalbegrip, maar ook taalproductie verbonden met woordenschat, en een grotere woordenschat is voor beide voordelig.

Woordenschatverwerving

De eerste drie hoofdstukken lieten zien dat er aanzienlijke variatie bestaat in de woordkennis van moedertaalsprekers van een taal, met name wanneer men kijkt naar mensen met verschillende achtergronden. Daarnaast werd vastgesteld dat die diversiteit invloed heeft op taalverwerking: een grotere woordenschat wordt geassocieerd met sneller en nauwkeuriger taalbegrip en taalproductie. Een vraag die hieruit volgt is wat de oorzaak is van de variatie in woordenschat. Waarom leren sommige mensen meer woorden dan anderen, waardoor ze uiteindelijk een grotere woordenschat hebben? Zoals hierboven beschreven worden zowel leeftijd als onderwijs gekoppeld aan verschillen in woordenschat. Bovendien is er een toenemend aantal studies naar de woordenschatverwerving van kinderen, die aantonen dat zowel de hoeveelheid blootstelling aan taal als de snelheid waarmee taal verwerkt wordt tot variatie leiden in de woordenschat van kinderen (Fernald et al., 2006; Jones & Rowland, 2017; Marchman & Fernald, 2008). Er is echter, ondanks het feit dat we weten dat ook volwassenen nieuwe woorden blijven bijleren, weinig onderzoek verricht naar welke factoren hierop van invloed zijn. Deze vraag werd behandeld in Hoofdstuk 5 en 6 van dit proefschrift.

In het experiment in Hoofdstuk 5 leerden proefpersonen nieuwe woorden. Ze kwamen naar het lab om 39 woorden te leren die ieder vergezeld waren van een kleurenfoto met daarop een afbeelding van de betekenis van het woord. Alle woorden waren zeer ongebruikelijke woorden van het Nederlands en als zodanig onbekend aan 90% van de Nederlanders en daarmee ook aan de proefpersonen in het experiment. Alle proefpersonen werden getraind op deze woorden en meteen daarna getoetst, alsook een week later. Tijdens de training kwamen woorden ofwel drie keer voor, ofwel negen of 16 keer. Het doel van het experiment was met name om de verbanden te onderzoeken tussen leerprestaties, slaap, de mate van blootstelling aan nieuwe woorden en ook andere cognitieve vaardigheden. De cognitieve vaardigheden waarop getoetst werd waren woordenschat, algemene intelligentie, verwerkingssnelheid en kortetermijngeheugen. Woordenschat werd geëvalueerd in twee van de eerder gebruikte testen en algemene intelligentie werd gemeten met een onderdeel van een bekende IQ-test. Verwerkingssnelheid werd gemeten in drie taken waarbij proefpersonen zo snel mogelijk een knop moesten indrukken na het zien van een plaatje op een computerscherm of het horen van een piepgeluid door de koptelefoon. Het kortetermijngeheugen waarin we geïnteresseerd waren heet het *fonologische kortetermijngeheugen* en duidt op de vaardigheid om gesproken getallenreeksen of letterreeksen te onthouden. In de taak waarin dit werd getoetst hoorden proefpersonen gesproken getallenreeksen variërend tussen de twee en negen cijfers lang en moesten deze vervolgens intypen.

De resultaten van het woordenschatexperiment lieten zien dat proefpersonen woorden beter onthielden als ze daar vaker op waren getraind. Bovendien waren prestaties beter in de tweede test, een week na de training, dan onmiddellijk na de training. Dit is een typisch resultaat: slaap helpt om nieuw geleerde woorden beter te onthouden. Ten slotte leerden proefpersonen met betere non-verbale intelligentie en een grotere woordenschat in hun moedertaal gemakkelijker nieuwe woorden.

Maar waarom is een grotere woordenschat of hogere intelligentie voordelig voor het leren van nieuwe woorden? Dit werd onderzocht in Hoofdstuk 6 aan de hand van een computationeel model dat de cognitieve processen van taalbegrip en taalproductie simuleert. Het voordeel van computationeel modelleren is dat factoren of vaardigheden die moeilijk te meten en te controleren zijn in mensen (zoals de mate van taalblootstelling en verwerkingssnelheid) expliciet kunnen worden gemanipuleerd. In ons model werd een aantal factoren gemanipuleerd dat mogelijk leidt tot verschillen in woordenschatverwerving. Het model simuleerde bijvoorbeeld variatie in algemene intelligentie en verwerkingssnelheid. Deze studie toonde aan dat zowel een snellere verwerking als een hogere algemene intelligentie een grotere woordenschat tot gevolg hebben. Sterker nog, de netwerken met hogere verwerkingssnelheden en met hogere intelligentie behaalden dit voordeel juist omdat ze niet alleen de makkelijke, maar ook de moeilijkere woorden leerden.

Het model werd ook gebruikt om de bevinding nader te analyseren dat individuen met een grotere woordenschat een snellere taalverwerking hebben en op volwassen leeftijd beter zijn in het leren van nieuwe woorden. Het model liet zien dat het niet per se een grotere woordenschat is die leidt tot een snellere taalverwerking en betere woordenschatverwerving op volwassen leeftijd. Integendeel is het misschien juist toe te schrijven aan onderliggende factoren, zoals een grotere mate van blootstelling, snellere verwerking of hogere intelligentie. Een woordenschattoets meet bij mensen mogelijk niet alleen de grootte van de woordenschat, maar ook deze onderliggende factoren. Dit is wellicht de reden dat er een sterk verband is tussen de variatie in taalverwerking en die in woordenschatverwerving.

Samenvatting

Samengevat toont mijn onderzoek aan dat er veel variatie is in de grootte van de woordenschat van moedertaalsprekers van een taal. Deze verschillen in woordkennis beïnvloeden taalverwerking: mensen met een grotere woordenschat zijn sneller en nauwkeuriger in het produceren en het begrijpen van woorden. Daarnaast is aangetoond dat het noodzakelijk is niet alleen universitaire bachelorstudenten te testen, die doorgaans de proefpersonen vormen in psycholinguïstisch onderzoek. Niet alle bevindingen uit studies met universitaire studenten zijn te generaliseren naar groepen met uiteenlopender achtergronden. Het spectrum van vaardigheden zoals woordenschatscores en de variatie in taalverwerking worden onderschat wanneer de focus alleen ligt op universitaire studenten.

Ten slotte is hier onderzocht wat de oorzaak is van variatie in woordenschatgrootte. Waarom leren sommige mensen meer woorden dan andere mensen en hebben ze zodoende een grotere woordenschat? Deze studies tonen aan dat individuen met een grotere woordenschat gemakkelijker nieuwe woorden leren. Dit is waarschijnlijk toe te schrijven aan onderliggende factoren, zoals de mate van blootstelling, intelligentie en verwerkingssnelheid. Het gebruikte computationele model liet zien dat een grotere mate van blootstelling, een hogere intelligentie en een snellere taalverwerking allemaal leidden tot het leren van een groter aantal woorden enerzijds, en anderzijds tot snellere taalverwerking en betere prestaties bij het leren van nieuwe woorden. Zodoende belicht dit proefschrift complexe relaties tussen verscheidene cognitieve vaardigheden, woordenschatverwerving en woordenschatgrootte en taalverwerking. Toekomstig onderzoek zal deze aspecten van één van de meest fascinerende vaardigheden van de mens, namelijk het vermogen om te leren en taal te gebruiken, verder moeten onderzoeken.

