

Running head: SUBLEXICAL CORRESPONDENCES

Quantifying the reliance on different sublexical correspondences in German and English

Xenia Schmalz, Eva Marinus, Serje Robidoux, Sallyanne Palethorpe, Anne Castles &

Max Coltheart

ARC Centre of Excellence in Cognition and its Disorders

Department of Cognitive Science

Macquarie University

Sydney, Australia

Abstract

The type of the sublexical correspondences employed during nonword reading has been a matter of considerable debate in the past decades of reading research. Nonwords may be read either via small units (graphemes), or large units (orthographic bodies). In addition, grapheme-to-phoneme correspondences may involve context-sensitive correspondences, such as pronouncing an "a" as /ɔ/ when preceded by a "w". Here, we use an optimisation procedure to explore the reliance on these three types of correspondences in nonword reading. In Experiment 1, we use vowel length in German to show that all three sublexical correspondences are necessary and sufficient to predict the participants' responses. We then quantify the degree to which each correspondence is used. In Experiment 2, we present a similar analysis in English, which is a more complex orthographic system.

Keywords: Reading, Sublexical processing, Optimisation.

Quantifying the reliance on different sublexical correspondences in German and English

How print is converted to speech is an important question, both from a theoretical and practical perspective. Sublexical translation processes have a central role in all current models of reading aloud (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Perry, Ziegler, & Zorzi, 2010; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989). The exact nature of this sound-to-speech conversion procedure, however, has been under considerable debate since the 1970s. In particular, the debate revolves around the question of whether this conversion relies predominantly on small units, such as graphemes, or larger units, such as orthographic bodies (e.g., "-ord") (Andrews, 1982; Coltheart et al., 2001; Cortese & Simpson, 2000; Glushko, 1979; Jared, 2002)¹. To a lesser extent, the literature has also drawn a distinction between context-sensitive and context-insensitive grapheme-to-phoneme correspondences (GPCs) and addressed the possibility that rather than relying purely on single-grapheme correspondences, in some cases the preceding or succeeding letters may provide a cue to the reader about the correct pronunciation of a grapheme (Perry, Ziegler, Braun, & Zorzi, 2010; Treiman, Kessler, & Bick, 2003; Treiman, Kessler, Zevin, Bick, & Davis, 2006).

Thus, the literature reports three different types of correspondences that may be involved in sublexical decoding: context-insensitive GPCs, context-sensitive GPCs, and body-rime correspondences. Here, we propose a mathematical model based on an optimisation procedure that will allow us to fit the degree of reliance on each of the three types of correspondences. We begin with two experiments in German, where the language structure allows us to assess the independent contribution of each of the three types of correspondences. In two further experiments, we apply the same methodology

to the English grapheme "a", which allows us to disentangle the reliance on context-sensitive GPCs compared to context-insensitive GPCs.

GPCs describe the relationship between graphemes and phonemes. The phoneme is the basic unit in spoken language, and a grapheme is the letter or letter cluster that corresponds to a single phoneme. The definitions of GPCs are straightforward in some cases; for example, the grapheme "b" always maps onto the phoneme /b/. This is an example of a context-insensitive GPC: regardless of the letters that precede or succeed the grapheme, its assigned phoneme does not change. However, this gets more complicated when we consider the GPC for a grapheme such as "a". In English, context-insensitive correspondences would dictate that "a" should be pronounced as in "cat". Using this correspondence, words like "was" and "false" would be considered irregular, meaning that the correct pronunciation is inconsistent with the GPC. Yet, upon closer inspection, the pronunciations of "was" and "false" are entirely predictable when the context of the grapheme "a" is taken into account: in "was", the "a" is preceded by a "w", which in most cases changes the pronunciation to /ɔ/, as in "wad" and "swan"². This context-sensitive correspondence can be written as "[w]a" → /ɔ/. The pronunciation of the vowel in "false" can be similarly predicted by a complex context-sensitive GPC, namely "[C]a[l][C]" → /o:/ (hereafter: a[l]-correspondence). It is worth noting that these context-sensitive correspondences are still GPCs, as they relate a single grapheme (in this case, "a") to the pronunciation of a single phoneme. Thus, GPCs can be subdivided into context-sensitive GPCs ("[w]a" → /ɔ/) and context-insensitive GPCs ("a" → /æ/).

The concept of GPCs is important for the classical computational model of the dual-route framework, the DRC (Coltheart et al., 2001). This model has a sublexical route which converts print to speech via a set of GPCs that are explicitly specified.

49 The sublexical route contains some context-sensitive correspondences ($N = 28$ - though
 50 the exact numbers vary according to the version of the DRC), but operates mostly on
 51 single-letter (e.g., "b" \rightarrow /b/; $N = 40$) and multi-letter (e.g., "th" \rightarrow /θ/; $N = 165$)
 52 context-insensitive GPCs.

53 There is also experimental evidence that stresses the importance of
 54 context-sensitive correspondences. One study reported the case of a patient with
 55 acquired surface dyslexia (Patterson & Behrmann, 1997): since this patient could not
 56 correctly read irregular words like "colonel" and "yacht", it was thought that her
 57 lexical system was heavily damaged. However, not all irregular words were a problem:
 58 she was unimpaired with words that could be resolved by the context-sensitive "[w]a"
 59 \rightarrow /ɔ/ correspondence, such as "wad" or "swan". This demonstrates the presence of
 60 such a context-sensitive correspondence in the sublexical system. Furthermore, studies
 61 of nonword reading have shown that there is psychological reality to context-sensitive
 62 correspondences (Treiman et al., 2003, 2006): both adults and children tend to
 63 pronounce nonwords such as "twamp" with the vowel as in "swan", whereas control
 64 items such as "glamp" are pronounced via the context-insensitive GPC, \rightarrow /æ/. This
 65 further suggests that the context-insensitive correspondence "a" \rightarrow /æ/ does not fully
 66 reflect the strategies used during nonword reading.

67 In addition to context-insensitive and context-sensitive GPCs, readers have been
 68 shown to rely on body-rime correspondences. Body-rime correspondences are the
 69 sublexical links between bodies and rimes, where bodies are defined as the vowel and
 70 optional final consonant(s) of a monosyllabic word (e.g., "-ark" in the word "bark").
 71 The rime is the phonological equivalent to the orthographic body. A linguistic analysis
 72 has shown that bodies are a reliable predictor of vowel pronunciation in English
 73 (Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995).

74 Full reviews of the psychological reality of body-rime correspondences can be
 75 found elsewhere (Goswami & Bryant, 1990; Ziegler & Goswami, 2005). Most relevant
 76 in the current context are nonword reading studies addressing this issue, because these
 77 allow for a systematic exploration of the non-lexical correspondences that participants
 78 rely on when lexical information is not available. In English, nonwords can be created
 79 that would yield different responses depending on whether GPCs or body-rime
 80 correspondences are used. This is done by manipulating the regularity and consistency
 81 of the base word. A base word that conforms to the context-insensitive GPCs is said to
 82 be regular, while words that violate the correspondences are considered to be irregular.
 83 The concept of regularity only matters if reading occurs at least in part via GPCs. If
 84 nonlexical reading occurs only via body-rime correspondences, then the reliability (or
 85 lack thereof) of the GPC information should not influence reading at all; rather, only
 86 inconsistency of body-rime correspondences should affect reading latencies and
 87 accuracy (e.g., the two ways of pronouncing "-ave" in "have" and "save", see Ziegler,
 88 Stone, and Jacobs (1997)).

89 Nonword reading studies that aim to estimate the reliance on GPCs versus
 90 body-rime correspondences can use the regularity and consistency of a base word to
 91 generate nonwords that predict different responses depending on the types of
 92 correspondences that are used by the participant. Such studies are important, because
 93 nonword reading data can shed light on processing underlying sublexical information,
 94 while minimising confounds from lexical processing. Understanding this process has
 95 strong theoretical implications, because sublexical print-to-speech conversion
 96 mechanisms play an important role in all prominent models of reading.

