The influence of individual differences on reading in readers of consistent compared to inconsistent orthographies

Julia M H Birchenough

Oxford Brookes University

This thesis is submitted in partial fulfilment of the
requirements of the award of Doctor of Philosophy

September 2019

Supervisors:
Prof Vincent Connelly
Dr Robert A I Davies

Acknowledgements

This research endeavour has been a long journey, and without the help of many others whom I met along the way this thesis would not have been possible. I would therefore like to take the opportunity to thank everybody for your interest and support! There are some I would like to mention in particular. Above all, I would like to thank my supervisors Dr Robert Davies and Prof Vincent Connelly for their support and patience throughout the duration of my research, and the invaluable guidance and advice that they have shared with me. Special mention also goes to Wakefield Morys-Carter, who helped me enormously with Qualtrics and with whom I was able to discuss different approaches to data analysis. I am also very grateful to Prof Dr Guido Nottbusch at the University of Potsdam, who kindly hosted me during data collection for the naming study in Germany. During my time in Germany, members of Prof Nottbusch's department also greatly facilitated my research activity. In particular, I would like to thank Dennis Pauly for setting up a website for participant sign-up, and to Frau Giesch, who made it very easy for me to find my way around the University. Furthermore, a very, very big thank-you goes to Frau Petra Schienmann of the Potsdam University EyeLab for all her help with participant recruitment. Moreover, I would like to thank the many online platforms, who have kindly let me advertise my AoA study. I would also like to express my deepest gratitude to the R community in general, who make it possible for new users to learn about R on websites such as Stack Overflow, and so generously give their time and expertise to answering questions online. Thanks a lot also to my fellow research students for sharing tips and advice, and for making it a pleasure to be in the PhD room. Specifically, I would like to mention Sarah Camp, Sarah Mansbridge, Annabel Molyneaux, Tracy Mcateer and Amanda Hyne. Nothing however, would have been possible without the participants who took part in this study – I am really very grateful to you all, and thank you for your time and interest. Last but not least, I thank my husband and children for their unwavering belief in me and continued encouragement to keep going. Without you, I could not have travelled this far.

Abstract

Languages differ in terms of how consistently they reflect spelling-sound relationships, and research has found that this may lead to differences in naming (reading aloud) processes. Readers themselves differ from each other in terms of relevant task performance, such as nonword decoding ability, vocabulary knowledge, spelling ability and reading experience. Such tasks tap into individual differences which have also been shown to influence the reading aloud process. The present study investigated whether language-related differences in reading aloud persisted even when reading-related individual differences were taken into account, and how effects of individual differences may vary between languages. The comparison necessitated a number of preparatory tasks to facilitate cross-language comparison. This included the computation of spelling-sound consistencies for both languages, the collection of German age-of-acquisition ratings and the creation of comparable measures to capture reading experience in both languages. For the naming study, reading aloud reaction times (RTs) on a set of 85 cognates were compared between skilled readers of English and German. Readers also completed tasks to assess individual differences. Linear mixed-effects modelling analysis showed that language differences remained, but that individual differences contributed additionally to explaining reading performance. To further examine how individual differences may impact differently on naming RTs between languages, the same data set was split four times into those who had scored higher and lower in each of the four individual differences (ID) tasks. Each ID group was then analysed separately. This resulted in eight different analyses. The language effect remained significant for all ID groups. Variations in effect patterns between different ID groups were observed. Effect patterns were more similar between languages for those readers who had scored higher in the ID tasks. Strong nonword decoders emerged as the fastest reader group for both languages, indicating that nonword decoding indexes a vital processing mechanism for skilled readers of different languages. As no significant interactions were found involving language or language and IDs for this group, strong decoders seemed to be most similar in their naming across the two languages. Although semantics were used by readers of both languages, person-level semantic knowledge was more beneficial for readers of the opaque script, especially when decoding skills were weaker. Good spelling ability facilitated naming in both languages, but differences between languages became apparent in weaker spellers, as those reading English were more influenced by other IDs, such as decoding skill. Unexpectedly, print exposure was not the strongest modulator out of all individual differences. Together the results suggest that alongside language differences, individual differences are important factors to be considered to account for a universal process of reading aloud.

Table of contents

List of Tables

List of Figures

List of Abbreviations

AoA	age-of-acquisition
DRC model	Dual Route Cascaded Model
GPCs	grapheme-phoneme correspondences
IDs	individual differences
IMG	imageability
LQH	Lexical Quality Hypothesis
ODH	Orthographic Depth Hypothesis
O-P	orthography-to-phonology
O-P-S	orthography-to-phonology-to-semantics
O-S	orthography-to-semantics
O-S-P	orthography-to-semantics-to-phonology
SR	semantic reliance
PDP model	Parallel Distributed Processing Model
PGST	Psycholinguistic Grain Size Theory

1 Introduction

The consistency with which sound is reflected in written language has a measurable impact on the process of reading aloud (Frost, Katz, & Bentin, 1987; Katz & Frost, 1992; Ziegler & Goswami, 2005). Specifically, developing readers of languages with unambiguous letter-sound relationships and few exceptions learn to read faster and make fewer errors than readers of languages such as English, which offer many different spellings for the same sound, and many different sounds for the same spelling (Seymour, Aro, & Erskine, 2003). Discernible differences between readers of different orthographies also remain in skilled readers (Rau, Moll, Snowling, & Landerl, 2015; Ziegler, Perry, Jacobs, & Braun, 2001).

However, individuals' reading processes also vary in other regards – with tangible and lasting behavioural differences in reading aloud. For example, people vary considerably in their ability to read unknown items, or nonwords. Nonword reading deficiency has been recognised as an acquired impairment in patients (Funnell, 1983), and as a developmental difficulty in children (A. W. Ellis, 2016). Notably, variation in nonword reading has also been reported for skilled readers (e.g., Baron & Strawson, 1976). Second, readers vary in their semantic knowledge. As word meaning assists reading aloud performance (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996), this may also influence the reading aloud process. Third, individuals differ in spelling ability. Some evidence suggests that better spelling is related to better word reading (e.g., Martin-Chang, Ouellette, & Madden, 2014), and thus seems to reflect another aspect of how people's reading systems can vary. Fourth, the functioning of the reading system improves with the training it receives, and consequently reading experience shapes the reading system measurably (e.g., Stanovich & West, 1989).

Simulations of the two most prominent reading models, the dual route model (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) and the connectionist parallel distributed model (Plaut et al., 1996; Seidenberg & McClelland, 1989), have been able to simulate many of the above mentioned individual differences in reading performance. However, models have also been criticised that their constructs of how reading works do not adequately reflect universal lexical processing. Specifically, reading research has been said to be too focused on the English language, which seems to be a special case due to its richness in spelling-sound mappings. The disproportionate amount of reading research in English may have influenced modelling in such a way to accommodate the ‘outlier’ language English, rather than describing a universal reading process independent of language specifics. (Share, 2008) More recently, there has been the call for models to focus on integrating all differences between readers to allow for a more comprehensive and universal understanding of the reading process. (Rueckl, 2016) This integrated approach, which would encompass differences in experience (language, teaching methods, reading acquisition and reading practice), as well as variation within the reading system itself, could then also shed light on the interdependencies and interrelations of these different components.

Notwithstanding the feasibility of an implementation of a more comprehensive and unifying reading model, any model simulation will need to be compared to behavioural data to evaluate its performance. The current study aims to make a substantial and novel contribution in this regard by comparing naming times across two different languages whilst taking account of a number of pertinent individual differences. Accordingly, naming latencies for a set of cognates were collected for two languages differing in spelling-sound consistency, English and German. Participants were additionally asked to complete a set of tasks in order to assess their ability to read nonwords (decoding), vocabulary (semantic) knowledge, spelling (orthographic) knowledge, and amount of reading experience. The first aim of the present study was to investigate, whether reading differed between speakers of languages with contrasting spelling-sound systems when individual differences (other than language) were taken into account. If individual differences affected reading aloud differently across languages, the second aim was to establish how reading performance varied.

1.1 Structure

The next chapter will introduce the two main reading models for reading aloud, the dual route model, and the connectionist PDP model, thereby laying down the theoretical frameworks for understanding reading aloud and the influence of individual differences.

The third chapter introduces two theories explaining reading differences due to the varied consistency of spelling-sound relationships, the Orthographic Depth Hypothesis (ODH) and the Psycholinguistic Grain Size Theory (PGST). Proponents of the ODH envisage a different distribution of reading mechanisms for readers of different languages. Specifically, languages with straightforward and consistent spelling-sound relationships are read with greater reliance on serial decoding, whereas languages with less consistent spelling-sound relationships rely more on lexical information. The PGST assumes that differences result from the need for more varied reading unit sizes in less transparent languages. The effects of reading unit sizes have been explored using the word length and neighbourhood effects of larger reading units. This chapter will review findings of these effects so far for two languages, which differ in the way that they reflect sound in spellings. The review will show that behavioural studies have reported mixed findings.

The fourth chapter presents how reading may differ in terms of the readers' reading system. Readers differ in their ability to read nonwords (decoding skill). Within the connectionist reading model, this difficulty lies within the efficiency to map orthography to phonology, and may lead to increased use of the semantic pathway in reading. People may also differ in their breadth and depth of semantic knowledge, or how easily this is accessible. According to the Lexical Quality Hypothesis, readers also differ in the quality of the orthographic components of their lexical representation. These individual differences have been captured with vocabulary and spelling tests, respectively. Finally, readers also differ in the amount of time they spend reading, and it has been shown to improve the reading process.

The present study will bring together readers from two different languages, and compare their reading whilst taking into consideration that they also differ in terms of decoding skill, semantic knowledge, spelling performance and reading experience. In order to meaningfully evaluate reading differences, stimuli had to be qualified in terms of spelling-sound consistency. This measure had to be computed prior to stimuli selection and is described in Chapter 5.

Chapter 6 describes the creation and analysis of print exposure measures, which could be employed for both language groups.

Chapter 7 reports the method and results from the reading aloud analyses. The main aim will be to see if language differences will remain, when allowing for participant variable variation other than language.

In Chapter 8, the focus will be to unpack the influence of individual differences further by dividing the data into sets which reflect weaknesses and strengths in individual difference tasks. This means, that the same data set will be split into higher and lower achievers of each individual difference, in order to better understand how each individual difference varies within and between languages.

The general discussion in Chapter 9 will present the main contributions of the present investigation, and how the results could be understood within the current understanding of reading aloud.

2 Reading aloud models

The present investigation aims to bring together different lines of reading research, namely theoretical considerations on reading in different languages, and considerations of other factors pertaining to individuals which impact on the reading aloud processes. The inclusion of all these factors in one behavioural investigation will hopefully be a useful contribution to understanding the interplay of these factors. As different models have contributed to our understanding of reading aloud, this chapter will briefly outline the two main existing reading models.

2.1 What is reading aloud?

Reading aloud refers to the process of translating a written into an oral code, or in other words, to process a string of letters and produce its respective pronunciation. It is also often referred to as the *naming* or *pronunciation* or *oral reading task*. Typically, in reading research, the naming task involves showing a test item (word or nonword) on a screen and recording participants' reactions from the moment of stimulus presentation to the start of the word pronunciation. The length of the resulting naming latency varies with different kinds of stimulus properties, such as item frequency, length in letters and the relationship between orthography and phonology, and is taken to reflect the underlying cognitive processes which lead to stimulus pronunciation. Two main theoretical accounts of the structure of the reading system have been proposed to explain these word property effects in the oral reading task.

2.2 Dual route model

The dual route approach (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart et al., 2001) proposes a dual structure which divides reading into a lexical (*word* or *direct*) and a sublexical (*assembled*) route. The lexical orthographic part of the model is based on the interactive activation model (Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981). The dual route structure is borne out of observations in neuropsychological studies of reading following brain injury, that either the capacity to read words with irregular spelling-sound correspondences, so-called *exception words*, or the capacity to read words with regular spelling-sound correspondences and made-up words can be independently impaired whilst the other remains preserved (Bub & Kertesz, 1985; Funnell, 1983; Marshall & Newcombe, 1973). In the dual route account then, words have an entry in the orthographic lexicon and can therefore generally be read by the lexical route, whereas nonwords (made-up words which have no lexical entry) can only ever be read via the sublexical route.

An outline of the implemented dual route model is shown in Figure 2.1. When reading through the lexical route, features activate letters in parallel, which in turn access a word lexicon with whole orthographic word representations. The activation of orthographic representations then leads to the activation of phonological lexical representations. Activation is cascaded, which means that as soon as there is activation, this is transmitted to the next level. All activation can be excitatory, i.e. activate the next level. However, connections in the lexical route between letters and orthographic representations, and between phonological representations and phonemic output, can also be inhibitory. For example, a letter activates all word units with the same letter in the same position, and inhibits all other word units. Finally, activation rises faster with frequent words leading to faster pronunciation for more frequent words.

The sublexical route shares the feature-to-letter activation system with the lexical route. It then successively converts graphemes into phonemes and thereby assembles words according to grapheme-phoneme correspondence rules (GPC rules; Rastle & Coltheart, 1999). This occurs in a serial, from left to right manner. Within the dual route framework, longer naming latencies with increasing letter length can be rooted in the serial grapheme – phoneme matching mechanism of the sublexical route.

Figure 2.1 has been removed from
this version of the thesis due to
copyright restrictions

Figure 2.1. Dual-route cascaded model of reading aloud in Coltheart, Rastle, Perry, Langdon and Ziegler (2001)

Naming occurs when all phonemes corresponding to the letter string have been activated. Importantly, both routes are activated each time that an item is being read, and both routes contribute to reading aloud. Depending on the item, the one or the other route will contribute more to the final output. Thus, high frequency words are read faster via the lexical route, whilst nonwords, which do not have a lexical entry, are processed via the sublexical route. The joint contribution of the two routes within the dual route theory is illustrated by the frequency x regularity interaction. Irregular words, which are also low-frequency, take longer to name than irregular high-frequency words (Seidenberg, Waters, Barnes, & Tanenhaus, 1984). According to the dual route model (Coltheart et al., 2001), this occurs because the irregular phoneme will receive two pronunciation activations, the corresponding irregular one from the lexical route, and the competing regular one from the GPC route. The competition will slow the process down, resulting in slower naming for low-frequency irregular words. For high-frequency irregular words, however, this competition will not occur, because it will have been activated much faster, thereby foregoing any competition resulting from sublexical route activation.

The dual-route cascaded model of reading is a skilled readers model, and was not designed to simulate acquiring or developing skilled reading (Coltheart et al., 2001). The lack of the dual route model to show how its parameter settings have developed has been considered arbitrary and as lacking legitimacy by its critics (Plaut et al., 1996; Rueckl, 2016). Thus, importantly, the DRC model is not a learning model. This has been addressed to some extent by the more recent connectionist dual process model (Perry, Ziegler, & Zorzi, 2007; Ziegler, Perry, & Zorzi, 2013; Zorzi, Houghton, & Butterworth, 1998), of which the most recent implementation in English is the CDP++ model (Perry, Ziegler, & Zorzi, 2010). The CDP+ model has the ability to learn orthography-phonology relationships via its sublexical route, and is thereby able to model the reading acquisition process. However, for the purpose of the current study, the dual route account will refer to the classical DRC model (Coltheart et al., 2001).

2.3 PDP approach

Borne out of the connectionist approach to model cognitive processes in analogy to neurobiological processes, the parallel distributed model of lexical processing (*PDP* or *connectionist* or *triangle model*) envisages a single framework to complete all tasks of lexical processing (Plaut et al., 1996; Seidenberg & McClelland, 1989). Thus, in contrast to the dual route account, all words as well as nonwords are read with the same mechanism.

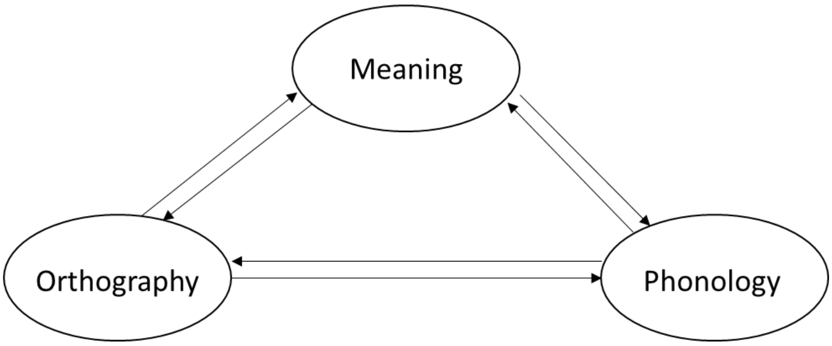


Figure 2.2. Triangle Model following Seidenberg and McClelland (1989).

Figure *2.2* shows a depiction of the triangle model, with oval shapes representing groups of orthographic, phonological and semantic units. Arrows represent the connections between these groups of units via hidden units.

In the implemented connectionist triangle model (Plaut et al., 1996), orthographic, phonological and semantic representations are understood as activation patterns over the respective processing units, where similar word items are characterised by similar patterns of activation. Reading aloud involves the transformation of activation patterns from orthographic to phonological units, or via semantic units. The transformation from one set of units to another is dependent on the weights of the connections. These weights are adjusted by an automatic learning procedure, such as for example back-propagation (Rumelhart, Hinton, & Williams, 1986), depending on the items which are presented to the network. Weights of the connections between units therefore reflect the statistical structure of items learned over time. Thus, frequency and similarity of items guide the learning procedure, which increases weights of connections for items which are presented more frequently and are more similar to other items. This way more frequent and more similar words are read faster than less frequent and less similar words.

Importantly then, the connectionist architecture allows for statistical learning to shape the reading system, which also enables the model to simulate reading acquisition (P. Monaghan & Ellis, 2010) and reading experience (Zevin & Seidenberg, 2006). Thus, as it is able to model how reading developed from its beginnings into a skilled reader model, the distributed connectionist approach possibly constitutes a more workable option to explore and understand the impact of variation in the reading system. (see Rueckl, 2016)

2.4 Regularity and consistency

Both models accept the crucial importance of spelling-sound relationships, but differ in how these have been conceptualised. Early studies quantified the relationship between spelling and sound in terms of correspondence rules between graphemes (smallest orthographic unit, i.e. single letter or letter cluster, which refers to a sound unit) and phonemes, which are single sound units (Marshall & Newcombe, 1973; Forster & Chambers, 1973). Words which could be pronounced according to the grapheme – phoneme correspondence rules (GPC rules) were considered regular. Those words for which the GPC rules could not be applied were termed *exception words* (Baron & Strawson, 1976) or *irregular*. As a rule, then, the correspondence between a grapheme and a phoneme which appeared most often was considered regular.

Words were thus dichotomously classified as either regular or irregular. In reading aloud, irregular words were found to have longer latencies than regular words (Baron & Strawson, 1976)[1]. This so-called regularity effect has been found to be stronger in low-frequency words than high-frequency items (e.g., Andrews, 1982). The concept of regularity is compatible with the assumption of the dual route approach that reading aloud occurs via

---

[1] Baron & Strawson (1976) defined regular words as those which adhered to the pronunciation rules as set out by Venezky, R.L. (1970) The structure of English orthography. The Hague: Mouton, 1970.

two separate mechanisms, either lexically (mandatory for exception words) or sublexically (obligatory for nonwords). Hence, the distinction between regular and irregular words is a classification of spelling-sound correspondences which is congruent with the construct of the dual route model.

Spelling-sound correspondences have also been defined in how consistently words with the same spelling pattern are pronounced in the same or a different way. Specifically, Glushko (1979) showed that consistency affected reading independently from regularity on naming latencies such that regular inconsistent words (*wave*; which has the inconsistent neighbour *have*) took longer to pronounce than regular consistent words (*wade* where the rime is always pronounced as in *made*).

Whilst regularity classifies words into two categories (regular vs irregular), consistency is a continuous measure with degrees of consistency strength (Jared, McRae, & Seidenberg, 1990). This is because within a connectionist network, the consistency effect stems from the similarity of activity patterns across orthographic and phonological units for consistent items (Plaut et al., 1996). When presenting a consistent item to a connectionist network, it will benefit from previous presentations of similar words which have already strengthened connection weights which these two words share, and this will facilitate processing. In contrast, inconsistent items will take longer to process as they are not able to benefit from extant strengthened connections. Rather, items will be slowed down by their inconsistent neighbours.

The importance of spelling – sound consistency for naming has been supported by recent evidence from large-scale studies. In a naming study of 2,428 monosyllabic words, Balota, Cortese, Sergent-Marshall, Spieler, & Yap (2004) reported a facilitatory body-rime consistency effect on naming. Additionally, in a large – scale study on 6,115 monomorphemic multisyllabic words, Yap & Balota (2009) reported a facilitatory consistency effect for naming times. More consistent words were pronounced faster. Spelling-sound consistency (in terms of summed frequency of word body neighbours) has been shown to affect naming of low – frequency words (Jared, 2002) and high – frequency

(Jared, 1997). Although consistency affected both high – and low-frequency words, the effect was significantly stronger for low-frequency words.

Spelling – sound consistency which captures larger, body - sized regularities in words has been found to have a greater effect on naming than regularity, which accounts for regularities at the grapheme-phoneme level (Cortese & Simpson, 2000). In fact, the regularity effect for exception words seems to have been carried by the words' inconsistency (Jared, 2002).

As the present investigation compares reading aloud in orthographies differing in spelling-sound relationships, it seemed opportune to quantify this relationship. When comparing the two concepts regularity and consistency, it seemed that consistency was the more appropriate measure given that it was graded (rather than categorical) and therefore possibly a more sensitive measure to capture how sounds relate to spellings. Furthermore, the concept of consistency is closely linked to the PDP modelling approach. As this theoretical framework is adaptive rather than needing parameter settings changes, it was felt that it may be the better at investigating individual differences (see Chapter 4). Therefore, in the current investigation, the relationship between spelling and sound has been quantified in term of consistency (see Chapter 5).

2.5 Chapter summary

Both reading models have contributed greatly to our understanding of reading aloud. The two differ on a number of assumptions, namely the DRC envisages two routes, a lexical and a sublexical one, whereas the PDP model envisages the same system for all items. Second, but related to the first, whilst spelling-sound correspondences have been conceptualised by the DRC model in terms of the binary concept regularity, it is understood as a statistical occurrence of similarity by the PDP model. Third, in contrast to the DRC, the connectionist nature of the PDP model makes it a learning model. This means that changes in the reading system due to differential developmental or experiential trajectories can be modelled, whereas the dual route model employs fixed parameter settings, which lack the legitimacy of its occurrence/development, i.e. do not explain how the change or difference has come about. Arguably, this may make the connectionist model more suitable to consider changes or variations of the reading process between individuals. Thus, although the current investigation will refer to both models to understand findings, models are referred to at different points of the present investigation. In line with trends in the reading literature, effects of individual differences have mostly been discussed within the framework of the PDP model. By contrast, effects of language transparency refer also to the DRC. However, spelling-sound relationships have been quantified in line with the PDP model. This approach mirrors the fact that the purpose of the study was not to adjudicate between model approaches, but to create a behavioural data set which may elucidate when and how reading aloud is influenced by the language and person-level characteristics.

3 Reading aloud in different languages

Alphabetic languages differ in terms of how reliably letters or letter clusters map onto the smallest sound units called phonemes (Frost et al., 1987; Katz & Frost, 1992; Liberman, Liberman, Mattingly, & Shankweiler, 1980). Languages, which preserve phonological invariance, are termed *transparent* or *shallow* languages and in its purest form graphemes and phonemes have a one-to-one correspondence. Italian and German orthographies are considered to be shallow or transparent, (e.g., Aro & Wimmer, 2003; N. C. Ellis & Hooper, 2001; Seymour et al., 2003; Share, 2008). Other languages are less consistent in the way that phonology is expressed in the writing system. This inconsistency may derive from maintaining morphological invariance, i.e. morphemes (smallest meaning units) are preserved at the expense of phonology, such as *heal* and *health* (Chomsky & Halle, 1968), or it may be due to other factors such as spelling reforms (Seidenberg & McClelland, 1989; Seidenberg, 1992). With increasing orthographic depth, the one-to-one congruency gives way to less reliable letter - sound correspondences. For instance, in English the letter *a* sounds different in the words *cat*, *car*, and *cape*. Amongst the alphabetic languages, English is considered a *deep* or *opaque* orthography.

Given that languages differ in the way they reflect spelling-sound relationships, it has been suggested that this may lead to differences in item processing. The aim of the present study is to examine if differences in reading aloud due to language are still apparent when other individual differences are taken account of, and if this is the case, how these individual differences impact on the reading process.

In the following chapter, two theories, the Orthographic Depth Hypothesis and the Psycholinguistic Grain Size Theory, will be introduced. Both have proposed frameworks to explain differences in reading across alphabetic languages. Whilst the Orthographic Depth Hypothesis presupposes that readers of different orthographies rely on different contributions of reading mechanisms, the Psycholinguistic Grain Size Theory presumes that different orthographies afford different optimal reading unit sizes for reading aloud. Therefore, this chapter will review behavioural findings with regard to differently sized reading units.

3.1 The orthographic depth hypothesis

The orthographic depth hypothesis (Frost et al., 1987; Katz & Frost, 1992) proposes that the transparency of a script influences the reader's reading processes. Shallow languages, where phonology has been transcribed into the writing system, would favour a process whereby phonology is assembled sublexically. Deep orthographies, on the other hand, would require access to lexical information to derive a pronunciation. The relative use of either process would be determined by the orthographic depth of the language. The predominant reading strategy, then, will depend on the relative transparency of the script. The ODH predicts that readers of transparent scripts would be more prone to using the sublexical pathway, whereas readers of opaque scripts would require the predominant use of the lexical route.

Behavioural support for the ODH was presented by Katz & Feldman (1983). They compared reading aloud and lexical decision performance in two languages: Serbo-Croatian, which is considered a shallow orthography, and English, a deep writing system. The lexical decision task is a task in which readers are asked to distinguish between words and nonwords. The aim was to find out if skilled readers of Serbo-Croatian showed increased use of phonological coding to access word pronunciations compared to readers of English. Phonological coding is the process of using letter-sound correspondences to assemble word pronunciations. In both the naming and the lexical decision tasks, target words were preceded by either a semantically related or a semantically unrelated word. Priming with a semantically related word should result in a faster response if the target word representation had been accessed lexically. In the lexical decision task, a task which is thought to require access to the lexicon to retrieve word meaning, readers showed effects of semantic priming. As expected, this was true for both languages, as the task required access to meaning in both languages. Crucially however, in reading aloud, in which pronunciation does not necessarily require access to meaning, only readers of English showed faster responses to semantically primed target words. These results supported the hypothesis that orthographic depth mediated the code which was accessed for reading, as they found a measurable predominance of the use of visual—orthographic code for English. Thus, the results clearly supported the assumption that readers in English accessed words lexically, whereas this was not the case for readers of Serbo-Croatian.

Further evidence was presented in a seminal study by Frost et al. (1987), who compared reading in three languages, Serbo-Croatian, English and Hebrew. The Hebrew script does not specify vowels, and hence can be considered even more ambiguous than English. If deep orthographies were read with more direct access to the lexical entry in the internal lexicon, and shallow orthographies more with recourse to prelexical phonological assembly, then lexical factors such as word frequency and lexicality (word-nonword distinction) should show a greater effect in deep languages than in shallow languages. Specifically, high-frequency words should be read faster than low-frequency words, and words should be read faster than nonwords. These effects should be greater for readers of deep orthographies. To test this, word stimuli across the three languages were matched in frequency and number of phonemes. Participants completed naming and lexical decision tasks in their respective languages. In accordance with the ODH, the results supported the assumption that the deeper the orthography, the more reading occurred through lexical mediation. Lexicality affected naming in Hebrew most, followed by English, and showed very little influence in Serbo-Croatian. Thus, with increasing orthographic depth, lexicality showed a stronger effect on naming times. By contrast, lexicality effects were similar across languages in the lexical decision task, which requires access to lexical information to determine if an item is a word or a nonword. In fact, the authors found that naming and lexical decision times for readers of Hebrew were very similar. The authors also noted that although the frequency effect showed the same result pattern and direction as the lexicality effect, the effect was in fact not significantly different between English and Serbo-Croatian. If only frequency had been employed as a measure for lexical mediation, then the difference between English and Serbo-Croatian would not have been detected. Therefore, it was foremost the lexicality effect (words read faster than nonwords) which supported the assumption that lexical mediation increased with increasing orthographic depth.

3.1.1 Section summary

In sum, the findings that semantic priming accelerated naming times in English but not in Serbo-Croatian, and the finding that the difference between word and nonword reading was smallest for Serbo-Croatian compared to the more opaque scripts of English and Hebrew, were taken as support for the ODH assumption that readers of transparent orthographies relied more on phonology, and readers of opaque scripts relied more on lexical access. As the ODH essentially proposes two mechanisms for reading, a sublexical and a lexical process, the assumption of the ODH is easily compatible with the dual route model.

3.2 The psycholinguistic grain size theory (PGST)

An alternative account to the ODH is provided by the psycholinguistic grain size theory (PGST; Ziegler & Goswami, 2005), which aims to deliver a theoretical framework for understanding reading acquisition in the context of impaired reading (dyslexia) and skilled reading across different languages. Like the ODH, the psycholinguistic grain size theory proposes that the reliability of spelling-sound mappings will influence reading processes. As seen above, the ODH suggests that readers may predominantly use either the lexical or the sublexical pathway within a dual route framework depending on the orthography they are reading. The PGST, by contrast, specifies that it is the size of the reading units which will vary with the reliability of spelling-sound correspondences. Transparent writing systems can be read using the smallest-sized reading units, the phonemes, but writing systems, in which grapheme -phoneme correspondences can be ambiguous, will require the reader to adopt larger sized reading units, such as for example body-rime units, to reliably read a word. Body-rime units are made up of the vowel and subsequent consonants in a syllable, such as *-at* in *cat*. With growing reading experience and concomitant increase of phonological and orthographic knowledge, these word-specific grain sizes will adapt and change as the reader learns about and adjusts to the correspondences between phonology and orthography of their respective language. Therefore, in terms of the PGST, skilled reading is understood as a direct continuation and result of the reading acquisition process, as the reading system is shaped by the spelling-sound consistency of the orthography. The PGST therefore suggests that reading development will leave its developmental footprint in skilled reading, and that readers of opaque orthographies will have learned to use reading units of multiple sizes, whilst readers of transparent orthographies will have developed a tendency for small-unit processing.

3.2.1 PGST: developing readers

As the PGST is largely a theory of reading acquisition, it seems fitting to first review behavioural findings from developing readers. It has repeatedly been reported that children learning to read opaque orthographies are found to be slower and less accurate in learning to read than developing readers of more transparent orthographies. For example, English primary school children were reported to make more errors and display slower reading speed in nonword reading than their German counterparts, even though some of them had had more years of reading instruction (Frith, Wimmer, & Landerl, 1998; H. Wimmer & Goswami, 1994). Equally, in a large comparative study of beginner readers in thirteen European languages, word and nonword reading of English was substantially less accurate and slower than word reading in more consistent orthographies after one year of schooling (Seymour et al., 2003). The PGST suggests that this acquisition delay in deep scripts like English is due to the unreliable grapheme – phoneme correspondences, which require young readers to learn correspondences at unit sizes other than the grapheme-phoneme level in order to achieve reliable phonological decoding. It is the additional number of spelling – sound correspondences at larger unit sizes that young readers of opaque orthographies have to learn additionally which slows down reading acquisition and makes it more error prone.

There is more direct evidence to support this assumption of differential reading size usage by developing readers in different orthographies. In a comparative study the majority of errors made by beginner readers of the transparent orthography Welsh tended to be nonword substitutions which had large overlaps with the test word, whilst a comparable sample of English children produced more null responses and false whole word substitutions. The tendency for English children to substitute a misread word with another real word was taken to indicate a lexical access approach, i.e. to recognise the word in its entirety. Correspondingly, the tendency of readers of the transparent orthography Welsh to mispronounce words as nonwords, hinted at a small grain sized phonological recoding strategy. Hence, developing readers of English and Welsh seemed to be using different unit sizes for reading. (N. C. Ellis & Hooper, 2001) In a further study, children reading deep orthographies have been shown to switch between different unit sizes, whilst beginner readers of consistent orthographies applied the same small unit size when reading. Goswami, Ziegler, Dalton, & Schneider (2003) presented children with small-unit and large-unit nonwords either in mixed or uniform lists. Whereas German children showed no difference in reading accuracy, English children displayed a disadvantage for the mixed list. This was interpreted as illustrating switching costs between using small and large units in English children, but not in German children. English children switched between reading unit sizes and hence made more mistakes, whilst German children presumably continued to rely on the smallest unit size, i.e. grapheme-phoneme, and thereby demonstrated higher accuracy. It therefore seems that in line with the PGST, beginner readers differ in the default grain size used for reading.

3.2.2 PGST: skilled adult readers

The PGST holds that developing readers of different orthographies use different unit sizes for their reading depending on how reliably sound is expressed in spellings. According to the PGST, these differences may still be evident in skilled readers. One study in particular has been considered to show strong support for the notion of the PGST that the adult reading system retains a developmental footprint. In a seminal study, Ziegler et al. (2001) reported that English and German skilled readers exhibited tendencies to use differently sized reading units in a speeded pronunciation task using the same words and nonwords for both languages. The authors capitalised on the fact that German and English share a common heritage, and hence some words have remained very similar across both languages. Yet, whilst German is fairly transparent in terms of spelling-sound correspondences, English is notoriously unreliable. Ziegler and colleagues identified 80 test words, which share the same meaning and exist in the same or nearly the same orthographical form (so-called *cognates*). All words were regular, i.e. they consistently followed the grapheme-to-phoneme conversion rules, and were not exception words. Words were monosyllabic, 3 to 6 letters long, and had an average frequency of 100 per million (range between 2 – 1,035 per million). Additionally, 80 nonwords were chosen. Half of the words and nonwords had a large orthographic body neighbourhood (body-*N*), the other half had a small body-*N*. Orthographic body neighbourhood refers to the words which share the same orthographic *rhyme* or *body*, for instance, *s-ea* and *fl-ea* are body neighbours of *t-ea*. The close matching of these cross-language stimuli meant that any differences found could more reliably be attributed to the orthographic depth of the respective language. German readers showed a stronger length effect for both words and nonwords compared to English readers, i.e. the longer the item, the longer the naming latency. In contrast, English readers more than German readers demonstrated a tendency to read items with large body neighbourhoods faster than items with small body neighbourhoods. The result was considered to support the assumption that the same orthographic items were read with recourse to reading units of different sizes, depending on the orthography. The length effect was considered to evidence small-unit processing, whilst the body-*N* effect was a marker for larger unit processing. These two effects will be considered in turn.

3.2.3 Section summary

The PGST (Ziegler & Goswami, 2005) suggests that in deep orthographies developing readers take longer to learn to read because appropriate reading unit sizes will be more varied, whilst readers of transparent scripts can rely on the smallest grain size. Studies reporting that English-speaking developing readers showed slower and more error-prone nonword reading (e.g., Frith et al., 1998), were more likely to substitute words for nonwords (N. C. Ellis & Hooper, 2001) and showed greater switching costs in word lists mixing small- and large unit sized nonwords (Goswami et al., 2003), seem to give support to this assumption. Although a theory of reading acquisition, the PGST assumes that the tendency for small-unit processing remains with skilled readers of transparent scripts, whilst readers of opaque scripts will be more prone to employ multiple unit sizes. In support, Ziegler et al. (2001) reported a body-*N* effect in reading aloud for English-speaking skilled readers and a length effect for German-speaking skilled readers.

3.3 Length effects in different languages

The length effect has been considered as an index for serial processing in transparent scripts. In the following, the findings regarding the length effect for nonword and word naming will be reviewed, first for developing readers, and then for skilled adult readers. As outlined above, the PGST predicts more small-unit processing for readers of transparent scripts, and although this originates in the reading acquisition process, it should still be present to some degree in skilled readers. To anticipate the review outcome, investigations on the length effect have not been congruous in their results. A brief review of how current reading models explain the length effect gives some indication that the effect may vary between individuals not only on account of the language that is being read.

3.3.1 The length effect in young readers of different orthographies

Regarding developing readers, it seems that young readers of transparent scripts use more small-unit processing than developing readers of opaque scripts. Specifically, word length has been found to explain more variance in reading aloud latencies of words in transparent orthographies than in opaque scripts for young readers (N. C. Ellis et al., 2004). Similarly, in a comparison of English-speaking and German-speaking primary school children, Ziegler, Perry, Ma-Wyatt, Ladner, & Schulte-Körne (2003) reported a length effect in both word and nonword reading for all children, but this was stronger for the German sample. This result supported the idea that compared to their English peers, the German children applied a small-unit processing strategy to arrive at the correct pronunciation.

A recent study (Rau et al., 2015) measured eye-movements in sentence reading of German and English primary school children matched for reading ability. A stronger length effect was found for German than English children in first-pass reading, which most likely reflected early processes such as letter identification, encoding and graphemic parsing (Perry et al., 2010). Given that letter identification and encoding is unlikely to differ between languages, Rau et al. (2015) suggested that the stronger length effect for German children reflected the longer graphemic parsing process. According to the dual route account (Perry et al., 2010), graphemic parsing is the process of identifying graphemes in the letter string, which operates in a serial left-to-right manner and is indicative of small-unit processing. Thus, stronger length effects in the German language group indicated more small-unit processing, whilst English children were assumed to make greater use of larger units for this main reading process (first-pass reading). However, for English children, a stronger length effect was found in rereading times for low-frequency words and nonwords compared to German children. This indicated English children seemed to employ a more serial reading strategy when they had not been able to recognise the item in the first instance, as reflected by the fact that they took longer to reread items, albeit this secondary reading mechanism seemed somewhat less effective than it was for the German children. (Rau et al., 2015)

In sum, evidence points to serial processing as indexed by the length effect for developing readers of German, but less so for those reading English.

3.3.2 The length effect in skilled readers of different orthographies

The present study, however, is concerned with skilled reading. The PGST predicts that small-unit processing should to some degree still be observable in adult reading. Yet, results of studies reporting on word and nonword length effects have been mixed.

Although Ziegler et al. (2001) reported a stronger length effect for German readers for both word and nonwords compared to their English-speaking counterparts, this has not been unequivocally replicated. In a naming study for 3-8 letter words in Italian, a transparent orthography, adult readers showed no significant length effect (Spinelli et al., 2005). Moreover, an unexpected result was reported from another cross-language experiment. Rau et al. (2015) compared eye movements in English and German adult readers reading high -

and low-frequency words and nonwords embedded in sentences. Items were either considered short (3-5 letters) or long (6-8 letters). Gaze duration was taken to reflect general word recognition (Juhasz & Rayner, 2003). No differences between English and German readers were found for high - and low-frequency words, but crucially and unexpectedly nonwords were read significantly slower by English adults. The authors suggested that for nonword reading both groups resorted to the use of small units, but that German readers were more experienced in using this process. This experience helped German readers to read the nonwords at a faster pace, whilst English readers took comparatively longer, as evidenced by a length effect in nonword reading for the English readers. Rau and colleagues concurred with the view that differences in reading due to orthographic consistency were still apparent in skilled adult readers, implying that skilled reading of nonwords reflected traces of developmental responses to orthographic challenges. Thus, Ziegler et al. (2001) and Rau et al. (2015) came to the same conclusion that challenges due to orthographic depth were still measurable in reading processes of skilled readers. Yet, Rau et al. reported a length effect in nonword reading for English readers, whilst Ziegler and colleagues reported a length effect for German readers for both words and nonword naming. Although these studies had admittedly employed different methods, namely naming latencies and gaze duration respectively, it is still surprising that opposite observations were interpreted as evidence for more small-unit processing in German readers.

Other studies reported mixed results for English only (for a recent comprehensive review see J. J. Barton, Hanif, Eklinder Björnström, & Hills, 2014). Using 3-6 letter items, Weekes (1997) observed a length effect for nonword, but not for word reading. In contrast, in two large-scale studies in English, length effects were reported for 2-8-letter-long monosyllabic words (Balota et al., 2004) and multisyllabic words with a mean of 6.7 letters (Yap & Balota, 2009), with longer words eliciting longer naming reaction times. The effect was significantly stronger for low-frequency than for high-frequency words in monosyllabic words (Balota et al., 2004).

Some of the variation in the findings regarding the word length effect may stem from the item lengths themselves. New, Ferrand, Pallier, & Brysbaert (2006) demonstrated that in a lexical decision task the relationship between word length and recognition latencies was a quadratic one. Short words showed a negative correlation with length, whilst length had no effect on medium-sized words (5-8 letter words). When words were longer than 8 letters, length became inhibitory, i.e. words were recognized more slowly. Thus, words of middle lengths received the shortest RTs. Crucially, Yap & Balota (2009) showed that this quadratic relationship was present in both lexical decision and naming. In line with New et al. (2006), they suggested that the recognition efficiency for medium-sized words might be due to these word lengths being the most common, i.e. that the lengths for the most common words would also then constitute the optimal length for reading. As average word length is slightly different between languages, this ‘optimal perceptual span’ (New et al., 2006; Yap & Balota, 2009) may vary with language (Yap & Balota, 2009).

However, quadratic length effects may not account for all of the findings reported here. In fact, most studies reviewed above included words of short to medium-sized lengths, and thus should have reported an inhibitory length effect for the shorter words at least, which was not the case.

In summary, although the length effect has been considered as an index for small-unit processing, studies with adult readers have reported mixed findings on when length effects occur. Whilst words generally seem to be named faster than nonwords across all alphabetic scripts (Frost et al., 1987; Ziegler et al., 2001, 2003), neither words nor nonwords are consistently read slower by readers of transparent orthographies compared to readers of opaque orthographies. This weakens the assumption of the PGST, that skilled readers of transparent scripts should still show a developmental footprint of relative more small-unit processing than readers of more opaque scripts. However, it has been suggested that the length effect may be a function of individuals’ variation of the ‘optimal perceptual span’ which may vary with language (Yap & Balota, 2009).

3.3.3 Theoretical explanations of the length effect

3.3.3.1 The length effect in the DRC model

For the DRC (Coltheart et al., 2001), the length effect is important as it supports the assumption of the dual route system. In dual route systems, the length effect results from serialisation of phonological assembly. For the DRC, this assembly is the matching of graphemes onto phonemes according to a set of grapheme – phoneme correspondence rules (GPC rules), and this is set in the sublexical route. The phonological assembly process occurs serially from left to right and consequently longer items take longer to pronounce than shorter items. Models simulating a dual route reading system have produced the length effect in monosyllabic nonwords (Coltheart et al., 2001; Perry et al., 2007) and polysyllabic nonwords (Perry et al., 2010).

However, within the framework of the DRC, the length effect may also be a result of the whammy effect, which is equally a serial effect (Rastle & Coltheart, 1998). According to this account, the sublexical route assembles letter by letter. Sometimes, several phonemes could be applicable to letter correspondents. For this reason, a rule order establishes which match is considered and applied first. If a letter has been wrongly matched with a phoneme, then this needs to be revised in the next cycle. This revision process will slow down nonword assembly, and previous activation of the wrong phoneme will interfere with the settling of the correct activation.

Pertinent for the current investigation, Perry & Ziegler (2002) explored the impact of vowel length revision in German, when comparing computational DRC models in English (Coltheart et al., 2001) and German (Ziegler, Perry, & Coltheart, 2000). Using the word and nonword items from the Ziegler et al. (2001) study, Perry & Ziegler (2002) simulated a greater nonword length effect in the German DRC than the English DRC, when employing the vowel length revision process in German. In general, German vowels are pronounced short if followed by a single consonant, but long, if two or more consonants follow. The DRC model applied this as a GPC rule. Whilst the German DRC assigned a short vowel by default, it had to correct this into a long vowel if it found that the vowel was followed by two or more consonants. This revision process took time and slowed down pronunciation. Longer monosyllables (3 letters +) are more likely to be subject to this revision process, as they are more likely to have two consonants follow their vowel. Thus, longer nonwords were subject to the whammy effect and this created the nonword length effect.

For words, however, the same simulation did not show a length effect for German. The authors reasoned that the German word length effect may be due to a stronger propensity for shallow orthographies to use sublexical processing. They therefore changed the parameter setting of the German DRC to weaken the lexical route and to strengthen the sublexical route. With these changed parameter settings, they also found a length effect for German words. The nonword length effect for German had reduced, but was still stronger than the nonword length effect in English.

In sum, for the dual route account the length effect is an important indicator of serial or small-unit processing and it has been fairly successful at simulating behavioural data. The word length effect simulation in the DRC models is reminiscent of the ODH which also assumes that different orthographies were handled through variation in the relative strength of contribution of lexical and sublexical routes. This account of the length effect thus proposes different proportional input of the two routes towards the final pronunciation and suggests that languages may differ in terms of the route contributions to derive a pronunciation by using more of the one than the other route.

3.3.3.2 The length effect in the PDP model

In the PDP model, the length effect cannot be so readily explained as in the dual route model, as it assumes parallel mechanisms of the network of assembling a pronunciation for all phonemes. Nevertheless, Plaut et al. (1996) reported a small word length effect in their third naming simulation. Plaut et al. (1996) suggested that parallel processing of letters may be dependent on the individual in such a way that the number of letters, which are captured by the window of parallel computation, fluctuates between participants as a function of experience.

In effect, this account concedes some form of serialisation as the origin of the length effect. Plaut (1999) proposed an alternative connectionist model with a particular view to processing polysyllabic words. As in previous models, words were presented in their entirety. However, in this model, letters were position-specific, and generated output as a sequence of phonemes rather than a static representation of an entire item. When encountering a difficulty, the model was able to refixate its attention focus on the difficult grapheme. The network would then retry to produce the correct pronunciation for the difficult grapheme, and the rest of the word. The model aimed to pronounce graphemes taking into account word context. When the model was in its early training, it needed to refixate more often, as it came across graphemes which it did not know in the given context. With increased practice, refixations became less common, and the larger parts or even the whole word could be pronounced at once. Thus, within this model, which in theory would be able to read any letter string of any length, the refixation mechanism and the serialisation of the phoneme output produced the length effect.

Importantly, however, this account also accommodates individual differences as an explanatory factor for the variation of the length effect. The window of parallel computation (Plaut et al., 1996) may be the same as the optimal perceptual span (Yap & Balota, 2009) based on a suggestion by New et al. (2006) that the length of the most common words ultimately created an optimal item length for a reader. This ‘most common length’ may vary with language. (Yap & Balota, 2009) Alternatively, Plaut et al. (1996) had suggested that the optimal window may vary with readers' experience.

However, it also needs to be noted that the length effect may occur due to processing effects outside of the PDP reading model. (Seidenberg & Plaut, 1998) Some support for the assumption that the length effect may stem from visual encoding of letters has been received from a simulation study (Chang, Furber, & Welbourne, 2012). According to this account, PDP models have not been able to replicate the length effect, because they did not model the process of letter encoding. Unfamiliar, or unknown, word items would take longer to be visually encoded, and this effect would increase with item length. Initially, the authors added a visual input layer to the PDP model of Plaut et al. (1996), but it did not produce the length effect. However, when the pre-defined orthographic representations of the same extended model were substituted by an extra hidden layer, which essentially made it possible for the model to compute its own orthographic representations, a length effect was produced for both words and nonwords. The length effect was greatest for nonwords, less for low-frequency words and smallest for high-frequency words. The authors concluded that the length effect resulted from the process of transferring visual representations into orthographic representations. It is conceivable that this notion of the length effect is compatible with the concept of an optimal window of parallel computation, which is variable on an individual level.

In sum, the theoretical account associated with the DRC model assumes that the word and nonword length effects reflect a stronger contribution from the sublexical pathway relative to the lexical pathway, and this propensity is thought to be stronger for more transparent languages. This view is congruent with the ODH. It is also compatible with the PGST which expects smaller reading units for readers of transparent orthographies. With regard to PDP models, they have been able to show some lengths effects when conceding some serialisation in the item uptake due to refixation. Alternatively, or possibly concurrently, it has been suggested that length effects could emerge due to individual variation in the process of visual letter encoding. This suggestion may be compatible with the less defined notion of an optimal perceptual window.

3.3.4 Section summary

The length effect has been taken to indicate serial processing in nonword reading (e.g., Coltheart et al., 2001) and small-unit processing in readers of transparent languages (Ziegler & Goswami, 2005). There has been supporting evidence showing that developing readers of German, but less so developing readers of English, show a length effect in oral reading (Frith et al., 1998; H. Wimmer & Goswami, 1994). However, in skilled readers, findings of the length effect were more mixed. Whereas words were reliably named faster than nonwords (Frost et al., 1987; Ziegler et al., 2001, 2003), neither words nor nonwords were consistently read slower by readers of transparent orthographies compared to readers of opaque orthographies (Rau et al., 2015; Spinelli et al., 2005; Ziegler et al., 2001). Thus, skilled readers of transparent scripts do not seem to have reliably retained a developmental footprint of more small-unit processing than readers of opaque orthographies, as suggested by the PGST.

Alternatively, from a PDP perspective the length effect may be due to participants' variation of the 'optimal perceptual span' which may vary with language (Yap & Balota, 2009). The PDP model for reading aloud has been able to accommodate length effects when allowing for some serialisation through re-fixation to occur.

3.4 Effects of larger grain sizes in different languages

As reviewed in the previous section (see section 3.3), although the length effect has been taken as an indicator of small-unit processing in readers of transparent languages, there are both contradictory behavioural and theoretical accounts, which make it difficult to accept the length effect as a reliable marker for small-unit processing. The alternative approach is to explore markers for large-unit processing. Whilst lexicality and frequency effects have been employed as such markers within the ODH (Frost et al., 1987), body-*N* effects were explored in the seminal study by Ziegler et al. (2001). This section will review research which has investigated body-*N* effects as markers for larger reading units in different languages. It will then introduce the alternative, but related psycholinguistic body-rime *N* effect. Finally, there will be a short overview of a study which compared the use of several grain sizes in nonword reading by German and English readers and suggested that the use of specific grain sizes may not be uniquely tied to the readers' mother tongue (Schmalz et al., 2014).

3.4.1 Body-*N*

Body-*N* are the number of words in a given corpus which have the same orthographic body as the target word. For example, *t-ea* and *p-ea* share the same orthographic body with *s-ea,* but not with *b-ee*. They are orthographic body neighbours, or body-*N*. In contrast, *b-ee* only shares the same phonological rime, but not the same orthographic body. Ziegler et al. (2001) reported greater body-*N* effects for readers of English than readers of German and suggested that these indicated the use of larger sized reading units (refer to section 3.2.2).

However, other studies with skilled readers were not able to trace differences across languages in body-*N* effects in reading aloud. Schmalz, Robidoux, Castles, Coltheart, & Marinus (2017) sought to replicate the findings by Ziegler et al. (2001) with the original data averaged across items. Schmalz and colleagues used two different statistical methods (multiple regression and Bayesian analysis) and changed the body-*N* variable from discrete to continuous. They were not able to replicate Ziegler et al. (2001)'s findings: there was neither evidence for a body-*N* effect, nor for a body-*N* x language interaction.

Schmalz et al. (2017) then investigated whether the body-*N* effect could be found for the same items but with another group of participants. For this purpose, they extracted data for 79 words of the English stimuli set used in Ziegler et al. (2001) from skilled adult readers, who had participated in the English Lexicon Project (Balota et al., 2007). The data was submitted to a linear mixed-effects analysis (Baayen, Davidson, & Bates, 2008) with random effects for participants and items, and a fixed effect for body-*N*. No evidence for a body-*N* effect was found. The result was also confirmed by means of a subsequent Bayes factor analysis. Thus, using data from different participants for the original Ziegler et al. item set in English, Schmalz and colleagues found no evidence of a body-*N* effect. Words which possessed the same body as many other words were not named faster than words which shared their orthographic body with few other words. As Schmalz et al. note, although this analysis did not compare two languages, it did not reproduce the results reported for the English stimuli, even though the body-*N* effect was reported to be stronger for English than for German. If the result for the English stimuli set could not be replicated, then this renders cross-language differences in body-*N* effects uncertain.

3.4.2 Phonographic body-rime neighbours

Andrews (1997) observed that word bodies may combine orthographic and phonological *N* effects. Peereman & Content (1997) investigated the contribution of the different kinds of neighbours: orthographic but not phonological neighbours, phonological but not orthographic neighbours, and finally, words which are both orthographic and phonological neighbours. Peereman & Content (1997) referred to words which differed in only one letter and in only one phoneme collectively as phonographic neighbours (RACE – FACE, RICE, RATE). They found that only phonographic neighbours emerged as predictors for (residual corrected) naming times. In fact, body-rime *N* (e.g. RACE – PACE) were the only group of phonographic neighbours which produced significantly shorter naming times for nonwords and words. Their results suggested that body-rime *N* effects carried the orthographic *N* effect in French. More recently, Arduino & Burani (2004) presented evidence that nonword naming in Italian was facilitated when nonwords had a large neighbourhood. The majority of word neighbours differed from the nonword stimuli in the first letter. The stimuli manipulation thus implies that these neighbours were in fact mostly body-rime *N*. The facilitating effect of body-rime *N*, therefore, seems to extend to readers of Italian, a very transparent language.

Adelman & Brown (2007) presented four large – scale regression analyses on English word naming latencies, which included a large number of neighbourhood variables as predictors: orthographic *N* and phonographic *N* (which included all possible neighbours); so-called friends (words, which share an orthographic and phonological rime, i.e. the presently discussed body-rime *N*), so-called enemies (rime neighbours which look like they rhyme, but are pronounced differently) and consistency (friends divided by friends plus enemies). Only the phonographic *N* effect emerged in all four regression analyses. The 'number of friends' variable, which is the variable of interest in the current discussion (in the present investigation: body-rime *N*), did not emerge as an effect, except in one of the analyses, and in this case, it seemed to have been inhibitory to naming speed. However, it is possible that the facilitating effect of body-rime *N* may have already been subsumed by the consistency effect.

Thus, although there have been suggestions that body-rime *N* may carry the body-*N* effect, research findings do not suggest an unequivocal facilitating effect for naming. Whilst a facilitating effect has been reported for French and Italian, results do not suggest such a (facilitating) effect for English.

3.4.3 Use of differential reading unit sizes

However, there is some evidence that skilled readers vary in terms of reading unit sizes as a result of the language they read. Schmalz, Marinus, Robidoux, Palethorpe, Castles and Coltheart (2014) explored nonword reading in English and German. Nonwords were manipulated so that it was possible to determine which grain size was used by the reader for reading aloud: depending on grain size used, nonwords would receive different pronunciations. This way, the authors were able to determine whether items had been read via strict grapheme-phoneme conversion (as the smallest reading units), context-sensitive grapheme-phoneme conversion which took into account letter context, or 3) body-rime units (codas which shared both an orthographic body and a phonological rime, e.g. *-ea* in *tea* and *sea*).

The results showed that English readers seemed to make very sparse use of the smallest unit size, the context-insensitive grapheme-phoneme correspondences (GPCs). They mostly applied context-sensitive GPCs, but the design was such that is was not fully possible to reliably distinguish between the use of context-sensitive GPCs and body-rime units for English. Importantly, however, a large number of responses were in fact other, unforeseen pronunciations, indicating that the three grain sizes investigated did not encompass all options used by participants.

German participants living in an English-speaking environment (German bilinguals) mostly made use of context- insensitive GPCs, i.e. they used the smallest unit size, but they also employed the two larger unit sizes (body rimes and context-sensitive rules). German speakers who lived in Germany (German monolinguals) made use of all three unit-sizes. In comparison to the German bilinguals, they derived their pronunciations less through context-insensitive GPCs (smallest units), and used more context-sensitive GPCs. The authors suggested that as German monolinguals were more exposed to German texts, they attained a higher level of proficiency than the German bilinguals, and subsequently were able to make more use of the bigger units, the context-sensitive GPCs. Both German-speaking groups made comparable use of rime units. Importantly, however, within the two German reader groups there also seemed to be a lot of variation as to which unit sizes were employed.

In a further experiment, German bilinguals were asked to read the English nonwords as if they were unfamiliar English words, and the authors found that in fact the bilinguals tended to use the rime units more often than English monolinguals. Moreover, like the English monolinguals, the German bilinguals did not seem to employ the context-insensitive GPCs (smallest units). Both groups employed the context-sensitive GPCs, although the German bilinguals perhaps to a lesser degree. Schmalz et al. (2014) concluded that German speakers read nonwords relying on all unit sizes. When reading in their own language, Germans tended to use more small units that English readers, although all groups relied on units of all sizes. When English – German bilinguals read English nonword items as if they were new English words, the strategy employed changed to a stronger reliance on context-sensitive GPCs (larger units), and the smallest unit size was hardly used at all.

Although broadly in line with the PGST, which predicts that German readers would read with greater reliance on the smallest units and English readers with greater reliance on larger units, the authors pointed out that the results did not support the idea of a developmental footprint of reading acquisition in a language. Instead, they argued that it was the orthography itself which determined which reading units were employed, rather than which language the participant grew up with.

3.4.4 Section summary

Although Ziegler et al. (2001) reported evidence for greater effects of body-*N* as markers for large-unit processing in English than German, this has not been replicated so far. It has been suggested that phonographic body-rime *N* (body-*N* which are also phonological neighbours) may carry body-*N* effects (Peereman & Content, 1997). Studies, which took into account both orthographic and phonological *N* in the form of body-rime *N*, reported a facilitating effect for naming in French (Peereman & Content, 1997) and Italian (Arduino & Burani, 2004), but not in English (Adelman & Brown, 2007) These observations do not support increased larger grain size use by readers of opaque orthographies than by readers of transparent scripts. Some recent evidence from nonword reading suggests that on average skilled readers adjust their reading unit sizes according to the requirements of the orthography they are reading. Notably, however, even within language reader groups there seems to be some variation as to which unit sizes are employed by readers, indicating that there are some individual differences between readers. (Schmalz et al., 2014)

3.5 Chapter summary

As languages differ in the way that orthography portrays phonology, the orthographic depth hypothesis (ODH; Frost et al., 1987; Katz & Frost, 1992) posits that all readers use both lexical and sublexical reading routes, but that the balance in the contribution of the outputs of the routes to the activation of lexical phonology varies between languages. Consequently, in deep orthographies, reading aloud is more dependent on the lexical route due to the opaqueness of the spelling-sound relationship.

As an alternative explanation, the psycholinguistic grain size theory (Ziegler & Goswami, 2005) suggests that orthographic depth determines the linguistic grain size of the reader. Due to the increased complexity and unpredictability of more opaque orthographies, readers are forced to use larger units to derive at reliable pronunciation accuracy. The PGST suggests that the use of differential grain size units carries over to skilled reading.

However, behavioural studies with skilled readers which have used length effects as markers for small-unit processing and body-*N* as markers for large-unit processing have provided mixed results. More recent investigations point to a more complex use of differential grain sizes in both transparent and opaque languages, which seems to be influenced by the script's transparency rather than which language the readers grew up with.

Importantly, even within reader language groups variations of different grain size units were noted.

The present study will compare naming for a set of cognates by English and German participants, thereby following previous work by Ziegler et al. (2001) and Schmalz et al. (2017). In order to link the results to previous literature, psycholinguistic predictor variables will include both length in letters and phonographic body-rime *N*. However, as will become clear in the next chapter, the present study will also add a further dimension to the investigation. For the first time in a cross-language study, the analysis will also include a set of participant variables. This will make it possible to explore if any psycholinguistic effects such as length and phonographic body-rime *N* are modulated by language and participant characteristics. It may be that individual differences variables will clarify some of the mixed findings which have been reported for effects indicated small – and larger-unit processing. The individual differences will be introduced in the next section. The main aim will be to see if language differences will remain, when allowing for participant variable variation other than language. The second aim will be to observe in which way individual differences may influence processing differently in languages differing in spelling-sound consistency.

4 Individual differences

Language transparency is only one dimension on which reading can vary between people. Other factors have been shown to influence the reading system (Butler & Hains, 1979; Katz & Frost, 1992; Mason, 1978; Seidenberg, 1992). The tendency for research to average across readers and not to take individual differences into account may have resulted in effects having been cancelled out between reader groups (Andrews, 2012). Increasingly research has edged towards exploring how individuals' reading may differ and to address this variability within reading models (Plaut et al., 1996; Zevin & Seidenberg, 2006; Ziegler, Perry, & Zorzi, 2014). In fact, expectations have been raised that individual differences as well as variation due to the language used should be part of constructing universal models of reading (Rueckl, 2016). The current study is therefore timely, as it aims to find whether differences due to language are still evident when other individual differences have been controlled, and if so, how individual differences manifest themselves in reading across different orthographies.

Yet, which individual differences should be taken into account? In the current study, the choice has been guided by how reading models have conceptualised individual variability. Thus, the study focusses on individual difference in nonword reading (or direct orthographic – phonological pathway efficiency), use of semantic knowledge in naming, lexical quality of orthographic representations and differential degrees of print exposure. Whilst the first three can be considered as variations of the individual reading system, the latter is dependent on experience. However, experience itself will in turn have shaped the reading system, and it is thus clear, that even though the current study explores these individual differences as separate aspects of reading, they are also shaped by each other.

The current approach is by necessity selective. An important individual difference which is not being taken into account in this study is reading instruction. There is evidence that teaching practices have a measurable impact on reading development (Connelly, Thompson, Fletcher-Flinn, & McKay, 2009; Goswami, 2008). Specifically, there is some evidence that phonics instruction is beneficial for a wide range of reading skills in younger readers (Ehri, Nunes, Stahl, & Willows, 2001; Landerl, 2000) When including differences in reading instruction, this has also improved computational models of reading. When simulating reading development in English and German, Hutzler, Ziegler, Perry, Wimmer, & Zorzi (2004) showed that the different trajectories of reading development between languages were best captured when different teaching methods (phonics for German children; whole-word teaching for English) were included in the model. However, the scope of the current project, and the fact that adult readers might be unreliable in correctly remembering the methods by which they were taught, precluded instruction method as a variable to use in this study.

As mentioned briefly above, individual differences (IDs) refer to reading model accounts of individual variation. It seems opportune at this point to mention that apart from early investigations on lexical versus sublexical readers, which can be accommodated by the dual route framework (Coltheart et al., 2001), most advances into the field of reading-related individual differences have been made within the framework of the PDP model (Plaut et al., 1996; Seidenberg & McClelland, 1989; but see Adelman, Sabatos-DeVito, Marquis, & Estes, 2014). This predominant reliance on the PDP model as a theoretical framework will be reflected in the following account of individual differences, and in our naming study analyses (see Chapters 7 and 8).

However, individual differences in skilled reading have also been framed by the Lexical Quality Hypothesis (LQH; Perfetti & Hart, 2002; Perfetti, 2007), a word-level concept of how reading can vary by the reader. This account allows for maximal variation between readers, as it focusses on the variability of the lexical representation itself for each word. Perfetti and colleagues consider the word as an entity made up of three constituents: the orthographic, the phonological and the semantic component. For each word that a person has in her vocabulary, each of the three components may be variably developed, and readers may vary generally on how well specified word representations are. Generally, unskilled readers will have lower quality representations than skilled readers, and this systematic variation between readers would account for the individual differences in measuring reading performance. However, Perfetti and colleagues stress that the real difference between good and poor quality representations is indeed at the word level. Skilled readers are those with many good quality representations, and poor readers are those with representations where the constituents lack some quality. The current study will take specific recourse to this account in terms of lexical quality of orthographic and semantic knowledge. However, it needs to be noted that whilst the aim of the present study is to investigate the interplay of item, language and participant characteristics, the LQH focusses on individual word level representations as the locus of participant differences. It is unlikely that the present data and the planned analysis will be able to address these differences with sufficient detail or accuracy. Moreover, the focus of the LQH is to understand individual differences in text comprehension, whereas the present study is concerned with word naming. Finally, the LQH is not a reading model as such, but addresses the exploration of individual differences in reading with a concept that cannot (yet) be computationally implemented. For these reasons, although the notion of lexical quality will be addressed, and individual differences can be understood in the framework of the LQH, it seems difficult to discuss findings with specific address to the LQH.

In this chapter, each individual difference (ID) will be introduced in turn. This will include a description of how reading models have explained the occurrence of the ID, how effects have been measured and whether behavioural findings are supportive of each concept of ID.

4.1 Individual differences in nonword reading proficiency

4.1.1 Lexical vs sublexical readers

Skilled reading is marked by seemingly effortless word reading ability, a capacity that extends to novel items such as nonwords. Decoding skills, which are the ability to read nonwords, are a crucial prerequisite for successful reading acquisition (self-teaching hypothesis; Jorm & Share, 1983; Share, 1995), which has received a lot of support from behavioural studies (e.g. Cunningham, Perry, Stanovich, & Share, 2002; Nation, Angell, & Castles, 2007; Share, 1999). The ability to read nonwords is therefore an essential component of reading. Reading research has used nonwords to capture the ability of readers and models to read novel words.

Reported cases of acquired reading impairment have given rise to the possibility that word reading and nonword reading could be independently impaired. Cases have been reported of patients, who produced many regularization errors for irregular words, but showed normal nonword reading (McCarthy & Warrington, 1986). This reading impairment pattern has been referred to as *surface dyslexia*. The opposite pattern has been reported for so-called *phonological dyslexics* who exhibited impaired nonword reading skill, but could pronounce all other words (Funnell, 1983).

An early exploration of the impact of individual differences on skilled reading investigated if unimpaired readers could be categorised as lexical or sublexical readers (Baron & Strawson, 1976). Two subgroups were identified from a large sample of university students based on their scores on two tests. The first test asked participants to identify nonwords which sounded like words from a group of made-up words. It was assumed that this would assess the ability to read sublexically, as participants would have to apply knowledge of spelling-sound conversion rules. The second test measured spelling ability and spelling recognition ability, with the assumption that participants with better lexical knowledge would be better at these tasks than participants with less complete lexical knowledge. It was assumed that good lexical knowledge would also entail a greater reliance on lexical mechanisms relative to sublexical processes. Two groups of participants were then chosen: the *Phoenician readers*, who were inferred to be good sublexical but poor lexical readers, and the *Chinese readers*, who showed the opposite pattern. Participants were then asked to read lists of exception words and regular words. The Phoenician group was faster at reading regular words compared to irregular words, whereas the Chinese group read irregular words faster than regular ones. The authors concluded that having been able to identify groups of readers who seemed to be better at using spelling-sound conversion rules over whole-word recognition or *vice versa* showed that proficient readers varied in the balance of their reliance on reading mechanisms.

Reader profiles such as those of acquired dyslexics and Baron & Strawson's Chinese and Phoenician readers can easily be accommodated by the dual route model of reading aloud (Coltheart et al., 2001). It offers the explanation of a singular impairment or reduced efficiency of the sublexical route for those with difficulties reading nonwords, and damage to the lexical route for those who experience difficulties in irregular word reading. (Coltheart et al., 1993)

However, the trade-off between lexical versus sublexical skilled readers has been difficult to replicate. Brown, Lupker, & Colombo (1994) were not able to separate their participants into Chinese and Phoenicians readers, although they tried with two different methods. The first classification method followed Baron & Strawson by using a spelling test and a pseudohomophone detection task. Participants were then asked to complete a set of four naming tasks. The Phoenicians were expected to show faster nonword reading, and faster mixed-list word-nonword reading. The Chinese readers were expected to show faster reading of irregular word lists, as well as a reduced frequency effect. In a fourth task, Seidenberg et al. (1984) stimuli varying in regularity and frequency should see smaller regularity and frequency effects for Chinese compared to Phoenician readers. However, the results from the series of naming tests showed that although Phoenicians were better at tasks emphasising sublexical reading, Chinese readers did not show an advantage for naming tasks with irregular words. In fact, contrary to predictions, they showed a larger frequency effect and larger frequency x regularity interaction than the Phoenicians. For all naming tasks, Phoenicians emerged as the faster readers for both nonword and word stimuli. The authors therefore created a new classification scheme by also matching reaction times across the two reader groups. The results from this matched group analysis showed that the two reader groups did not exhibit any significant differences with one exception: naming RTs on the Seidenberg et al. (1984) stimuli showed that the Chinese readers again produced a larger frequency effect and a larger regularity effect, which was contrary to expectations. The authors concluded that readers do not classify as Phoenicians or Chinese readers.

It is important to draw the attention again to Brown et al. (1994)'s finding that, compared to Chinese readers, Phoenicians were not only faster at naming regular words and nonwords, but also at naming irregular words. In other words, when categorised according to Baron and Strawson's scheme, Phoenicians emerged as the faster readers of all items. Since this early investigation, research seems to converge on the observation that faster nonword reading is indeed associated with faster word reading. (Aaron et al., 1999; P. Brown et al., 1994; Davies, Arnell, Birchenough, Grimmond, & Houlson, 2017; Martin-Chang et al., 2014; Stanovich & West, 1989; Torgesen, Wagner, & Rashotte, 1999).

4.1.2 Deficient phonological representations

The ability to read nonwords is a central skill for readers, because it constitutes the capacity to read unfamiliar or novel words, which is a common task. Although readers may not fall into sublexical or lexical reader subcategories, individuals still differ in their nonword reading ability.

Difficulties in nonword reading have mostly been reported within dyslexia research (e.g., see Vellutino, Fletcher, Snowling, & Scanlon, 2004). Dyslexic beginner readers have difficulties in analysing and manipulating sounds (phonological awareness) and to acquire the alphabetic principle, which links letters to specific sounds (Anthony & Francis, 2005; Bentin, 1992; Liberman, Shankweiler, & Liberman, 1989). These difficulties can still be found in adult dyslexics (Callens, Tops, Stevens, & Brysbaert, 2014). Nonword reading is a skill related to both phonological awareness and the alphabetic principle, and thus a common measure to identify reading impairment (Rack, Snowling, & Olson, 1992, for a review see Colenbrander, Nickels, & Kohnen, 2011).

The difficulty of nonword reading is most central to phonological dyslexia, which describes the impairment of nonword reading, whilst exception word reading is largely retained (Castles & Coltheart, 1993; Manis, Seidenberg, Doi, McBride-Chang, & Peterson, 1996). Theoretical accounts of phonological dyslexia are therefore informative for understanding how variation in nonword reading may occur in skilled readers. As described above, the dual route model (Coltheart et al., 2001) foresees two separate (but simultaneous) mechanisms for reading. Exception words have to be read via the lexical route, nonwords via the sublexical route. Impaired nonword reading is due to limitations in the functioning of the sublexical route for the dual route model.

The triangle model, on the other hand, assumes that a singular mechanism is sufficient for reading words and nonwords. However, the first implementation of the direct pathway of the triangle model, the SM89 (Seidenberg & McClelland, 1989), was criticised mainly for its poor nonword reading compared to behavioural data from skilled readers (Besner, Twilley, McCann, & Seergobin, 1990). The authors had used Wickelfeatures (Wickelgren, 1969) to represent phonological codes. This method, which creates context-sensitive phoneme triplets as phonological representations, did not pick up position-independent regularities between letters and phonemes. Thus, similarities in spelling-sound correspondences across different positions of a word were not taken account of during training, and consequently generalisation from word reading to nonword reading suffered. This limitation was known as the dispersion problem. The next instantiation of the direct pathway of the triangle model (Plaut et al., 1996) was marked by improved phonological representations, which led to much improved nonword reading and thus compared better to skilled reading behaviour. Specifically, the authors introduced a system which maximally condensed spelling-sound correspondences to increase generalisation: they reduced the number of phoneme candidates by making use of the phonotactic limitations of the English language to group phonemes into sets which would then be read out from left to right. Correspondingly, sets of (single or multi-letter) graphemes were also constructed according to graphotactic rules. In this way, the regularities of spelling-sound co-occurrences had been captured in the most efficient way, so that generalization to new stimuli would be possible. Importantly, the improvement of nonword reading due to better phonological representations showed that nonword reading did not need to be handled by a separate mechanism as in the dual route model, but that words and nonwords could be read by the same direct route between graphemes and phonemes. It was thus the enhanced generalisation ability due to improved phonological representations which led to nonword naming performance akin to human data.

Harm & Seidenberg (1999) showed how the same connectionist model could also simulate the features of both phonological and surface dyslexia. Whilst phonological dyslexia is characterised by impaired nonword reading, but intact exception word reading, surface dyslexia describes the reverse symptoms, namely the ability of exception word reading, but difficulties in nonword reading (Castles & Coltheart, 1993; Manis et al., 1996). Based on previous implementations (Plaut et al., 1996), the model used a phonological attractor network to represent phonological knowledge. This network was able to reduce the dispersion problem even further by connecting phonological units between each other and to an additional layer of clean-up units, so that dependencies across features and segments could better be appropriated by the network. This architecture enabled the model to further improve generalisations from known items to new items such as nonwords, and to repair or complete noisy or degraded input to fall into its most suitable output pattern. To simulate phonological dyslexia, Harm & Seidenberg (1999) implemented a mild and a strong form of impairment of the model's phonological representations. In the mild condition, they reduced the weights in the phonological attractor, and thereby weakened the phonological attractors. In the strong condition, they additionally removed the clean-up units and halved the number of connections between phonological units. The simulations with these two impaired networks showed that the mild modification produced a 'pure' phonological dyslexic profile with reduced nonword reading ability but retained exception word reading. The stronger modification led to a further deterioration in nonword reading together with an additional reduction in exception word reading. The results showed that phonological dyslexia could be simulated with a connectionist network, and that the degree to which the phonological network had been disabled correlated with the intensity of the deterioration in reading performance, consistent with the view that phonology-focused reading impairments can vary on a continuum (see also Crisp & Lambon Ralph, 2006)

Although the results from the computational implementation of the connectionist approach (Harm & Seidenberg, 1999) were quite specific in identifying non-optimal phonological representations as a possible cause for decoding difficulties, it seems prudent to understand weaker decoding skills within the broader concept of a less efficient direct orthography to phonology pathway. This more general approach is useful, as for the purpose of the present investigation it conceptually and operationally separates the orthography – phonology (O-P) mapping from the orthography – semantics – phonology (O-S-P) mapping and links well to other pertinent studies which will be addressed below (Strain & Herdman, 1999; Woollams, Lambon Ralph, Madrid, & Patterson, 2016).