(Translated from English by Merel Maslowski)

Deutsche Zusammenfassung

Wortwissen oder Vokabular ist ein wichtiger Aspekt unseres Sprachvermögens. Sowohl schulischer als auch beruflicher Erfolg sind zu einem bedeutenden Teil abhängig davon, ob wir dazu in der Lage sind, geschriebene und gesprochene Sprache zu verstehen (etwa in Aufgabenstellungen oder Anweisungen) und unsere Gedanken oder Meinungen deutlich zu kommunizieren.

Als Muttersprachler einer Sprache kennen wir eine beeindruckende Anzahl von Wörtern. Im Durchschnitt kennt ein 20 Jahre alter Muttersprachler des Amerikanischen Englisch 42.000 Lemmas, das sind nicht gebeugte Wortgrundformen (Brysbaert et al., 2016b). Das Erlernen von Wörtern ist aber nicht irgendwann im jungen Erwachsenenalter abgeschlossen. Ganz im Gegenteil: Die Größe des Vokabulars nimmt mit dem Alter zu (Brysbaert et al., 2016b; Schroeder & Salthouse, 2004). Über die Jahre sind wir einer beachtlichen Anzahl verschiedener Wörter ausgesetzt, von denen wir viele behalten. Im Alter zwischen 20 und 60 Jahren lernt ein Erwachsener schätzungsweise 6.000 neue Wörter. Das bedeutet, dass wir in dieser Zeit etwa alle zwei Tage ein neues Wort lernen (Brysbaert et al., 2016b; Keuleers et al., 2015).

Unser Vokabular ist also ein ständig wachsender Wortschatz, der es uns erlaubt, uns über die Welt um uns herum und Gedanken und Gefühle auszutauschen. Dabei sind für unterschiedliche Menschen die unterschiedlichsten Dinge relevant – es liegt also nahe anzunehmen, dass Menschen sich auch darin unterscheiden, welche und wie viele Worte sie kennen. Fachworte sind ein sehr intuitives Beispiel dafür, wie stark sich der Wortschatz verschiedener Menschen mit verschiedenen Erfahrungen unterscheidet. Die Worte *Synkope* oder *lunge* (engl.) sind zum Beispiel Menschen mit bestimmten Berufen, Interessen oder Hobbies bekannt, wohingegen andere diese Ausdrücke noch nie gehört haben.[2] Ein weiterer Faktor, von dem wir wissen, dass er die Größe des Wortschatzes von Menschen allen Alters beeinflusst, ist Bildung. Verschiedene Studien haben gezeigt, dass höhere Bildungsniveaus im Zusammenhang stehen mit einem größeren Vokabular. Das ist unter anderem darauf zurückzuführen, dass beispielsweise die Ausbildung an einer

[2] *Synkope* ist ein Wort, dass Musiker oder vielleicht auch Tänzer kennen. Es beschreibt eine Veränderung des Betonungsschemas einer Musikstücks, bei dem die eigentlich unbetonten Schläge betont werden. *Lunge* (engl.) ist der Name einer Tanzbewegung bzw. -position, bei der ein Bein gebeugt und das andere gestreckt ist.

Universität mit einem umfangreichen Lesepensum verbunden ist, welches wiederum die Vokabulargröße beeinflusst (Brysbaert et al., 2016b; Keuleers et al., 2015).

Wortwissen

Studien haben gezeigt, dass Unterschiede in der Größe des Wortschatzes nicht nur Auswirkungen darauf haben, welche seltenen und sehr spezifischen Worte Menschen kennen und verstehen (etwa *Serendipität*, was so viel bedeutet wie *glücklicher Zufall*). Auch die Verarbeitung von sehr häufig auftretenden Worten, die im Wortschatz fast aller Erwachsener enthalten sind (etwa *Zufall*), variiert abhängig davon wie viele Worte ein Erwachsener kennt. Menschen mit einem umfangreicheren Vokabular zeigen eine bessere Sprachverarbeitung als Menschen mit geringerem Wortwissen (Bent et al., 2016; Janse & Jesse, 2014; Rodriguez-Aranda & Jakobsen, 2011; Yap et al., 2012). Das bedeutet, dass Erwachsene mit einem großen Vokabular in Aufgaben zur Messung von Sprachverstehen und -produktion gewöhnlich weniger Fehler machen und schneller sind. Im ersten Teil meiner Doktorarbeit habe ich das Verhältnis zwischen Vokabelwissen und Sprachverarbeitung näher untersucht (Kapitel 2, 3 & 4).

Wenn man Wortwissen und Sprachverarbeitung erforschen möchte, muss man sich eine wichtige Frage stellen: Wie misst man Vokabular am besten? In den meisten der früheren Studien mussten die Testpersonen nur einen einzigen Vokabeltest absolvieren. Da Wortwissen aber sehr komplex und nicht einfach zu messen ist, haben verschiedene Forscher argumentiert, dass ein einzelner Test zur Messung von Vokabular nicht ausreichend ist. Stattdessen wurde vorgeschlagen, dass verschiedene Tests, etwa Multiple-Choice und offene Vokabeltests, erforderlich sind. Aus diesem Grund habe ich sieben verschiedene Vokabeltests verwendet, von denen ich fünf selbst entwickelt habe.

In den Kapiteln 2 und 3 habe ich das Verhältnis zwischen Wortwissen und Sprachverstehen untersucht. Zur Messung von Sprachverstehen wurde die in der psycholinguistischen Forschung weit verbreitete Lexikalische Entscheidungsaufgabe verwendet. In dieser Aufgabe sehen die Testpersonen immer je eine Buchstabenreihe auf dem Bildschirm. Sie müssen entscheiden, ob es sich dabei um ein sinnvolles Wort oder eine sinnlose Aneinanderreihung von Buchstaben handelt, und dieses Urteil per Druck auf eine Reaktionstaste abgeben. Insgesamt stehen die Ergebnisse meiner Experimente in Einklang mit denen früherer Studien: Personen mit besserem Vokabular schnitten in dieser Aufgabe zur Messung von Sprachverstehen besser ab, das heißt sie waren schneller als Personen mit geringerem Wortwissen.