97 In order to disentangle the different sublexical processes that take place during
 98 reading, the first step is to create nonwords for which different types of

correspondences make different predictions. For example, from a regular and consistent word such as "fact", the onset can be changed to create a nonword, for example, "ract". In this case, both large and small correspondences make the same predictions for pronouncing this nonword. However, if we take an irregular, but consistent word, such as "talk", and change the onset to create the nonword "ralk", we can use the readers' pronunciations of this nonword to determine whether they relied on context-insensitive GPCs (in which case the item would be pronounced to rhyme with "talc") or body-rime correspondences (where it would rhyme with "talk"). Such studies have shown that GPCs cannot fully account for the types of pronunciations that participants give to such nonwords, but neither do body-rime correspondences (Andrews & Scarratt, 1998; Brown & Deavers, 1999; Perry, Ziegler, Braun, & Zorzi, 2010; Pritchard, Coltheart, Palethorpe, & Castles, 2012).

Thus, there is evidence for reliance on the three different types of print-to-speech correspondences, but there are still questions that remain to be answered. Firstly, previous studies do not distinguish between the reliance on context-sensitive GPCs and body-rime correspondences. For example, if a participant pronounces the nonword "palse" to rhyme with "false", it may be that a context-sensitive correspondence, "a[l]" \rightarrow /o:/ has been used to derive the pronunciation, rather than the body-rime correspondence that "-alse" \rightarrow /o:ls/. As will be discussed later, this is a problem in the English language, as body-rime correspondences and context-sensitive correspondences are confounded.

Secondly, even though such studies can establish the psychological reality of certain types of correspondences, examining between-item differences cannot provide any estimation of the *relative degree* to which each type of correspondence plays a role. As previous literature has demonstrated the psychological reality of context-insensitive

GPCs, context-sensitive GPCs, and body-rime correspondences, it is likely that all three correspondence types help the sublexical route to determine the pronunciation of a nonword. How such a conflict between different types of correspondences may be resolved by the cognitive system is addressed in detail in the General Discussion. The possibility of parallel activation of several sublexical correspondences raises the question of whether it is possible to quantify the degree to which each plays a role in determining the pronunciation of a novel word, which is a natural next step after demonstrating a sublexical correspondence's psychological reality. As discussed below, more sophisticated analyses are needed to estimate the relative importance of each type of correspondence.

In addition to establishing the psychological reality of different types of sublexical correspondences, a considerable body of research has explored cross-linguistic differences in the reliance on GPCs versus body-rime correspondences (Goswami, Gombert, & De Barrera, 1998; Goswami, Porpodas, & Wheelwright, 1997; Goswami, Ziegler, Dalton, & Schneider, 2003; Ziegler, Perry, Jacobs, & Braun, 2001; Ziegler, Perry, Ma-Wyatt, Ladner, & Schulte-Körne, 2003). The psycholinguistic grain-size theory, a cross-linguistic theory of reading development and skilled reading, proposes that the degree of reliance on sublexical correspondences of different types varies across languages (Ziegler & Goswami, 2005). In particular, the reliance on body-rime correspondences has been reported to be stronger in English than German (Ziegler et al., 2001, 2003). This is argued to be because in English, large units (i.e., bodies) are a better predictor of the pronunciation of a word than GPCs (Treiman et al., 1995): for a word like "calm", the pronunciation is inconsistent with the GPCs ("kælm") but can be derived from its body neighbours (palm, balm, etc.). In German, on the other hand, the GPCs are highly reliable, meaning that there are few exceptions

149 to the correspondences (Ziegler, Perry, & Coltheart, 2000), therefore smaller units are
 150 the preferred grain-size of German readers. In other words, there is a theoretical
 151 framework which predicts differences in the reliance on the units across languages.
 152 Therefore, it is desirable to develop a mathematical model quantifying the degree of
 153 reliance in different languages.

154 In summary, previous literature has shown reliance on three different types of
 155 correspondences in English: context-insensitive GPCs, context-sensitive GPCs, and
 156 body-rime correspondences. The psycholinguistic grain-size theory proposes that the
 157 reliance on the different types of correspondences differs across languages (Ziegler &
 158 Goswami, 2005). In the present experiments, we introduce a new method of
 159 quantifying the reliance on each type of correspondence. In the first two experiments
 160 (1A and 1B), we used German nonwords to assess the degree of reliance on each type
 161 of correspondence. In Experiment 2A and 2B we extend the procedure to a more
 162 complex orthographic system, namely English.

163 Experiment 1A

164 The German language allows us to neatly assess the independent contributions of
 165 context-insensitive GPCs, context-sensitive GPCs, and body-rime correspondences in a
 166 nonword reading paradigm: It is possible to create a set of items which generate
 167 different predictions for vowel pronunciation, depending on which strategy is used.

168 In German, there is relatively little ambiguity in print-to-sound correspondences,
 169 compared to English. What little ambiguity there is stems mostly from vowel
 170 pronunciation (Ziegler et al., 2000): Each vowel can be pronounced as either long or
 171 short (e.g., "Schal" /ʃa:l/ versus "Schall" /ʃal/). In monosyllabic words, vowel length is
 172 often signalled by context. Some context-sensitive correspondences allow the reader to

173 unambiguously determine vowel length; for example, any vowel followed by an "h" is
 174 pronounced long ("V[h]" → long vowel). Other context-sensitive correspondences are
 175 less transparent. These correspondences are described by a German implementation of
 176 Coltheart et al.'s (2001) DRC (Ziegler et al., 2000). To allow the sublexical route to
 177 determine vowel length, it contains a set of context-sensitive super-rules: any vowel
 178 which is followed by only one consonant elicits a long vowel response (e.g., "Wal"), and
 179 a vowel which is followed by two or more consonants is pronounced short (e.g.,
 180 "Wald"). These two rules can be summarised as follows: "V[C]" → long vowel, and
 181 "V[C][C]" → short vowel.

182 Although these super-rules capture the overall statistical distribution, there are
 183 also some exceptions, or words that would be irregular according to the German DRC
 184 (Ziegler et al., 2000). The word "Magd", for example, is pronounced with a long vowel;
 185 conversely the word "Bus" is pronounced with a short vowel. The presence of several
 186 bodies which consistently break the super-rules allows us to orthogonally manipulate
 187 the number of consonants in the body of a nonword, and the pronunciation of the
 188 base-words. Thus, we create a situation where the different types of correspondences
 189 (i.e., super-rules and body analogy) make different predictions about the pronunciation
 190 of the vowel.

191 For the present experiment, we can make a set of simple predictions if we assume
 192 that readers generally use only one type of correspondence: If only context-insensitive
 193 GPCs are used for German nonword reading, we expect that the likelihood of a short
 194 vowel pronunciation should be independent of any other orthographic features of the
 195 nonword. Such a GPC would predict many more short than long vowels, as the
 196 majority of vowels in German have the short pronunciation (Perry, Ziegler, Braun, &
 197 Zorzi, 2010). If a context-sensitive super-rule is used, vowel length should be solely

determined by the number of consonants following the vowel. In this case, even a nonword based on the irregular but consistent word such as "Magd" (e.g., "blagd") should be pronounced with a short vowel. These irregular-base-word items can distinguish between reliance on super-rules compared to body-rime correspondences: if body-rime correspondences are used, nonwords based on irregular consistent words should be pronounced to rhyme with their real-word counterparts.

Methods

Participants were 12 German native speakers who were staff or postgraduate students at Macquarie University, or members of the university's German society. As they lived in Australia, they were also fluent in English - a point which we will discuss in a later section. With one exception, all participants had completed secondary education in Germany and 10 had also attended German tertiary education. One participant had moved to Australia at the age of 5, but had attended a German-speaking school for 7 years.

The nonwords that were used for this experiment are listed in Appendix A. There were 30 nonwords in each of three conditions. The nonwords were created by changing the onsets of real words. All base-words were taken from a list of consistent German words (J. Ziegler, personal communication, 2012). The first condition used base-words with V[C] bodies which were pronounced with a long vowel ("Jod" → "FOD"); the second condition was based on V[C][C] words with a short vowel ("Saft" → "BLAFT"). The third condition was derived from irregular words, which had either a V[C] body but a short vowel ("mit" → "GIT") or a V[C][C] structure and a long vowel ("Jagd" → "BAGD"). The three conditions were matched on orthographic N (the number of real words that can be created by substituting one letter): V[C] items

222 had an average orthographic N-size of 1.73 (SD = 1.46), V[C][C] items had a mean of
 223 2.10 (SD = 1.69), and items with irregular base-words had a mean of 1.83 (SD = 1.90).
 224 The mean body-N (number of real words with the same body) for the three conditions
 225 is 1.93 (SD = 1.87), 2.40 (SD = 1.98), and 1.37 (SD = 1.00) respectively.