4.1.3 Nonword reading ability varies with orthography

In languages with transparent orthographies, readers diagnosed with phonological dyslexia have been observed to present different symptom patterns compared to readers of opaque scripts (see Goswami, 2002). English-speaking children diagnosed with dyslexia have been found to perform the nonword reading task at a lower performance level than typically developing readers (Rack et al., 1992). In contrast, German-speaking dyslexic children showed relatively high accuracy in both word and nonword reading. For these children, the greatest difficulties were encountered in reading speed and spelling accuracy (Landerl, 2002). In a cross-language study comparing reading of English and German children, both language groups made more errors reading nonwords than words. Yet, English dyslexic children made many more errors and were slower in reading low frequency words and nonwords compared to their German counterparts. However, both groups showed comparable deficits in phonological skills, as assessed by a spoonerism task. It seemed, therefore, that the underlying deficit was the same, but that the symptoms for reading impairment presented different patterns in the two orthographies (Landerl, Wimmer, & Frith, 1997).

In typically developing readers, the differences in nonword reading in transparent and opaque scripts are also evident. A comparison between German and English young readers showed that English children made more errors in nonword and low frequency word reading. At age eight, word and nonword naming latencies were longer for the English-speaking cohort. Equally, for nonword compared to word reading, and with increasing syllable length, English children showed significantly longer latencies compared to German children. By age 12, naming latency differences between the two groups were no longer observed. However, the English 12-year olds still made more errors when reading nonwords. (Frith et al., 1998)

For adult skilled readers, a comparison of naming latencies in words and nonword naming found that German readers were slower at naming all items compared to English readers, although this difference was not significant. In both languages, words were read faster than nonwords. There was no language x lexicality interaction, suggesting that language groups did not differ in their advantage to name words over nonwords. Error rates for both language groups were comparable, with English readers producing slightly fewer errors than German readers (E: 1.8% and G: 3.7%). (Ziegler et al., 2001)

In contrast, in their seminal study, Frost et al. (1987) found that naming latencies for nonwords decreased with increasing orthographic transparency. Readers of transparent Serbo-Croatian were faster at nonword naming than English readers, and even faster than readers of unpointed Hebrew. Unpointed Hebrew lacks information about vowels and is therefore considered a very opaque script. However, the authors reported more errors for Serbo-Croatian than English readers for nonword reading (E: 2.6%, S-C: 5%, H: 7.5%), which is comparable to the Ziegler et al. (2001) study. For all three languages, words were read faster than nonwords.

Altogether these results suggest that developing readers of English seem to make more errors (especially in nonword reading) and have longer naming latencies than readers of more transparent orthographies. By the time that readers have matured, this pattern is no longer reliably observable. In fact, fewer errors are reported for readers of English than of a more transparent language. Furthermore, speed comparisons in nonword naming yield conflicting results. Faster nonword naming has been reported for English compared to German readers, but faster nonword naming has also been reported for Serbo-Croatian readers over English readers.

4.1.4 Deficit in phonological skills may lead to a different division of labour

As mentioned previously, differences in nonword reading ability are prevalent within language groups of unimpaired skilled readers (e.g., Stanovich & West, 1989; Torgesen et al., 1999). It has been suggested that this variation in phonological skills within skilled reader groups leads to an adjusted division of labour between the two pathways within the triangle model, the direct orthography-to-phonology and the indirect via semantics pathway. Readers who are found to be less efficient in processing in the direct pathway (as evidenced by slower nonword reading and other phonological processing tasks) make more use of other information available to arrive at a pronunciation, such as semantic information. Thus, the functioning of the direct pathway, as indexed by nonword decoding, may determine the use of semantics in reading aloud. Difficulties in using the direct pathway could result in increased employment of the indirect pathway via semantics. (Strain & Herdman, 1999; Woollams et al., 2016)

4.1.5 Section summary

In summary, nonword decoding is a basic skill, which has been identified as key for reading acquisition. Although there has been no unequivocal evidence that readers may fall into decoder (sublexical) readers and whole-word (lexical) readers, research findings suggest that skilled readers do differ in the degree of nonword reading fluency. It seems that slower nonword reading is also related to slower word reading. This would seem to be consistent with a single route model of the reading system as suggested by the PDP approach.

Reading models give different explanations of nonword reading variation. Accounts of phonological dyslexia give insight into how models explain nonword reading. From a dual route model perspective, poor decoding skills result from deficient functioning of the sublexical route. In contrast, triangle model simulations point to impaired phonological representations, which lead to inefficient functioning of the O-P pathway.

Across different languages, phonological dyslexia exhibits different reading patterns. In English, dyslexia is marked by slow and error-prone nonword reading. In orthographies like German, which have a more transparent spelling-sound relationship, reading performance of phonological dyslexics is more similar to that of normal readers. German dyslexic children generally show accurate nonword reading, but are slower at reading both words and nonwords compared to unimpaired German peers.

For skilled readers, it is less clear how spelling-sound transparency impacts on nonword reading latencies. Studies have reported both comparatively faster and slower nonword naming for English compared to more transparent languages. Generally, however, it seems that across languages nonwords are read slower than words.

Importantly, skilled readers show variation in decoding skill within their own language group, and it therefore remains an important individual difference to consider. Within the context of the triangle model, it has been suggested that difficulties in phonological skills may prompt an increased usage of the indirect semantic pathway for oral reading, as the direct pathway seems to be less efficient.

In the present investigation, nonword decoding will be considered as assessing O-P pathway efficiency.

4.2 Individual differences in semantics

As mentioned in the previous section, it has been suggested that deficient phonological processing may lead to increased use of semantics in reading (see Strain & Herdman, 1999; Woollams et al., 2016). The present section will give a brief overview of studies exploring the use of semantics in reading aloud. This will be followed by a description of the division of labour hypothesis within connectionist modelling, which proposes that people vary in their use of the phonological and the semantic pathways (Plaut et al., 1996). The present study will explore individual differences in semantic knowledge. Greater semantic knowledge may support increased use of semantics, or alternatively, may make semantic effects more detectable.

4.2.1 Are semantics used in reading aloud?

The use of semantics in reading aloud has been extensively investigated. Effects of variables associated with semantic information are often observed in lexical decision, and such effects may be observed more prominently in lexical decision than naming. (Balota et al., 2004; Chumbley & Balota, 1984)

Imageability has been used to measure semantic involvement in naming tasks. Ratings of imageability are measures of the capacity to evoke the image of a word referent independent of the task. Strain, Patterson, & Seidenberg (1995) argued that more imageable words have more stable semantic representations, possibly due to a higher number of defining features, lesser dependency on context for understanding, or greater anchorage in sensory-motor memory, and that imageability was therefore an adequate indicator for semantic involvement in reading aloud. They reported an advantage for more imageable words when naming exception words of low frequency. They did not find this advantage with regular words of low frequency. Their result suggested that for words, which are more difficult to decode phonologically, the slower processing gives semantics more time to influence the naming process. However, the imageability effect in reading aloud has not been an unequivocal finding. Although some reading aloud studies have since reported imageability effects in reading aloud (Balota et al., 2004; Cortese & Schock, 2013), others have not (Cortese, Yates, Schock, & Vilks, 2018; J. Monaghan & Ellis, 2002).

More supporting evidence for semantic involvement in reading aloud has been reported with the use of other semantic proxies. For example, Hino & Lupker (1996) explored the effects of the number of meanings per word. They found that low-frequency words with several meanings were read faster than low-frequency words with fewer meanings. This facilitatory effect was attributed to increased semantic feedback from items with many meanings which then accelerated the translation into a phonological code. Another paradigm exploited the activation of synonyms as a marker for semantic engagement to show that meaning is activated during reading aloud. Words without synonyms were read faster than those with several synonyms, suggesting that the activation of other words with the same meaning fed back to the processing loop and slowed down the build-up of the phonological code (Hino, Lupker, & Pexman, 2002; Pecher, 2001). Finally, in a novel word paradigm, inconsistent nonwords which were attributed a meaning were named faster than inconsistent nonwords without a meaning (McKay, Davis, Savage, & Castles, 2008).

Thus, it seems that meaning is inevitably activated during word processing, even in naming, although this is more easily detectable in tasks which benefit from access to word meaning, such as lexical decision (Balota et al., 2004; Yap & Balota, 2009).

4.2.2 The triangle model and the division of labour

With regard to reading models, the influence of semantics has been investigated within the PDP approach. The triangular structure of the PDP model of reading aloud allows for pattern activation to pass from orthography to phonology, and also to semantics. Plaut et al. (1996) investigated pronunciation production of words with the contribution of the semantic pathway by training a feedforward network of the direct pathway with additional input from semantics. They observed that – given the quasi-regular nature of the spelling-sound mappings in English - the direct orthography – phonology (O-P) pathway initially contributed most to naming in a developing reading system. With increasing reading experience, the semantic pathway developed and contributed increasingly more, especially to reading exception words. Eventually, the O-P pathway most prominently computed very consistent items, whilst the semantic pathway read inconsistent words, thereby creating a redistribution of labour between the pathways. Although the specialisation of the O-P pathway to reading consistent words is reminiscent of the dual route approach, Plaut et al. (1996) stressed that - in contrast to the dual route model - in their simulation the O-P pathway still contributed to reading some exception words, mostly to those of high frequency.

In the first full implementation of the triangle model, Harm & Seidenberg (2004) aimed at simulating reading for meaning. The model used both pathways O-S and O-P-S simultaneously and cooperatively to compute meaning. The authors suggested that the dual pathway activation to compute meaning from print was also valid for producing sound from print, but that the division of labour between the two pathways may not be as complementary as for the two pathways used to compute meaning. Harm & Seidenberg (2004) pointed out that when computing meaning from print, the O-S pathway was initially slower to learn. However, when it was trained, it provided the faster way to access meaning. In contrast, the O-P-S pathway contributed more at the beginning of training and less in a more skilled model. As the phonological representation has to be stable before the semantic representation can be activated, it eventually became the slower route. Harm & Seidenberg (2004) stressed that there was a decisive difference when looking at naming. In naming, the direct O-P route was the more consistent route, as it involved quasi-regular spelling to sound conversion. The long route via semantics was in fact the more arbitrary route, as mappings between orthography and semantics, and again between semantics and phonology are more or less arbitrary. They therefore suggested that in the case of reading aloud, the help from the long route via semantics may only be used in very special cases, for example, when the spelling – sound mappings were very inconsistent. Thus, in contrast to the division of labour hypothesis as formulated in Plaut et al. (1996), which foresaw a separation of the routes not unlike the dual route model, the division of labour as conceptualised in Harm & Seidenberg (2004) was now considered to have a stronger O-P route and a less dominant assisting semantic route, which would be available for reading more difficult (e.g. inconsistent) words.

An important contribution to understanding how the division of labour could vary between individuals is the study of Dilkina, McClelland, & Plaut (2008), who simulated semantic dementia patients' data with a computational PDP model. Semantic dementia is a form of cerebral atrophy, which is characterised by progressive loss of word meaning in both the receptive and productive vocabulary of patients (Hodges, Patterson, Oxbury, & Funnell, 1992; Hodges & Patterson, 2007; Snowden, Goulding, & Neary, 1989). Although most of semantic dementia patients also develop a reading impairment at some stage as their disease progresses, there have been reports of unimpaired reading skills, leading to the assumption that the reading system is separate from the semantic system. (Coltheart, 2004) In contrast, Dilkina and colleagues proposed that a single computational reading system would be able to account for the observed differences between patients. They suggested that the seemingly different cases of impairment were in fact the result of several factors which varied within each individual: individual variation in direct pathway strength, reading experience, and the individual patients' brain damage due to atrophy. Although their simulations showed that simulated brain damage was the most important factor to account for inter-individual variation, importantly, both patients' reading experience before the disease onset and the relative greater strength of the direct O-P pathway further contributed to model success in simulating patients' retention of the ability to read aloud after disease onset. The results of the study implied that the reading system was less reliant on semantics when it had had more practice (reading experience) and a stronger O-P pathway.

The findings by Dilkina and colleagues will be discussed again in the next chapter on print exposure. As the focus of the present chapter is the involvement of semantics in reading, it is important to acknowledge that the work of Dilkina and colleagues implemented a computational model which simulated patient data by manipulating the division of labour between a phonological and a semantic pathway, demonstrating that this division of labour could be affected differently in patients varying on person-level dimensions, notably including the strength of the phonological pathway.

4.2.3 Varied contribution between the two pathways?

Even within PDP modelling approaches, the proportional contribution of semantics in reading aloud is still a matter of debate. Whilst an instantiation of the triangle model with semantic contribution showed a shift from direct to indirect route contribution with increasing training (see simulation 4; Plaut et al., 1996), other researchers foresee semantic contribution primarily in cases where the item to read is difficult, for example, inconsistent (Harm & Seidenberg, 2004). Hoffman, Lambon Ralph, & Woollams (2015) presented an fMRI study which demonstrated how participants varied in terms of their use of semantics in reading aloud. As the triangle model predicts that the semantic route contributes most for reading inconsistent words, the authors reasoned that a greater consistency effect for an individual would indicate greater reliance on the semantic route. They therefore grouped participants into high semantic reliance (SR) readers and low SR readers. For this, participants were asked to read both consistent and inconsistent low-imageable words, and the difference in reading performance (consistency effect) was computed per participant. A larger difference between reading of consistent vs inconsistent words of low imageability would be defined as a larger consistency effect. It was assumed that readers with a larger consistency effect for low imageable items would use the semantics pathway relatively more often (high SR reader). In contrast, readers, who would show less differences between reading consistent and inconsistent items when reading low imageable words, would be more prone to use the direct pathway (low SR reader). During an fMRI scan, participants were then asked to complete a reading aloud task with stimuli varying in frequency and regularity. Results showed activity in an area of the anterior temporal cortex which is linked to semantic processing. Greater activation of the area was measured when participants had been identified as relying more heavily on semantics for reading low-imageable exception words. Conversely, participants who had been identified as relying less on semantics showed greater activation of the premotor cortex associated with phonological processing. As such, it seemed that some readers rely more on semantics, whereas others may rely more on the phonological pathway.

However, Strain & Herdman (1999) found that semantics usage was apparent when phonological reading ability was less efficient. In this study, imageability was used as a semantic referent. The expectation was that participants with less efficient O-P mappings would resort to using semantic information to produce the correct pronunciation. Furthermore, for participants with strong phonological reading skill, the imageability effect would only show in LF exception words. In contrast, for participants with weaker phonological reading skill, the imageability effect would occur for all LF words, both regular and exception. Participants were grouped into three skill groups based on their combined performance in nonword reading and sound blending. The authors found an imageability effect for all participants across all ability groups for low frequency exception words compared to low frequency regular words. Importantly, this result was modulated by individual differences in phonological recoding ability. For strong decoders, the imageability effect was stronger for exception words than for regular words, for medium decoders this interaction was not significant, but *post hoc* examinations showed that this group showed larger imageability effects for exception words than for regular words. For weak decoders, the authors did find an imageability main effect, but no regularity effect or an interaction between the two main effects. The authors concluded that for this last group, O-P mappings for all words regardless of their regularity were affected by semantics, which was in line with the prediction. A group comparison showed that the less participants were skilled in phonological recoding, the more they relied on semantics (IMG effect), and the less they relied on phonological recoding (regularity effect). The results presented by Strain & Herdman (1999) supported the assumption that participants resorted to the use of semantics when the ability to phonologically recode was not optimal. Highly skilled decoders used semantics more for low frequency exception words than for regular low frequency words, whereas participants with decreasing skill tended to use semantics to improve reading of all low frequency words, regardless of their regularity status.

Woollams et al. (2016) presented further evidence that greater reliance on semantics in reading aloud was due to insufficient functioning of the phonological route. As in Hoffman et al. (2015), participants were grouped according to their reliance on semantics when reading words with low imageability: Readers were considered high SR readers when they showed a relatively stronger consistency effect in reading low-imageable words. If participants were either direct route or semantic route readers, as is suggested by the division of labour hypothesis (Plaut et al., 1996), then it would be expected that low SR readers would be more proficient at using the direct route than high SR readers. Contrary to the authors' expectations, they found that readers with a higher reliance on semantics for inconsistent words were also slower at nonword reading. Moreover, the semantic reliance measure was not associated with other semantic processing tasks, thus suggesting that the reliance on semantics in reading aloud low imageable words was not due to general semantic processing capacity, but rather that semantic reliance was associated with difficulties in phonological processing tasks. Thus, Woollams et al. (2016) suggested that the semantic reliance in reading aloud is driven by a deficit in phonological processing rather than a better performance in semantic processing tasks, and that semantic readers in fact compensate for phonological processing deficits.

Combined it seems that whilst people may vary in their respective use of the pathways, increased contribution of the O-S-P pathway seems to be compensatory, rather than complementary. The results by Woollams et al. (2016) qualify the division of labour hypothesis in so far as they seem to suggest that although the O-P and the O-S-P pathways both contribute to reading aloud, inefficient processing by the O-P pathway cannot be fully compensated by the O-S-P pathway. This is reminiscent of studies which have reported that poor nonword reading (as an index of less efficient O-P mapping) is related to poor word reading (e.g. Aaron et al., 1999; P. Brown et al., 1994; Davies et al., 2017; Martin-Chang et al., 2014; Stanovich & West, 1989; Torgesen et al., 1999), and identified the O-P pathway as the crucial pathway for skilled reading. Finally, these results provide behavioural evidence that individual differences in phonological decoding shape the reading system towards greater semantic reliance.

This is an important consideration for the present study. Harm & Seidenberg (2004) noted that the division of labour between the two pathways will vary with orthographies, but also with other factors such as stress, intonation and morphology. The present study will be the first to attempt a direct comparison between two languages in this respect by including a psycholinguistic variable (age-of-acquisition) to index semantic involvement in reading aloud. In this way, it is hoped that the study will contribute to our understanding how semantic contribution varies between readers of two languages differing in spelling-sound consistency who also have different sets of reading skills.

4.2.4 Age-of-acquisition as a semantic marker

As seen in previous sections, behavioural studies have employed a number of markers for semantic contribution, such as the difference in reading performance of reading consistent and inconsistent low-imageable words (Woollams et al., 2016), imageability ratings (Strain & Herdman, 1999; Strain et al., 1995), and number of word meanings (Hino & Lupker, 1996). There is some evidence which suggests that age-of-acquisition effects may be a marker for (some) semantic involvement. Early acquired words are named faster than late acquired words (G. D. Brown & Watson, 1987; Cortese & Khanna, 2007; Cortese &

Schock, 2013). This effect seems independent of frequency (G. D. Brown & Watson, 1987; Cortese & Khanna, 2007), even though the two variables are highly correlated (Brysbaert & Ghyselinck, 2006; Zevin & Seidenberg, 2002).

The effect of AoA has been found to be larger in size and more clearly distinct from frequency effects in tasks that require a mapping between language and semantics in English (Cortese & Khanna, 2007; Cortese & Schock, 2013), and in the more transparent Spanish (Cuetos & Barbón, 2006; Davies, Wilson, Cuetos, & Burani, 2014). The semantic locus hypothesis (Brysbaert, Wijnendaele, & Deyne, 2000; van Loon - Vervoorn, 1989) understands AoA effects as a reflection of the system according to which words' meanings are organised in the semantic system. Steyvers & Tenenbaum (2005) have presented a model of semantic growth which attributes to early acquired words central organising roles. These words have usually acquired more connections to other words and are therefore processed at a faster rate. According to this account, AoA effects are due to words' meanings being better connected, and therefore reflect the underlying semantic network.

However, a strong competitor to this proposition is the hypothesis that AoA reflects the way that the reading system has developed. Within the framework of the PDP model, the network plasticity hypothesis (A. W. Ellis & Lambon Ralph, 2000) proposes that the age at which words are acquired shapes the reading system. As the system learns more and more words, knowledge about words is encoded in connection weights. Whilst the system is still malleable at the beginning, its ability to adapt with later acquired words shrinks. The system becomes less plastic. Later acquired words therefore find it harder to shape the reading system, and to be securely encoded. Support for this theory comes from Ellis & Lambon Ralph's (2000) series of simulations, where a connectionist system was presented with patterns for learning in a cumulative and interleaved fashion. With increasing exposure of input patterns, the system adapted by strengthening connections of earlier presented patterns, and consequently these items shaped the network in their favour. The simulations showed that frequency of item presentation, and crucially, also the order of item presentation led to later-acquired items to be learnt less securely than earlier acquired items. P. Monaghan &

Ellis (2010) presented further support with model simulations which mimicked children's word learning. The model was presented with items in an incremental and cumulative fashion. The authors found that early presented words had been shaping the system more reliably than late acquired words, as the reading system gradually lost its plasticity. Consequently, early items were learned more robustly than late items. However, neither the model of Ellis & Lambon Ralph (2000) nor that of P. Monaghan and Ellis (2010) simulated the complete triangle model with semantics. Therefore, their simulations do not contradict the assumption that AoA effects may at least in part reflect semantic connectivity and involvement.

In summary, AoA seems to be an order effect of the way spelling-sound mappings were mapped out. It also seems that AoA effects are more pronounced when words are spelling-sound inconsistent. Yet, as AoA effects seem stronger in tasks that require access to meaning, the semantic locus hypothesis purports that AoA effects reflect the underlying semantic connectivity. The latter would mean that AoA can be recognised as a marker for semantic involvement in reading.

4.2.5 Readers' semantic knowledge

In the previous sections, it has been outlined how and when semantics may be involved in reading aloud. Importantly for the present study, however, it also needs to be established whether differences between individuals in semantic knowledge also impact on reading performance.

In their study, Woollams et al. (2016) reported that greater semantic reliance was only associated with less efficient phonological processing, but was not associated with tasks which tapped into greater semantic knowledge. These tasks were to judge if word pairs were synonyms, and to enumerate member items of a category. This led Woollams and colleagues to the conclusion that the division of labour was driven uniquely by inefficient O-P processing.

However, there is nevertheless reason to assume that readers' semantic knowledge may still influence the reading process. The lexical quality hypothesis (Perfetti & Hart, 2002; Perfetti, 2007) proposes that readers differ in word-specific quality of the three lexical components orthography, phonology and semantics and that inter-individual variation in word recognition and comprehension stems from the individual strength and weaknesses of these three word constituents combined with their reading experience. A less well-defined semantic representation may then slow down reading performance.

Moreover, there are a number of studies which document the impact of better semantic knowledge on word recognition. Semantic knowledge is often tested via vocabulary tests. Pexman & Yap (2018) found that participants with higher vocabulary scores were faster at making semantic decisions compared to participants with lower vocabulary scores, suggesting that they had better access to semantic information. Andrews & Lo (2013) reported stronger priming for morphologically transparent primes for participants with a 'semantic profile' than participants with an 'orthographic profile'. Furthermore, individuals with greater vocabulary knowledge showed more efficient lexical decision and naming performance (Mainz, Shao, Brysbaert, & Meyer, 2017; Yap, Balota, Sibley, & Ratcliff, 2012).

Individual differences in vocabulary knowledge may therefore tap into more readily available, or more complete semantic information. It seems prudent then to acknowledge this semantic advantage in individuals with greater vocabulary knowledge when exploring individual differences in reading systems in the current study. Of course, it needs to said that - as will be discussed in the next section - any effects of vocabulary knowledge could also be considered as a product of reading experience, rather than semantic effects, or both. Vocabulary knowledge has been used as a proxy for reading experience (Yap et al., 2012) and word learning has been shown to increase with print exposure (e.g. Nation et al., 2007). By controlling for print exposure in the present study, any effects arising from vocabulary knowledge are more likely to be due to semantic knowledge. Individual differences in print exposure will be discussed in the next section.

4.2.6 Section summary

The division of labour hypothesis holds that with increasing reading experience, the reading system undergoes a shift in contributions of the phonological and the semantic pathways to reading aloud. Whilst early investigations surmised a specialisation of the phonological route for consistent and of the semantic route for inconsistent items (Plaut et al., 1996), more recent theoretical contemplations view the direct O-P route as the main reading route with additional support from the O-S-P route for difficult (e.g. inconsistent) items (Harm & Seidenberg, 2004). In accordance with this, a simulation study found that the division of labour between O-P and O-S-P pathways may be influenced by the individual's strength of the direct pathway, reading experience and accidental damage to the semantic part of the reading system (Dilkina et al., 2008). Moreover, evidence from behavioural studies also suggested that the increased use of the indirect pathway may be due to insufficient functioning of the direct pathway (Strain & Herdman, 1999; Woollams et al., 2016). This suggestion is congruent with the repeated reporting of a correlation between poor nonword and word reading, which identifies the direct pathway as the crucial route for skilled reading (e.g., P. Brown et al., 1994). The division of labour may be different in languages other than English (Seidenberg, 1992), and the present study will be well-placed to address semantic effects in reading aloud in different orthographies whilst taking into account individual differences.

In the literature, semantic involvement has typically been measured using imageability ratings, or other semantic referents, such as number of word meanings. There is some reason to assume that age-of-acquisition effects may partly reflect semantic involvement, as the meanings of early learned words may be anchored more robustly in the reading system (Brysbaert et al., 2000; van Loon - Vervoorn, 1989). In the absence of other measures, it seems justifiable to consider AoA a referent for semantic involvement.

Yet, individuals also differ in terms of their semantic knowledge, and it therefore seems useful to take into account semantic knowledge as an individual difference (e.g. Andrews & Lo, 2013). This is often measured by vocabulary tests. Individuals with greater vocabulary scores may have greater, or faster access to semantic knowledge which may assist the reading process, and show a semantic effect.

4.3 Individual differences in print exposure

People differ from each other in the amount of reading experience. Due to their ability to learn and adapt, connectionist models are able to take into account experience as a contributing factor to individual differences in reading. As mentioned in the previous section (see section 4.2.2), connectionist model simulations of semantic dementia patient data showed that the reading system of patients with more reading experience was more robust to the decline in reading performance than for patients with less reading experience (Dilkina et al., 2008). Furthermore, there is ample behavioural evidence from skilled readers that more reading experience leads to lasting and performance enhancing effects in reading tasks, such as reading aloud, but also in vocabulary knowledge, spelling ability and general knowledge (e.g., Mol & Bus, 2011; Sears, Siakaluk, Chow, & Buchanan, 2008; Stanovich & West, 1989). Thus, print exposure has surfaced as an important individual difference in reading.

This section will briefly review the extent of individual differences in reading experience and which effects have been reported in this regard. This will be followed by a short introduction of measures used to capture reading experience. This chapter will then show how print exposure effects have been convincingly modelled with connectionist networks. Special attention will be paid to the word frequency effect, as this robust psycholinguistic effect has been found to reduce in participants with greater print exposure.

4.3.1 How do people vary in the amount they read?

Time spent reading varies considerably between people. In a questionnaire, adolescents in the UK aged between 11 and 16 years were reported to have read on average between 30 min and 1h over the previous weekend (McGeown, Duncan, Griffiths, & Stothard, 2014). Diary surveys with adult participants reported an average reading time of

2h 48 minutes per day (Smith, 2000), and one study even found an average 4.5h a day spent on reading related activities (White, Chen, & Forsyth, 2010). It has been estimated that American middle school children read between 100,000 to 10,000,000 words a year (Nagy & Anderson, 1984). It seems intuitively right then, that print exposure should have a measurable impact on reading skill, if reading skill can be assumed to continue to develop in association with ongoing reading practice.

4.3.2 Effects of print exposure measures on reading aloud

Print exposure has been associated with faster word naming in children (Spinelli et al., 2005), and predicted naming of so-called strange and exception words (Stanovich & West, 1989) and nonwords (Chateau & Jared, 2000) in skilled adult readers. In fact, there is evidence that participants with more print exposure are generally faster at naming words of all frequencies than participants who are less familiar with words (Lewellen, Goldinger, Pisoni, & Greene, 1993). Faster naming speed is not the only observable effect of print exposure. Readers with more reading experience become so proficient at the reading process that words are processed very similarly, regardless of their properties. In a large-scale analysis of speeded pronunciation data of 1,289 participants, Yap et al. (2011) showed that more experienced readers were less influenced by lexical characteristics than less experienced readers. They found that vocabulary knowledge (as an index for print exposure) was negatively correlated to structural word characteristics (e.g. length, number of syllables), both phonological and orthographic neighbourhood size as well as word frequency and semantic properties. In other words, the more participants had read, the less lexical characteristics influenced the reading process.

4.3.3 Measures of print exposure

Research has drawn on a variety of print exposure measures. As mentioned above, vocabulary knowledge has been used as a proxy measure by some (Yap et al., 2012). Research has shown that vocabulary knowledge measurably increased with growing reading experience (Burt & Fury, 2000; Stanovich & Cunningham, 1992). Other researchers have assumed that older participants have read more in their life time than younger participants, therefore taking age as an indicator for reading experience (Davies et al., 2017). A further approach to index reading experience has been to combine several measures to yield a measure of how familiar participants were with words. For example, Lewellen et al. (1993) determined participants' familiarity with words by the performance on a vocabulary test, a language experience questionnaire and individuals' word familiarity ratings.

More direct measures have also been devised, such as reading questionnaires and the author recognition test (ART; Stanovich & West, 1989). Whilst self-reported measures such as questionnaires have been criticised for potentially being subject to self - serving bias, the ART is considered a more objective print exposure measure, as participants have to identify real authors from a long list of names which includes as many false as genuine author names. Any false identification will reduce the points gained for correct author recognitions. These measures will be discussed further in Chapter 7.

4.3.4 Print exposure in reading models

Due to the nature of its network architecture, the PDP model lends itself well to examining the effect of print exposure on reading. Within the PDP reading framework, reading experience translates into increased training of the reading system. Thus, in reading aloud, the recurring presentation of stimuli strengthens the weights on connections between orthographic and phoneme layers in ways that are helpful to producing the presented stimuli.

Through learning algorithms, the match or mismatch between the target output activation pattern and the activation pattern at the phonological level is used to adjust connection weights so as to reduce the discrepancy between the phoneme activity and the desired output activity. The frequency with which items are presented therefore determines the rate of reduction in distance between the target word pattern and the actual output pattern.

4.3.5 Modelling individual differences in print exposure

As mentioned previously (see section 4.2.2), the impact of increased network training on the network's functioning and organisation has been demonstrated in network simulations of semantic dementia patients. Dilkina et al. (2008) showed how variable reading experience contributed to explaining different reading performance patterns in semantic dementia patients. Although for many semantic dementia patients the deterioration of conceptual knowledge is also associated with an impairment to read, some have retained unimpaired reading ability whilst exhibiting semantic deficits (for references please refer to Dilkina et al., 2008). The authors ran several model simulations which had been manipulated on three variables: lesion bias, direct pathway strength and prior network training. Whilst lesion bias emerged as the strongest predictor for a reading deficit, this was followed by pre-morbid network training. More prior training seemed to reduce reading performance decline after lesioning the system. Thus, these modelling results suggested that patients with more reading experience retained the ability to read longer than others, as the reading experience prior to disease onset had made their reading system more robust.

It thus seems that continued reading experience renders the reading system robust to damage and makes it very efficient. One marker for reading efficiency seems to be the reduction of the word frequency effect. More frequent words are named faster than less frequent words (Schilling, Rayner, & Chumbley, 1998), and word frequency accounts for more variance in naming than any other variable (Balota et al., 2004; Brysbaert et al., 2011;

Yap & Balota, 2009). It has also been reported for a number of other alphabetic languages, such as German (Schröter & Schroeder, 2017), Italian (Burani, Arduino, & Barca, 2007), and Spanish (Davies et al., 2014). Crucially, Yap et al. (2012) showed that reading experience (as indexed by vocabulary knowledge) led to smaller frequency effects for individuals with higher vocabulary scores. In other words, more reading experience reduced the word frequency effect. This is an important finding, as the word frequency effect is a robust effect in word naming.

The decrease of the word frequency effect with increased reader experience was further investigated by P. Monaghan, Chang, Welbourne, & Brysbaert (2017), who trained a connectionist triangle model to simulate the changes on the word frequency effect with increasing print exposure on naming performance. They found a U-shaped curve, with small word frequency effects for little print exposure, larger frequency effects with increasing training, which was then followed by a decrease when the system had reached a certain exposure saturation point. The authors also teased apart whether the reduction of the frequency effect was due to print exposure, or whether it was due to the size of the vocabulary that the model had acquired with training. The results showed that for all models differing in vocabulary size, the frequency effect decreased with extended training. However, larger vocabularies meant that models struggled for longer to form efficient mappings between orthography and phonology. Word frequency effects were therefore apparent for longer in the larger-vocabulary models, but eventually, when learning had occurred sufficiently, the word frequency effect started to reduce as well. Thus, models with larger vocabularies showed a word frequency effect for longer. Consequently, the reduction in the word frequency effect was due to print exposure, but larger vocabulary sizes meant that it took longer for the word frequency effect to be reduced.

4.3.6 Section summary

Behavioural evidence suggests that individual differences in print exposure lead to observed differences in reading aloud (Chateau & Jared, 2000; Stanovich & West, 1989). This manifests itself as shorter naming latencies and reduced psycholinguistic effects, such as word frequency (Yap et al., 2012). Within the framework of the PDP model, these observations are due to increased strengthening of the relevant activation patterns, making the reading process more efficient. The efficiency is evidenced by the reduction of the robust word frequency effect due to increased training of the reading system. The reduction of the word frequency effect is delayed when larger vocabularies were acquired by the reading system. (P. Monaghan et al., 2017)

The present study will investigate the impact of individual differences in print exposure across two orthographies. The expectation is that higher print exposure should lead to faster naming performance, and it should lead to a reduction of the word frequency effect, as the reading system should have become highly efficient. This should be observable across both languages, independent of their orthographic transparency.

4.4 Individual differences in orthographic representation quality

The current project also addresses a last individual difference, the quality of orthographic representations. The lexical quality hypothesis (LQH; Perfetti & Hart, 2002; Perfetti, 2007) proposes that lexical representations have orthographic, phonological and semantic knowledge components, which can vary in quality on a word-by-word basis. Importantly for the current study, high quality for orthographic word knowledge is manifested by fully specified letter knowledge (Perfetti, 2007).

Within the framework of the LQH, spelling has been considered an appropriate index for precise orthographic knowledge. This is compatible with the suggestion that poor spelling may be due to incomplete analysis of word stimuli, which leads to poor encoding and storage of lexical information (Frith, 1980, 1985; Holmes & Ng, 1993). The concept of low-quality orthographic representations can also be accommodated with the view that some readers may have difficulties in extracting spelling regularities (Fischer, Shankweiler, & Liberman, 1985). Orthographic knowledge may to a large degree be acquired through statistical learning (Deacon, Conrad, & Pacton, 2008; Treiman & Kessler, 2006). Every language has certain spelling regularities, which means that some letter sequences are admissible and others inadmissible (letters which never occur next to each other in a word). The implicit learning of these statistical probabilities is often referred to as orthographic redundancy (Andrews, 1992; Chetail, 2015), and is central to network learning as instantiated by the PDP approach. Importantly for this study, whatever the underlying cause for spelling differences between individuals may be, the LQH provides a vehicle to explore how variant orthographic lexical quality impacts on reading performance.

The assumption of the LQH is that differential lexical quality is the root for individual differences in reading, and specifically reading comprehension. Individuals differ in the way that their representations have developed into precise, redundant and meaningful constructs. Words which have very precise orthographic information will be spelled and read faster than words with lower quality information. In support of this, Martin-Chang et al.

(2014) reported that students who were better spellers also tended to be faster readers. This was true on a word-by-word basis, i.e. they found a relationship between a word that a participant found hard to spell and their reading speed. Although the LQH focusses on explaining text comprehension, the concept of lexical quality has been found to be helpful in providing a framework for individual differences in other reading tasks (Andrews & Bond, 2009; Andrews & Hersch, 2010; Andrews & Lo, 2013; Andrews, 2012; Hersch & Andrews, 2012). In the present study, the LQH will be used as the framework to explore whether and if so, how, individuals' variable quality of orthographic representations will affect naming in two different languages.

4.4.1 Do reading and spelling share orthographic representations?

Studies have used spelling as an index for the quality of stored orthographic information (e.g., Andrews & Bond, 2009; Martin-Chang et al., 2014) under the assumption that orthographic reading and spelling representations overlap (Jones & Rawson, 2016; Monsell, 1987) or are even the same.

The assumption that the word form that is learned through spelling is also used for reading has been supported by a recent meta-analysis which found that overall spelling instruction improves children's word reading skill (Graham & Santangelo, 2014). Investigating the cross-modality effect in skilled readers, Burt & Tate (2002) found that words which had been misspelled were also recognised slower and less accurately in a lexical decision task, thereby suggesting that the same orthographic information was activated for both tasks. Further evidence for a shared orthographic representation has been provided through a spelling recognition task. For example, in Holmes & Carruthers (1998), students were able to discriminate between misspellings and correct spellings, which they had previously produced correctly themselves. However, when they were not able to correctly spell the word, they could not discriminate between their own misspelling and the correct word. If they had had two separate orthographic representations for each word, it would have been expected that they would have picked their own misspelling more often than the other misspelling.

However, it needs to be noted that double dissociations between reading and spelling have been reported in children/beginner readers (Hepner, McCloskey, & Rapp, 2017; Moll & Landerl, 2009; Wimmer & Mayringer, 2002), and in neurological patients (Hanley & Kay, 1992; Patterson, 1986). These dissociations have been taken as support for separate systems for reading and spelling. However, the current study adopts the assumption that the representations are partly or wholly shared.

4.4.2 Individual differences in orthographic knowledge

There is convincing evidence that individuals differ in their quality of orthographic representations. In a masked priming lexical decision task, Andrews & Lo (2013) explored if individuals differed in terms of their use of semantics and orthographic knowledge when reading morphological complex words. Morphemes play an important part in reading as they constitute the smallest meaning – bearing units in language. In line with previous research (Rastle, Davis, & New, 2004), targets and prime relations were divided into three groups: 1) transparent, i.e. semantically and orthographically related (e.g. cleaner – CLEAN), 2) opaque, i.e. semantically unrelated but orthographically related (e.g. corner – CORN), or 3) form-related, i.e. did not share a meaning or a morpheme (brothel – BROTH). Participants were divided into an ‘orthographic profile’ group (better spellers/lower vocabulary knowledge) and a ‘semantic profile’ group (higher vocabulary/ poorer spellers). Andrews & Lo (2013) found that participants with a ‘semantic profile’ showed stronger priming for morphologically transparent primes. Participants with an ‘orthographic profile’, on the other hand, exhibited equally strong priming for transparent and opaque primes. The authors concluded that individual differences in orthographic and semantic lexical quality modulated early stages of the reading process.

4.4.3 Orthographic lexical quality impacts in reading performance

There is some evidence that better and poorer spellers show differences in reading aloud. Martin-Chang et al. (2014) investigated the relationship between orthographic quality and naming reading speed. Participants were asked to complete a naming task. Seven days later they were asked to spell the same words five times. Participants were not able to refer to earlier spellings of the same word. Participants read words faster if they had spelled them consistently accurately across the five spellings. If words had been misspelled consistently in the same fashion over five spellings, then they were read faster than words which were misspelled in different ways. The authors concluded that increased orthographic lexical quality as indexed by consistent accurate spelling was related to faster word reading. Similarly, Rossi, Martin-Chang, & Ouellette (2019) reported that teenage participants with better spelling ability also read words faster. Their research further showed both accuracy as well as stability of the spelling (same spelling across all test instances) were predictive of faster reading speed.

4.4.4 Orthographic similarity

Words differ in orthographic overlap with other words. Orthographic neighbourhood has been shown to influence word naming and spelling. Words with more orthographic neighbours have been named (Davis, 2012; Spieler & Balota, 2000, for a review see Andrews, 1997) and spelled aloud (Roux & Bonin, 2009) faster by adult participants than words with fewer neighbours. This was also true for multisyllabic words in

English (Yap & Balota, 2009). Thus, generally in naming and spelling, denser orthographic neighbourhood has been found to be facilitating. This concurs with the framework of the PDP model, in which orthographic similarity contributes to stronger weights of shared hidden unit connections and thus to faster activation (Plaut et al., 1996).

Orthographic similarity has most often been operationalised in terms of substitution neighbours (C. R. Brown & Rubenstein, 1961; Landauer & Streeter, 1973), which define neighbours as words of the same length which differ in one letter only to the target word. Coltheart's *N* (Coltheart, Davelaar, Jonasson, & Besner, 1977) has been commonly used as the measure for this substitution neighbourhood. Words which are orthographically similar to many other words of the same length have a higher Coltheart's *N* than words which have fewer orthographic neighbours. Coltheart's *N* interacts with length, as shorter words have more orthographic neighbours than longer words (Balota et al., 2004; Frauenfelder, Baayen, Hellwig, & Schreuder, 1993). In fact, Andrews (1997) reported estimations from a 30,000 word sample and found that 4-letter words had on average 7.2 neighbours, whilst 5-letter words had 2.4 neighbours and 6-letter words had only 1.1 neighbour on average.

Yarkoni, Balota, & Yap (2008) further developed the measure of orthographic similarity to include not only substitution, but also deletion and insertion. This measure, called orthographic Levensthein distance 20 (old20), is the average number of changes (substitution, deletion, and insertion) that are needed to change the target word into its 20 nearest orthographic neighbours. Consequently, words with many neighbours have low old20 values, because on average they only need few changes to be transformed into their nearest 20 neighbours. Conversely, words with few neighbours have high old20 values. Whereas Coltheart's *N* only includes words of the same length within its definition of neighbourhood, old20 extends this definition to words of all length and is therefore more sensitive to orthographic similarity for longer words. Accordingly, old20 explained more variance than Coltheart's *N* in naming multisyllabic words (Yap & Balota, 2009). However, old20 and Coltheart's *N* seem to be very similar when assessing shorter monosyllabic words (Yarkoni et al., 2008).

As mentioned above (see section 4.4), there is some evidence that orthographic knowledge is acquired through statistical learning (Deacon et al., 2008; Treiman & Kessler, 2006), and this is congruent with the assumptions of the PDP model (Plaut et al., 1996). When considering individual differences in spelling (as a marker for orthographic knowledge) and their impact on naming, orthographic similarity may serve as a marker of the reading system to be able to exploit the statistical regularities of the script. It is conceivable that better spellers should be able to exploit the orthographic neighbourhood more than poorer spellers.

4.4.5 Section summary

According to the lexical quality hypothesis (LQH; Perfetti & Hart, 2002; Perfetti, 2007), readers may differ in the lexical quality of word representations. The orthographic component can vary in terms of the specific letter knowledge. Research has suggested that less specified orthographic representations lead to slower naming times (Martin-Chang et al., 2014).

It seems reasonable to believe that the best way to assess the quality of orthographic knowledge is by assessing spelling ability (Andrews & Lo, 2013). Although not unequivocal, extensive literature exists to support the idea that spelling and reading rely on the same orthographic representation, hence justifying the use of spelling tasks to probe the orthographic knowledge of readers (e.g., Holmes & Carruthers, 1998).

Spelling ability is subject to a wide spectrum of variation, and ordinarily studies refer to good versus poor spellers. Investigations into the reasons underlying spelling difficulties in poor spellers point to the incomplete representation hypothesis (Frith, 1980) which holds that poor spellers are less efficient at analysing and encoding orthographic information leading to incomplete lexical representations. This is in line with the lexical quality hypothesis (Perfetti & Hart, 2002; Perfetti, 2007). Other research has suggested that some individuals may find it more difficult to pick up statistical regularities of letter co-occurrences. (Deacon et al., 2008; Treiman & Kessler, 2006) It may be that this specific deficit in statistical learning leads to less concrete orthographic representations. Orthographic similarity may be a useful marker to explore this possibility further.

The current study will include spelling ability as a marker for individual differences in orthographic knowledge in its naming experiments. In line with previous research it is expected that better spellers will also be faster readers.

4.5 Chapter summary

Individuals' reading aloud is influenced by a number of factors, which shape their individual reading system. Within the connectionist modelling approach, differences in nonword reading ability have been considered as variations in the efficiency of phonological processing, or direct O-P pathway usage. There is also some evidence that difficulties in processing in the direct pathway may result in increased usage of the indirect semantic pathway as a compensatory measure. This concurs with the frequent observation that slower nonword reading is correlated with slower word reading. However, individuals also differ in terms of their semantic knowledge. As such, better semantic knowledge may result in faster semantic contributions to reading aloud. Equally, readers also differ in terms of precision and quality of the orthographic representations, as indexed by their spelling ability. Finally, the amount of training that a reading system has received also improves the functioning of the reading system.

4.6 The present investigation

The aim of the present investigation was to combine two research strands which have previously been investigated separately, namely research focusing on differences in reading due to spelling-sound consistency and research examining variation in reading due to individual differences. To the author's knowledge, the present study is the first inter-language behavioural study to include both a large set of psycholinguistic variables as well as a greater number of individual difference variables to predict reading aloud reaction times on a number of words shared in both languages. This places this study in a unique place to examine whether language effects uphold when individual differences are controlled.

The primary aim of the present investigation was to see if differences between languages in reading aloud remain, if individual differences are accounted for. The PGST suggests that readers of more opaque scripts require a greater variety of larger sized reading units to achieve accurate pronunciations. (Ziegler & Goswami, 2005) The quantification of reading units has typically been achieved by examining effects of word length to index small-unit processing and orthographic body neighbourhood to identify effects of larger unit sizes. (Ziegler et al., 2001) However, other studies suggest that the size of reading units may be influenced by other factors, including individual differences (e.g., Schmalz et al., 2014). The present study therefore also aims to examine whether effects for small and/or large grain sizes occur when individual differences have been accounted for.

In a second instance, the study aims to unpack interactions between psycholinguistic item-level effects and person-level effects of individual differences in reading aloud in the different language groups. There are a number of individual differences which may be more or less influential on the reading process, depending on the orthography.

First, the present investigation will explore the role of nonword reading ability when comparing readers of two different scripts. The ODH (e.g., Frost et al., 1987) proposes that the contribution of lexical and sublexical pathways varies with language transparency, but attempts to categorise readers into primarily lexical and sublexical readers within a language group have not been unequivocally successful (Baron & Strawson, 1976; Brown et al., 1994). Yet, it is generally acknowledged, that nonword decoding skills vary considerably between individuals (e.g., Torgesen et al., 1999). The triangle model rests on the assumption of the division of labour between the phonological and the semantic pathway to contribute to accurate pronunciations. Individuals may vary in the way that each pathway contributes. (Plaut et al., 1996) The present study will attempt to create greater insight into the importance of nonword reading ability and whether this differs between languages.

Second, the study will explore the contribution of semantics to reading aloud. The ODH suggests that readers of more opaque scripts show greater semantic effects because they rely more on lexical reading units. However, greater use of semantics has also been explored within the PDP model approach. Here, readers may differ in the division of labour between the phonological and semantic pathways. Greater contributions from the semantic pathway may stem from an inefficient functioning of the phonological route (Strain & Herdman, 1999; Woollams et al., 1996). Furthermore, not only do individuals vary in decoding ability, they also differ in their vocabulary knowledge, which could contribute to differential semantic effects. In the present study, semantic contribution in reading aloud will be investigated by examining the patterns for semantic effects at the item-level (AoA) and at the person-level (vocabulary knowledge) in two different languages, whilst taking into account person-level phonological route functioning (nonword decoding). The study explores in how far semantic effects are a result of the language itself, person-level attributes or both.

Third, the study will present a comparative investigation into the role of individual differences in the quality of orthographic representations between readers of two languages. Participants with greater quality lexical representations (as indexed by spelling ability) show better reading aloud performance (Martin-Chang et al., 2014; Rossi et al., 2019). It is assumed that this effect is due to more complete orthographic representations (Frith, 1980), which is compatible with the lexical quality hypothesis (Perfetti & Hart, 2002; Perfetti, 2007). The weakness in orthographic representations may derive from a specific deficit in statistical learning (Deacon et al., 2008; Treiman & Kessler, 2006). The present study will investigate if the quality of orthographic representations has a similar effect across the two languages.

Fourth, the study addresses the role of print exposure as a prime individual difference in reading performance. Print exposure has been identified as an important predictor of reading performance (e.g., Lewellen et al. 1993; Yap et al., 2011). Within the PDP model approach, print exposure can be seen as a measure of practice of the reading system. One marker of a mature reading system is the reduced effect of word frequency (e.g., Brysbaert et al., 2011; Yap et al., 2011; Monaghan et al., 2017). The present investigation will investigate the impact of reading practice (as measured by print exposure) across both languages.

In order to undertake such a comprehensive cross-language comparison, some measures had to be established prior to testing. First, this comparison of two languages differing in spelling-sound consistency had to be qualified by determining spelling-sound consistency of item stimuli. Some English consistency measures exist, but are only available for a small set of items (Ziegler, Stone, & Jacobs, 1997), or are not generally available (Yap & Balota, 2009). For German, no consistency measures have been published. The present investigation therefore entailed the computation of spelling-sound consistency measures for English and German. Second, a cross-language author recognition test and reading questionnaire suitable for both German and English participants were created based on the original work of Stanovich and colleagues (Stanovich & West, 1989; Stanovich & Cunningham, 1992; Stanovich, West, & Harrison, 1995). Third, whilst a large number of AoA estimates exist for English (e.g., Kuperman, Stadthagen – Gonzalez, & Brysbaert, 2012), very few are available for German (Schröder, Gemballa, Ruppin, & Wartenburger, 2012). For this reason, a larger set of AoA estimates were collected, and have since been published (Birchenough, Davies & Connelly, 2017).

Data will be analysed using mixed effect models (Baayen, Davidson, & Bates, 2008). Mixed models are able to model the variation due to predictor variables (fixed effects) whilst taking into account the variation within predictor variables (random effects) (Winter, 2013). They also provide a reliable statistical method for larger datasets even when data points are missing as often occurs with naming data (Baayen et al. 2008).

Results will be discussed in relation to extant theories of reading aloud. However, the study's objective is not to make the case for one over other reading models; rather theories are used to explain the findings and to put them into the context of current theoretical discussions. The aims of this study will be to see if differences in reading aloud across languages remain if individual differences are accounted for, and if individual differences play out similarly in these two languages.

In a first step, this comparison of two languages differing in spelling-sound consistency has to be qualified by determining spelling-sound consistency of item stimuli. Some English consistency measures exist, but are only available for a small set of items (Ziegler, Stone, & Jacobs, 1997), or are not generally available (Yap & Balota, 2009). For German, no consistency measures have been published. The present investigation therefore entailed the computation of spelling-sound consistency measures for English and German. This will be described in the next chapter.

5 Creation of consistency measures

The comparison of two languages differing in the reliability of their spelling- sound relationships requires a measure which can be applied to quantify this relationship. As outlined in Chapter 2, some studies have employed the concept of regularity. If the most common grapheme – phoneme conversion is used for reading a word, then the word is deemed regular, otherwise it is considered irregular (see for example, Ziegler et al., 2000). The irregularity concept is elementary to dual route models as it provides the rationale for the two routes structure to explain the reading of irregular exception words by means of the lexical route (Coltheart et al., 2001). However, given the decision to focus on PDP models of reading as a primary theoretical framework for the present investigation, the present study opted to use the concept of consistency, which has been central to the connectionist models of reading. Words which share the same orthography and phonology with many other words are considered more consistent than words who differ from many other words in either orthography or phonology. This concept ensures that the learning algorithm of the connectionist network is able to pick up statistical regularities between orthographic and phonological units, and that connection weights are adjusted to reflect these regularities. This means that reading of consistent words can benefit from the structure of a reading system which has been moulded by similar words (Plaut et al., 1996; Seidenberg & McClelland, 1989; Seidenberg, 1992). It is therefore important to establish an appropriate measure for the spelling-sound consistency of words to take it into account when comparing languages. As the aim of the current investigation was to compare two languages, it seemed that the creation of comparable measures based on similar computations and word corpora would be desirable.

The present chapter will give a brief overview how antecedent studies have quantified consistency in English, and how spelling-sound relationships have been described for German. Following this, the preparation of consistency measures for English and German will be presented, and results will be compared to previous consistency estimates.

Although the concept of consistency could also be applied to other linguistic unit sizes (Coltheart, 2012; J. Monaghan & Ellis, 2002; Plaut et al., 1996; Strain, Patterson, & Seidenberg, 2002), it has typically been operationalised in terms of syllabic body-rime consistency (Balota et al., 2004; Glushko, 1979; Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995; Yap & Balota, 2009). Take for example the (oft – cited) monosyllabic word *pint*, which can be parsed into an onset, i.e. the initial consonant (P), a vowel (I) and a coda, i.e. the final consonants (NT). The vowel and the coda taken together (-INT) form the phonological *rime* (Treiman & Chafetz, 1987), or orthographic word *body* (Kay & Bishop, 1987; Patterson & Morton, 1985). Body-rimes can be consistent in the way that a particular spelling will always result in the same pronunciation, for example, *-at* will always be pronounced /æt/ as in *cat, hat, mat*. These words then have consistent body-rimes. The word *pint*, on the other hand, is considered inconsistent, because its body twins are pronounced differently, as in *mint* and *hint*.

A number of studies have constructed *token* consistency measures for rime-bodies by summing the frequency of their 'friends' and 'enemies'. *Friends* are those words in a given corpus which are spelled and pronounced the same as the target word. Correspondingly, an *enemy* is a word which is spelled the same but pronounced differently. In their seminal study, Jared et al. (1990) showed that naming latencies were longest for words with high frequency enemies and low frequency friends, and shortest for words with low frequency enemies and high frequency friends. The choice of stimuli was based on the frequency of friends and enemies based on all uninflected monosyllabic words of the Kučera & Francis (1967) corpus to reflect the relative proportion of high- and low frequency friends and enemies.

Ziegler et al. (1997) presented a spelling-sound consistency database for English, basing the spelling – sound consistency calculations for English body-rimes on all monosyllabic, monomorphemic words available in the Kucera & Francis (1967) database. The phonemic transcriptions were based on the pronunciations given by the Webster Pocket Dictionary. For all spelling bodies, they extracted all corresponding rimes. Correspondingly, they extracted all spelling bodies possible for each phonological rime. A body was considered feedforward consistent if it mapped onto only one possible pronunciation. Likewise, a rime was considered feedback consistent if it could only be spelled in one possible way. Ziegler et al. (1997) noted that this would make uniquely spelled words such as *yacht* seem feedforward consistent, as they had only one corresponding phonological body. Yet, when considering how many spellings matched the phonological form of these words, it often showed that these words were highly feedback inconsistent.

More recently, consistency computations have been based on larger corpora and have employed both token and type measures. Type measures count the instances of occurrence of a particular unit, rather than summing the frequency of the words in which units occur. Balota et al. (2004) based their consistency calculations on the 4,444 monosyllabic words of the ELP for which frequency estimates (Zeno, Ivens, Millard, & Duvvuri, 1995) were available. The authors argued that the larger number of words as a basis for computations than previous studies improved the sensitivity of the consistency measure. Although Balota et al. (2004) reported effects on consistency measures, which had been computed based on the frequencies of friends and enemies, they noted that consistency measures based on type counts yielded the same consistency effects.

An even larger corpus was employed by Yap (2007), and by extension Yap & Balota, (2009). These consistency calculations, for the first time, included also multisyllabic words, and were based on all monomorphemic 9,643 words of the ELP. Not only were these measures based on the hitherto largest word corpus, but they also included consistency measures of onset and rimes of all word syllables, and introduced a new averaged consistency estimate across onsets and rimes of a word syllables, the composite consistency measure. Moreover, Yap (2007) found that type and token consistency performed almost identically in predicting word recognition, and that position-specific were preferred to position-independent computations. Hence, consistency estimates reported in Yap & Balota (2009) were based on position-specific type counts.

Whilst several studies have presented consistency computations for English, there were none available for German at the time of planning the present study. However, multiple studies have of course explored spelling-sound relationships in German.

In terms of its grapheme-phoneme correspondence, German is considered more regular than English. In the German implementation of the dual route model (DRC), Ziegler et al. (2000) used grapheme-phoneme correspondences (GPC) of German monosyllabic words to extract the corresponding GPC rules. They found that only about 10.3 % of the monosyllabic words could not be read according to the GPC rules (compared to 18% in English). The authors identified three main reasons for irregularity in German. The most common was due to unexpected vowel length, as for example in *hoch* (*high, up*), which is pronounced with a long vowel, even though on most occasions a vowel receives a short pronunciation when followed by two consonants. Secondly, unexpected pronunciations could be found with loan words, such as *Jazz* and *Job*. Finally, a very small number of other irregular words could not be categorized in any meaningful manner, such as *Volt* or *zig* (*umpteen*). Interestingly, Ziegler et al. (2000) pointed out that the share of words which were deemed irregular according to GPC rules was bigger than commonly expected.

Seymour et al. (2003) reported a classification of 16 European languages, based on the work of a network of European researchers (Niessen, Frith, Reitsma, & Ohngren, 2000). They classified orthographies on two dimensions, namely syllabic complexity and orthographic depth. The latter was defined as shallow when graphemes and phonemes had one-to-one correspondence, and was considered deeper if this relationship was complicated by multi-letter rules, context-dependent rules and other irregularities (including those due to morphology). Within this framework, German was considered orthographically shallow, along with, for example, Greek and Italian. Only Finnish was considered shallower.

However, it has been pointed out that although grapheme – phoneme mappings mostly have a one-to-one correspondence when reading German, this regularity is reduced in spelling, as sounds may be spelled in several ways (Landerl, 2002; Moll, Fussenegger, Willburger, & Landerl, 2009). When considering body-rime consistency, this imbalance between feedforward (spelling to sound) and feedback (sound to spelling) consistency seems to persist. Wimmer & Mayringer (2002) reported a personal communication by J. Ziegler, stating that German was 84% feedforward consistent but only 47% feedback consistent. In comparison, Ziegler et al. (1997) computed 69.3 % feedforward consistency and 27.7 % feedback consistency for English.

A similar trend was observed in *H* values (Fitts & Posner, 1967). Ranging between 0 and 1, higher *H* values capture greater variation in pronunciation for a particular spelling-sound unit. For the word body, Treiman et al. (1995) reported values of .14 by type and .22 by token. For German, Perry & Ziegler (2002) computed .07 by type and .03 by token, thereby evidencing less variation in spelling-sound relationships.

Feedback (FB) inconsistency makes spelling more difficult. The effect for reading, however, is more contentious. Stone, Vanhoy, & Van Orden (1997) reported that – based on the Kucera & Francis word corpus 76.7 % of all English monosyllabic words were FB inconsistent. They suggested that previously found small feedforward (FF) consistency effects in lexical decision tasks might have been reduced by uncontrolled FB inconsistency. This was replicated in French and extended to naming, although the effect was weaker (Ziegler, Montant, & Jacobs, 1997). However, Peereman, Content, & Bonin (1998) were not able to replicate these results with French items. Instead, they found that FB inconsistency resulted in slower and more error prone writing, but did not affect LDT or naming. Replicating Ziegler et al.'s 1997 study whilst controlling for subjective frequency, they found no significant FB consistency effect. Yet, more large-scale investigations with consistency computations based on larger word corpora reported FB consistency effects in naming for English, even after controlling for objective and subjective frequency (Balota et al., 2004; Yap & Balota, 2009).

The present study presents consistency computations for onset and rime units for German and English mono - and multisyllabic words. The aim was to create two comparable measures based on similar word corpora in order to maximise comparability. The method of computation closely followed (Yap, 2007, and by extension Yap & Balota, 2009) to ensure validity. Given the proposition that inconsistency in German may be stronger in the FB than the FF direction, computations are presented for both directions.

5.1 Method

The present consistency computations used the syllabified phonological transcriptions of the Celex database for German and English (Baayen, Piepenbrock, & Gulikers, 1995), which both comprise over 50,000 words. However, such large databases may lead to increasing the effect of rare words with unusual spelling-sound combinations on the computations. For this reason, only those Celex items were included which also appeared in the Clearpond database (Marian, Bartolotti, Chabal, & Shook, 2012), a collection of word corpora specifically compiled for cross-language comparison. It comprises the 27,751 most frequent words for five languages, including German and English.

In order to capture how reliably spelling is reflected in sound and *vice versa*, it is necessary to parse words and their phonological transcriptions into the units to be compared. This presents two potential difficulties for the computations. First, depending on the phonological transcription, single sounds may consist of several signifiers, and thus make the parsing process more difficult when using computer software. Using an example from the Celex German Linguistic Guide, IPA uses /ts/ to transcribe the sound for the first letter in the German word *Zahl* (*number*). This means that 2 signs are needed to describe a single sound, which makes the matching between letter and sound more difficult. However, Celex provides the machine-readable phonological DISC transcription. Contrary to for example the IPA transcriptions, DISC uses a single sign for each sound, which makes it especially suitable for computing with programming languages. The DISC code for /ts/ is ‘=’. Thus, computations were calculated based on the phonological DISC code using the software R (R Core Team, 2014).

Second, parsing into syllables and onset and rime units within syllables is not without challenges, as sounds of syllables may not as clearly separate into onsets and rimes as letters. With view to this difficulty of parsing words and phonological transcriptions into comparable units, the researcher chose the parsing of the phonological transcriptions in Celex as the default. Orthographic words were then parsed to fit the phonological parsings. For example, the word *c-ute* was phonologically parsed as *k-jut*, rather than *kj-ut*. Recall that Yap (2007) had found that consistency measures based on words, which were parsed either according to orthographic or phonological principles, performed very similarly. However, given that syllables are phonological units, Yap (2007) opted to report results based on phonological parsing. In accordance with these results, multisyllabic words in the present database were also split according to phonological principles. The guiding rule was to match orthography as closely as possible to the phonological syllables. Sometimes, phonological parsings as given by Celex were also adjusted. Where possible, the Duden Aussprachewörterbuch (1990) or online dictionaries were used to update some pronunciations, or to adjudicate if the researcher disagreed with a Celex entry. It needs to be clear that other researchers may have resolved some parsing questions differently, and that the decisions taken will naturally have influenced the resulting consistency measures.

German

For German, 11,192 words were extracted as common words in both Celex and Clearpond databases, thereby necessarily excluding very rare or uncommon words. Following this, double entries of words were deleted unless they were homographs with different pronunciations, as for example *Rentier*, which could be /rɛnˈti̯eː/ (*annuitant*) or /ˈreːntiːɐ/ (*reindeer*).

The resulting German corpus included 10,885 words. In terms of phonological syllables, it comprised 1,029 monosyllabic words, 4,049 bi-syllabic, 3,882 tri-syllabic and 1,493 words with four syllables, 351 with five, 68 with six, 11 with 7 syllables, and one each with eight and nine syllables. The mean number of syllables was 2.66. The mean number of letters per word was 8.21 (SD = 2.76) with a minimum of 2 and a maximum of 27 letters.

English

The English versions of the Celex and the Clearpond database had 14,084 words in common, which also had an entry in the SUBTLEX-US frequency database (Brysbaert & New, 2009). As with the German database, words were parsed according to the phonological syllables given by Celex, but on occasion the researcher adjusted these. If several phonological syllable parsings were offered by Celex, the first was selected.

The resulting database had 14,084 word entries. The mean number of phonological syllables was 1.37, which was less than for the German database, even though the English database had 3,199 more words. The corpus comprised 2,830 monosyllabic words, 5,854 bi-syllabic, 3,415 tri-syllabic and 1,475 words with four syllables, 445 with five, 58 with six, and 7 with seven syllables. English words were somewhat shorter than German words. The mean number of letters per word was 7.27 (SD = 2.37, min = 1, max = 19).

In line with previous studies (Yap & Balota, 2009; Yap, 2007), consistencies were computed as position-specific type measures for all onset and rime units.

Feedforward consistency was based on the

unit + units with same spelling + pronunciation / all units with same spelling

and feedback consistency was based on the

unit + units with same spelling + pronunciation / all units with same pronunciation.

Consistency values ranged between 0 and 1 with higher values indicating greater consistency.

5.2 Results and Discussion

The present study provides consistency computations for onset and rime units for 10,885 German and 14,084 English words. These are provided for all phonological syllables. Composite consistency measures, which constitute the consistency average across all syllables for a given word (Yap & Balota, 2009), have also been added. Appendix 11.1 gives overviews of the first six entries for each database.

To provide more insight and transparency into the computations, Appendix 11.2 lists the words for each language, which according to the presented databases are the most inconsistent of the following units: feedforward onsets for syllables 1 and 2 (FFO1 and FFO2), feedback onsets for syllables 1 and 2 (FBO1 and FBO2), feedforward rimes for syllables 1 and 2 (FFR1 and FFR2), feedback rimes for syllables 1 and 2 (FBR1 and FBR2). These lists also show the syllable parsing of each word, and the corresponding phonological DISC transcriptions on which the computations were based. Appendix 11.3 gives similar lists for words with the most inconsistent composite consistencies for FFO, FBO, FFR and FBR, and provide details of the consistency values for the first four relevant units, which the composite consistencies are based on.

As outlined above, the present consistency computations followed the method of Yap (2007) and Yap & Balota (2009) very closely. Computations were based on the same formula and also used the Celex database for phonological transcriptions and syllabification. Equally, orthographic syllable parsing was adjusted to conform to phonological parsing, although due to the element of arbitrariness of the task, the outcome may have been slightly different. The main difference between the two databases was the choice of words. Whilst Yap & Balota (2009) presented their consistency computations for 9,639 monomorphemic words of the ELP (Balota et al., 2007), the present study extracted all words common between the Clearpond and Celex databases. The motivation for this method was the aim of language comparison. The Clearpond database was specifically assembled for the purpose of inter-language comparison and consists of the 27 thousand most common words in several languages. By choosing words from the Clearpond database, it was hoped to make the consistency computations more comparable between German and English.

Table 5.1

Comparative table for means (and standard deviations) of feedforward and feedback onset and rime consistency computations for present study and Yap & Balota (2009)

	English		German
	present study (n = 14084)	Yap & Balota (2009) (n = 9639)	present study (n = 10885)
Syllable 1			
FF onset consistency	0.96 (0.14)	0.93 (0.18)	0.98 (0.10)
FF rime consistency	0.60 (0.31)	0.47 (0.28)	0.76 (0.27)
FB onset consistency	0.92 (0.19)	0.91 (0.19)	0.92 (0.18)
FB rime consistency	0.62 (0.31)	0.62 (0.31)	0.84 (0.25)
Syllable 2			
FF onset consistency	0.86 (0.24)		0.97 (0.12)
FF rime consistency	0.61 (0.32)		0.74 (0.27)
FB onset consistency	0.67 (0.30)		0.82 (0.25)
FB rime consistency	0.55 (0.33)		0.83 (0.25)
Comp FF onset consistency	0.90 (0.15)	0.84 (0.16)	0.97 (0.08)
Comp FF rime consistency	0.67 (0.22)	0.54 (0.20)	0.79 (0.17)
Comp FB onset consistency	0.81 (0.20)	0.75 (0.18)	0.86 (0.16)
Comp FB rime consistency	0.60 (0.23)	0.53 (0.20)	0.84 (0.18)

Note . Consistency is expressed as a decimal figures with 0 = not consistent and 1 = completely consistent; FF = feedforward (orthography to phonology), FB = feedback (phonology to orthography); Comp = composite (averaged across all onsets or rimes in database respectively)

Table 5.1 shows the means and standard deviations for the present German and English consistency computations, and those presented by Yap & Balota (2009) have been added for comparative purposes. The present consistency computations for English were very similar to those reported by Yap & Balota, providing some support for their validity. However, Yap & Balota (2009) reported a distinctly lower feedforward rime consistency for the first syllable (FFR1). It is likely that this discrepancy is due to the choices made when parsing words into syllables, onsets and rimes according to phonology.

All in all, German seems more consistent than English at all levels. Importantly, however, FB rime consistency was greater than FF rime consistency for German. This seems contradictory to previous literature as reviewed above (Landerl, 2002; Moll et al., 2009).

In order to examine FF and FB rime consistencies further, Table 5.2 gives averages for rime consistencies for each syllable when grouped into mono-, bi- and trisyllabic words. Body-rime consistency for the third syllable in trisyllabic words was not reported, as the stimuli in the dataset used in Chapter 8 and 9 had at the most two syllables. Remember, however, that whilst this table represents averages for only a subgroup of words, the consistency value for each unit was still based on the complete datasets.

Table 5.2

Comparative table for means (and standard deviations) of feedforward and feedback rime consistency computations for words with one, two and three syllables.

	English	German
Monosyllabic words	(n = 2830)	(n = 1029)
FF rime consistency	0.84 (0.26)	0.89 (0.24)
FB rime consistency	0.62 (0.33)	0.71 (0.34)
Bisyllabic words	(n = 5854)	(n = 4049)
Syllable 1		
FF rime consistency	0.56 (0.30)	0.71 (0.29)
FB rime consistency	0.59 (0.31)	0.81 (0.27)
Syllable 2		
FF rime consistency	0.74 (0.28)	0.84 (0.21)
FB rime consistency	0.59 (0.33)	0.87 (0.24)
Trisyllabic words	(n = 3415)	(n = 3882)
Syllable 1		
FF rime consistency	0.52 (0.27)	0.78 (0.25)
FB rime consistency	0.64 (0.29)	0.88 (0.22)
Syllable 2		
FF rime consistency	0.47 (0.31)	0.66 (0.31)
FB rime consistency	0.49 (0.32)	0.78 (0.28)

Note. Consistency is expressed as a decimal figure with 0 = not consistent and 1 = completely consistent; FF = feedforward (orthography to phonology), FB = feedback (phonology to orthography); body-rime consistency for the third syllable in trisyllabic words is not reported, as stimuli in the naming studies had a maximum of two syllables.

There are two interesting observations which follow from Table 5.2. The first observation relates to the FB rime consistency of the first syllable in German. FB rime consistency seemed to be lower than FF rime consistency for monosyllables, but not for multisyllabic words. Remember that when averaging across all words in the German database, first syllable rimes were more FB than FF consistent, which was contrary to expectations (see Table 5.1). The finding that this may not apply to monosyllables could help to explain this puzzling result. It is possible that previous estimates of FF and FB rime consistency in German were largely based on monosyllabic words, and hence found that FF rime consistency was greater than FB rime consistency, whilst in the current database monosyllabic words were only a small subset of all words included.

The second observation pertains to FF rime consistency in the first syllable. In English, the FF consistency for bi- and tri-syllabic words was lower than for monosyllables. This may be an important observation, as the greater rime consistency for monosyllabic words may speed up word recognition for these words. Furthermore, it will lead to lower composite consistencies for words with more than 1 syllable. Although first syllable FF rime consistency in German multisyllabic words also seemed somewhat lower than in monosyllables, overall consistency values varied less across syllable-word groups. This is likely to improve recognition performance for longer words in German compared to English.

The present consistency measures for English seem comparable to those presented by Yap & Balota (2009). By extension, it could be argued that the German computations must be equally valid, as the same method was applied. However, it would be preferable to compare the present German consistency measures directly to others. As outlined previously, there are currently no German database for consistency measures available. However, in a personal communication from February 2001 J Ziegler estimated that 84% of all German monosyllabic words were FF rime consistent and 47% were FB rime consistent (Wimmer & Mayringer, 2002). In order to compare the present consistency data to these figures, FF and FB rime consistency measures for all 1,029 German monosyllabic words were summarised in Table 5.3. For this purpose, rimes which had scores of 1 were considered consistent. All rimes with scores lower than 1 were considered inconsistent. Thus, in the present database, 73% of German monosyllabic words were FF rime consistent, and 39% were FB rime consistent (see bold figures in Table 5.3). These results seem to approximate Ziegler's estimations.

Table 5.3 also shows that of German FF consistent monosyllabic words more than half were categorised as FB inconsistent. However, an even stronger tendency for FB inconsistency appeared for FF inconsistent words. Thus, for German monosyllabic words it seems that there is a general tendency for FB inconsistency, which is even more pronounced for FF inconsistent words.

Table 5.3

Numbers (and percentages) of German monosyllabic words categorised according to feedforward and feedback body-rime consistency

	Feedback					
	Consistent		Inconsistent		Total	
	n	%	*n*	%	*n*	%
Feedforward						
Consistent	333	32	420	41	753	**73**
Inconsistent	68	7	208	20	276	27
Total	401	**39**	628	61	1029	100

A similar comparison to previous consistency estimates can be presented for English. In fact, in order to facilitate easy comparison, the format of Table 5.3 was deliberately chosen to be identical to Ziegler, Stone, et al. (1997)'s table of consistency estimates of English monosyllables. To compare the present to Ziegler et al.'s estimates, Table 5.4 presents the summarised consistencies for FF and FB rimes for English monosyllables of the current database. As before with the German data, all 2,830 monosyllabic words from the English consistency database were selected and categorised into consistent and inconsistent. Only words with a rime consistency of 1 fell into the consistent category. Ziegler, Stone, et al. (1997) reported 69% of English monosyllables as feedforward consistent and 28% as feedback consistent. In the present study, English monosyllables were 59% FF consistent and 22% (see bold figures in Table 5.4). This is a reassuring finding, as figures show the same trend as in Ziegler, Stone, et al. (1997).