Ein wichtiger Unterschied zwischen meinen Experimenten und früheren Untersuchungen ist, dass ich nicht nur Universitätsstudenten (Kapitel 2) getestet habe, sondern auch junge Erwachsene unterschiedlicherer Hintergründe, genauer gesagt

Berufsschüler (Kapitel 3). Der Großteil dessen, was wir über Sprachverarbeitung wissen, basiert auf Studien, in denen Universitätsstudenten getestet wurden. In dem Zusammenhang kann man sich aber fragen, ob sich das, was wir aus diesen Studien gelernt haben, auf den durchschnittlichen jungen erwachsenen Muttersprachler verallgemeinern lässt. Der Grund dafür ist, dass Universitätsstudenten beispielsweise sehr viel mehr geschriebener Sprache ausgesetzt sind und vermutlich ein höheres Lesepensum haben als der durchschnittliche junge Erwachsene, was wiederum die Sprachfertigkeiten dieser jungen Menschen beeinflusst.

Mein Experiment in Kapitel 3 hat gezeigt, dass die Spanne der Ergebnisse bei den Vokabeltests und in der Lexikalischen Entscheidungsaufgabe (d.h. die Anzahl an Fehlern und Schnellheit) in der Gruppe der Berufsschüler deutlich größer war, als in der Gruppe der Universitätsstudenten. Wenn wir also nur Universitätsstudenten testen, übersehen wir eine ganze Menge an Varianz, die Menschen verschiedener Hintergründe in bestimmten Fähigkeiten aufweisen. Zusätzlich war das Verhältnis zwischen Vokabular und Sprachverarbeitung (größeres Vokabular = schnellere und akkuratere Sprachverarbeitung) zwar im Großen und Ganzen ähnlich, unterschied sich im Detail aber doch zwischen den verschiedenen Gruppen von Testpersonen. Die Take-Home Message ist also, dass es notwendig ist, nicht nur die "typischen" Testpersonen (Universitätsstudenten) zu untersuchen, sondern dass wir uns bemühen sollten, auch Testpersonen vielfältigerer Hintergründe in psychologische und psycholinguistische Experimente einzubeziehen.

In Kapitel 4 habe ich untersucht, ob ein großes Vokabular auch positive Auswirkungen auf Sprachproduktion hat. Die Testpersonen in diesem Experiment haben, wie vorher beschrieben, sieben Vokabeltests gemacht. Zusätzlich dazu haben sie eine Sprachproduktionsaufgabe absolviert, in der sie nacheinander verschiedene Bilder auf dem Computerbildschirm sahen, die sie benennen mussten. In dieser Aufgabe haben wir die Benennungsgeschwindigkeit gemessen. Die Ergebnisse dieses Experiments haben gezeigt, dass Personen mit besserem Vokabular schneller darin sind, die Bilder zu benennen. Das macht deutlich, dass Menschen mit größerem Wortwissen nicht nur im Wortverstehen (wie in der Lexikalischen Entscheidungsaufgabe) schneller und fehlerfreier abschneiden, sondern auch schneller in der Wortproduktion sind (wie beim Benennen von Bildern).

Wortlernen

Die ersten drei Kapitel haben gezeigt, dass die Muttersprachler einer Sprache große individuelle Unterschiede im Wortwissen aufweisen, vor allem wenn man Testpersonen unterschiedlicher Bildungshintergründe betrachtet. Außerdem wurde gezeigt, dass diese

Unterschiede im Vokabular einen Einfluss auf Sprachverarbeitungsgeschwindigkeit und -präzision haben: Wer ein größeres Vokabular hat, ist im Verständnis und der Produktion von Worten schneller und präziser. Was sind aber die Gründe für diese so großen Unterschiede im Wortschatz von erwachsenen Muttersprachlern? Warum lernen manche Menschen mehr Worte als andere und haben daher schließlich ein größeres Vokabular als andere? Wie zuvor angemerkt wissen wir, dass sowohl das Alter als auch die Bildung im Zusammenhang stehen mit Unterschieden in der Größe des Vokabulars. Desweiteren gibt es viele Studien zum Wortlernen bei Kindern. Es wurde gezeigt, dass Unterschiede in der Menge der Spracherfahrung von (Klein-)Kindern, sowie Unterschiede in ihrer Sprachverarbeitungsgeschwindigkeit einen Einfluss auf das Wortlernen und die Größe des Wortschatzes von Kindern haben (Jones & Rowland, 2017; Fernald et al., 2006; Marchman & Fernald, 2008). Obwohl bekannt ist, dass wir im Erwachsenenalter weiter ständig neue Worte lernen, gibt es nur wenige Untersuchungen, die sich mit den Einflussfaktoren auf das Wortlernen in Erwachsenen beschäftigen. Dieser Forschungsfrage habe ich mich in den Kapiteln 5 und 6 meiner Doktorarbeit gewidmet.

Das Experiment in Kapitel 5 war eine Wortlern-Studie. In diesem Experiment haben die Testpersonen 39 Worte gelernt, deren Bedeutung jeweils durch ein Foto repräsentiert wurde. Alle Worte waren sehr seltene niederländische Worte, die laut einer Studie weniger als 10% der Muttersprachler in den Niederlanden kennen. Die Testpersonen haben verschiedene Trainingsaufgaben absolviert. Ihre Kenntnis der neu erlernten Worte wurde sowohl direkt nach dem Training als auch eine Woche später getestet. Während des Trainings wurden einige Wort-Bild Paare nur drei Mal gezeigt, wohingegen andere acht Mal und wieder andere sechzehn Mal gezeigt wurden. Das Ziel des Experimentes war es, die Auswirkungen von Schlaf, der Worthäufigkeit (Anzahl der Wiederholungen der einzelnen Worte) und von verschiedenen kognitiven Fähigkeiten auf den Lernerfolg zu untersuchen. Die kognitiven Fähigkeiten, die untersucht wurden, waren Vokabular, allgemeine Intelligenz, allgemeine Verarbeitungsgeschwindigkeit und Kurzzeitgedächtnis. Vokabular wurde in zwei Tests gemessen, die ich auch in früheren Studien schon verwendet habe, und zur Messung von Intelligenz mussten die Testpersonen einen Teil eines bekannten IQ Tests absolvieren. Allgemeine Verarbeitungsgeschwindigkeit wurde in drei verschiedenen Tests beurteilt. In diesen Tests müssen die Testpersonen so schnell wie möglich eine Taste drücken, sobald sie ein bestimmtes Bild auf dem Bildschirm sehen oder einen Piep-Ton hören. Die Art des Kurzzeitgedächtnisses, an dem ich interessiert war, nennt man *phonologisches Kurzzeitgedächtnis*. Das beschreibt die Fähigkeit, eine Abfolge gesprochener Zahlen oder Buchstaben im Kurzzeitgedächtnis zu behalten. In der Aufgabe, die ich verwendet habe, um das *phonologische Kurzzeitgedächtnis* zu messen, hören die Testpersonen gesprochene Abfolgen von Zahlen, die zwischen zwei und neun Zahlen lang sind, und müssen diese in den Computer eintippen.