226 Participants were tested individually in a quiet room. Instructions were given in
 227 German by a native speaker. The participants were told that they would be asked to
 228 read nonwords which were created using German orthographic rules. The instructions
 229 emphasised that accuracy was more important than speed to discourage quick lexical
 230 processing, which might result in lexicalisation errors.

231 The items were presented using the DMDX software package (Forster & Forster,
 232 2003) in random order. Each trial consisted of a fixation cross, which remained in the
 233 centre of the screen for 500 ms, followed by the item, which remained on the screen
 234 until the voice-key was triggered. Ten practice nonwords preceded the experiment. As
 235 all nouns in German are spelled with capital initial letters, presenting nonwords in all
 236 lower-case would provide an indication of word class of a nonword. Previous research
 237 has shown that information on the likely word class of a nonword affects its
 238 pronunciation (Campbell & Besner, 1981). Therefore, all items were presented in
 239 upper case.

240 *Results*

241 Six trials (0.6%) were excluded due to poor sound quality or premature voice-key
 242 triggering. The rest of the trials were scored by a German native speaker as
 243 pronounced with a long vowel, a short vowel, or incorrectly. For identifying incorrect
 244 responses, we used a lenient marking criterion: if a participant's response was
 245 consistent with a possible pronunciation of the GPCs, it was marked as correct (e.g.,

246 "spic" was marked as correct regardless of whether it was pronounced as /spik/ or as
 247 /ʃpik/ - while, in German, "s" is typically pronounced as /ʃ/ before "p" or "t", there
 248 are a few instances, such as loanwords, where it is assigned the pronunciation /s/).
 249 Overall, 1.1% of all responses were classified as incorrect and excluded from subsequent
 250 analyses. Of primary interest were the proportions of long and short vowel responses
 251 and how they differed across condition. We split the Irregular-base-word condition by
 252 whether the bodies had one (hereafter referred to as V[C] Irregular; N = 17) or two
 253 (V[C][C] Irregular, N = 13) consonants. Note that the two "irregular" conditions did
 254 not differ dramatically on any item characteristics: the mean number of letters was
 255 3.88 (SD = 0.34) and 4.36 (SD = 0.63) for the V[C] and V[C][C] conditions
 256 respectively, orthographic N was 1.88 (SD = 2.19) and 1.79 (SD = 1.58) respectively,
 257 and body-N was 1.56 (SD = 1.15) and 1.29 (SD = 0.61) respectively. The proportions
 258 of short vowel responses for each of the four item types (V[C] Regular, V[C][C]
 259 Regular, V[C] Irregular and V[C][C] Irregular) are listed in Table 1, along with the
 260 predictions according to each of the three types of correspondences.

[Table 1 about here]

261 In order to make the predictions more specific, we can use a corpus analysis to
 262 determine the percentage of times a vowel is pronounced as long or short under certain
 263 circumstances. For example, overall, 78.02% of all monosyllabic words are pronounced
 264 with a short vowel (Perry, Ziegler, Braun, & Zorzi, 2010), therefore if German readers
 265 rely on context-insensitive GPCs, we expect them to give around the same percentage
 266 of short vowel responses. Among words with a single-consonant coda, 24.53% are
 267 pronounced with a short vowel, so we expect about the same percentage of short vowel
 268 responses to V[C] nonwords, if only super-rules are used to determine vowel length. In
 269 Table 1, we present the predicted vowel lengths for each of the four conditions and by

each of the three types of correspondences. For the context-insensitive GPCs and super-rules, these are calculated from the analyses presented in Perry, Ziegler, Braun, and Zorzi (2010). The predictions of the body-rime correspondences depend on the consistency ratio of the body. In the current study, we used only consistent items, where the body has only one pronunciation in real words. This means that if participants rely solely on body-rime correspondences, 100% of the pronunciations should be consistent with the base word vowel length.

The obtained percentages of long and short vowels (Table 1) are not consistent with the predictions of any one strategy we described above: vowel length responses are neither predominantly short in all four conditions, nor completely dependent on the number of consonants following the vowel, nor the vowel length of the base word. This is a clear indication that German readers rely on more than one type of correspondence for reading nonwords. Moreover, a closer look at Table 1 shows that no combination of two types of correspondences can account for the results, either: If context-insensitive GPCs and context-sensitive correspondences were the sole determiners of vowel length, we would not expect to find different proportions for the V[C] Regular and V[C] Irregular items - but we do. If context-insensitive GPCs and body-rime correspondences were the only predictors of vowel length, we would find no difference between the V[C] Regular and the V[C][C] Regular items - and we do. If only context-sensitive correspondences and body-rime correspondences were used, we should observe less than 25% short vowel responses - which is not supported by the data.

Modelling Vowel Pronunciations

As shown above, it is not possible for a single or even a pair of types of correspondences to adequately fit the empirical data. It may be, however, that some

combination of all three types of correspondences provides a good fit. Here we introduce a mathematical modelling approach that allows us to uncover more complex relationships between the types of correspondences. The goal is to weight the three strategies³ in a way that optimally fits the empirical data. More formally, we are seeking a set of β weights that best satisfy the following mathematical model (one pair of equations for each item):

$$\begin{aligned} P_i(Short) &= \beta_{gpc} \times GPC_{short,i} + \beta_{csc} \times CSC_{short,i} + \beta_{brc} \times BRC_{short,i} \\ P_i(Long) &= \beta_{gpc} \times GPC_{long,i} + \beta_{csc} \times CSC_{long,i} + \beta_{brc} \times BRC_{long,i} \end{aligned} \quad (1)$$

where $GPC_{[length],i}$ is the probability of item i being pronounced with a vowel of the corresponding length according to the corpus analysis when using only context-insensitive (single-letter) GPCs as a predictor, $CSC_{[length],i}$ is the probability according to context-sensitive super-rules, and $BRC_{[length],i}$ is the probability according to the body-rime correspondence rules. Table 1 provides the average predictions for each condition, but the predictions from each correspondence were calculated separately for each item in the experiments. $P_i([length])$ is the empirically observed proportion of the vowel length in Experiment 1A.⁴

At a first glance, this would appear to be a simple regression problem (with no intercept term). Linear regression would optimally select β values that minimised the prediction error for (1) (indexed by the residual sum of squares). However, there are several reasons why this should not be thought of as regression. First, since the β values are thought of as the degree to which a strategy applies in reading the items in Experiment 1, negative values would be uninterpretable. This means that all of our β parameters must exceed 0. This constraint can not be guaranteed by standard linear regression using ordinary least squares (Monfort, 1995).

Even with only positive β s, there are two ways to interpret the weights. One

could think of them as the contribution of each strategy to some sort of blending process that ultimately chooses the vowel pronunciation. In which case, we can simply fit the model in (1) above with the constraint that $\beta_i \geq 0, \forall i$. Alternately, one can think of the weights as the probabilities of adopting the vowel prediction from a given strategy. We prefer the latter interpretation (and discuss some evidence for it later), but it requires two further constraints: the β weights must fall below 1, and, since we assume the three strategies (GPCs, super-rules, and body-rime correspondences) are exhaustive, the three β s must sum to 1. The model can be formalised as:

$$\begin{aligned}
 P_i(short) &= \beta_{gpc} \times GPC_{short,i} + \beta_{csc} \times CSC_{short,i} + \beta_{brc} \times BRC_{short,i} \\
 P_i(long) &= \beta_{gpc} \times GPC_{long,i} + \beta_{csc} \times CSC_{long,i} + \beta_{brc} \times BRC_{long,i} \\
 &\text{where } \beta_j \in [0, 1] \text{ and } \Sigma \beta_j = 1, \forall j \in \{gpc, csc, brc\}
 \end{aligned} \tag{2}$$

that is, we are seeking a set of probabilistic weights on the three strategies that minimises the prediction error of the model. The challenge here is to both efficiently search the available parameter space, and satisfy the $\Sigma \beta_j = 1$ constraint. The first problem is a well-studied one in computer science and solutions are available that solve it. The second problem is largely solved by introducing an additional equation that can only be satisfied if $\Sigma \beta_j = 1$, and giving that equation a strong influence on the final parameter set. The interested reader can find a fuller discussion of the implementation details in Appendix B.