As with the German data, estimates proposed by Ziegler, Stone, et al. (1997) were somewhat more consistent than they appear in the present computations. The reason for this may be the greater number of words in the present database. Ziegler and colleagues based their computations on 2,694 words. In the present database, consistency was computed on the complete database of 14,084 words. Given that words were considered inconsistent as soon as they had one body-rime enemy, it is probable that a larger corpus also included more words which deviated from the most common pronunciation or spelling in the first body-rime. Therefore, it would be expected that the percentage of inconsistent words in the present database should be greater than those reported by Ziegler and colleagues. Thus, all in all the present results seem to be in line with reports of English rime consistency by Ziegler, Stone, et al. (1997).

Table 5.4

Numbers (and percentages) of English monosyllabic words categorised according to feedforward and feedback body-rime consistency

	Feedback					
	Consistent		Inconsistent		Total	
	n	%	*n*	%	*n*	%
Feedforward						
Consistent	454	16	1204	43	1658	**59**
Inconsistent	169	6	1003	35	1172	41
Total	623	**22**	2207	78	2830	100

5.3 Summary

In summary, two large databases for FF and FB consistency measures of onset and rime units have been presented for English and German. The computations of these measures were based on similar corpora, and the methods employed were the same, which was to enhance the inter-language comparability. For German, this is the first database of this kind. The presented consistency measures seem to be quite similar to those presented previously by Yap & Balota (2009) for English, except for FF rime consistency estimates, which seems more consistent. This may be due to parsing decisions taken. The German consistency measures generally appear to reflect previous reports of spelling-sound relationships.

The next chapter is dedicated to the creation and analysis of print exposure measures designed to be comparable across two languages. Chapter 7 will then report the naming study.

6 Creation of print exposure measures

As reviewed in section 4.3, one central individual difference is the amount of reading practice over the lifetime. For skilled readers, print exposure has been measured by a variety of means, often in the form of a self-report in which participants are asked to share their own estimation of their reading habits. Common self-reports are reading diaries (Smith, 2000; White et al., 2010) and reading questionnaires (Clark & Foster, 2005; McGeown et al., 2014; Noor, 2011; Schmidt & Retelsdorf, 2016; Stanovich & Cunningham, 1992; White et al., 2010). However, Stanovich & West (1989) argued that self-report measures were subject to a social desirability confound, and proposed an author recognition test (ART) as a more objective measure to gauge people's reading. In the ART, participants are asked to identify authors from a list of author names mixed with an equal number of foil author names. Participants are advised that scores will be lost for each non-author falsely identified as a real author, thereby correcting for guessing. The number of correctly identified authors (less the number of erroneously identified foil authors) gives an indication of the person's reading experience compared to other participants taking part in the same study.

Since its first inception a number of ART versions have been created and adapted to cater for children (Stainthorp, 1997) and adults (Acheson, Wells, & MacDonald, 2008; Burt & Fury, 2000; Masterson & Hayes, 2007; Stainthorp, 1997; Stanovich & West, 1989). The reason for continued adaption lies with the ART's cultural sensitivity (Stainthorp, 1997). The changing popularity of authors over time may reduce older ART versions' sensitivity for reading experience. Acheson et al. (2008) ran a pilot test on the authors included in the original Stanovich & West (1989) ART, and found that many of authors seemed unfamiliar to participants in the noughties, concluding that the ART was a measure which needed updating.

At the time of devising the present study, there was no ART version directed at a German readership, and the last ART version for the UK had been published in 2007

(Masterson & Hayes, 2007). It therefore seemed beneficial for the present investigation to create a new, adapted version of the ART. Given that the current study investigated reading across two different languages, only one ART version for both language groups was created. Although cultural differences in author popularity between English and German speakers were to be expected, a common test was thought to increase comparability.

The present chapter describes the preparation and analysis of a newly created ART, based on the original by Stanovich & West (1989), which could be administered to both English and German speaking participants. The findings from the present ART version (hence referred to as the GE - ART) will then be compared to results from an adapted version of the Reading and Media Habits Questionnaire by Stanovich & West (1989), which will be referred to as the GE-RQ.

6.1 Author recognition test

6.1.1 Methods

6.1.1.1 Participants

Participants who completed the GE-ART and the GE-RQ were the same as in the naming study described in Chapter 7. To avoid unnecessary duplicate reporting of information, the full description of participant characteristics, data collection and data preparation can therefore be found in Chapter 7 (please refer to section 7.2). Here, only a brief outline of the most relevant information will be given.

Participants were aged between 18 and 25 years and did not report any reading impairment. Participants were students or had completed their studies within the last three years. The following analysis includes data from 104 English – speaking and 104 German-speaking participants.

6.1.1.2 Materials

In order to create a single author recognition test (GE-ART), which would be sensitive to reading practices in both languages, fifty authors were chosen, who had all been published in both languages. 14 authors had appeared at least once on both the UK and the DE Amazon bestseller lists of the years 2013, 2014 and 2015. Another seven authors had been on an Amazon bestseller list in the UK, and a further 6 had featured on the German one. All other authors had either been awarded a prize, or had sold considerable numbers of copies over the years. Wolfgang Herrndorf was added as an author because of his recent popularity, and to counterbalance a bias towards English language authors, as in the compiled list far fewer authors were originally published in German compared to English. Authors were selected from a wide range of genres: autobiography, business, children's fiction, crime/mystery, fantasy, fiction, history, horror, humour, science fiction, IT, nutrition, psychology, romance, self-help, teenage, thriller and travel.

Fifty foil authors were added randomly to the author list. All names of non-authors were made up, and internet searches of the names confirmed that these were not authors who were unknown to the researcher.

Instructions were taken from Stanovich & West (1989), and were also translated into German. Previously Martin-Chang & Gould (2008) had found that ART scores based on authors whose work participants had read (primary print knowledge, PPK) were stronger predictors of reading rate (number of words read within one minute), vocabulary knowledge and self-reported reading than ART scores based on mere recognition of authors without having read their work (secondary print knowledge). The original ART (Stanovich & West, 1989) had not taken into account how the author was known. The present GE- ART was therefore adapted to be able to distinguish between different levels of author familiarity.

Specifically, apart from indicating whether participants had (a) recognised the author's name, participants were also asked to specify if (b) the author had been recognised, but had not been read, (c) a book by an author had been started, but not finished, and (d) at least one book by this author had been read.

The resulting GE-ART therefore included 50 real author names and 50 made-up names, and was aimed at both language groups. To counterbalance any order effects, three versions of the GE-ART were created with items in a different random order. One version can be seen in appendix 11.5.

6.1.1.3 Procedure

The GE-ART was one of several tasks given to each participant in the study. These tasks included a speeded naming task, a nonword reading test, a spelling test, a vocabulary test, as well as the GE-Q and GE-ART. Except on two occasions when the GE-ART was completed first (for 1 English and 1 German participant), the naming task was always the first task to be administered, whilst all other tasks including the GE-ART were completed afterwards in a random order.

6.1.2 Results

6.1.2.1 Data preparation

As described above, participants were expected to indicate author recognition (box 1), and then to specify how the author was known (boxes 2 - 4). A number of participants, however, interpreted the four columns as distinct options, thereby distinguishing between box 1 and box 2. In cases where only box 1 but none of the other boxes (2-4) had been ticked, ticks in box 1 were also scored as ticks in box 2. In case of future use of the GE-ART, it is therefore recommended to only give boxes 2 to 4 as options.

Names of foil authors who received many ticks were reinvestigated. Despite previous internet searches, it was found that the name M C Smith did in fact belong to two authors. However, they did not seem to fulfil the criteria of having been published in both languages, or to be known to a greater number of readers (as evidenced by prizes or bestseller lists). Moreover, none of the participants specified to have read a book by an author of this name. Therefore, M C Smith remained a non-author name for the purpose of this study. However, for possible future uses of this test, it may be necessary to reconsider the inclusion of this name.

In the German GE-ART, the author Hermann Hesse was erroneously misspelled as Herman Hesse. Until this error was rectified, the test had been completed twenty-four times, and Hermann Hesse had been recognised as an author 19 times. A Fisher Exact test confirmed that there was no association between whether or not Hermann Hesse had been recognised as an author and the spelling (p = .776), and thus will not have influenced recognition scores.

In line with previous literature, the following three indicators of print exposure were extracted for each participant:

1) general author recognition (= number of authors recognised);

2) primary print knowledge (PPK, = number of authors who had been read by the participant). This was based on the PPK measure promoted by Martin-Chang & Gould (2008);

3) SW89 – ART score (= number of authors recognised less the number of mistakenly recognised foil authors). This measure was based on the scoring used in the original ART by Stanovich & West (1989).

6.1.2.2 English data

The data of 104 participants were eligible for analysis. The mean number of authors recognised were 11.09 (*SD* = 5.92) with a minimum of 2 and a maximum of 31. The mean number of non-authors checked were 1.02 (*SD* = 1.85) with a minimum of 0 and a maximum of 10. The SW89 - ART score, based on the scoring used by Stanovich & West (1989), was computed by subtracting the number of foils incorrectly identified as real from the number of correctly identified (real) author names. In the present study, this resulted in a mean score of 10.07 (*SD* = 5.57) with a range of 0 to 30. These results are comparable to Stanovich & West (1989) (*M* = 9.3, *SD* = 5.7, Exp 1) and Masterson & Hayes (2007) (*M* = 9.44, *SD* = 7.75). Figure 6.1 shows the density plot of SW89 – ART scores of the GE-ART for English and German.

For the English language groups, the three versions of GE-ART had been employed 35, 34, and 35 times respectively. No differences were found between the SW89 - ART scores in the three GE - ART versions (*F*(1,102) = 0.663, *p* = .417).

6.1.2.3 German data

The data of 104 participants contributed to the German data set. German participants recognised on average 11.8 (*SD* = 5.59) authors with a range of 1 to 28. On average, 1.24 (*SD* = 2.11) non-authors were ticked, ranging from 0 to maximal 9. The SW89-ART score was 10.56 (*SD* = 5.22, range 0 - 28). These scores were very close to the results reported for the English language set. In Figure 6.1 it can be seen, that the distributions of SW89-ART scores for both language groups were very similar.

For the German language group, GE-ART versions were used 34, 35, and 35 times. As with the English language group, no differences were found between SW89 - ART scores of the three different versions of the ART (*F*(1,102) = 0.744, *p* = .39).

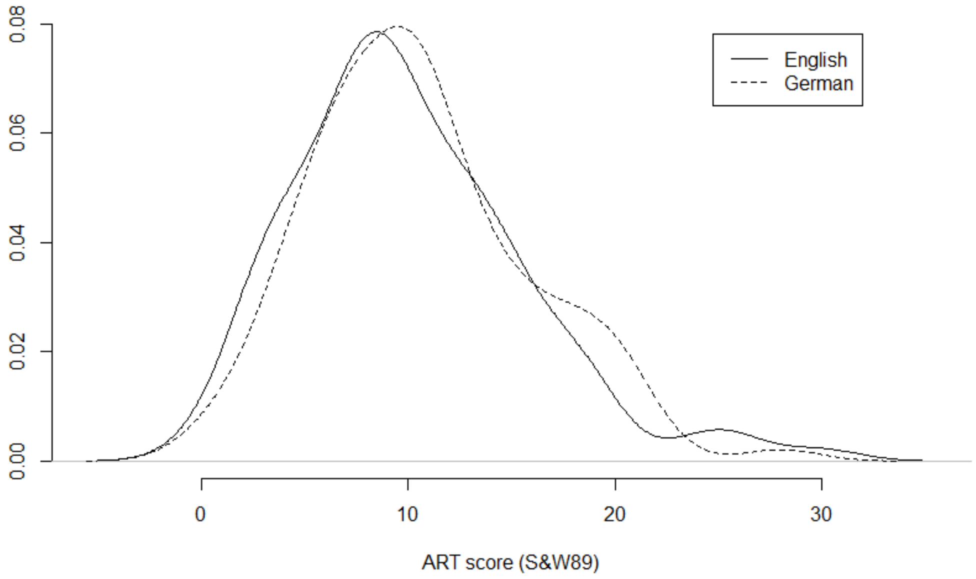


Figure 6.1. Density plots for ART score distributions per language group. ART scores were computed according to Stanovich & West (1989).

6.1.2.4 Primary Print Knowledge (PPK)

Results for primary print knowledge (PPK) as defined by dividing the number of authors read by the number of true authors were also investigated. Martin-Chang & Gould (2008) had found that PPK correlated more strongly with self-reported reading behaviour than secondary print knowledge. They had used the proportion of number of authors out of all real authors in the ART. However, as all participants had been filling in the same GE-ART, it did not seem necessary to present the measure in this way. Rather, the number of authors whose work had been read was employed as the PPK measure. Please note that this measure did not counterbalance for a social desirability bias by deducting points for wrongly identified non-authors. Martin-Chang & Gould (2008) reported that none of their participants had claimed to have read a book written by a non-author. Similarly, in the present study, only one participant in the German data set and none in the English data set declared to have read a book by a non-author. These results indicate that participants were answering truthfully, and that correction for guessing may not be a necessary requirement for the ART.

On average, the English language group reported having read 4.6 authors (*SD* = 3.66, range 0 - 19) or a proportion of 0.09 (*SD* = 0.07, range 0 – 0.38). This is comparable to the results reported by Martin – Chang et al. (2008) who reported a proportion of 0.07 (SD = .06, range 0-0.32). The German language group had read on average 5.03 (*SD* = 3.23, range = 0-14) authors or a proportion of 0.1 (*SD* = 0.06, range 0 – 0.28). Although the average for the German group was slightly higher, this difference was not significant (W = 5981.5, p = .185).

6.1.2.5 Reliability

To measure reliability, participants were split into even and odd numbered participants. For each group, the number of times an author was correctly recognised was computed. For the English groups, the number of times that each author was correctly identified very highly correlated with $r = .984$ (Spearman-Brown correction: $r = .992$). For the two German groups, the correlation was equally high at $r = .983$ (Spearman-Brown correction: $r = .991$).

Reliability for PPK scores was similarly strong for the English group ($r = .971$; with Spearman – Brown correction: $r = .985$) and the German group ($r = .944$, with Spearman – Brown correction: $r = .971$).

6.1.2.6 Were the same authors recognised across languages?

As the GE-ART was used for both language groups, there was a legitimate interest to find out, how author recognition compared across languages.

Authors were recognised on average 23.06 (SD = 27.41) times (22%) for the English group, and 24.54 (SD = 29.36) times (24%) for the German group. This is comparable to (Moore & Gordon, 2015), whose mean selection rate was 24%. Appendices 11.6 and 11.7 list the number of times authors were a) recognised, and b) their books had been read by participants for the English and German language groups respectively. Within the English language group, the number of times an author was recognised and the number of times an author has been read, correlated very strongly ($r_s = .883$, $p < .001$). Within the German language group, the number of times an author was recognised and the number of times an author has been read equally correlated very strongly ($r_s = .9$, $p < .001$).

In order to find out if the same authors were recognised similarly across languages, recognition times per author were correlated, which resulted in a surprisingly large association ($r = .664$). Figure 6.2 shows the corresponding scatter plot. Only Cornelia Funke, Hermann Hesse and Ken Follet seemed to have been recognised far more often by German speakers than English speakers. Figure 6.3 visualises the correlation between language groups regarding the authors whose work had been read in both language groups (PPK scores), which was also strongly, positively correlated (r = .735). Even when J K Rowling was removed, the correlation was still very strong ($r = .610$).

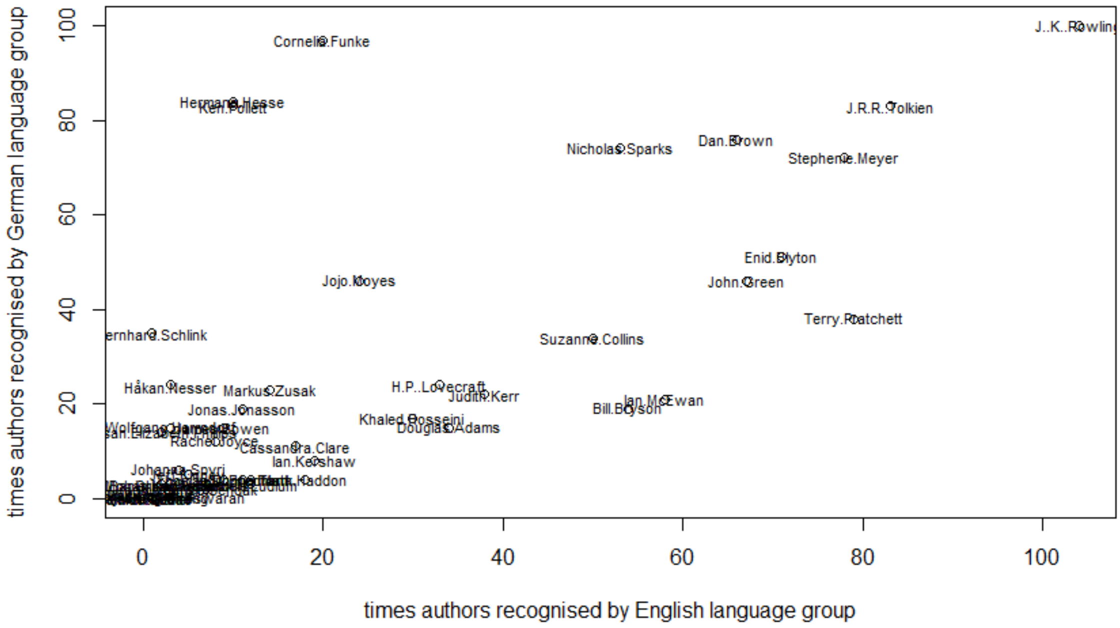


Figure 6.2. Scatterplot showing times authors were recognised by German and English readers

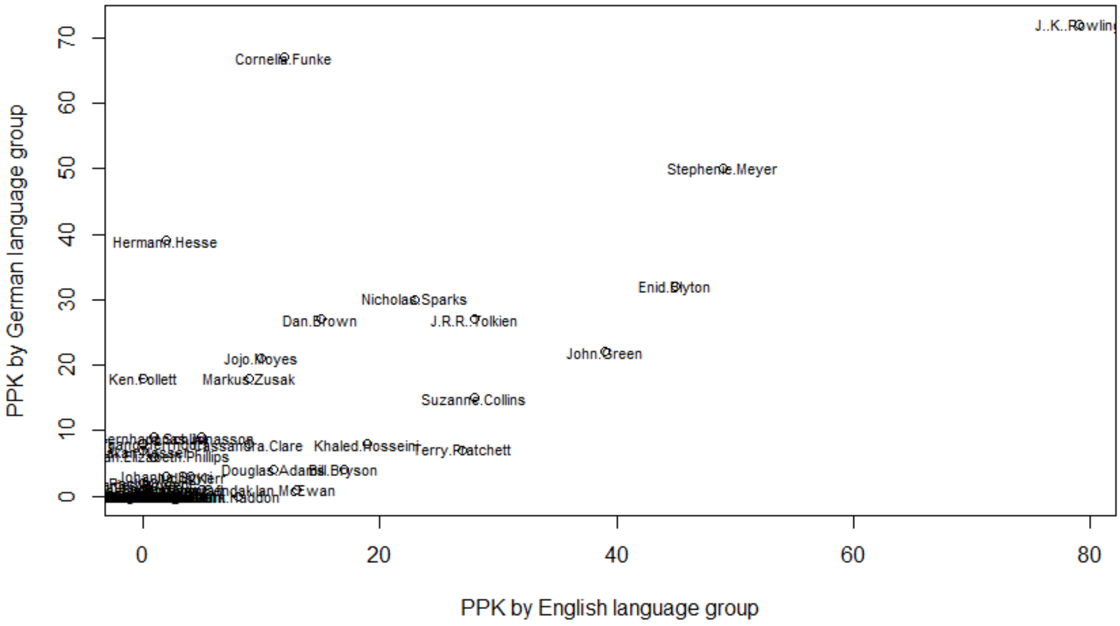


Figure 6.3. Authors whose work has been read by participants in both language groups (PPK scores).

6.1.3 Discussion

The GE-ART presented in this study was employed to extract three measures of print exposure, namely 1) the number of authors recognised, 2) the number of authors whose work had been read, and 3) the SW89 – ART score which corrected for guessing by subtracting the number of falsely recognised foils from the number of correctly identified authors. All three measures seemed to perform well in capturing reading experience variation across two language groups. Not only were the results very close to those reported by previous ART results, but they were also very similar across the two language groups. High internal reliability was evident for number of authors recognised, and for PPK scores in both language sets. These results are reassuring and support the usage of these measures in order to index reading experience in the current study.

6.2 Reading questionnaire

The previous section showed that scores from the GE-ART seemed very comparable across the two languages, and could thus be used as a measure for print exposure across both languages. However, it still remains to be seen, which one of the scores available (number of authors recognised, ART-SW89 or PPK) better captures print exposure and should thus be employed as a print exposure measure in the naming study. To compare and investigate the measures' validity, the present study sought to compare the GE-ART results with participants' reported reading habits. Even though there is a recognised general tendency for people to describe themselves approximating the socially desirable norm (e.g., Krueger, 1998), self-reported reading behaviour remains one of the most important validating measures of the ART. In fact, as can be seen in Table 6.1, positive associations of self-report measures and the ART in previous studies were typically of a moderate to large size. These self-report measures included different types of questionnaires. Reading and media habits questionnaires typically required participants to indicate how much they engaged in reading-related and media-related activities. Activity preference questionnaires asked participants forced choice questions to select the preferred activity from a choice of reading-related and not-reading-related activities. The favourite author question was part of the reading and media habits questionnaires, which inquired into participants' two favourite authors.

Table 6.1

Correlations between the author recognition test (ART) and self-report measures of reading.

Study	Variable	*N*	ART
Stanovich & West (1989), Study 1	reading habits questionnaire	61	.38
	favourite author question		.33
Stanovich & Cunningham (1992)	activity preference	300	.42
	reading habits questionnaire		.36
	favourite author question		.44
Stanovich, West, & Harrison (1995)	activity preference	133	.50
	Reading habits composite		.46
Martin-Chang and Gould (2008)	activity preference	171	.41
Acheson, Wells, & MacDonald (2008)	comparative reading habits questionnaire (CRH)	99	.44

Note . All correlations reported were significant. ART was scored according to Stanovich & West (1989).

As the main aim of the present investigation was a cross-language comparison, and given that reading habits and related behaviour are subject to change over time, it seemed sensible to create a new, updated reading questionnaire which would be suitable for both language groups. In the following, the preparation of the new reading questionnaire (GE-RQ) will be described. This will be followed by an investigation into which one of the three GE-ART measures is most suitable for measuring print exposure.

6.2.1 Methods

For the purpose of this study, the Reading and Media Habits Questionnaire (Stanovich & West, 1989) was shortened and modified. Specifically, the present reading questionnaire GE-ART included four of the 12 original questions in Stanovich & West (1989) (‘read for pleasure’, ‘how many books’, ‘library membership’, ‘visit bookshop’), and two further questions (‘favourite author’, ‘read newspaper’) were modified. Four new questions were added. These modifications will be described in more detail below.

Please note, that there were six questions in the original Stanovich & West (1989) questionnaire, which were not included in the present version. Four of the excluded questions probed television watching habits. Media-related questions had also been omitted in questionnaire versions by Stanovich & Cunningham (1992) and Stanovich, West, & Harrison (1995). Another question which was dropped inquired into the name of the last bookshop visited. It was felt that with the rise of online book-buying, this question would no longer be relevant. Finally, the question on magazine subscription was deleted, because it was found that short text reading was not associated with reading skills (Anderson et al. 1988, Spear-Swerling, Brucker & Alfano, 2010; McGeown et al. 2014). Please note, however, that the question on frequency of newspaper reading ('read newspaper') was retained, because arguably this could be regarded as an extended text.

Two questions were added to tap into online reading. The first gathered information on the frequency of downloading free e-books (as an alternative to a library visit) and the online purchase of books (as an alternative to visiting bookshops). The original question on the frequency of newspaper reading was also adapted to include online access to newspapers. As the internet had not yet been available at the time of creation of the Stanovich & West questionnaire, it seemed that this development should be reflected in the current questionnaire by including questions on online reading.

A new question was added to complement the reading questionnaire, asking the participant to indicate whether they carried a book with them at the moment that they were reading for pleasure and if they could indicate the author's name. It seemed plausible that prolific readers were more likely to carry their pleasure reads with them.

The question to name the two favourite authors was modified to ask for three favourite authors instead, thereby following Stanovich & Cunningham (1992). Greater possibility to expand on reading preferences would make the measure more sensitive to discrimination between more avid readers.

The Stanovich & West questionnaire did not directly ask participants to report on how much they read, presumably to avoid the social desirability confound. However, there is some evidence that self-report does not yield less reliable results than more objective measures. Schulte-Körne, Deimel & Remschmidt (1997) investigated whether self-report measures for spelling and reading ability would yield similar results to individual psychometric tests measuring spelling (dis)ability. Their participants completed a 12-point questionnaire based on Finucci et al. (1984), which asked participants to report spelling and reading difficulties, their attitude towards reading and spelling and to report on reading habits. They found that self-report of spelling ability (good, not so good, poor and very poor) was predictive of spelling ability. The current questionnaire was therefore extended to include a direct self-report question on frequency of reading. On a 7- point Likert scale, participants could choose to estimate their reading behaviour ranging from “I avoid reading” to “I read all the time”. The reliability of this direct self-report measure was enhanced by reiterating the need for a truthful response in the questionnaire instructions. Gordon (1987) suggested that stressing the importance and need for honest responses reduced the social desirability bias in surveys. Accordingly, written instructions to participants were extended with the request to answer as accurately and honestly as possible. This instruction was repeated on the second page of the questionnaire.

Appendices 11.8 and 11.9 include the final version of the modified GE - Reading Questionnaire in English and German, respectively.

Responses to questionnaire questions were scored so that more points were credited for answers which indicated more reading. For example, responses to the question ‘I read for pleasure’ were attributed 1 point for the answer ‘almost never’ and 5 points for the answer ‘once or more a day’. In the GE-RQ, a total number of 46 points could be attained.

6.2.2 Results

6.2.2.1 Descriptive summary

Participants in both groups received similar average score results (English language group: *M* = 24.89 (*SD* = 6.88, min = 10, max = 42) points, German language group: *M* = 26.35 (*SD* = 4.82, min = 12, max = 38) points). Figure 6.4 shows the density plots of the summed reading questionnaire scores per language group. There were slightly heavier tails to the distribution of the English language group indicating more participants scoring in the higher and the lower ranges. However, kurtosis and skew values were within generally accepted levels (English language group: skew = 0.16, kurtosis = -0.57; German language group: skew = -0.41, kurtosis = 0.36).

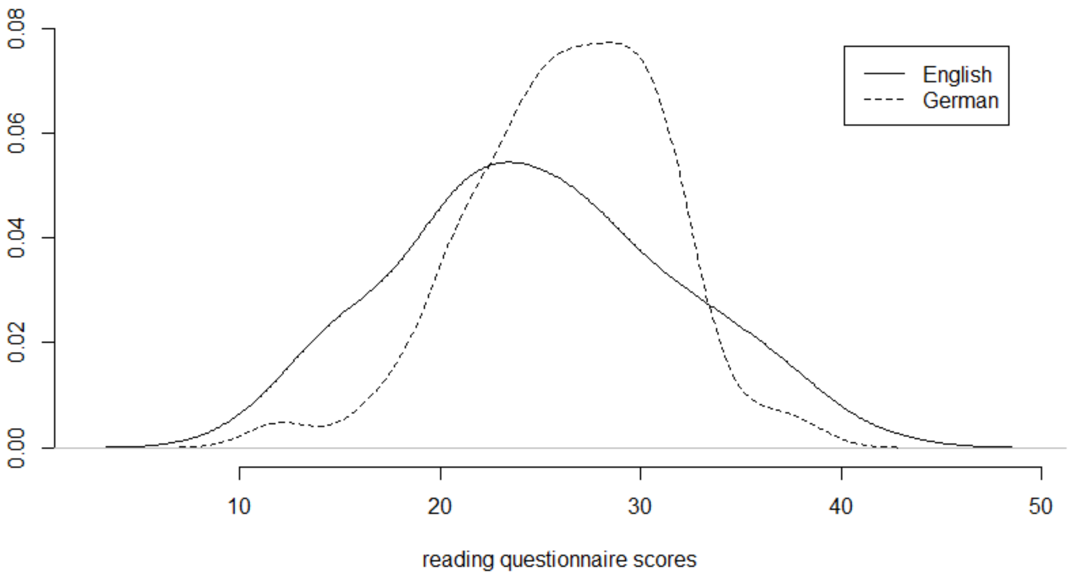


Figure 6.4. Density plots for summed reading questionnaire scores per language groups.

Table 6.2 gives an overview of the average, minimal and maximal points per question attained for each language group. It can be seen that minimal and maximal points were scored in almost all questions for both language groups. This indicates that the questionnaire was able to capture a very broad spectrum of reading behaviour. Exceptions were only found in the German language group, where no-one had chosen the highest option for 'buy online books' and the lowest option for 'self-reported reading'.

Table 6.2

Average GE-Reading Questionnaire scores per question for each language group.

		English language group				German language group			
GE-Reading Questionnaire questions	GE scores	*M*	*SD*	min	max	*M*	*SD*	min	max
read for pleasure	5	3.23	1.28	1	5	3.75	1.15	1	5
how many books per year	5	2.93	1.04	1	5	3.08	0.86	1	5
library membership	3	1.61	0.70	1	3	1.63	0.70	1	3
download free ebooks	5	1.68	0.97	1	5	1.46	0.81	1	5
visit bookshop	5	2.39	0.96	1	5	2.56	0.85	1	5
buy books online	5	2.10	0.88	1	5	1.86	0.67	1	4
three favourite authors	3	2.08	1.12	0	3	2.13	1.01	0	3
read newspaper	5	2.66	0.96	1	5	2.57	0.93	1	5
self-reported reading	7	4.48	1.66	1	7	5.36	1.34	2	7
currently carry book	3	1.74	0.90	1	3	1.96	0.92	1	3

Note . GE scores = maximal number of scores available in GE-Reading Questionnaire per question.

6.2.2.2 Correlation results

Strong moderate to strong positive correlations were found between GE-RQ and the three GE-ART scores (see Table 6.3), suggesting that the GE-ART captured reading experience in both languages. Of the three GE-ART scores, the PPK scores were most strongly associated with GE-RQ scores. This suggests that PPK scores should be an adequate index for reading experience for the present study.

It is noteworthy, that the associations between GE-RQ and GE-ART scores were even somewhat stronger than comparable associations presented by previous studies (see Table 6.1). On a speculative note, it may be that the plea for honest responses on the questionnaire reduced participants' tendency to answer in a way that was more socially desirable.

Table 6.3

Spearman correlations between GE-ART scores and GE-Reading Questionnaire scores

	English	German
authors recognised	0.51***	0.47***
PPK	0.65***	0.57***
ART (S&W 1989)	0.51***	0.45***

Note. * p < .05, ** p < . 01, *** p < .001. Authors recognised = number of authors recognised, PPK = number of authors read, i.e. primary print knowledge according to Martin-Chang & Gould (2008); ART = number of authors recognised minus number of foils checked, original ART scoring method according to Stanovich & West (1989).

Table 6.4 shows the correlations between the individual questions of the GE-RQ as a combined data set and per language group. Almost all correlations were positive, indicating that generally, higher scores on one question indicated higher scores on other questions in the questionnaire. The strongest correlation for all three samples was between 'reading for pleasure' and 'how many books read per year'. Interestingly, an almost similarly strong positive correlation was found between ‘self-reported reading’ and ‘reading for pleasure’ for the English language group. This was somewhat lower for the German language group. Newspaper reading seemed least associated with all other questions in the questionnaire in both language groups.

There were some interesting differences between the language groups. First of all, in the German language sample, intercorrelations generally seemed lower and fewer. Whilst ‘reading for pleasure’ and ‘how many books read per year’ were positively correlated with all other questions apart from ‘newspapers reading’ in the English set, there were also no correlations found with ‘library membership’ and ‘downloading free e-books’ in the German language group.

Another difference seemed to emerge in relation to self-reported reading. For the English language set, this did not correlate with ‘newspaper reading’, but did so for the German set. Conversely, self-reported reading did not correlate with ‘downloading free e-books’ and the ‘favourite author’ questions for the German language group, but did so for the English language group. Taken together, these differences may indicate that the question about self-reported reading was interpreted more widely by the German language group to include all kind of reading (including newspaper reading), whilst for the English language group understood the question to be more focused on literature reading.

Finally, the question ‘currently carry book’, which hypothesised that readers were more likely than non-readers to carry reading material with them, correlated with all but ‘library membership’, ‘free e-books’ and ‘newspaper reading’ in the English language group. In contrast, it correlated only with ‘reading for pleasure’ and the ‘self-reported reading’ for the German language group.

Table 6.4
Polychoric correlations of ten reading questionnaire questions.

	Variable	1	2	3	4	5	6	7	8	9
	Combined group									
1	read for pleasure									
2	how many books per year	.67***								
3	library membership	.13	.25***							
4	download free e-books	.26***	.37***	.1						
5	visit bookshop	.38***	.38***	.17*	.14*					
6	buy books online	.37***	.46***	.08	.54***	.27***				
7	three favourite authors	.44***	.44***	.18*	.14*	.31***	.25***			
8	read newspaper	.08	-	.12	.03	.15*	-.03	.02		
9	self-reported reading	.63***	.55***	.23***	.27***	.44***	.37***	.34***	.19**	
10	currently carry book	.43***	.33***	.13	.12	.1	.18**	.21**	.04	.31***
	English language group									
1	read for pleasure									
2	how many books per year	.74***								
3	library membership	.24*	.35***							
4	download free e-books	.40***	.52***	.19						
5	visit bookshop	.47***	.49***	.22*	.18					
6	buy books online	.44***	.55***	.13	.64***	.30**				
7	three favourite authors	.56***	.56***	.27**	.27**	.48***	.34***			
8	read newspaper	.06	-	-	-	.17	-.02	.08		
9	self-reported reading	.72***	.66***	.26**	.44***	.56***	.54***	.53***	.17	
10	currently carry book	.49***	.45***	.14	.14	.26**	.25**	.29**	-.02	.35***
	German language group									
1	read for pleasure									
2	how many books per year	.56***								
3	library membership	-.01	.13							
4	download free e-books	.16	.19	.01						
5	visit bookshop	.24*	.24*	.12	.12					
6	buy books online	.38***	.37***	.03	.35***	.28**				
7	three favourite authors	.29**	.27**	.07	-.02	.1	.14			
8	read newspaper	.13	.02	.25*	.05	.14	-.05	-.05		
9	self-reported reading	.45***	.40***	.19*	.14	.25**	.27**	.08	.27**	
10	currently carry book	.33***	.19	.11	.13	-.1	' .15	.12	.13	.23*

Note . * p < .05, ** p < .01, *** p < .001

6.2.2.3 Factor Analysis

These differences in correlation patterns between language groups are not easily explained, and it seemed necessary to investigate whether the questionnaire had tapped into different factors in the two language groups. An exploratory factor analysis was employed to explore the underlying factors of the reading questionnaire, and to determine if these factors correlated with the three ART scores. This was done for each language group separately, and also jointly. Given that the questionnaire was exactly the same for both language groups, it would be expected that it would probe the same reading behaviour for both language groups. However, given the different intercorrelation patterns between the ten questions, this was called into question.

Using the psych package (Revelle, 2017) an exploratory factor analysis using the weighted least square method with varimax rotation on all three groups (combined language groups and each language group separately) was performed. For all three groups two underlying factors were extracted. The factor loadings are reported in Table 6.5.

Although the joint and the respective language groups supported a two-factor structure, the factors seemed to be somewhat different between the two language groups. For the English language group, there seemed to be a main factor tapping into general reading, and another which seemed to have a digital reading component. Notably, ‘newspaper reading’ had the lowest loadings on both factors.

For the German language group, there also seemed to be a general reading factor. As with the English data set, it included ‘read for pleasure’ and ‘how many books read per year’. However, ‘buy books online’ and ‘self-reported reading’ also loaded heavily on this general factor. The second factor included ‘self-reported reading’, and further included ‘newspaper reading’, and ‘library membership’. This combination seemed to represent non-fiction or possibly University course reading.

In sum, whereas the English group showed a general reading factor and a second digital reading factor, the German language group seemed to have a second factor which was less clearly defined, but possibly tapped into non-fiction reading. The general factor did not have exactly the same loadings in both language groups.

Table 6.5

Factor loadings for exploratory factor analysis of the reading questionnaire GE-RQ for the combined sample, the English language group, and the German language group.

Variables		Factor loadings					
		Combined group		English language group		German language group	
		Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
1	read for pleasure	**.79**	.35	**.82**	.33	**.79**	.24
2	how many books per year	**.70**	**.48**	**.76**	**.47**	**.75**	.16
3	library membership	.35	.05	.37	.14	.05	**.42**
4	download free ebooks	.14	**.69**	.27	**.85**	.34	.02
5	visit bookshop	**.52**	.19	**.71**	.05	.34	.15
6	buy books online	.24	**.78**	.37	**.72**	**.65**	-.02
7	three favourite authors	**.55**	.22	**.77**	.21	.34	-.02
8	read newspaper	.22	-.12	.18	-.15	-.06	**.73**
9	self-reported reading	**.70**	.31	**.75**	.37	**.47**	**.49**
10	currently carry book	**.45**	.19	**.51**	.20	.31	.26

Note. Factor loadings > .4 are in bold.

In order to investigate how the different factors correlated with the three different GE-ART scores, the latter were converted into z-scores. Table 6.6 shows the correlation coefficients for the relationship between the three types of GE-ART scores and the factors which emerged from the exploratory factor analysis for the combined set of participants, and for each language group separately. Whereas the GE-ART scores all correlated strongly and positively with the first factor in both the combined set and the English set, the GE-ART scores from the German set correlated with both factors. This suggests that the reading behaviour that the questionnaire tapped into, is somewhat different for the German language group than it is for the English language group.

Table 6.6

Pearson correlations between scaled ART scores and the factors underlying the reading questionnaire.

Variable (z-scores)	Correlation coefficients					
	Combined group		English language group		German language group	
	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
authors recognised	.53***	.08	.59***	-.1	.40***	.36***
PPK	.60***	.20**	.67***	.04	.49***	.38***
ART (S&W 1989)	.53***	.04	.59***	-.14	.41***	.37***

Note . * p < .05, ** p < . 01, *** p < .001. Variables are z-scores. Authors recognised = number of authors recognised, PPK = number of authors read, i.e. primary print knowledge according to Martin-Chang & Gould (2008); ART = number of authors recognised minus number of foils checked, original ART scoring method according to Stanovich & West (1989).

6.2.3 Discussion

An adapted version of the Stanovich & West (1989) Reading Habits Questionnaire, the present GE-Reading Questionnaire was used to compare the results to the three GE-ART scores with the two aims to establish whether the GE-ART captured reading experience, and which one of the three GE-ART measures available would be most successful at doing so. Correlation results between GE-RQ results and GE-ART measures qualified the GE-ART as a valid measure for print exposure. Importantly, outcomes were very similar across the two language groups. PPK scores (number of authors whose work had been read) were most closely related to GE-RQ results, suggesting that print exposure as measured by the GE-ART should be based on PPK scores.

Factor analysis of the GE-RQ results for each language group showed that factor loadings followed different patterns for each language. Although two factors emerged for both language groups, there was no single common factor with the same loadings for both language groups. Additionally, the second factor did not correlate at all with GE-ART scores for the English language set, but did so for the German language set. This means that the GE-RQ tapped into different reading behaviours in the two language groups, despite having the same wording (albeit in translation). This may be because the GE-RQ is more culture-sensitive than the GE-ART. This is plausible because reading behaviour may vary within culture groups. However, it may also be the case that written questions (as a defining feature of any questionnaire) are more prone to interpretation within a cultural context which could result in different answers for different groups. It is not possible for the present study to investigate this further. However, it can be concluded that the GE-RQ should be more useful for studies exploring differences rather than studies which aim to capture similar reading experiences of different groups.

This notwithstanding, the high correlations between the GE-ART scores and the first factor in the English set, and both factors in the German set still validates the GE-ART as a measure of print exposure. Of the three measures investigated, the PPK was most strongly related to general reading behaviour, thereby supporting the decision to employ the GE-ART PPK measure as the index measure for print exposure in the current investigation.

The next chapter will report the naming study of the two language groups. Both groups will have also completed the test battery for individual differences. The PPK score of the GE-ART was used as a measure for print exposure.

7 Cross-language comparison study

7.1 Introduction

Languages which differ to which degree spelling-sound relationships have been expressed in their writing system may afford differential reading processes. This may involve more prelexical processing for more transparent scripts, as suggested by the ODH (Frost et al., 1987; Katz & Frost, 1992). Alternatively, the PGST (Ziegler & Goswami, 2005) proposes that reading unit size will vary with language spelling-sound consistency. More opaque scripts will necessitate larger grain sizes to achieve correct reading. This is particularly true for beginner readers. As the grain size for achieving accurate reading adapts with increased reading experience and accrued knowledge of spelling-sound relationships, the reading system will aim to optimise the size of the reading units to suit the language's relative transparency. Thus, readers of transparent orthographies will rely on small reading units, as these are conceived as reliable reading unit sizes, whereas readers of opaque languages like English will find that they need to employ larger units, varying in size, to achieve accurate reading.

Whilst behavioural studies have found support for these assumptions for developing readers, results are less clear for skilled readers. A seminal study found evidence for more small-unit processing as indexed by the length effect for readers of transparent German, and for relatively more large-unit processing as marked by the body-*N* effect for opaque English (Ziegler et al., 2001). However, other studies were not able to replicate a body-*N* effect for identical English items (Rau et al., 2015; Schmalz et al., 2017), or reported the opposite pattern for the length effect, with English readers exhibiting a greater length effect in nonword reading than German readers (Rau et al., 2015). The size of reading units may depend on more than the language you learn to read. Schmalz et al. (2014), for example, observed great variation in unit sizes used for reading nonwords within a language group and suggested that individual differences should be further explored.

There have been calls to move away from English-language centred reading research (Share, 2008) and also to more fully integrate individual differences (additionally to language) into reading models (Rueckl, 2016). In fact, there has been an immense effort to understand the implications of individual differences in reading. Early investigations focussed on exploring the possibility of sublexical versus lexical readers, but were not able to capture reading differences within the general reading population (Baron & Strawson, 1976; P. Brown et al., 1994). However, there has been the repeated observation of a positive correlation between faster nonword and word reading (e.g., Brown et al.; 1994, Davies et al., 2017; Stanovich & West, 1989; Torgesen et al., 1999; Wagner & Torgesen, 1987). Hence, even though it seems that people cannot be categorised into either lexical or sublexical readers, it appears that individuals differ in terms of their nonword reading (decoding) ability, and that better decoding skills lead to faster reading aloud reaction times.

Within the connectionist triangle model (Plaut et al., 1996), the division of labour between the direct (phonological) and the indirect (semantic) pathway may vary between people. It has been suggested that greater use of semantic knowledge in reading aloud is in fact driven by a deficit in phonological processing as indexed by nonword reading ability (Strain & Herdman, 1999; Woollams et al., 2016). This assumption seems to agree with the finding that faster decoding is also associated with faster word reading (e.g., Brown et al., 1994; Davies et al., 2017; Torgesen et al., 1999), and extends the importance of decoding skill to skilled readers, and beyond its role as a vital prerequisite for successful reading acquisition (Share, 1995). However, as people will vary in terms of their semantic knowledge, this should also be taken into account when investigating the division of labour. It is possible that people with greater semantic knowledge may be more prone to receive contributions from the semantic pathway, or alternatively, that these contributions are more easily measurable than in individuals with less semantic knowledge.

Within the connectionist model, reading practice is important to becoming a skilled reader. Participants with more print exposure produce faster naming times (Chateau & Jared, 2000; Stanovich & West, 1989), and have shown attenuated lexical and sublexical psycholinguistic effects (Butler & Hains, 1979; Yap et al., 2012). In computational parallel distributed reading aloud models, reading experience is an integral part in the form of network training (A. W. Ellis & Lambon Ralph, 2000; Plaut et al., 1996), and simulations have successfully shown that differential reading experience can shape the reading system (e.g., P. Monaghan et al., 2017).

Each individual strength or weakness will add to the individuality of each reading system. Dilkina et al. (2008) reported results from a computational model of data from semantic dementia patients differing greatly in how far their reading was impaired in addition to their semantic deficit. The authors hypothesised that the variability shown in the patients would be imitable by a single computational reading system, but which was able to vary the size of the direct pathway (to mimic differential use of distribution of pathways), the training of the network (to model individual pre-morbid engagement with reading experience), and differences in the spatial distribution of the atrophy (to simulate differences in patients' lesions). Although the lesion bias was in fact the strongest predictor, network training and to a lesser degree direct pathway size predicted variance in the reading deficit above and beyond other predictors when semantic impairment was controlled. Thus, the model simulations suggested that when semantic impairment was held constant, reading performance was more likely retained the smaller the lesion, the more reading experience and, with the smallest effect, the stronger the O-P pathway of the individual. Not only do these results point to differential reliance of semantics for individuals. They also suggest that for those individuals who have strong O-P pathways, and those who have a lot of reading practice, the employment of semantics in the reading process may be less essential.

Finally, it has been suggested that readers can also differ in terms of the quality of their orthographic representations. According to the Lexical Quality Hypothesis, less well-developed orthographic representations could contribute to differential reading performance between readers. Support for this assumption has come from research using spelling as a marker for orthographic knowledge. Better spelling has been shown to lead to faster naming performance (Martin-Chang et al., 2014).

The first aim of the present study was to investigate whether differences in reading aloud in readers of a spelling-sound consistent language versus readers of an inconsistent language were still apparent when reading-related individual differences were taken into account. The current study thus aims to combine inter-language and inter-individual reading research. Specifically, the present study will compare naming performance by English and German skilled readers for a set of cognates. English and German lend themselves well as examples of less and more spelling-sound transparent languages, whilst sharing a number of words which are identical in form and meaning (Ziegler et al. 2001). The present study will also take into account key individual differences relating to reading performance. For this purpose, participants were asked to complete a set of tasks to measure naming latencies, decoding ability, vocabulary knowledge, print exposure, and spelling ability, tapping into the four central aspects of reading: phonology, semantics, orthography and reading practice. Decoding ability was tested to gauge the phonological efficiency of the direct route. Importantly, it has also been proposed that inefficient functioning of the direct pathway results in increased use of semantics in reading. The current study will also take into account participants' semantic knowledge as indexed by a vocabulary score. Better semantic knowledge is likely to yield faster access to semantics. Moreover, print exposure measures were designed to tap into participants' reading experience, which is likely to improve general reading processes. Finally, spelling ability was used to test the quality of the orthographic representations.

7.2 Method

7.2.1 Participants

Originally, 109 English speakers and 106 German speakers took part in the study. However, five participants were excluded from the English and two from the German sample (see details in section 7.2.1.2), leaving 104 participants in both language groups.

7.2.1.1 Data collection

Data was collected in Germany between October 2015 and January 2016 at the University of Potsdam, and in the UK at Oxford Brookes University between November 2015 and November 2016. All participants were aged between 18 and 25 years and reported to have no history of dyslexia. They received an Amazon voucher or course credit for participation. The Oxford Brookes Ethics committee gave ethical approval for the implementation of the study in both countries based on the letter of invitation from the German host that the ethics approval from the UK University was sufficient for testing in Germany (UREC registration no. 130730). Informed consent was received from all participants tested within the framework of this study, and responses are stored securely and separately from data which could identify individual contributors.

Initially, participants were asked to fill in a short questionnaire about general demographics (see appendix 11.4 for the two language versions of the demographics questionnaire). One participant from the English language group did not disclose their age. This participant was attributed the age that was most often given by participants who were in the same year of study. Participants were also asked to fill in information about their gender, and whether they had been raised multi-lingually. Participants were asked for both their current year of study and their current semester of study. In the German questionnaire, the phrasing of the question about the current year of study seemed not sufficiently clear.

For this reason, where the answers to current year of study and current semester did diverge, the current semester was considered the correct answer.

Following the completion of the demographics questionnaire, participants were asked to complete the naming task. The naming task was always the first task to be completed, except on two occasions when for technical reasons the ART was completed before naming with one English and one German participant. The order of test presentation of the individual differences tasks which followed was varied randomly.

7.2.1.2 Data cleaning

Any participant who was found to be an outlier in one task, was subsequently excluded from the whole study. This was to ensure that all participants had taken part in all tasks to ensure transparency and consistency in the subsequent naming study analysis.

In the UK, a total of 109 young people participated. Initially, three participants were excluded. In two cases this was due to experimenter error. In another case, it was not sufficiently clear if the participant could be classified as a native speaker.

Data were then screened to identify any outliers. In the naming data, no responses faster than 200 msec were found. Having removed all errors from the data, all data points 2.5 *SD* above or below the mean were excluded. Mean naming reaction times were then computed for each participant for this trimmed data set. One participant showed an average reaction time above 2.5 *SD* of the trimmed data, and was consequently removed from the data set. The data sets for the spelling test, the vocabulary test and the TOWRE nonword reading test were also screened for outliers. However, in the case of these tasks, only those whose mean performance was 2.5 *SD* below the mean were considered outliers, to avoid a ceiling in terms of task performance. In other words, outliers who performed 2.5 *SD* above the mean on these three tasks remained in the data set. Only in the nonword decoding task, one participant fulfilled the exclusion criteria of performing below 2.5 *SD* of the mean and was subsequently excluded from the study. This resulted in a total of 104 English-speaking participants.

In Germany, 106 young adults participated. One erroneously participated twice and the second recording of data was therefore removed from the dataset. Analogous to the English data set, the data was screened for outliers. There were no reaction times shorter than 200 msec. One further participant was excluded because the mean naming reaction time was more than 2.5 *SD* above the mean of the trimmed data set. This resulted in 104 German-speaking participants.

7.2.1.3 Participant descriptives

For the English language group, 104 young people between 18 – 25 years were included in the final analysis. The mean age was 20.74 years (*SD* = 1.88, *Mdn* = 21). As English students tend to be a little younger than their German counterparts, six young people within the age range who had completed their studies within the last three years were also included. The other 97 were current full-time students and one participant studied on a part-time basis. The majority (34 participants, or 33%) studied in their first year. For those participants who had already completed their studies, another year of study for each year that they had left university was added. Thus, the distribution was as follows: 14 (13%) in Year 2, 25 (24%) in Year 3, 15 (14%) in Year 4, 14 (13 %) in Year 5, and 1 (< 1 %) each in Year 6 and Year 7. Overall, 82 were female, 21 male and 1 other. Fourteen participants reported to have been raised with another language other than English, of which four with two additional languages.

The 104 German speaking participants were all students aged between 18 and 25 years (*M* = 21.63, *SD* = 2.15, *Mdn* = 22 years). Figure 7.1 shows the age distribution of both English and German participants. The English language group tended to have higher numbers in the lower age groups, whereas there were more equal numbers of participants across year groups in the German language group. This age difference between the two language groups was significantly different (Wilcoxon rank sum test, $W = 4128$, $p = .003$). Of the German speakers, 24 (23%) were in Year 1, 25 (24% in Year2, 17 (16%) in Year 3, 21 (20%) in Year 4, 10 (9%) in Year 5, 5 (5%) in Year 6 and 2 (almost 2%) in Year 7. The last three year groups were regrouped, and a chi-squared test showed, that there was no significant association between year of study and language group ($\chi^2 = 7.3808$, $df = 4$, $p = .117$). Of the German speakers, 85 were female and 19 were male. Although only four of the German speakers had grown up with an additional language, including one with two additional languages, there was no significant association between language group and multilingualism ($\chi^2 = 4.9263$, $df = 1$, $p = .026$). However, although this was not specifically assessed, it needs to be noted that on account of the general curricula in both countries, German participants will have been more likely to have knowledge of English, than English participants knowledge of German.

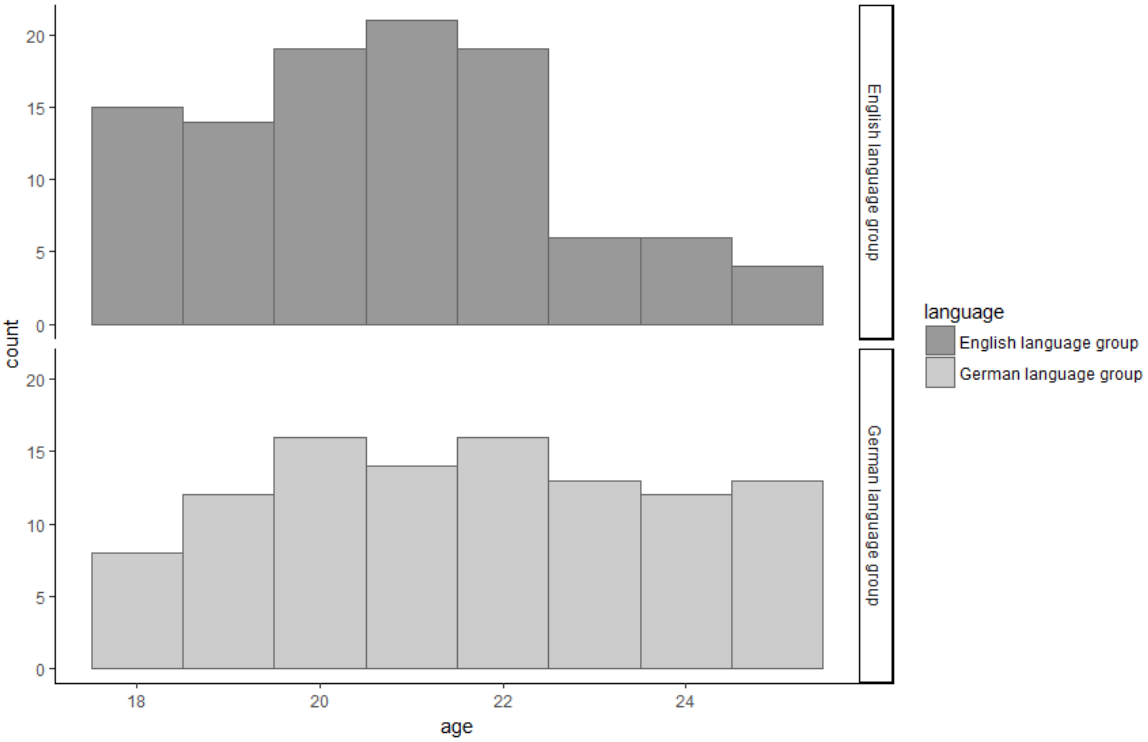


Figure 7.1. Participants' age distribution according to language group.

7.2.2 Naming task

7.2.2.1 Materials and stimuli

For the naming task, 301 German and 326 English words were selected from two pools of words for which a large number of psycholinguistic variables were known (see appendices 11.10 for English and 11.11 for German word stimuli). Of these, there was a small subgroup of 89 cognate pairs (see appendix 11.12). The other word stimuli were part of three further subsets matched across languages for future analyses. There were fewer German than English items (301 vs 326) because some German items belonged to more than one subgroup, and therefore the overall number of German stimuli was lower. However, the present study will only report on the results of the cognate subgroups, although participants completed the naming task on the complete sets of 301 German or 326 English words, respectively.

The following psycholinguistic variables were available for the complete word sets. For the summary descriptive statistics of the final cognate set used for the present analysis please refer ahead to section 7.2.2.5.

7.2.2.1.1 Onsets

Onsets were categorised into 15 categories as in previous studies ((Balota et al., 2004; Spieler & Balota, 1997; Treiman et al., 1995). These categories were affricative, alveolar/postalveolar, bilabial, dental, fricative, glottal, labiodental, approximant, nasal, palatal, plosive, velar, and voiced. By adding short vowel and long vowel/diphthong, two vowel categories for those words without an onset were also added. The phonemes of onsets have been shown to explain a large proportion of variance in reading aloud, for example (R^2 = .299 in Spieler & Balota, 1997), and R^2 = .28 for monosyllables and R^2 = .043 for multisyllabic words (Yap & Balota, 2009).

7.2.2.1.2 Stress

Stress was added as another categorical variable indicating stress on the first or second syllable. Monosyllabic words were automatically categorised as having the stress on their first (and only) syllable.

7.2.2.1.3 Frequency

Having been collected from subtitles of films and TV series in both languages, the subtitle frequencies SUBTLEX for English (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014) and German (Brysbaert et al., 2011) seemed the most similar frequency word counts across the two languages. Frequencies were subsequently converted to Zipf frequencies, which are the log10 of the frequencies per million (fpmw) + 3. The transformation to Zipf makes it easier to include words with very low frequencies, as their frequency value will be easier to interpret than a log10(fpmw) transformation which results in negative numbers for words with a frequency of lower than 1 fpmw (Van Heuven et al., 2014).

7.2.2.1.4 Age-of-acquisition

For English, AoA ratings compiled by Kuperman, Stadthagen-Gonzalez, & Brysbaert (2012) were used. For German, the ratings collected by Birchenough, Davies, & Connelly (2017) were employed. They had been collected specifically for this study and had used a similar collection and preparation method to Kuperman et al. (2012). Please see Appendix 11.17 for the published paper.

7.2.2.1.5 Word length

7.2.2.1.5.1 Number of syllables

For both languages, the number of phonological syllables were based on the syllable parsing for the consistency computations (see Chapter 5).

7.2.2.1.5.2 Number of letters and number of phonemes

Number of letters and phonemes were counted using R base software (R Core Team, 2014). Phonemes were expressed in DISC code, a single digit phonetic transcription developed and used by CELEX (Baayen et al., 1995). DISC has the advantage that every sound is expressed by a single and unique sign which makes it possible for computer software to easily identify it.

7.2.2.1.6 Onset and rime consistency

Consistency measures were based on the computations presented in this study (see Chapter 5). In total, there were 16 consistency measures: feedforward (FF) and feedback (FB) measures for each syllable onset and body-rime, as well as composite consistency measures which were averages across words for either FF or FB onsets or rimes.

7.2.2.1.7 Neighbourhood

7.2.2.1.7.1 Coltheart's N and old20

Orthographic distance measures (Coltheart's *N* and old20) were computed on the 27k words from the Clearpond database in German and English. For these computations in R, the vwr package (Keuleers, 2013) was used.

7.2.2.1.7.2 Body-N

Body-*N* counts were extracted from the data compiled for consistency computations. Body-*N* are the number of orthographic rime neighbours, which were used to compute consistency, minus 1 (the count for the word itself).

7.2.2.1.7.3 Phonographic neighbours

Phonographic *N* of the first rime (PNR1) were also extracted from the consistency computations. They are the number of words which are the same orthographically and phonologically minus 1 (the count for the word itself).

As this measure is new, it requires some prior consideration. As reviewed in section 3.4 of the Introduction, research has sought to capture the use of larger grain sizes by measuring the number of rime neighbours of words. Larger neighbourhood generally led to faster naming times. This was most prominent when the measure included neighbours which shared both orthography and phonology of the target word.

However, the present study employed not only monosyllabic stimuli, but also bisyllabic words. In the absence of a better measure, the present study attempted to capture the use of larger grain size units by using the number of phonographic neighbours of the first rime (PNR1) of the stimuli. The first rime was chosen simply because both mono- and bisyllabic words have a first rime. Nevertheless, it seems reasonable to explore the measure further in order to ascertain that PNR1 is likely to capture larger reading units for both mono- and bisyllabic words.

Within the current stimuli set, the English cognate set included 28 monosyllabic words and 57 bisyllabic words, whilst the German set comprised 24 monosyllabic words and 61 bisyllabic words. The length in letters of each rime can be gleaned from Table 7.1. It becomes clear that monosyllabic words have longer rimes in letters than bisyllabic words. For example, the bi-syllabic word *alarm* has a first rime length of 1 (the letter *a*), whilst the monosyllable *arm* has 3 letters as its rime. If rimes are the larger reading units to be captured, then the first rime of bisyllabic words does not seem to consist of large units.

Table 7.1.

Overview of length of first rimes for mono- and bisyllabic words in German and English

	Length in letters of first rime							
	German				English			
	1	2	3	4	1	2	3	4
monosyllabic words	-	2	21	1	-	2	24	2
bisyllabic words	46	15	-	-	41	16	-	-

Furthermore, all alphabetic languages have many more bisyllabic words than monosyllabic words. When counting rime neighbours then, counts for PNR1 of monosyllables should be much lower compared to those for bisyllabic words, thereby creating a bias for larger PNR1 for bisyllabic words.

Combined this means that if words have more than one syllable, they typically have shorter rimes (in letters) but a larger rime neighbourhood of the first rime (PNR1). Does the PNR1 measure then still reflect larger grain sizes for bisyllabic words? The author of the present study argues that for reading the rimes in bisyllabic words larger grain sizes are required, because the pronunciation needs to be ambiguated. Disambiguation can only occur, when some of the letters following the rime have been processed, thereby making the reading unit larger than the first rime. Thus, if rimes can be pronounced in more than one way, this should entail the processing of letters which follow the rime, in order to arrive at the correct pronunciation. In other words, the more pronunciations are possible, the larger a reading unit (above and beyond the rime) is required. Figure 7.2 shows the relationship of the PNR1 and the number of rime 1 pronunciations possible for the first rimes of the mono – and bisyllabic cognate stimuli for each language. The number of pronunciations possible was based on the database used for the consistency computations in Chapter 5. Rimes of monosyllables have lower PNR1, but also have less uncertainty about their pronunciation, as indicated by the lower number of pronunciations possible. First rimes of bisyllabic words tend to have a larger phonographic neighbourhood, but also more uncertainty about their pronunciation. This larger uncertainty would require larger reading units in order to be pronounced correctly. The positive relationship between PNR1 and number of possible pronunciations is true for both languages (polyserial correlation of .556 for German and .600 for English).

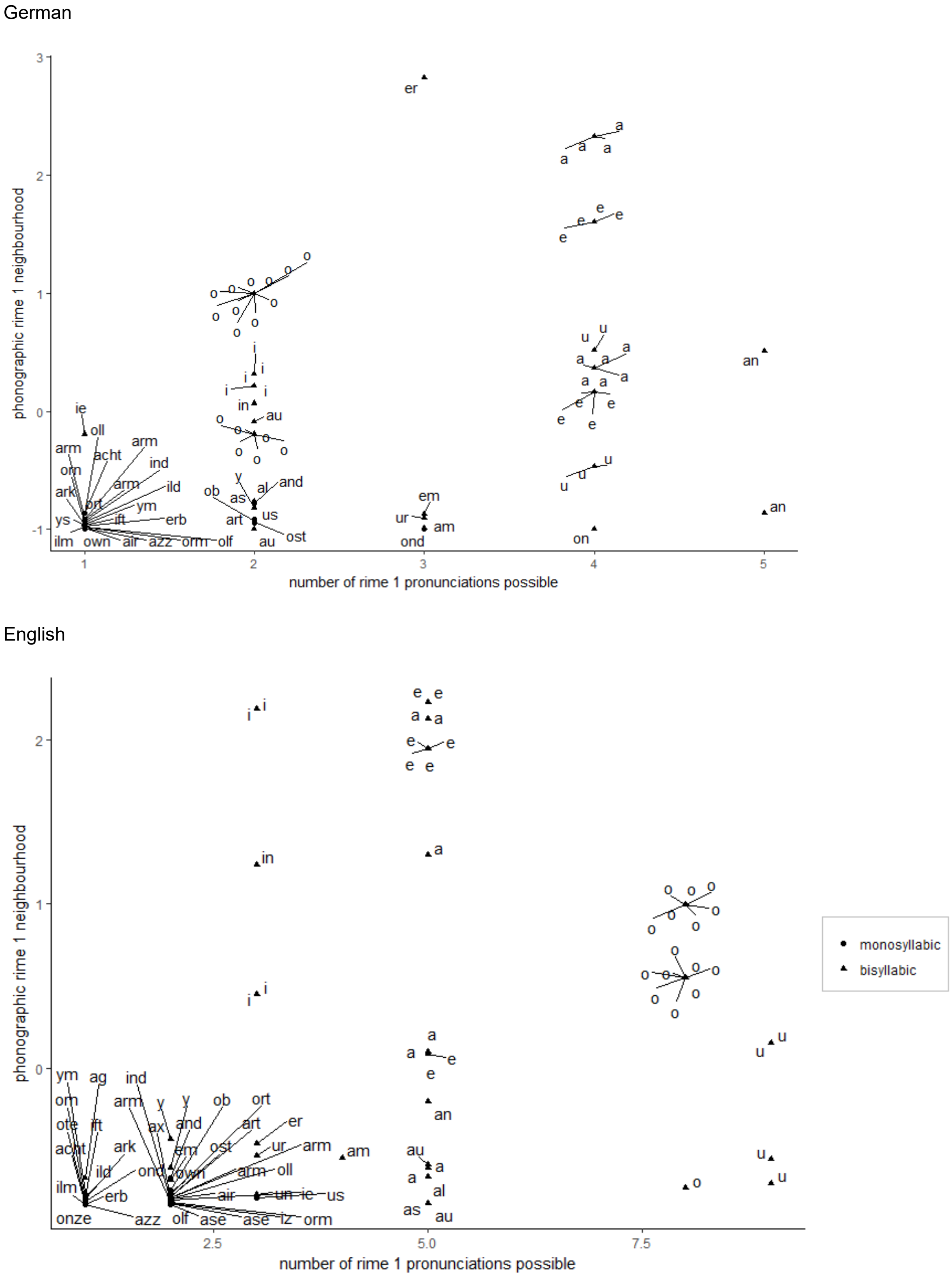


Figure 7.2. Scatterplot showing relationship between PNR1 and number of possible pronunciations for each first rime of monosyllabic- and bisyllabic words in German and English.

For bisyllabic words it thus seems that PNR1 captures larger grain sizes, because these shorter rimes require disambiguation. There is a positive relationship between syllable number and letter length for the current stimuli set (German: r_s = .58, English: r_s = .59), so bisyllabic words tend to be longer than monosyllabic words. This opens up the possibility that the grain sizes used for disambiguation could be much larger than what is possible for monosyllables. Of course, what we do not know is the size of these reading units. One possibility of conceptualising these larger reading units in first rimes of bisyllabic words is the sublexical unit BOSS or BOB, which comprises the first vowel and all subsequent consonants of polysyllabic words, and for which there has been some empirical support as a reading unit for polysyllabic words (e.g., Taft 1987, 1992).

Importantly, however, it needs to be noted that, whilst body-rime neighbours for monosyllabic words in previous work was understood as a facilitatory factor (larger rimes which are the same and occur more often), the current measure is slightly different for bisyllabic words. It implies the use of larger grain sizes for disambiguation, but it also implies more uncertainty for decoding. In that sense, it is not the same measure as phonographic rime neighbours for monosyllables. However, in the current study, they appear within the one measure (PNR1). In the absence of a better alternative, this measure can be used to index larger grain size usage, but future research may be able to find a measure which is better at identifying the use of larger reading units in both mono- and bisyllabic words.

7.2.2.2 Procedure naming task

Participants were seated in front of a Hewlett Packard Envy17 Laptop with a 17'' screen. DMDX (Forster & Forster, 2003) was used for stimuli presentation. Participants were told that words would appear in the middle of the screen and they were asked to read them aloud as quickly as possible. By pressing the space bar, participants started off the first word presentation block. A white blank screen appeared for 500 msec, followed by an asterisk for another 500 msec. Then the first item appeared and remained on screen for 2s. The next trial followed immediately afterwards in the same fashion.

Participants were offered a break after each block, and were free to press the space bar when they were ready to continue. The 301 German stimuli were randomly grouped into 9 blocks of 30 and 1 block of 31 stimuli. The 326 English word stimuli were randomly assigned to 10 blocks of 27 words and 2 blocks of 28 words. The order of presentation of the blocks and the words within the blocks were randomised. All stimuli were presented in Courier New font.

Responses were recorded using the head microphone Shure SM35 connected to a Focusrite Scarlett Solo preamp. For each response, voice onset was determined by the experimenter using CheckVocal (Protopapas, 2007).

English words were presented in small letters only, whilst in line with German spelling rules German words were presented with a capitalised first letter if they were nouns. As English nouns are not usually capitalised, one could argue that this presents a difference between languages in stimulus presentation method. However, previous cross-language studies have frequently presented German nouns with a capitalised first letter, whilst English counterpart stimuli were shown in lower-case. Thus, this format of presentation was employed for reading aloud and lexical decision tasks for children in English and German (Frith et al., 1998), children reading aloud in several European languages (Seymour et al., 2003) and a primed lexical decision task with adult participants in German (Sonnenstuhl, Eisenbeiss, & Clahsen, 1999). In fact, Jacobs, Nuerk, Graf, Braun, & Nazir (2008) presented evidence that German words were best recognized when presented in their most natural and frequent form. In visual identification tasks, German participants recognized nouns with an initial capital better than nouns presented in all lower-case and all upper-case. As such, whereas the format of presentation may appear to be different, in fact stimuli were presented in their most natural form in both languages, thereby ensuring comparability across languages.

7.2.2.3 Data cleaning for complete data sets

Data cleaning was accomplished for the responses recorded to the complete sets of stimuli, 301 for German and 326 for English, and is therefore reported for the complete sets of stimuli. Initially, there were 33,904 data points for 104 participants for 326 English word stimuli, and 31,304 data points for 104 participants for 301 German word stimuli. Four items were excluded from the complete data sets, because most oral responses did not coincide with the phonetic transcriptions the CELEX database used to compute spelling-sound consistency. Thus, in the English set, the word *ego* was excluded because it had only been pronounced as /ˈɛːɡəʊ/ (CELEX pronunciation) by 7 participants. Most other participants had pronounced it as /ˈiːɡəʊ/, which is of course also correct. However, these responses were not accepted, because for the purpose of the present study only CELEX pronunciations, which had been used to compute the spelling-sound consistency, were accepted. As the German partner word for *ego* in the other stimuli subsets was always *Ego*, the German word *Ego* was also excluded from the complete dataset, as it would not be used in the current or any future analyses. In the German data set, three more words were excluded, because the pronunciations given did mostly not conform with those in the CELEX database, which had been used in the present consistency computations: *Steak*, *Song* and *Chance* received the CELEX pronunciations only 4,5, and 9 times respectively. These word exclusions resulted in reduced data sets: 33,800 data points for 104 participants for 325 English words and 30,888 data points for 104 participants for 297 German words.

For both language naming data sets, error responses were excluded and treated as missing values. Responses were coded as errors if they included false starts, repetitions and misplaced stress, or if the response was an incorrect word or indeed a nonword. Equally, responses were considered errors when external noise, unclear speech or interjections made the latency difficult to determine. Finally, utterances which were unduly protracted were also excluded. All errors were categorised, and an overview can be found in Table 7.2. A brief discussion on the error data can be found in section 7.2.2.4.

7.2.2.3.1 English

Of 33,800 remaining data points, 1,408 were errors or non-responses (4.17%), which were removed, leaving 32,392 data points. A further 11 responses (0.033% of 33800 data points) were removed as the RTs were recorded at 2,000 msec. These responses were made to eleven different items by ten different participants, indicating that these outlier responses were not connected to a particular word stimulus or participant. They were categorised as hesitation errors. For the analysis, errors were treated as missing values (NAs).

All in all, of the 33,800 remaining datapoints, 32,381 were valid RTs and 1,419 (4.20 %) were missing values (NAs). The mean RT was 479.16 msec (*SD* = 103.92). The six items with the slowest mean reaction times were *tibia* (690.49 msec), *cello* (676.69 msec), *depot* (667.16 msec), *banshee* (608.82 msec), *chauffeur* (605.92 msec) and *opium* (602.83 msec). The six items with the most missing values were *depot* (68% NAs), *chauffeur* (65% NAs), *moped* (65% NAs), *psyche* (63% NAs), *vase* (51% NAs), and *hangar* (44% NAs).

7.2.2.3.2 German

From the remaining data with 30,888 datapoints for 297 stimuli, 889 data points were non-responses or errors. One further response was identified as having been recorded at 2,000 msec, and was thus also excluded. For the analysis, errors were treated as missing values (NAs).

The remaining dataset of 30,888 observations had 29,998 RTs and 890 missing values (2.88%). The average response time was $M = 514.09$ msec ($SD = 88.78$). The words with the slowest average reaction times were *Jockey* (697.65 msec), *Cello* (675.06 msec), *Taifun* (645.03 msec), *Jazz* (644.07 msec), *Terrain* (638.77 msec) and *Tyrann* (634.85 msec). The six items with most missing values were *Pfennig* (68% NAs), *Hockey* (59% NAs), *Detail* (44%NAs), *Hangar* (37% NAs), *Bronze* (37% NAs), and *Jockey* (34% NAs). The high percentage of missing values for these items can be explained by the strict enforcement of only accepting the pronunciations used in the consistency computations.