Die Wortlernstudie hat gezeigt, dass Testpersonen Wörter besser behalten, je öfter sie diese im Laufe des Trainings hören. Außerdem haben die Testpersonen in dem zweiten Test, der eine Woche nach dem Training stattgefunden hat, besser abgeschnitten, als im Test direkt nach dem Training. Das ist eine ganz typische Beobachtung: Schlaf unterstützt die Gedächtnisbildung neu erlernter Worte. Zusätzlich haben Testpersonen mit besserer allgemeiner Intelligenz und größerem Vokabular eine größere Anzahl der neuen Wörter gelernt.

Aber warum haben ein größeres Vokabular oder eine größere Intelligenz positive Auswirkungen auf das Erlernen von Wörtern? In Kapitel 6 habe ich die Computation Modelling Technik verwendet, um diese Frage zu beantworten. Ich habe ein Modell verwendet, das den kognitiven Prozess des Wortlernens simuliert; also ein Modell, das das Erlernen von Wortverstehen und Wortproduktion simuliert. Der Vorteil des Computational Modellings ist, dass es uns erlaubt, Faktoren oder Fähigkeiten, die in Testpersonen sehr schwierig zu kontrollieren und zu messen sind (wie etwa die Menge an Sprachinput, die ein Mensch bekommt, oder seine Verarbeitungsgeschwindigkeit), explizit zu manipulieren. In unserem Modell haben wir einige Faktoren manipuliert, die potentiell zu Unterschieden im Wortlernen führen können. So wurde das Modell beispielsweise dazu verwendet, um Unterschiede in der Intelligenz und der allgemeinen Verarbeitungsgeschwindigkeit zu simulieren. Die Studie hat gezeigt, dass schnellere Verarbeitungsgeschwindigkeit und höhere Intelligenz tatsächlich zu einem erfolgreicheren Wortlernen und damit zu einem größeren Vokabular führen. Interessanterweise scheint es, als ob die höhere Verarbeitungsgeschwindigkeit und Intelligenz deshalb vorteilhaft für den Wortlernprozess sind, weil sie dazu führen, dass nicht nur leicht zu erlernende Wörter gelernt werden, sondern auch schwieriger zu erlernende Wörter.

Zusätzlich wurde das Modell verwendet, um eine andere Beobachtung aus den Studien mit menschlichen Testpersonen weiter zu untersuchen; nämlich die Beobachtung, dass Personen mit einem größeren Vokabular schneller in der Sprachverarbeitung sind und im Erwachsenenalter erfolgreicher neue Wörter lernen. Das Modell hat gezeigt, dass es wahrscheinlich nicht Vokabulargröße per se ist, die zu besserer Sprachverarbeitungs- und Lernfähigkeit führt. Stattdessen scheint es so, als führten verschiedene Faktoren (mehr Sprach-Input, schnellere allgemeine Verarbeitungsgeschwindigkeit, höhere Intelligenz) zu verbesserter Sprachverarbeitung und erfolgreicherem Wortlernen im Erwachsenenalter. Wenn wir menschliche Testpersonen untersuchen, misst ein Vokabeltest nicht nur die Größe des Vokabulars, sondern auch diese zugrundeliegenden Faktoren. Das könnte ein

Grund dafür sein, warum Vokabeltestergebnisse in einem so starken Zusammenhang mit Unterschieden in der Sprachverarbeitung und dem Wortlernen zu stehen scheinen.

Zusammenfassung

Meine Forschung hat gezeigt, dass es große Unterschiede in der Vokabulargröße von Muttersprachlern einer Sprache gibt. Diese Unterschiede im Wortschatz beeinflussen die Sprachverarbeitung: Menschen mit einem größeren Wortwissen sind schneller und präziser in der Produktion und im Verständnis von Wörtern. Außerdem habe ich gezeigt, dass es notwenig ist, nicht nur Universitätsstudenten – die "typischen" Testpersonen in psycholinguistischen Studien – zu testen. Nicht alle Beobachtungen aus Studien mit dieser Gruppe von Testpersonen lassen sich auf Testpersonen mit vielfältigeren Bildungshintergründen verallgemeinern. Zudem unterschätzen wir deutlich, wie groß die Varianz in der Größe des Wortschatzes und der Sprachverarbeitungsgeschwindigkeit von Muttersprachlern ist, wenn wir uns auf die Untersuchung von Universitätsstudenten konzentrieren.

Im zweiten Teil meiner Doktorarbeit habe ich die Frage untersucht, was der Ursprung für diese Varianz in der Vokabulargröße von Muttersprachlern ist. Warum haben manche Menschen einen größeren Wortschatz als andere? Meine Studien haben gezeigt, dass Menschen mit einem größeren Vokabular besser darin sind, neue Worte zu erlernen. Das liegt jedoch wahrscheinlich an anderen zugrundeliegenden Faktoren, wie beispielsweise der Menge an sprachlichem Input, Intelligenz oder Verarbeitungsgeschwindigkeit. Das Computational Model hat gezeigt, dass mehr Input, höhere Intelligenz und erhöhte Verarbeitungsgeschwindigkeit einerseits dazu führen, dass das Modell während der frühen Entwicklung mehr Worte lernt, und andererseits auch schnellere Sprachverarbeitungsgeschwindigkeit und verbessertes Wortlernen im Erwachsenenalter zur Folge hat. Diese Doktorarbeit hat also geholfen, einige der komplexen Verhältnisse zwischen kognitiven Fähigkeiten, Wortlernen, Vokabular und Sprachverarbeitung zu beleuchten. Zukünftige Studien sind nötig, um weitere Erkenntnisse zu einer der faszinierendsten Fähigkeiten des Menschen zu sammeln, nämlich der Fähigkeit Sprache so scheinbar mühelos zu erlenen und zu verwenden.

English summary

The knowledge of words is an important part of our command of a language. Both educational and professional success are to a great extent dependent on whether we are able to understand written and spoken instructions and can clearly communicate our thoughts and opinions.

The number of words that we know is quite impressive. An average 20-year-old native speaker of American English is estimated to know 42,000 lemmas, i.e. uninflected word forms from which all inflections are derived (Brysbaert et al., 2016b). However, word learning does not stop at some point in early adulthood. On the contrary, vocabulary size tends to improve with ageing and across the entire lifespan (Brysbaert et al., 2016b; Schroeder & Salthouse, 2004). Thus, over the years, we are exposed to, continue to learn, and importantly also remember, a remarkable number of words. It has been estimated that an average adult learns approximately 6,000 new words between the ages of 20 and 60 years. That means in these 40 years, on average one new lemma is learned every two days (Brysbaert et al., 2016b; Keuleers et al., 2015).