Optimal weights in Experiment 1A. In Experiment 1A, we collected the proportion of short and long vowel responses to 90 items, and for each item we have the predicted probability of a short or long vowel pronunciation according to each of the three strategies. The strategy predictions were obtained from the corpus analysis undertaken by Perry, Ziegler, Braun, and Zorzi (2010).

Using the technique described above, the native German readers in Experiment 1A appear to be relying most on GPCs ($\hat{\beta}_{gpc} = 0.56$), and to a lesser extent on super-rules ($\hat{\beta}_{csc} = 0.19$), and body-rime correspondences ($\hat{\beta}_{brc} = 0.26$). See Table 2 for a summary of the modelling results across all of the present experiments.

[Table 2 about here]

The above analysis contains a theoretically supported but strong assumption that readers use *only* the three strategies described in the introduction when reading nonwords. It is possible that other sources of information are used by German native speakers to determine vowel length. We can provide a simple test of this possibility by relaxing some of the constraints on the model, and observing how critical those constraints were to the optimisation results. To do this we removed the $\sum \beta_j = 1$ constraint, and allowed the β s to take on any positive weights in the fitting process. That is, we fit the following alternative model (some subscripts indicating length and item have been omitted for simplicity):

$$P(\text{length}) = \beta_{gpc} \times GPC + \beta_{csc} \times CSC + \beta_{brc} \times BRC, \text{ where } \beta_i > 0 \ \forall i \quad (3)$$

If readers are adopting other strategies that are not well described by the GPC, super-rules and body-rime correspondence strategies, the incomplete nature of the model should be reflected in these alternate weights. The weights that optimise (3) were $\hat{\beta}_{gpc} = 0.58$, $\hat{\beta}_{csc} = 0.14$, and $\hat{\beta}_{brc} = 0.24$. These values sum to 0.96, suggesting that there is little need for a fourth strategy to describe the data. This does not conclusively rule out a role for any other strategies, but provides some evidence that the three strategies already tested are sufficient. That said, there is one additional strategy that could be playing a role: anti-body correspondences, or the probability of a vowel being pronounced as long or short based on the *onset* of the word. In this

corpus of nonwords, the predictions from antibody-rime correspondences and context-insensitive GPCs are highly correlated, so it is difficult to disentangle the two strategies entirely, but it may be that anti-body rime correspondences are more important than context-insensitive GPCs and thus are a better predictor. To test whether or not anti-body correspondences were important for determining vowel pronunciations, we added a component to model (2):

$$P(\text{Length}) = \beta_{gpc} \times GPC + \beta_{csc} \times CSC + \beta_{brc} \times BRC + \beta_{abc} \times ABC \quad (4)$$

where $\beta_j \in [0, 1]$ and $\sum \beta_j = 1$

where the addition of ABC represents the predictions from anti-body correspondences, and β_{abc} is the associated weight. Fitting (4) produced the same weights that resulted from (2) where the antibody-rime correspondences were not included. That is, $\hat{\beta}_{abc} = 0$, giving little reason to believe that any other strategies are being used in Experiment 1A.

Model fits. The optimisation procedure presented here is only useful if it arrives at a model that fits the data better than alternatives. To determine the effectiveness of the model, we calculated the correlation between the model predictions and the observed response patterns. For comparison, we did the same for the GPCs, context-sensitive correspondences (CSC), and body-rime correspondences (BRC) individually. As can be seen in Table 3, the optimisation process outperforms the other three alternatives in all four samples presented here. In experiment 1A, the correlation is .844 while the next best model (based on context-insensitive GPCs) correlates at .714.

[Table 3 about here]

380 *Discussion*

381 In Experiment 1A, we successfully used an optimisation procedure to quantify
 382 the degree of reliance on three types of sublexical correspondences: context-insensitive
 383 GPCs, context-sensitive GPCs, and body-rime correspondences. This can be achieved
 384 with the German language, because it is possible to create items where different
 385 correspondence types make different predictions about the vowel length pronunciation.

386 Importantly, we found that all three types of correspondences are both *necessary*
 387 and *sufficient* to predict vowel length responses in a sample of German native speakers.
 388 Context-insensitive correspondences appear to be the strongest predictor. This is in
 389 line with the psycholinguistic grain size theory, which argues that the smallest unit size
 390 is favoured by readers of a language with predictable GPCs, such as German (Ziegler
 391 & Goswami, 2005).

392 Experiment 1A has some limitations. It could be argued that the results are
 393 unreliable, firstly due to the small sample size and secondly because the participants
 394 were bilingual, and very fluent in English. It is unclear how fluency in English may
 395 affect the reliance on different types of correspondences in German. Even though we
 396 took care to only include German participants who learned to read and write in
 397 German from a young age, there is a possibility that their exposure to German reading
 398 material has been diminished by residing in an English-speaking country. It is also
 399 possible that their knowledge of English would change the preferred unit in their native
 400 language: for example, psycholinguistic grain size theory predicts that readers of
 401 English rely more heavily on larger grain sizes than readers of German (Ziegler &
 402 Goswami, 2005), though it does not make any statements about sublexical processing
 403 in bilinguals. We address these concerns in Experiment 1B.

Experiment 1B

In Experiment 1B we collected data with two different samples of German native speakers who live in Germany and are not exposed to English on an everyday basis. We hereafter refer to them as monolingual Germans, even though they are not strictly monolingual: due to globalisation, it would be difficult if not impossible to find Germans who have no knowledge of English. Having collected data with two different samples of monolingual Germans allows us to test the reliability of the modelling method described here. If our model arrives at similar weights for two independent samples from the same population, we can be more confident that our modelling procedure is stable and reliable.

Methods

The methods were almost identical to Experiment 1A. One item was replaced (due to a typo, the original item set contained an inconsistent item, "blen", which was changed to "blem" in Experiment 1B).

The first sample consisted of 10 German native speakers who were staff or students at the Freie Universität in Berlin. All had completed their schooling in Germany. The second sample consisted of 26 undergraduate students at Potsdam University. Again, all were native German speakers and had completed their education in Germany.

Results

The scoring procedure was identical to Experiment 1A. For the Berlin sample, there were two non-responses (0.22%) and 15 errors (1.67%). The Potsdam sample made 2.3% errors. A series of t-tests showed that the percentages of long and short

vowel responses did not differ significantly for any of the conditions across the two samples, all $p > 0.4$. Furthermore, fitting each sample separately using the model described in (2) produced very similar weights. For the participants from Berlin, the weights were $\hat{\beta}_{gpc} = 0.40$, $\hat{\beta}_{csc} = 0.33$, and $\hat{\beta}_{brc} = 0.27$. For the participants from Potsdam they were $\hat{\beta}_{gpc} = 0.37$, $\hat{\beta}_{csc} = 0.35$, and $\hat{\beta}_{brc} = 0.28$. This result is comforting, suggesting that the method introduced here is reliable across different samples from similar populations. Since there was little difference between the two samples, we collapsed across them yielding a sample of 36 native German monolinguals. Using this collapsed sample, our model produces $\hat{\beta}_{gpc} = 0.38$, $\hat{\beta}_{cdc} = 0.35$, and $\hat{\beta}_{brc} = 0.27$. As in Experiment 1A, the optimal parameter set outperforms the alternatives in fitting the observed data (Table 3).

German/English bilingual vs. German monolingual readers. Since Experiments 1A and 1B are based on the same set of items, we have the opportunity to compare how the bilingual readers differed from the monolingual readers. The critical question is whether or not the smaller $\hat{\beta}_{gpc}$ and larger $\hat{\beta}_{csc}$ for monolinguals represents a real difference, or simply random variation. In the usual context of a linear regression model, this would be a simple matter of including the language status of the participants (bilingual vs. monolingual) in the model, and testing for an interaction between language status, and the GPC and/or CSC estimates. However, our modelling strategy violates many of the assumptions that allow for straightforward t-tests of the parameter estimates (given the constraints of our model, the parameter estimates are unlikely to be well-behaved, statistically). Instead we turn to a bootstrapping methodology to allow us to use the data to conduct non-parametric tests of the variability in our estimates.