Table 7.2

Types, frequency and proportion of errors in naming reaction times per language group

	English (325 items)		German (297 items)	
Error types	*n*	%	*n*	%
incorrect pronunciation (produced incorrect nonword)	439	30.94	223	25.06
incorrect pronunciation (produced incorrect word)	153	10.78	64	7.19
pronunciation not according to CELEX transcription	279	19.66	288	32.36
hesitations and self-corrections	320	22.55	202	22.70
pronounced as word from another language	1	0.07	31	3.48
pronounced as homograph (used alternative meaning and pronunciation)	43	3.03	0	0.00
incorrect stress	89	6.27	27	3.03
external noise or unclear speech, technical fault, no response	95	6.69	55	6.18
Total	1419	100	890	100

7.2.2.4 *Error analysis for the complete naming sets*

Table 7.2 summarizes all errors made during the pronunciation of all admissible stimuli items, 325 for English and 297 for German. Participants in the two language groups made similar proportions of hesitations and self-corrections (E: 22.55% and G: 22.70%). There were also similar proportions of technical errors (E: 6.69% and G: 6.18%). The similarity of these figures is reassuring as it points to a similar testing environment, conditions and error marking. The largest difference was that German participants more often used pronunciations which were not congruent with the CELEX transcription. These were mostly due to dialectal or regional differences in pronunciations. English participants, on the other hand, more often used an alternative pronunciation with a new meaning. For example, as in the case of *moped*, participants read /məʊpd/ instead of /ˈməʊpɛd/. Another difference between the types of errors between the languages was that German readers tended to mispronounce words as foreign words more often than their English counterparts. It is likely that this occurred because participants were aware of the cross-language aspect of the study, and thus German participants were more likely to pronounce a word with an English pronunciation, than English participants would be to pronounce a word with a German pronunciation.

The most interesting error data from our perspective, however, is the incorrect pronunciation of items as words or as nonwords. Recall that N. C. Ellis & Hooper (2001) found that beginner readers of English tended to mispronounce words as other words (44.8%) rather than nonwords (24.5%), whereas beginner readers of the transparent Welsh script more often misread a word as a nonword (72.5%) than a word (24.5%). These error differences have been interpreted as mirroring different reading mechanism, i.e. whole word recognition for English readers and phonological recoding for Welsh readers. In our mature reader groups, English readers made more mispronunciations in the form of words and nonwords compared to German readers. However, for both groups, the mispronunciation as nonwords occurred more often than the mispronunciation as an incorrect word. Thus, the trend for beginner readers of different orthographies to use different mechanisms did not seem to be in place in skilled readers.

7.2.2.5 *Descriptive statistics for 85 cognate stimuli*

The final stimuli set for the present study consisted of 85 cognates. Recall that of the original 89 cognates, 4 (*ego*, *chance*, *steak* and *song*) had been excluded. Table 7.3 gives a brief overview of categorical descriptive data for the two language sets.

Table 7.3

Counts for categorical psycholinguistic variables for 85 cognate stimuli

	English								German							
	1	2	3	4	5	6	7	9	1	2	3	4	5	6	7	9
stress count	79	6							65	20						
letter count			2	26	28	22	6	1			2	26	28	22	6	1
phoneme count		1	14	33	25	11	1				4	40	27	13	1	
syllable count	28	57							24	61						

Table 7.4 presents the full unstandardized descriptive statistics for the two language stimuli sets and also gives the results of the statistical comparisons of all matching variables. Cognates are chosen for language comparison because the same words exist in both languages, therefore being identical in meaning and form, including length in letters as can be seen in Table 7.3. When examining Table 7.4, it becomes clear that the two cognate sets were also very similar in AoA. Cognates seemed to be somewhat more frequent in English than in German, but the difference was not quite significant. Although letter count was of course identical, phoneme count and syllable count were not. The lower phoneme count in English suggests that there are more occasions in English when several letters are subsumed under one sound compared to German. The slightly higher syllable count in German was due to the words *bronze/Bronze*, *note/Note*, *phase/Phase* and *vase/Vase* which are all bisyllabic in German, but monosyllables in English. The English cognates also tended to have more orthographic neighbours than German cognates. Although on average German cognates seemed to have less body-rime neighbours and phonographic rime neighbours of the first rime (PNR1), these differences were not significant. An important observation is,

that the two cognate sets differed on almost all consistency measures, mostly at the rime level. Thus, the same words were more spelling-sound consistent in German than in English.

Table 7.4

Comparison of cognate stimuli sets in English and German

	n	*M (SD)*	median	min	max	*n*	*M (SD)*	median	min	max	test statistic	df	*p*
	English					German							
Zipf word frequency	85	4.15 (0.69)	4.22	2.83	5.77	85	3.95 (0.64)	3.89	2.92	5.59	t = 1.92	168	.056
AoA	85	7.67 (2.51)	7.53	3.26	13.94	85	7.21 (2.65)	7.11	2.54	13	t = 1.17	168	.242
letter count	85	5.09 (1.06)	5	3	9	85	5.09 (1.06)	5	3	9	$\chi^2 = 0$	3	1
phoneme count	85	4.40 (0.99)	4	2	7	85	4.61 (0.85)	4	3	7	$\chi^2 = 7.27$	3	.064
syllable count	85	1.67 (0.47)	2	1	2	85	1.72 (0.45)	2	1	7	$\chi^2 = 0.25$	1	.618
Coltheart's *N*	85	2.93 (3.32)	2	0	12	85	1.94 (2.68)	1	0	11	W = 4219.5		.050
OLD20	85	1.94 (0.51)	1.85	1.05	4.25	85	2.12 (0.51)	1.95	1.25	4.05	W = 2802		.011 *
body *N* rime 1	85	567.35 (615.32)	143	0	1608	85	401.14 (405.2)	370	0	1294	W = 3924		.331
phonographic *N* rime 1	85	190.32 (230.57)	51	0	703	85	178.81 (179.88)	144	0	686	W = 3467		.651
FFO1	80	0.96 (0.18)	1	0	1	80	0.91 (0.27)	1	0.07	1	W = 3197		.991
FBO1	80	0.90 (0.23)	0.99	0.02	1	80	0.83 (0.33)	1	0.01	1	W = 3082.5		.673
FFR1	85	0.52 (0.31)	0.44	0.02	1	85	0.64 (0.32)	0.71	0.01	1	W = 2971		.045 *
FBR1	85	0.62 (0.31)	0.67	0	1	85	0.84 (0.24)	0.94	0.01	1	W = 1874		< .001 ***
FFO2	53	0.89 (0.22)	0.99	0.06	1	58	0.97 (0.14)	1	0	1	W = 974		< .001 ***
FBO2	54	0.61 (0.34)	0.81	0.01	1	58	0.69 (0.36)	0.85	0.01	1	W = 1186		.027 *
FFR2	57	0.57 (0.27)	0.57	0.04	1	61	0.62 (0.28)	0.68	0.02	1	W = 1564.5		.350
FBR2	57	0.40 (0.29)	0.31	0	1	61	0.69 (0.37)	0.87	0	1	W = 895.5		< .001 ***
composite FFO	84	0.94 (0.15)	1	0.14	1	84	0.93 (0.19)	1	0.07	1	W = 3106		.144
composite FBO	84	0.81 (0.24)	0.90	0.02	1	84	0.77 (0.29)	0.91	0.01	1	W = 3496		.920
composite FFR	85	0.58 (0.26)	0.57	0.09	1	85	0.67 (0.24)	0.65	0.18	1	W = 2877		.022 *
composite FBR	85	0.54 (0.25)	0.54	0	1	85	0.78 (0.24)	0.88	0.04	1	W = 1719		< .001 ***

Note . Consistency measures are notated as FF for feedforward and FB for feedback, O = onset and R = rime; OLD20 = orthographic Levenshtein distance 20 .

* = *p* < .05, ** = *p* < .01, *** = *p* < .001

All psycholinguistic variables for the 85 cognates were standardized within language groups. Figure 7.3 and Figure 7.4 (see next page) show the distributions of variables used in the analysis per language group. Apart from the consistency variables, distributions seemed very similar across the two languages. The cognates tended to be more composite FB and FF rime inconsistent in English than in German.

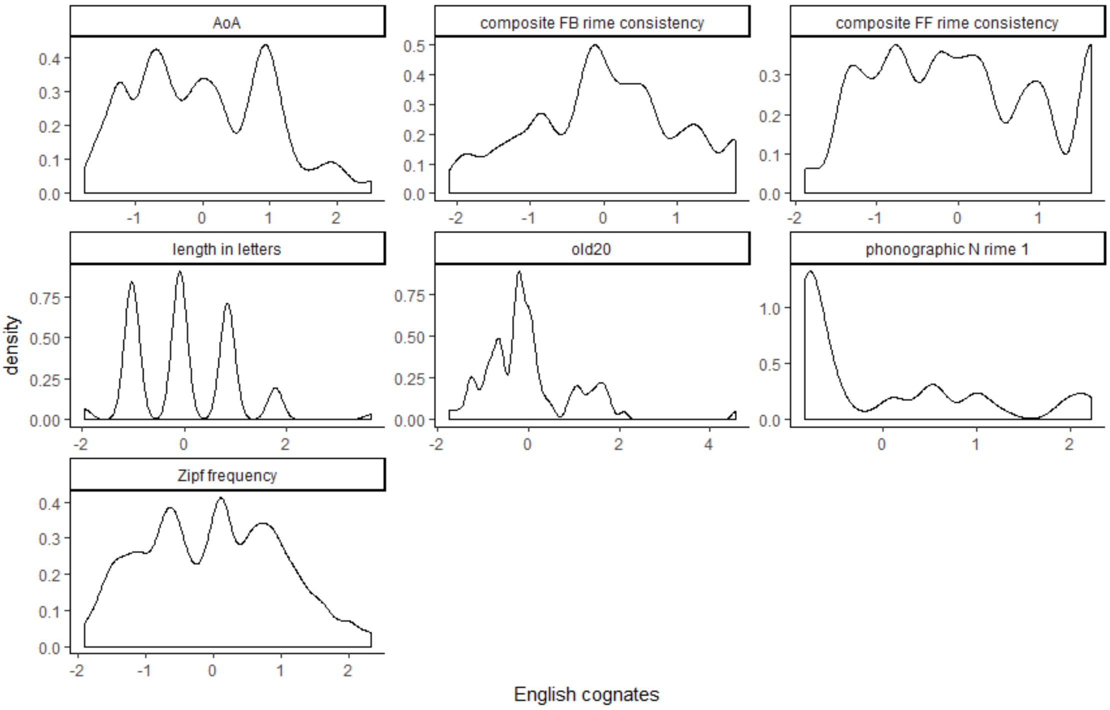


Figure 7.3. Density plots for standardised psycholinguistic variables in 85 English cognates. Scores were standardised within the language group.

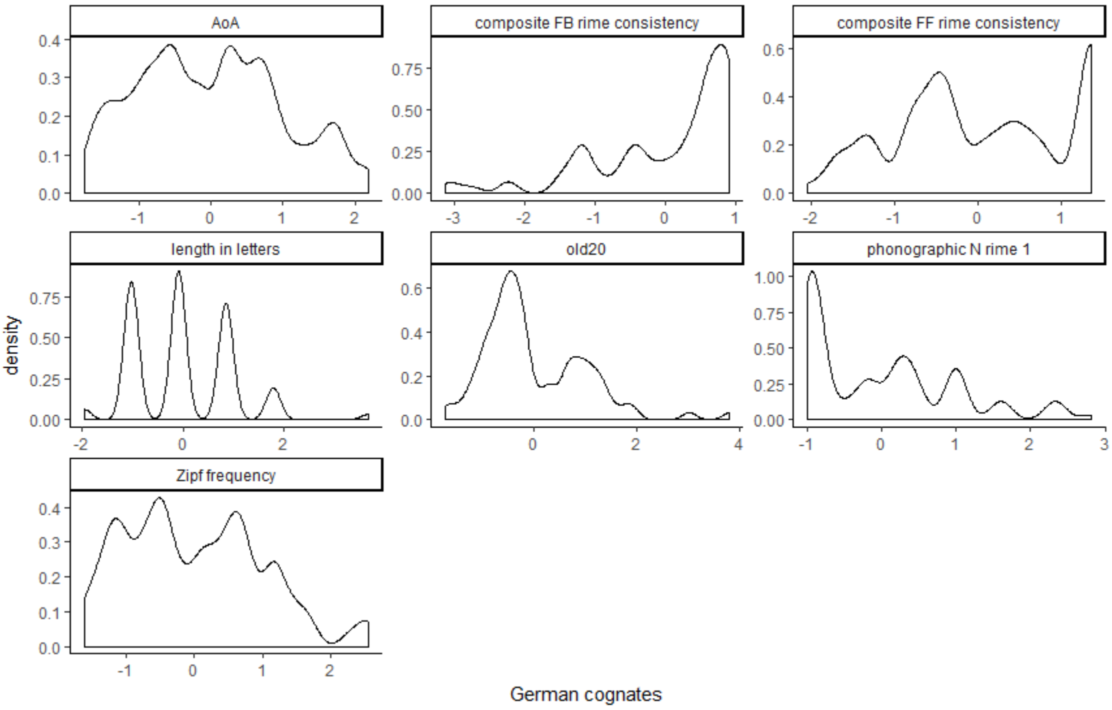


Figure 7.4. Density plots for standardised psycholinguistic variables in 85 German cognates. Scores were standardised within the language group.

7.2.3 Individual differences tasks

7.2.3.1 Materials and procedure

Tasks to measure individual differences in print exposure, vocabulary knowledge, decoding ability and spelling performance are described below. An overview of the raw and standardised results of the ID tasks can be found in Table 7.5 and Figure 7.5 in the Results section (see section 7.3.1).

7.2.3.1.1 Print exposure

Print exposure was measured using the GE-ART (see Chapter 6), an updated version of the ART (Stanovich & West, 1989), asking participants to identify authors from a list of made-up and real author names. To counteract order effects, three versions of the ART, each with a different name order, had been presented to participants. Version 1 was completed by 35 English and 34 German speakers, version 2 by 34 English and 35 German speakers and version 3 was completed by 35 participants in each language group. In neither language, average naming reaction times differed due to the ART version used (English: $F(1,102) = 0.663$, $p = .417$), German: ($F(1,102) = 0.744$, $p = .39$). The ART was validated using the GE-RQ (see Chapter 7), a revised version of the Reading and Media Habits Questionnaire (Stanovich & West, 1989). Primary print knowledge (PPK; Martin-Chang & Gould, 2008) emerged as the best predictor for print exposure ahead of the original ART score according to Stanovich & West (1989) and the number of authors recognised. PPK scores were standardised within each language group, and was therefore used as the print exposure measure in the present study.

7.2.3.1.2 Spelling

Due to lack of comparable tests across languages, different spelling-to-dictation assessments were administered in each language. Given that participants were University students, spelling-to-dictation tests were chosen, as these have been found to be more difficult than recognition tests (Bosman & de Groot, 1992). English-speaking participants completed the blue WRAT3 spelling test (Wilkinson, 1993), which required participants to spell up to 40 words of increasing difficulty. The test was administered by the experimenter. For each word, the experimenter first pronounced the word, then read out a sentence containing this word, and then repeated the word once again. The participant then wrote down the word without having seen it. In line with scoring test instructions, participants were credited an extra 15 points, as the administration of the name and letter writing section was not necessary. Thus, the highest score participants could attain was 55 points.

The German-speaking participants completed the *Rummelplatz* dictation of the Rechtschreibungstest (Kersting & Althoff, 2004). In this test, participants were given a text with gaps. The experimenter read the text aloud, including the words which were missing in the gaps. Whilst the experimenter was reading, the participant filled in the missing words. This test was chosen as it can be administered irrespective of whether the participant has learned the old or the new spelling rules for German. In order to have a comparable raw score to the English test, the number of items filled in correctly were used as raw scores. At most, participants could attain 60 points. For both tests, scores were standardised within each language group.

It is important to highlight a potentially important difference between the tests. In the English test, all errors were strictly based on the correct choice and sequence of letters to make up a word. In the German test, words were also deemed incorrect if they incorrectly used capitalisation or non-capitalisation, if several words were incorrectly joined into one, or if they were incorrectly partitioned into several words. These differences diminish the comparability of the spelling tasks across languages. It is therefore also not surprising that the distributions of the spelling tasks were quite dissimilar across the two languages, as can be seen from Figure 7.5 in section 7.3.1. However, other cross-language spelling tests available at the time of testing were not considered more suitable.

7.2.3.1.3 Nonword decoding

English-speaking participants completed the TOWRE-2 Form A nonword reading test (Torgesen, Wagner, & Rashotte, 2012). For this test, participants read as many nonwords as they could from a list of 66 items in 45 seconds. The ratio of words per time in seconds was used as the measure for nonword reading ability, so that a higher ratio indicated a better performance. For example, if a participant read 30 words within 45 seconds, then they would receive a score of 30/45 = 0.67. If a participant read twice as fast, that person would score 60/45 = 1.34.

To ensure comparability across languages, a German version of the TOWRE-2 test was devised by the researcher. Test items were created by identifying rimes which – based the present consistency computations – had similar consistencies to those used in the English TOWRE-2-test.

For analysis, all scores were standardised within language groups.

7.2.3.1.4 Vocabulary

The Vocabulary subtest of the Wechsler Adult Intelligence Scale IV (WAIS IV) has been published for both English (Wechsler, 2008) and German (Wechsler, 2012), and was therefore well-suited for this inter-language comparison study. This test probes semantic knowledge by asking participants to give definitions of words of increasing difficulty. The highest score possible was 57 points. Although standardised scores were available for both groups, it was decided to standardise scores within present language groups, analogously to the other tasks. Standardised scores were then used for analysis.

7.2.3.2 Data cleaning individual differences tasks

Only participants who had completed all tasks were included in the analysis. For more details on data cleaning for ID tasks, please refer back to section 7.2.1.

7.3 Results

7.3.1 Individual tasks results

Table 7.5 shows the raw scores for the individual tasks. However, as mentioned above, for comparability purposes, scores were standardised to be used in the analysis. Figure 7.5 shows the standardized score distributions for the four ID tasks per language.

Table 7.5

Unstandardised results for individual differences tasks

	M	*SD*	min	max
English language group				
PPK	4.6	3.66	0	19
spelling	46.09	2.64	40	53
nonword reading	1.18	0.2	0.69	1.72
vocabulary	40.34	7.8	20	57
German language group				
PPK	5.03	3.23	0	14
spelling	51.6	5.11	33	59
nonword reading	1.22	0.18	0.82	1.73
vocabulary	44.75	4.73	29	55

Note . PPK = Primary Print Knowledge (number of authors read; Martin-Chang & Gould, 2008); spelling = number of words spelled correctly (different tests for each language group); nonword reading = ratio of number of items read per 45 seconds; vocabulary = WAIS IV vocabulary test point scores.

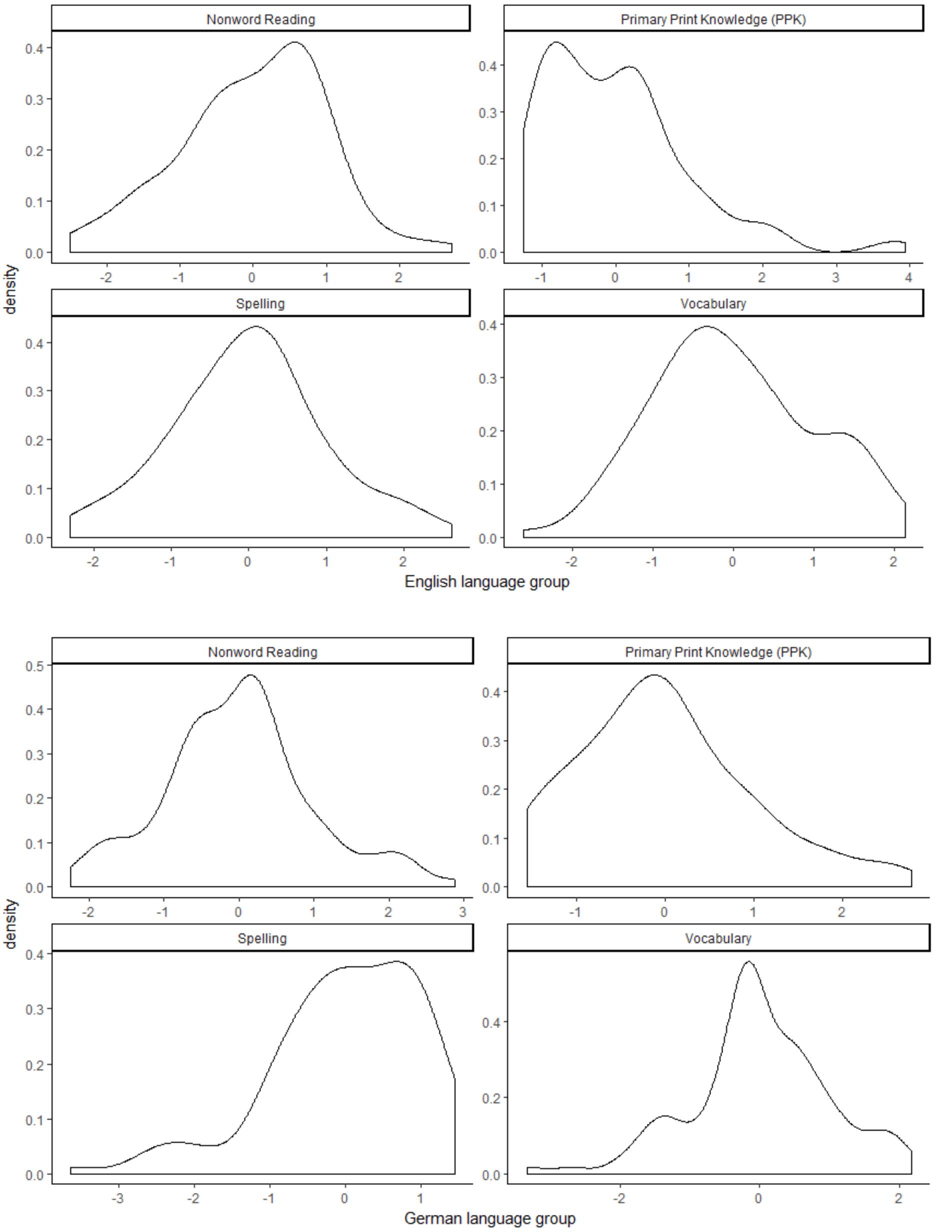


Figure 7.5 Density plots of four standardised individual difference variables per language group.

7.3.2 Naming task reaction time distributions

For the 85 cognate items, 8,200 valid responses (missing = 640) were collected for the English items, and 8,421 valid responses (missing = 419) were collected for the German items. The average RT for the combined cognate data set was 502.04 msec (*SD* = 102.29).

Considering the language sets separately, mean RTs for English items were 480.71 msec (*SD* = 106.5) and for German items 522 msec (*SD* = 93.46 msec). Figure 7.6 shows the distribution of naming RTs in msec per language group.

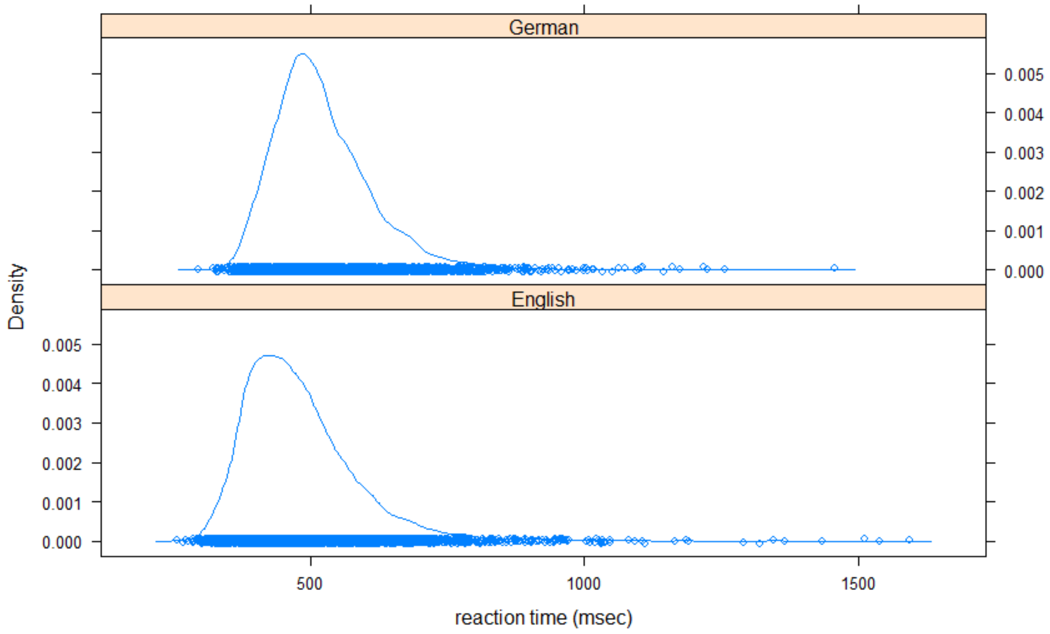


Figure 7.6. *Density plots for English and German naming times in msec for 85 cognate stimuli.*

As noted previously, due to differences in national curricula, it is likely that German participants spoke English as a second language, whereas English participants would have been less likely to have learned German at school. The longer naming times for German participants begs the question whether this might be due to the fact that they were aware of the cross-language nature of the study, and therefore experienced a slowing down in their reaction times due to simultaneous activation of the English pronunciations additionally to the German one. In fact, the current study results showed that German participants gave foreign pronunciations to items much more often than English participants (see Table 7.2, G: 31 times, E: 1 time), indicating that foreign language knowledge was clearly activated in some cases. However, Jared and Kroll (2001) showed that readers were not generally influenced by the knowledge of a second language when only reading in their first language.

Only if they had a relatively good command of the second language and had been exposed to words of this second language immediately before testing, reading aloud times tended to be slightly longer, indicating a disruptive process in assembling pronunciations. In the current study it was not assessed how well German participants spoke English, nor were they asked if that had been exposed to English text just prior to taking part in the study. However, given that German participants were tested in Germany, and only German was spoken during the testing sessions, it is unlikely that any prior exposure to English should have had an impact on naming latencies.

Yet, Jared and Kroll (2001) pointed out that the effect may be different for cognates, as they obviously exist in several languages. However, Friesen, Jared and Haigh (2014) did not find longer latencies for naming cognates compared to matched control words in English-French bilinguals who were naming words in their first language (English). In fact, in a further experiment with French – English bilinguals who were living/studying in the second language environment (English), the authors found a facilitatory cognate effect in first language naming. It thus seems unlikely that German participants were slower than English participants at naming cognates because of competition between two word pronunciations. Nevertheless, it would be beneficial in a future analysis to revisit this question and to compare the reading times for cognates of both language groups to a matched set of control words.

7.3.3 Correlations between naming RTs and psycholinguistic and ID variables

Table 7.6 presents the correlations between all standardised variables and naming RTs for both language sets separately. As the data set was relatively large, almost all correlations are significant.

For the English cognate set, reaction times were most strongly correlated with the psycholinguistic variables AoA (r_s = .20), frequency (r_s = -.17) and orthographic neighbourhood (r_s = .16). For the German cognate set, these correlations were the same in direction but turned out somewhat stronger, frequency (r_s = -.26), AoA (r_s = .23) and orthographic neighbourhood (r_s = .22). Additionally, for the German cognates only, composite FB rime consistency reached an almost moderate correlation size (r_s = -.21).

In both language sets, nonword reading was correlated strongest with RTs (E: r_s = -.21; G: r_s = -.2). In English, vocabulary knowledge (r_s = -.19) and reading experience (r_s = -.15) reached small correlations with RTs, whilst in German spelling ability was somewhat correlated to RTs (r_s = -.12).

Variables were also correlated with each other. Most notably, frequency and AoA were strongly correlated in both languages (E: r_s = -.54; G: r_s = -.66). This was to be expected, as it reflects previous research findings. Another important set of correlations regards composite FF and FB rime consistencies. In German, words which were more FFR consistent also tended to be shorter (r_s = -.46), be more FBR consistent (r_s = .38) and have more orthographic neighbours (r_s = -.45). These relationships were much more attenuated in English. Equally, in German, words which were more FBR consistent also tended to be shorter (r_s = -.38) and have more orthographic neighbours (r_s = -.55). Again, these relationships were less pronounced in English. It is possible that the stronger inter-variable relationships in German partly explain why the relationships between RTs and consistency seemed stronger in German than in English.

Phonographic N for rime 1 (PNR1) were included in the study as a marker for larger grain sizes in reading. For the current cognate set in both languages, the relationship with RTs was very small (E: r_s = - .09; G: r_s = .03). In both languages, words with more phonographic rime 1 neighbours tended to be less FFR consistent (E: r_s = -.24, G: r_s = -.38), be longer (E: r_s = .43, G: r_s = .2), and have less orthographic neighbours (E: r_s = .25, G: r_s = .19). These relationships are likely to stem from the fact that the cognate sets included mostly words with two syllables (E: 57; G: 61 out of 85 words). An informal investigation with

2,830 monosyllabic words from the database which was used to compute spelling-sound consistency showed that monosyllables showed a different correlation pattern: English monosyllabic words with more phonographic rime neighbours tended to have no relationship with FFR consistency (r = -.04), be shorter (r = -.3), and have more orthographic neighbours (r = -.24).

With regard to the ID variables, Table 7.6 also shows some inter-correlation patterns. In the English group, the strongest positive relationship was between print exposure and vocabulary knowledge (r_s= .56), so that high-print individuals seemed to have greater vocabularies. This reflects previous research findings. The difference however seems to lie in their relationship with spelling. Whereas vocabulary knowledge was also moderately positively correlated to spelling ability (r_s = .42), the relationship between spelling and print exposure seemed less strong (r_s = .24). Decoding skill seemed to have its strongest relationship with spelling: good spellers also tended to be good decoders (r_s = .4).

For the German language group, the inter-correlation pattern for ID tasks was slightly different to the English language group. The strongest relationship for decoding ability was with vocabulary knowledge (r_s = .27). However, vocabulary was also positively correlated to all other individual difference variables to almost the same degree. High-print individuals also tended to have greater vocabularies (r_s = .24) and better spelling ability (r_s = .19), but did not seem to be better decoders (r_s = .02).

For each language group, scatterplots showing naming RTs for the 85 cognates by standardised psycholinguistic variables per language group can be found in Appendix 11.13 and by standardised individual differences variables in Appendix 11.14.

Table 7.6

Spearman correlations between RTs and standardised psycholinguistic and individual differences variables for 85 cognate items in English and German

Variable	RTs	Zipf	AoA	comp FFR	comp FBR	letter length	old20	vocabulary	spelling	NW reading
English										
Zipf word frequency	-.17***									
AoA	.20***	-.54***								
comp FFR	-.06***	.11***	-.30***							
comp FBR	.01	.15***	-.16***	.18***						
letter length	.11***	-.21***	.29***	-.11***	-.12***					
old20	.16***	-.38***	.49***	-.15***	-.31***	.73***				
phonographic N rime 1	.09***	-.18***	.17***	-.24***	.16***	.43***	.25***			
vocabulary	-.19***	-	-	-	-	-				
spelling	-.09***	-	-	-	-	-		.42***		
NW reading	-.21***	-	-	-	-	-		.15***	.40***	
PPK	-.15***	-	-	-	-	-		.56***	.24***	.14***
German										
Zipf word frequency	-.26***									
AoA	.23***	-.66***								
compFFR	-.15***	.21***	-.22***							
compFBR	-.21***	.35***	-.34***	.38***						
letter length	.15***	-.17***	.23***	-.46***	-.38***					
old20	.22***	-.27***	.40***	-.45***	-.55***	.73***				
phonographic N rime 1	.03	-.06*	.18***	-.38***	.00***	.20***	.19***			
vocabulary	-.09***	-	-	-	-	-				
spelling	-.12***	-	-	-	-	-		.29***		
NW reading	-.20***	-	-	-	-	-		.27***	.15***	
PPK	-.08***	-	-	-	-	-		.24***	.19***	.02***

Note . RTs = naming reaction times in milliseconds, AoA = age-of-acquisition (years), comp FFR = composite feedforward rime consistency, comp FBR = composite feedback rime consistency, NW reading = nonword reading, PPK = Primary Print Knowledge. * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

7.4 Analysis I: Mixed-effects analysis on complete cognate data set

Linear mixed-effects modelling analysis (LMM, e.g., Baayen et al., 2008)) was chosen as the statistical method of data analysis. In psycholinguistic studies, linear mixed-effects modelling has enjoyed increasing usage over the past few years, as it is able to account for error variance given the loss of independence between residuals resulting from the collection of data under repeated measures designs (Bates, Mächler, Bolker, & Walker, 2015). Like ANOVA or regression analysis, LMMs fit predictor variables onto a dependent variable. Predictor variables are known as fixed effects. The additional advantage of LMMs is that they also take into account the random or unexplained deviations between the grand mean of the dependent variable and the mean latency of response of sampled participants or items, that is, the random intercepts of subjects or items. Moreover, LMM can also allow for random variation between subjects or items in the size and direction (the slope) of predictor effects, that is, taking into account the error variance associated with random slopes. Finally, when both random intercepts and random slopes have been specified, LMMs may also take into account the correlations between random deviations in intercepts and random deviations in slopes, the random covariances. Thus, linear mixed-effects modelling includes terms corresponding to subject and item variation in its estimations, thereby accounting for the fact that individuals and stimuli vary between each other randomly with respect to intercepts or to predictor effect slopes.

The specification of random slopes is referred to in terms of the random effects structure and, with regard to successful modelling, is the most crucial element for modelling data with LMM analysis (e.g., Barr, Levy, Scheepers, & Tily, 2013). In fact, Barr et al. (2013) advocate to use the maximal random effects structure to lower the risk of finding an effect when there is none, and to forego effect cherry-picking by the researcher. However, keeping the random effects structure maximal incurs the cost of power loss, reducing the capacity to detect effects that are present in the data (Bates, Kliegl, Vasishth, & Baayen, 2015; Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). Additionally, often, maximal models are so complex that model fitting algorithms are unable to converge on a set of effects estimates (Barr et al., 2013; Bates et al., 2015). In this context, Bates et al. (2015) offer a guideline on how to construct a parsimonious LMM. Guided by theoretical reasoning on which fixed effects and corresponding random effects are to be included, a step-by-step process reduces this maximal model to include only those variance components which are supported by the data.

7.4.1 Model building

Before starting with the analysis, the cognate data sets were submitted to a simple mixed-effects model with random effects of subjects and items on intercepts in order to check whether any order of presentation effects were present. For both the English and the German cognate RTs, *t*-values were below 2 (E: $t = 1.07$, G: $t = -1.02$), indicating that item order had not made a significant contribution to reaction times (Baayen, 2008).

The two data sets for the two languages were then combined to one single data set. Language was added as a categorical variable. Following Bates et al. (2015), a first maximal model was then built using a single combined data set. The maximal model included onset and stress variables, a categorical variable to indicate language (English and German), a selection of psycholinguistic variables which were pertinent to the present investigation (frequency, AoA, composite FF and FB rime consistency, length in letters, old20, phonographic N for rime 1), and the individual differences variables PPK, vocabulary, spelling and nonword reading. The model syntax was such that the effect of language was allowed to interact with the effects of psycholinguistic variables and individual differences. The full model did not converge. The model was thus slimmed by excluding the random effects correlation parameters. With the help of the RePsychLing package (Baayen, Bates, Kliegl, & Vasishth, 2015), the variance components (random slopes) spelling and decoding were identified as not adding to improve the model and were also remove. This model constituted a zero-correlation parameter model, and it did converge. This model is the one reported in the analysis I (see section 7.4.2).

It needs to be reported that there was an attempt to further improve the model, but that this attempt was not successful. Specifically, in order to improve the model, the correlation parameters were re-introduced, but this meant that once again the model did not converge. Consequently, the random effects structure was therefore reduced in this model: as identified with the RePsychLing package, all random slopes apart from frequency were found not to improve this correlation-model, and therefore removed. This resulted in a very reduced random effects structure model with random effects of participants on intercepts and on the slope of the frequency effect, and with random effects of items on intercepts. This reduced random effects model was then compared to the zero-correlation parameter model (see previous paragraph). In a likelihood ratio test, the zero-correlation-parameter model showed a better goodness of fit than the reduced random effects structure model (χ^2 (7, $N =$ 208) = 63.013, $p < .001$). Consequently, the present study reports the zero-correlation parameter model.

Table 7.7 in the Results section shows the zero-correlation-parameter model. It is however noteworthy that the direction of the coefficients of the reduced random effects structure model were the same as in the reported model. This reassuring observation suggests that the findings of the current reported zero-correlation parameter model would not be greatly changed, if a different way of identifying a model had been used.

7.4.2 Results

Before reporting the results, it is important to note that visualisations of all significant effects and interactions can be found in Appendix 11.15. An exception are three-way interactions, which have been reported in-text. In fact, these specific interactions have been presented visually twice. The plots in-text show the effects plot of the reported model (*Language * subject variables * item variables*). The 3-way interaction plots in Appendix 11.15 show the same effect plots from the same model but with a different variable order (*Language * item variables * subject variables*). The change in variable order leads to the same model results, but offers a different view of the relationship between the variables involved, and thus enhances the understanding of the results. All graphs were created using the R package 'effects' (Fox, 2003).

Table 7.7

Summary of the combined English and German reading aloud data model using mixed effects modelling.

	Estimate	SE	df	*t*	*p*
(Intercept)	601.125	13.978	282	43.006	< . 001 ***
bilabial *(present)*	-129.713	12.606	234	-10.29	< . 001 ***
labiodental *(present)*	-113.430	14.085	169	-8.053	< . 001 ***
alveolar and postalveolar *(present)*	-127.976	8.789	4297	-14.562	< . 001 ***
palatal *(present)*	-81.282	26.895	94	-3.022	.003 **
velar *(present)*	-96.116	17.125	96	-5.613	< . 001 ***
glottal *(present)*	-121.417	15.765	123	-7.702	< . 001 ***
plosive *(present)*	12.813	10.939	94	1.171	.244
nasal *(present)*	18.574	9.874	896	1.881	.060 .
fricative *(present)*	-26.776	12.599	79	-2.125	.037 *
approximant *(present)*	71.607	20.771	98	3.447	.001 ***
voiced *(present)*	26.933	3.854	971	6.989	< . 001 ***
affricate *(present)*	40.978	20.756	69	1.974	.052 .
short vowel *(present)*	-84.275	19.025	119	-4.43	< . 001 ***
long vowel or diphtong *(present)*	-79.180	19.107	119	-4.144	< . 001 ***
stress *(2nd syllable)*	-18.838	3.556	4853	-5.297	< . 001 ***
Language *(G)*	38.286	7.387	202	5.183	< . 001 ***
PPK	-7.598	6.462	198	-1.176	.241
vocabulary	-13.470	7.060	200	-1.908	.058 .
NW reading	-18.419	5.841	198	-3.153	.002 **
spelling	5.403	6.588	198	0.82	.413
Zipf	-6.193	1.931	1312	-3.206	.001 **
comp FFR cons	-5.709	1.435	1445	-3.979	< . 001 ***
comp FBR cons	-2.992	1.407	1172	-2.127	.034 *
AoA	10.001	2.012	1434	4.971	< . 001 ***
length	4.401	4.206	127	1.046	.297
old20	9.195	2.781	3208	3.306	.001 ***
phonographic N rime 1	4.477	1.587	1148	2.822	.005 **
Language (*G*) × PPK	2.176	8.551	198	0.254	.799
Language (*G*) × vocabulary	10.056	9.125	198	1.102	.272
Language (*G*) × nonword reading	3.439	7.953	198	0.432	.666
Language (*G*) × spelling	-13.033	8.635	198	-1.509	.133
Language (*G*) × Zipf	3.210	1.761	528	1.823	.069 .
Language (*G*) × comp FFR cons	-0.257	1.730	888	-0.149	.882
Language (*G*) × comp FBR cons	0.993	1.884	986	0.527	.598
Language (*G*) × AoA	2.409	2.014	618	1.196	.232
Language (*G*) × length	5.541	2.295	1203	2.415	.016 *
Language (*G*) × old20	1.601	2.408	1465	0.665	.506
Language (*G*) × phonographic N	-7.555	1.866	600	-4.049	< . 001 ***
PPK × Zipf frequency	2.109	1.294	199	1.629	.105
PPK × comp FFR cons	0.713	1.095	175	0.651	.516
PPK × comp FBR cons	-1.102	1.169	192	-0.943	.347
PPK × AoA	0.263	1.475	238	0.178	.859
PPK × length	-4.084	1.857	372	-2.199	.028 *
PPK × old20	0.640	1.924	369	0.333	.740
PPK × phonographic N	0.226	1.364	230	0.166	.869
vocabulary × Zipf	0.555	1.530	271	0.363	.717
vocabulary × comp FFR cons	1.943	1.302	278	1.493	.137
vocabulary × comp FBR cons	-1.126	1.377	295	-0.818	.414
vocabulary × AoA	-0.565	1.728	314	-0.327	.744
vocabulary × length	5.510	2.211	467	2.492	.013 *
vocabulary × old20	-5.080	2.284	546	-2.224	.027 *
vocabulary × phonographic N	-0.426	1.587	323	-0.268	.789
NW reading × Zipf	1.158	1.175	266	0.986	.325
NW reading × comp FFR cons	0.815	0.999	257	0.816	.415
NW reading × comp FBR cons	-0.590	1.066	268	-0.553	.580
NW reading × AoA	-1.759	1.343	313	-1.31	.191
NW reading × length	5.044	1.678	885	3.006	.003 **
NW reading × old20	-1.505	1.729	1010	-0.87	.384
NW reading × phonographic N	-2.160	1.244	300	-1.736	.084 .

spelling × Zipf	0.878	1.315	259	0.668	.505
spelling × comp FFR cons	-1.450	1.122	253	-1.292	.197
spelling × comp FBR cons	0.394	1.197	264	0.329	.743
spelling × AoA	-1.352	1.515	313	-0.892	.373
spelling × length	-3.142	1.887	874	-1.665	.096 .
spelling × old20	1.305	1.941	987	0.672	.502
spelling × phonographic N	-1.637	1.391	290	-1.177	.240
Language *(G)* × PPK × Zipf	-1.757	1.748	266	-1.005	.316
Language *(G)* × PPK × comp FFR cons	-1.302	1.494	206	-0.871	.385
Language *(G)* × PPK × comp FBR cons	2.224	1.595	224	1.394	.165
Language *(G)* × PPK × AoA	-1.298	1.954	290	-0.664	.507
Language *(G)* × PPK × length	2.004	2.351	663	0.852	.394
Language *(G)* × PPK × old20	1.016	2.414	657	0.421	.674
Language *(G)* × PPK × phonographic N	-0.610	1.763	230	-0.346	.730
Language *(G)* × vocabulary × Zipf	0.637	1.922	311	0.331	.740
Language *(G)* × vocabulary × comp FFR cons	-3.631	1.709	318	-2.125	.034 *
Language *(G)* × vocabulary × comp FBR cons	1.337	1.811	328	0.738	.461
Language *(G)* × vocabulary × AoA	0.191	2.142	334	0.089	.929
Language *(G)* × vocabulary × length	-5.583	2.574	792	-2.169	.030 *
Language *(G)* × vocabulary × old20	3.044	2.660	885	1.144	.253
Language *(G)* × vocabulary × phonographic N	-0.119	1.977	304	-0.06	.952
Language *(G)* × NW reading × Zipf	0.552	1.635	290	0.337	.736
Language *(G)* × NW reading × comp FFR cons	-0.101	1.396	286	-0.072	.942
Language *(G)* × NW reading × comp FBR cons	0.473	1.485	294	0.319	.750
Language *(G)* × NW reading × AoA	0.869	1.822	309	0.477	.634
Language *(G)* × NW reading × length	-7.337	2.185	745	-3.358	.001 ***
Language *(G)* × NW reading × old20	3.034	2.229	825	1.361	.174
Language *(G)* × NW reading × phonographic N	2.641	1.647	269	1.604	.110
Language *(G)* × spelling × Zipf	-0.016	1.767	285	-0.009	.993
Language *(G)* × spelling × comp FFR cons	1.505	1.511	282	0.996	.320
Language *(G)* × spelling × comp FBR cons	0.702	1.610	292	0.436	.663
Language *(G)* × spelling × AoA	-0.445	1.981	310	-0.225	.822
Language *(G)* × spelling × length	0.648	2.380	755	0.272	.785
Language *(G)* × spelling × old20	-1.820	2.433	838	-0.748	.455
Language *(G)* × spelling × phonographic N	2.206	1.782	266	1.237	.217

() indicate level shown for categorical variables, such as (G) = German; NW reading = nonword reading/decoding, Zipf = Zipf frequency, comp FFR cons = composite feedforward rime consistency, comp FBR cons = composite feedback rime consistency, length = length in letters, old20 = orthographic Levensthein distance 20, phonographic N= phonographic neighbourhood for rime 1; *** $p < .001$, ** $p < .01$, * $p < .05$, . $p < .1$

Several main effects were found. The main effect of language indicated significantly longer latencies for German speakers than English speakers (coefficient = 38.286, $p < .001$). Nonword reading was the only main effect for individual differences with faster responses for better nonword reading ability (coefficient = -18.419, $p = .002$). In line with previous research, the results showed that across both languages responses were faster when words were of higher frequency, more feedforward and feedback consistent, were learned at an earlier age, and had more orthographic neighbours (old20). The frequency effect seemed to be stronger for English than for German (coefficient = 3.210, $p = .069$, see Figure 7.7), whereas the length effect was stronger for German than for English (coefficient = 5.541, $p = .016$, see Figure 7.8).

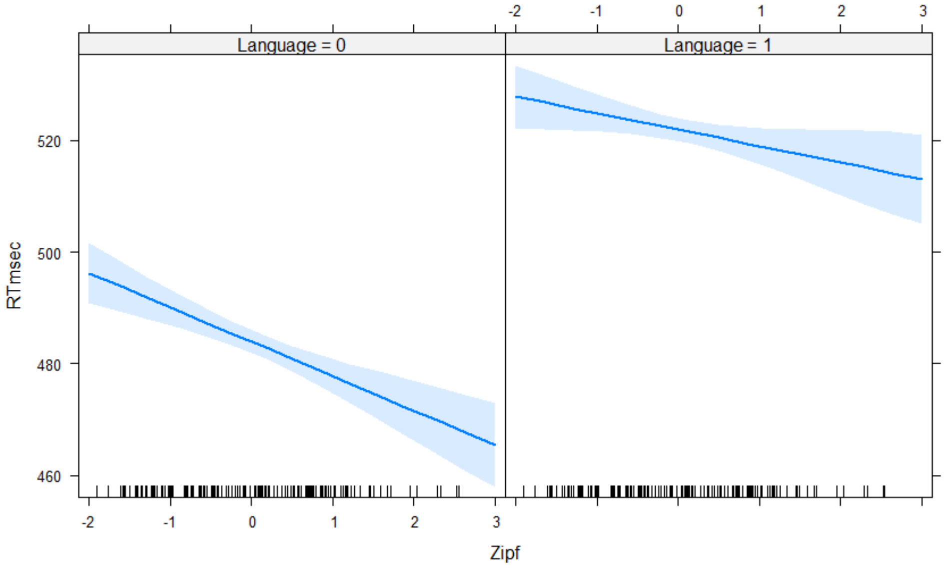


Figure 7.7. Graphs of language x Zipf frequency (coefficient = 3.210, p = .069). Language 0 = English; Language 1 = German. Zipf values were standardised within language group and higher values indicate greater frequency.

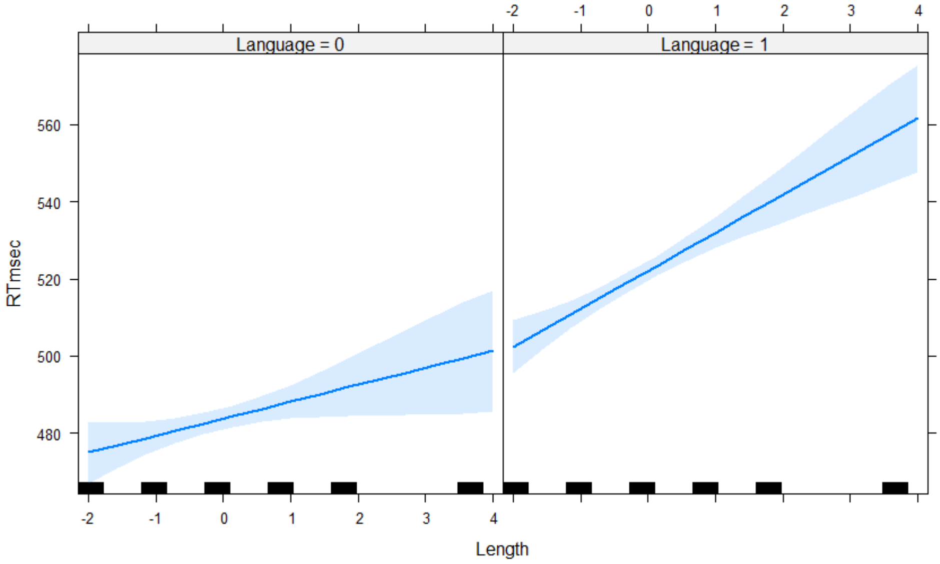


Figure 7.8. Graphs of language x length in letters (coefficient = 5.541, p = .016). Language 0 = English; Language 1 = German. Length in letters indicates number of letters of word items.

Phonographic neighbours of the first rime of the word (PNR1) interacted with language: for English, latencies were shorter for words with fewer PNR1, whilst this was reversed for German (coefficient = -7.555, p < .001, see Figure 7.9). Thus, larger phonographic rime 1 neighbourhood was slowing down pronunciation in English, whereas it was facilitatory for German.

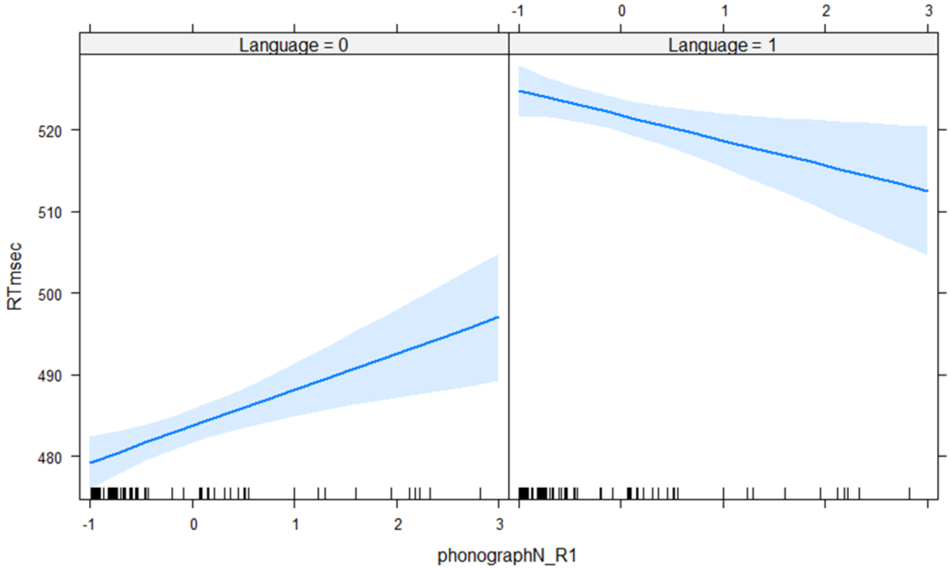


Figure 7.9. Graphs of language x phonographic rime 1 neighbours (PNR1, coefficient = -7.555, p < .001). Language 0 = English; Language 1 = German. PNR1 values were standardised within language group and higher values indicate fewer neighbours.

Length interacted with all individual difference variables. Across languages, greater print exposure sped up RTs when words were longer (coefficient = -4.084, *p* = .028). Vocabulary knowledge seemed to affect naming differently across languages. There was a facilitatory effect of vocabulary knowledge for short but not long words, which was only apparent in English but not German (coefficient = -5.583, *p* = .030, see Figure 7.10). In terms of decoding skill, the 3-way interaction language x nonword reading x length in letters indicated that in English, decoding skills facilitated responses to shorter words more than for longer words, whilst in German, decoding skills seemed to facilitate responses more for longer words than for shorter words (coefficient = -7.337, *p* < .001, see Figure 7.11). It may be of interest, that - although the interaction did not quite reach significance - there was a general tendency across both languages for spelling ability to accelerate RTs when words were longer, but not when they were shorter (coefficient = -3.142, *p* = .096).

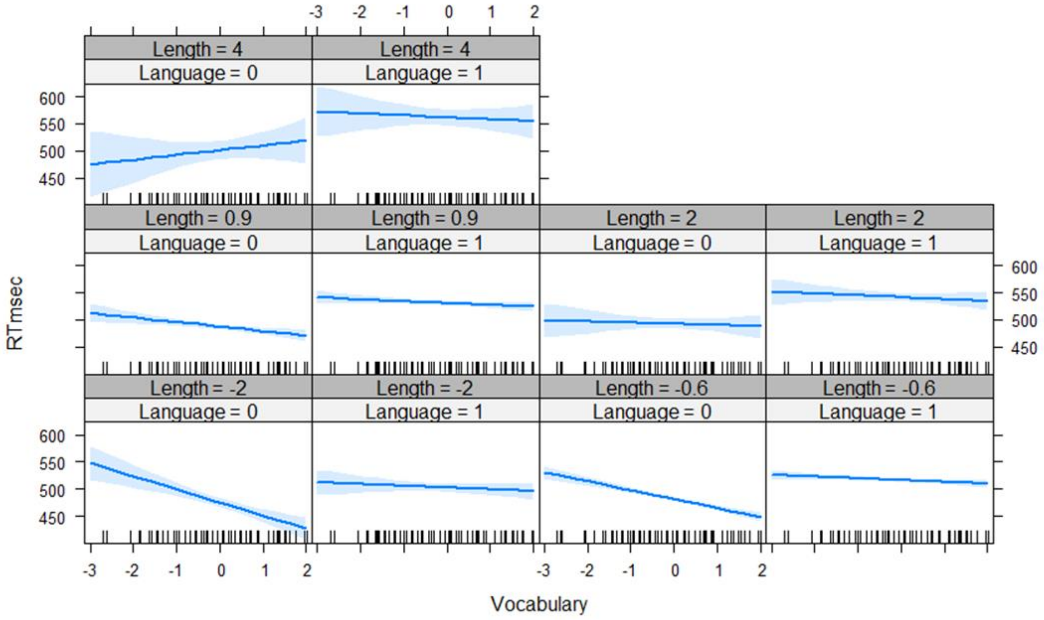


Figure 7.10. Graphs of language x vocabulary x length in letters interactions (coefficient =-5.583, p = .030). Language 0 = English; Language 1 = German. Vocabulary scores were standardised within language group and higher values indicate greater knowledge. Length in letter is indicated as number of letters.

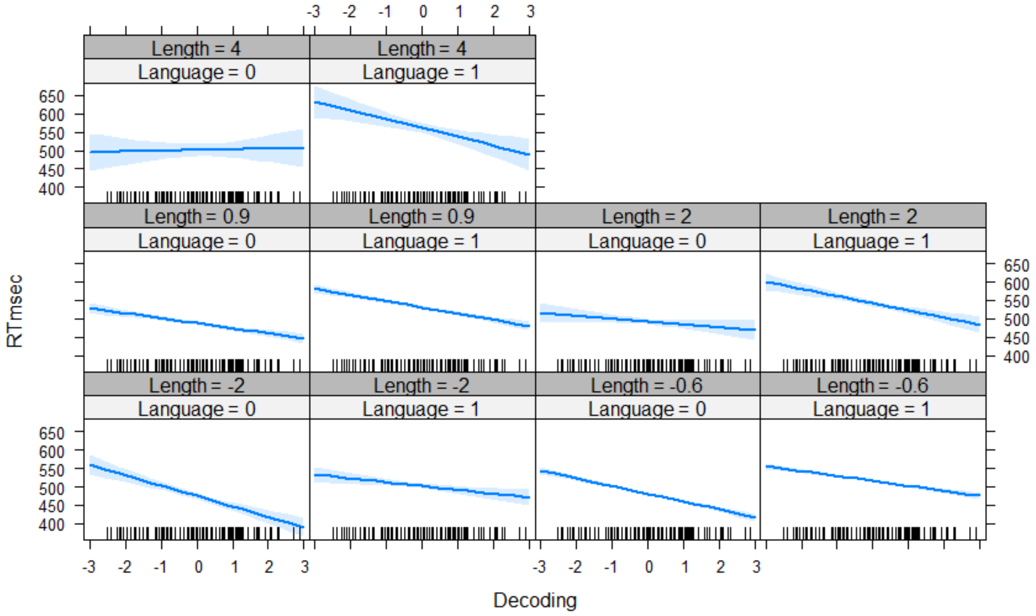


Figure 7.11. Graph of language x decoding x letter length interactions (coefficient = -7.337, p < .001). Language 0 = English; Language 1 = German. Decoding scores were standardised within language group and higher values indicate greater knowledge. Length indicates number of letters.

A significant 2-way interaction between vocabulary and old20 indicated that vocabulary knowledge was more facilitating for words with fewer orthographic neighbours (coefficient = -5.080, *p* = .027, see Figure 7.12). The 3-way interaction with language was not significant, suggesting that this was the same across languages.

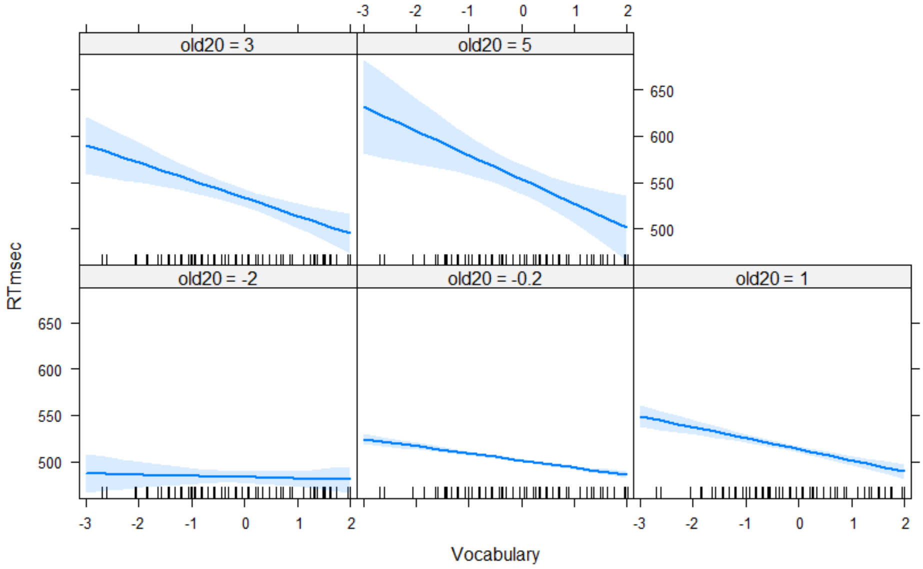


Figure 7.12. Graph of vocabulary x old20 (coefficient = -5.080, p = .027). Language 0 = English; Language 1 = German. Vocabulary scores were standardised within language groups and higher values indicate greater knowledge. Higher old20 values indicate fewer orthographic neighbours.

Finally, a significant 3-way interaction between language, vocabulary knowledge and composite FFR consistency suggested that in German, vocabulary knowledge seemed to be somewhat more facilitating when words grew more consistent, although this effect seemed to be very weak. In English, on the other hand, vocabulary knowledge was facilitating for words of all levels of consistency, but this effect seemed stronger for less consistent words (coefficient = -3.631, *p* = .034, see Figure 7.13).

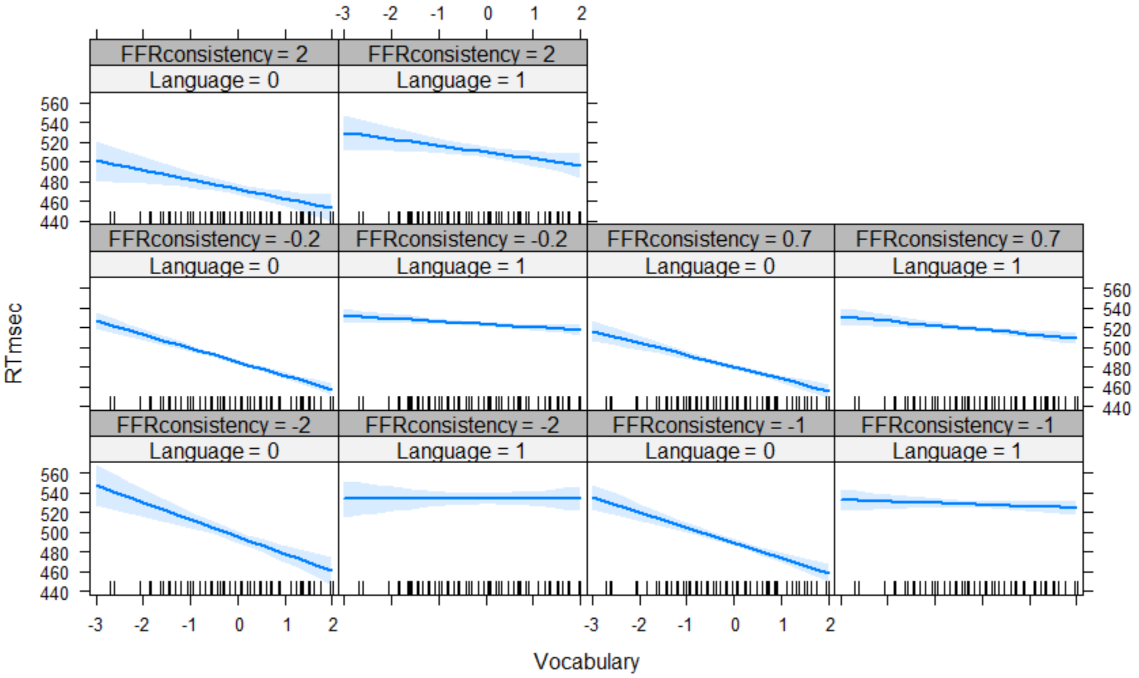


Figure 7.13. Graph of language x vocabulary x composite FFR consistency interactions (coefficient = 3.631, p = .034). Language 0 = English; Language 1 = German. Vocabulary scores were standardised within language groups and higher values indicate greater knowledge.

Individual differences clearly added to account for the variance of reading aloud latencies. Using the MuMIn package (K. Barton, 2018), two R^2 values can be obtained to indicate what proportion of variance is explained by the effects: the marginal R^2 values (R^2_m) pertain to fixed effects only, whilst the conditional R^2 values (R^2_c) also include variance explained by random effects. The model without the individual differences had an R^2_m of .196, and an R^2_c of .567. Including individual differences as fixed and random effects in the model increased the R^2_m to .247, and R^2_c to .574, and a likelihood ratio test confirmed the improved goodness of it (χ^2 (66, $N = 208$) = 195.91, $p < .001$).

Although all predictors had been standardised to reduce multicollinearity, high values for the variance inflation factor (VIF) and kappa (VIF = 17.65, kappa = 39.27; code from https://github.com/aufrank/R-hacks/blob/master/mer-utils.R) attested that some variables were still highly correlated. When excluding the non-standardised dummy variables for onsets, VIF and kappa values were reduced to more acceptable levels (VIF = 3.92, kappa = 18.11).

Alternatively, it was important to check if the removal of any psycholinguistic variables would reduce multicollinearity. For this purpose, variable clusters using the Hmisc package (Harrell & Dupont, 2018) were visualised as shown in Figure 7.14. In a step-by-step approach, one variable of each variable cluster with high correlations was excluded. However, to anticipate results, this did not reduce multicollinearity indicator values to acceptable levels. In the following, VIF and kappa values after the exclusion of the following variables from the model are reported: old20 (VIF = 17.71, kappa = 29.19), old20 and AoA (VIF = 16.93, kappa = 28.66), old20, AoA and vocabulary (VIF = 16.87, kappa = 26.67), old20, AoA, vocabulary and composite FFR consistency (VIF = 18.46, kappa = 24.63), old20, AoA, vocabulary, composite FFR consistency and spelling (VIF = 18.41, kappa = 26.65). Thus, although the reported model had higher than recommended levels of multicollinearity, these were predominately due to onset variables, and therefore remained in the model. Multicollinearity carried by onset variables was deemed acceptable, as the individual contribution of each onset variable was not the focus of the investigation. (cf. Cohen, Cohen, West, & Aiken, 2003)

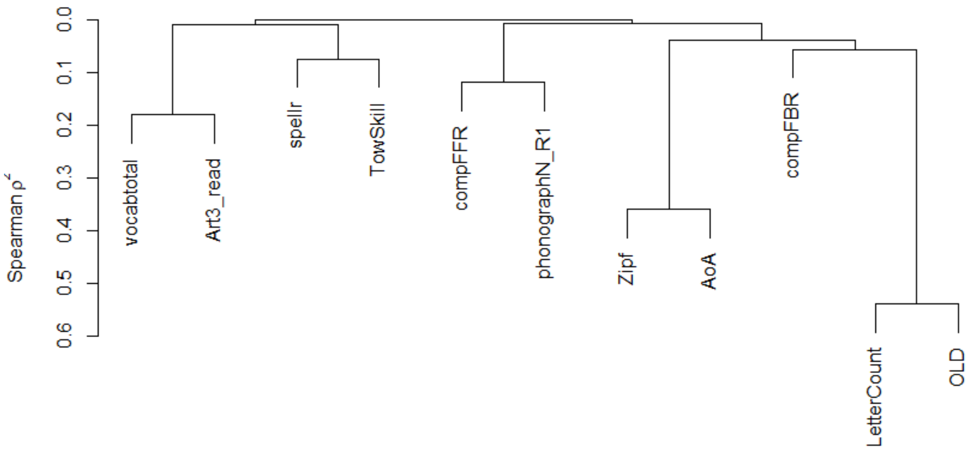


Figure 7.14. Visualisation of numeric standardised predictor variable used for cognate model.

7.4.3 Discussion

The present study's aim was to examine whether or how language differences were observed when taking into account the effects of individual differences in reading aloud. A mixed-effects model which included both psycholinguistic item variables and individual difference participant variables produced a number of interesting findings.

7.4.3.1 Language differences remain when IDs were controlled

Language was found to influence the reading process. The same cognate items were read significantly slower by German readers than by English readers. Language thus remained a significant main effect determining reading aloud even when important individual differences were considered.

In controlling for the first time for item consistency in a language comparison study, the influence of consistency was more precisely measured. Importantly, the composite FF rime consistency effect was present in both languages, evidencing that item consistency still influences reading performance even in more transparent languages.

Given that FF and FB rime consistency emerged as significant predictors additionally to language, it seems that the impact of language on reading aloud is not merely due to the difference in spelling-sound rime consistency of the stimuli. It is an important question to raise what exactly has been isolated with the language variable, given that composite FFR consistency was controlled. It is likely that the language effect comprises other aspects of language-level transparency, which are not being captured by the composite FFR and RBR measure. For the purpose of this investigation, the variable language will be considered to reflect general language-level transparency beyond composite rime consistency. Of course, there is also the possibility that the language effect is at least partly to do with teaching practices. It is beyond the scope of the present study to investigate this further, but future studies may usefully address this.

7.4.3.2 Differential reading units

In line with the Ziegler et al. (2001) study, the present study also reported a stronger length effect for German than English, which has frequently been interpreted as a marker of small-unit usage in reading. The stronger frequency effect in English (though the interaction is not quite significant) may point to stronger bigger unit usage for English than German. This seems to support the assumption of the PGST of differential reading of the two languages, with German tending to use small grain sizes, and English using larger grain sizes.

However, phonographic N for rime 1 (PNR1), another marker for larger unit sizes, did not support the assumption for the use of larger unit sizes for English readers. Contrary to Ziegler et al. (2001), it was found that larger PNR1 was inhibitory for English readers, but facilitatory for German readers. In fact, these results seem more in line with reports of facilitating phonographic body-rime *N* effects in French (Peereman & Content, 1997) and Italian (Arduino & Burani, 2004), which are both more transparent than English. Thus, the current results do not support larger grain size use by English readers as measured by PNR1. Rather this measure seems to better capture reading units used by readers of more transparent orthographies. Yet, the interpretation of these results needs to be treated with caution. As outlined in section 7.2.2.7.7.3, the PNR1 was clearly confounded by syllable number. Items with greater PNR1 were more likely to be bisyllabic than monosyllabic. It is therefore also possible that larger PNR1 were inhibitory for English readers because of their greater length in letters (as bisyllabic words are typically longer than monosyllabic words).

7.4.3.3 Different length x ID effect pattern across languages

The results also showed that the length effect was modulated by individual differences, and that to some degree these effects differed between languages. It is therefore possible that previous findings of the length effect were mixed because studies had not controlled for individual differences.

Let us have a closer look at the language x length x ID interactions. For English readers, naming times for longer words benefitted from more print exposure and better spelling ability. Naming times for shorter words were faster with higher vocabulary knowledge and better decoding skill. For German readers on the other hand, naming for longer words were faster with more print exposure, spelling ability and decoding skill. Vocabulary knowledge did not seem to interact with word length for German naming times. In other words, two IDs seemed to interact with length similarly in both languages. Reading practice (as indexed by print exposure) seemed to be universally beneficial to readers of both languages, and showed greater effects when items were longer. Similarly, spelling ability (or orthographic lexical quality), which almost reached significance in the present analysis, seemed to facilitate reading of longer words in both languages.

In contrast, a different pattern emerged between languages for the interactions of the two other IDs decoding and vocabulary knowledge with length. Let us discuss each of these IDs in turn.

7.4.3.3.1 Vocabulary knowledge

English readers reading short words seemed to benefit from better vocabulary knowledge, but not when reading long words. In contrast, for German readers, vocabulary knowledge did not seemingly interact with word length. If vocabulary knowledge is an indicator for an individual's quality and/or breadth of semantic knowledge, then the present results suggest that this knowledge may be facilitating for English for short words only, but not notably for German readers. Thus, it seems that German readers were less likely to use semantics in reading aloud, even if their semantic knowledge was well developed. In reference to the division of labour hypothesis within the PDP framework, this suggests that readers of transparent orthographies may not show semantic involvement, because given the transparency of the orthography, the direct O-P pathway is still more efficient. For English readers with greater vocabulary knowledge, however, the semantic route may be a useful alternative route, as it may be faster than the direct route due to the opaqueness of the script.

7.4.3.3.2 Nonword decoding skill

The other ID which interacted with length and language was decoding skill. Before discussing the interaction, it is important to highlight that decoding skill was the only ID which emerged as a main effect in the analysis. This is an important finding, as it emphasises the importance of decoding even within a skilled reader group and across languages differing in spelling-sound consistency. Previous research has mostly emphasised the importance for decoding in reading acquisition (e.g., Share, 1995), although some have argued for continued relevance (see Andrews, 2012). The present study shows that strong decoders have generally faster naming RTs than weak decoders, and that decoding ability continues to be a major predictor of reading RTs – in both languages.

However, as described above, whilst German readers seemed to benefit more from better decoding skills with increasing word length, for English readers better decoding skills seemed to be more beneficial for shorter than longer words. These findings are more easily explained with the PGST than the ODH. In terms of the ODH, the results are only partly compatible. The ODH assumes that readers of transparent orthographies use more small-unit processing than readers of opaque orthographies. For readers of German who are not good decoders, this would mean relatively slower naming RTs for longer words, which is what was observed. For English readers, however, decoding skills did not make a difference when words were long, but decoding skills were facilitating when words were short. This seems contradictory to the ODH, which would predict that decoding skills should not make (much of) a difference to reading in English, because words of all lengths would be accessed lexically. The results are more compatible with the PGST. The PGST assumes that readers of transparent orthographies tend to use small grain sizes, whereas readers of opaque orthographies use varied reading unit sizes. German readers would use small unit sizes, and good decoders would profit from their skills considerably more when reading longer words. English good decoders may be faster at finding the best grain size for shorter than for longer words and hence show facilitation for shorter words. For longer words, the difficulty of finding the correct grain size may slow down the reader to such an extent that beyond a certain length, reaction times are slowed down even for the most skilled readers.

7.4.3.4 Other findings

The present analysis also found an interaction between language, vocabulary knowledge and FFR consistency. In English, vocabulary knowledge facilitated naming for words of all consistencies, but more so for less consistent words. In comparison, for German it seemed that vocabulary had a less facilitating effect on naming, but that there was a tendency for more consistent words to be named faster when vocabulary knowledge was greater. At present, no explanation can be offered. Drawing on the extant literature, there seems to be no reason why in German vocabulary knowledge should be more facilitating as words become more consistent.

7.4.3.5 Implications for reading models

As far as it is reasonable without actual simulation work, the overall results from the current analysis can be considered within the framework of the PDP approach. As reviewed above, decoding skill can be taken as an indicator for the efficiency of the direct route within the triangle model framework (e.g., Strain & Herdman, 1999). German readers may tend to utilise the direct O-P pathway, as it remains a reliable strategy in a transparent script, even for more difficult or longer words. For readers of a transparent orthography, better decoders experience real benefit from decoding skills when reading longer words. For good decoders in English, O-P activation may be faster than activation via semantics when words are short, but this advantage may no longer be detectable when words are longer. Hence, in English, better decoding skill seemed to facilitate shorter words more than longer words. Longer words may require refixation (e.g., Plaut, 1999). The opaqueness of the orthography may then require a longer time to find the next grain size, as proposed by the PGST. This leads to the supposition that better decoding skills do not seem to facilitate reading of longer words in English because of the need to partition the longer words.