Hence, our vocabularies are ever-growing inventories of words to describe the world around us, a collection of expressions to describe all kinds of experiences and information that is relevant to us. It might be assumed that there is variation in how many and which words people know. Specialist vocabulary is one example for the strong effects of experience on the type of words known by different individuals. The words *syncopation* or *lunge* are, for instance, known by individuals with certain occupations, interests, or hobbies, and completely opaque to others.[3] Another factor that affects the number and types of words we know is education. It has been shown that educational level significantly affects vocabulary size across the entire age range. Higher educational levels are associated with greater vocabulary knowledge. This is likely due to effects of extensive reading and studying – which are associated with formal education – on vocabulary growth (Brysbaert et al., 2016b; Keuleers et al., 2015).

[3]In this case, *syncopation* is a word known by musicians or perhaps also dancers and it refers to a change to the rhythm of a piece of music where the stressed beats become unstressed and the unstressed ones become stressed. *Lunge* is a dance movement where one leg is in a bent position and the other leg is extended.

Vocabulary knowledge

Previous research has indicated that variation in the number of words people know does not only affect which uncommon words they know and understand. Also the processing of very common words that are included in almost all adults' vocabularies has been shown to be influenced by how many words an adult knows. Individuals with better vocabulary knowledge have generally been reported to be better at processing language (Bent et al., 2016; Janse & Jesse, 2014; Rodriguez-Aranda & Jakobsen, 2011; Yap et al., 2012). That means they were usually faster and/or more accurate in performing different tasks assessing language comprehension and production. In the first part of my dissertation (Chapters 2, 3, & 4) I investigated this relationship between vocabulary and language processing more closely.

One important question to ask when studying vocabulary and language processing is: How is vocabulary knowledge best measured? Most of the earlier studies have used only single vocabulary tests. It has, however, been argued that in order to assess a skill as complex as word knowledge, it is necessary to use multiple tests, including for instance multiple-choice and open tests. Therefore, I did not only use one vocabulary measure but seven different tests, five of which I developed in preparation for my first experiments.

In Chapters 2 and 3, I examined the relationship between variation in vocabulary and language comprehension. The latter was assessed in a task typically used in psycholinguistic research, called *lexical decision task*. In this task, participants are presented with strings of letters on a computer screen and they are asked to decide whether or not what they see is an existing Dutch word, by pressing a button. Overall, the findings from this experiment are in line with what has been found previously: Individuals with better vocabulary knowledge performed better on the language comprehension task.

However, one important way in which my experiments extend earlier studies is that I did not only test university students (Chapter 2) but also young adults from more diverse backgrounds (Chapter 3). Most of what is known about language processing is based on studies testing undergraduate university students. One might ask whether what we know from studies on undergraduate university students can be generalised to the average young adult. The reason is that university students are, for example, presumably exposed to more written language than the average young adult, which in turn likely affects their language abilities.

My experiment in Chapter 3 demonstrated that the range of vocabulary scores and lexical decision performances (i.e. accuracy rates and speed of making the decisions) was much larger in the group of non-university students. Consequently, testing only university students means that we miss a lot of variation in skills that exists in humans.

In addition, the relationship between vocabulary size and language processing (larger vocabularies, better language processing) was similar but not exactly the same in the non-university as in the university students. Hence, the take-home message from this is that it is necessary to not just test the "typical" group of participants (university students) but strive for including participants from more diverse backgrounds in psychological and psycholinguistic studies.

In Chapter 4, I examined whether having a larger vocabulary is also beneficial for language production. Participants in this experiment again completed the seven vocabulary tests that were mentioned before. In addition, they did a task where they had to name pictures and we measured their naming speed. The results of this experiment show that individuals with better vocabulary scores were faster in naming the pictures. Hence, not only word comprehension but also word production is related to vocabulary and knowing more words is beneficial for both.

Vocabulary learning

The first three chapters have shown that there is considerable variation in vocabulary knowledge among the native speakers of a language, especially when looking at individuals from diverse backgrounds. Additionally, these differences have an effect on language processing: Larger vocabularies are associated with faster and more accurate word comprehension and production. A question arising is what the origins are of this variation in vocabulary size. Why do some people learn more words than others, thus ending up having a larger vocabulary than others? As indicated above, both age and education have been related to differences in vocabulary (Brysbaert et al., 2016b; Keuleers et al., 2015). In addition, there is a growing body of research on word learning in children. Differences in the amount of exposure to language and in language processing speed have been shown to lead to variation in word learning and vocabulary size in childhood (Fernald et al., 2006; Jones & Rowland, 2017; Marchman & Fernald, 2008). However, despite the fact that we know that word learning continues across the life span, very little research has investigated which factors affect word learning in adulthood. This question was addressed in Chapters 5 and 6 of my dissertation.

The experiment in Chapter 5 was a word learning experiment. Participants came to the lab and learned 39 words, each of which was paired with a coloured photograph depicting the word's meaning. All words were very uncommon Dutch words and therefore unknown to 90% of the native speakers of Dutch in the Netherlands, thus, also to the participants in the experiment. All participants were trained on these words and tested immediately after training and one week later. During the training tasks, some words were presented three times, others nine times, and others 16 times. The purpose of the experiment was in particular to examine the relationships between word learning success, sleep, the amount of exposure to novel words, and different cognitive abilities. The cognitive abilities we looked at were vocabulary, general intelligence, processing speed, and short-term memory. Vocabulary was assessed in two of the tests which I had used before, and general intelligence was measured with a sub-part of a well-known IQ test. Processing speed was measured in three tasks, where participants had to press a button as quickly as possible upon seeing a picture on the screen or hearing a beep-sound over headphones. Finally, the type of short-term memory I was interested in is called *phonological short-term memory*and it describes the ability to keep strings of spoken digits or letters in short-term memory. In the task to measure this, participants hear spoken strings of digits, ranging from two-digit to nine-digit strings, and are asked to type them in.

The results of the word learning experiment showed that participants remembered words better that they had seen more often during training. In addition, their performance was better in the second test, one week after the training had taken place, than right after training. This is a typical finding: Sleep supports the formation of memories of newly learned words. Furthermore, participants with better nonverbal intelligence and better vocabularies in their native language learned more words.

But why is a larger vocabulary or greater intelligence beneficial for word learning? In Chapter 6, this issue was addressed using a computational model that simulates the cognitive process of learning to produce and comprehend words. The advantage of computational modelling is that constructs that are very difficult to measure and control in humans (such as the amount of language exposure or processing speed) can be manipulated explicitly. In our model, a number of factors that potentially lead to differences in word learning were manipulated. The model was, for example, used to simulate variation in general intelligence and processing speed. The study showed that faster processing speed and greater general intelligence do indeed lead to larger vocabularies. Importantly, it seemed that the networks with increased processing speed and those with higher intelligence had advantages in word learning because they learned not only the easy words but also more difficult words.

In addition, the model was used to further examine the finding that individuals with larger vocabularies are faster in processing language and better at learning novel words in adulthood. The model demonstrated that it might not be a larger vocabulary per se, but rather underlying factors (such as more exposure, faster processing speed, or greater intelligence) that lead to faster language processing and better word learning in adulthood. When testing human participants, a measure of vocabulary probably captures not only vocabulary size but also these other underlying factors. That might be the reason why vocabulary has been shown to strongly relate to variation in language processing and also variation in word learning.