To establish the reliability of the difference in the $\hat{\beta}_{gpc}$ and $\hat{\beta}_{csc}$ estimates, we

repeatedly resampled 90 items (with replacement) from the data set, and estimated the $\hat{\beta}_i$ s for both the bilingual and monolingual participants with each sample of items. Of 10,000 such samples, 9,890 (98.9%) produced a larger GPC weight for the bilingual subjects than for the monolingual subjects (95%CI of the difference: 0.019 to 0.327). Similarly, 9,634 (96.2%) samples produced a larger CSC weight for the monolingual participants than for the bilingual participants (95%CI: -0.011 to 0.317). This suggests that the difference in the GPC weights is robust, while the difference in the CSC weights is slightly more tenuous. The difference in the BRC weights was not at all significant: 3,454 (34.5%) of the samples produced larger BRC weights for bilinguals than for monolinguals (95%CI: -.058 to .089). We also took advantage of these bootstrap samples to estimate the variability in the correlations from the optimal parameters in Table 3.

To summarise the results so far, the reliance on body-rime correspondences did not differ between monolingual and bilingual readers, but there was a very stable difference in the reliance on context-insensitive GPCs and a somewhat stable difference in the role of context-sensitive super-rules. Monolinguals relied less on context-insensitive GPCs and somewhat more on super-rules than bilinguals.

Individual differences. There is some ambiguity in interpreting the weights: as we collapsed across participants, the weightings do not give us any information about inter-individual participant variability. Theoretically, it is possible that all participants rely on the same strategies to the same extent, or that the weightings are reflective of the percentage of participants who rely on a particular strategy only. To address this, we generated the weightings for each individual participant in Experiments 1A and 1B. These are summarised in Figure 1. This figure shows that there is individual variability, but most participants rely on a combination of the three strategies.

[Figure 1 about here]

477 *Discussion*

478 As in the previous experiment, we were able to quantify the degree of reliance on
 479 each of the three types of correspondences in two samples of monolingual German
 480 native speakers. Even though there is individual variation, we found, on average,
 481 almost identical reliance on the three strategies in two independent samples of German
 482 readers, suggesting that the procedure we introduced is reliable. The overall pattern of
 483 results was also broadly consistent with the findings from Experiment 1A, showing
 484 that reliance on all three types of correspondences is both necessary and sufficient to
 485 explain the vowel length pronunciations in German, and that context-insensitive
 486 correspondences are the major predictor of the vowel responses.

487 While the bilingual and monolingual participants' response patterns were similar,
 488 we did find some significant differences in terms of reliance on context-sensitive versus
 489 context-insensitive correspondences: bilingual participants show stronger reliance on
 490 context-insensitive correspondences and less reliance on context-sensitive
 491 correspondences. Two possible causes of the difference between German/English
 492 bilinguals and German monolinguals are the influence of English proficiency on reading
 493 in the bilingual sample, or a general difference in German reading proficiency.

494 According to the psycholinguistic grain size theory, if the difference in weights is due to
 495 the influence of English (L2) on the choice of correspondences in German (L1), we
 496 would expect bilinguals to rely more on larger correspondences (context-sensitive
 497 correspondences or body-rime correspondences as opposed to context-insensitive
 498 correspondences). Developmental studies have shown that reliance on larger units
 499 differs as a function of reading efficiency, as younger children rely to a greater extent

on context-sensitive rules (Treiman et al., 2006). In Experiment 1B, we found that bilingual participants rely more on context-insensitive rules, which is more in line with a proficiency explanation - bilinguals may be less proficient in reading German than monolinguals, as they are less exposed to German texts. As a result, they rely to a greater extent on the context-insensitive correspondences. ⁵

Experiment 2A

The majority of prior research on the use of GPCs, context-sensitivity and body-rime correspondences has been conducted in English. In contrast to German, the English letter-to-sound correspondence system is highly complex, as a large set of correspondences on different levels are required to describe the relationship between print and speech (Venezky, 1970). In Experiment 2, we aimed to explore whether it is possible to apply the methodology which we introduced in Experiment 1 to quantify the degree of reliance on the same three strategies in a more complex system.

English, like German, contains some context-sensitive correspondences. However, there are no super-rules, or correspondences which apply to all vowels, as in German. Therefore, we concentrated solely on the grapheme "a", as its correct pronunciation can often be disambiguated by taking into account its context. By default, "a" is pronounced as in "cat" in Australian English, but there are several context-sensitive and multi-letter GPCs that can modify its pronunciation. The context-sensitive correspondence of interest here is the correspondence that an "a" preceded by a "qu" or "w" is pronounced as /ɔ/. We chose this correspondence to assess reliance on context-sensitivity for two reasons: Firstly, previous research has shown that there is some psychological reality to this correspondence (Patterson & Behrmann, 1997; Treiman et al., 2003). Secondly, unlike other context-sensitive GPCs (e.g., "a[l]" →

524 /o:/), this correspondence is not confounded with body-rime analogy, as the modifier is
 525 located in the onset, before the vowel. This is therefore one of the few English
 526 context-sensitive correspondences that allows us to independently assess effects of
 527 context-sensitivity.

528 In order to create an item set equivalent to the German nonwords used in
 529 Experiment 1, we isolated English bodies with the vowel grapheme "a" which are
 530 consistently pronounced irregularly (Ziegler et al., 1997). There are five such bodies:
 531 "-alse", "-att", "-alk", "-alt", and "-ald". With one exception, they are confounded
 532 with the "a[l]" \rightarrow /o:/ correspondence: the body "-att" only occurs in the word "watt"
 533 and therefore only has the /ɔ/-pronunciation. As a result, and in contrast to the
 534 German experiment, the degree of reliance on body-rime correspondences cannot be
 535 assessed using this paradigm, because it is almost perfectly confounded with reliance
 536 on the "a[l]" context-sensitive correspondence.

537 In short, there are three possible pronunciations indicative of reliance on
 538 different types of correspondences. If English participants rely on context-insensitive
 539 GPCs, we should find that the majority of nonwords are pronounced with the
 540 /æ/-vowel. If context-sensitive correspondences are used, then in the conditions where
 541 a "qu" or "w" precedes the vowel we should find many /ɔ/-responses. If either
 542 body-rime correspondences or the "a[l]"-correspondence are used, the conditions with
 543 the consistently irregular bodies should be pronounced with an /o:/.

544 *Methods*

545 The participants were 19 undergraduate students at Macquarie University who
 546 were all native speakers of English.

547 We created four conditions of 18 words each (listed in Appendix A). All were

monosyllables containing the single vowel grapheme "a". The first condition was created by taking consistently regular bodies (Ziegler et al., 1997) and adding an onset which does not change the pronunciation of the vowel (i.e., any onset that does not contain "w" or "qu"), resulting in nonwords like "hact" (this condition is hereafter referred to as CS+BR+, as both the context-sensitive correspondences, CS, and the body-rime correspondences, BR, agree with the context-insensitive GPC "a" \rightarrow /æ/). The second condition (CS+BR-, e.g., "halse") was based on bodies where the "a" is consistently pronounced as /o:/ (or /ɔ/ for the body "-att"), and "normal" onsets, as in the first condition. Here, the body-rime correspondences predict an /o:/ pronunciation, and therefore disagree with the context-insensitive correspondence. The items in the third condition (CS-BR+, e.g., "wact") were based on regular bodies and onsets containing "w" or "qu", meaning that the context-sensitive "[qu,w]"a-correspondence contradicted the context-insensitive GPC while the body-rime correspondences did not. The fourth condition (CS-BR-, e.g., "qualse") had items with irregular bodies and onsets with "w" or "qu" - here both the context-sensitive correspondence and the body disagree with the context-insensitive GPC. As filler items, we used a set of unrelated nonwords.

The presentation was identical to Experiment 1, with items presented in random order and in upper case letters. As with Experiment 1, participants were instructed to read the items as accurately as possible, without putting them under time pressure.