It has been suggested that readers who are more prone to utilise semantics for reading aloud do so, because of inefficient decoding skills (Strain & Herdman, 1999; Woollams et al., 2016). This is supported by the triangle model structure through the division of labour between the O-P and the O-S-P routes. (Plaut et al., 1996) The present results showed that German readers' vocabulary knowledge did not seem to interact with length, whereas good vocabulary knowledge in English readers seemed to facilitate naming of shorter words, but not for longer words. Unfortunately, it is not possible to deduce from the current analysis if the facilitating effect for shorter words in English was stronger for those participants with weaker decoding skills. For this, the analysis would have to be even more sensitive to variation of IDs within these reader groups. This will be addressed in the next chapter.

7.4.4 Summary

The present analysis suggests that language influences reading aloud, but that individual differences also play a role. Specifically, better decoding ability gives readers a clear advantage across languages even at a skilled reader level. The length effect, which has often been interpreted as a marker for small-unit processing, has been particularly affected by individual differences. This may be one of the reasons why previous studies found mixed results on length effects. Across both languages, more reading experience and better spelling ability seemed to be more beneficial with increasing word length. In contrast, the interactions between length and decoding skills and vocabulary knowledge seemed to show different patterns for readers of different languages. The interactions observed between language, IDs and length can be explained within the context of the PDP model and with the assumption of differential grain sizes for readers of different orthographies as assumed by the PGST. Readers of transparent German showed an advantage when they were strong decoders for words of increasing length. If readers of transparent orthographies can rely on the O-P pathway due to the language transparency, then stronger semantic knowledge may not induce advantages, as the O-P network remains the most reliable reading route. Readers of opaque English, on the other hand, showed an advantage for shorter words when they were either strong decoders or had strong vocabulary knowledge. It is conceivable that in participants with these strengths, either the direct or indirect pathway respectively were especially efficient, and this became evident for shorter words only. The advantage was no longer evident when words were longer, possibly because of the added burden of having to find the appropriate grain size for several word segments. This points to the fact that partitioning of longer words is a more disruptive process for readers of opaque than transparent orthographies.

Thus, it is clear that both language and individual differences influence the reading aloud process, and it appears that - at least in the case of decoding ability and vocabulary knowledge – they do not influence in exactly the same way. It therefore seems opportune to look at individual differences more closely.

8 Analysis II: Mixed-effects analysis on individual differences group

The previous analysis suggested that both language and individual differences influenced the reading process. Of the individual differences in question, decoding ability seemed to have the strongest facilitating effect in both languages. Moreover, it seemed to emerge that some IDs influenced reading of words of various lengths similarly across languages, such as reading practice and orthographic lexical quality. Other IDs, namely decoding skill and semantic knowledge, seemed to have differential impact between languages depending on words' lengths. The psycholinguistic variable length in letters was found to be most strongly influenced by language and IDs. Length in letters has previously been interpreted as an effect of small-unit processing typical of transparent orthographies. It seemed sensible to investigate these findings in more detail. In particular, it was considered that the rich dataset may have been too complex and effects may have been averaged out (cf. Andrews, 2012). The next analysis therefore aimed at creating a better insight into how individual differences modulate reading aloud differently in the two languages by creating less diverse data sets.

It seemed therefore opportune to subdivide the same data set four times according to relative strength and weakness of each individual difference and analyse each one separately. The subgrouping process is described in more detail below. As this approach yielded eight different analyses, the results section includes merely model specifications and estimates. More detailed findings, including the examination of interactions, will be presented in the discussion. This deviation from conventional reporting of psychological studies seemed necessary to ensure sufficient clarity and readability of the results. Other studies with an exploratory nature (e.g., Balota et al., 2004; Yap & Balota, 2009) have also veered from strictly separating results and discussion, presumably for readability purposes. As such, the discussion of the present chapter will first highlight general trends observed, then address findings pertaining to the use of differential grain sizes in the two languages, and finally record the findings specific to each subgroup.

Following the current chapter, the General Discussion in Chapter 9 will expound on the contributions of the present findings to the current understanding of reading aloud.

8.1 Method

8.1.1 Participant grouping into ID groups

In order to examine how each individual difference would shape the reading process, and given the complexity of the model and the limited number of stimuli and participants, the data set was split into stronger and weaker performers in each ID task. Specifically, for each ID the data for 104 German and 104 English participants was separated into better performers, who had scored above the mean of standardised ID scores, and relatively poorer performers, who had scored at or below the mean. Thus, each analysis included the same participants, but always grouped according to their strength and weakness in one of the individual differences tasks.

Please note that participants were the same as in Analysis I (see section 7.2.1). For details of data cleaning, please refer back to section 7.2.1.2. Naming data and performance data were the same as in Analysis I.

Table 8.1

Table of group categorisations for individual differences groups

hp	hv	hd	hs	lp	lv	ld	ls	all	E	G	diff
+	-	-	-	-	+	+	+	11	6	5	1
+	-	-	+	-	+	+	-	5	0	5	-5
+	-	+	-	-	+	-	+	8	3	5	-2
+	-	+	+	-	+	-	-	7	6	1	5
+	+	-	-	-	-	+	+	**12**	**11**	**1**	**10**
+	+	-	+	-	-	+	-	11	3	8	-5
+	+	+	-	-	-	-	+	9	7	2	5
+	+	+	+	-	-	-	-	25	13	12	1
-	-	-	-	+	+	+	+	32	18	14	4
-	-	-	+	+	+	+	-	14	7	7	0
-	-	+	-	+	+	-	+	20	10	10	0
-	-	+	+	+	+	-	-	13	6	7	-1
-	+	-	-	+	-	+	+	6	1	5	-4
-	+	-	+	+	-	+	-	7	4	3	1
-	+	+	-	+	-	-	+	7	3	4	-1
-	+	+	+	+	-	-	-	**21**	**6**	**15**	**-9**
								208	**104**	**104**	

Note. Groups of high scores are hp = high print exposure, hv = high vocabulary knowledge, hd = high nonword reading/decoding score, hs = high spelling score; groups of low scores are lp = low print exposure, lv = low vocabulary knowledge, ld = low nonword reading/decoding score, ls = low spelling score; E = English participants, G = German participants, diff = difference in number of participants (E-G); + = member of individual difference score group, - = not a member of individual difference score group

Before analysing the data sets, it was important to look at whether there were any evident patterns in terms of participant grouping. It could have been possible that distinct reader groups would emerge. Table 8.1 shows how many participants shared group memberships in the ID score groups. The eight columns on the left of the table represent the ID groups, for example 'hp' stands for high-print group, 'hv' denotes strong vocabulary knowledge and so on. Membership in one of the groups is indicated by a plus sign (+). The right – hand columns indicate the number of participants with each combination of ID score group memberships. For example, in the first line, out of 208 participants, 11 were categorised as high-print, but also fell into the low vocabulary, weak decoder and weak speller groups. This was true for 6 participants from the English and 5 from the German language group. This makes the difference between languages '1'. The very last column on the right hand-side 'diff' indicates this difference in number of people between the languages who share this combination. For most combinations, the number of participants were fairly similar across languages.

Overall, some tendencies towards specific groupings could be observed. Notably, out of all combinations, the largest ones were those where participants were in either all of the lower score groups (32 participants in total, 18 in English, 14 in German) or in all of the higher score groups (25 participants, 13 in English, 12 in German). Moreover, when adding up the number of participants who shared either 3 or 4 group memberships in the lower score groups, then 40% of all participants (40% English, 39% German) were either in all four or three of the lower score groups. Similarly, 35% of all participants (34% English, 37% German) fell into all 4 or 3 of the higher score groups. Only 25% of all participants (26% English, 24% German) had a truly mixed 'reader profile' with memberships in two high-score ID groups, and two low-score ID groups. Thus, it seems that although no distinct reader profiles emerged, there was a tendency for readers to either perform generally higher or generally lower on most ID tasks.

As indicated above, the last column ('diff') reports the difference in participant numbers for each ID combination between languages. For most combinations, the number of participants were fairly similar across languages. Yet, two inter-language differences are worth mentioning. First, the group of German participants who performed above the standardised score mean in vocabulary, decoding and spelling, but were also classified as belonging to the low-print group was larger (15) than the respective English group (6). Please see figures in bold in the lowest line of Table 8.1. Tentatively, it may be argued that this bolsters the assumption that in German reading experience is not as tightly linked to relative better performance on reading-related tasks. Second, participants who were classified as high-print and with strong vocabulary knowledge, but had scored lower on decoding and spelling was larger in English (11 participants) than in German (1 participant, see bold figures in fifth line from top in Table 8.1). It may be possible that this difference arises because of the different reading processes employed by the two language groups. For readers of German, low decoding skills and difficulties with spelling may deter from reading and further vocabulary learning, because the decoding process is cumbersome. For readers of English, who may employ more varied reading unit sizes and strategies for reading in English, the reduced decoding ability does not represent such a strong hindrance for these participants to still become avid readers, and subsequently vocabulary learners. Alternatively, this may of course be a result of the way that the groups were divided, which arguably may have left some border line cases with the ‘wrong’ classification.

Table 8.2 reports the mean scores each group attained in the four ID tasks, as well as the means for all participants combined for reference purposes. Generally, lower score groups had lower means in all ID tasks, whereas higher score groups tended to have higher scores across all tasks. This was true for both languages and concurs with the finding that participants tended to be either in more high-score or more low-score ID groups.

Table 8.2

Descriptive summary of individual differences standardised task scores and mean naming reaction times (msec) of cognate items, when participants were split according to above or at/below mean standardised score in individual differences tasks of each language group

	n	*M*	SD	min	max	*n*	*M*	SD	min	max
	English					German				
all participants										
vocabulary	104	0	1	-2.61	2.14	104	0	1	-3.33	2.17
spelling	104	0	1	-2.3	2.62	104	0	1	-3.64	1.45
PPK	104	0	1	-1.25	3.93	104	0	1	-1.56	2.78
NW reading	104	0	1	-2.5	2.72	104	0	1	-2.24	2.88
mean RT cognates	8200	480.71	106.5	257.4	1592.7	8421	522.8	93.46	295.1	1456.6
high print exposure	(hp)									
vocabulary	49	0.51	0.88	-1.07	2.14	39	0.31	0.95	-1.85	2.17
spelling	49	0.25	0.92	-1.55	2.24	39	0.28	0.77	-1.88	1.45
PPK	49	0.81	0.86	0.11	3.93	39	1.02	0.72	0.3	2.78
NW reading	49	0.22	0.9	-1.49	2.72	39	0.09	0.96	-1.74	2.88
mean RT cognates	3907	469.18	96.68	257.4	1290.6	3168	514.34	85.97	295.1	1106.1
low print exposure	(lp)									
vocabulary	55	-0.45	0.88	-2.61	2.01	65	-0.18	0.99	-3.33	1.96
spelling	55	-0.23	1.03	-2.3	2.62	65	-0.17	1.09	-3.64	1.45
PPK	55	-0.72	0.34	-1.25	-0.16	65	-0.61	0.53	-1.56	-0.01
NW reading	55	-0.2	1.05	-2.5	2.31	65	-0.05	1.03	-2.24	2.3
mean RT cognates	4293	491.21	113.72	269.1	1592.7	5253	527.9	97.35	329.8	1456.6
high vocabulary	(hv)									
vocabulary	48	0.88	0.61	0.09	2.14	50	0.78	0.6	0.05	2.17
spelling	48	0.44	0.95	-1.17	2.62	50	0.27	0.88	-2.66	1.45
PPK	48	0.54	1.13	-0.98	3.93	50	0.15	1.05	-1.56	2.47
NW reading	48	0.19	0.9	-2.5	2.72	50	0.29	1.01	-1.86	2.88
mean RT cognates	3851	465.26	97.53	269.1	1344.3	4051	517.65	88.43	295.1	1256.4
low vocabulary	(lv)									
vocabulary	56	-0.76	0.54	-2.61	-0.04	54	-0.04	1.11	-2.61	1.75
spelling	56	-0.38	0.89	-2.3	1.1	54	-0.07	0.96	-2.3	2.24
PPK	56	-0.46	0.56	-1.25	1.2	54	-0.04	1.04	-1.25	3.66
NW reading	56	-0.16	1.06	-2.39	2.31	54	0.02	0.87	-2.39	1.43
mean RT cognates	4349	494.4	112.1	257.4	1592.7	4370	527.57	97.66	322.5	1456.6

high decoding skill (hd)										
vocabulary	54	0.14	1.05	-1.84	2.14	56	0.16	1.08	-3.33	2.17
spelling	54	0.3	1.06	-1.93	2.62	56	0.16	0.82	-2.27	1.45
PPK	54	0.13	1.11	-1.25	3.93	56	-0.01	0.94	-1.56	2.47
NW reading	54	0.77	0.52	0.08	2.72	56	0.7	0.71	0.02	2.88
mean RT cognates	4281	462.08	96.08	257.4	1344.3	4523	508.8	90.02	295.1	1256.4
low decoding skill (ld)										
vocabulary	50	-0.15	0.93	-2.61	1.75	48	-0.19	0.88	-2.06	1.75
spelling	50	-0.33	0.83	-2.3	1.86	48	-0.19	1.16	-3.64	1.45
PPK	50	-0.14	0.85	-1.25	2.02	48	0.02	1.07	-1.56	2.78
NW reading	50	-0.83	0.67	-2.5	-0.03	48	-0.82	0.57	-2.24	-0.1
mean RT cognates	3919	501.06	113.4	286.9	1592.7	3898	539.04	94.75	328.8	1456.6
high spelling score (hs)										
vocabulary	45	0.26	1.07	-1.84	2.14	58	0.36	0.92	-1.85	2.17
spelling	45	0.88	0.63	0.35	2.62	58	0.7	0.42	0.08	1.45
PPK	45	0.11	1.18	-1.25	3.93	58	0.14	1.07	-1.56	2.78
NW reading	45	0.42	0.99	-2.5	2.72	58	0.11	0.88	-2.11	2.3
mean RT cognates	3601	472.24	100.45	257.4	1592.7	4722	511.94	87.21	295.1	1256.4
low spelling score (ls)										
vocabulary	59	-0.2	0.9	-2.61	1.62	46	-0.46	0.91	-3.33	1.11
spelling	59	-0.67	0.63	-2.3	-0.03	46	-0.88	0.81	-3.64	-0.12
PPK	59	-0.08	0.83	-1.25	2.02	46	-0.18	0.88	-1.56	2.47
NW reading	59	-0.32	0.89	-2.39	1.22	46	-0.14	1.12	-2.24	2.88
mean RT cognates	4599	487.35	110.57	269.1	1537.4	3699	536.66	99.18	329.8	1456.6

Note . Vocabulary = vocabulary knowledge, spelling = spelling ability, PPK = Primary Print Knowledge (measure of print exposure), NW reading = nonword reading/decoding ability

Table 8.3 gives a rank order for ID groups in naming times of the cognate stimuli within each language group. It is important to notice, that strong decoders were the fastest readers in both languages (English group: M = 462.08 (SD = 96.08) msec; German group: M = 508.8 (SD = 90.02) msec). Weak decoders turned out to be the slowest to name words in both English and German. The difference between the two languages seemed to be the ranking place of the vocabulary and spelling groups. In the English language group, the high vocabulary group was second fastest, and the low vocabulary group second slowest. In the German language group, the vocabulary groups held the middle ranks. Instead, the strong spellers group was second fastest and the low spellers group second slowest.

Table 8.3

Average naming reaction times for each ID group per language, ordered from shortest to longest

English language group			German language group		
ID score group	*n*	RT (msec)	ID score group	*n*	RT (msec)
hd	54	462.08	hd	56	508.80
hv	48	465.26	hs	58	511.94
hp	49	469.18	hp	39	514.34
hs	45	472.24	hv	50	517.65
ls	59	487.35	lv	54	527.57
lp	55	491.21	lp	65	527.90
lv	56	494.4	ls	46	536.66
ld	50	501.06	ld	48	539.04

Note. n = number of participants per group; RT(msec) = average naming reaction time in milliseconds; Groups of high scores are hp = high print exposure, hv = high vocabulary knowledge, hd = high nonword reading/decoding score, hs = high spelling score; groups of low scores are lp = low print exposure, lv = low vocabulary knowledge, ld = low nonword reading/decoding score, ls = low spelling score.

In sum, although group memberships were combined in all possible ways, there still seemed to be some grouping tendencies for individual differences. Within both language groups, there was a tendency for participants to either fall mostly into all lower score ID groups, or all higher score ID groups. Mean RTs for all higher score ID groups were generally shorter than mean RTs for lower score ID groups. When ranking the ID groups within languages, the strong decoder group had the fastest average naming RT, and the weak decoder group the slowest. This was true for both languages. The languages differed in terms of the ranking of the vocabulary ID groups. In English, the high-vocabulary group was second fastest, and the low-vocabulary group was second slowest. In contrast, in German these places were held by the spelling groups. A further difference between languages was found in the ID group membership combinations. German participants were more often classified as low-print but high score on all other tasks, whereas English readers were more often classified as high-print and with strong vocabulary knowledge but low on decoding and spelling ability.

8.2 Results

Each of the eight individual differences group was submitted to the same model as reported in the previous analysis (see section 7.4). The mixed models included onset and stress variables, a categorical variable to indicate language (English and German), a selection of the most important psycholinguistic variables (frequency, AoA, composite FF and FB rime consistency, length in letters, old20, phonographic *N* for rime 1), and the individual differences variables print exposure (PPK), vocabulary, nonword reading (decoding) and spelling. All models included random intercepts of items and participants. Random slopes were included as long as they improved model fit, as estimated using the RePsychLing package (Baayen et al., 2015). For all eight models, zero-correlation parameter models are reported, i.e. the correlation parameters were not included in the models. The model specifications for each group can be found in Appendix 11.16.

Table 8.4 gives an overview of the complete results in terms of estimates and significance levels. The results will be discussed in the next section. Please note that only those effects which have reached significance at the alpha = .005 level have been flagged with asterisks.

Table 8.4

Estimates of zero-correlation mixed models on naming RTs of 85 cognates for groups split according to performance above or at/below the standardised means of individual differences tasks' performances.

	High print	Low print	High vocabulary	Low vocabulary	Strong decoders	Weak decoders	Strong spellers	Weak spellers
	n = 88	*n* = 120	*n* = 98	*n* = 110	*n* = 110	*n* = 98	*n* = 103	*n* = 105
(Intercept)	602.468 ***	573.842 ***	573.800 ***	592.803 ***	571.214 ***	608.180 ***	569.173 ***	600.488 ***
bilabial *(present)*	-127.622 ***	-104.475 ***	-109.400 ***	-118.356 ***	-114.860 ***	-116.006 ***	-106.920 ***	-120.506 ***
labiodental *(present)*	-104.330 ***	-93.266 ***	-94.750 ***	-100.774 ***	-94.426 ***	-103.791 ***	-88.847 ***	-107.189 ***
alveolar and postalveolar *(present)*	-126.785 ***	-107.429 ***	-109.100 ***	-121.770 ***	-115.604 ***	-117.792 ***	-108.047 ***	-122.740 ***
palatal *(present)*	-79.093 **	-47.871	-83.290 **	-43.451	-81.302 **	-44.907	-78.927 **	-42.528
velar *(present)*	-99.444 ***	-74.604 ***	-84.280 ***	-86.035 ***	-88.253 ***	-83.604 ***	-82.024 ***	-86.345 ***
glottal *(present)*	-124.072 ***	-99.228 ***	-104.700 ***	-113.983 ***	-106.923 ***	-114.142 ***	-94.024 ***	-125.223 ***
plosive *(present)*	20.155 .	5.125	19.410 .	4.643	16.962	6.216	15.552	6.449
nasal *(present)*	24.998 *	1.172	16.790	5.315	13.265	9.800	13.950	6.539
fricative *(present)*	-24.418 *	-33.243 *	-21.440 .	-37.176 *	-29.656 *	-29.718 *	-32.073 **	-27.086 .
approximant *(present)*	73.586 ***	41.977 .	63.720 **	48.151 *	68.243 **	43.987 .	60.495 **	49.929 *
voiced *(present)*	24.141 ***	25.352 ***	18.740 ***	28.816 ***	22.977 ***	25.130 ***	22.736 ***	24.808 ***
affricate *(present)*	35.274 .	44.069 *	39.870 *	42.004 .	46.786 *	35.110	43.848 *	40.848 .
short vowel *(present)*	-87.160 ***	-65.561 **	-74.360 ***	-75.350 **	-72.595 ***	-79.328 ***	-67.409 ***	-82.231 ***
long vowel or diphtong *(present)*	-67.795 **	-69.751 **	-53.970 **	-83.932 ***	-68.401 ***	-71.999 **	-65.362 ***	-74.561 **
stress *(2nd syllable)*	-11.651 *	-22.627 ***	-14.030 **	-20.130 ***	-19.671 ***	-14.286 **	-16.779 ***	-17.552 ***
Language *(G)*	40.580 *	56.833 **	59.280 **	35.416 *	57.703 ***	41.092 *	43.973 *	39.843 *
PPK	-12.790	-21.146	-10.220	-3.844	-15.630 .	2.667	-13.726	0.326
vocabulary	-15.525	-11.887	-9.709	-20.481	0.429	-31.617 **	-2.505	-28.663 *
NW reading	-15.866 .	-21.219 **	-7.583	-24.881 **	-3.531	-6.774	-16.221 .	-16.167 .
spelling	8.366	3.629	11.620	-1.062	5.225	4.364	13.231	6.215
Zipf	-5.494 *	-7.755 .	-4.836	-7.477 .	-6.687 *	-7.047 .	-9.119 **	-9.692 **
comp FFR cons	-5.962 **	-1.482	-7.271 **	-6.110 *	-2.500	-11.242 ***	-5.318 .	-10.523 ***
comp FBR cons	-1.396	-7.588 *	-1.793	-3.577	-1.146	-4.209	-2.693	1.175
AoA	6.711 *	5.194	9.046 **	6.286 .	4.873	10.629 **	2.955	11.281 ***
length	5.029	1.348	0.706	9.194	5.768	1.283	-0.844	6.118
old20	5.793	13.869 *	10.900 *	6.594	15.314 ***	4.118	13.632 **	9.008 .
phonographic N rime 1	4.616 .	8.386 *	6.729 **	1.594	-0.074	5.612 .	5.058 .	0.657
Language (*G*) × PPK	0.996	21.955	-4.855	12.049	2.057	-0.946	5.056	-1.245
Language (*G*) × vocabulary	2.784	14.981	-6.455	9.303	-3.595	27.967 .	6.803	17.577
Language (*G*) × nonword reading	-6.753	11.160	-11.590	10.570	-16.722	10.423	2.741	-2.758
Language (*G*) × spelling	-14.598	-14.226	-7.371	-14.597	-11.621	-15.191	-16.032	-9.339
Language (*G*) × Zipf	1.993	6.745	1.475	4.275	2.210	3.616	0.763	7.997 .
Language (*G*) × comp FFR cons	3.583	-7.486 .	3.049	-5.152	-4.781	4.399	4.977	-3.116
Language (*G*) × comp FBR cons	5.609	2.824	1.185	-3.221	0.516	-1.706	-1.463	-3.657
Language (*G*) × AoA	0.479	11.411 *	0.743	3.567	4.999	1.917	1.342	2.008
Language (*G*) × length	8.485 .	10.789 .	8.610 .	-1.235	5.505	7.350	10.462 .	-0.042
Language (*G*) × old20	1.597	-3.203	1.744	4.196	-3.924	3.839	-0.392	-0.383
Language (*G*) × phonographic N	-7.735 *	-12.745 **	-10.540 **	-4.202	-3.166	-7.601 *	-8.046 *	-2.640
PPK × Zipf frequency	2.166	2.427	2.013	3.536	2.689 .	1.799	1.311	4.563 .
PPK × comp FFR cons	1.622	6.629 .	-0.163	3.875	0.964	-0.048	-0.102	2.361
PPK × comp FBR cons	-1.379	-6.385	-0.183	-4.027	0.245	-2.839	1.435	-6.090 **
PPK × AoA	2.346	-4.234	1.527	-3.179	1.127	-1.361	0.756	-0.456
PPK × length	-3.419	-9.160	-3.763 .	-4.280	-3.987 .	-6.065 .	-4.790 *	-5.726 .
PPK × old20	1.628	6.287	0.110	1.975	2.407	1.289	3.233	-0.482
PPK × phonographic N	1.727	4.173	-0.009	1.102	-1.929	1.800	-2.379	4.185 *
vocabulary × Zipf	-1.307	1.484	-1.336	1.298	-1.174	2.333	-1.294	0.964

vocabulary × comp FFR cons	0.531	3.545 *	3.264	1.888	-0.718	5.547 **	3.430 .	0.061
vocabulary × comp FBR cons	-0.842	-1.581	-0.987	-0.420	-1.884	-0.287	-2.482	0.366
vocabulary × AoA	-1.273	-0.111	-2.283	0.706	-1.440	1.279	-1.118	-1.804
vocabulary × length	2.968	8.058 **	7.284 *	9.260 *	4.273 .	8.579 *	5.523 .	6.905 *
vocabulary × old20	-3.472	-6.154 *	-6.427	-8.461 .	-5.971 *	-5.842	-5.967 .	-5.952 .
vocabulary × phonographic N	-1.977	1.245	-1.505	-1.644	0.449	-1.704	3.681 .	-4.595 *
NW reading × Zipf	0.165	1.733	-1.476	3.014	-2.352	-1.115	0.050	1.849
NW reading × comp FFR cons	0.261	1.616	-1.694	2.630 .	-0.802	-1.791	1.309	0.644
NW reading × comp FBR cons	-0.548	-0.959	-1.341	-0.099	-0.698	-2.157	-1.507	-0.518
NW reading × AoA	-2.553	-1.442	-3.684 *	-0.100	0.152	-2.356	-3.733 .	-0.063
NW reading × length	5.039 *	4.750 *	6.991 **	4.017 .	3.390	4.164	8.131 **	2.151
NW reading × old20	-0.711	-1.560	-4.222 .	0.135	-5.336	-6.867 .	-5.519 *	1.735
NW reading × phonographic N	-2.827 .	-1.391	-2.734 .	-1.617	2.635	-1.011	0.448	-3.364 *
spelling × Zipf	0.738	0.759	2.026	-0.342	1.466	0.664	4.080	-2.529
spelling × comp FFR cons	-0.185	-2.985 .	1.458	-4.694 **	0.062	-3.634 .	-0.562	-4.201 .
spelling × comp FBR cons	1.454	-0.185	-0.296	1.238	-1.190	2.300	0.115	4.927 *
spelling × AoA	-0.963	-1.660	1.387	-3.914 *	-2.297	-0.424	3.271	-0.231
spelling × length	-1.656	-4.729 .	-4.016 .	-2.343	0.024	-8.571 **	-0.386	-2.275
spelling × old20	1.156	1.216	1.943	0.630	-2.574	8.835 *	-0.804	2.614
spelling × phonographic N	0.533	-3.927 *	1.795	-5.286 **	0.080	-5.365 *	-3.106	-5.931 *
Lang *(G)* × PPK × Zipf	-3.342	2.143	-1.423	-3.228	-2.338	-1.435	-2.021	-2.522
Lang *(G)* × PPK × comp FFR cons	-5.538 .	-9.974 *	-0.207	-3.985	0.485	-2.785	-0.333	-2.275
Lang *(G)* × PPK × comp FBR cons	-1.981	8.789 .	-0.621	7.636 *	-0.662	4.754	-1.954	9.774 ***
Lang *(G)* × PPK × AoA	-4.099	10.045 .	-0.549	-0.072	1.046	-2.097	-2.098	-0.493
Lang *(G)* × PPK × length	-2.794	10.379	-0.004	4.327	2.865	3.941	1.194	7.583 .
Lang *(G)* × PPK × old20	1.417	-3.552	3.567	-2.078	-1.577	-0.149	-0.949	0.828
Lang *(G)* × PPK × phonographic N	-1.443	-7.484	-0.235	-1.726	1.363	-2.347	2.496	-5.604 *
Lang *(G)* × vocabulary × Zipf	0.962	1.147	0.914	0.796	1.885	1.122	2.494	1.412
Lang *(G)* × vocabulary × comp FFR cons	0.094	-6.522 **	-4.779	-7.479 *	-0.713	-7.700 **	-5.874 *	-3.244
Lang *(G)* × vocabulary × comp FBR cons	3.109	1.197	2.123	-2.868	-0.482	5.051	5.422 *	-2.451
Lang *(G)* × vocabulary × AoA	3.212	-1.015	1.183	-1.286	1.576	-2.143	1.424	1.536
Lang *(G)* × vocabulary × length	-2.819	-8.127 *	-7.971 .	-10.650 *	-4.663	-10.555 *	-5.929	-9.144 *
Lang *(G)* × vocabulary × old20	3.507	3.125	2.507	6.267	1.606	9.630 *	3.641	5.146
Lang *(G)* × vocabulary × phonographic N	-0.143	-1.764	2.430	0.667	1.253	-2.023	-5.877 *	5.329 *
Lang *(G)* × NW reading × Zipf	3.582	-0.994	2.326	-0.569	0.942	0.049	1.862	-0.228
Lang *(G)* × NW reading × comp FFR cons	0.292	-1.200	3.015	-3.502 .	4.347	2.849	-0.109	-0.054
Lang *(G)* × NW reading × comp FBR cons	-2.339	1.783	-0.919	1.361	-0.239	-0.619	-1.433	1.664
Lang *(G)* × NW reading × AoA	3.505	-0.408	3.574	-2.068	-4.478	-0.271	3.304	-1.173
Lang *(G)* × NW reading × length	-8.517 **	-6.667 *	-7.939 **	-8.226 **	-7.708	-6.051	-8.475 *	-6.128 *
Lang *(G)* × NW reading × old20	2.657	3.033	4.555	2.682	8.980 .	6.505	4.552	2.137
Lang *(G)* × NW reading × phonographic N	3.787 .	1.690	4.148 *	1.484	-1.535	2.873	0.435	4.062 *
Lang *(G)* × spelling × Zipf	0.272	0.260	-1.003	1.029	-1.322	0.294	-0.676	2.971
Lang *(G)* × spelling × comp FFR cons	0.383	3.080	-2.591	5.255 *	-0.665	4.913 *	-1.475	1.665
Lang *(G)* × spelling × comp FBR cons	-1.993	1.339	0.450	0.361	1.761	-0.900	0.621	-3.972
Lang *(G)* × spelling × AoA	0.560	0.033	-1.208	0.819	0.684	-1.260	-1.136	-2.260
Lang *(G)* × spelling × length	-1.315	2.144	3.964	-1.849	0.291	4.373	-0.398	-2.506
Lang *(G)* × spelling × old20	0.586	-2.098	-5.939 .	1.082	0.459	-7.791 .	-2.270	-4.466
Lang *(G)* × spelling × phonographic N	0.261	4.399 *	-3.562 .	7.380 ***	-2.818	7.587 **	4.688	6.831 *

Note . () indicate level shown for categorical variables, such as (G) = German; NW reading = nonword reading, Zipf = Zipf frequency, comp FFR cons = composite feedforward rime consistency, comp FBR cons = composite feedback rime consistency, length = length in letters, old20 = orthographic Levensthein distance 20, phonographic N= phonographic neighbourhood for rime 1, Lang = language;

*** p < .001, ** p < .01, * p < .05, . p < .1

The inclusion of individual differences in the eight score model groups significantly added to the model. Table 8.5 shows the results of the likelihood ratio models when comparing R^2 between the models without and with the predictors print exposure, vocabulary, decoding ability and spelling ability. All models apart from the high print model improved significantly.

Table 8.5

R^2 for zero-correlation mixed models of naming RTs on 85 cognates for individual differences groups excluding and including individual differences in the models

ind diff group	R^2 without ind diff	R^2 with ind diff	R^2 increase	LRT result
High print	.19	.27	.08	χ^2(65, N = 88) = 78.98
Low print	.20	.23	.03	χ^2(64, N = 120) = 126.65 ***
High vocabulary	.20	.25	.05	χ^2(64, N = 98) = 95.96 **
Low vocabulary	.20	.26	.06	χ^2(64, N = 110) = 160.37 ***
Strong decoders	.21	.25	.04	χ^2(66, N = 110) = 120.51 ***
Weak decoders	.19	.23	.04	χ^2(65, N = 98) = 144.42 ***
Strong spellers	.18	.22	.04	χ^2(66, N = 103) = 93.93 *
Weak spellers	.22	.28	.06	χ^2(64, N = 105) = 151.66 ***

Note . Ind diff = individual differences (print exposure, vocabulary knowledge, nonword reading, spelling), R^2 = marginal R^2 for fixed effects (Barton, 2018), LRT = likelihood ratio test. * = $p < .05$, ** = $p < .01$, *** = $p < .001$

8.3 Discussion

The present set of analyses was designed to investigate in which way individual differences modulated reading in two languages differing in spelling-sound transparency. The individual differences investigated were print exposure (as a measure for reading practice), vocabulary knowledge (as an indicator for semantic knowledge), nonword reading (as an index for O-P pathway efficiency) and spelling ability (to tap into orthographic lexical quality).

The previous analysis reported in this thesis (see section 7.4, Analysis I) suggested that language remained a strong predictor in reading aloud, even when individual differences were taken into account. Thus, both language and individual differences influenced the reading process. Word length emerged as a particular psycholinguistic variable of interest with regard to modulation by IDs. Some IDs (reading practice and orthographic lexical quality) influenced reading similarly across languages for words of various lengths, whereas other IDs (decoding skill and semantic knowledge) seemed to have differential impact between languages. Specifically, whilst German readers seemed to benefit more from better decoding skills with increasing word length, for English readers better decoding skills seemed to be more beneficial for shorter than longer words. Moreover, English readers who had better vocabulary knowledge seemed to benefit from this when reading shorter rather than longer words. In contrast, for German readers, vocabulary knowledge did not seemingly interact with word length.

The present set of mixed-effects models aimed at looking more closely at how reading differed across languages within groups which showed either strengths or weaknesses in a particular ID. This approach was chosen to detect the finer relationships between language, psycholinguistic effects and IDs (see, for example, Andrews, 2012). For this purpose, the same data set of 104 English-speaking and 104 German-speaking skilled readers was used to divide participants into groups of their relative strength or weakness in individual differences tasks. This resulted in eight separate analyses. For each individual difference task, participants had been divided into high and low achievers on this particular task. This meant that the data of each participant was submitted for analysis four times, once for each individual difference task in either the high- or the low-score group. Participants, who scored above the standardised mean were considered high-score individuals, whilst all others fell into the respective lower-score groups. No distinct reader profiles were found. However, there was a tendency for readers to belong to either most of the high-score or most of the low-score groups.

In the following, the results of the eight mixed-effects models for the ID score groups will be discussed. The wealth of results dictates that some choices had to be made; these have been guided by the literature review to this thesis, and will be reported in the following order: First of all, overall trends will be highlighted (see sections 8.3.1 and 8.3.2). Second, results will be discussed in reference to findings from previous cross-language research (see section 8.3.3). Third, pertinent result patterns will be discussed for the different ID groups (sections 8.3.4 to 8.3.8).

8.3.1 Language and individual differences modulate reading aloud performance

The main focus of the present study was to examine whether cross-language differences in reading aloud could be offset by individual differences, and analysis I (see section 7.4) showed that this was indeed not the case. Crucially, in the present set of analyses, a language main effect emerged in all eight analyses of participants grouped according to their ID strengths and weaknesses. This underlines once more that even when looking at less varied participant groups, the difference between languages was still present in so far that German naming times were longer than English naming times.

However, likelihood ratio model comparisons also exposed that for all ID groups – except for the high- print group - the inclusion of individual differences as fixed effects led to a significantly improved model to explain naming RTs (see Table 8.5). Thus, even when participants were grouped according to their relative strengths or weaknesses, the models were still improved when taking into account of how they performed on the other ID tasks. This suggests that ‘reader profiles’ are complex and that each reader characteristic plays a role in the overall outcome.

The high-print group was the only group where the inclusion of individual differences as fixed effects did not improve the model. It is possible that participants with relative more reading experience were at an advantage over all other ID groups due to their greater practice, and therefore effects from other IDs may not have been advantageous enough to improve the model. This would be in line with the fundamental assumption of connectionist models that practice or frequency of activation shapes and improves model structures (see e.g., Plaut et al., 1996). However, the data did not support this explanation entirely. It was the strong decoder group, which emerged as fastest readers, not the high-print group. It thus seems more likely, that since the high-print group had the least number of participants out of all ID groups, this may have reduced the power to detect the effects of the individual differences. Only 88 participants belonged to this group, whereas participant numbers ranged between 98 and 120 for all other ID groups (see Table 8.5).

In sum, within the ID groups, language differences remained and in almost all groups task performance on other individual differences contributed significantly to explaining reading aloud.

8.3.2 More similarities across languages for skilled readers compared to less skilled readers

Table 8.6 gives an overview of psycholinguistic effects modulated either by language or by both language and individual differences in the different ID groups. The significance values given are based on computations with the R lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017). Interactions of language with psycholinguistic effects and/or individual differences were more numerous in the low-score groups than in the high-score groups. It thus seems that for readers in the high-score groups, reading was more similar across languages, whereas for readers in the low-score groups, reading performance differed in terms of language and/or individual differences effects. This indicates that relatively more skilled readers are more alike across languages than less skilled readers. Research has found that more skilled readers show fewer psycholinguistic effects (e.g., Yap et al., 2012). The current study extends this by showing that the reduction in psycholinguistic effects can also be observed in skilled readers of different languages.

Table 8.6

Significant interactions of language with psycholinguistic variables and/or individual differences

	hp	lp	hv	lv	hd	ld	hs	ls
Lang (*G*) × PPK								
Lang (*G*) × vocabulary								
Lang (*G*) × decoding								
Lang (*G*) × spelling								
Lang (*G*) × Zipf								
Lang (*G*) × comp FFR cons								
Lang (*G*) × comp FBR cons								
Lang (*G*) × AoA		*						
Lang (*G*) × length								
Lang (*G*) × old20								
Lang (*G*) × PNR1	*	**	**			*	*	
Lang *(G)* × PPK × Zipf								
Lang *(G)* × PPK × comp FFR cons		*						
Lang *(G)* × PPK × comp FBR cons				*				***
Lang *(G)* × PPK × AoA								
Lang *(G)* × PPK × length								
Lang *(G)* × PPK × old20								
Lang *(G)* × PPK × PNR1								*
Lang *(G)* × vocabulary × Zipf								
Lang *(G)* × vocabulary × comp FFR cons		**		*		**	*	
Lang *(G)* × vocabulary × comp FBR cons							*	
Lang *(G)* × vocabulary × AoA								
Lang *(G)* × vocabulary × length		*		*		*		*
Lang *(G)* × vocabulary × old20						*		
Lang *(G)* × vocabulary × PNR1							*	*
Lang *(G)* × decoding × Zipf								
Lang *(G)* × decoding × comp FFR cons								
Lang *(G)* × decoding × comp FBR cons								
Lang *(G)* × decoding × AoA								
Lang *(G)* × decoding × length	**	*	**	**			*	*
Lang *(G)* × decoding × old20								
Lang *(G)* × decoding × PNR1			*					*
Lang *(G)* × spelling × Zipf								
Lang *(G)* × spelling × comp FFR cons				*		*		
Lang *(G)* × spelling × comp FBR cons								
Lang *(G)* × spelling × AoA								
Lang *(G)* × spelling × length								
Lang *(G)* × spelling × old20								
Lang *(G)* × spelling × PNR1		*		***		**		*

Note . Groups of high scores are hpr = high print exposure, hv = high vocabulary knowledge, hd = high decoding score, hs = high spelling score; groups of low scores are lp = low print exposure, lv = low vocabulary knowledge, ld = low decoding score, ls = low spelling score. Variables are Lang = language, PPK = primary print exposure, decoding = nonword reading ability, comp FFR cons = composite feedforward rime consistency, comp FB cons = composite feedback rime consistency, length = length in letters, PNR1 = phonographic rime neighbours of rime 1. * = p < .05, ** = p < .01, *** = p < .001

Importantly, for the different ID groups, there were also different interaction patterns between language x psycholinguistic variables x ID variables, highlighting how reading aloud differs depending on the ID strengths and weaknesses. Andrews (2012), amongst others, has called for the inclusion of individual differences in reading models, as the models which simulate average reading behaviour may conceal important differences to improve our understanding of word recognition and reading. As will become clearer in the following, the present results demonstrate very clearly that the results from participants with a particular strength and weakness produce different effect patterns for each one of these groups.

8.3.3 The language x length interaction is modulated by decoding skill in both languages

The word length effect has been taken to indicate small-unit processing in readers of transparent languages (PGST; Ziegler & Goswami, 2005). As reviewed in Chapter 3, evidence suggests that developing readers of German, but less so developing readers of English, show a length effect in oral reading (Frith et al., 1998; H. Wimmer & Goswami, 1994). However, in skilled readers there exists contradictory evidence for slower word (or nonword) naming by readers of transparent orthographies compared to readers of opaque orthographies (e.g., Rau et al., 2015; Spinelli et al., 2005; Yap & Balota, 2009; Ziegler et al., 2001). Thus, skilled readers of transparent scripts do not seem to have reliably retained a developmental footprint of more small-unit processing than readers of opaque orthographies, as suggested by the PGST.

In the first analysis on the complete data set (see Chapter 7), it emerged that length in particular was subject to modulation by language and IDs. More reading experience and better spelling ability seemed to be more beneficial with increasing word length in both language groups. However, decoding skills and vocabulary knowledge interacted differently with length in the two language groups. Whilst vocabulary knowledge seemed to be mostly beneficial to English readers for shorter words, word length did not seem to interact with vocabulary knowledge in German. In contrast, German readers benefitted from decoding skills at all word lengths, but English readers seemed to mostly benefit from better decoding skills when reading shorter words.

These results were generally replicated, when the data set was split four times into weaker and stronger ID score groups. In the eight mixed – effects models on the same participants grouped according to their strengths and weaknesses of IDs, no length main effect was found, nor did any of the language x length interactions reach significance. However, the analysis did find that all low-score reading groups showed a significant 3-way interaction between language x vocabulary knowledge x length. As before, greater semantic knowledge did not have a measurable effect at any word length for German low-score participants. For English low-score participants, on the other hand, greater semantic knowledge facilitated shorter rather than longer word reading. This applied to all low-score groups, namely low – print, low-vocabulary, weak decoders and weak spellers.

Why would semantic knowledge only be facilitating for shorter words for English low-score groups? As these readers are reading an opaque script, they are more likely than readers of a transparent script to take recourse to semantics for reading (as assumed, for example, in the ODH). It is conceivable that for these low-score reader groups reading an opaque script, semantic knowledge is activated fast enough for shorter words to act as a facilitator, but too slow for longer words. This would lead to greater semantic knowledge facilitating shorter rather than longer word reading.

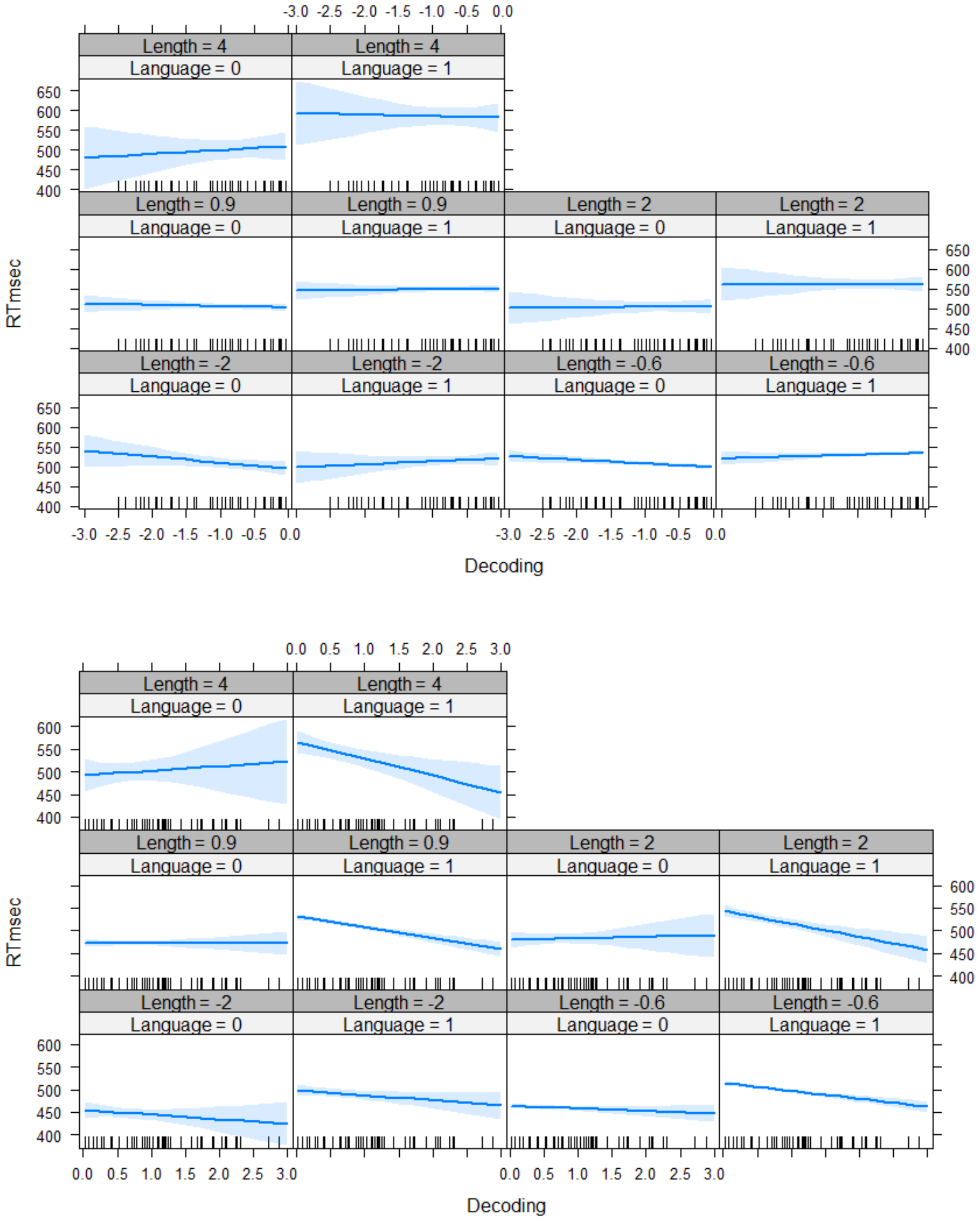


Figure 8.1. Graph of language x decoding x length interactions for the weak decoder group (upper panel, coefficient = -6.051, ns) and the strong decoder group (lower panel, coefficient = -7.708, ns). Language 0 = English; Language 1 = German. Decoding skill was standardised within language group and higher values indicate greater skill.

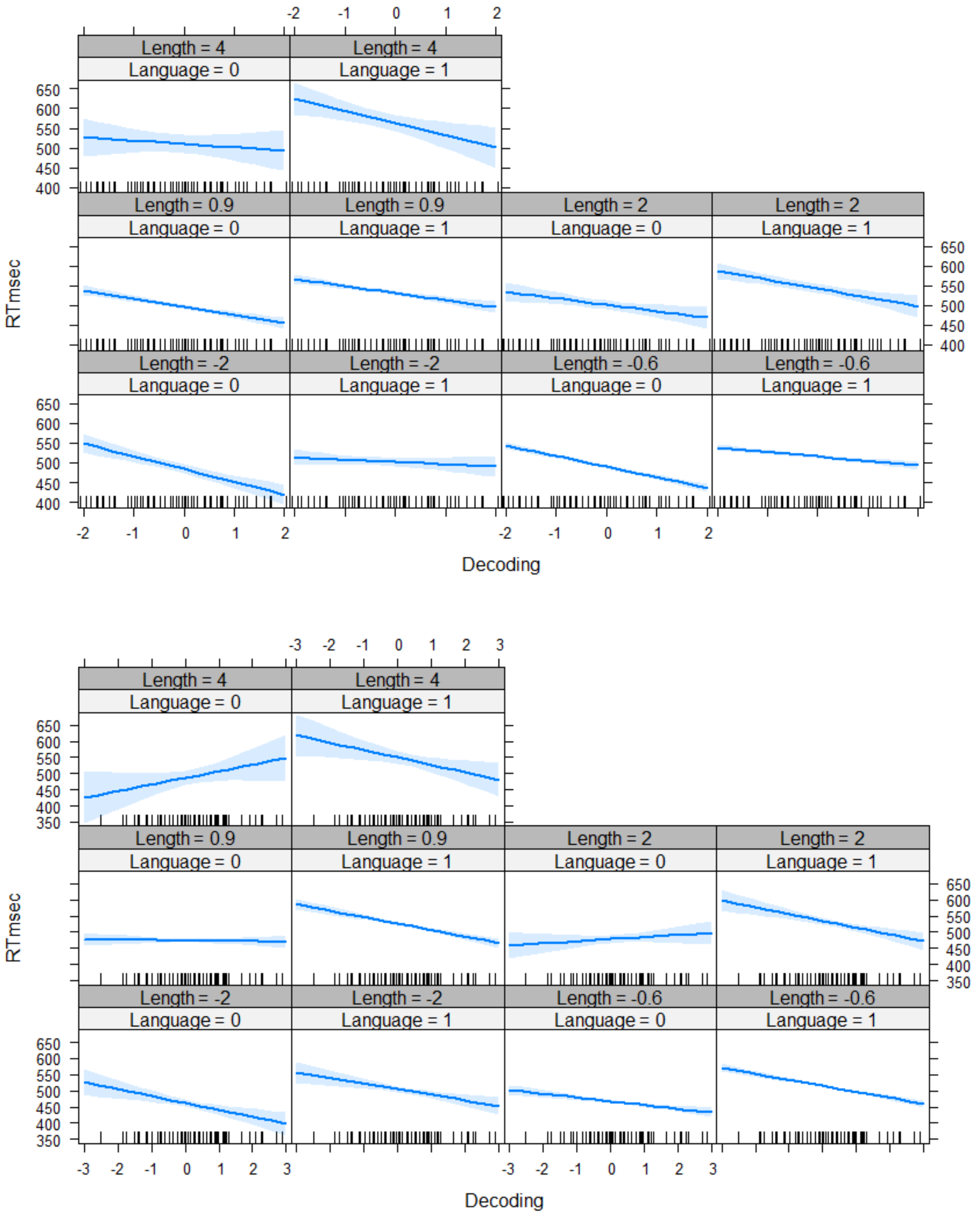


Figure 8.2. Graph of language x decoding x length interactions for the weak vocabulary group (upper panel, coefficient = -8.226, p = .007) and the strong vocabulary group (lower panel, coefficient = -7.939, p = .006). Language 0 = English; Language 1 = German. Decoding skill was standardised within language group and higher values indicate greater skill.

Apart from the two models based on strong and weak decoding scores (see Figure 8.1), all models also showed significant language x decoding x length interactions (see as an example Figure 8.2 for the two vocabulary groups). In English, better decoding skill facilitated naming of short words for all these groups. Only for the high-vocabulary and strong speller groups did decoding ability also inhibit naming for longer words in English. In German, decoding facilitated naming of all lengths for the high-print, high-vocab, low-vocab and strong speller groups. Only in the case of German low-print and weak spellers, better decoding ability facilitated reading of long words only. To summarize this rather complicated set of results, it seems that generally across all reader groups, decoding skills were helpful for short words for English participants, and for all words for German participants.

However, there was an exception to this finding. No language x decoding x length interaction was found in the strong and weak decoder groups which had been categorized according to nonword reading ability. In fact, given that the language x decoding x length interactions within the strong and weak decoder groups did not reach significance, two important possibilities need to be considered. First, it seems that the decoding skill rather than language (transparency) determines the emergence of a length effect. This is interesting, because it points towards decoding as the most crucial modulator of the length effect, even overriding language differences. Reading across languages seems to be more similar in terms of word processing when taking into account individual decoding skills. Second, it is possible that good decoders are better at finding the correct grain size than weaker decoders in both languages, which is why the language x decoding x length interaction did not reach significance in this group. Strong decoding, then, for English participants becomes a measure of expertise in finding the correct grain size. The results presented here are then compatible with the PGST in so far that it assumes that readers of transparent orthographies tend to use small grain sizes, whereas readers of opaque orthographies use more varied reading unit sizes, which is harder to achieve. If decoding skill taps into the ability to find the optimal grain size, then the present results add a person-level element to the PGST as the strong decoders' expertise of finding the correct grain sizes use (which very likely also includes word segmentation into optimal grain sizes) influences the reading process and modulates the length effect.

In sum, the PGST suggests that readers of transparent orthographies use smaller unit sizes for reading than readers of opaque scripts. Stronger length effects (reading times increase with word length) for the former have been considered as evidence supporting this claim. The present analysis showed that the language x length interaction was modulated by decoding (nonword reading) skill and vocabulary knowledge. In the latter case, English low-score groups benefitted from greater vocabulary knowledge when reading shorter words, whereas there seemed to be no effect for German participants. In the case of decoding, English participants benefitted from better decoding skills for shorter words, and German participants for all word lengths. Importantly, when participants were grouped into stronger and weaker decoders, the language x decoding x length interaction was no longer significant. Decoding skill then becomes the central reason for the occurrence of the length effect, ahead of language differences. This could mean that good decoders have similar unit processing across languages. Alternatively, and with reference to the PGST, it may imply that decoding skill determines the proficiency of finding the correct grain size, thus adding an individual-level element to the account of the PGST, which assumes that grain sizes are determined by the language transparency.

8.3.4 Strong nonword decoders as most efficient readers

Strong nonword decoding seems to indicate most efficient reading skill even amongst skilled readers and irrespective of language. The present study has produced a number of results which support this claim.

First, in the present analysis, the strong decoder group emerged as the group with the lowest number of psycholinguistic effects and the least number of interactions with either language or IDs or both. They were also the fastest readers out of all ID groups. By contrast, the weak decoder group appeared to have the slowest latencies. These results were true for both languages. This is congruent with previous findings that participants, who are better at nonword reading, are generally faster readers for both words and nonwords (Aaron et al., 1999; P. Brown et al., 1994; Davies et al., 2017; Martin-Chang et al., 2014; Stanovich & West, 1989; Torgesen et al., 1999). The current study confirms this for the first time for two languages differing in spelling-sound transparency.

Second, the case for the importance for decoding skill to mark skilled reading can also be supported by the fact that only when participants were grouped as strong decoders, no effect of FFR consistency could be detected. In other words, for this group, inconsistent words were not read significantly more slowly than consistent words – in neither language. As outlined in Chapter 2, PDP model simulations have shown that as the reading system is built by learning to map orthography to phonology, words with inconsistent spelling-sound mappings take longer to learn (e.g., Monaghan & Ellis, 2010). These model simulations were based on networks learning to map orthography to phonology, thereby locating the consistency effects in the direct route from orthography to phonology. Consequently, very skilled usage of O-P mappings is reflected in a reduction of the consistency effect (e.g., Strain & Herdman, 1999). The fact that the strong decoder group was the only ID group where a consistency effect could not be detected indicates that only strong decoders had assumed such an automated processing from orthography to phonology that the consistency effect did not emerge. Importantly, this seemed to be true for readers of both languages.

The above results are important. By not detecting a consistency effect for this groups, this indicates that they had a very efficient direct route. The fact that this group were also the fastest readers indicates that the efficiency of the direct route seems the most crucial element to excel in skilled reading. These findings suggest that decoding skills are not only important for learning to read, but continue to provide the most efficient basis for mastering skilled reading. It suggests that the O-P pathway seems to remain the dominant reading pathway in both transparent and opaque languages.

Some may argue that the absence of an effect cannot be considered as evidence. This concern can be addressed in two ways. First of all, the absence of the consistency effect was interpreted as an index for a very efficient direct route in reference to other separate analyses, each one sharing some of the same participants. Thus, the absence of the consistency effect in one group (strong decoders), but its existence in others (high-print, strong vocabulary), will necessarily have to be interpreted with reference to each other, and it needs to be outlined both what it means to find a consistency effect, and correspondingly, what it means when the effect is not apparent. Secondly, critics may argue that a consistency effect would be detected for the strong decoder group when power was increased, for example by including a larger number of participants. This may be the case, but it does not change the argument, because in that case the author would expect that the consistency effect would also increase for the other groups, if comparable analyses were executed. As such, the importance of the result does not lie with the null result as such, but in the comparative effect patterns of participant groups who were selected for their strengths and weaknesses of a particular ID task.

Returning to the task on hand of enumerating supporting evidence of the assumption that strong nonword decoding seems to indicate the most efficient reading skill amongst skilled readers, there is a third piece of evidence which should be mentioned. As discussed previously (see section 4.2.4), AoA effects can be considered a marker for semantic involvement, as studies have reported AoA effects when tasks required access to meaning (e.g., Cortese & Khanna, 2007; Davies et al., 2014). The strong decoder group did not show an AoA effect, hence suggesting that this reader group did not measurably need to resort to the indirect pathway via semantics to read stimuli. These results are in congruence with findings by Strain & Herdman (1999), who found that participants with strong phonological skills (nonword reading and blending) were less likely to show semantic effects in reading aloud. The authors had argued that their O-P pathways were more efficient than of those participants who had lower phonological reading ability. The present results seem to support this assumption.

In sum, the combined results of fastest reading latencies, and lack of measurable consistency and AoA effects point to a very efficient O-P pathway, which produces the fastest readers and is not in need of semantic assistance in both language groups.

8.3.5 Weak decoders show differences between languages and vocabulary knowledge

The weak decoders, on the other hand, showed two important interactions with language and individual differences, both of which were reminiscent of the relevant interactions observed in Analysis 1.

First, the weak decoder group showed the language x vocabulary x FFR consistency interaction, as can be seen in the upper panel of Figure 8.3. In English, words were read faster by participants with high vocabulary knowledge, but the effect was reduced with increasing consistency. In German, there did not seem to be an effect. The lower panel of Figure 8.3 shows the same interaction for the strong decoder group for reference purposes. In the strong decoder group, this 3-way interaction was not significant.

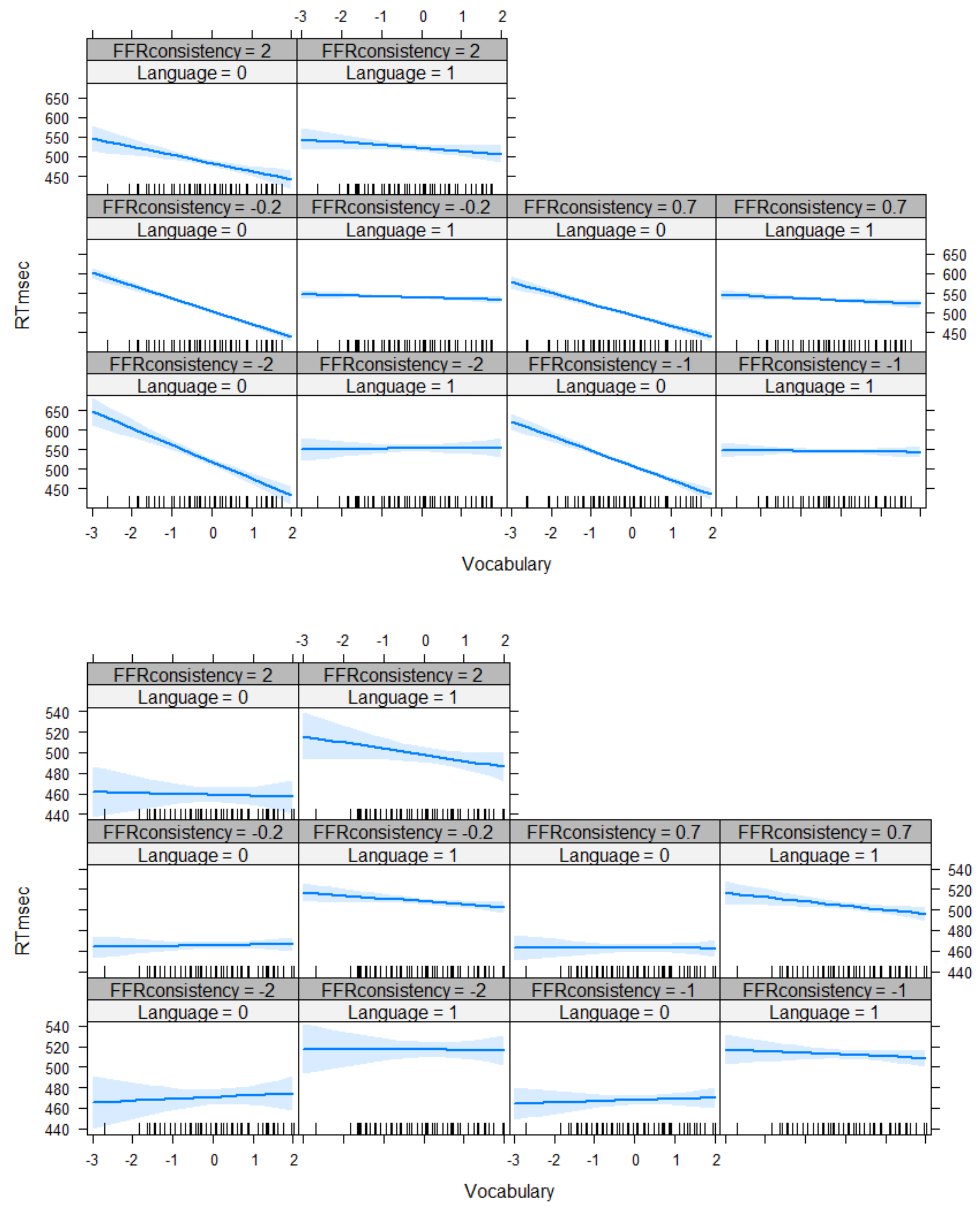


Figure 8.3. Graph of language x vocabulary x FFR consistency interactions for the weak decoder group (upper panel, coefficient = -7.700, p = .008) and the strong decoder group (lower panel, coefficient = -0.713, ns). Language 0 = English; Language 1 = German. Vocabulary knowledge was standardised within language group and higher values indicate higher scores.

Focusing on the interaction of the weak decoder group in the upper panel, it seems that for readers of transparent German, stronger semantic knowledge did not seem to be facilitating when reading words differing in composite spelling-sound rime consistency. For weak decoders of opaque English, however, better vocabulary knowledge seemed to facilitate reading aloud for words of all consistencies. Only when these weak decoders also had low vocabulary knowledge did higher consistency seem to benefit reading aloud. These results indicate that - if available - the English weak decoder group had a tendency to use semantics for reading aloud.

As mentioned previously, it has been suggested that readers with less efficient decoding skills may tend to employ semantics for reading aloud (Strain & Herdman, 1999; Woollams et al., 2016), and the division of labour hypothesis of the PDP models supports this possibility (Plaut et al., 1996). The present results could be seen to support this assumption, as for English weak decoders relatively better vocabulary knowledge resulted in faster naming times. However, for those who had both weaker decoding skills and less vocabulary knowledge, there seemed to be an effect of FF rime consistency, which points to the use of the O-P pathway. Thus, in absence of an efficient semantic network, the less efficient O-P pathway becomes once again the default pathway.

This observation seems only to pertain to the English, but not to the German language group. However, at this point it is important to acknowledge that the analysis of weak decoders also showed an AoA effect which was not modulated by language. If an AoA effect reflects semantic involvement (Brysbaert et al., 2000; van Loon - Vervoorn, 1989), then it seems that reading in both languages requires semantic information for processing, but – as our data showed - only English weak decoders measurably benefitted from better semantic knowledge. In terms of the division of labour hypothesis, this means that deficient decoding results in greater use of the semantic pathway in both languages, but only readers of opaque English benefitted when semantic knowledge was more developed or accessible.

The second important interaction, which will be addressed is that the weak decoder group showed a language x vocabulary x length interaction (see upper panel of Figure 8.4). In English, better vocabulary knowledge resulted in faster RTs when words were shorter, but this advantage diminished with increasing word length. In German, vocabulary knowledge did not seem to interact with length. The lower panel of Figure 8.4 also shows the same interaction for the strong decoder group for reference purposes. In the strong decoder group, this interaction was not significant.

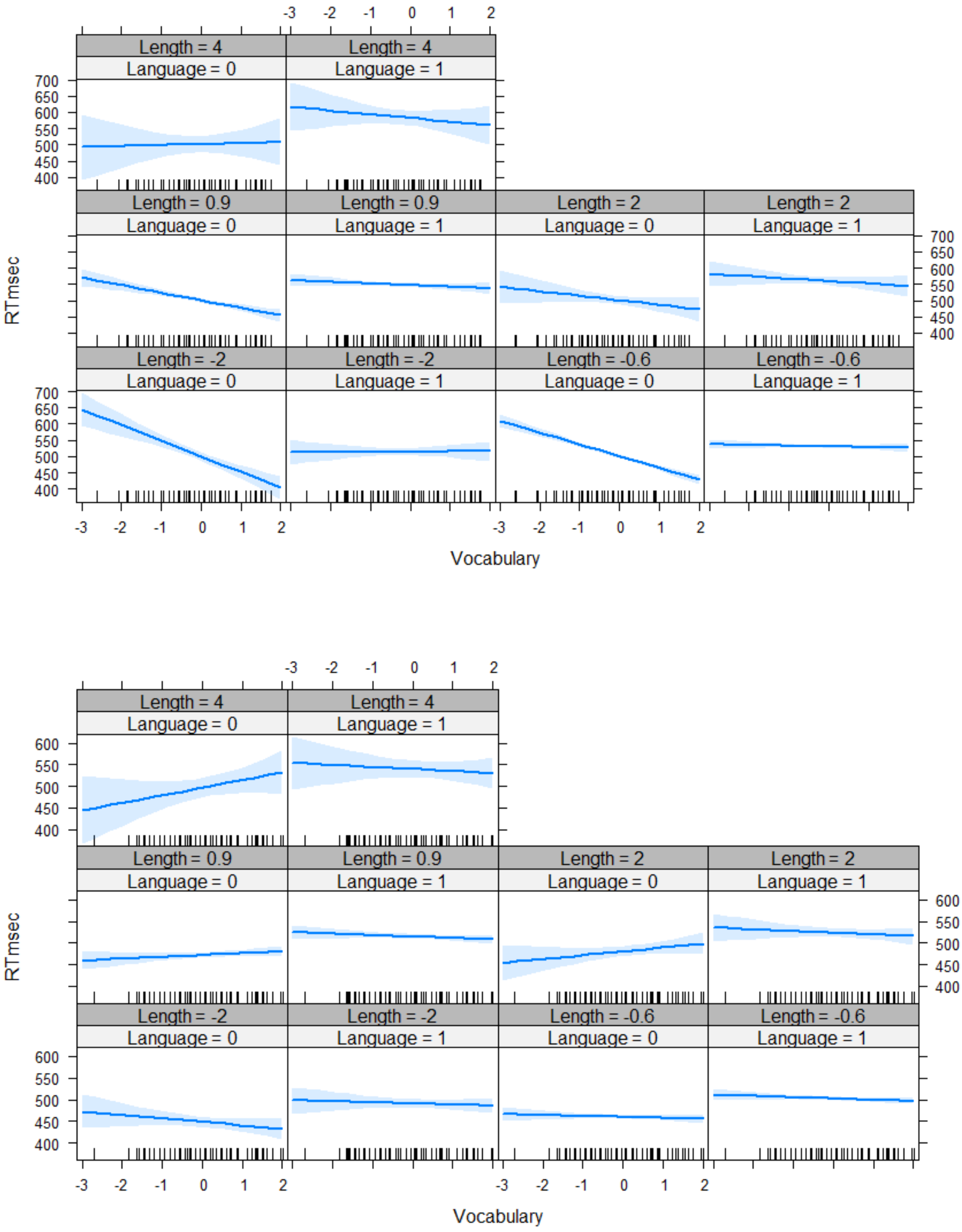


Figure 8.4. Graph of language x vocabulary x letter length interactions for the weak decoder group (upper panel, coefficient = -10.555, p = .012) and the strong decoder group (lower panel, coefficient = -4.663, ns). Language 0 = English; Language 1 = German. Vocabulary knowledge was standardised within language group and higher values indicate higher scores.

For the transparent language German, vocabulary knowledge did not seem to modulate the psycholinguistic effect of length even in the weak decoder group. This suggests that – even if semantics may be used for reading as evidenced by the AoA effect – the transparency of the language is still a bigger advantage for reading aloud than increased semantic knowledge, so that any benefit derived from better semantic knowledge does not yield a measurable impact at any word length. Thus, for German weak decoders, the direct O-P pathway reading seems still more effective than the route via semantics, even when it does not work at its greatest efficiency level.

In contrast, in English weak decoders, relatively greater vocabulary knowledge seemed to reduce naming latencies and interacted with length. For participants with relatively better vocabulary knowledge, RTs increased with increasing length. For participants with weak decoding skills and less vocabulary knowledge, however, there seemed to be some tendency for shorter RTs for longer words. The latter observation initially seems counter-intuitive, but it may be useful to recall that Yap & Balota (2009) found a quadratic effect of length with longer latencies for very short words. In accordance with New et al. (2006), they suggested that the shortest latencies were observed for the most common word length. The present data suggest that the modulation by vocabulary knowledge and nonword reading ability may also contribute to the quadratic length effect in reading aloud of opaque languages such as English.

Analysis I (see section 7.4) reported mixed-model effects results from the complete data set, i.e. participants had not been partitioned into better or weaker performers on ID tasks. It was observed that in the case of English readers good vocabulary knowledge resulted in facilitation for shorter, but not for longer words. Strain & Herdman (1999) and others had suggested that participants with weaker phonological skills may resort to increased use of semantics. It was not possible to address this question in analysis I, as the relevant interaction had not been part of the specified model. However, the present analysis outcomes from the weak and the strong decoder groups may give some indication in this regard. As seen above, the language x vocabulary knowledge x length interaction was significant for weak decoders, but not for strong decoders. This implies that for English weak decoders, strong semantic knowledge was helpful for naming. The present results thus suggest that the ability to take recourse to strong semantic knowledge is specifically beneficial to weaker decoders of opaque scripts, but not measurably so for weaker decoders of transparent scripts.

In sum, in weak decoders an AoA effect indicated the use of semantic knowledge in reading aloud in both languages. Yet, person-level semantic knowledge still seemed to be more beneficial for readers of the opaque script, and specifically for those who had lower decoding skills. Furthermore, it may be possible that the quadratic length effect found in previous studies may have been subject to the interplay of decoding ability and use of semantics, as the present study found that combined weaker vocabulary and weaker decoding skills lead to shorter RTs for longer words in English readers. It may be useful to explore these relationships in future studies as an important difference between languages.

8.3.6 Vocabulary knowledge facilitates detection of semantic effects

In the present investigation, strong vocabulary knowledge in participants signified well-developed semantic knowledge, which in terms of the triangle model could be used for reading via semantics. Thus, the high-vocabulary group may be more apt at using semantics for reading effectively when words become more difficult (e.g. inconsistent), or alternatively, it may be easier to detect semantic effects in such a group where semantics are well-developed.

In concert with this assumption, the high-vocabulary group would not to be expected to necessarily have a maximally efficient O-P pathway, and therefore would show a consistency effect. This would distinguish the high-vocabulary group from the strong decoder group, for example, who would be expected to have an extremely proficient O-P pathway. Congruently, the mixed-effects model showed that the high-vocabulary group indeed displayed a consistency effect with shorter RTs for more consistent words. Notably, this effect was not modulated by language.

The division of participants with strong and weaker vocabulary knowledge into separate groups makes it possible to investigate what happens across the spectrum of vocabulary (semantic) knowledge. In the previous analysis on the complete data set (see section 7.4), an interaction between language, vocabulary knowledge and FFR consistency was found, which could not be explained in the light of previous literature. In English, vocabulary knowledge facilitated naming for words of all consistencies, but more so for less consistent words. In comparison, for German it seemed that vocabulary had a less facilitating effect on naming, but – surprisingly - there was a tendency for more consistent words to be named faster when vocabulary knowledge was greater.

Figure 8.5 shows the graphs for the language x vocabulary knowledge x composite FFR consistency interaction for both the low-vocabulary group (upper panel) and the high-vocabulary group (lower panel). Only for the low-vocabulary group was this interaction significant: for all English low-vocabulary participants, greater vocabulary knowledge led to shorter naming times, irrespective of consistency. For German low-vocabulary participants, greater vocabulary knowledge was only facilitating when words were most consistent. Thus, it seems that the locus of the language interaction lies with those participants who have the smallest vocabulary knowledge. Specifically, the participants with the lowest vocabulary knowledge in the German language group were not facilitated by their vocabulary knowledge when words were very inconsistent, whereas this was still the case with their English counterparts. It seems then, that readers of opaque orthographies are prone to take recourse to semantic information for reading when words become more inconsistent even when they have relatively less semantic knowledge. For very low-vocabulary readers of transparent orthographies, this seems not to be the case. It is possible that these readers have developed lesser ability to employ semantics, given the transparency of their script. However, clearly this interaction cannot be explained satisfactorily within the current study, but certainly merits further investigation in future research.