Summary

To sum up, my research demonstrated that there is a lot of variation in vocabulary size among the native speakers of a language. These differences in word knowledge influence language processing: Individuals with better vocabularies are faster and more accurate at producing and comprehending words. In addition, I showed that it is necessary to not only test undergraduate university students – the typical participants in psycholinguistic research. Not all findings from studies on university students extend to groups of participants from more diverse backgrounds. The range of skills, in my case vocabulary scores, and the variation in language processing performance, are underestimated when focusing on university students only.

Finally, I asked what the origin of variation in vocabulary size is. Why do some people learn more words than others and therefore end up having larger vocabularies than others? My studies showed that individuals with larger vocabularies are more successful in learning novel words. However, this is likely due to underlying factors, such as exposure, intelligence, or processing speed. The computational model demonstrated that increased exposure, greater intelligence, and higher processing speed all lead to larger numbers of words being learned on the one hand, and to faster language processing speed and better novel word learning on the other hand. Hence, this dissertation shed light on some of the complex relationships between different cognitive abilities, vocabulary learning and size, as well as language processing. Future research is needed to further examine these aspects of one of humans' most fascinating ability, namely the ability to learn and use language.

Acknowledgements

When I came to Nijmegen to do my PhD at the MPI, I had no idea what lay ahead of me. I was thrilled to get the chance to work at this outstanding institute, excited to start a new chapter in my life, looking forward to being challenged and to learning a lot, and I was a little bit scared. The fear disappeared very quickly, excitement and happiness remained – of course interspersed with moments of frustration. It would not have been the same without the people who accompanied me on this journey over the last 3.5 years.

First and foremost, I would like to thank my supervisors Antje, Zeshu, and Alastair. Antje, I learned a lot from you and became a better writer and better researcher through your feedback and guidance. You taught me how to clearly formulate questions and work towards answering them in an efficient way. I greatly appreciate that you were always fast in providing feedback and responding to emails, and that your door was always open for me to walk in. Thank you for all the long meetings and fruitful discussions. Importantly, your door was not only open to discuss planned experiments, results, or conference submissions, but also when I was having a hard time. Thank you for your understanding when I felt a little lost and frustrated, and for caring not only for the output I produced as part of my PhD but also for my emotional well-being.

Thank you also to you, Zeshu. For the first half of my PhD, your office just across the corridor was always open for me to walk in, whether I had a question about my stimuli or was in doubt about the random effects structure of my mixed-effects model. I learned a lot from you about designing neat experiments and analysing data. Thanks for being a daily supervisor with an open ear for any issue and for helping me have a very smooth start to my PhD.

Alastair, you came on board a little later in my PhD journey. I learned a lot from your structured and thorough way of preparing experiments and modelling investigations, and your attention for all the important details. You challenged me in many stimulation discussions, which I greatly appreciated. In endless meetings you taught me about computational modelling, scribbling graphs and formulas on the white board, walking me through code, and with enormous patience answering all the many questions I had. At least as important as this was your open ear, understanding, and patience in the more difficult months. You believed in the project at times when I was losing confidence and struggled. You helped me to structure my work when I did not see the woods for trees, and motivated me to keep on working towards my goal. I am incredibly grateful for this support, which made the hard times a little easier and helped me gain back my confidence.

I would also like to thank the members of the reading committee, Prof. Ardi Roelofs, Prof. James McQueen, and Prof. Marcus Damian, for devoting their time and efforts to reading and evaluating this dissertation.

My dear paranymphs, thank you very much for being at my side. Sara, the first time we met was on the day of your job interview. That is how it all started and in the end you became one of my best friends and luckily even my office mate. I loved the long talks with you about really anything; not only work, but also our shared passion for music, or yoga. I learned a lot from talking to you. Many problems turned out to be smaller and less scary after having shared them with you, many days got lighter and sunnier by laughing together. Thank you for being the way you are, a beautiful soul and true friend.

Dear Greta, meine Seelenfreundin, our story began on the bus ride to the March for Science in Amsterdam. That day we talked a lot and I think we both went home in the evening with the feeling of having met a very special person and friend for life. What we share is having a deep, soul- and heart-filling passion, which can really not be described with words. Thank you for all the inspiring talks about music, singing, and dancing, for taking me to one of your singing lessons, and for letting me show you how to improve your *tendus* or *pliés*. Thanks for listening to me and making me feel understood, whether it is about the arts or work or anything else.

I really enjoyed my time as part of the Psychology of Language department. Thank you to all members of the department – past and present – for their scientific input and the fun chats during lunch breaks: Agnieszka, Alastair, Amie, Andrea, Annelies, Caitlin, Conny, Eirini, Evelyn, Falk, Fan, Federica, Florian, Greta, Hans Rutger, Jeroen, Joe, Johanne, Laurel, Limor, Markus, Merel M., Merel W., Miguel, Phillip, Renske, Sara, Saoradh, Shiri, Suzanne, Will, Xuefei, and Zeshu. Many of you were not only colleagues but friends with whom I enjoyed many dinners and drinks. Thanks to every one of you.

In particular, I would like to thank the office mates I had over the years: Florian, Suzanne, Limor, and Sara. Limor, thank you for being at my side as my long-term office mate and friend. Two people sharing an office could hardly be more different in personality than we are but we had a wonderful time together. Thank you for all the fun, inspiring, eye-opening, comforting conversations we had. Eirini, thanks for many lovely chats in the morning and over lunch. I will continue sending you photos of unicorns and goats. Thanks also to my friends Caitlin, Charlotte, Emma, and Sophie for lunch breaks and dinners and many laughs.

Thank you also to the operations team, TG, and library for supporting us in any way they can to ensure that business trips can happen, computers are running, experiment labs are working, articles can be published, and books are available. I would also like to thank the IMPRS coordinators Kevin, Dirkje, and Els, for giving us the opportunity to participate in so many enriching activities and courses. Thank you for your dedication and enthusiasm.

I had a wonderful time in Nijmegen, which is due to all the lovely friends from the MPI and beyond, but also due to many hours spent dancing in stimulating, inspiring, and relaxing dance classes by Matthias Löffler and Percy Kruythoff. I got the chance to grow, not only as a researcher but also as a dancer. I am grateful for all the opportunities I got, for being able to take classes in Amsterdam, for being on stage, and for meeting a lot of passionate and talented dancers.

Thank you also to the many lovely people from Aachen (especially Julia, Judith, Caro, Conny, Vera, Suse, and Sirin), for being at my side for many years already. I am very happy to call you my friends.