Results

The results were scored by the fourth author (SP), a native Australian English speaker and an experienced transcriber, with the aid of spectral analysis using the EMU speech database System and associated speech analysis tools (Cassidy &

Harrington, 2001). SP was unaware of the aims of the experiment while she was transcribing the data. Unlike the German data, scoring the responses as correct or incorrect was more complicated. For the grapheme "a", there are at least five plausible pronunciations: as in "cat", as in "false", as in "what", as in "cake", and as in "car". We considered only the first three responses, as they were predicted either by the context-insensitive GPC, "a" → /æ/, the context-sensitive GPC, "[qu,w]a" → /ɔ/, or the body-rime correspondence "a[l]" → /o:/ context-sensitive correspondence. Other responses and errors made up 4.09% of the CS+BR+ condition, 24.85% of the CS+BR- condition, 6.43% of the CS-BR+ condition, and 20.76% of the CS-BR- condition, and were excluded from the subsequent analyses. The percentage of "other" responses is particularly high for the BR- conditions, partly because in English, a post-vocalic "l" creates ambiguity in the pronunciation of the vowel, such that a long /o:/ may become indistinguishable from the phoneme /əʊ/. The percentages of /æ/, /o:/ and /ɔ/ responses are presented in Table 4, with the results from Experiment 2B for comparison.

[Table 4 about here]

Modelling Vowel Pronunciations in English

The modelling strategy for Experiment 2A and 2B required a small modification from that employed in Experiments 1A and 1B. In German, there are only two available vowel pronunciations for "a": short and long. In Australian English, there are three pronunciations available for items of Experiment 2. This means that we now

592 need three equations per item:

$$\begin{aligned}
 P(\text{æ}) &= \beta_{gpc} \times GPC_{\text{æ}} + \beta_{csc} \times CSC_{\text{æ}} + \beta_{brc} \times BRC_{\text{æ}} \\
 P(\text{ɔ}) &= \beta_{gpc} \times GPC_{\text{ɔ}} + \beta_{csc} \times CSC_{\text{ɔ}} + \beta_{brc} \times BRC_{\text{ɔ}} \\
 P(o:) &= \beta_{gpc} \times GPC_{o:} + \beta_{csc} \times CSC_{o:} + \beta_{brc} \times BRC_{o:}
 \end{aligned} \tag{5}$$

where $\beta_j \in [0, 1]$ and $\Sigma \beta_j = 1$

593 where each of the subscripted strategies indicates the likelihood of the
 594 subscripted pronunciation under that strategy; for example, $GPC_{\text{æ}}$ indicates the
 595 likelihood of an /æ/ response under the GPC strategy. The end result is a set of $\hat{\beta}_i$ s
 596 that fit all three pronunciations simultaneously.

597 The weightings are shown in Table 2. The role of context-sensitive
 598 correspondences appears to be the most important in predicting the pronunciation of
 599 the grapheme "a", with, $\hat{\beta}_{csc} = 0.69$. Body-rime correspondences also appear to
 600 contribute significantly, $\hat{\beta}_{brc} = 0.26$, while the reliance on context-insensitive
 601 correspondences is very small, $\hat{\beta}_{gpc} = 0.05$. Indeed, the bootstrapping procedure
 602 produced $\hat{\beta}_{gpc} = 0$ in 43.3% of the samples, and $\hat{\beta}_{gpc} < 0.1$ in 82.0%, suggesting that
 603 the reliance on context-insensitive correspondences does not differ significantly from
 604 zero. Here again, the model is outperforming each of the independent strategies at
 605 predicting response patterns on an item by item basis (see Table 3), but when
 606 considering the model's ability to predict cell means (Table 4), it's clear this approach
 607 is less successful in English than it was in German.

608 *Discussion*

609 We quantified the reliance on different types of correspondences for English
 610 nonwords with the grapheme "a", using the same modelling technique we introduced in
 611 Experiment 1 for German, with some minor modifications. Although the results were

less clear-cut than in German, we show that the procedure can be applied to a more complex orthography. The model fits in Table 4 indicate that the English orthography is not best suited for such an analysis. In particular, the poor model fits are due to many /ɔ/-responses, even when these were not predicted by us. This may be a result of the complex phonology of English: the phonemes /ɔ/ and /o:/ are very similar, therefore it is possible that the participants had a tendency to shorten /o:/-responses, which then became indistinguishable from the vowel /ɔ/. The second possibility is that another source of information is used to determine vowel pronunciations in English which we did not take into account.

Despite these limitations, there are several conclusions that can be drawn from the results. Firstly, the weightings showed that in English the three strategies are neither necessary nor sufficient to predict the pronunciation of the grapheme "a". In contrast to German, we obtained a relatively high percentage of "other" responses for the English data, or pronunciations that were implausible according to any of the correspondences that we thought participants may use. Such a heterogeneity of nonword reading aloud responses has also been reported elsewhere (Andrews & Scarratt, 1998; Pritchard et al., 2012). While this would be an interesting topic to pursue in further research, for our purposes we discarded the unusual pronunciations as we were interested in quantifying the reliance on the same three types of correspondences we showed to be critical to nonword reading in German. This high percentage of "other" responses shows that it is likely that other strategies, such as more complex context-sensitive correspondences or lexical analogy, are used during nonword reading in English. In other words, the three types of correspondences we described in the introduction are not sufficient to explain vowel responses to the grapheme "a" in English - which is in contrast to the findings we report for German.

637 Secondly, a striking finding is that the context-insensitive correspondences are
 638 hardly used at all to derive the pronunciation of the grapheme "a". Rather, English
 639 readers rely heavily on the context-sensitive GPC, which can often be used to derive
 640 the correct pronunciation for English words.

641 These results imply that in the special case of the grapheme "a", it may not be
 642 necessary to rely on all three types of sublexical correspondences to explain the pattern
 643 of vowel responses. However, we consider it highly unlikely that context-insensitive
 644 GPCs are not used at all for reading in English. We relied solely on nonwords with the
 645 grapheme "a" to derive the weightings in Experiment 2, and its correct pronunciation
 646 can often be predicted by context. Arguably, this may falsely bias the weightings
 647 towards an apparent greater reliance on context-sensitive correspondences than we
 648 would observe if we used different graphemes for this procedure. However, we consider
 649 it likely that context-sensitivity plays an equally important role for other vowels in
 650 English: as is the case for the grapheme "a", vowel pronunciations in English are
 651 generally inconsistent, but can be often resolved context-sensitive correspondences
 652 (Treiman et al., 1995). Nonword reading studies have also provided evidence for the
 653 psychological reality of context-sensitive correspondences determining vowel
 654 pronunciation in English, other than the "[qu/w]a"-correspondence (Treiman et al.,
 655 2003, 2006). As described above, we focussed on the "[qu/w]a"-correspondence only
 656 because it is not confounded with body-rime correspondences - if we used any other
 657 context-sensitive correspondence we would be unable to distinguish it from reliance on
 658 body analogy.

659 However, we do stress that the almost exclusive reliance on context-sensitive
 660 correspondences in Experiment 2 is unlikely to generalise to the processing of more
 661 consistent graphemes in English, such as consonants. If, linguistically,

context-insensitive correspondences are generally predictive of the correct pronunciation, there is no pressure on the readers to take into account the surrounding letters for those particular graphemes.

As discussed in the introduction, the body-rime correspondences of English are confounded with context-sensitive correspondences. Instead of the German super-rules, we used an English context-sensitive correspondence that is not located in the body, namely the "[qu,w]a" \rightarrow /ɔ/ correspondence. However, we cannot fully disambiguate the reliance on body-rime correspondences and the "a[l]"-correspondence. Future studies using nonword reading should bear in mind that body-rime correspondences and context-sensitive correspondences are heavily confounded, and that an apparently irregular pronunciation of a nonword may show reliance on either context-sensitive correspondences or body-rime correspondences.

Experiment 2B

In Experiment 2B, we tested a sample of German/English bilingual speakers on the English item set. As with Experiment 1B, this will allow us to verify the weightings in a different sample, and explore potential differences between mono- and bilingual participants.

In Experiment 1, we argued that the differences that we found between the two samples are more consistent with an account based on reading proficiency rather than one based on the influence of acquiring a language with a deeper orthography. However, it may be that an early acquired L1 shapes the cognitive system in a way that biases the processing of subsequently learnt languages towards familiar types of correspondences. If so, this would predict a difference between participants reading English nonwords depending on whether their first language was English (as in

686 Experiment 2A) or German.

687 *Methods*

688 The participants were 13 native German speakers living in Australia
 689 (undergraduate and graduate students at Macquarie University, academic staff, family
 690 and friends). Eight of them had also participated in Experiment 1A several months
 691 earlier, but did not know that the two studies were related. In this sample, all
 692 participants had lived in Germany for at least 18 years before moving to an
 693 English-speaking country. The items and procedure were identical to Experiment 2A.
 694 The participants were told that they would see English nonwords, and were asked to
 695 pronounce each item as if it were an English word that they are unfamiliar with.