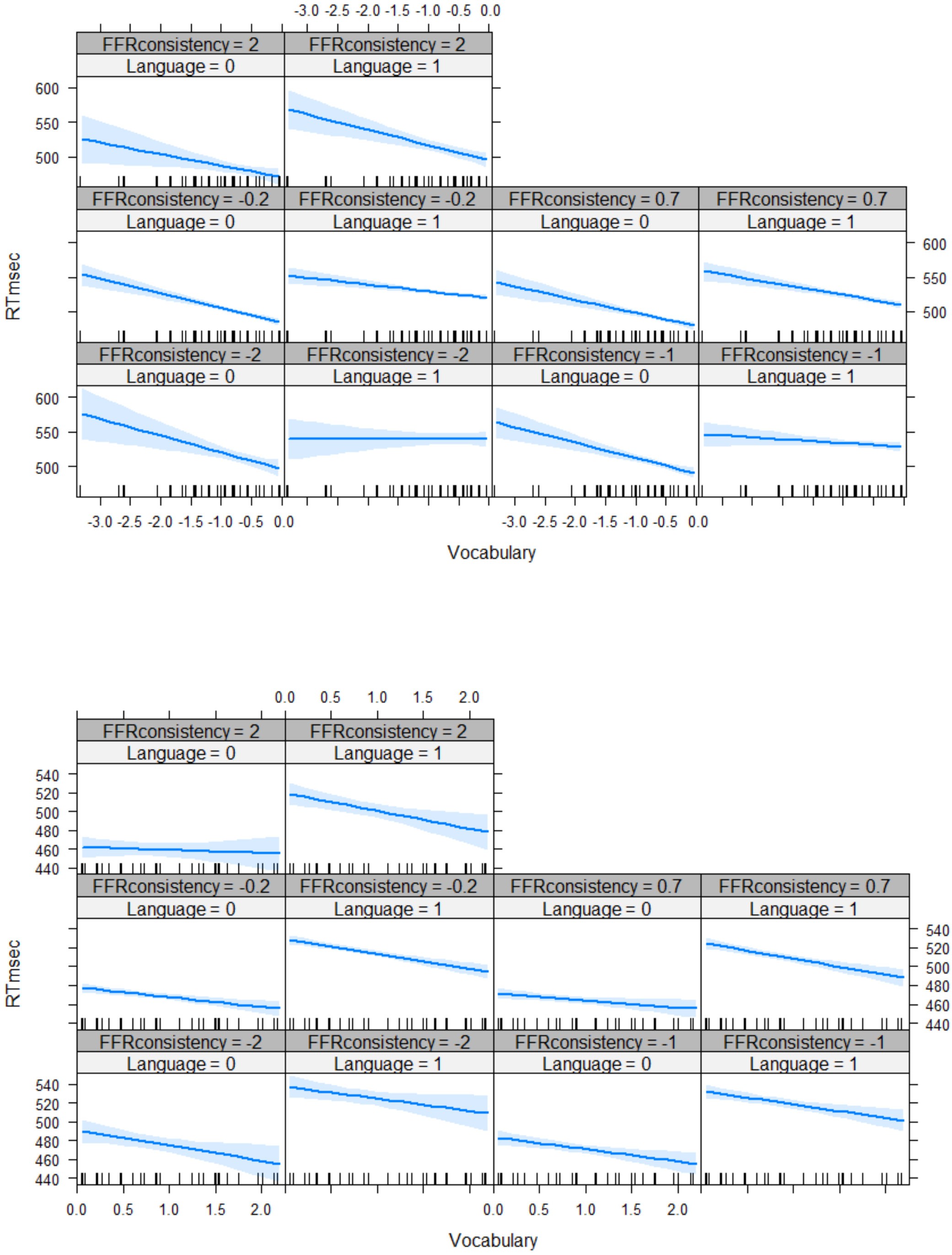


Figure 8.5. Graph of language x vocabulary knowledge x composite FFR consistency interactions for the low-vocabulary group (upper panel, coefficient = -7.479, p = .030) and the high-vocabulary group (lower panel, coefficient = -4.779, ns). Language 0 = English; Language 1 = German. Vocabulary knowledge was standardised within language group and higher values indicate greater knowledge.

As mentioned above, in the present investigation it was assumed that the high-vocabulary group comprised participants with a more accessible semantic system, which may lead to more use of semantics or to easier detection of semantic effects. Indeed, the analysis of the high-vocabulary group found an AoA effect which was modulated by decoding skill (see Figure 9.6). If we accept that AoA effects have a partial semantic component, then the occurrence of an AoA effect with the high-vocabulary reader group points at semantic involvement in naming for this group. The AoA x decoding skill interaction showed that participants with less efficient decoding skill were more disadvantaged at naming late than early acquired words. Importantly, this interaction was not modulated by language. This interaction is interesting, because it shows that in a sample of readers with strong semantic knowledge, the AoA effect was stronger in those participants with weaker decoding skill in both languages. Recall the findings from the previous section (see section 8.3.5), which showed that weak decoders in both languages showed an AoA effect. Together, these results suggest that semantics were employed in weak decoders of both language reader groups.

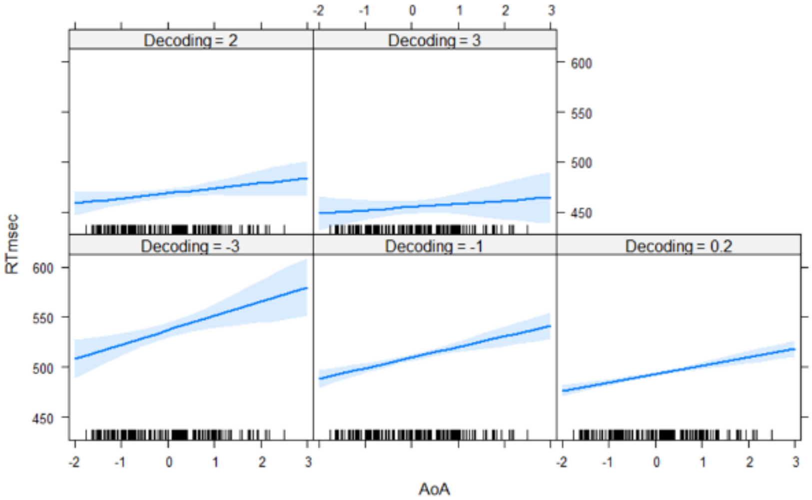


Figure 8.6. Graph of AoA x decoding skill interaction for the high-vocabulary group (coefficient = -3.684, p < .05). Decoding skill was standardised within language group and higher values indicate better skill. AoA measures were standardised with higher number indicating later acquired words.

In sum, in the present study it was assumed that vocabulary knowledge was a valid indicator of semantic knowledge. The high-vocabulary group showed a consistency effect independent of language, indicating that this group included participants with different O-P pathway efficiency levels. However, for the low-vocabulary group, the consistency effect also interacted with language. German participants with the weakest vocabulary knowledge had longer RTs for less consistent words, evidencing that this group did not make efficient use of their semantic knowledge for reading. This effect was not observed with their English counterparts. Thus, whilst even relatively reduced semantic knowledge lead to benefits when reading inconsistent items in an opaque language, this was not the case for the readers of the transparent script. However, the data also showed that semantics were employed by both language groups (AoA effect), and that the increased usage of semantics for participants with weaker decoding skills was found in both languages.

8.3.7 Spelling ability differences lead to different effects patterns in reading aloud

In the present study, it was assumed that participants also differed in terms of the quality of their orthographic representations. Within the framework of the lexical quality hypothesis (LQH; Perfetti & Hart, 2002; Perfetti, 2007), spelling has been considered an appropriate index for precise orthographic knowledge. In the following, it will become evident that stronger and weaker spellers showed quite different effect patterns, including interactions with psycholinguistic variables which index different reading units. Hence, the quality of orthographic knowledge can be considered an important individual difference.

8.3.7.1 Strong spellers name faster than weak spellers in both languages

As shown in Table 8.2, mean RTs for weak spellers (E: 487.35 msec, G: 536.66 msec) were longer than those for stronger spellers (E: 472.24; G: 511.84 msec) in both languages. Mann-Whitney U tests confirmed that strong spellers named cognates significantly faster than weak spellers in English (W = 7,576,700, $p < .001$) and German (W = 7,356,000, $p < .001$). This finding is in line with previous research which found that better spellers are also faster at reading aloud (Martin-Chang et al., 2014), and the LQH which supposes that less precise orthographic representations will hamper word recognition. The present study extends these findings to a transparent language (German).

8.3.7.2 Strong spellers show orthographic neighbourhood effect independent of language

Previous research had found that in reading aloud, words with larger orthographic neighbourhoods would be named faster than words with sparse neighbourhoods. Thus, generally, larger orthographic neighbourhoods were considered facilitatory. The current study included orthographic neighbourhood as one of the predictor variables. In the present study, strong spellers showed a facilitatory main effect of orthographic neighbourhood (old20, coefficient = 13.632, $p = .007$). This effect was not modulated by language.

For weaker spellers, the old20 main effect did not reach significance. Why was naming facilitated by old20 for strong but not weak spellers? It may be that the weaker spellers did not show an old20 effect because they were not able to exploit word similarity in their processing. If poor spelling is a measure of unspecified orthographic knowledge, then weak spellers may not be able to use the orthographic information (old20) efficiently enough to identify items.

In fact, the assumption that stronger spellers are more likely to exploit orthographic similarity is reminiscent of previous research by Andrews and colleagues, who found that weaker and stronger spellers were impacted differently by orthographic neighbourhood in masked form priming lexical decision. Specifically, Andrews & Hersch (2010) found that only poor spellers showed facilitatory priming of high and low *N* words. Better spellers also showed facilitatory priming for low *N* target words, but for high *N* target words the priming effect was reduced. Within the framework of the LQH, the authors suggested that better spellers had more precise orthographic representations which would be activated faster than those of poor spellers. In this sense, greater spelling ability would increase lexical competition at the letter perception level as suggested by the IA model (McClelland & Rumelhart, 1981). The faster activation may then cause lexical inhibition of orthographic neighbours, including the target word, which would lead to a reduced priming effect. Andrews & Hersch (2010) therefore suggested that whilst all readers benefitted from word priming due to sublexical overlaps between prime and target, precise orthographic representations (as indexed by good spelling ability) would reduce the effect through faster lexical competition. Andrews & Lo (2012) reported convergent evidence from a masked priming lexical decision task with transposed-letter (TL) primes (e.g. sung SNUG) and neighbour nonword primes (snag SNUG). The authors hypothesised that individual differences in lexical quality would result in increased lexical competition for those participants with more precise lexical representations. TL primes were used to measure the effect of lexical competition. Neighbour nonword primes were employed to gauge sublexical facilitation. As expected, poor spellers showed facilitation from both primes, whilst good spellers showed more inhibition, in particular for the TL primes.

However, it needs to be stressed that the masked priming task captures effects which occur very early in the reading process, when letters are activated (e.g., Evett & Humphreys, 1981). For the naming task, which is used in the current study, on the other hand, activation has progressed so far that articulation of the correct pronunciation of the letter string has commenced. Any effects of lexical competition would then unlikely to still be measurable.

Yet, the similarity of the present to Andrews et al.'s findings still suggests that the person-level quality of the word representations impacts on the reading process, possibly at different stages throughout the process.

As mentioned previously (see section 4.4.4), it has been suggested that orthographic knowledge is acquired through statistical learning (Deacon et al., 2008; Treiman & Kessler, 2006). This view can be accommodated by the connectionist approach (Plaut et al., 1996), which is based on the principles of statistical learning. When considering individual differences in spelling (as a marker for orthographic knowledge) and their impact on naming, an effect of orthographic similarity may indicate that the reading system is able to exploit the statistical regularities of the script. The present results support this assumption in so far as the facilitating effect of greater orthographic similarity reduced naming latencies in good spellers, but not weaker spellers.

8.3.7.3 Decoding skill also affects naming in strong spellers

It is important to note that the old20 effect in strong spellers was also modulated by decoding skill in such a way that the old20 effect reduced with increasing decoding skill (coefficient = -5.519, *p* = .038, see Figure 8.7). This interaction was not significant with the weaker spellers group, but the graph of the latter shown in the upper panel of Figure 8.7 hints at some facilitation for better decoders within the weaker speller group when orthographic word neighbourhood was denser. This is to be expected as participants were split by the mean, and therefore border cases would demonstrate some behaviour similar to the strong speller group.

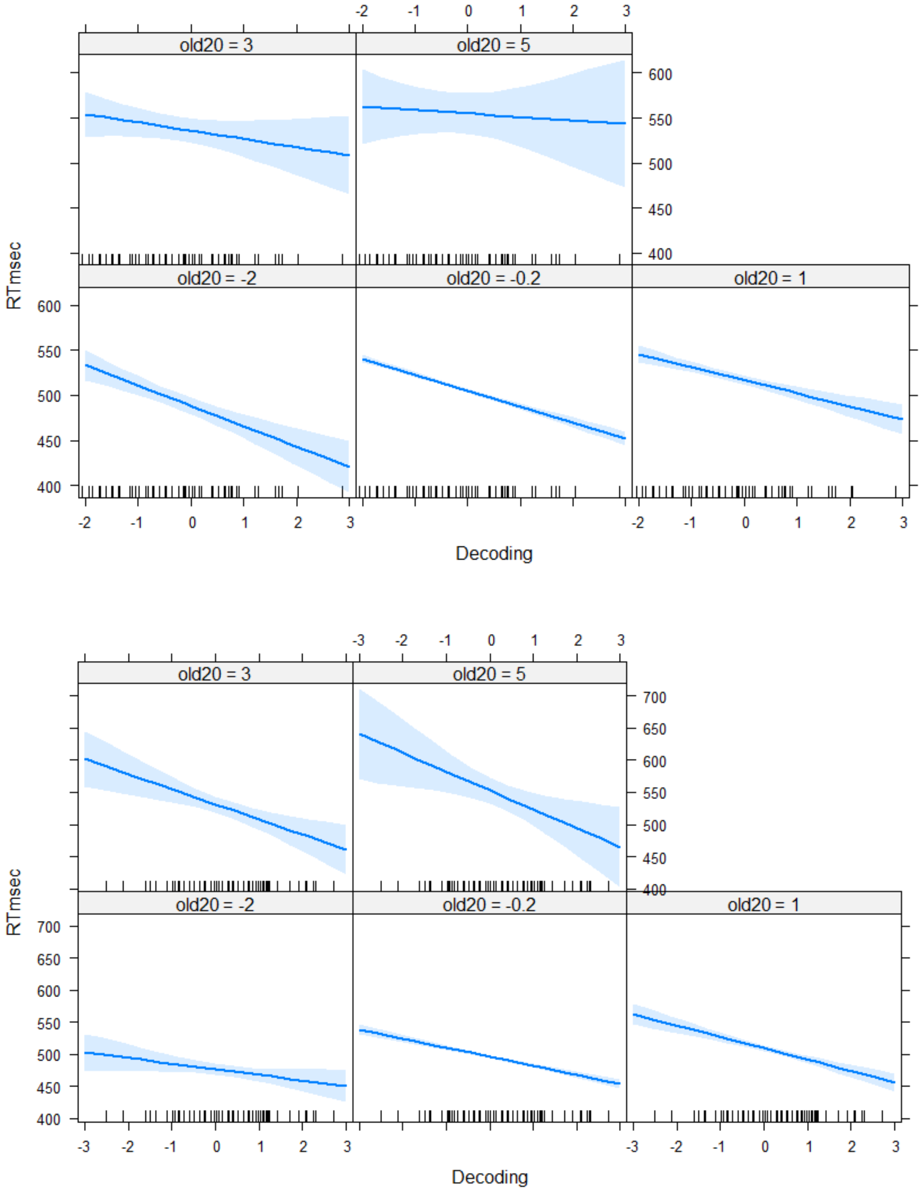


Figure 8.7. Graph of decoding x old20 interactions for the weak speller group (upper panel, coefficient = 1.735, ns) and the strong speller group (lower panel, coefficient = -5.519, p = .038). Decoding scores were standardised within language group and higher values indicate greater knowledge. Old20 scores were standardised within language group and higher values indicate sparser neighbourhood.

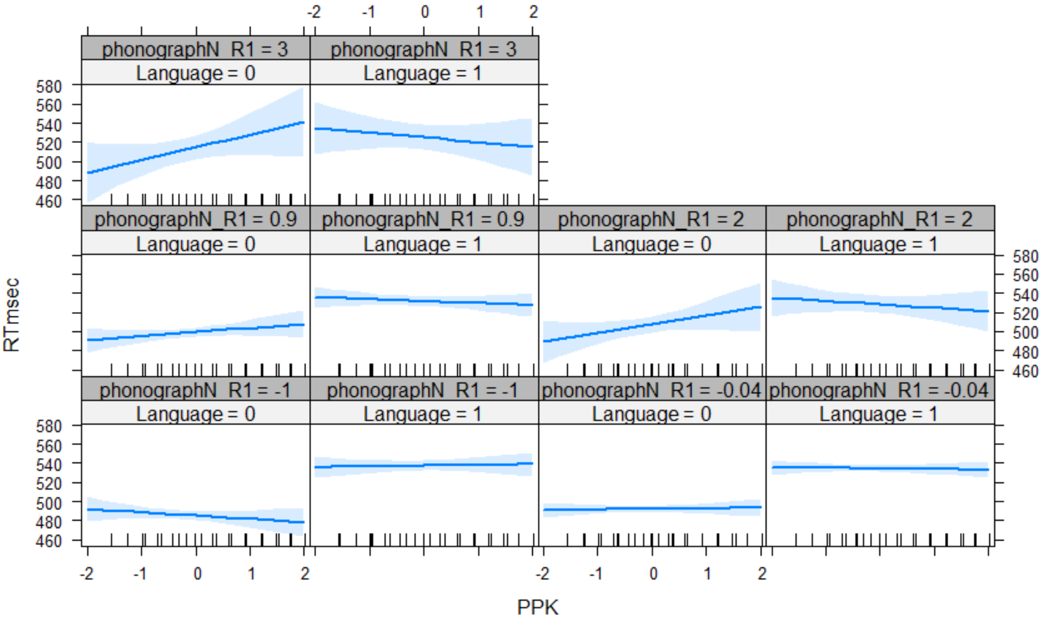
8.3.7.4 Weaker spellers show PNR1 effects modulated by language and IDs

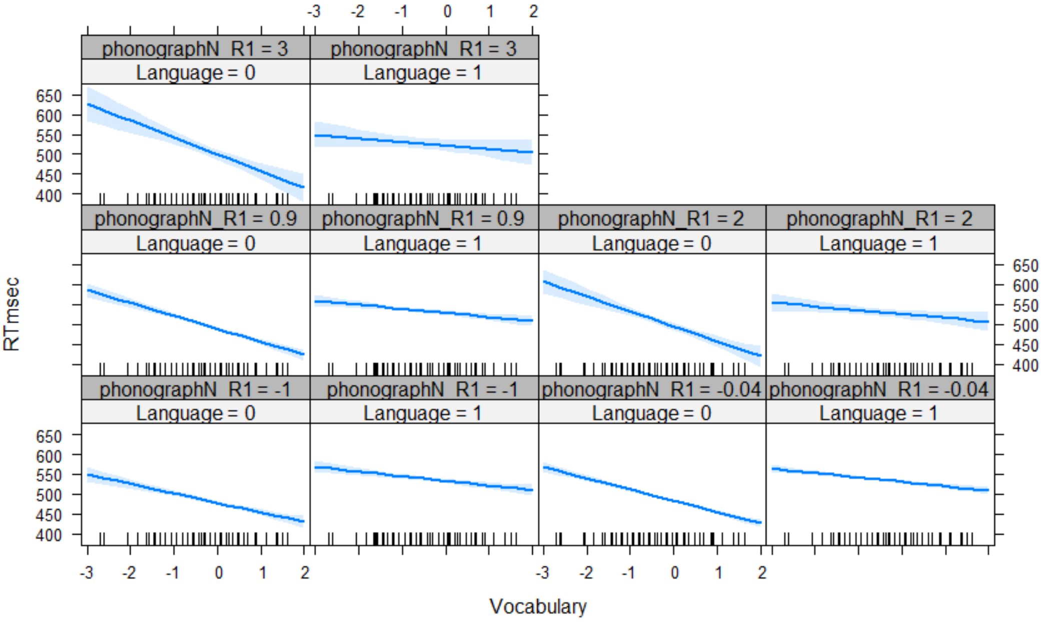
As discussed above, the present results suggest that weaker spellers may not be able to use the orthographic information (old20) efficiently enough to identify items. Instead, the data seems to suggest that weak spellers made more use of information which links orthographic and phonological information, namely spelling – sound consistency and phonographic rime 1 neighbourhoods (PNR1). Thus, the weaker speller group showed a spelling-sound consistency main effect, which was independent of language (coefficient = -10.523, p = <.001). This also suggests that in weaker spellers of both languages the O-P pathway is less efficient.

The weaker spellers group also displayed several 3-way interactions between language, phonographic rime 1 neighbours (PNR1) and each ID task (see Figure 8.8). PNR1 are assumed to index larger grain sizes for reading. Thus, for weaker spellers reading opaque English, it seemed that larger grain sizes were more usefully exploited when participants had more semantic knowledge or better decoding skills, or if they were relatively better spellers within the weaker spelling group. Moreover, relatively better print knowledge was found to be inhibitory for naming speed. Weaker spellers of the more transparent German language were less affected or were not affected.

It is an interesting finding that IDs modulate the PNR1 effect in (mostly English) weaker spellers. This opens up the possibility that – depending on their individual strengths or weaknesses – readers may form their orthographic knowledge differently, and specifically so in opaque scripts. Take the fact that within the weak speller group relatively more print exposure slowed down naming speed (see first panel of Figure 8.8). This may reflect the fact that orthographic knowledge is not complete at a word-level, but that many words have been encountered, which leads to increased competition and slows down RTs. Equally it is plausible that better decoders within this group are at an advantage to identify different grain sizes, as decoding skill is partly reflected in better ability to segment words for reading (see discussion on individual differences and the length effect in section 8.3.3). Relatively better spellers within this group may be able to detect larger unit sizes, because of their more complete orthographic representations, whilst the weakest spellers in the English-speaking weaker speller group found these larger grain sizes generally not facilitating. In fact, the weaker the spelling skills, the more inhibitory PNR1 seemed to be (see last panel in Figure 8.8). The very reduced naming latencies for words with very few PNR1 points to the possibility, that these unique words were instead recognised for their rarity. Finally, and speculatively, weaker spellers with more semantic knowledge might be able to focus on morphological units. As reviewed in section 4.4.2, Andrews & Lo (2013) found that university students with a ‘semantic profile’ showed stronger priming for morphologically transparent primes, and Kuperman & Van Dyke (2011) showed that base morphemes facilitated whole word recognition in poor readers in an eye-tracking study. Interestingly, Pagliuca & Monaghan (2010) demonstrated in a PDP model simulation with Italian multisyllabic words that morphological effects appeared even in the absence of an implemented semantic route. The reading system thus may to be able to identify statistical regularities present in morphological units and employ these units independently of their meaning reference. It is of course beyond the scope of the present study to explore these ideas further with the present data. However, the bottom line is that larger phonographic units used for reading may be extracted by a person’s reading system depending on their individual strengths or weaknesses.

It needs to be stressed that the interpretation of these results is to be treated with caution, as the PNR1 measure effectively combines two measures in one, one for monosyllables and one for bisyllabic words (see discussion in 7.2.2.1.7.3).


phonographN_R1 = 3
Language = 0
Language = 1
phonographN_R1 = 0.9
phonographN_R1 = 2
phonographN_R1 = -1
phonographN_R1 = -0.04
RTmsec
580
560
540
520
500
480
460
-2
-1
0
1
2
PPK


phonographN_R1 = 3
Language = 0
Language = 1
phonographN_R1 = 0.9
phonographN_R1 = 2
phonographN_R1 = -1
phonographN_R1 = -0.04
RTmsec
650
600
550
500
450
400
-3
-2
-1
0
1
2
Vocabulary

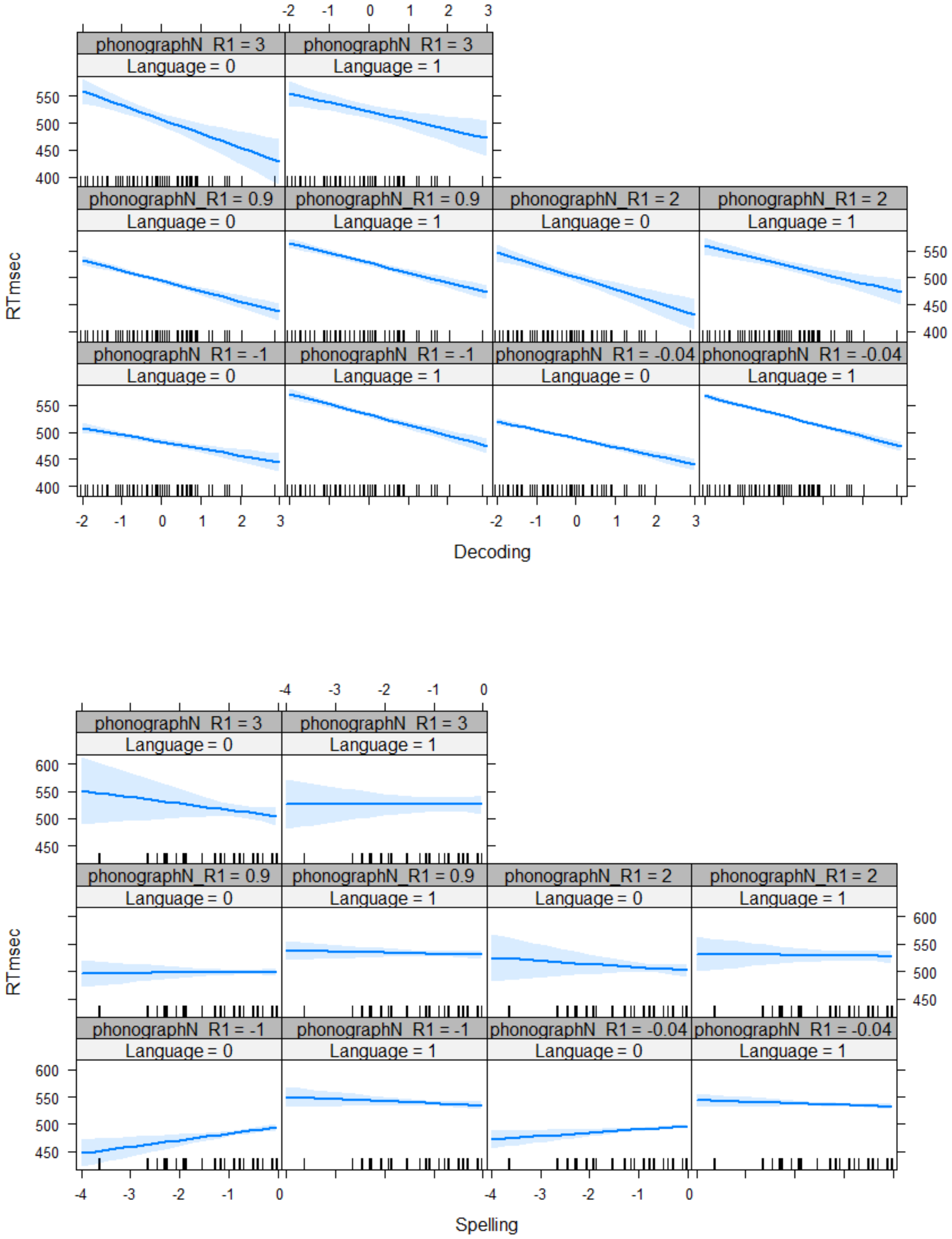


Figure 8.8. Four graphs showing the 3-way interactions involving language and PNR1 in the weak speller group. All were significant (p < .05). Language 0 = English; Language 1 = German. Spelling scores were standardised within language group and higher values indicate greater PPK scores, vocabulary knowledge, decoding skills and spelling ability, respectively.

In sum, across both languages, stronger spellers seem to have been able to efficiently exploit orthographic similarity (old20) in naming, possibly evidencing their greater ability to extract statistical regularities in the orthography. This effect was reduced when decoding ability was stronger. Decoding ability thus seemed to override the effect of orthographic similarity. Weaker spellers did not show an effect of orthographic similarity which suggests that they were not able to make as effective use of orthographic information. Weaker spellers also showed a consistency effect pointing to a less efficient O-P pathway. However, it seemed that English (but not, or to a lesser degree German) weaker spellers benefitted from the occurrence of larger phonographic reading sized units (PNR1). As this effect was modulated by IDs, it seemed plausible that the identification of larger reading units may be dependent on individuals' different strength and weaknesses.

The above findings revise the results reported in the previous analysis (Analysis I, see section 7.4). There, effects of spelling had been limited to faster naming for longer words in both languages, although this result had not quite been statistically significant. The present findings imply that spelling ability is related to the ability to segment words into various constituents, and that for weaker spellers the ability to detect larger grain sizes may be influenced by other IDs. These results seem to be true for English weaker spellers, but weaker or absent for German weaker spellers. At this point it needs to be noted that the spelling tasks used in the present investigations were the least comparable of all tasks. Whilst the task used for the English-speaking participants was focused on lexical precision by testing the accuracy of the letter string, the task for the German speakers also probed other rules of orthography, such as capitalisation and phrase segmentation. It would be useful to investigate if the same findings were obtained under more similar task demands.

8.3.8 Effects of print exposure across languages

The large body of research on the impact of print exposure on reading aloud (e.g., Mol & Bus, 2011; Stanovich & Cunningham, 1992; Stanovich & West, 1989; Yap et al., 2012) , as well as the underlying fundamental assumption of connectionist network models that frequency of exposure should be one of the main predictors to explaining reading (e.g. Plaut et al., 1996; Seidenberg & McClelland, 1989) has fuelled the expectation that the high-print group should show attenuated psycholinguistic effects, and efficient naming performance – perhaps more so than any other high-score ID group. However, the review of the descriptive statistics already indicated that high-print individuals did not emerge as the fastest readers (see Table 8.2 which shows that strong decoders were fastest on average). In fact, PPK was not a significant main effect in any of the eight models. This might imply that print exposure is not such an influential variable at skilled reader level after all.

When examining the results from the mixed-effects model for high-print individuals, three psycholinguistic main effects emerged. Naming was facilitated by higher word frequency, greater FFR consistency and earlier AoA. As outlined above, however, the expectation had been that psycholinguistic effects should have been reduced. This will be discussed below in sections 8.3.8.1 and 8.3.8.2.

With regard to differences in effects of print exposure across languages, it was found in the previous analysis (Analysis I, see section 7.4) that practice had a similar effect in both languages. Reading practice (as indexed by print exposure) seemed to be universally beneficial to readers of both languages, and showed greater effects when items were longer. Equally, in the present set of results, the language x print exposure x length interactions were not significant, concurring with the view that the effect of print exposure is not significantly different across languages.

However, in the current set of results both the high-and the low-print models showed a significant language x decoding x length interaction. As will be discussed in section

8.3.8.3, this interaction showed that decoding ability is a more influential individual difference than print exposure.

8.3.8.1 High-print group comprises different kinds of readers

The results pattern for the high-print group suggests foremost that this ID group includes a diverse reader group. First of all, being an avid reader does not necessarily seem to imply, that the reading process has also reached the highest level of efficiency (as indexed by shortest naming RTs). The FFR consistency effect hints at some variation in terms of O-P pathway efficiency, and the AoA effect proposes some involvement of semantics. Crucially, these effects seemed to be true for both languages, implying that these variations within readers are true for both languages.

8.3.8.2 Frequency effect apparent due to stimuli frequency

Within the connectionist framework, the word frequency effect occurs as a result of strengthened connection weights between orthography to phonology mappings. P. Monaghan et al. (2017) showed in connectionist model simulations that the frequency effect reduced once mappings had achieved a certain efficiency, resulting in a quadratic word frequency effect: when a certain amount of print exposure had been reached, the word frequency effect in naming reduced again. Moreover, model simulations with larger vocabularies found a longer increase in the word frequency effect compared to models with smaller vocabularies, which meant that the reduction in the word frequency was delayed for models learning to read a larger vocabulary.

Contrary to these findings from previous modelling endeavours, the present data found significant word frequency effects in the high-print group. Given the findings from the literature, a reduced or even absent word frequency effect would have been expected. Additionally, the high-vocabulary group did not show a word frequency effect, even though larger vocabularies should have resulted in a longer lasting word frequency effect. How can the present findings be explained?

It is possible that the apparent contradiction between the literature and the present findings can be resolved when taking into account the frequency of the stimuli. The word frequencies (in Zipf) for the cognates were mostly medium to high frequency (for E: *M*(*SD*) = 4.15(0.69), *Median* = 4.22, min = 2.83, max = 5.77, G: *M*(*SD*) = 3.95(0.64), *Median* = 3.89, min = 2.92, max = 5.59). As a guideline, Zipf values between 0 – 3 could be considered low frequency (Van Heuven et al., 2014). Recently presented data by Brysbaert, Mandera, & Keuleers (2018) suggested that depending on the vocabulary size of the models, the word frequency effect would be apparent for differing word frequencies of the stimuli. The authors suggested, that for participants with very large vocabularies, the frequency effect may only appear for word stimuli in the frequency range of 2 to 4 Zipf. As such, the cognate stimuli may have been too frequent for the high-vocabulary participants to show a frequency effect. Analogously, the high-print group comprised individuals with very varied vocabulary scores (see Table 8.2). Standardised vocabulary scores for high-print individuals varied between -1.85 and 2.17, indicating that some of the participants had large vocabularies, whereas others had vocabularies in the lower range. This will have lowered the Zipf threshold at which the word frequency effect for the present stimuli set would be expected. Brysbaert et al. proposed that medium vocabulary sizes would be expected to show word frequency effects at 3 to 5 Zipf, and very small vocabulary sizes for all Zipf frequencies above 4. As mentioned above, the present stimuli had frequencies ranging between 2.83 and 5.77. Thus, although high-print individuals should not have shown a frequency effect due to increased print exposure, this was possibly off-set by their varied vocabulary knowledge. The smaller vocabularies of some of the participants may have led to the emergence of a frequency effect, as participants with smaller vocabularies may show frequency effects when naming medium to large range word frequency stimuli. The varied vocabulary knowledge in the high-print group may therefore be the cause for the word frequency main effect in the high-print model.

8.3.8.3 Does print exposure affect reading across languages differently?

In Analysis I (see section 7.4) it was found that reading practice influenced reading similarly across languages for words of various lengths. In the separate models for high- and low-print participants, the relevant interaction language x print exposure x length were also not significant. The interactions are visualised in Figure 8.9. These suggest that reading practice has a similar beneficial influence on reading longer words in both languages across the entire spectrum of print exposure.

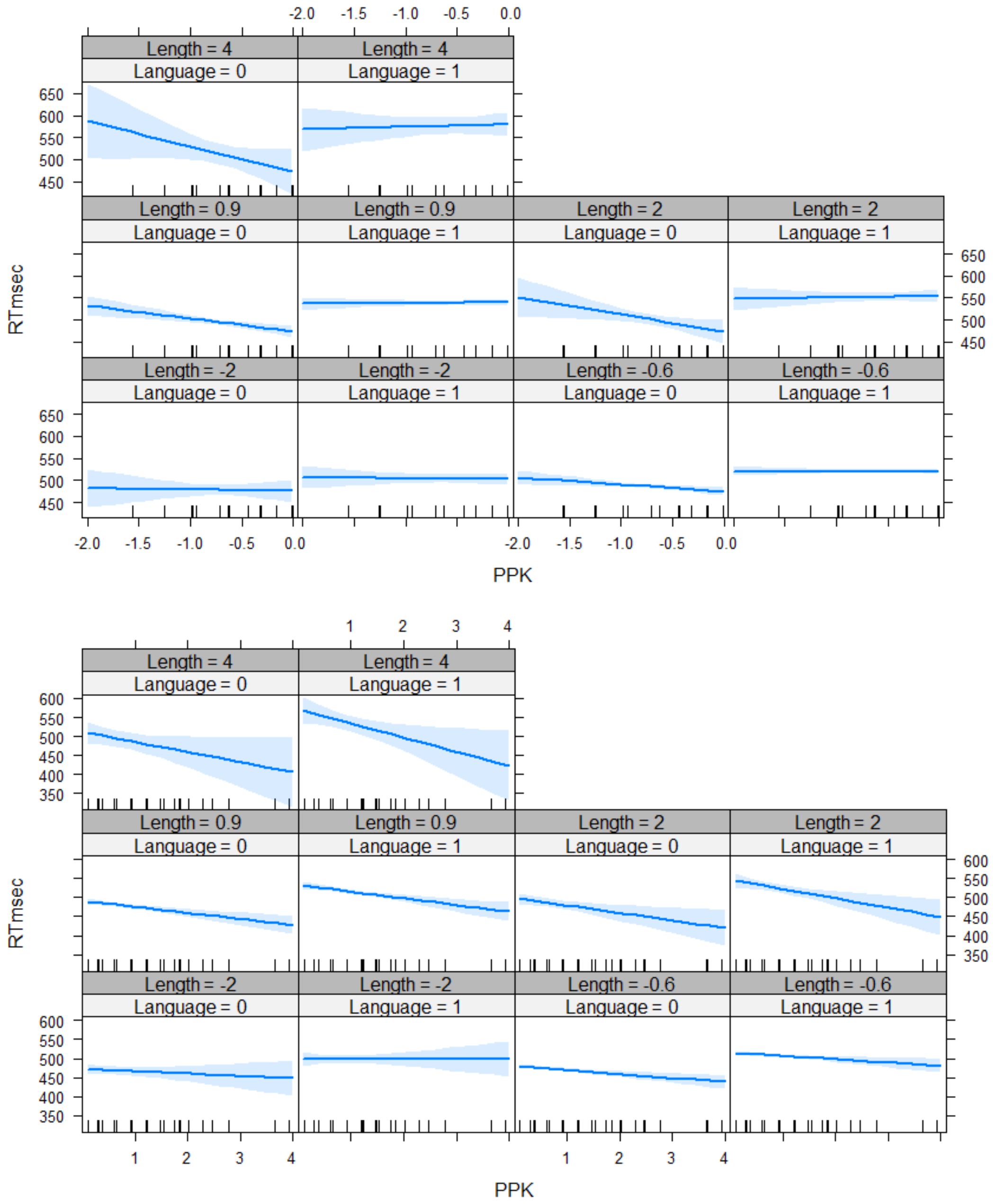


Figure 8.9. Graph of language x print exposure x length interactions for the low-print group (upper panel, coefficient = 10.379, ns) and the high-print group (lower panel, coefficient = -2.794, ns). Language 0 = English; Language 1 = German. Print exposure was standardised within language group and higher values indicate greater print exposure.

However, when also taking into account participants' decoding skill, this skill seems to vary in both print exposure groups differently across languages: a significant language x decoding x length interaction was reported for both the low- and the high-print groups (see Figure 8.10). It can be seen that for German readers, greater decoding skills lead to shorter reading times, and that this effect increased with longer words. The effect seems somewhat stronger in high- than low-print German readers. English readers showed a different pattern: greater decoding skill lead to shorter reading times when words were short. This effect disappeared when words were longer. The effect patterns seemed very similar for both high- and low-print English participants. These results suggest foremost, that decoding skill differences remain a strong influence on naming RTs irrespective of participants' print exposure, and that the relationship between decoding and length differs according to language. Whilst for German participants, decoding skill reduced latencies at all lengths, for English participants, greater decoding skill facilitated reading aloud mostly for shorter words – irrespective of print exposure.

Thus, it seems that print exposure has a facilitating effect in both languages when reading longer words. However, when decoding skill is also being considered, then language differences appear due to decoding skill variation, and not due to print exposure differences.

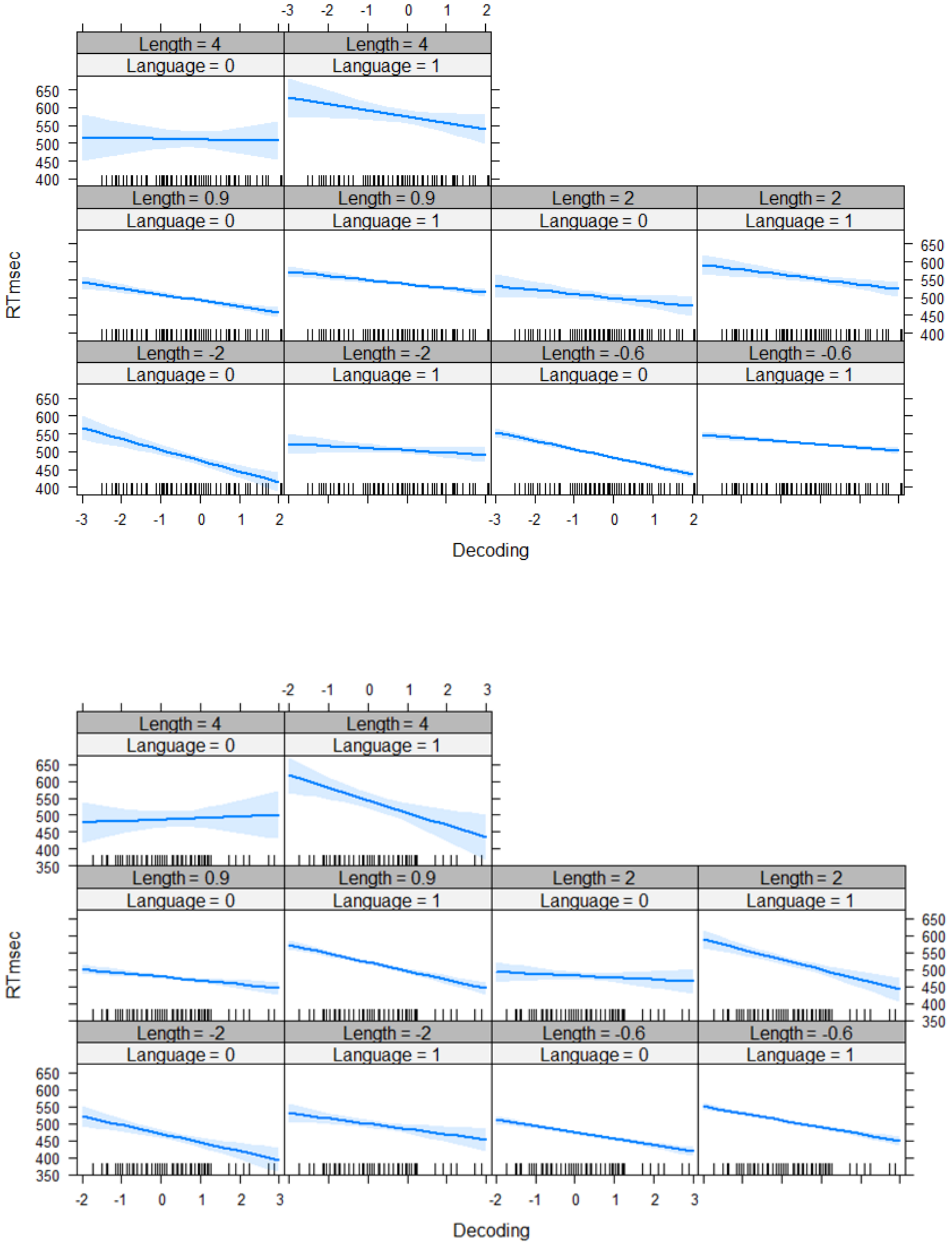


Figure 8.10. Graph of language x decoding x length interactions for the low-print group (upper panel, coefficient = -8.517, p < .01) and the high-print group (lower panel, coefficient = -6.667, p < .05). Language 0 = English; Language 1 = German. Decoding skill was standardised within language group and higher values indicate greater skill.

In sum, reading practice had a beneficial effect for reading longer words in both languages. However, although the reading literature generally expects high-print individuals to be proficient readers and to show reduced psycholinguistic effects, this expectation was only partly met in the present data set. First, the present data suggest that high-print individuals are a heterogeneous reader group as consistency and AoA effects pointed to differential efficiency in the use of O-P mappings. This was true for both languages. It is therefore feasible that high-print readers still vary on other important dimensions, which makes them a heterogeneous reader group. Second, an unexpected word frequency effect was found, also pointing to less efficient reading processes. It is feasible that the reduction of the word frequency effect in the high-print group was offset by the variable vocabulary knowledge of participants. Third, when examining the significant language x decoding x length interactions in both the low-and high-print reader groups, it became evident that decoding skill was a more influential ID than print exposure as language differences became apparent. For German participants, decoding skills facilitated naming at all lengths, whereas for English participants, better decoding skills were helpful for naming mostly shorter words – irrespective of print exposure. This points to decoding skill as the most important individual difference to consider, even ahead of print exposure.

9 General discussion

The present investigation explored whether differences in reading aloud in languages which vary in spelling-sound relationships persist if individual differences (IDs) are taken into account. Specifically, variation in decoding skill, vocabulary knowledge, spelling skill and print exposure were measured for 104 English-speaking and 104 German – speaking participants. They also named the same set of cognate words. A mixed-model analysis revealed that language (English or German) remained a significant predictor, but that naming latencies were also modulated by IDs. A second analysis, in which the same participants were grouped according to their strengths and weaknesses of the four IDs, explored in more detail when and how psycholinguistic effects were modulated by language and/or IDs. To the author's knowledge, the present study is the first to have compared reading in different languages whilst including a large number of item properties as well as person-level characteristics as reading aloud predictors.

In the following, the main research questions addressed in this study will be discussed in reference to current theoretical assumptions and debates.

9.1 Language and individual differences

The first purpose of the study was to find out if language differences remained when individual differences were taken into account. This is what was found, when analysing the complete dataset, as well as when examining the data of participants grouped according to their ID strengths and weaknesses. Thus, language remained an important determinant of reading performance in all data analyses presented as part of this investigation.

However, whilst extant theories on how language differences influence the reading process have argued for differential reading unit sizes to vary with language transparency, the results from the present study suggested that the ability to find the optimal grain size for reading units depends on both language and decoding skill.

The PGST (Ziegler & Goswami, 2005) suggests that opaque languages require more varied and larger grain sizes for reading than transparent languages, which can reliably be read aloud using small unit sizes. Although the PGST primarily applies to developing readers, it accepts that this difference leaves a footprint in adult skilled reading performance. The length effect has often been interpreted as a measure for small-unit processing. Ziegler et al. (2001) reported a length effect in word and nonword naming for German but not English readers, which suggested that readers of transparent languages used smaller reading units than readers of opaque languages. However, other studies did not find a length effect in word oral reading for Italian readers (Spinelli et al. 2005), and length effects have been reported for word naming in English (Balota et al, 2004; Yap & Balota, 2007). Further behavioural findings suggested that individuals within language groups vary in the size of reading units (e.g., Schmalz et al., 2014).

The present study showed that the language x length interaction was modulated by nonword decoding ability and vocabulary knowledge. For the purpose of this discussion, only the interaction involving nonword decoding will be discussed, but for reasons of completeness, please note that vocabulary knowledge facilitated naming in English for short words, and did not have a measurable impact on naming in German. Focusing on nonword decoding ability, higher levels of decoding ability facilitated naming in German at all word sizes, and naming in English for short words. Importantly, no interaction between language, decoding and length was found when readers were grouped into stronger and weaker decoders, suggesting that the language difference was no longer measurable when participants across language groups were similar in their decoding skill.

These results imply that decoding skill influences the ability to optimally adjust grain sizes for reading. First of all, German readers seemed to benefit more from better decoding skills with increasing word length. This is congruent with the PGST that readers of a transparent orthography benefit from this transparency when reading words of all lengths. It adds a further dimension by showing that the better these participants are at decoding, the easier it is to read longer words. In contrast, English readers benefitted from better decoding skills when reading shorter rather than longer words. It is conceivable that for longer words, the difficulty of finding the correct grain size may have slowed down the readers to such an extent that beyond a certain length, decoding ability did not reduce reaction times even for the most skilled readers. The difficulty of finding the correct grain sizes for longer items may be because the word has to be parsed. In an opaque language, this segmentation process may prove more difficult than in a transparent orthography, because – as suggested by the PGST – more possibilities of matching orthography to phonology are available, making it harder to settle for the correct mapping at the correct unit size. Decoding skill, then, also reflects this ability to find the optimal grain size(s), which becomes most crucial for readers of opaque scripts when reading longer words.

This assumption is supported by the fact that the strong and the weak decoder groups did not show a language x decoding x length interaction. When grouped according to this essential ID, the ability of finding the optimal grain size (segmentation) is more similar across languages, and limits the scope for an interaction.

The PGST assumes that the script transparency dictates a tendency of reading unit sizes in readers: smaller sizes for readers of transparent orthographies and more varied sizes of reading units for readers of opaque orthographies. If decoding skill is understood as the ability to find the optimal grain size (as suggested by the present results), then this adds an additional, person-level variation to the PGST, as finding the optimal grain size may be dependent on decoding skill of the reader. The ability to find the optimal grain sizes, then, depends on both language and decoding skill.

It thus seems that studies exploring the language x length interaction may not have always produced similar results, because they did not control for participants' decoding skill. Whilst the language transparency undoubtedly provides conditions which favour smaller or larger reading units (as suggested by the PGST), participants' ability to segment items into optimal reading units (decoding skill) heavily influences whether a language x length interaction can be detected. Previous behavioural reports have pointed to individual differences in the use of different unit sizes (Schmalz et al., 2014). Importantly, proponents of the PDP model (Plaut et al., 1996) have also suggested that readers differ in the size of the 'optimal window of parallel computation', and that lengths effects may stem from this variation. It has been suggested that readers' experience (Yap & Balota, 2009) or language (Plaut et al., 1996) could create the variation. This view of the length effect seems compatible with the current findings in so far as variations in length effects vary with both language and decoding skill. Thus, the present study has found that language itself shapes the reading process, but that psycholinguistic effects such as the length effect seem to be modulated by individual differences (nonword decoding).

9.2 Nonword decoding skills

The ODH (e.g. Frost et al. 1987) proposes that readers of transparent orthographies make greater use of the sublexical route compared to the lexical route. In the context of the ODH, nonword decoding is taken as a proxy for sublexical route contribution, and is compatible with the dual route model. However, behavioural studies have not consistently yielded evidence for categories of lexical and sublexical readers (Baron & Strawson, 1976; Brown et al., 1994). Instead, it has been shown that people vary considerably in their nonword reading ability (e.g., Torgesen et al., 1999) and it has been reliably reported that readers who are slower at nonword naming are also likely to be slower at word naming (Aaron et al., 1999; P. Brown et al., 1994; Davies et al., 2017; Martin-Chang et al., 2014; Stanovich & West, 1989; Torgesen et al., 1999). This last observation has been important for the conception of a reading model with a connectionist network system like the triangle model, which explains difficulties in nonword reading as a deficiency in the phonological pathway. Importantly, the division of labour suggests that such a phonological pathway deficiency may result in an increased use of the semantic pathway. The present study explored whether the quality of the phonological pathway in readers (as indexed by nonword decoding ability) led to a similar or different reading pattern in each language.

Strong decoders emerged in both languages as the fastest readers. This finding fits well with previous research observing that individuals who were faster at nonword reading were also faster at word reading (e.g., Stanovich & West, 1989). The current study extends this finding to German, as strong decoders were found to be the fastest readers irrespective of language.

Strong decoders also showed the most similar result patterns across languages. Within the framework of the triangle model and the assumption of a division of labour between the phonological and semantic pathways, the fact that in the present study strong decoders were the fastest reader group, and showed the least psycholinguistic effects and interactions, points to the phonological pathway as the royal road for reading aloud. Whilst Plaut et al. (1996) foresaw that the semantic route would increasingly contribute more, whilst the phonological route would specialise on processing very consistent items, Harm & Seidenberg (2004) suggested that the semantic route would only be used in difficult cases, whereas the shortest and most direct route between orthography and phonology would serve as the standard route. The present results are in line with the assumption of Harm & Seidenberg (2004) and recognises the O-P pathway in the role of the prominent or default route in reading aloud. Importantly, the present study shows that this seems to be the case for both languages, which makes a dual route approach favouring the predominant use of one pathway for one and another pathway for another language (cf. Perry & Ziegler, 2002) less likely. Instead, the results support the assumption of a universal reading system for oral reading with a default O-P route, which can be complemented with semantic information, when the language environment, or the personal reader characteristics so require.

Thus, being able to decode unknown items using spelling-sound knowledge creates a lasting advantage in reading aloud even at the skilled reader level. Although this study did not investigate reading acquisition as such, the finding of the continued relevance of nonword decoding skill in skilled readers may be an interesting result for those engaged in research on reading development. Explicit teaching of the spelling-sound relationships through phonics has been found to be very successful in reading acquisition of developing readers (e.g., Ehri et al., 2001, but see, for example, Thompson et al., 2008). It seems very likely that the explicit teaching of phonics will provide a sound basis for decoding skills and therefore make an impact on lifelong reading.

In sum, the present study found that nonword decoding skill seems the most important individual difference in both languages, suggesting that this seems to be a core reading route in alphabetic languages of differing spelling-sound consistency.

9.3 Semantics

Semantic effects in reading aloud have been interpreted differently in the two main reading models. Within the framework of the DRC, the ODH suggests that readers of less transparent orthographies will make greater use of the lexical route. Greater semantic effects in reading opaque scripts, then, have been interpreted as lexical reading in contrast to sublexical reading. The alternative reading model, the triangle model, explains semantic effects as a shift in contributions from the phonological and the semantic reading routes (division of labour hypothesis), and that the contribution of routes may vary between individuals. Greater semantic effects may then be interpreted as a greater contribution of the semantic pathway due to insufficient functioning of the phonological pathway (Strain & Herdman, 1999; Woollams et al, 2016). Moreover, there is some evidence that individual variation in participants' semantic knowledge may influence their reading process, as greater semantic knowledge of individuals may lead to more efficient use of semantics in reading (cf. Andrews & Lo, 2013).

In the present study, it was possible to explore if readers of two different orthographies showed different reading patterns in relation to the above reported effects. Age-of-acquisition (AoA) was used to index item-level semantic effects. Vocabulary knowledge was a proxy for person-level semantic effects, and nonword-decoding skill was an indicator for the efficiency of the phonological route. To the author's knowledge, this is the first study to distinguish between semantic effects at the item- and at the person-level.

The results showed that semantic effects varied at the item - and at the person-level, and that there were differences between the two languages. The AoA effect patterns suggested that, for weaker decoders, semantics at the item-level were activated in both languages. The vocabulary effects showed that for weak decoders relatively better vocabulary knowledge (i.e. stronger semantic information at the person-level) resulted in faster naming times, but only when reading opaque English, not when reading more transparent German.

Congruently, when participants were grouped according to their vocabulary knowledge, the high-vocabulary group showed an AoA x decoding skill interaction, which was not modulated by language. The item-level AoA effect was stronger in those strong-vocabulary participants with weaker decoding skill in both languages, evidencing that item-level (AoA) and person-level (vocabulary) semantic effects were separate effects which interacted with each other. In the low-vocabulary group neither the AoA main effect nor the AoA x decoding interaction were significant.

Thus, the study showed that semantics are a reading component in both languages, and that it is useful for weaker decoders in both languages. However, in very weak decoders, relative better semantic knowledge seemed to be particularly advantageous for readers of opaque but not transparent scripts. Possibly, this is because in transparent scripts, for weak decoders, the O-P pathway is more efficient than the semantic pathway. Supporting this idea, it was found that for the low-vocabulary group, the consistency effect also interacted with language. Specifically, German low-vocabulary participants did not benefit from relatively better vocabulary knowledge when reading inconsistent words, whereas their English counterparts did. It seems plausible that the German low-vocabulary group did not make efficient use of their semantic knowledge for reading inconsistent words. Thus, whilst even relatively reduced semantic knowledge facilitated reading inconsistent items in English, this was not the case for German. It seems that for readers who are used to reading transparent scripts, the reading system does not so readily resort to the semantic pathway, compared to readers used to reading opaque scripts. Therefore, whilst language consistency clearly plays a role in the use of semantics, there are person-level considerations which need to be taken into account to establish a more accurate picture of how readers use semantics in different language contexts.

The results support increased use of semantics for readers of opaque scripts, as suggested by the ODH (Frost et al., 1987; Katz & Feldman, 1983), and thereby reinforces the assumption that the language inconsistency increases the use of semantic units for reading aloud. Yet, the ODH was designed to explain differences in reading between languages, and did not consider individual differences. However, the findings also showed that the use of semantics varied with phonological decoding skill and vocabulary knowledge, and that this varied between languages. The ODH as such therefore cannot accommodate these findings.

In contrast, it seems likely that a PDP model would be able to simulate the pattern of effects. Contributions from the semantic pathway would be evident in both languages (AoA effects), but are stronger in readers with less decoding skill (shift from phonological to semantic pathway). Language differences become evident in readers where semantics are less accessible or developed (less developed semantic pathway makes differences in phonological pathway more pronounced, and readers of opaque scripts would show greater slow-down). Whilst readers of opaque scripts are able to benefit even from relatively low levels of semantic knowledge to read inconsistent items (due to language transparency, more contribution from semantic pathway), this is not true for readers of transparent scripts. Presumably, for very inconsistent words, the predominant phonological pathway use of readers of transparent scripts cannot be compensated sufficiently by the semantic pathway when the semantic knowledge is not as accessible or developed.

9.4 Orthographic representations

The present study also investigated the impact of individual differences in the quality of orthographic representations on reading aloud in both languages. The lexical quality hypothesis (LQH; Perfetti & Hart, 2002; Perfetti, 2007) suggests that readers differ in the lexical quality of word representations and that poorer quality will lead to delays in word recognition. In the present study, spelling skills were taken as a proxy for lexical quality. Although spelling ability did not modulate results as strongly as decoding skill and vocabulary, the results from the separate analyses from strong spellers and weak spellers pointed to some important similarities and differences across the two languages.

Across both languages, stronger spellers were found to name words faster than weaker spellers. This suggests that better lexical representations facilitate reading aloud performance irrespective of language transparency.

Also across both languages, stronger spellers seemed to efficiently exploit orthographic similarity (old20) in naming, as greater orthographic neighbourhood was found to be facilitating for the stronger spellers group. Importantly, this effect was not evident for weaker spellers, but was observed across both languages. Within the framework of connectionist modelling, it has been suggested that orthographic knowledge can be accrued through statistical learning (e.g., Deacon et al., 2008). For the present result, this would suggest that stronger spellers may have had better orthographic knowledge because of more efficient statistical learning processes, and that this was irrespective of language transparency.

By extension, this finding can be interpreted to show that the processes underlying skilled reading do not differ between languages as suggested by proponents of a dual route model. The ODH, which is based on a dual route structure, foresees that sublexical processes contribute more to reading transparent scripts, whereas the lexical route contributes more to reading opaque scripts. In contrast, the finding supports the assumptions that reading in both languages occurs in a single route system based on distributed representations which have been acquired through interactive network learning. In both languages, it seems, skilled readers were able to exploit the orthographic similarities between words to acquire good orthographic representations.

The present study found some language differences apparent for weaker spellers. English (but not or to a lesser degree German) weaker spellers benefitted from the occurrence of phonographic *N* for rime 1 (PNR1), a marker for larger reading unit sizes. As this effect was modulated by IDs, it was suggested that people differ in how larger reading units are captured by their reading systems, and that this was linked to their individual strengths and weaknesses. However, as discussed elsewhere, this result has to be treated with caution as the PNR1 measure was not coherent across mono- and bisyllabic words. Whilst the PNR1 for monosyllables is most likely reflective of a genuine neighbourhood effect for the body-rime, and therefore reflects the use of larger grain sizes, the PNR1 effect for polysyllables most likely reflects the need for disambiguation of the first rime by means of using larger sized reading units. Consequently, the different reasons for assuming the involvement of larger grain sizes for mono-and bisyllabic words in this measure surely add some additional noise to this measurement.

Yet, even though the measure may be noisy, the results may still rightfully point to more differences between languages in weaker spellers than in stronger spellers. Whilst stronger spellers may find it easier to extract statistical regularities, this may be harder for weaker spellers, and hence each individual may have a different ‘strategy’ to find commonalities between items. As orthographies offer different sets of orthographical regularities, language becomes a greater factor for weaker spellers than for stronger spellers.

9.5 Print exposure

As mentioned earlier, strong decoders emerged in both languages as the fastest readers. This finding was not anticipated, as decoding skill is mostly prominent in theories of reading development (Ehri, 2005; Share, 1995). Instead, the tacit expectation was that at the level of reading skill as investigated in the current study, greater print exposure would be the distinguishing individual difference. This assumption was implicit given the surge of investigations into the effect of print exposure over the last few decades. Findings had shown that participants with more print exposure were faster at naming words of all frequencies than participants with less print exposure (Lewellen et al., 1993). Moreover, experienced readers were found to show less psycholinguistic effects than less experienced readers (Yap et al., 2012). PDP model simulations on semantic dementia patients had found that whilst lesion bias emerged as the strongest predictor for a reading deficit, this was followed by pre-morbid network training. By contrast, the strength of the O-P pathway was the weakest predictor for retained reading ability (Dilkina et al., 2008). Together, these findings had fuelled the implicit expectation that print exposure was the distinguishing individual difference between lower and higher levels of skilled reading.

Contrary to expectations, the high-print participants seemed to be a very heterogenous reader group. The main effects of consistency and AoA revealed variation within this group in terms of O-P and O-S-P mapping efficiency, suggesting that individuals varied in their respective route contributions. It thus becomes clear that greater reading practice does neither result in the most efficient reading process, nor does high print exposure drive every reading system towards the same structure.

Importantly, in the present study differences in decoding skill were still apparent within the high-print reader group, indicating that decoding skill was relevant even within this highly experienced group. In fact, the present study showed that decoding skill had the same effect across the high- and the low-print group: facilitatory for German readers at all lengths, and for English readers at short word lengths. The fact that readers were either high-print or low-print did not alter this effect. This suggests then that individual differences in decoding skill are more influential in the skilled reading process than individual differences in reading practice.

Moreover, recall that the strong decoder group showed the most similar reading effects patterns and emerged as the fastest readers in both languages. Together with the aforementioned, these results point to decoding skill as the most important individual difference to consider, even ahead of print exposure. Print exposure, then, does not offer to make readers most proficient, but just more proficient at the way they read, given their individual strengths and weaknesses. Whilst print exposure does not seem to drive systems towards similarities, nonword decoding skill does.

Although surprising, these results are in fact compatible with the triangle model of reading, as training (reading experience) improves performance, and this effect is not different across languages. However, the effect of training is limited to influence the reading process within the parameters given by individual differences in decoding skill.

9.6 Consistency databases for English and German

The above results would not have been possible without a number of preparatory tasks, the first of which is the creation of two methodologically coherent and comprehensive databases for onset, rime and composite feedforward and feedback consistencies for both English and German. In the case of the latter, this is the first of its kind. The fact that these databases were developed in a very similar manner with the objective to create comparable databases can be seen as an additional benefit for future cross-language research.

The creation of these two consistency databases also revealed some other interesting findings relevant to cross-language comparisons. In the present study, German was chosen to represent more transparent orthographies and English as an example of an opaque orthography. The results from the analysis of the consistency measures as computed within this study showed that German seemed indeed to be more spelling-sound consistent than English at both onset and rime level, with greater differences between the languages at the rime level.

Quantifying the language differences in terms of spelling-sound consistency as instantiated in the present computations was done under the assumption that spelling-sound consistency most closely reflected the differences between the two languages. It was argued that because spelling-sound consistency was expressed on a continuous scale, that it was nuanced and therefore a sensitive measure, and that the computation process reflected with some similarity the statistical mapping process between spelling and sound.

Given these assumptions, there remains an element of surprise that in the cognate naming study language emerged as a strong effect even though spelling-sound consistency was controlled. The important question to ask is what this language grouping variable comprised that was not captured by spelling-sound consistency? Although it is beyond the current study to determine what this language variable may include, the author would like to put forward some considerations on this matter.

First of all, from a technical viewpoint, it may be possible that despite successful reliability checks, the present spelling - sound consistency estimates were too crude. Specifically, whilst general trends in consistency computations were shown to be comparable between the present and previous computations (Yap & Balota, 2009; Ziegler, Stone, et al., 1997; personal communication from February 2001 of J Ziegler in Wimmer & Mayringer, 2002), estimates showed nevertheless some differences. This is expected as computations depend on the parsing procedure, corpus size (cf. Brysbaert & New, 2009 and their discussion of the importance of corpus size on word frequency estimates), as well as other properties of the words in the corpus, such as the number of syllables. Generally, computations based on larger corpora are expected to yield more reliable estimates, but some degree of uncertainty will of course remain. This may have contributed to the strong language effect.

Second, the language effect may comprise language transparency at other levels than the body-rime, such as letter-sound mappings, and morphemes. Alternatively, it could also include easiness of segmentation into reading units. It may however also involve the teaching methods predominant for the language. Further research is certainly warranted to assess what the language variable includes when controlling for onset and rime consistency.

To the author's knowledge, the present study constitutes the first cross-language investigation to control for spelling-sound consistency in such a specific and quantifiable manner. An important result is therefore that the language variable remains an important predictor for naming, beyond the spelling – sound consistency as it has been quantified in the current study. The language variable seems to subsume the environment which each language provides for its readers. The present study shows that this language environment remains a crucial element in reading aloud performance even when spelling-sound consistency is controlled.

9.7 The new GE-ART and GE-RQ measures

The second preparatory contribution of the present study is the creation of a new version of the ART and an adaptation of the Reading Habits Questionnaire (Stanovich & West, 1989). Both new measures, GE-ART and GE-RQ were created with the aim to be employed for cross-language research for English and German. The GE-ART seemed to be successful at capturing reading experience in both language groups. This is the first time that an ART was developed for two language groups. The GE-RQ was useful as a validation measure for the GE-ART. However, a factor analysis found that factor loadings were not the same for the two language groups. For the English language group, there seemed to be a general reading factor and a digital reading component. For the German language group, a general factor with somewhat different factor loadings to the English language group and a non-fiction reading factor seemed to emerge. It is possible that the questionnaire tapped into different reading cultures in the two language groups, which were dissimilarly related to reading experience as measured by the GE-ART. Whilst this was not the focus of the present study, it may be interesting in future investigations to explore this further. The difference in factor loadings on the GE-RQ results between the two language groups is also a reminder of the difficulty to create comparable measures, for cross-language research, which are equally valid in both languages.

9.8 AoA ratings for German

A third contribution of the present project was the collection of age-of-acquisition ratings for over 3,000 German words. Although published as a separate article (Birchenough et al., 2017) and added as an addendum to the thesis presented here, the collection was part and parcel of this project. Ratings for words shared with extant English estimates (Cortese & Khanna, 2008; Kuperman et al., 2012; Stadthagen-Gonzaelz & Davies, 2006) correlated highly (r_s between .7 and .77). The ratings will now enable other researchers to readily access AoA ratings for German.

9.9 Limitations

There are, of course, some limitations to the present study. The salient limitation is the relatively small number of word stimuli used. Cognates were presented in order to permit stricter control over the matching of stimulus sets between languages. However, the findings from the present investigation will need to be tested in a replication study, using more stimuli and other participant samples, to assure the reliability of the results. The author hopes that this will be possible sooner because of the ground work completed for the investigation.

Whilst every effort was made to make the measures of individual differences as comparable as possible between languages, this was not necessarily achieved with regard to spelling. Although the tests of spelling knowledge that were used both probed the relative quality of individuals' knowledge, future analyses would benefit from establishing equivalence between languages in tests. In order to investigate orthographic quality across languages, it would be worthwhile to devise a spelling test which is developed to examine orthographic knowledge in an exactly equivalent manner in both languages, preferably focusing on the accurate recall of the letter string.

9.10 Theoretical considerations and future directions

The findings of the present study bring into view, for the first time, an empirical picture of a reading system which varies between individuals and languages. The present study clearly demonstrates that language transparency is a major shaper of the reading system, but that individual differences also modulate the reading aloud process, and that language and IDs interact with each other. These individual differences seem to revolve around the ability to analyse and segment words into appropriate reading units. The recourse to semantics seems to be taken more quickly and successfully by readers of an opaque script, but is also governed by the necessity (among some individuals) to compensate for weaknesses in the decoding process. However, differences in the capacity of the decoding process appears to be paramount in shaping the responses of readers of both languages to words in the oral reading task. The theoretical implications of these key observations must be discussed in terms of potential extensions of existing accounts of the reading system.

Of the theoretical accounts that are influential in reading research today, those associated with the PDP framework seem the most likely to be capable of explaining the findings reported in this thesis. Within a PDP account, the primacy of individual differences in decoding ability is naturally accommodated by the assumption that the response of the reading system to words in word naming depends primarily on the functioning of the phonological (orthography-to-phonology) route. Specifically, there is a match between the findings of the present study and the account proposed by Harm & Seidenberg (2004) which retains the role of the phonological pathway as the default pathway for reading most items. In contrast, the assumptions of the DRC account (Coltheart et al., 2001) imply that skilled readers would make use of the lexical pathway, rather than the sublexical pathway, for all words represented in the vocabulary, and this expectation is simply not supported by the findings in the present study. Whether the effects of variables that are taken to be markers of lexical or sublexical processing were observed, depended critically on individual differences in decoding ability, vocabulary knowledge and spelling knowledge - even in a sample of skilled adult readers. Such a pattern of results could potentially be accommodated by modifying the parameters in instantiations of the DRC, but because such modification would be entirely post hoc, it is difficult to see how a dual route account can explain the present findings in a principled manner. The principles of the PDP approach (e.g., Seidenberg & Plaut, 2006), which link gradual adaptation in network structure and functioning to the distribution of training experiences (e.g., Monaghan & Ellis, 2010), and to the nature of the challenges posed by the learning environment (e.g., Harm & Seidenberg, 2004; Plaut et al., 1996), furnish a theoretical framework which makes it considerably easier to imagine how a PDP model would adapt in response to differences in language, reading experience, and the strength of semantic, phonological and orthographic processing capacities (cf. Dilkina et al., 2008), giving rise to the patterns of behaviour observed in the present study.

In short, it can be concluded that the results of the present study appear to favour PDP approaches to modelling reading behaviour; however, there is a caveat, that must be considered with this conclusion. Rueckl (2016) suggested that even the triangle model in its current form needs a broader framework to adequately address variation within the reading system as posed by individual differences and language. He proposes a framework that encompasses a set of control parameters which define and constrain the space within which the reading system exists. The control parameters are based on factors which are either inherent to the environment (e.g. writing system or instruction method), or inherent to the person (e.g. biological factors). As a result of these constraining factors, the reading system will have adopted an internal organisation. This organisation may then be further adjusted in its structure across its lifetime in response to experience. This proposed framework seems to be a very promising basis for developing a future account of reading across language and individual differences.

The present study is one of the first studies to have attempted to include several of the constraining factors at the same time: language along with individual differences in decoding, vocabulary, print exposure and spelling. It is not possible at this time to say how these factors would fit into the framework suggested by Rueckl (2016), nor whether the effects examined here correspond to the ones envisaged by Rueckl (2016). However, what is important is that there is the accepted need for understanding the reading system in its greater context and to understand the impact of influencing factors on its organisation. What is required next is to use the framework suggested by Rueckl (2016) to create a clearer picture of how the reading system changes systematically, with each constraining factor, and along pre-specified dimensions. In practice, this approach may be limited in its effectiveness by computational capacities and possibilities. It may also face challenges with regard to the choice of constraining factors and dimensions. Yet, despite the difficulties that such a project may encounter, the current study clearly demonstrates that the evidence requires it: readers were affected by both language and individual differences, and that needs to be taken into account to understand reading aloud as a universal skill.

9.11 Conclusion

This study has brought together two research areas regarding reading aloud: differences due to language transparency and differences due to individual differences. The effect of differences between languages is highly salient but individual differences are also associated with important effects, and these unfold differently depending on whether the reading system is reading opaque English or more transparent German. Importantly, although there were tendencies that participants who were weaker in one ID task were also weaker in others, no distinct reader profiles became apparent. Differences between readers, in the effects of psycholinguistic variables, are found to diminish among readers who achieve higher levels of reading expertise. More reading experience improves reading systems regardless of the language being read, but avid readers do not all seem to be reading in the same way. Decoding skill seems to be the most influential individual difference, as strong decoders are the fastest readers in both languages. Readers with strong decoding skills also showed effects patterns most similar across the two languages. Readers of English showed more facilitating effects from greater vocabulary knowledge than readers of German. Good spelling ability was faciliatory in naming for both language groups. The results imply that spelling ability is related to the ability to segment words into various constituents, and that for weaker spellers the ability to detect larger grain sizes may be modulated by other IDs, specifically for readers of opaque English. Thus, although readers' reading system will be influenced by the structure of the language they read, and the strengths and weaknesses they bring in terms of IDs, there is still a great variability between readers, which highlights that all these factors are important when modelling reading as a universal skill.

10 References

Aaron, P., Joshi, R., Ayotollah, M., Ellsberry, A., Henderson, J. & Lindsey, K. (1999). Decoding and sight-word naming: Are they independent components of word recognition skill? *Reading and Writing*, *11*, 89–127.

Acheson, D. J., Wells, J. B. & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods*, *40*, 278–289.

Adelman, J. S. & Brown, G. D. (2007). Phonographic neighbors, not orthographic neighbors, determine word naming latencies. *Psychonomic Bulletin & Review*, *14*, 455–459.

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

Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? *Memory & Cognition*, *10*, 565–575.

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 234 –254.