Finally, I want to thank my family for all their love, support, and encouragement. Lisa, you are not only my sister but my soulmate and best friend. You have been at my side for all of my life, always listening to me, believing in me, laughing with me. Thank you for being there, whenever I need you, and a big thank you also for drawing the picture for my dissertation cover. Thank you also to my brother-in-law, Martin, and my two beautiful little nephews, Ben and Max, for all the laughs and good times together. Mum and dad, many thanks to you, for being the most amazing parents and for your unconditional trust, love, and support. You allowed me to follow my dreams, listened to my doubts and worries, encouraged me to believe in myself, and celebrated my achievements with me. I cannot thank you enough for being the wonderful humans you are.

Last but not least, I would like to thank my partner Stefan. Thank you for listening to me when I was excited and enthusiastic or frustrated and upset, for believing in me and giving me confidence, for making me laugh, for all the fun facts, for dreaming and travelling and exploring together, and for the most wonderful weekends in Frankfurt, Nijmegen, Aachen, and Düsseldorf.

Curriculum Vitae

Nina Mainz was born in 1991 in Aachen, Germany. She obtained her bachelor's degree in Linguistics and Communication Studies, and English Studies from Rheinisch-Westfälische Technische Hochschule (RWTH) Aachen in 2013. This was followed by a Master's degree in Linguistics at Queen Mary University of London, which she completed in 2014. In January 2015, Nina began her PhD project in the Psychology of Language Department at the Max Planck Institute for Psycholinguistics.

Publications

Mainz, N., Smith, A.C., & Meyer, A.S. (submitted). An exploratory study of individual differences in adult novel word learning.

Mainz, N., Shao, Z., Brysbaert, M., & Meyer, A.S. (2017). Vocabulary knowledge predicts lexical processing: Evidence from a group of participants with diverse educational backgrounds. *Frontiers in Psychology, 8*, 1164. doi: 10.3389/fpsyg.2017.01164.

MPI Series in Psycholinguistics

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing. *Miranda I. van Turennout*
2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography. *Niels O. Schiller*
3. Lexical access in the production of ellipsis and pronouns. *Bernadette M. Schmitt*
4. The open-/closed class distinction in spoken-word recognition. *Alette Petra Haveman*
5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach. *Kay Behnke*
6. Gesture and speech production. *Jan-Peter de Ruiter*
7. Comparative intonational phonology: English and German. *Esther Grabe*
8. Finiteness in adult and child German. *Ingeborg Lasser*
9. Language input for word discovery. *Joost van de Weijer*
10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe. *James Essegbey*
11. Producing past and plural inflections. *Dirk J.Janssen*
12. Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea. *Anna Margetts*
13. From speech to words. *Arie H. van der Lugt*
14. Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language. *Eva Schultze-Berndt*
15. Interpreting indefinites: An experimental study of children's language comprehension. *Irene Krämer*
16. Language-specific listening: The case of phonetic sequences. *Andrea Christine Weber*
17. Moving eyes and naming objects. *Femke Frederike van der Meulen*

18. Analogy in morphology: The selection of linking elements in Dutch compounds. *Andrea Krott*

19. Morphology in speech comprehension. *Kerstin Mauth*

20. Morphological families in the mental lexicon. *Nivja Helena de Jong*

21. Fixed expressions and the production of idioms. *Simone Annegret Sprenger*

22. The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria. *Birgit Hellwig*

23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies. *Fermín Moscoso del Prado Martín*

24. Contextual influences on spoken-word processing: An electrophysiological approach. *Danielle van den Brink*

25. Perceptual relevance of prevoicing in Dutch. *Petra Martine van Alphen*

26. Syllables in speech production: Effects of syllable preparation and syllable frequency. *Joana Cholin*

27. Producing complex spoken numerals for time and space. *Marjolein Henriëtte Wilhelmina Meeuwissen*

28. Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction. *Rachèl Jenny Judith Karin Kemps*

29. At the same time...: The expression of simultaneity in learner varieties. *Barbara Schmiedtová*

30. A grammar of Jalonke argument structure. *Friederike Lüpke*

31. Agrammatic comprehension: An electrophysiological approach. *Marijtje Elizabeth Debora Wassenaar*

32. The structure and use of shape-based noun classes in Miraña (North West Amazon). *Frank Seifart*

33. Prosodically-conditioned detail in the recognition of spoken words. *Anne Pier Salverda*

34. Phonetic and lexical processing in a second language. *Mirjam Elisabeth Broersma*

35. Retrieving semantic and syntactic word properties: ERP studies on the time course in language comprehension. *Oliver Müller*

36. Lexically-guided perceptual learning in speech processing. *Frank Eisner*

37. Sensitivity to detailed acoustic information in word recognition. *Keren Batya Shatzman*

38. The relationship between spoken word production and comprehension. *Rebecca Özdemir*
39. Disfluency: Interrupting speech and gesture. *Mandana Seyfeddinipur*
40. The acquisition of phonological structure: Distinguishing contrastive from non-constrative variation. *Christiane Dietrich*
41. Cognitive cladistics and the relativity of spatial cognition. *Daniel Haun*
42. The acquisition of auditory categories. *Martijn Bastiaan Goudbeek*
43. Affix reduction in spoken Dutch: Probabilistic effects in production and perception. *Mark Plumaekers*
44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence. *Valesca Madalla Kooijman*
45. Space and iconicity in German sign language (DGS). *Pamela M. Perniss*
46. On the production of morphologically complex words with special attention to effects of frequency. *Heidrun Bien*
47. Crosslinguistic influence in first and second languages: Convergence in speech and gesture. *Amanda Brown*
48. The acquisition of verb compounding in Mandarin Chinese. *Jidong Chen*
49. Phoneme inventories and patterns of speech sound perception. *Anita Eva Wagner*
50. Lexical processing of morphologically complex words: An information-theoretical perspective. *Victor Kuperman*
51. A grammar of Savosavo: A Papuan language of the Solomon Islands. *Claudia Ursula Wegener*
52. Prosodic structure in speech production and perception. *Claudia Kuzla*
53. The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension. *Sarah Schimke*
54. Studies on intonation and information structure in child and adult German. *Laura de Ruiter*
55. Processing the fine temporal structure of spoken words. *Eva Reinisch*
56. Semantics and (ir)regular inflection in morphological processing. *Wieke Tabak*
57. Processing strongly reduced forms in casual speech. *Susanne Brouwer*
58. Ambiguous pronoun resolution in L1 and L2 German and Dutch. *Miriam Ellert*

59. Lexical interactions in non-native speech comprehension: Evidence from electroencephalography, eye-tracking, and functional magnetic resonance imaging. *Ian FitzPatrick*

60. Processing casual speech in native and non-native language. *Annelie Tuinman*

61. Split intransitivity in Rotokas, a Papuan language of Bougainville. *Stuart Payton Robinson*

62. Evidentiality and intersubjectivity in Yurakaré: An interactional account. *Sonja Gipper*

63. The influence of information structure on language comprehension: A neurocognitive perspective. *Lin Wang*

64. The meaning and use of ideophones in Siwu. *Mark Dingemanse*

65. The role of acoustic detail and context in the comprehension of reduced pronunciation variants. *Marco van de Ven*

66. Speech reduction in spontaneous French and Spanish. *Francisco Torreira*

67. The relevance of early word recognition: Insights from the infant brain. *Caroline Mary Magteld Junge*

68. Adjusting to different speakers: Extrinsic normalization in vowel perception. *Matthias Johannes Sjerps*

69. Structuring language: Contributions to the neurocognition of syntax. *Katrien Rachel Segaert*

70. Infants' appreciation of others' mental states in prelinguistic communication: A second person approach to mindreading. *Birgit Knudsen*