696 *Results*

697 The same scoring system was used as for Experiment 2A. The proportions of
 698 /æ/, /ɔ/ and /o:/ responses for both Experiment 2A and 2B are presented in Table 4.
 699 German native speakers overall gave more "other" nonword responses, or vowel
 700 responses that were inconsistent with our predictors, compared to the English
 701 monolinguals in Experiment 2A: 15.74%, 23.61%, 17.95%, and 8.80% for the CS+BR+,
 702 CS+BR-, CS-BR+ and CS-BR- conditions respectively.

703 We repeated the optimisation technique to derive strategy weights for this
 704 Experiment. Table 2 summarises the weights for each of the three strategies in
 705 Experiments 1A, 1B, 2A and 2B. The results of Experiment 2B mirror the findings
 706 from Experiment 2A: Again, we find strongest reliance on context-sensitive
 707 correspondences, robust reliance on body-rime correspondences, and negligible reliance
 708 on context-insensitive correspondences. Numerically, the reliance on context-sensitive
 709 correspondences appear to be larger ($\hat{\beta}_{csc} = 0.61$) than in the monolingual sample

710 ($\hat{\beta}_{cdc} = 0.69$). Here again, the optimal parameters outperform the alternatives with a
 711 correlation of .717 (see Table 3).

712 *Comparing bilingual to monolingual English readers.* Using the same
 713 bootstrapping technique described in Experiment 1, we confirmed that the
 714 German-English bilingual participants relied more on body-rime correspondences
 715 (BRCs) than did the English monolinguals. In 9,998 (99.98%) of the samples, $\hat{\beta}_{brc}$ was
 716 larger for bilinguals than monolinguals (95%CI of the difference: 0.046 to 0.150). The
 717 two samples did not differ significantly in their reliance on context-insensitive (GPC)
 718 rules, but there is some evidence that the monolinguals may rely more on
 719 context-sensitive correspondences (91.72% of the samples, 95%CI: -0.039 to 0.160).

720 *Discussion*

721 In Experiment 2B we collected data on English nonword pronunciation from
 722 German/English bilingual participants, which we then compared to the
 723 "a"-pronunciations of English monolinguals in Experiment 2A. Again, we find that the
 724 fits of the model are somewhat discrepant with the data, suggesting that the
 725 pronunciation of the letter "a" depends also on sources of information that are not
 726 included in our model. As in Experiment 2A, we found no reliance on
 727 context-insensitive GPCs in either group, and only a nonsignificant trend towards
 728 larger reliance on body-rime correspondences or the "a[l]" \rightarrow /o:/ correspondence in
 729 English monolinguals than the German/English bilinguals.

730 We found broadly the same pattern among two different groups of participants;
 731 here, we once again demonstrate the reliability of the optimisation procedure. The
 732 significant difference in the reliance on body-rime correspondences suggest that
 733 German native speakers, even when they are highly proficient in English, continue to

734 rely less on these large units than English monolingual participants. Thus, the native
 735 orthography appears to leave small but noticeable footprints in the cognitive processes
 736 underlying reading in a second language.

737 **General Discussion**

738 In four experiments, we explored the reliance on three different sublexical
 739 correspondence types in different populations. In Experiments 1A and 1B, we found
 740 that German native speakers relied on all three strategies: the greatest weighting was
 741 found for context-insensitive GPCs, followed by context-sensitive GPCs (super-rules)
 742 and body-rime correspondences when reading German-derived nonwords. In
 743 Experiments 2A and 2B, we applied the same procedure to quantify the types of
 744 correspondences that participants rely on to derive the pronunciation of the grapheme
 745 "a" in English. We found strong reliance on context-sensitive GPCs, some reliance on
 746 body-rime correspondences, and little evidence that context-insensitive GPCs play a
 747 large role in determining the pronunciation of the grapheme "a".

748 *Cross-Linguistic Differences in the Choice of Sublexical Correspondences: Comparing* 749 *Experiments 1 and 2*

750 Previous theoretical work predicts cross-linguistic differences in the reliance on
 751 different units in German and English (Ziegler & Goswami, 2005). Unfortunately, with
 752 the experiments in the current study it is impossible to make a direct quantitative
 753 comparison across the two languages as we are comparing two differently structured
 754 orthographic correspondences. An alternative approach is to conduct the analyses
 755 within the languages and point out the differences between them on a descriptive level.

756 Our data suggest that given a grapheme where context is very important in
 757 English (i.e., "a"), context-sensitivity becomes very important compared to German,

758 where context-insensitive correspondences are the major predictor. This is true even
 759 for a situation where there are statistical regularities at the level of context-sensitive
 760 correspondences. This is broadly in line with the psycholinguistic grain size theory
 761 (Ziegler & Goswami, 2005): as the context is often an important predictor of the
 762 correct pronunciation of English words, readers are forced to rely on larger units. Our
 763 data emphasises the importance of context-sensitive GPCs in an inconsistent
 764 orthography such as English. In German, on the other hand, context-insensitive
 765 correspondences are mostly sufficient to derive the correct pronunciation of an
 766 unfamiliar word, therefore this level of correspondences is preferred.

767 The reality of the cross-linguistic differences becomes more evident in a
 768 comparison of Experiments 1A and 2B. This is partly a within-subject design, and
 769 involves bilingual participants reading both the English and the German item sets.
 770 The differences between the weightings in these two experiments were remarkable, with
 771 the pattern of results being more similar to that of the monolinguals of the respective
 772 language. This shows that the language is the determining factor for the reliance on
 773 different unit sizes, rather than the language background of the participants.

774 From this comparison, we conclude that the language that a participant is asked
 775 to read in matters more than the participant's language background: comparing the
 776 participants in Experiments 1A and 2B shows that bilinguals rely on the three types of
 777 correspondences almost to the same extent as monolinguals do in their respective
 778 language. Thus, we conclude that the cross-linguistic differences in sublexical
 779 processing are language-specific: acquiring a deep versus shallow orthography from
 780 childhood does not shape the cognitive system, but rather encourages the reader to
 781 rely on certain types of correspondences above others in that particular orthography.
 782 Those preferences do not seem directly transferable to a later acquired orthography;

783 instead, a reader develops a sensitivity to the most advantageous combination of
 784 strategies in the new language.

785 *Models of Reading*

786 The current study shows that both in English and in German, several
 787 correspondence types are used in parallel. There are multiple verbal models that
 788 postulate such a scenario (LaBerge & Samuels, 1974; Patterson & Morton, 1985; Taft,
 789 1991, 1994; Ziegler & Goswami, 2005). The theoretical contribution of the current
 790 paper is proposing a method to quantify the degree to which these are used, which can
 791 be used as a benchmark for computational models.

792 An open question then is whether the current computational models can
 793 simulate the obtained results. The parallel processing of various correspondences poses
 794 a computational problem: whenever there are conflicts between the pronunciations
 795 predicted by various correspondences, the system needs a way to resolve these. In
 796 English, this is important, because there are often cases where different sublexical
 797 correspondences provide conflicting information.

798 In Table 5, we provide the percentages of regular responses from two models
 799 which have been implemented both in English and in German, namely the DRC
 800 (Coltheart et al., 2001; Ziegler et al., 2000) and the CDP+ (Perry, Ziegler, & Zorzi,
 801 2007; Perry, Ziegler, Braun, & Zorzi, 2010). For English, there is a newer version of the
 802 CDP+, namely the CDP++ (Perry, Ziegler, & Zorzi, 2010), which differs from the
 803 CDP+ in several points: it has been trained on a larger word set, contains some
 804 parameter changes, and can also deal with polysyllabic words. We provide the
 805 simulation data from both versions of the model.