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

Andrews, S. (2012). Individual differences in skilled visual word recognition and reading. In Adelman, J.S. (Ed). *Visual Word Recognition Volume 2: Meaning and Context, Individuals and Development*, (pp. 151 – 172). Hove, East Sussex, England: Psychology Press.

Andrews, S. & Bond, R. (2009). Lexical expertise and reading skill: Bottom-up and top-down processing of lexical ambiguity. *Reading and Writing*, *22*, 687–711.

Andrews, S. & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General*, *139*, 299–318.

Andrews, S. & Lo, S. (2012). Not all skilled readers have cracked the code: Individual differences in masked form priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 152.

Andrews, S. & Lo, S. (2013). Is morphological priming stronger for transparent than opaque words? It depends on individual differences in spelling and vocabulary. *Journal of Memory and Language*, *68*, 279–296.

Anthony, J. L. & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science*, *14*, 255–259.

Arduino, L. S. & Burani, C. (2004). Neighborhood effects on nonword visual processing in a language with shallow orthography. *Journal of Psycholinguistic Research*, *33*, 75–95.

Aro, M. & Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics*, *24*, 621–635.

Baayen, R. H., Bates, D., Kliegl, R. & Vasishth, S. (2015). *RePsychLing: Data sets from Psychology and Linguistics experiments*.

Baayen, R. H., Davidson, D. J. & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Baayen, R. H., Piepenbrock, R. & Gulikers, L. (1995). *The CELEX lexical database (Release 2). Linguistic Data Consortium, University of Pennsylvania*. Philadelphia.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H. & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*, 283–316.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., et al. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445–459.

Baron, J. & Strawson, C. (1976). Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 386 – 393.

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*, 255–278.

Barton, J. J., Hanif, H. M., Eklinder Björnström, L. & Hills, C. (2014). The word-length effect in reading: A review. *Cognitive Neuropsychology*, *31*, 378–412.

Barton, K. (2018). *MuMIn: Multi-Model Inference*. Retrieved from https://CRAN.R-project.org/package=MuMIn

Bates, D., Kliegl, R., Vasishth, S. & Baayen, R. H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1 – 48.

Bentin, S. (1992). Phonological awareness, reading, and reading acquisition: A survey and appraisal of current knowledge. *Haskins Laboratories Status Report on Speech Research*, *SR-111|112*, 167–180.

Besner, D., Twilley, L., McCann, R. S. & Seergobin, K. (1990). On the association between connectionism and data: Are a few words necessary? *Psychological Review*, 432–446.

Birchenough, J. M., Davies, R. & Connelly, V. (2017). Rated age-of-acquisition norms for over 3,200 German words. *Behavior Research Methods*, *49*, 484–501.

Bosman, A. M. & de Groot, A. M. (1992). Differential effectiveness of reading and non-reading tasks in learning to spell. In Satow, F. and Gatherer, B. (Ed.), *Literacy without frontiers* (pp. 279–289). Widnes, Cheshire: United Kingdom Reading Association.

Brown, C. R. & Rubenstein, H. (1961). Test of Response Bias Explanation of Word-Frequency Effect. *Science*, *133*, 280–281.

Brown, G. D. & Watson, F. L. (1987). First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory & Cognition*, *15*, 208–216.

Brown, P., Lupker, S. J. & Colombo, L. (1994). Interacting sources of information in word naming: A study of individual differences. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 537–554.

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J. & Böhl, A. (2011). The word frequency effect. *Experimental Psychology (formerly Zeitschrift für Experimentelle Psychologie)*, *58*, 412–424.

Brysbaert, M. & Ghyselinck, M. (2006). The effect of age of acquisition: Partly frequency related, partly frequency independent. *Visual Cognition*, *13*, 992–1011.

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

Brysbaert, M. & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990.

Brysbaert, M., Wijnendaele, I. V. & Deyne, S. D. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica*, *104*(2), 215–226.

Bub, D. C. A. & Kertesz, A. (1985). Whole-word and analytic translation of spelling to sound in a non-semantic reader. In Patterson, K.E., Marshall, J.C., Coltheart, M. (Ed.), (pp. 15–34). Hove, UK: Lawrence Erlbaum Associates, Ltd.

Burani, C., Arduino, L. S. & Barca, L. (2007). Frequency, not age of acquisition, affects Italian word naming. *European Journal of Cognitive Psychology*, *19*, 828–866.

Burt, J. S. & Fury, M. B. (2000). Spelling in adults: The role of reading skills and experience. *Reading and Writing*, *13*, 1–30.

Burt, J. S. & Tate, H. (2002). Does a reading lexicon provide orthographic representations for spelling? *Journal of Memory and Language*, *46*, 518–543.

Butler, B. & Hains, S. (1979). Individual differences in word recognition latency. *Memory & Cognition*, *7*, 68–76.

Callens, M., Tops, W., Stevens, M. & Brysbaert, M. (2014). An exploratory factor analysis of the cognitive functioning of first-year bachelor students with dyslexia. *Annals of dyslexia*, *64*, 91–119.

Castles, A. & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition*, 47, 149-180.

Chang, Y.-N., Furber, S. & Welbourne, S. (2012). “Serial” effects in parallel models of reading. *Cognitive psychology*, *64*, 267–291.

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

Chetail, F. (2015). Reconsidering the role of orthographic redundancy in visual word recognition. *Frontiers in psychology*, *6*, 645.

Chomsky, N. & Halle, M. (1968). *The sound pattern of English.* New York: Harper & Row.

Chumbley, J. I. & Balota, D. A. (1984). A word’s meaning affects the decision in lexical decision. *Memory & Cognition*, *12*, 590–606.

Clark, C. & Foster, A. (2005). *Children’s and young people’s reading habits and preferences: The who, what, why, where and when*. National Literacy Trust London.

Cohen, P., Cohen, J., West, S. & Aiken, L. (2003). *Applied Multiple Regression/Correlation for the Behavioral Sciences, 3rd edn. Mahwah, New Jersey: L.* Erlbaum.

Colenbrander, D., Nickels, L. & Kohnen, S. (2011). Nonword reading tests: A review of the available resources. *Australasian Journal of Special Education*, *35*, 137–172.

Coltheart, M. (2004). Are there lexicons? *The Quarterly Journal of Experimental Psychology Section A*, *57*, 1153–1171.

Coltheart, M. (2012). Dual-route theories of reading aloud. In Adelman, J.S. (Ed.), *Visual word recognition: Models and methods, orthography and phonology* (Vol. 1, pp. 3–27). Hove, East Sussex, UK: Psychology Press.

Coltheart, M., Curtis, B., Atkins, P. & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, *100*, 589–608.

Coltheart, M., Davelaar, E., Jonasson, T. & Besner, D. (1977). Attention and Performance VI. In S. Dornic (Ed.), (Vol. VI, pp. 535–555). New York: Academic Press.

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

Connelly, V., Thompson, G. B., Fletcher-Flinn, C. M. & McKay, M. F. (2009). Does the type of reading instruction have an influence on how readers process print? In Wood, C. and Connelly, V. (Ed.), (Vol. Contemporary Perspectives on Reading and Spelling, pp. 239–253). London and New York: Routledge.

Cortese, M. J. & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *The Quarterly Journal of Experimental Psychology*, *60*, 1072–1082.

Cortese, M.J. & Khanna, M.M. (2008). Age of acquisition ratings for 3000 monosyllabic words. *Behavior Research Methods*, *40*, 791 – 794.

Cortese, M. J. & Schock, J. (2013). Imageability and age of acquisition effects in disyllabic word recognition. *The Quarterly Journal of Experimental Psychology*, *66*, 946–972.

Cortese, M. J. & Simpson, G. B. (2000). Regularity effects in word naming: What are they? *Memory & Cognition*, *28*, 1269–1276.

Cortese, M. J., Yates, M., Schock, J. & Vilks, L. (2018). Examining word processing via a megastudy of conditional reading aloud. *Quarterly Journal of Experimental Psychology*, 1–19.

Crisp, J. & Lambon Ralph, M. A. (2006). Unlocking the nature of the phonological-deep dyslexia continuum: The keys to reading aloud are in phonology and semantics. *Journal of Cognitive Neuroscience*, *18*, 348–362.

Cuetos, F. & Barbón, A. (2006). Word naming in Spanish. *European Journal of Cognitive Psychology*, *18*, 415–436.

Cunningham, A. E., Perry, K. E., Stanovich, K. E. & Share, D. L. (2002). Orthographic learning during reading: Examining the role of self-teaching. *Journal of Experimental Child Psychology*, *82*, 185–199.

Davies, R., Arnell, R., Birchenough, J. M., Grimmond, D. & Houlson, S. (2017). Reading through the life span: Individual differences in psycholinguistic effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 1298–1338.

Davies, R., Wilson, M., Cuetos, F. & Burani, C. (2014). Reading in Spanish and Italian: Effects of age of acquisition in transparent orthographies? *The Quarterly Journal of Experimental Psychology*, *67*, 1808–1825.

Davis, C. J. (2012). The orthographic similarity of printed words. In Adelman, J.S. (Ed.), *Visual word recognition: Models and Methods, Orthography and Phonology* (Vol. 1, pp. 185–206). London and New York: Psychology Press.

Deacon, S. H., Conrad, N. & Pacton, S. (2008). A statistical learning perspective on children's learning about graphotactic and morphological regularities in spelling. *Canadian Psychology/Psychologie Canadienne*, *49*, 118–124.

Duden (1990). Duden Aussprachewörterbuch: *Wörterbuch der deutschen Standardaussprache, 3. völlig neu bearbeitete und erweiterte Auflage. bearbeitet von Max Mangold in Zusammenarbeit mit der Dudenredaktion.* Mannheim, Wien, Zürich: Dudenverlag

Dilkina, K., McClelland, J. L. & Plaut, D. C. (2008). A single-system account of semantic and lexical deficits in five semantic dementia patients. *Cognitive Neuropsychology*, *25*, 136–164.

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of reading*, *9*, 167–188.

Ehri, L. C., Nunes, S. R., Stahl, S. A. & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research*, *71*, 393–447.

Ellis, A. W. (2016). *Reading, writing and dyslexia: a cognitive analysis*. Psychology Press.

Ellis, A. W. & Lambon Ralph, M. A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1103–1123.

Ellis, N. C. & Hooper, A. M. (2001). Why learning to read is easier in Welsh than in English: Orthographic transparency effects evinced with frequency-matched tests. *Applied Psycholinguistics*, *22*, 571–599.

Ellis, N. C., Natsume, M., Stavropoulou, K., Hoxhallari, L., Daal, V. H., Polyzoe, N., Tsipa, M.-L., et al. (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading Research Quarterly*, *39*, 438–468.

Evett, L. J. & Humphreys, G. W. (1981). The use of abstract graphemic information in lexical access. *The Quarterly Journal of Experimental Psychology*, *33*(4), 325–350. Taylor & Francis.

Fischer, W. F., Shankweiler, D. & Liberman, I. Y. (1985). Spelling proficiency and sensitivity to word structure. *Journal of Memory and Language*, *24*, 423–441.

Fitts, P. M. & Posner, M. I. (1967). *Human performance.* Belmont, CA: Brooks/Cole.

Forster, K. I. & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*, 116–124.

Fox, J. (2003). Effect Displays in R for Generalised Linear Models. *Journal of Statistical Software*, *8*(15), 1–27. Retrieved from http://www.jstatsoft.org/v08/i15/

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

Friesen, D.C., Jared, D. & Haigh, C.A. (2014). Phonological Processing Dynmaics in bilingual word naming. *Canadian Journal of Experimental Psychology*, 68, 179-193.

Frith, U. (1980). Unexpected spelling problems. In Frith, Uta (Ed.), *Cognitive processes in spelling* (pp. 495–515). London: Academic Press.

Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. Patterson, J. Marshall & M. Coltheart (Ed.), *Surface dyslexia* (pp. 301–330). London: Erlbaum.

Frith, U., Wimmer, H. & Landerl, K. (1998). Differences in phonological recoding in German-and English-speaking children. *Scientific Studies of Reading*, 31–54.

Frost, R., Katz, L. & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: a multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 104–115.

Funnell, E. (1983). Phonological processes in reading: New evidence from acquired dyslexia. *British Journal of Psychology*, *74*, 159–180.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 674–691.

Gordon, R. A. (1987). Social desirability bias: A demonstration and technique for its reduction. *Teaching of Psychology*, *14*, 40-42.

Goswami, U. (2002). Phonology, reading development, and dyslexia: A cross-linguistic perspective. *Annals of Dyslexia*, *52*, 139–163.

Goswami, U. (2008). Learning to read across languages: The role of phonics and synthetic phonics. In Goouch, K. & Lambirth, A. (Ed.), *Understanding Phonics and the Teaching of Reading: critical perspectives* (pp. 124–143).

Goswami, U., Ziegler, J. C., Dalton, L. & Schneider, W. (2003). Nonword reading across orthographies: How flexible is the choice of reading units? *Applied Psycholinguistics*, *24*, 235–247.

Graham, S. & Santangelo, T. (2014). Does spelling instruction make students better spellers, readers, and writers? A meta-analytic review. *Reading and Writing*, 1–41.

Grainger, J. & Jacobs, A. M. (1996). Orthographic Processing in Visual Word Recognition: A Multiple Read-Out Model. *Psychological Review*, *103*, 518–565.

Hanley, J. R. & Kay, J. (1992). Does letter-by-letter reading involve the spelling system? *Neuropsychologia*, *30*, 237–256.

Harm, M. W. & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106*(3), 491–528.

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*, 662–720.

Harrell, F. E. J. & Dupont, C. (2018). *Hmisc: Harrell Miscellaneous*. Retrieved from https://CRAN.R-project.org/package=Hmisc

Hepner, C., McCloskey, M. & Rapp, B. (2017). Do reading and spelling share orthographic representations? Evidence from developmental dysgraphia. *Cognitive Neuropsychology*, *34*, 119–143.

Hersch, J. & Andrews, S. (2012). Lexical quality and reading skill: Bottom-up and top-down contributions to sentence processing. *Scientific Studies of Reading*, *16*(3), 240–262. Taylor & Francis.

Van Heuven, W. J., Mandera, P., Keuleers, E. & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*, 1176–1190.

Hino, Y. & Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming: An alternative to lexical access accounts. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 1331–1356.

Hino, Y., Lupker, S. J. & Pexman, P. M. (2002). Ambiguity and synonymy effects in lexical decision, naming, and semantic categorization tasks: Interactions between orthography, phonology, and semantics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 686–713.

Hodges, J. R. & Patterson, K. (2007). Semantic dementia: a unique clinicopathological syndrome. *The Lancet Neurology*, *6*, 1004–1014.

Hodges, J. R., Patterson, K., Oxbury, S. & Funnell, E. (1992). Semantic dementia: Progressive fluent aphasia with temporal lobe atrophy. *Brain*, *115*, 1783–1806.

Hoffman, P., Lambon Ralph, M. A. & Woollams, A. M. (2015). Triangulation of the neurocomputational architecture underpinning reading aloud. *Proceedings of the National*

Academy of Sciences, *112*(28), E3719–E3728.

Holmes, V. M. & Carruthers, J. (1998). The relation between reading and spelling in skilled adult readers. *Journal of Memory and Language*, *39*, 264–289.

Holmes, V. M. & Ng, E. (1993). Word-specific knowledge, word-recognition strategies, and spelling ability. *Journal of Memory and Language*, *32*, 230–257.

Hutzler, F., Ziegler, J. C., Perry, C., Wimmer, H. & Zorzi, M. (2004). Do current connectionist learning models account for reading development in different languages? *Cognition*, *91*, 273–296.

Jacobs, A. M., Nuerk, H.-C., Graf, R., Braun, M. & Nazir, T. A. (2008). The initial capitalization superiority effect in German: Evidence for a perceptual frequency variant of the orthographic cue hypothesis of visual word recognition. *Psychological Research*, *72*, 657–665.

Jared, D. (1997). Spelling-sound consistency affects the naming of high-frequency words. *Journal of Memory and Language*, *36*, 505–529.

Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming. *Journal of Memory and Language*, *46*, 723–750.

Jared, D. & Kroll, J.F. (2001) Do bilinguals activate phonological representations in one or both of their languages when naming words? *Journal of Memory and Language*, 44, 2-31.

Jared, D., McRae, K. & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language*, *29*, 687–715.

Jones, A. C. & Rawson, K. A. (2016). Do reading and spelling share a lexicon? *Cognitive Psychology*, *86*, 152–184.

Jorm, A. F. & Share, D. L. (1983). Phonological recoding and reading acquisition. *Applied Psycholinguistics*, *4*, 103–147.

Juhasz, B. J. & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1312–1318.

Katz, L. & Feldman, L. B. (1983). Relation between pronunciation and recognition of printed words in deep and shallow orthographies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 157–166.

Katz, L. & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In Katz, Leonard and Frost, Ram (Ed.), *Orthography, Phonology, Morphology and Meaning* (pp. 67–84). Amsterdam: Elsevier.

Kay, J. & Bishop, D. (1987). Anatomical differences between nose, palm, and foot, or, the body in question: Further dissection of the processes of sub-lexical spelling-sound translation. In Coltheart, M. (Ed.), *Attention and performance: The Psychology of Reading* (p. 449–469). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Kersting, M. & Althoff, K. (2004). *Rechtschreibungstest: RT*. Göttingen: Hogrefe.

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

Krueger, J. (1998). Enhancement bias in descriptions of self and others. *Personality and Social Psychology Bulletin*, *24*, 505–516.

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

Kuperman, V. & Van Dyke, J. A. (2011). Individual differences in visual comprehension of morphological complexity. In L. Carlson, C. Hoelscher, T. Shipley (Ed.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 1643–1648). Austin, TX: Cognitive Science Society.

Kuperman, V., Stadthagen-Gonzalez, H. & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*, 978–990.

Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*.

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

Landerl, K. (2000). Influences of orthographic consistency and reading instruction on the development of nonword reading skills. *European Journal of Psychology of Education*, *15*, 239–257.

Landerl, K. (2002). Word recognition deficits in German: More evidence from a representative sample. *Dyslexia*, *7*, 183–196.

Landerl, K., Wimmer, H. & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German-English comparison. *Cognition*, *63*, 315–334.

Lewellen, M. J., Goldinger, S. D., Pisoni, D. B. & Greene, B. G. (1993). Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision, and semantic categorization. *Journal of Experimental Psychology: General*, *122*, 316–330.

Liberman, I. Y., Liberman, A. M., Mattingly, I. & Shankweiler, D. (1980). Orthography and the beginning reader. In Kavanagh, J.F. and Venezky, R.L. (Ed.), *Orthography, Reading, and Dyslexia* (pp. 137–153). Baltimore: University Park Press.

Liberman, I. Y., Shankweiler, D. & Liberman, A. M. (1989). The alphabetic principle and learning to read. In Shankweiler, Donald and Liberman, Isabelle Y (Ed.), *Phonology and Reading Disability: Solving the Reading Puzzle* (pp. 1–33). Ann Arbor, MI: The University of Michigan Press.

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.

Manis, F.R., Seidenberg, M.S., Doi, L.M., McBride-Chang, C., & Peterson, A. (1996) On the basis of two subtypes of developmental dyslexia. *Cognition*, 58, 157-195.

Marian, V., Bartolotti, J., Chabal, S. & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PloS one*, *7*(8), e43230.

Marshall, J. C. & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research*, *2*, 175–199.

Martin-Chang, S. L. & Gould, O. N. (2008). Revisiting print exposure: exploring differential links to vocabulary, comprehension and reading rate. *Journal of Research in Reading*, *31*, 273–284.

Martin-Chang, S. L., Ouellette, G. & Madden, M. (2014). Does poor spelling equate to slow reading? The relationship between reading, spelling, and orthographic quality. *Reading and Writing*, *27*, 1485–1505.

Mason, M. (1978). From print to sound in mature readers as a function of reader ability and two forms of orthographic regularity. *Memory & Cognition*, *6*, 568–581.

Masterson, J. & Hayes, M. (2007). Development and data for UK versions of an author and title recognition test for adults. *Journal of Research in Reading*, *30*, 212–219.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H. & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.

McCarthy, R. & Warrington, E. K. (1986). Phonological reading: Phenomena and paradoxes. *Cortex*, *22*(3), 359–380.

McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*, 375–407.

McGeown, S. P., Duncan, L. G., Griffiths, Y. M. & Stothard, S. E. (2014). Exploring the relationship between adolescent's reading skills, reading motivation and reading habits. *Reading and Writing*, 1–25.

McKay, A., Davis, C., Savage, G. & Castles, A. (2008). Semantic involvement in reading aloud: evidence from a nonword training study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1495–1517.

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

Moll, K., Fussenegger, B., Willburger, E. & Landerl, K. (2009). RAN is not a measure of orthographic processing. Evidence from the asymmetric German orthography. *Scientific Studies of Reading*, *13*, 1–25.

Moll, K. & Landerl, K. (2009). Double dissociation between reading and spelling deficits. *Scientific Studies of Reading*, *13*, 359–382.

Monaghan, J. & Ellis, A. W. (2002). What exactly interacts with spelling-sound consistency in word naming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 183–206.

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.

Monaghan, P. & Ellis, A. W. (2010). Modeling reading development: Cumulative, incremental learning in a computational model of word naming. *Journal of Memory and Language*, *63*, 506–525.

Monsell, S. (1987). Nonvisual orthographic processing and the orthographic input lexicon. In Coltheart, M. (Ed.), (Vol. Attention and peformance XII: Reading, pp. 299–323). Hove, UK: Lawrence Erlbaum Associates, Inc.

Moore, M. & Gordon, P. C. (2015). Reading ability and print exposure: item response theory analysis of the author recognition test. *Behavior Research Methods*, *47*, 1095–1109.

Nagy, W. E. & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, 304–330.

Nation, K., Angell, P. & Castles, A. (2007). Orthographic learning via self-teaching in children learning to read English: Effects of exposure, durability, and context. *Journal of Experimental Child Psychology*, *96*, 71–84.

New, B., Ferrand, L., Pallier, C. & Brysbaert, M. (2006). Re-examining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review*, *13*, 45–52.

Niessen, M., Frith, U., Reitsma, P. & Ohngren, B. (2000). Learning disorders as a barrier to human development 1995-1999. *Evaluation report. Technical Committee COST Social Sciences*.

Noor, N. M. (2011). Reading habits and preferences of EFL post graduates: A Case Study. *Indonesian Journal of Applied Linguistics*, *1*, 1–9.

Pagliuca, G. & Monaghan, P. (2010). Discovering large grain sizes in a transparent orthography: Insights from a connectionist model of Italian word naming. *European Journal of Cognitive Psychology*, *22*(5), 813–835.

Patterson, K. (1986). Lexical but nonsemantic spelling? *Cognitive Neuropsychology*, *3*, 341–367.

Patterson, K. E. & Morton, J. (1985). *From orthography to phonology: An attempt at an old interpretation*. London: Erlbaum.

Pecher, D. (2001). Perception is a two-way junction: Feedback semantics in word recognition. *Psychonomic Bulletin & Review*, *8*, 545–551.

Peereman, R. & Content, A. (1997). Orthographic and phonological neighborhoods in naming: Not all neighbors are equally influential in orthographic space. *Journal of Memory and Language*, *37*, 382–410.

Peereman, R., Content, A. & Bonin, P. (1998). Is perception a two-way street? The case of feedback consistency in visual word recognition. *Journal of Memory and Language*, *39*, 151–174.

Perfetti. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*, 357–383.

Perfetti, & Hart. (2002). The lexical quality hypothesis. In Verhoeven, Ludo and Elbro, Carsten and Reitsma, Pieter (Ed.), *Precursors of Functional Literacy* (pp. 67–86). Amsterdam/Philadephia: John Benjamins Publishing.

Perry, C. & Ziegler, J. C. (2002). Cross-language computational investigation of the length effect in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 990–1001.

Perry, C., Ziegler, J. C. & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud. *Psychological Review*, *114*, 273–315.

Perry, C., Ziegler, J. C. & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology*, *61*, 106–151.

Pexman, P. M. & Yap, M. J. (2018). Individual Differences in Semantic Processing: Insights from the Calgary Semantic Decision Project. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 1091–1112.

Plaut, D. C. (1999). A connectionist approach to word reading and acquired dyslexia: Extension to sequential processing. *Cognitive Science*, *23*, 543–568.

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*, 56.

Protopapas, A. (2007). Check Vocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, *39*, 859–862.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Retrieved from http://www.R-project.org/

Rack, J. P., Snowling, M. J. & Olson, R. K. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly*, 29–53.

Rastle, K. & Coltheart, M. (1998). Whammies and double whammies: The effect of length on nonword reading. *Psychonomic Bulletin & Review*, *5*, 277–282.

Rastle, K. & Coltheart, M. (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 482–503.

Rastle, K., Davis, M. H. & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, *11*, 1090–1098.

Rau, A. K., Moll, K., Snowling, M. J. & Landerl, K. (2015). Effects of orthographic consistency on eye movement behavior: German and English children and adults process the same words differently. *Journal of Experimental Child Psychology*, 92–105.

Revelle, W. (2017). psych: Procedures for psychological, psychometric, and personality research. *Northwestern University, Evanston, Illinois*.

Rossi, M., Martin-Chang, S. & Ouellette, G. (2019). Exploring the Space Between Good and Poor Spelling: Orthographic Quality and Reading Speed. *Scientific Studies of Reading*, *23*, 192–201.

Roux, S. & Bonin, P. (2009). Neighborhood effects in spelling in adults. *Psychonomic Bulletin & Review*, *16*, 369–373.

Rueckl, J. G. (2016). Toward a theory of variation in the organization of the word reading system. *Scientific Studies of Reading*, *20*, 86–97.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*, 533–536.

Schmalz, X., Marinus, E., Robidoux, S., Palethorpe, S., Castles, A. & Coltheart, M. (2014). Quantifying the reliance on different sublexical correspondences in German and English. *Journal of Cognitive Psychology*, *26*, 831–852.

Schmalz, X., Robidoux, S., Castles, A., Coltheart, M. & Marinus, E. (2017). German and English bodies: No evidence for cross-linguistic differences in preferred grain size. *Collabra: Psychology*, *3*(1), 5.

Schmidt, F. T. & Retelsdorf, J. (2016). A New Measure of Reading Habit: Going Beyond Behavioral Frequency. *Frontiers in Psychology*, *7*, 1364.

Schröter, P. & Schroeder, S. (2017). The Developmental Lexicon Project: A behavioral database to investigate visual word recognition across the lifespan. *Behavior Research Methods*, 1–21.

Schulte-Körne, G., Deimel, W., & Remschmidt, H. (1997). Can self-report data on deficits in reading and spelling predict spelling disability as defined by psychometric tests? *Reading and Writing*, *9*, 55-63.

Sears, C. R., Siakaluk, P. D., Chow, V. C. & Buchanan, L. (2008). Is there an effect of print exposure on the word frequency effect and the neighborhood size effect? *Journal of Psycholinguistic Research*, *37*, 269–291.

Seidenberg, M. S. (1992). Beyond orthographic depth in reading: Equitable division of labor. In Frost, R., and Katz, L. (Ed.), *Orthography, Phonology, Morphology, and Meaning* (pp. 85–118). Amsterdam: Elsevier.

Seidenberg, M. S. & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523–568.

Seidenberg, M. S. & Plaut, D. C. (1998). Evaluating word-reading models at the item level: Matching the grain of theory and data. *Psychological Science*, *9*, 234–237.

Seidenberg, M. S. & Plaut, D. C. (2006). Progress in understanding word reading: Data fitting versus theory building. In Andrews, S. (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (pp.25-49). Hove, UK: Psychology Press.

Seidenberg, M. S., Waters, G. S., Barnes, M. A. & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behaviour*, *23*, 383–404.

Seymour, P. H. K., Aro, M. & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*, 143–174.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, *55*, 151–218.

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, *72*, 95–129.

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: the perils of overreliance on an “outlier” orthography. *Psychological Bulletin*, 584–615.

Smith, M. C. (2000). The real-world reading practices of adults. *Journal of Literacy Research*, *32*, 25–52.

Snowden, J. S., Goulding, P. J. & Neary, D. (1989). Semantic dementia: A form of circumscribed cerebral atrophy. *Behavioural Neurology*, *2*, 167–182.

Sonnenstuhl, I., Eisenbeiss, S. & Clahsen, H. (1999). Morphological priming in the German mental lexicon. *Cognition*, *72*, 203–236.

Spear-Swerling, L., Brucker, P. O., & Alfano, M. P. (2010). Relationships between sixth-graders' reading comprehension and two different measures of print exposure. *Reading and Writing*, *23*(1), 73-96.

Spieler, D. H. & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, 411–416.

Spieler, D. H. & Balota, D. A. (2000). Factors influencing word naming in younger and older adults. *Psychology and Aging*, *15*, 225–231.

Spinelli, D., De Luca, M., Di Filippo, G., Mancini, M., Martelli, M. & Zoccolotti, P. (2005). Length effect in word naming in reading: Role of reading experience and reading deficit in Italian readers. *Developmental Neuropsychology*, *27*, 217–235.

Stadthagen-Gonzalez, H. & Davis, C.J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods, 38*, 598–605.

Stainthorp, R. (1997). A children's author recognition test: A useful tool in reading research. *Journal of Research in Reading*, *20*, 148–158.

Stanovich, K. E. & Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory & Cognition*, *20*, 51–68.

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

Stanovich, K. E., West, R. F. & Harrison, M. R. (1995). Knowledge growth and maintenance across the life span: The role of print exposure. *Developmental Psychology*, *31*, 811–826.

Steyvers, M. & Tenenbaum, J. B. (2005). The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive Science*, *29*, 41–78.

Stone, G. O., Vanhoy, M. & Van Orden, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language*, *36*, 337–359.

Strain, E. & Herdman, C. M. (1999). Imageability effects in word naming: an individual differences analysis. *Canadian Journal of Experimental Psychology*, *53*, 347–359.

Strain, E., Patterson, K. & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1140–1154.

Strain, E., Patterson, K. & Seidenberg, M. S. (2002). Theories of word naming interact with spelling-sound consistency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 207–214.

Taft, M. (1987). Morphographic processing. The BOSS re-emerges. In M. Coltheart (Ed.), *Attention & performance XII* (pp 265-279), Hillsdale, NJ: Erlbaum.

Taft, M. (1992). The body of the BOSS: Subsyllabic units in the lexical processing of polysyllabic words. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 1004 – 1014.

Thompson, G. B., McKay, M. F., Fletcher-Flinn, C. M., Connelly, V., Kaa, R. T. & Ewing, J. (2008). Do children who acquire word reading without explicit phonics employ compensatory learning? Issues of phonological recoding, lexical orthography, and fluency. *Reading and Writing*, *21*, 505–537.

Torgesen, J. K., Wagner, R. K. & Rashotte, C. A. (1999). Test of word reading efficiency. *Austin, TX: Pro-Ed.*

Torgesen, J. K., Wagner, R. K. & Rashotte, C. A. (2012). *TOWRE-2 Examiner's Manual. Austin, TX: Pro-Ed.*

Treiman, R. & Chafetz, J. (1987). Are there onset-and rime-like units in printed words. In Coltheart, M. (Ed.), *Attention and performance, XII: Reading* (pp. 281–298). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Treiman, R. & Kessler, B. (2006). Spelling as statistical learning: Using consonantal context to spell vowels. *Journal of Educational Psychology*, *98*, 642–652.

Treiman, R., Mullennix, J., Bijeljac-Babic, R. & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, *124*, 107–136.

Van Loon - Vervoorn, W. A. (1989). *Eigenschappen van basiswoorden*. Lisse: Swets and Zeitlinger.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J. & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, *45*, 2–40.

Venezky, R.L. (1970) The structure of English orthography. The Hague: Mouton, 1970.

Wagner, R. K. & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, *101*, 192–212.

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *50*, 439–456.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale - Fourth Edition*. San Antonio, TX: NCS Pearson.

Wechsler, D. (2012). *Wechsler adult intelligence scale-fourth edition (WAIS-IV). Deutschsprachige Adaptation der WAIS-IV von D. Wechsler.* Frankfurt: Pearson.

White, S., Chen, J. & Forsyth, B. (2010). Reading-related literacy activities of American adults: Time spent, task types, and cognitive skills used. *Journal of Literacy Research*, *42*, 276–307.

Wickelgren, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review*, *76*(1), 1–15.

Wilkinson, G. S. (1993). *WRAT-3: Wide range achievement test administration manual*. Wide Range, Incorporated.

Wimmer, H. & Goswami, U. (1994). The influence of orthographic consistency on reading development: Word recognition in English and German children. *Cognition*, *51*, 91–103.

Wimmer, H. & Mayringer, H. (2002). Dysfluent reading in the absence of spelling difficulties: A specific disability in regular orthographies. *Journal of Educational Psychology*, *94*, 272–277.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499. [http://arxiv.org/pdf/1308.5499.pdf]

Woollams, A. M., Lambon Ralph, M. A., Madrid, G. & Patterson, K. E. (2016). Do you read how I read? Systematic individual differences in semantic reliance amongst normal readers. *Frontiers in Psychology*, *7*. 1757.

Yap, M. J. (2007). *Visual Word Recognition: Explorations of Megastudies, multisyllabic words, and individual differences*. .

Yap, M. J. & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, *60*, 502–529.

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*, 53–79.

Yarkoni, T., Balota, D. & Yap, M. J. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*, 971–979.

Zeno, S., Ivens, S., Millard, R. & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science.

Zevin, J. D. & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language*, *47*, 1–29.

Zevin, J. D. & Seidenberg, M. S. (2006). Simulating consistency effects and individual differences in nonword naming: A comparison of current models. *Journal of Memory and Language*, *54*, 145–160.

Ziegler, J. C. & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological Bulletin*, *131*, 3–29.

Ziegler, J. C., Montant, M. & Jacobs, A. M. (1997). The feedback consistency effect in lexical decision and naming. *Journal of Memory and Language*, *37*, 533–554.

Ziegler, J. C., Perry, C. & Coltheart, M. (2000). The DRC model of visual word recognition and reading aloud: An extension to German. *European Journal of Cognitive Psychology*, *12*, 413–430.

Ziegler, J. C., Perry, C., Jacobs, A. M. & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science*, *12*, 379–384.

Ziegler, J. C., Perry, C., Ma-Wyatt, A., Ladner, D. & Schulte-Körne, G. (2003). Developmental dyslexia in different languages: Language-specific or universal? *Journal of*

Experimental Child Psychology, *86*, 169–193.

Ziegler, J. C., Perry, C. & Zorzi, M. (2013). Modelling reading development through phonological decoding and self-teaching: implications for dyslexia. *Philosophical Transactions of The Royal Society B*, *369*(10120397).

Ziegler, J. C., Perry, C. & Zorzi, M. (2014). Modelling reading development through phonological decoding and self-teaching: implications for dyslexia. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120397. The Royal Society.

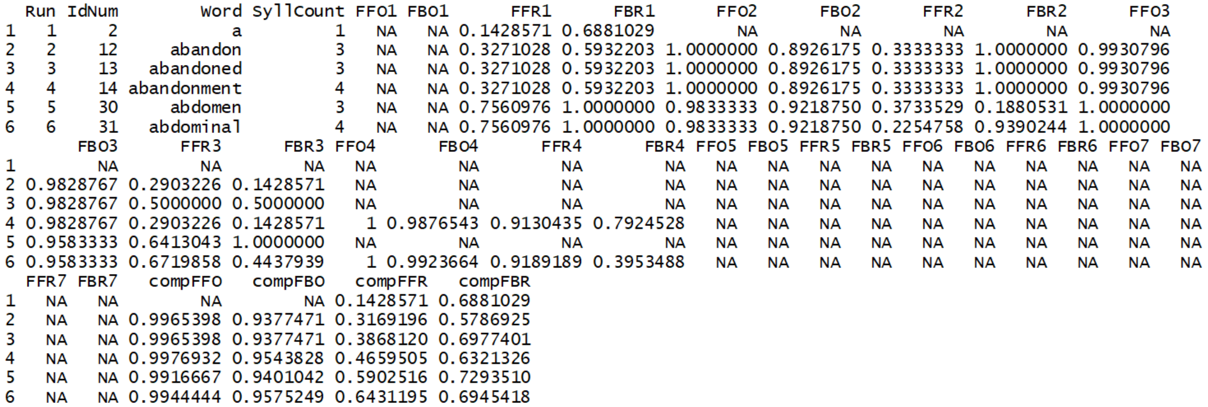
Ziegler, J. C., Stone, G. O. & Jacobs, A. M. (1997). What is the pronunciation for-ough and the spelling for/u/? A database for computing feedforward and feedback consistency in English. *Behavior Research Methods, Instruments, & Computers*, *29*, 600–618.

Zorzi, M., Houghton, G. & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1131–1161.

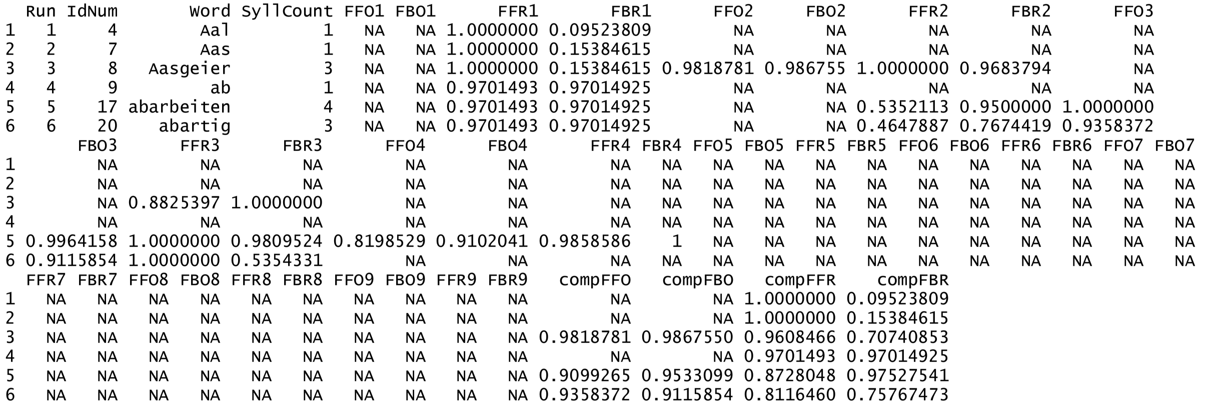
11 Appendices

11.1 Consistency databases

First entries of English (left) and German (right) consistency databases



```
  Run IdNum        Word SyllCount FFO1 FBO1      FFR1      FBR1      FFO2      FBO2      FFR2      FBR2      FFO3
1   1     2           a         1   NA   NA 0.1428571 0.6881029        NA        NA        NA        NA        NA
2   2    12     abandon         3   NA   NA 0.3271028 0.5932203 1.0000000 0.8926175 0.3333333 1.0000000 0.9930796
3   3    13   abandoned         3   NA   NA 0.3271028 0.5932203 1.0000000 0.8926175 0.3333333 1.0000000 0.9930796
4   4    14 abandonment         4   NA   NA 0.3271028 0.5932203 1.0000000 0.8926175 0.3333333 1.0000000 0.9930796
5   5    30     abdomen         3   NA   NA 0.7560976 1.0000000 0.9833333 0.9218750 0.3733529 0.1880531 1.0000000
6   6    31   abdominal         4   NA   NA 0.7560976 1.0000000 0.9833333 0.9218750 0.2254758 0.9390244 1.0000000
       FBO3      FFR3      FBR3 FFO4      FBO4      FFR4      FBR4 FFO5 FBO5 FFR5 FBR5 FFO6 FBO6 FFR6 FBR6 FFO7 FBO7
1        NA        NA        NA   NA        NA        NA        NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
2 0.9828767 0.2903226 0.1428571   NA        NA        NA        NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
3 0.9828767 0.5000000 0.5000000   NA        NA        NA        NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
4 0.9828767 0.2903226 0.1428571    1 0.9876543 0.9130435 0.7924528   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
5 0.9583333 0.6413043 1.0000000   NA        NA        NA        NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
6 0.9583333 0.6719858 0.4437939    1 0.9923664 0.9189189 0.3953488   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
  FFR7 FBR7   compFFO   compFBO   compFFR   compFBR
1   NA   NA        NA        NA 0.1428571 0.6881029
2   NA   NA 0.9965398 0.9377471 0.3169196 0.5786925
3   NA   NA 0.9965398 0.9377471 0.3868120 0.6977401
4   NA   NA 0.9976932 0.9543828 0.4659505 0.6321326
5   NA   NA 0.9916667 0.9401042 0.5902516 0.7293510
6   NA   NA 0.9944444 0.9575249 0.6431195 0.6945418
```



```
  Run IdNum       Word SyllCount FFO1 FBO1      FFR1       FBR1      FFO2     FBO2      FFR2      FBR2      FFO3
1   1     4        Aal         1   NA   NA 1.0000000 0.09523809        NA       NA        NA        NA        NA
2   2     7        Aas         1   NA   NA 1.0000000 0.15384615        NA       NA        NA        NA        NA
3   3     8   Aasgeier         3   NA   NA 1.0000000 0.15384615 0.9818781 0.986755 1.0000000 0.9683794        NA
4   4     9         ab         1   NA   NA 0.9701493 0.97014925        NA       NA        NA        NA        NA
5   5    17 abarbeiten         4   NA   NA 0.9701493 0.97014925        NA       NA 0.5352113 0.9500000 1.0000000
6   6    20    abartig         3   NA   NA 0.9701493 0.97014925        NA       NA 0.4647887 0.7674419 0.9358372
       FBO3      FFR3      FBR3      FFO4      FBO4      FFR4 FBR4 FFO5 FBO5 FFR5 FBR5 FFO6 FBO6 FFR6 FBR6 FFO7 FBO7
1        NA        NA        NA        NA        NA        NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
2        NA        NA        NA        NA        NA        NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
3        NA 0.8825397 1.0000000        NA        NA        NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
4        NA        NA        NA        NA        NA        NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
5 0.9964158 1.0000000 0.9809524 0.8198529 0.9102041 0.9858586    1   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
6 0.9115854 1.0000000 0.5354331        NA        NA        NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
  FFR7 FBR7 FFO8 FBO8 FFR8 FBR8 FFO9 FBO9 FFR9 FBR9   compFFO   compFBO   compFFR    compFBR
1   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA        NA        NA 1.0000000 0.09523809
2   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA        NA        NA 1.0000000 0.15384615
3   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA 0.9818781 0.9867550 0.9608466 0.70740853
4   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA        NA        NA 0.9701493 0.97014925
5   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA 0.9099265 0.9533099 0.8728048 0.97527541
6   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA 0.9358372 0.9115854 0.8116460 0.75767473
```

Note. Run = running number, IdNum = IdNum from Celex database, SyllCount = number of phonological syllables, FF = feedforward, FB = feedback, O = onset, R = rime, comp = composite.

11.2 Most inconsistent entries in consistency databases

German

10 most inconsistent feedforward onsets in syllable 1

```
> GFFO1  <- GCF3[order(GCF3$FFO1),]
> GFFO1[1:50,c("Word","FFO1", "GO1","GR1","GO2", "GR2","GO3","GR3",
+                "GPO1","GPR1", "GPO2", "GPR2", "GPO3", "GPR3")]
              Word        FFO1 GO1   GR1  GO2  GR2  GO3  GR3 GPO1 GPR1 GPO2 GPR2 GPO3 GPR3
6052       Musical 0.002066116   m     u    s    i    c   al   mj    u    z    i    k   @l
10604          Zen 0.003174603   z    en <NA> <NA> <NA> <NA>    z   En <NA> <NA> <NA> <NA>
3561     Gentleman 0.004901961   g entle    m   an <NA> <NA>    _ Entl    m   @n <NA> <NA>
3701           Gin 0.004901961   g    in <NA> <NA> <NA> <NA>    _   In <NA> <NA> <NA> <NA>
8369       Steward 0.005050505  st    ew <NA>  ard <NA> <NA>   st   ju <NA>  @rt <NA> <NA>
8208          Spot 0.008928571  sp    ot <NA> <NA> <NA> <NA>   sp   Ot <NA> <NA> <NA> <NA>
3507         Gelee 0.009803922   g     e    l   ee <NA> <NA>    Z    @    l    e <NA> <NA>
3557         Genie 0.009803922   g     e    n   ie <NA> <NA>    Z    e    n    i <NA> <NA>
3560         Genre 0.009803922   g    en    r    e <NA> <NA>    Z    q    r    @ <NA> <NA>
3700        Gigolo 0.009803922   g     i    g    o    l    o    Z    i    g    o    l    o
```

10 most inconsistent feedback onsets in syllable 1

```
> GFBO1  <- GCF3[order(GCF3$FBO1),]
> GFBO1[1:50,c("Word","FBO1", "GO1","GR1","GO2", "GR2","GO3","GR3",
+               "GPO1","GPR1", "GPO2", "GPR2", "GPO3", "GPR3")]
               Word        FBO1 GO1  GR1  GO2  GR2  GO3  GR3 GPO1 GPR1 GPO2 GPR2 GPO3 GPR3
4816           Khan 0.002217295  kh   an <NA> <NA> <NA> <NA>    k   an <NA> <NA> <NA> <NA>
6903     Quarantäne 0.002217295  qu    a    r   an    t    ä    k    &    r   &n    t    )
3691         Ghetto 0.002481390  gh    e   tt    o <NA> <NA>    g    E    t    o <NA> <NA>
10604           Zen 0.002793296   z   en <NA> <NA> <NA> <NA>    z   En <NA> <NA> <NA> <NA>
1878        Celsius 0.003174603   c   el    s    i <NA>   us    =   El    z    i <NA>   Us
10340       Whiskey 0.004048583  wh   is    k   ey <NA> <NA>    v   Is    k    i <NA> <NA>
10341        Whisky 0.004048583  wh   is    k    y <NA> <NA>    v   Is    k    i <NA> <NA>
7187          Rhein 0.004889976  rh  ein <NA> <NA> <NA> <NA>    r   Wn <NA> <NA> <NA> <NA>
7188     rhetorisch 0.004889976  rh    e    t    o    r isch    r    e    t    o    r   IS
8021            Ski 0.005050505  sk    i <NA> <NA> <NA> <NA>    S    i <NA> <NA> <NA> <NA>
```

10 most inconsistent feedforward rimes in syllable 1

```
> GFFR1  <- GCF3[order(GCF3$FFR1),]
> GFFR1[1:50,c("Word","FFR1","GO1","GR1","GO2", "GR2","GO3","GR3",
+ "GPO1","GPR1", "GPO2", "GPR2", "GPO3", "GPR3")]
            Word        FFR1  GO1 GR1  GO2  GR2  GO3  GR3 GPO1 GPR1 GPO2 GPR2 GPO3 GPR3
1088        Baby 0.001176471    b   a    b    y <NA> <NA>    b    e    b    i <NA> <NA>
5652  Management 0.001176471    m   a    n    a  gem  ent    m    {    n    I   _m  @nt
7943     Service 0.001379310    s  er    v  ice <NA> <NA>    z   |r    v   Is <NA> <NA>
1196       Bebop 0.001544402    b   e    b   op <NA> <NA>    b    i    b   Op <NA> <NA>
2018    Designer 0.001544402    d   e    s   ig    n   er    d    i    z    W    n   @r
1841      Budget 0.002695418    b   u   dg   et <NA> <NA>    b    Y    _    e <NA> <NA>
4633        Jury 0.002695418    j   u    r    y <NA> <NA>    Z    y    r    i <NA> <NA>
7307         Run 0.003105590    r  un <NA> <NA> <NA> <NA>    r   &n <NA> <NA> <NA> <NA>
10530     Yankee 0.003246753    y  an    k   ee <NA> <NA>    j   EN    k    i <NA> <NA>
1319       beige 0.003401361    b  ei    g    e <NA> <NA>    b    e    Z    @ <NA> <NA>
```

10 most inconsistent feedback rimes in syllable 1

```
> GFBR1  <- GCF3[order(GCF3$FBR1),]
> GFBR1[1:50,c("Word","FBR1", "GO1","GR1","GO2", "GR2","GO3","GR3",
+               "GPO1","GPR1", "GPO2", "GPR2", "GPO3", "GPR3")]
              Word        FBR1  GO1  GR1  GO2  GR2  GO3  GR3 GPO1 GPR1 GPO2 GPR2 GPO3 GPR3
5987      Monsieur 0.001620746    m   on    s    i <NA>  eur    m    @    s    i <NA>    |
1088          Baby 0.001876173    b    a    b    y <NA> <NA>    b    e    b    i <NA> <NA>
1319         beige 0.001876173    b   ei    g    e <NA> <NA>    b    e    Z    @ <NA> <NA>
3857      Guerilla 0.001876173    g   ue    r   il    l    a    g    e    r   Il    j    a
8537       Sweater 0.001876173   sw   ea    t   er <NA> <NA>   sv    e    t   @r <NA> <NA>
5760       Meeting 0.002564103    m   ee    t  ing <NA> <NA>    m    i    t   IN <NA> <NA>
7394         Sauce 0.002610966    s   au    c    e <NA> <NA>    z    o    s    @ <NA> <NA>
8247         Squaw 0.002610966  squ   aw <NA> <NA> <NA> <NA>  skv    o <NA> <NA> <NA> <NA>
10680          Zoo 0.002610966    z   oo <NA> <NA> <NA> <NA>    =    o <NA> <NA> <NA> <NA>
3690        Geysir 0.003154574    g   ey    s   ir <NA> <NA>    g    W    z   Ir <NA> <NA>
```

10 most inconsistent feedforward onsets in syllable 2

```
> GFFO2  <- GCF3[order(GCF3$FFO2),]
> GFFO2[1:50,c("Word","FFO2", "GO1","GR1","GO2", "GR2","GO3","GR3",
+                "GPO1","GPR1", "GPO2", "GPR2", "GPO3", "GPR3")]
                Word        FFO2  GO1 GR1 GO2  GR2  GO3  GR3 GPO1 GPR1 GPO2 GPR2 GPO3 GPR3
8650        Teamwork 0.003048780    t eam   w  ork <NA> <NA>    t   im    w   3k <NA> <NA>
1804          Bronze 0.003831418   br  on   z    e <NA> <NA>   br    ~    s    @ <NA> <NA>
1633         Billard 0.004926108    b  il   l  ard <NA> <NA>    b   Il    j  &rt <NA> <NA>
1798        brillant 0.004926108   br  il   l  ant <NA> <NA>   br   Il    j  &nt <NA> <NA>
8575          Taille 0.004926108    t ail   l    e <NA> <NA>    t   &l    j    @ <NA> <NA>
6412       Orchester 0.012195122 <NA>  or  ch   es    t   er <NA>   Or    k   Es    t   @r
8215      Sprechchor 0.012195122  spr ech  ch   or <NA> <NA>  Spr   Ex    k   or <NA> <NA>
1319           beige 0.018121911    b  ei   g    e <NA> <NA>    b    e    Z    @ <NA> <NA>
3344            Gage 0.018121911    g   a   g    e <NA> <NA>    g    a    Z    @ <NA> <NA>
4451       Ingenieur 0.018121911 <NA>  in   g    e    n    i <NA>   In    Z    e    n    i
```

10 most inconsistent feedback onsets in syllable 2

```
> GFBO2  <- GCF3[order(GCF3$FBO2),]
> GFBO2[1:50,c("Word","FBO2", "GO1","GR1","GO2", "GR2","GO3","GR3",
+                "GPO1","GPR1", "GPO2", "GPR2", "GPO3", "GPR3")]
               Word        FBO2  GO1 GR1 GO2  GR2  GO3  GR3 GPO1 GPR1 GPO2 GPR2 GPO3 GPR3
7990     Silhouette 0.001394700    s   i  lh   ou <NA>    e    z    I    l    u <NA>    E
1839         Buddha 0.001855288    b   u ddh    a <NA> <NA>    b    U    d    a <NA> <NA>
6891         Puzzle 0.002036660    p   u zzl    e <NA> <NA>    p    U    z   @l <NA> <NA>
5692        Marquis 0.002475248    m  ar  qu   is <NA> <NA>    m   &r    k    i <NA> <NA>
239     Afghanistan 0.003311258 <NA>  af  gh    a    n   is <NA>   &f    g    a    n   Is
4602        Joghurt 0.003311258    j   o  gh  urt <NA> <NA>    j    o    g  Urt <NA> <NA>
6412      Orchester 0.004950495 <NA>  or  ch   es    t   er <NA>   Or    k   Es    t   @r
8215     Sprechchor 0.004950495  spr ech  ch   or <NA> <NA>  Spr   Ex    k   or <NA> <NA>
1804         Bronze 0.006024096   br  on   z    e <NA> <NA>   br    ~    s    @ <NA> <NA>
675       Arthritis 0.007936508 <NA>  ar thr    i    t   is <NA>   &r   tr    i    t   Is
```

10 most inconsistent feedforward rimes in syllable 2

```
> GFFR2  <- GCF3[order(GCF3$FFR2),]
> GFFR2[1:50,c("Word","FFR2", "GO1","GR1","GO2", "GR2","GO3","GR3",
+                "GPO1","GPR1", "GPO2", "GPR2", "GPO3", "GPR3")]
                   Word         FFR2  GO1   GR1  GO2 GR2  GO3  GR3 GPO1 GPR1 GPO2 GPR2 GPO3 GPR3
6891             Puzzle 0.0007358352    p     u  zzl   e <NA> <NA>    p    U    z   @l <NA> <NA>
3346            Galaxie 0.0015105740    g     a    l   a    x   ie    g    &    l   &k    s    i
8659           Teenager 0.0015105740    t   een <NA>   a    g   er    t   in <NA>    e    _   @r
703               Athen 0.0020120724 <NA>     a   th  en <NA> <NA> <NA>    &    t   en <NA> <NA>
6621            Phosgen 0.0020120724   ph    os    g  en <NA> <NA>    f   Os    g   en <NA> <NA>
4818          kidnappen 0.0030211480    k    id    n   a   pp   en    k   It    n    E    p   @n
4819          Kidnapper 0.0030211480    k    id    n   a   pp   er    k   It    n    E    p   @r
8764           Toilette 0.0032258065    t     o <NA>   i    l    e    t    o <NA>    &    l    E
8765    Toilettenpapier 0.0032258065    t     o <NA>   i    l    e    t    o <NA>    &    l    E
5651            managen 0.0060422961    m     a    n   a    g   en    m    E    n    I    _   @n
```

10 most inconsistent feedback rimes in syllable 2

```
> GFBR2  <- GCF3[order(GCF3$FBR2),]
> GFBR2[1:50,c("Word","FBR2", "GO1","GR1","GO2", "GR2","GO3","GR3",
+                "GPO1","GPR1", "GPO2", "GPR2", "GPO3", "GPR3")]
                 Word        FBR2  GO1   GR1  GO2  GR2  GO3  GR3 GPO1 GPR1 GPO2 GPR2 GPO3 GPR3
5692          Marquis 0.001626016    m    ar   qu   is <NA> <NA>    m   &r    k    i <NA> <NA>
3561        Gentleman 0.002173913    g entle    m   an <NA> <NA>    _ Entl    m   @n <NA> <NA>
8033           Slogan 0.002173913   sl     o    g   an <NA> <NA>   sl    o    g   @n <NA> <NA>
1874             Café 0.002481390    c     a    f    é <NA> <NA>    k    &    f    e <NA> <NA>
1909        Container 0.002481390    c    on    t   ai    n   er    k   On    t    e    n   @r
2850            Essay 0.002481390 <NA>     e   ss   ay <NA> <NA> <NA>    E    s    e <NA> <NA>
4600           Jockey 0.002481390    j     o   ck   ey <NA> <NA>    _    O    k    e <NA> <NA>
8659         Teenager 0.002481390    t   een <NA>    a    g   er    t   in <NA>    e    _   @r
2856             Etat 0.002949853 <NA>     e    t   at <NA> <NA> <NA>    e    t    a <NA> <NA>
1145        Barkeeper 0.003252033    b    ar    k   ee    p   er    b   ar    k    i    p   @r
```

English

10 most inconsistent feedforward onsets in syllable 1

```
> EFFO1[1:100,c("Word","FFO1", "O1","R1","O2", "R2","O3","R3",
+ "PO1","PR1", "PO2", "PR2", "PO3", "PR3")]
                Word        FFO1  O1   R1   O2   R2   O3   R3  PO1 PR1  PO2  PR2  PO3  PR3
1892         cellist 0.002120891   c    e   ll  ist <NA> <NA>    J   E    l  Ist <NA> <NA>
1893           cello 0.002120891   c    e   ll    o <NA> <NA>    J   E    l    5 <NA> <NA>
12169          sugar 0.004043127   s    u    g   ar <NA> <NA>    S   U    g   @R <NA> <NA>
12245           sure 0.004043127   s  ure <NA> <NA> <NA> <NA>    S  $R <NA> <NA> <NA> <NA>
12246         surely 0.004043127   s  ure    l    y <NA> <NA>    S   $    l    I <NA> <NA>
11698          spiel 0.008000000  sp  iel <NA> <NA> <NA> <NA>   Sp  il <NA> <NA> <NA> <NA>
5185        gendarme 0.011070111   g   en    d arme <NA> <NA>    Z  ~n    d   #m <NA> <NA>
5206           genre 0.011070111   g   en    r    e <NA> <NA>    Z   ~    r    @ <NA> <NA>
5244          gigolo 0.011070111   g    i    g    o    l    o    Z   I    g    @    l    5
5751            heir 0.019762846   h  eir <NA> <NA> <NA> <NA> <NA>  8R <NA> <NA> <NA> <NA>
```

10 most inconsistent feedback onsets in syllable 1

```
> EFBO1[1:100,c("Word","FBO1", "O1","R1","O2", "R2","O3","R3",
+ "PO1","PR1", "PO2", "PR2", "PO3", "PR3")]
                 Word        FBO1  O1   R1   O2   R2   O3   R3 PO1 PR1  PO2  PR2  PO3  PR3
6840            khaki 0.001048218  kh    a    k    i <NA> <NA>   k   #    k    I <NA> <NA>
9821            queue 0.002096436  qu  eue <NA> <NA> <NA> <NA>   k  ju <NA> <NA> <NA> <NA>
9823           quiche 0.002096436  qu iche <NA> <NA> <NA> <NA>   k  iS <NA> <NA> <NA> <NA>
7222            llama 0.002232143  ll    a    m    a <NA> <NA>   l   #    m    @ <NA> <NA>
9698      pterodactyl 0.002380952  pt    e    r    o    d   ac   t   E    r    5    d   {k
7831         mnemonic 0.003389831  mn    e    m    o    n   ic   n   i    m    Q    n   Ik
12346           sword 0.004645761  sw  ord <NA> <NA> <NA> <NA>   s  $d <NA> <NA> <NA> <NA>
12347       swordfish 0.004645761  sw  ord    f  ish <NA> <NA>   s  $d    f   IS <NA> <NA>
12348       swordplay 0.004645761  sw  ord   pl   ay <NA> <NA>   s  $d   pl    1 <NA> <NA>
12349       swordsman 0.004645761  sw ords    m   an <NA> <NA>   s $dz    m   @n <NA> <NA>
```

10 most inconsistent feedforward rimes in syllable 1

```
              Word         FFR1   O1 R1   O2    R2   O3   R3  PO1 PR1  PO2  PR2  PO3  PR3
4139         ennui 0.0006215040 <NA>  e   nn    ui <NA> <NA> <NA>   q    n   wi <NA> <NA>
13817     wherever 0.0006215040   wh  e    r     e    v   er    w   8    r    E    v   @R
8078         naive 0.0006675567    n  a <NA>   ive <NA> <NA>    n   2 <NA>   iv <NA> <NA>
2281        cognac 0.0009615385    c  o   gn    ac <NA> <NA>    k  Qn    j   {k <NA> <NA>
13629       voyeur 0.0009615385    v  o    y   eur <NA> <NA>    v  w#    j   3R <NA> <NA>
7196          lira 0.0009950249    l  i    r     a <NA> <NA>    l   7    r    @ <NA> <NA>
7177      lingerie 0.0018867925    l in    g     e    r   ie    l  0n    Z    @    r    i
13370      uranium 0.0019920319 <NA>  u    r     a    n  ium <NA>  jU    r    1    n   7m
8447          only 0.0035714286 <NA> on    l     y <NA> <NA> <NA>  5n    l    I <NA> <NA>
1394         bosom 0.0038461538    b  o    s    om <NA> <NA>    b   U    z   @m <NA> <NA>
```

10 most inconsistent feedback rimes in syllable 1

```
> EFBR1[1:100,c("Word","FBR1", "O1","R1","O2", "R2","O3","R3",
+ "PO1","PR1", "PO2", "PR2", "PO3", "PR3")]
               Word         FBR1   O1   R1   O2   R2   O3   R3  PO1 PR1  PO2  PR2  PO3  PR3
1490       breeches 0.0006858711   br   ee   ch   es <NA> <NA>   br   I    J   Iz <NA> <NA>
835            awry 0.0012106538 <NA>   aw    r    y <NA> <NA> <NA>   @    r    2 <NA> <NA>
5509      guerrilla 0.0012106538    g   ue   rr    i   ll    a    g   @    r    I    l    @
1650           busy 0.0013717421    b    u    s    y <NA> <NA>    b   I    z    I <NA> <NA>
1651       busybody 0.0013717421    b    u    s    y    b    o    b   I    z    I    b    Q
1202     blackguard 0.0014577259   bl  ack    g uard <NA> <NA>   bl   {    g   #d <NA> <NA>
5503      guarantee 0.0014577259    g   ua    r   an    t   ee    g   {    r   @n    t    i
10688        salmon 0.0014577259    s   al    m   on <NA> <NA>    s   {    m   @n <NA> <NA>
13556      vineyard 0.0018518519    v  ine    y  ard <NA> <NA>    v  In    j   @d <NA> <NA>
13897      windmill 0.0018518519    w  ind    m  ill <NA> <NA>    w  In    m   Il <NA> <NA>
```

10 most inconsistent feedforward onsets in syllable 2

```
> EFFO2  <- ECF3[order(ECF3$FFO2),]
> EFFO2[1:100,c("Word","FFO2", "O1","R1","O2", "R2","O3","R3",
+ "PO1","PR1", "PO2", "PR2", "PO3", "PR3")]
               Word        FFO2   O1   R1 O2     R2   O3    R3  PO1 PR1  PO2  PR2  PO3  PR3
10973      senorita 0.001706485    s    e  n      o    r     i    s   E   nj    $    r    i
8104        natural 0.001892148    n    a  t   ural <NA>  <NA>    n   {   Jr   @l <NA> <NA>
8105      naturally 0.001892148    n    a  t    ura   ll     y    n   {   Jr    @    l    I
8965     personally 0.001934236    p   er  s    ona   ll     y    p   3   sn    @    l    I
4943       fortress 0.002132196    f  ort  r    ess <NA>  <NA>    f   $   tr   Is <NA> <NA>
10089     reference 0.003134796    r    e  f erence <NA>  <NA>    r   E   fr  @ns <NA> <NA>
4310     excavation 0.003577818 <NA>   ex  c      a    v     a <NA>  Ek   sk    @    v    1
4337      excursion 0.003577818 <NA>   ex  c     ur    s   ion <NA>  Ik   sk    3    S    H
6569       interest 0.004730369 <NA>   in  t  erest <NA>  <NA> <NA>  In   tr  @st <NA> <NA>
6570     interested 0.004730369 <NA>   in  t    ere   st    ed <NA>  In   tr    @   st   Id
```

10 most inconsistent feedback onsets in syllable 2

```
> EFBO2  <- ECF3[order(ECF3$FBO2),]
> EFBO2[1:100,c("Word","FBO2", "O1","R1","O2", "R2","O3","R3",
+ "PO1","PR1", "PO2", "PR2", "PO3", "PR3")]
               Word         FBO2   O1  R1   O2   R2   O3   R3  PO1 PR1 PO2 PR2  PO3  PR3
12602       theatre 0.0009478673   th  ea   tr    e <NA> <NA>    T   7   t  @R <NA> <NA>
11247    silhouette 0.0010427529    s   i   lh   ou <NA> ette    s   I   l   u <NA>   Et
478          answer 0.0013531800 <NA>  an   sw   er <NA> <NA> <NA>  #n   s  @R <NA> <NA>
4353     exhibition 0.0013531800 <NA>  ex    h    i    b    i <NA>  Ek   s   I    b    I
6925        lacquer 0.0015503876    l   a  cqu   er <NA> <NA>    l   {   k  @R <NA> <NA>
14084      zucchini 0.0015503876    z   u  cch    i    n    i    z   U   k   i    n    I
3058        damning 0.0015600624    d   a   mn  ing <NA> <NA>    d   {   m  IN <NA> <NA>
2109      cirrhosis 0.0017152659    c   i  rrh    o    s   is    s   I   r   5    s   Is
8155         nephew 0.0020920502    n   e   ph   ew <NA> <NA>    n   E   v  ju <NA> <NA>
10724      sapphire 0.0022573363    s   a  pph    i    r    e    s   {   f   2 <NA>   @R
```

10 most inconsistent feedforward rimes in syllable 2

```
> EFFR2  <- ECF3[order(ECF3$FFR2),]
> EFFR2[1:100,c("Word","FFR2", "O1","R1","O2", "R2","O3","R3",
+ "PO1","PR1", "PO2", "PR2", "PO3", "PR3")]
                  Word        FFR2   O1  R1   O2  R2   O3   R3  PO1 PR1  PO2 PR2  PO3  PR3
4964             foyer 0.001136364    f  oy <NA>  er <NA> <NA>    f   4 <NA>   1 <NA> <NA>
4698           firearm 0.001297017    f   i    r   e <NA>  arm    f   2 <NA>  @r <NA>   #m
8722            papaya 0.001324503    p   a    p   a <NA>   ya    p   @    p   2 <NA>   I@
6919         labyrinth 0.001647446    l   a    b   y    r inth    l   {    b   @    r  InT
7378          madwoman 0.002928258    m  ad    w   o    m   an    m  {d    w   U    m   @n
13324             unto 0.002928258 <NA>  un    t   o <NA> <NA> <NA>  Vn    t   U <NA> <NA>
7357         luxurious 0.003690037    l  ux <NA>   u    r ious    l  Vg    Z   9    r   7s
1686              cafe 0.003891051    c   a    f   e <NA> <NA>    k   {    f   1 <NA> <NA>
8804              pate 0.003891051    p   a    t   e <NA> <NA>    p   {    t   1 <NA> <NA>
10534            rodeo 0.003891051    r   o    d   e <NA>    o    r   5    d   1 <NA>    5
```

10 most inconsistent feedback rimes in syllable 2

```
> EFBR2  <- ECF3[order(ECF3$FBR2),]
> EFBR2[1:100,c("Word","FBR2", "O1","R1","O2", "R2","O3","R3",
+ "PO1","PR1", "PO2", "PR2", "PO3", "PR3")]
               Word         FBR2   O1   R1   O2     R2   O3   R3  PO1 PR1  PO2 PR2  PO3  PR3
1974        chassis 0.0005241090   ch    a   ss     is <NA> <NA>    S   {    s   I <NA> <NA>
5521         guinea 0.0005241090    g   ui    n     ea <NA> <NA>    g   I    n   I <NA> <NA>
10235    rendezvous 0.0005241090    r   en    d     ez    v  ous    r  Qn    d   I    v    u
2547    connoisseur 0.0007374631    c    o   nn     oi   ss  eur    k   Q    n   @    s   3R
4896      foreigner 0.0007374631    f    o    r    eig    n   er    f   Q    r   @    n   @R
6919      labyrinth 0.0007374631    l    a    b      y    r inth    l   {    b   @    r  InT
8062          mynah 0.0007374631    m    y    n     ah <NA> <NA>    m   2    n   @ <NA> <NA>
8105      naturally 0.0007374631    n    a    t    ura   ll    y    n   {   Jr   @    l    I
8965     personally 0.0007374631    p   er    s    ona   ll    y    p   3   sn   @    l    I
10358    restaurant 0.0007374631    r    e   st     au    r  ant    r   E   st   @    r   ~N
```

11.3 Words with most inconsistent composite FF and FB consistency

German

10 most inconsistent composite FF onsets

```
> GcompFFO[1:25,c("Word", "compFFO", "FFO1", "FFO2","FFO3","FFO4")]
             Word     compFFO        FFO1       FFO2       FFO3 FFO4
10604         Zen 0.003174603 0.003174603         NA         NA   NA
3701          Gin 0.004901961 0.004901961         NA         NA   NA
8369      Steward 0.005050505 0.005050505         NA         NA   NA
8208         Spot 0.008928571 0.008928571         NA         NA   NA
266        Aktion 0.019813520          NA 0.01981352         NA   NA
877       Auktion 0.019813520          NA 0.01981352         NA   NA
6407       Option 0.019813520          NA 0.01981352         NA   NA
215         Achse 0.029702970          NA 0.02970297         NA   NA
2264        Echse 0.029702970          NA 0.02970297         NA   NA
6354        Ochse 0.029702970          NA 0.02970297         NA   NA
```

10 most inconsistent composite FB onsets

```
> GcompFBO  <- GCF3[order(GCF3$compFBO),]
> GcompFBO[1:25,c("Word", "compFBO", "FBO1", "FBO2","FBO3","FBO4")]
           Word     compFBO        FBO1       FBO2 FBO3 FBO4
4816       Khan 0.002217295 0.002217295         NA   NA   NA
10604       Zen 0.002793296 0.002793296         NA   NA   NA
7187      Rhein 0.004889976 0.004889976         NA   NA   NA
8021        Ski 0.005050505 0.005050505         NA   NA   NA
8722      Thron 0.008403361 0.008403361         NA   NA   NA
1912      Creme 0.010869565 0.010869565         NA   NA   NA
261      Akkord 0.012376238          NA 0.01237624   NA   NA
1880      Chaos 0.013303769 0.013303769         NA   NA   NA
1895       Chor 0.013303769 0.013303769         NA   NA   NA
2256       Ebbe 0.015100671          NA 0.01510067   NA   NA
```

10 most inconsistent composite FF rimes

```
> GcompFFR  <- GCF3[order(GCF3$compFFR),]
> GcompFFR[1:25,c("Word", "compFFR", "FFR1", "FFR2","FFR3","FFR4")]
           Word     compFFR        FFR1       FFR2       FFR3 FFR4
7307        Run 0.003105590 0.003105590         NA         NA   NA
2950        Fan 0.006493506 0.006493506         NA         NA   NA
8922        tun 0.009316770 0.009316770         NA         NA   NA
1902       Clan 0.019480519 0.019480519         NA         NA   NA
4816       Khan 0.019480519 0.019480519         NA         NA   NA
5130       Kran 0.019480519 0.019480519         NA         NA   NA
6655       Plan 0.019480519 0.019480519         NA         NA   NA
7785     Schwan 0.019480519 0.019480519         NA         NA   NA
7373   Sandwich 0.028119869 0.047619048 0.00862069         NA   NA
3772       Grab 0.029850746 0.029850746         NA         NA   NA
```

10 most inconsistent composite FB rimes

```
> GcompFBR  <- GCF3[order(GCF3$compFBR),]
> GcompFBR[1:25,c("Word", "compFBR", "FBR1", "FBR2","FBR3","FBR4")]
          Word     compFBR        FBR1        FBR2 FBR3 FBR4
8247     Squaw 0.002610966 0.002610966          NA   NA   NA
10680      Zoo 0.002610966 0.002610966          NA   NA   NA
7307       Run 0.003584229 0.003584229          NA   NA   NA
8212     Spray 0.003752345 0.003752345          NA   NA   NA
6938      Rain 0.005235602 0.005235602          NA   NA   NA
7959      Show 0.007832898 0.007832898          NA   NA   NA
5091     Korps 0.007936508 0.007936508          NA   NA   NA
5992      Moor 0.007936508 0.007936508          NA   NA   NA
9884      Vieh 0.010256410 0.010256410          NA   NA   NA
1088      Baby 0.011507192 0.001876173 0.021138211   NA   NA
```

English

10 most inconsistent composite FF onsets

```
> EcompFFO[1:50,c("Word", "compFFO", "FFO1", "FBO1", "FFR1", "FBR1", "FFO2", "FBO2",
+ "FFR2", "FBR2", "FFO3", "FBO3", "FFR3","FBR3")]
           Word     compFFO        FFO1        FBO1       FFR1        FBR1        FFO2        FBO2       FFR2        FBR2
12245      sure 0.004043127 0.004043127 0.017964072 0.14285714 0.033333333          NA          NA         NA          NA
6569   interest 0.004730369          NA          NA 0.90000000 0.883333333 0.004730369 0.033557047 1.00000000 0.250000000
6573  interests 0.004730369          NA          NA 0.90000000 0.883333333 0.004730369 0.033557047 1.00000000 1.000000000
11698     spiel 0.008000000 0.008000000 1.000000000 1.00000000 0.130434783          NA          NA         NA          NA
8539        our 0.010660981          NA          NA 0.39062500 0.357142857 0.010660981 1.000000000         NA          NA
5965       hour 0.015211913 0.019762846          NA 0.39062500 0.357142857 0.010660981 1.000000000         NA          NA
4415     extend 0.016083254          NA          NA 0.47239264 0.587786260 0.016083254 0.060070671 0.94444444 0.809523810
4419     extent 0.016083254          NA          NA 0.47239264 0.587786260 0.016083254 0.060070671 0.31168831 1.000000000
4424    extinct 0.016083254          NA          NA 0.47239264 0.587786260 0.016083254 0.060070671 1.00000000 1.000000000
4146     ensure 0.017408124          NA          NA 0.29670330 0.100000000 0.017408124 0.042452830 0.05172414 0.107142857
```

10 most inconsistent composite FB onsets

```
> EcompFBO[1:50,c("Word", "compFBO","FFO1", "FBO1", "FFR1", "FBR1", "FFO2", "FBO2",
+ "FFR2", "FBR2", "FFO3", "FBO3", "FFR3","FBR3")]
            Word     compFBO        FFO1        FBO1       FFR1        FBR1        FFO2        FBO2       FFR2       FBR2 FFO3
478       answer 0.001353180          NA          NA 0.03317536 0.636363636 1.000000000 0.001353180 0.62272727 0.82282282   NA
9821       queue 0.002096436 0.034482759 0.002096436 1.00000000 0.006711409          NA          NA         NA         NA   NA
9823      quiche 0.002096436 0.034482759 0.002096436 0.50000000 0.500000000          NA          NA         NA         NA   NA
116         acre 0.003100775          NA          NA 0.14285714 0.688102894 0.040000000 0.003100775 0.01297017 0.01501502   NA
13134 unanswered 0.003194888          NA          NA 0.93436293 0.937984496          NA          NA 0.08333333 1.00000000  0.5
8425        ogre 0.004291845          NA          NA 0.30673077 0.791563275 0.018518519 0.004291845 0.01297017 0.01501502   NA
13271    unknown 0.004491018          NA          NA 0.93436293 0.937984496 1.000000000 0.004491018 0.12500000 0.08333333   NA
8423       often 0.004514673          NA          NA 0.40480769 0.935555556 1.000000000 0.004514673 0.26890756 0.43537415   NA
12346      sword 0.004645761 0.064516129 0.004645761 0.87500000 0.350000000          NA          NA         NA         NA   NA
13100        two 0.004761905 0.100000000 0.004761905 0.01442308 0.080645161          NA          NA         NA         NA   NA
```

10 most inconsistent composite FF rimes

```
> EcompFFR  <- EC[order(EC$compFFR),]
> EcompFFR[1:50,c("Word", "compFFR", "FFO1", "FBO1", "FFR1", "FBR1", "FFO2", "FBO2",
+ "FFR2", "FBR2", "FFO3", "FBO3", "FFR3","FBR3")]
       Word     compFFR      FFO1        FBO1         FFR1        FBR1        FFO2       FBO2        FFR2        FBR2 FFO3
13679   war 0.005102041 1.0000000 0.840455840 0.0051020408 0.033333333          NA         NA          NA          NA   NA
6884   kiwi 0.011256169 1.0000000 0.078616352 0.0129353234 0.031100478 1.000000000 0.95454545 0.009577015 0.100840336   NA
8133    nee 0.012658228 1.0000000 0.888135593 0.0126582278 0.003215434          NA         NA          NA          NA   NA
11327   ski 0.012935323 1.0000000 0.327102804 0.0129353234 0.031100478          NA         NA          NA          NA   NA
3687     do 0.014423077 1.0000000 1.000000000 0.0144230769 0.080645161          NA         NA          NA          NA   NA
12761    to 0.014423077 1.0000000 0.992857143 0.0144230769 0.080645161          NA         NA          NA          NA   NA
13100   two 0.014423077 0.1000000 0.004761905 0.0144230769 0.080645161          NA         NA          NA          NA   NA
13845   who 0.014423077 0.1764706 0.023622047 0.0144230769 0.080645161          NA         NA          NA          NA   NA
3163  decor 0.015106299 1.0000000 1.000000000 0.0043505283 0.022508039 0.504472272 0.43720930 0.025862069 0.107142857   NA
5772    her 0.016260163 0.9802372 0.976377953 0.0162601626 0.166666667          NA         NA          NA          NA   NA
```

10 most inconsistent composite FB rimes

```
> EcompFBR  <- EC[order(EC$compFBR),]
> EcompFBR[1:50,c("Word", "compFBR", "FFO1", "FBO1", "FFR1", "FBR1", "FFO2", "FBO2",
+ "FFR2", "FBR2", "FFO3", "FBO3", "FFR3","FBR3")]
           Word     compFBR       FFO1        FBO1        FFR1        FBR1      FFO2      FBO2       FFR2        FBR2 FFO3 FBO3
5521     guinea 0.001290861 0.77490775 0.976744186 0.250000000 0.002057613 0.9982935 0.8757485 0.02631579 0.000524109   NA   NA
8643        owe 0.002481390         NA          NA 1.000000000 0.002481390        NA        NA         NA          NA   NA   NA
3986         eh 0.003215434         NA          NA 1.000000000 0.003215434        NA        NA         NA          NA   NA   NA
8133        nee 0.003215434 1.00000000 0.888135593 0.012658228 0.003215434        NA        NA         NA          NA   NA   NA
12884     trait 0.003215434 1.00000000 1.000000000 0.200000000 0.003215434        NA        NA         NA          NA   NA   NA
5991        huh 0.003333333 0.98023715 0.976377953 1.000000000 0.003333333        NA        NA         NA          NA   NA   NA
6443        inn 0.003703704         NA          NA 1.000000000 0.003703704        NA        NA         NA          NA   NA   NA
1982  chauffeur 0.003982891 0.11038961 0.101796407 0.031250000 0.004962779 0.9620253 0.1715576 0.40000000 0.003003003   NA   NA
11041       sew 0.004962779 0.99595687 0.858304297 0.068965517 0.004962779        NA        NA         NA          NA   NA   NA
2784       coup 0.005376344 0.90562036 0.895178197 0.250000000 0.005376344        NA        NA         NA          NA   NA   NA
```

11.4 Demographics questionnaires in English and German

Dear participant, please answer some questions about yourself.