71. Gaze behavior in face-to-face interaction. *Federico Rossano*

72. Sign-spatiality in Kata Kolok: How a village sign language of Bali inscribes its signing space. *Connie de Vos*

73. Who is talking? Behavioural and neural evidence for norm-based coding in voice identity learning. *Attila Andics*

74. Lexical processing of foreign-accented speech: Rapid and flexible adaptation. *Marijt Witteman*

75. The use of deictic versus representational gestures in infancy. *Daniel Puccini*

76. Territories of knowledge in Japanese conversation. *Kaoru Hayano*

77. Family and neighbourhood relations in the mental lexicon: A cross-language perspective. *Kimberley Mulder*

78. Contributions of executive control to individual differences in word production. *Zeshu Shao*

79. Hearing speech and seeing speech: Perceptual adjustments in auditory-visual processing. *Patrick van der Zande*

80. High pitches and thick voices: The role of language in space-pitch associations. *Sarah Dolscheid*

81. Seeing what's next: Processing and anticipating language referring to objects. *Joost Rommers*

82. Mental representation and processing of reduced words in casual speech. *Iris Hanique*

83. The many ways listeners adapt to reductions in casual speech. *Katja Pöllmann*

84. Contrasting opposite polarity in Germanic and Romance languages: Verum Focus and affirmative particles in native speakers and advanced L2 learners. *Giuseppina Turco*

85. Morphological processing in younger and older people: Evidence for flexible dual-route access. *Jana Reifegerste*

86. Semantic and syntactic constraints on the production of subject-verb agreement. *Alma Veenstra*

87. The acquisition of morphophonological alternations across languages. *Helen Buckler*

88. The evolutionary dynamics of motion event encoding. *Annemarie Verkerk*

89. Rediscovering a forgotten language. *Jiyoun Choi*

90. The road to native listening: Language-general perception, language-specific input. *Sho Tsuji*

91. Infants'understanding of communication as participants and observers. *Gudmundur Bjarki Thorgrímsson*

92. Information structure in Avatime. *Saskia van Putten*

93. Switch reference in Whitesands. *Jeremy Hammond*

94. Machine learning for gesture recognition from videos. *Binyam Gebrekidan Gebre*

95. Acquisition of spatial language by signing and speaking children: A comparison of Turkish sign language (TID) and Turkish. *Beyza Sumer*

96. An ear for pitch: On the effects of experience and aptitude in processing pitch in language and music. *Salomi Savvatia Asaridou*

97. lncrementality and Flexibility in Sentence Production. *Maartje van de Velde*

98. Social learning dynamics in chimpanzees: Reflections on (nonhuman) animal culture. *Edwin van Leeuwen*

99. The request system in Italian interaction. *Giovanni Rossi*

100. Timing turns in conversation: A temporal preparation account. *Lilla Magyari*

101. Assessing birth language memory in young adoptees. *Wencui Zhou*

102. A social and neurobiological approach to pointing in speech and gesture. *David Peeters*

103. Investigating the genetic basis of reading and language skills. *Alessandro Gialluisi*

104. Conversation electrified: The electrophysiology of spoken speech act recognition. *Rósa Signý Gisladottir*

105. Modelling multimodal language processing. *Alastair Charles Smith*

106. Predicting language in different contexts: The nature and limits of mechanisms in anticipatory language processing. *Florian Hintz*

107. Situational variation in non-native communication. *Huib Kouwenhoven*

108. Sustained attention in language production. *Suzanne Rosa Jongman*

109. Acoustic reduction in spoken-word processing: Distributional, syntactic, morphosyntactic, and orthographic effects. *Malte Viebahn*

110. Nativeness, dominance, and the flexibility of listening to spoken language. *Laurence Bruggeman*

111. Semantic specificity of perception verbs in Maniq. *Ewelina Wnuk*

112. On the identification of FOXP2 gene enhancers and their role in brain development. *Martin Becker*

113. Events in language and thought: The case of serial verb constructions in Avatime. *Rebecca Defina*

114. Deciphering common and rare genetic effects on reading ability. *Amaia Carrión Castillo*

115. Music and language comprehension in the brain. *Richard Kunert*

116. Comprehending Comprehension: Insights from neuronal oscillations on the neuronal basis of language. *Nietzsche H.L. Lam*

117. The biology of variation in anatomical brain asymmetries. *Tulio Guadalupe*

118. Language processing in a conversation context. *Lotte Schoot*

119. Achieving mutual understanding in Argentine Sign Language. *Elizabeth Manrique*

120. Talking Sense: the behavioural and neural correlates of sound symbolism. *Gwilym Lockwood*

121. Getting under your skin: The role of perspective and simulation of experience in narrative comprehension. *Franziska Hartung*

122. Sensorimotor Experience in Speech Perception. *William Lawrence Schuerman*

123. Explorations of beta-band neural oscillations during language comprehension: Sentence processing and beyond. *Ashley Lewis*

124. Influences on the magnitude of syntactic priming. *Evelien Heyselaar*

125. Lapse organization in interaction. *Elliott Hoey*

126. The processing of reduced word pronunciation variants by natives and foreign language learners: Evidence from French casual speech. *Sophie Brand*

127. The Neighbors Will Tell You What To Expect: Effects of Aging and Predictability on Language Processing. *Cornelia Moers*

128. The role of voice and word order in incremental sentence processing. Studies on sentence production and comprehension in Tagalog and German. *Sebastian Sauppe*

129. Learning from the (Un)Expected: Age and Individual Differences in Statistical Learning and Perceptual Learning in Speech. *Thordis Neger*

130. Mental representations of Dutch regular morphologically complex neologisms. *Laura de Vaan*

131. Speech production, perception, and input of simultaneous bilingual preschoolers: Evidence from voice onset time. *Antje Stoehr*

132. A holistic approach to understanding pre-history. *Vishnupriya Kolipakam*

133. Characterization of transcription factors in monogenic disorders of speech and language. *Sara Busquets Estruch*

134. Indirect request comprehension in different contexts. *Johanne Tromp*

135. Envisioning Language – An Exploration of Perceptual Processes in Language Comprehension. *Markus Ostarek*

136. Listening for the WHAT and the HOW: Older adults' processing of semantic and affective information in speech. *Juliane Kirsch*

137. Let the agents do the talking: On the influence of vocal tract anatomy on speech during ontogeny and glossogeny. *Rick Janssen*

138. Age and hearing loss effects on speech processing. *Xaver Koch*