1. Today's date __________________________
2. Please indicate your gender:
   - ☐ Female
   - ☐ Male
3. Are you a student?
   - ☐ Yes, at ____ (name of institution)
   - ☐ No
4. Are you a
   - ☐ Full-time student
   - ☐ Part-time student
   - ☐ Neither, my situation is a different one: _____________________________________
     ____
5. Counting all your years in Higher Education, in which year of study are you in?
   ____
6. What is your age (in years)? _______________
7. Have you been raised multi-lingually (second or third mother tongue)?

☐ No

☐ Yes, I have been raised with __________ more languages

Lieber Teilnehmer, bitte beantworten Sie ein paar Fragen über Ihre Person.

1. Heutiges Datum. __________________________
2. Bitte geben Sie Ihr Geschlecht an:
   - ☐ weiblich
   - ☐ männlich
3. Sind Sie Student?
   - ☐ Ja, an ____ (Name der Institution)
   - ☐ Nein
4. Sind Sie
   - ☐ Vollzeitstudent
   - ☐ Teilzeitstudent
   - ☐ Weder noch, meine Situation ist folgende: ___________________________________
     ____
5. Seit wievielen Jahre studieren Sie? ____
6. Bitte geben Sie Ihr Alter (in Jahren) an? _______________
7. Sind Sie mehrsprachig aufgewachsen (zweite oder dritte Muttersprache)?

☐ Nein

☐ Ja, ich bin mit _______________________ anderen Sprachen aufgewachsen.

11.5 GE-ART (English Version 1)

Below you will see a list of 100 names. Some of the people in the list are popular writers and some are not. Please read the names and tick the names that belong to real authors (box 1). After you have identified an author as being real, please also indicate if you have heard of the author, but not read any of the author's work (box 2); if you have started a book by the author, but not finished it (box 3); or if you have actually read at least one book by the author (box 4). Please do not guess. Some of the names do not belong to a popular writer, so guessing can easily be detected.

	BOX 1	**BOX 2**	**BOX 3**	**BOX 4**
NAMES	recognise name	heard of, but have not read anything	started a book, but have not finished it	have read at least one book
Theobald Wright				
Olive Silver				
Ron Ryder Wilde				
Paul Salm				
J.R.R. Tolkien				
Ana Acebo				
Anabel Floyd				
Judith Kerr				
Rachel Joyce				
Guillaume Laziere				
R.G. Rothkin				
Jacques Tebeau				
Felicitas Gummersberg				
Suzanne Collins				
Rüdiger Hofmeister				
Esmeralda Hope				
Khaled Hosseini				
Graeme Simsion				
Pietro Castaldi				
Rajesh Parameswaran				
Herman Hesse				
Ada Meers				
Stieg Olhouser				
Judith Hermann				
Cornelia Funke				
J. K. Rowling				
Mansana Biblington				
S. E. Spellmeyer				
Rose Hammersleigh				
Maurice Sendak				

	BOX 1	**BOX 2**	**BOX 3**	**BOX 4**
NAMES	recognise name	heard of, but have not read anything	started a book, but have not finished it	have read at least one book
Guiliano Vanucci				
Thomas L. Friedman				
Annabelle Casey				
Håkan Nesser				
Wolfgang Herrndorf				
Timothy Snyder				
Lorenzo di Antonio				
M. C. Smith				
Susan Elizabeth Phillips				
Rico Marcos da Silva				
Gunther Kramer				
Enid Blyton				
Nicholas Sparks				
Axl Nordahl				
Jed Goodman				
Robert Ludlum				
Douglas Adams				
Francis D. Porter				
H.P. Lovecraft				
Olamide Addae				
Ian Kershaw				
Katharina Alster				
Jenny Svensson-Marshall				
Johanna Spyri				
Ferdinand Högent				
Ruth Nickels				
Jojo Moyes				
Orhan Pamuk				
Flora Jenkins				
Jeff Cain				
Bill Bryson				
Roger Ferrand				
Serra Badem				
Daniel Kahnemann				
Dan Brown				

	BOX 1	**BOX 2**	**BOX 3**	**BOX 4**
NAMES	recognise name	heard of, but have not read anything	started a book, but have not finished it	have read at least one book
James Bowen				
Dr William Davis				
Lucas Seeler				
Jung Chang				
Terry Pratchett				
Zelda Zavon				
Jana Hanson				
Ken Follett				
Ian McEwan				
Gudrun Färber				
John Green				
Ajeet Bhatnagar				
Mohammed Hanif				
Elias Canetti				
Anita Sneller				
Atif Mian				
Nisse Stangeland				
Jaron Lanier				
Floris Keat				
Rolf Dobelli				
John Banville				
Frederick Marsting				
Markus Zusak				
Dominique Laurent				
Mark Haddon				
Bernhard Schlink				
Stephenie Meyer				
Donna Tartt				
Cassandra Clare				
Liu Min				
Jonas Jonasson				
Frank Redmoor				
Tim Strain				
Jeff Kinney				
Ursula Goldstein				

11.6 Authors recognised and authors read by the English language group

English language group					
Authors correctly recognised	N	%	Authors having been read	N	%
J K Rowling	104	100.00	J K Rowling	79	75.96
J R R Tolkien	83	79.81	Stephenie Meyer	49	47.12
Terry Pratchett	79	75.96	Enid Blyton	45	43.27
Stephenie Meyer	78	75.00	John Green	39	37.50
Enid Blyton	71	68.27	J R R Tolkien	28	26.92
John Green	67	64.42	Suzanne Collins	28	26.92
Dan Brown	66	63.46	Terry Pratchett	27	25.96
Ian McEwan	58	55.77	Nicholas Sparks	23	22.12
Bill Bryson	54	51.92	Khaled Hosseini	19	18.27
Nicholas Sparks	53	50.96	Bill Bryson	17	16.35
Suzanne Collins	50	48.08	Dan Brown	15	14.42
Judith Kerr	38	36.54	Ian McEwan	13	12.50
Douglas Adams	34	32.69	Cornelia Funke	12	11.54
H P Lovecraft	33	31.73	Douglas Adams	11	10.58
Khaled Hosseini	30	28.85	Jojo Moyes	10	9.62
Jojo Moyes	24	23.08	Cassandra Clare	9	8.65
Cornelia Funke	20	19.23	Markus Zusak	9	8.65
Ian Kershaw	19	18.27	Mark Haddon	8	7.69
Mark Haddon	18	17.31	Jonas Jonasson	5	4.81
Cassandra Clare	17	16.35	Maurice Sendak	5	4.81
Markus Zusak	14	13.46	Judith Kerr	4	3.85
Donna Tartt	12	11.54	H P Lovecraft	3	2.88
Robert Ludlum	12	11.54	Donna Tartt	3	2.88
Jonas Jonasson	11	10.58	Robert Ludlum	3	2.88
Judith Hermann	11	10.58	Ian Kershaw	2	1.92
Hermann Hesse	10	9.62	Hermann Hesse	2	1.92
Ken Follett	10	9.62	Johanna Spyri	2	1.92
James Bowen	9	8.65	Orhan Pamuk	2	1.92
Thomas L Friedman	9	8.65	Daniel Kahnemann	2	1.92
Rachel Joyce	8	7.69	Rachel Joyce	1	0.96
John Banville	7	6.73	John Banville	1	0.96
Maurice Sendak	7	6.73	Susan Elizabeth Phillips	1	0.96
Jeff Kinney	5	4.81	Bernhard Schlink	1	0.96
Johanna Spyri	4	3.85	Judith Hermann	0	0.00
Orhan Pamuk	4	3.85	Ken Follett	0	0.00
Daniel Kahnemann	3	2.88	James Bowen	0	0.00
Håkan Nesser	3	2.88	Thomas L Friedman	0	0.00
Jung Chang	3	2.88	Jeff Kinney	0	0.00
Rajesh Parameswaran	3	2.88	Håkan Nesser	0	0.00
Wolfgang Herrndorf	3	2.88	Jung Chang	0	0.00
Graeme Simsion	2	1.92	Rajesh Parameswaran	0	0.00
Mohammed Hanif	2	1.92	Wolfgang Herrndorf	0	0.00
Susan Elizabeth Phillips	2	1.92	Graeme Simsion	0	0.00
Atif Mian	1	0.96	Mohammed Hanif	0	0.00
Bernhard Schlink	1	0.96	Atif Mian	0	0.00
Jaron Lanier	1	0.96	Jaron Lanier	0	0.00
Dr William Davis	0	0.00	Dr William Davis	0	0.00
Elias Canetti	0	0.00	Elias Canetti	0	0.00
Olamide Addae	0	0.00	Olamide Addae	0	0.00
Rolf Dobelli	0	0.00	Rolf Dobelli	0	0.00

11.7 Authors recognised and authors read by the German language group

German language group

Authors correctly recognised	N	%	Authors having been read	N	%
J. K. Rowling	100	96.15	J. K. Rowling	72	69.23
Cornelia Funke	97	93.27	Cornelia Funke	67	64.42
Hermann Hesse	84	80.77	Hermann Hesse	50	48.08
J. R. R. Tolkien	83	79.81	J. R. R. Tolkien	39	37.50
Ken Follett	83	79.81	Ken Follett	32	30.77
Dan Brown	76	73.08	Dan Brown	30	28.85
Nicholas Sparks	74	71.15	Nicholas Sparks	27	25.96
Stephenie Meyer	72	69.23	Stephenie Meyer	27	25.96
Enid Blyton	51	49.04	Enid Blyton	22	21.15
John Green	46	44.23	John Green	21	20.19
Jojo Moyes	46	44.23	Jojo Moyes	18	17.31
Terry Pratchett	38	36.54	Terry Pratchett	18	17.31
Bernhard Schlink	35	33.65	Bernhard Schlink	15	14.42
Suzanne Collins	34	32.69	Suzanne Collins	9	8.65
H. P. Lovecraft	24	23.08	H. P. Lovecraft	9	8.65
Håkan Nesser	24	23.08	Håkan Nesser	8	7.69
Markus Zusak	23	22.12	Markus Zusak	8	7.69
Judith Kerr	22	21.15	Judith Kerr	8	7.69
Ian McEwan	21	20.19	Ian McEwan	7	6.73
Bill Bryson	19	18.27	Bill Bryson	7	6.73
Jonas Jonasson	19	18.27	Jonas Jonasson	6	5.77
Khaled Hosseini	17	16.35	Khaled Hosseini	4	3.85
Douglas Adams	15	14.42	Douglas Adams	4	3.85
James Bowen	15	14.42	James Bowen	3	2.88
Wolfgang Herrndorf	15	14.42	Wolfgang Herrndorf	3	2.88
Susan Elizabeth Phillips	14	13.46	Susan Elizabeth Phillips	2	1.92
Rachel Joyce	12	11.54	Rachel Joyce	2	1.92
Cassandra Clare	11	10.58	Cassandra Clare	1	0.96
Ian Kershaw	8	7.69	Ian Kershaw	1	0.96
Johanna Spyri	6	5.77	Johanna Spyri	1	0.96
Jeff Kinney	5	4.81	Jeff Kinney	1	0.96
Mark Haddon	4	3.85	Mark Haddon	1	0.96
Donna Tartt	4	3.85	Donna Tartt	0	0.00
Judith Hermann	4	3.85	Judith Hermann	0	0.00
Thomas L Friedman	4	3.85	Thomas L Friedman	0	0.00
Robert Ludlum	3	2.88	Robert Ludlum	0	0.00
Orhan Pamuk	3	2.88	Orhan Pamuk	0	0.00
Daniel Kahnemann	3	2.88	Daniel Kahnemann	0	0.00
John Banville	3	2.88	John Banville	0	0.00
Mohammed Hanif	3	2.88	Mohammed Hanif	0	0.00
Maurice Sendak	2	1.92	Maurice Sendak	0	0.00
Graeme Simsion	2	1.92	Graeme Simsion	0	0.00
Dr. William Davis	1	0.96	Dr. William Davis	0	0.00
Elias Canetti	1	0.96	Elias Canetti	0	0.00
Rolf Dobelli	1	0.96	Rolf Dobelli	0	0.00
Jung Chang	0	0.00	Jung Chang	0	0.00
Rajesh Parameswaran	0	0.00	Rajesh Parameswaran	0	0.00
Atif Mian	0	0.00	Atif Mian	0	0.00
Jaron Lanier	0	0.00	Jaron Lanier	0	0.00
Olamide Addae	0	0.00	Olamide Addae	0	0.00

Note . Following German writing conventions, initials and abbreviations like 'Dr' were followed by a full stop.

11.8 Reading Questionnaire (English version)

Please fill in this reading questionnaire by circling the alternative that is most accurate.
Please be as *honest* in your response as possible.

1. I read for pleasure
    a. almost never
    b. a couple of times a year
    c. a couple of times a month
    d. at least once a week
    e. once or more a day

2. *Not* including textbooks for your University courses, how many books do you read in a year?
    a. none
    b. one or two
    c. 3-10
    d. 10-40
    e. more than 40

3. *Excluding* the University library, which of the following is true?
    a. I have a library card for the local library.
    b. I do not have a library card.
    c. I have library cards for more than one library.

4. I download free e-books
    a. almost never
    b. a couple of times a year
    c. a couple of times a month
    d. at least once a week
    e. several times a week

5. I visit bookshops
    a. several times a week
    b. at least once a week
    c. a couple of times a month
    d. a couple of times a year
    e. almost never

6. I buy books and/or e-books online
    a. almost never
    b. a couple of times a year
    c. a couple of times a month
    d. at least once a week
    e. several times a week

7. If you answered b, c, d or e to Question 2, please name your three favourite authors/writers. If there are less than three, please list however many there are.

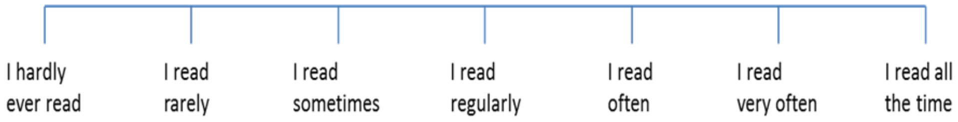
Please turn the page.

/page 2

Please continue to fill in this reading questionnaire by circling the alternative that is most accurate. Please be as *honest* in your response as possible.

8. How often do you read newspapers (online and/or paper versions)?
    a. I do not read newspapers.
    b. I rarely read a newspaper.
    c. I read newspapers occasionally.
    d. I read a newspaper every day.
    e. I read more than one newspaper a day.

9. Generally, taking all your reading together, how much do you read? Please indicate on the 7-point scale below, ranging from '*I hardly ever read*' to '*I read all the time*'.



10. Do you have a book with you at present that you are in the process of reading for pleasure? Can you indicate the author's name?

☐ no ☐ yes author's name: ______________________

Thank you very much for your participation.

11.9 Reading Questionnaire (German version)

Lesefragebogen

Bitte kreisen Sie die Antwort ein, die am ehesten auf Sie zutrifft.
Bitte beantworten Sie die Fragen so *ehrlich* wie möglich.

1. Ich lese zum Vergnügen
    a. fast nie
    b. ein paar Mal im Jahr
    c. ein paar Mal im Monat
    d. mindestens einmal pro Woche
    e. ein oder mehrmals am Tag

2. Wieviele Bücher lesen Sie im Jahr, wenn Sie Lehrbücher *nicht* mitzählen?
    a. keins
    b. eins oder zwei
    c. 3-10
    d. 10-40
    e. mehr als 40

3. Von der Universitätsbibliothek *abgesehen* , welche der folgenden Aussagen trifft auf Sie zu?
    a. Ich habe einen Ausweis für die örtliche Bibliothek.
    b. Ich habe keinen Bibliotheksausweis.
    c. Ich habe Ausweise für mehr als eine Bibliothek.

4. Ich lade E-Bücher herunter, die kostenlos sind.
    a. fast nie
    b. ein paar Mal im Jahr
    c. ein paar Mal im Monat
    d. mindestens einmal pro Woche
    e. mehrmals pro Woche

5. Ich gehe in Buchläden
    a. mehrmals pro Woche
    b. mindestens einmal pro Woche
    c. ein paar Mal im Monat
    d. ein paar Mal im Jahr
    e. fast nie

6. Ich kaufe Bücher online
    a. fast nie
    b. ein paar Mal im Jahr
    c. ein paar Mal im Monat
    d. mindestens einmal pro Woche
    e. mehrmals pro Woche

7. Wenn Sie auf Frage 2 mit b, c, d oder e geantwortet haben, nennen Sie bitte Ihre drei Lieblingsschriftsteller. Wenn Sie weniger als drei haben, dann zählen Sie einfach weniger als drei auf.

____________________ ____________________ ____________________

Blatt bitte wenden.

/Seite 2

Bitte kreisen Sie weiterhin die Antwort ein, die am ehesten auf Sie zutrifft.
Bitte beantworten Sie die Fragen so *ehrlich* wie möglich.

8. Wie häufig lesen Sie Zeitung (online und/oder in Papierform)?
   - a. Ich lese keine Zeitung.
   - b. Ich lese selten Zeitung.
   - c. Ich lese manchmal Zeitung.
   - d. Ich lese jeden Tag eine Zeitung.
   - e. Ich lese jeden Tag mehrere Zeitungen.

9. Generell, wenn Sie Ihr sämtliches Lesen miteinbeziehen, wieviel lesen Sie? Bitte nutzen Sie für Ihre Antwort die untenstehende 7-Punkte Skala von "*Ich lese fast nie* " bis "*Ich lese ständig* ".

Ich lese fast nie	Ich lese selten	Ich lese manchmal	Ich lese regelmäßig	Ich lese oft	Ich lese sehr häufig	Ich lese ständig

10. Haben Sie derzeit ein Buch bei sich, das Sie zum Vergnügen lesen? Können Sie den Namen des Autors angeben?

☐ Nein ☐ Ja Autorenname: ______________________________

Vielen Dank fürs Mitmachen.

11.10 326 English naming stimuli

acorn
alarm
album
arm
ask
balloon
bandit
banjo
banshee
barber
basis
bath
blame
bless
blond
bounce
brass
bronze
bubble
bus
butter
cabin
cadet
canal
candle
candy
canvas
capsule
cargo
cello
chain
chance
chaos
chauffeur
clap
claw
clown
coat
cocoon
coin
cold
combo
comet
comic
cool
cough
count
cradle
crater
crumb
crust
dark
denim
dent
depot
desk
dessert
detail
diesel
dirt
dish
dozen
drama
draw
drip
drop
drum
eel
ego
empty
error
erupt
fabric
fair
farm
female
fever
field
film
fish
fizz
flag
flame
fold
form
format
forum
fossil
frog
fruit
futile
garbage
gasp
gift
giraffe
glow
goblet
goblin
gold
golf
graph
gravy
growl
guru
gusto
half
halo
hamster
hand
hangar
harvest
haunt
hectic
help
hinge
hobby
hockey
hoop
horn
hotel
huge
humble
hunger
idol
ignore
jazz
jeep
jelly
jockey
jolt
jumbo
kilt
knowledge
ladder
lamb
lava
lift
limbo
limit
lobby
locket
loop
magnet
manual
maroon
marrow
mask
maze
medic
mellow
melon
melt
menu
mesh
mess
meter
milk
mirage
mobile
moment
mono
month
moped
moral
moth
motto
mouth
muffin
music
nanny
native
neck
nice
note
offer
office
omen
opium
panic
park
path
penny
pepper
person
petty
phase
pilot
pizza
planet
plasma
play
pledge
plenty
plot
plug
poet
pond
pony
post
prance
prep
problem
pronto
psyche
puke
punish
punt
puppet
puzzle
quail
queen
raccoon
radish
refrain
refuge
refuse
remote
rice
rinse
robot
rodent
room
rubber
sand
sauna
scarce
scoop
secret
sedan
self
send
serve
shampoo
shelf
shield
shy
sister
sleep
slogan
small
smart
smile
snap
snip
snob
snug
soccer
soda
sofa
soft
song
soothe
speak
spider
spin
spine
split
sport
spry
square
stab
stand
start
steak
stem
stench
stole
stop
strap
straw
strive
suburb
such
sulk
super
surf
swim
swoop
symbol
system
talent
tangle
tape
taxi
tempo
tempt
tennis
terrain
thermos
tibia
tiger
topic
town
tree
triumph
troll
trot
tuna
tunnel
turban
twin
unique
vase
velvet
venom
verb
vial
victim
villa
volley
volume
vulgar
warm
widow
wild
wind
winter
wipe
wish
world
yacht
yard
yoga
zebra
zoo
zoom

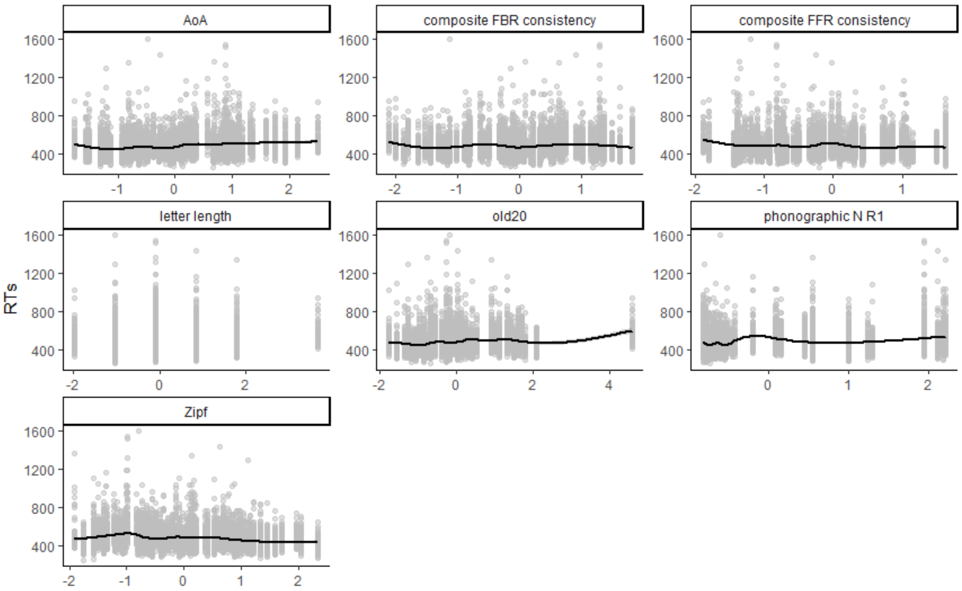
11.11 301 German naming stimuli

Aal
Alarm
Album
Arm
Bandit
Basis
Bein
Bengel
Besen
Beutel
Biest
Bilanz
Bild
Blei
Block
blond
Braut
Bronze
Brot
Bus
Butter
Cello
Chance
Chaos
Charme
Chauffeur
Clown
Dach
Darm
Depot
Dessert
Detail
Dieb
Diesel
Dock
Drama
Dreck
Ego
Eis
Essig
Ethik
ewig
fair
Farm
faul
Faust
Film
Flagge
Flamme
Flegel
Fleiß
Flöte
flott
Flotte
Flut
Form
forsch
Forum
Fracht
Frist
Gala
Gast
Gebiet
Gegend
Geist
Geld
Geste
Glück
Gnade
Golf
Gruß
Gurt
Guru
Hai
halb
Hangar
hart
heiß
Held
hell
Helm
Herr
Herz
Hirsch
Hobby
Hockey
Horde
Horn
Hotel
Hummer
Hunger
Hütte
Idol
Insel
Jazz
Jockey
Kaffee
Käfig
Kammer
Kessel
Kitsch
Klima
Klippe
Knabe
Kneipe
Knospe
Koffer
Komet
Komma
Korken
Kraft
Kredit
Kreis
Krimi
Krise
Krone
Kultur
Kunst
Kupfer
Kur
Küste
Lamm
Last
Lava
Leib
Leim
Licht
Lied
Lift
Limit
List
Liter
Lobby
Locke
Lohn
Lotto
Luft
Magen
Magnet
Mahl
Mais
Mantel
Markt
Marsch
Mast
Maul
Meer
Metall
Meter
mies
Milch
mild
Mist
Mittel
Moment
Moor
Moped
Moral
Motte
Motto
munter
Muschel
Neid
Nest
nett
Niveau
nobel
Note
Omen
Paket
Pakt
Panik
Park
passiv
Pedal
Person
Pfeffer
Pfennig
Phase
Pizza
Plasma
platt
Platz
Poet
Pony
Portal
Post
Posten
Preis
Problem
Profi
Profil
Psyche
Quote
Rabe
Raum
Rebell
Refrain
reich
reif
Reim
Reis
Rekord
Rente
Ring
Rost
Rubin
Rudel
Rune
Säbel
Salz
Sand
Satz
Sauna
Schal
Schuft
See
Sense
Seuche
Shampoo
Sieg
simpel
Sinn
Sitte
Skrupel
Slogan
Snob
Sofa
Sohn
Song
Spiel
Spion
Sport
Spott
Staat
stabil
Stamm
Start
Steak
steil
Stempel
steril
Stier
still
stur
Sumpf
super
Symbol
System
Taifun
Taktik
Talent
Taxi
Teil
Tempel
Tempo
Tennis
Terrain
Test
Theke
Titel
Tracht
Traum
treu
Triumph
Troll
Tropf
Trumpf
Truppe
Tunnel
Turban
Turm
Tyrann
Urteil
Vase
Vene
Verb
Vers
vier
Villa
Visum
voll
Wand
warm
weich
weit
Werk
wert
Wicht
wild
Wind
Winter
Wort
Yacht
Zahl
Zahn
zart
Zauber
Zepter
Ziel
Zorn

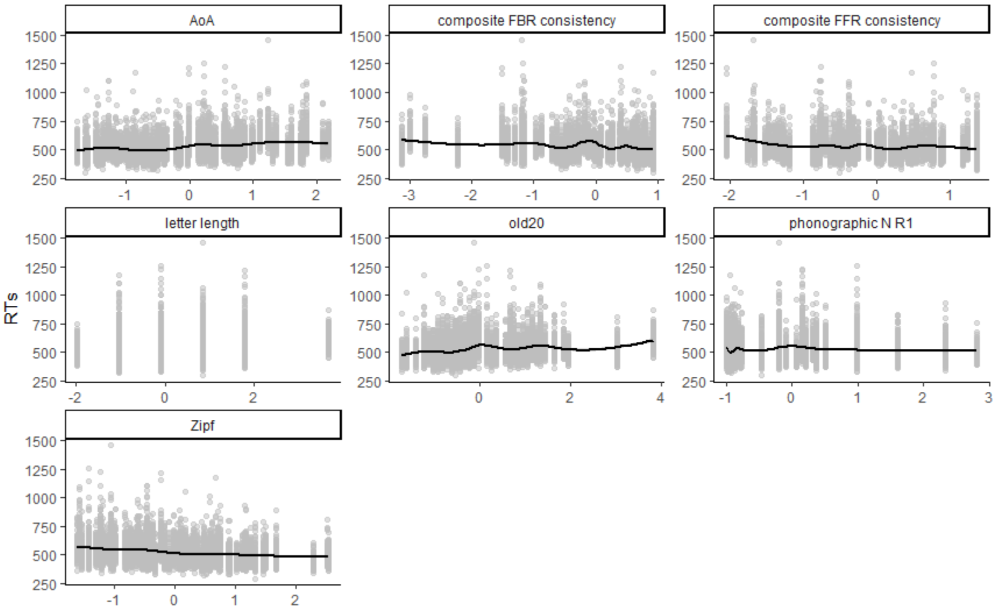
11.12 85 cognate stimuli

cognate naming stimuli			
English		German	
alarm	motto	Alarm	Motto
album	note	Album	Note
arm	omen	Arm	Omen
bandit	park	Bandit	Park
basis	person	Basis	Person
blond	phase	blond	Phase
bronze	pizza	Bronze	Pizza
bus	plasma	Bus	Plasma
butter	poet	Butter	Poet
cello	pony	Cello	Pony
chaos	post	Chaos	Post
chauffeur	problem	Chauffeur	Problem
clown	psyche	Clown	Psyche
depot	refrain	Depot	Refrain
dessert	sand	Dessert	Sand
detail	sauna	Detail	Sauna
diesel	shampoo	Diesel	Shampoo
drama	slogan	Drama	Slogan
fair	snob	fair	Snob
farm	sofa	Farm	Sofa
film	sport	Film	Sport
form	start	Form	Start
forum	super	Forum	super
golf	symbol	Golf	Symbol
guru	system	Guru	System
hangar	talent	Hangar	Talent
hobby	taxi	Hobby	Taxi
hockey	tempo	Hockey	Tempo
horn	tennis	Horn	Tennis
hotel	terrain	Hotel	Terrain
hunger	triumph	Hunger	Triumph
idol	troll	Idol	Troll
jazz	tunnel	Jazz	Tunnel
jockey	turban	Jockey	Turban
lava	vase	Lava	Vase
lift	verb	Lift	Verb
limit	villa	Limit	Villa
lobby	warm	Lobby	warm
magnet	wild	Magnet	wild
meter	wind	Meter	Wind
moment	winter	Moment	Winter
moped	yacht	Moped	Yacht
moral		Moral	

11.13 Scatterplots of naming RTs (msec) of 85 cognate stimuli by standardised psycholinguistic variables for each language group.

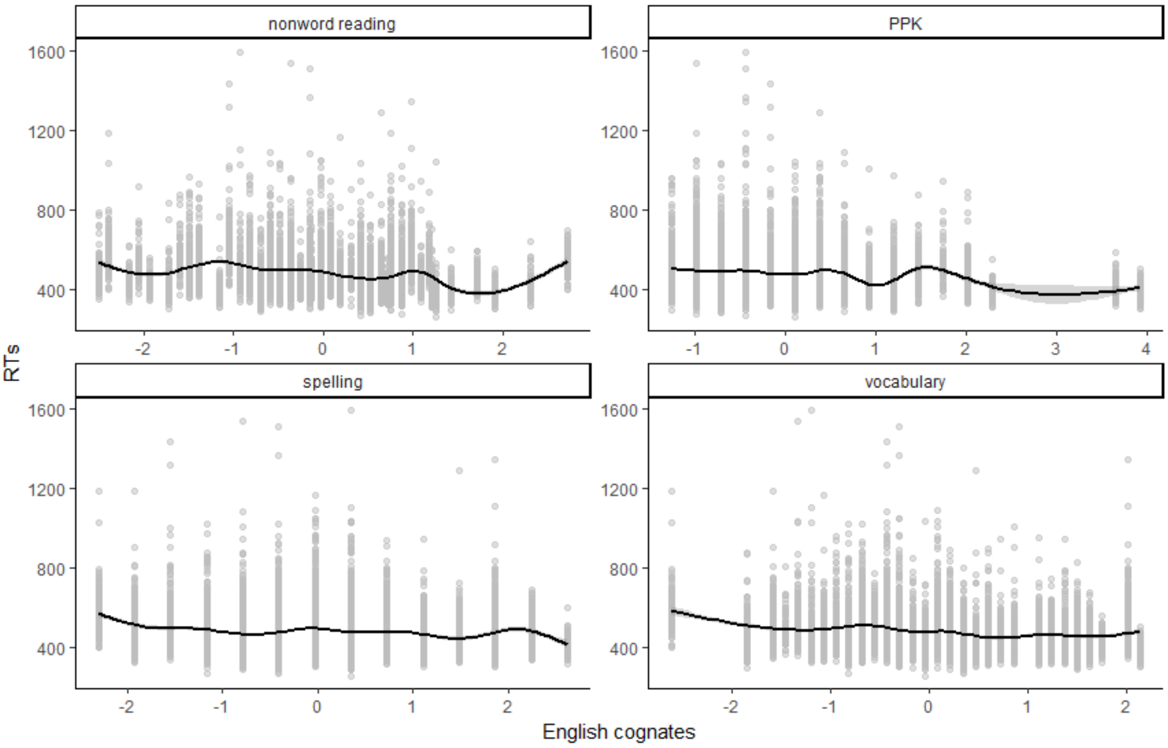


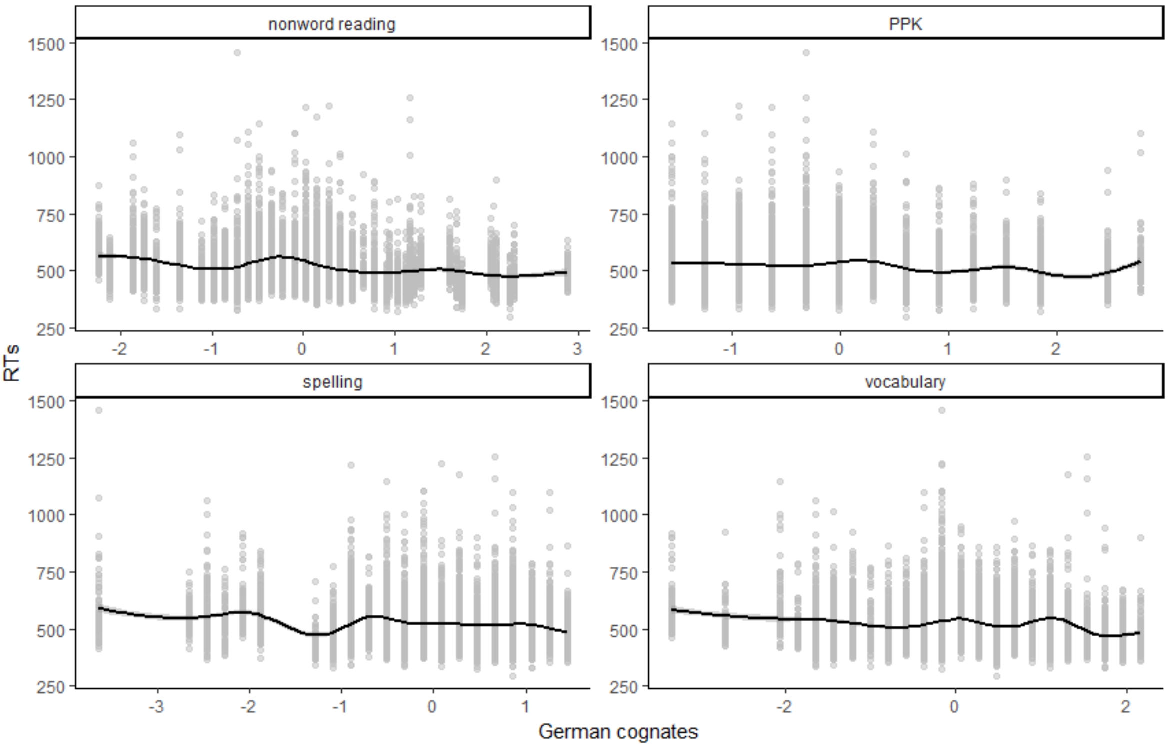
English cognates



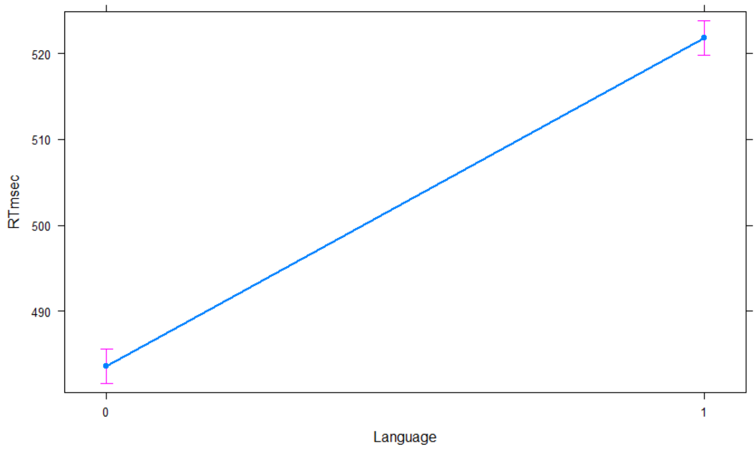
German cognates

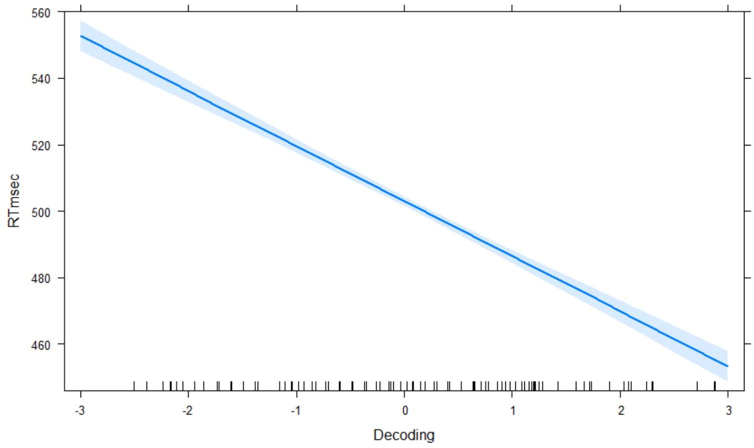
11.14 Scatterplots of naming RTs (msec) for 85 cognate stimuli by standardised individual differences variables

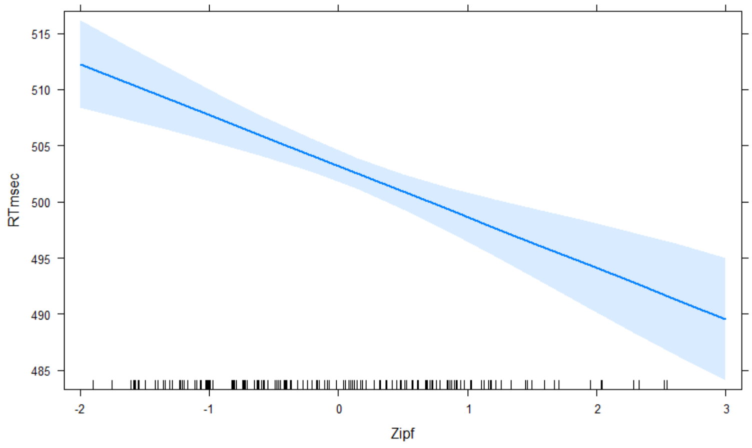


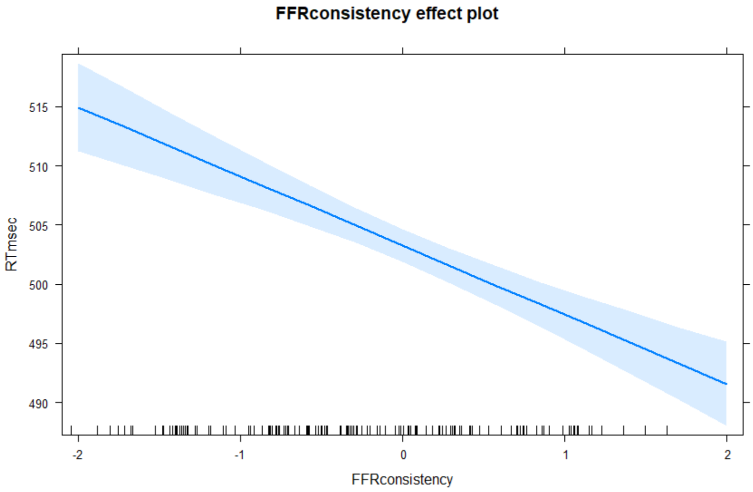


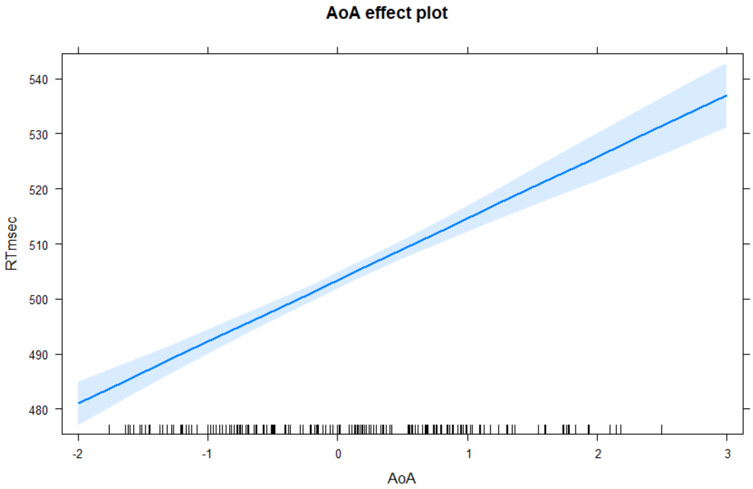
11.15 Analysis I: visualisations of effects of lmer model of combined cognate data in English and German using the R package ‘effects’ (Fox, 2003)

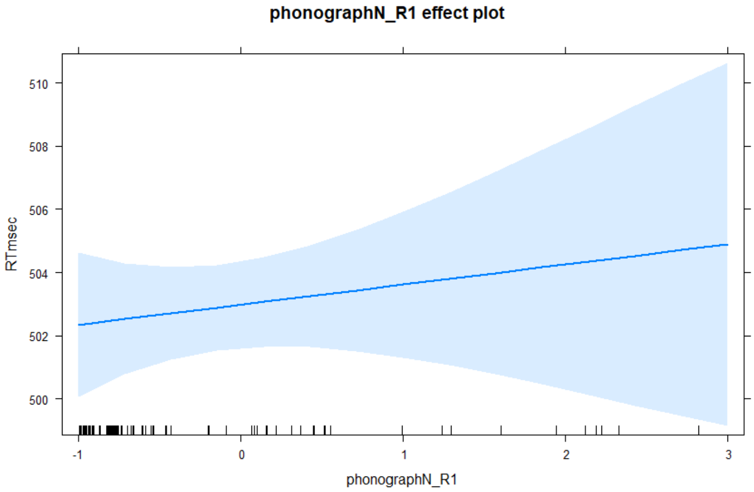

Language effect plot


Decoding effect plot

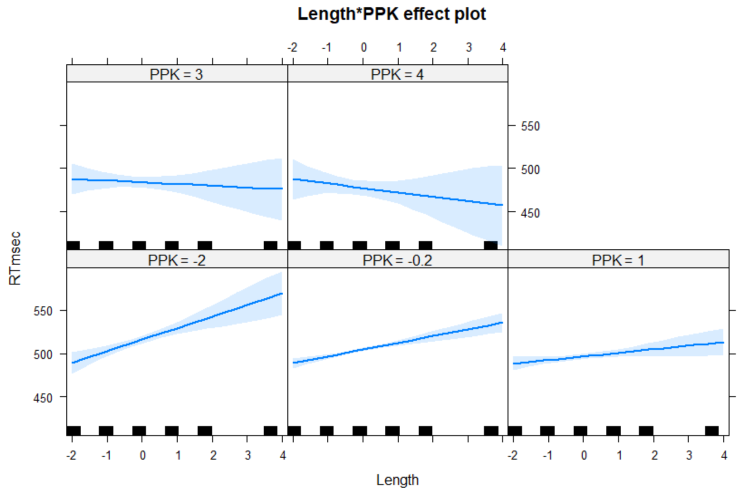

Zipf effect plot

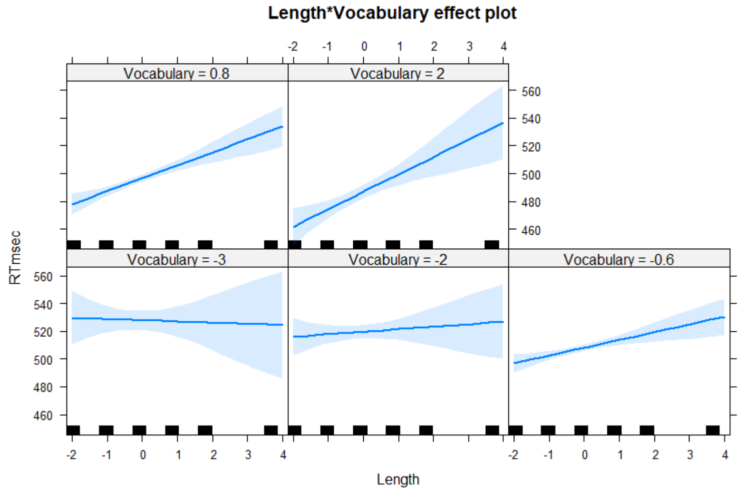

FFRconsistency effect plot
RTmsec
515
510
505
500
495
490
-2
-1
0
1
2
FFRconsistency

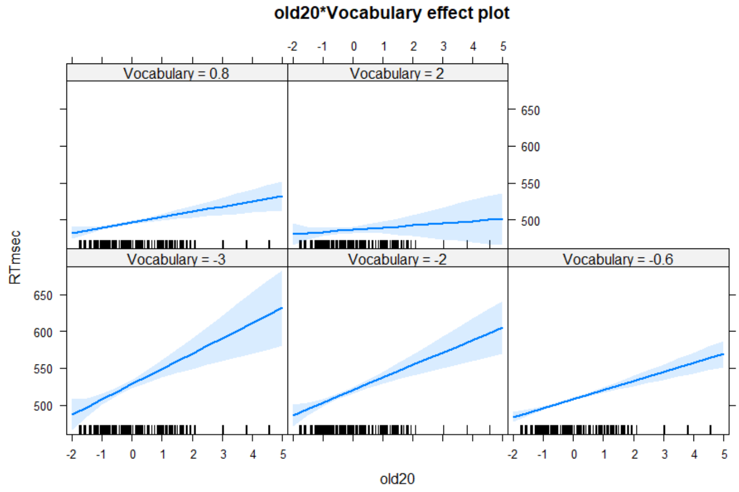

AoA effect plot
RTmsec
540
530
520
510
500
490
480
-2
-1
0
1
2
3
AoA

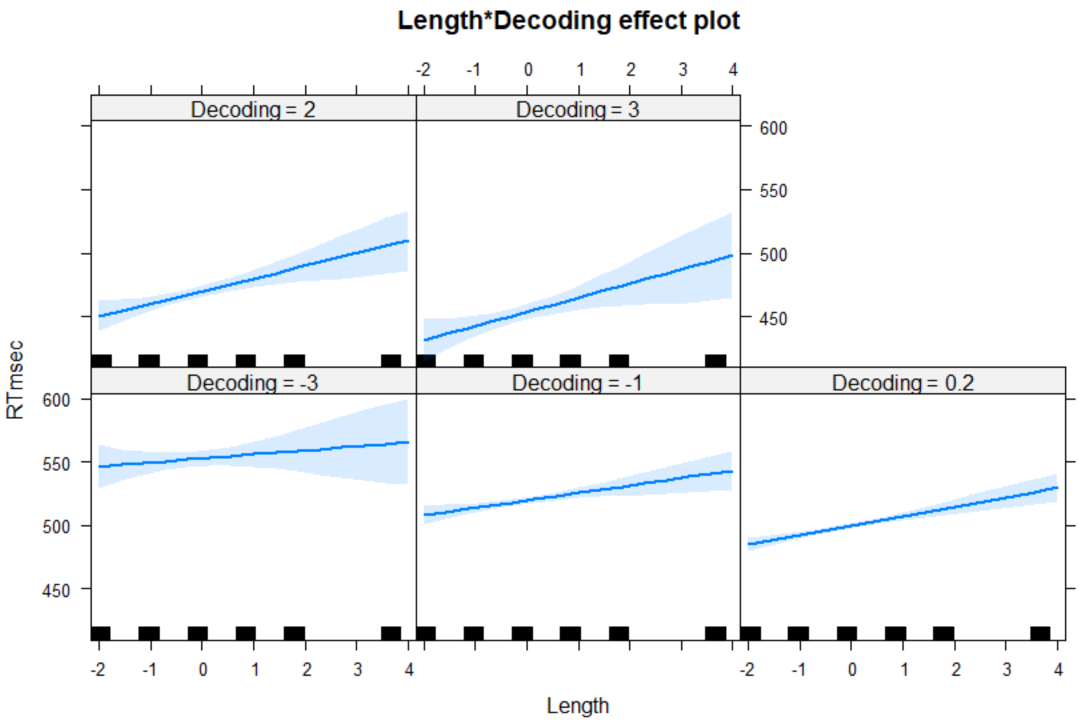

phonographN_R1 effect plot
RTmsec
510
508
506
504
502
500
-1
0
1
2
3
phonographN_R1

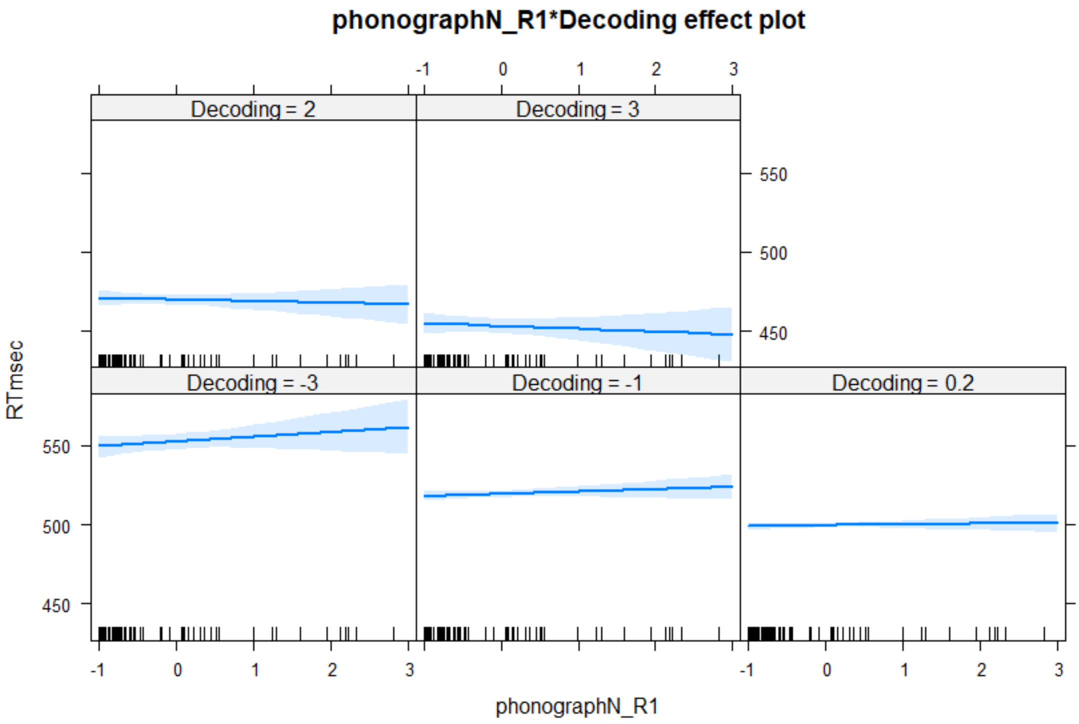
The following graphs are based on the same mixed-model, but with a different order. Whilst the reported model was based on *Language * ID variables * item variables*, the below graphs were obtained with the changed order specification on *Language * item variables * ID variables*. The graphs then give a different perspective of the relationship between item and ID variables. The author thought that it would add to the understanding of the 3-way interactions to be able to view the visualised relationships in both ways.

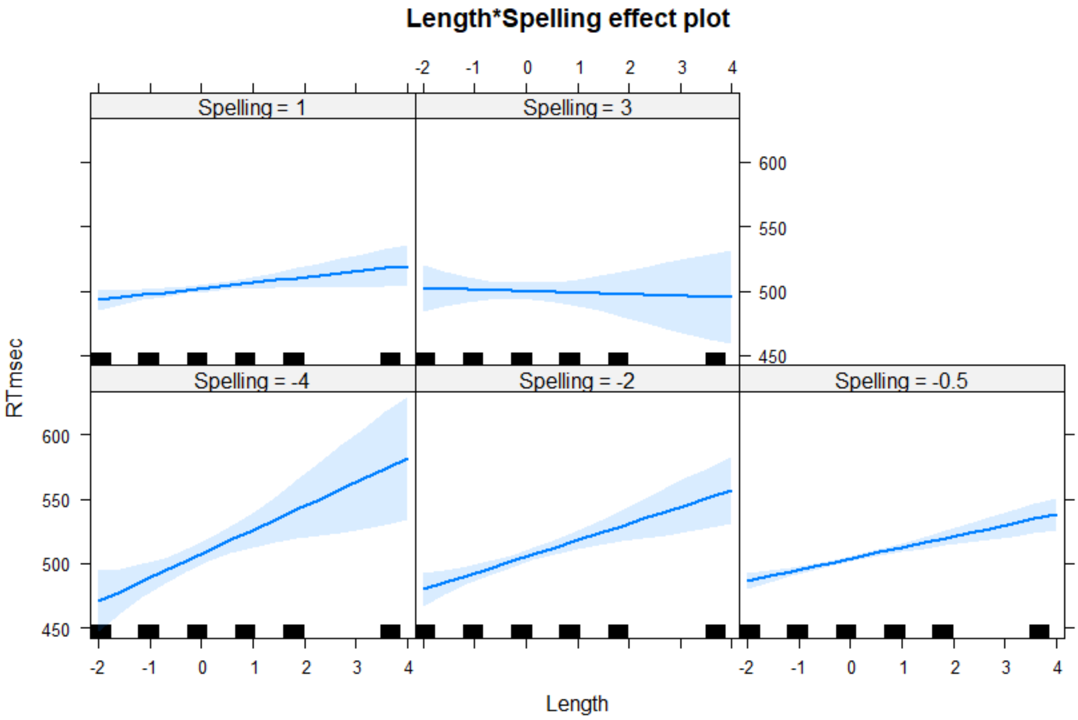


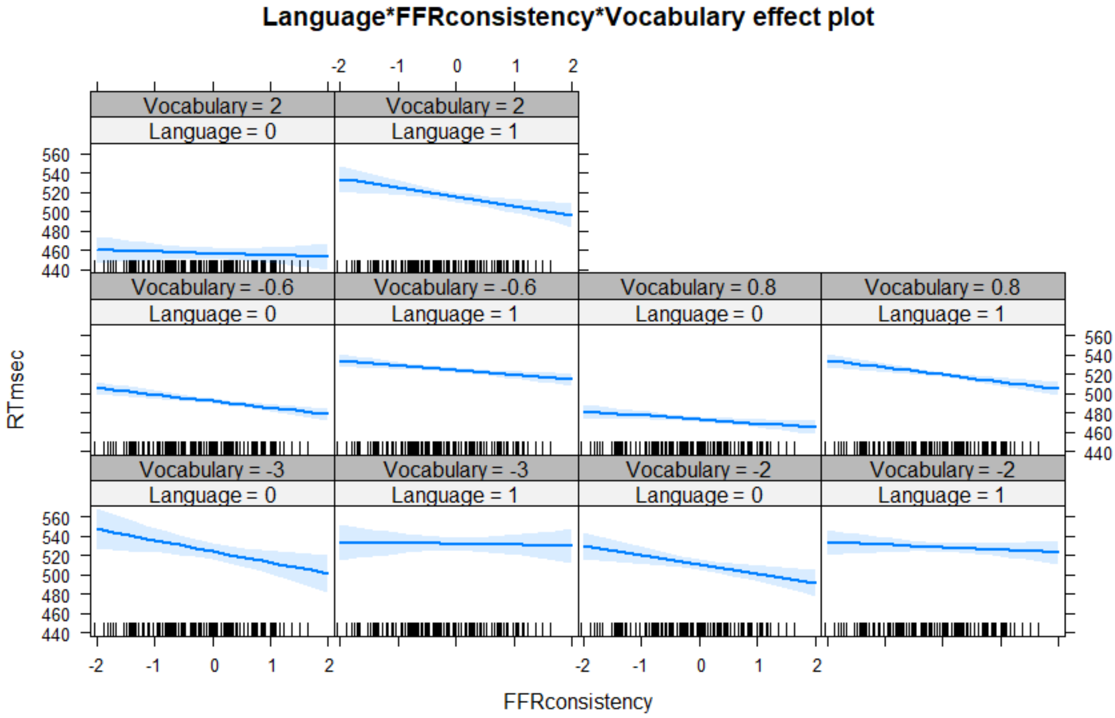


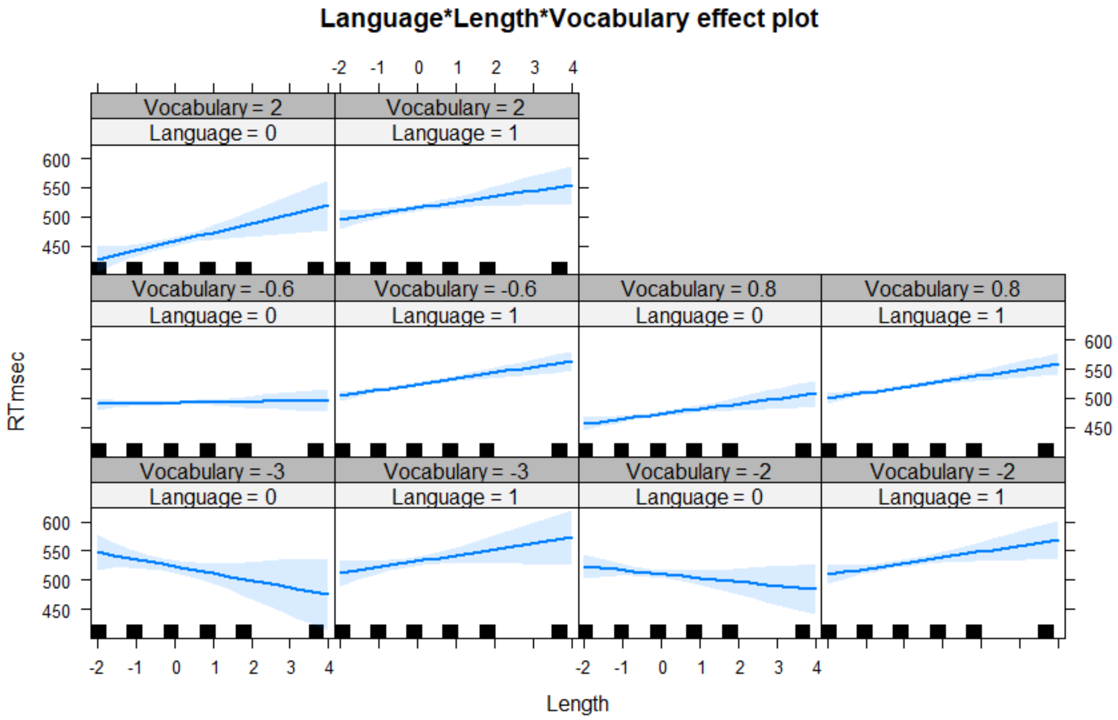


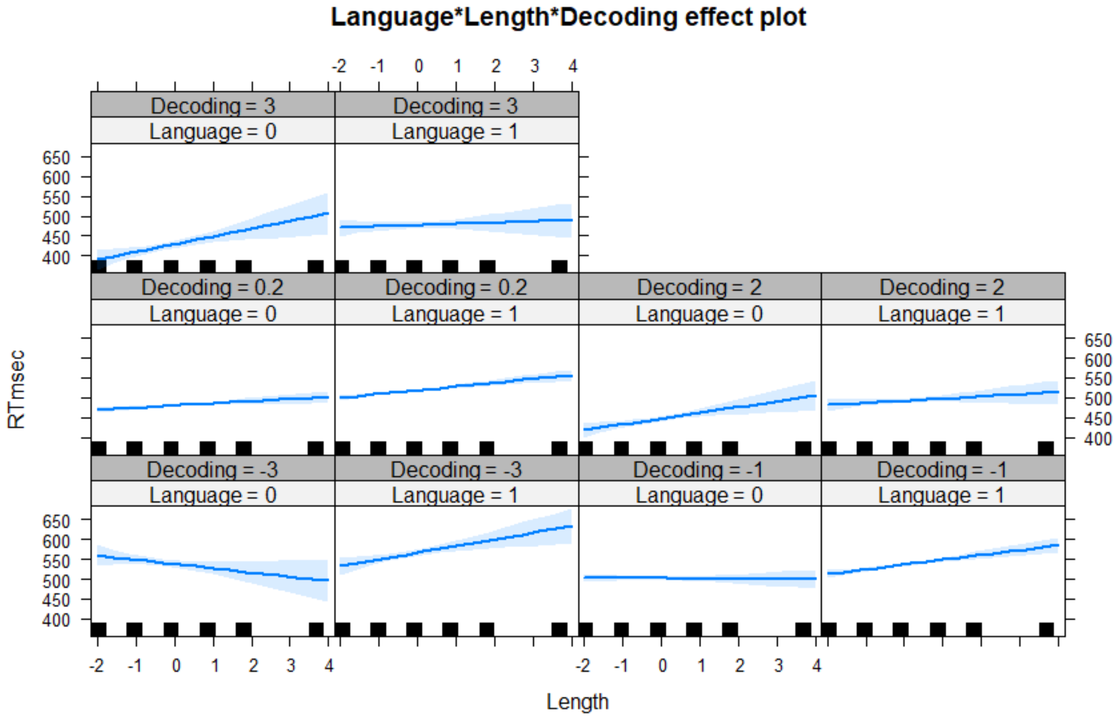

Length*Decoding effect plot
Decoding = 2
Decoding = 3
Decoding = -3
Decoding = -1
Decoding = 0.2
RTmsec
600
550
500
450
-2
-1
0
1
2
3
4
Length


phonographN_R1*Decoding effect plot
Decoding = 2
Decoding = 3
Decoding = -3
Decoding = -1
Decoding = 0.2
RTmsec
550
500
450
-1
0
1
2
3
phonographN_R1


Length*Spelling effect plot
Spelling = 1
Spelling = 3
Spelling = -4
Spelling = -2
Spelling = -0.5
RTmsec
Length
450
500
550
600
-2
-1
0
1
2
3
4


Language*FFRconsistency*Vocabulary effect plot
Vocabulary = 2
Language = 0
Vocabulary = 2
Language = 1
Vocabulary = -0.6
Language = 0
Vocabulary = -0.6
Language = 1
Vocabulary = 0.8
Language = 0
Vocabulary = 0.8
Language = 1
Vocabulary = -3
Language = 0
Vocabulary = -3
Language = 1
Vocabulary = -2
Language = 0
Vocabulary = -2
Language = 1
RTmsec
FFRconsistency
440
460
480
500
520
540
560
-2
-1
0
1
2


Language*Length*Vocabulary effect plot
Vocabulary = 2
Language = 0
Vocabulary = 2
Language = 1
Vocabulary = -0.6
Language = 0
Vocabulary = -0.6
Language = 1
Vocabulary = 0.8
Language = 0
Vocabulary = 0.8
Language = 1
Vocabulary = -3
Language = 0
Vocabulary = -3
Language = 1
Vocabulary = -2
Language = 0
Vocabulary = -2
Language = 1
RTmsec
450
500
550
600
Length
-2
-1
0
1
2
3
4


Language*Length*Decoding effect plot
Decoding = 3
Language = 0
Decoding = 3
Language = 1
Decoding = 0.2
Language = 0
Decoding = 0.2
Language = 1
Decoding = 2
Language = 0
Decoding = 2
Language = 1
Decoding = -3
Language = 0
Decoding = -3
Language = 1
Decoding = -1
Language = 0
Decoding = -1
Language = 1
RTmsec
400
450
500
550
600
650
Length
-2
-1
0
1
2
3
4

11.16 Analysis II model specifications

High – print group

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: RTNA ~ bilabial + labiodental + alveolar_postalveolar + palatal +
    velar + glottal + plosive + nasal + fricative + approximant +
    voiced + affricate + shortV + longV_diph + stress + (Lang) *
    (Art3_read + vocabtotal + TowSkill + spellr) * (Zipf + compFFR +
    compFBR + AoA + LetterCount + OLD + phonographN_R1) + (1 + Zipf + compFFR || Part) + (1 + Art3_
read || Upper)
```

Low – print group

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: RTNA ~ bilabial + labiodental + alveolar_postalveolar + palatal +
    velar + glottal + plosive + nasal + fricative + approximant +
    voiced + affricate + shortV + longV_diph + stress + (Lang) *
    (Art3_read + vocabtotal + TowSkill + spellr) * (Zipf + compFFR +
    compFBR + AoA + LetterCount + OLD + phonographN_R1) + (1 +
    Zipf + compFFR + compFBR + AoA + LetterCount + OLD || Part) + (1 | Upper)
```

High – vocabulary group

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: RTNA ~ bilabial + labiodental + alveolar_postalveolar + palatal +
    velar + glottal + plosive + nasal + fricative + approximant +
    voiced + affricate + shortV + longV_diph + stress + (Lang) *
    (Art3_read + vocabtotal + TowSkill + spellr) * (Zipf + compFFR +
    compFBR + AoA + LetterCount + OLD + phonographN_R1) + (1 + Zipf + compFFR + compFBR + AoA || Pa
rt) + (1 | Upper)
```

Low - vocabulary group

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: RTNA ~ bilabial + labiodental + alveolar_postalveolar + palatal +
    velar + glottal + plosive + nasal + fricative + approximant +
    voiced + affricate + shortV + longV_diph + stress + (Lang) *
    (Art3_read + vocabtotal + TowSkill + spellr) * (Zipf + compFFR +
    compFBR + AoA + LetterCount + OLD + phonographN_R1) + (1 + Zipf + compFFR + compFBR || Part) +
(1 | Upper)
```

Strong decoder group

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: RTNA ~ bilabial + labiodental + alveolar_postalveolar + palatal +
    velar + glottal + plosive + nasal + fricative + approximant +
    voiced + affricate + shortV + longV_diph + stress + (Lang) *
    (Art3_read + vocabtotal + TowSkill + spellr) * (Zipf + compFFR +
    compFBR + AoA + LetterCount + OLD + phonographN_R1) + (1 +
    Zipf + compFFR + compFBR + AoA + LetterCount || Part) + (1 + Art3_read + vocabtotal || Upper)
```

Weak decoder group

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: RTNA ~ bilabial + labiodental + alveolar_postalveolar + palatal +
    velar + glottal + plosive + nasal + fricative + approximant +
    voiced + affricate + shortV + longV_diph + stress + (Lang) *
    (Art3_read + vocabtotal + TowSkill + spellr) * (Zipf + compFFR +
    compFBR + AoA + LetterCount + OLD + phonographN_R1) + (1 +
    Zipf + compFFR + compFBR || Part) + (1 + Art3_read || Upper)
```

Strong speller group

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: RTNA ~ bilabial + labiodental + alveolar_postalveolar + palatal +
    velar + glottal + plosive + nasal + fricative + approximant +
    voiced + affricate + shortV + longV_diph + stress + (Lang) *
    (Art3_read + vocabtotal + TowSkill + spellr) * (Zipf + compFFR +
    compFBR + AoA + LetterCount + OLD + phonographN_R1) + (1 +
    Zipf + compFFR + compFBR + AoA + LetterCount || Part) + (1 + Art3_read + vocabtotal || Upper)
```

Weak speller group

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: RTNA ~ bilabial + labiodental + alveolar_postalveolar + palatal +
    velar + glottal + plosive + nasal + fricative + approximant +
    voiced + affricate + shortV + longV_diph + stress + (Lang) *
    (Art3_read + vocabtotal + TowSkill + spellr) * (Zipf + compFFR +
    compFBR + AoA + LetterCount + OLD + phonographN_R1) + (1 + Zipf + compFFR + compFBR || Part) +
(1 | Upper)
```

Note. RTNA = reaction times in msec, Lang = language, Art3_read = PPK (print exposure), vocabtotal = vocabulary knowledge, TowSkill = nonword reading/decoding skill, spellr = spelling skill, Zipf = frequency in Zipf, compFFR = composite feedforward rime consistency, compFBR = composite feedback rime consistency, AoA = age-of-acquisition, LetterCount = length in letters, OLD = old20, phonograph_R1 = phonographic N for rime 1, Part = participant, Upper = item.

11.17 Collection of German AoA ratings

Published paper. Please cite as

Birchenough, J. M., Davies, R. & Connelly, V. (2017). Rated age-of-acquisition norms for over 3,200 German words. *Behavior Research Methods*, *49*, 484–501.

The publication is available at

https://link.springer.com/article/10.3758/s13428-016-0718-0