[Table 5 about here]

Both the CDP+/CDP++ and the DRC are dual route models of reading, where nonwords are read purely via a sublexical procedure. Therefore, the current data are relevant to both models, as it concerns the nature of sublexical processing. The distinguishing feature between the two models is the way in which this procedure operates. The DRC has a set of sublexical GPCs, which are manually programmed into the sublexical route. A GPC in the DRC is defined as the most frequent phoneme that co-occurs with a given grapheme. As described in the introduction, the DRC contains context-sensitive correspondences as well as single-letter and multi-letter correspondences, but there is some ambiguity when it comes to deciding which context-sensitive correspondences to include in the model. The current version of the English DRC does not contain either a "[w]a"- or an "a[l]"-correspondence, therefore it provides the response /æ/ to all items (see Table 5). For the second DRC simulation, we added some more context-sensitive correspondences, however this does not seem to reflect the overall responses given by participants, either, as it now underestimates the number of regular (i.e., /æ/) pronunciations given by the participants. For the German DRC, the GPCs that are used to determine vowel length are the super-rules (Ziegler et al., 2000). It is clear, both from the present study (see Table 5) and from Perry, Ziegler, Braun, and Zorzi (2010) that the super-rules are not sufficient to explain German nonword pronunciations.

The CDP+/CDP++, like the DRC, is grapheme-based, but it develops context-sensitivity because the grapheme-to-phoneme correspondences are derived via a learning algorithm, which uses real word knowledge to obtain the most likely correspondences between print and speech (Zorzi, 2010). Yet, the CDP+ does not provide an optimal fit for either the German or the English data, as it often underestimates the number of regular pronunciations (see Table 5). In particular, the

English CDP+ and CDP++ seem to take context-sensitive correspondences into account more than the participants do, as they underestimates the number of /æ/-responses for the CS- conditions. In German, the biggest discrepancy between the CDP+ prediction and the behavioural data is in the BR- conditions, suggesting that CDP+ does not develop the same degree of reliance on body-rime correspondences that participants do.

As neither of the computational models is compatible with the behavioural results, these data cannot be used to adjudicate between the DRC and CDP+ approach. (Note that this was not the aim of the study to begin with.) We therefore turn to verbal models to provide a theoretical framework that can explain our obtained results. One such model which provides a means for the cognitive system to resolve conflicts between different sublexical correspondences has been proposed by Taft (1991, 1994). This interactive activation model states that activation passes hierarchically from the smallest units, through subsyllabic and syllabic units and morphemes to whole words, which then gives access to the semantic concept. There are additional feedback connections, which send activation from larger to smaller units.

Taft's (1994) model also makes some explicit statements about cross-linguistic differences: the salient sublexical correspondences differ depending on the orthographic and phonological properties of the language. For example, while English readers parse words into orthographic-syllabic units called BOSSes (Taft, 1979, 1992), French readers rely more on the phonological syllable (Taft & Radeau, 1995). In our experiments we found reliance on similar types of correspondences in English and German. Thus, the correspondences that have psychological reality in English and German appear to be very similar. It is noteworthy that English and German are very similar in terms of their phonological and orthographic structure, therefore we expect that the salient

sublexical correspondences do not differ greatly. The situation might be different in other languages. For example, when there is a tendency for words to be polysyllabic and to contain fewer consonant clusters, as is the case in languages like Italian, Spanish, or Russian, body-rime correspondences are unlikely to play a large role in reading (Duncan et al., 2013; Kerek & Niemi, 2012).

Limitations and Future Directions

The goal of the study was to identify an optimal combination of different sources of information in deciding which vowel pronunciation is most appropriate when there are two or more alternatives. A limitation of the model is that it makes no claims about the decision-making mechanisms that resolve the ambiguity, only that some sources of information are more influential than others. It may be that on each trial, the decision is based on a "winning strategy" in which case the weights represent the likelihood of a particular strategy winning. Alternately, it may be that all three sources of information are combined in a Bayesian sense of "what response is most likely correct given the mix of influences." In this case the model weights should be interpreted as the degree of influence that each strategy has on the decision process. The present study is not able to adjudicate between these alternatives (or any others that we may not have considered), so we refrain from making strong statements favouring one or the other. The extent to which nonword pronunciations remain stable in different situations, the factors that influence any variability, and the mechanisms that resolve ambiguity remain questions for future research. We do note, however that while there is considerably variability between subjects in terms of their strategy weights (see Figure 1), there is some recent evidence that readers can be grouped according to their choices (Robidoux & Pritchard, 2014), so there may be more

880 structure hiding within this variability.

881 A limitation of the paradigm as described in this paper is that it is better suited
 882 for across-subject comparisons than across-item comparisons, due to the small number
 883 of available items. This is a general problem with this approach: there are not many
 884 items where context-sensitive correspondences and body-rime correspondences can be
 885 dissociated, as these are intrinsically correlated. While it would be interesting to use
 886 the same paradigm for a different set of nonword or word items to explore systematic
 887 changes in the weightings associated with item characteristics such as frequency (for
 888 words) or word-likeness (as measured, e.g., by orthographic N), the small number of
 889 possible items prevents us from doing this in a meaningful way.

890 Arguably, the data reported in this paper are also limited by our focus on the
 891 grapheme "a" only. While this criticism applies to the English data, the German data
 892 can be generalised to predicting vowel length across different graphemes. The English
 893 results, and our conclusions based on these analyses, are therefore weaker than those
 894 from the German analyses. Nevertheless, understanding the principles underlying
 895 reading in languages other than English is essential for the long-term goal of describing
 896 all differences and similarities between reading in different languages, and thereby
 897 creating a universal model of reading (Frost et al., 2012). This is especially important
 898 given the focus of previous literature on English. English is considered to be an
 899 "outlier" orthography, therefore it is questionable to use it as a base for most models of
 900 skilled reading, reading development, and dyslexia (Share, 2008). While we
 901 acknowledge that, in the current context, the optimisation procedure works better for
 902 German than English, we argue that the English data provides a strong demonstration
 903 of the parallel use of different types of sublexical grain sizes, and in particular
 904 context-sensitive correspondences in English, new insights into cross-linguistic

905 differences associated with the reliability of print-to-speech correspondences, and a new
 906 benchmark for computational models of reading aloud.

907 We believe that this approach also has some utility when applied to other areas
 908 of psycholinguistics. In future research, the same paradigm can be used to
 909 systematically explore the sources of individual differences that we report in the
 910 current study. The paradigm can also be used with children: previous literature has
 911 debated for decades whether children start learning to read using large or small units
 912 first (Goswami, 2002; Goswami & Bryant, 1990; Hulme et al., 2002). Such explorations
 913 in group and individual differences are of theoretical and practical value. Future
 914 research can also apply the same mathematical procedure to any situation in which
 915 items can be created where different strategies yield different predictions. Other areas
 916 in psycholinguistics to which this paradigm can be extended could be topics such as
 917 stress assignment for polysyllabic words, because it has been shown that, in several
 918 languages, different cues are used by participants to determine the stress of a given
 919 nonword (Arciuli, Monaghan, & Seva, 2010; Burani & Arduino, 2004; Protopapas,
 920 Gerakaki, & Alexandri, 2006; Ševa, Monaghan, & Arciuli, 2009).

921 *Conclusions*

922 The current study contributes to the literature on cognitive processes underlying
 923 reading in several aspects. We show that context-insensitive GPCs, super-rules and
 924 body-rime correspondences are necessary and sufficient to explain the vowel length
 925 pronunciations in German; in English, context-insensitive GPCs play a smaller or
 926 negligible role in assigning the pronunciation of the grapheme "a". We introduce a
 927 method to quantify the degree of reliance on each of the three different sublexical
 928 correspondence types using statistical modelling. This technique can be used to test

929 other hypotheses by future studies.

References

- Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent?
Memory & Cognition, 10(6), 565-575.
- Andrews, S., & Scarratt, D. R. (1998). Rule and analogy mechanisms in reading
 nonwords: Hough dou peapel rede gnew wirts? *Journal of Experimental
 Psychology: Human Perception and Performance*, 24(4), 1052.
- Arciuli, J., Monaghan, P., & Seva, N. (2010). Learning to assign lexical stress during
 reading aloud: Corpus, behavioral, and computational investigations. *Journal of
 Memory and Language*, 63(2), 180-196.
- Brown, G. D., & Deavers, R. P. (1999). Units of analysis in nonword reading:
 Evidence from children and adults. *Journal of Experimental Child Psychology*,
 73(3), 208-242.
- Burani, C., & Arduino, L. S. (2004). Stress regularity or consistency? Reading aloud
 Italian polysyllables with different stress patterns. *Brain and Language*, 90(1),
 318-325.
- Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for
 bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5),
 1190-1208.
- Campbell, R., & Besner, D. (1981). This and THAP: constraints on the pronunciation
 of new, written words. *The Quarterly Journal of Experimental Psychology*,
 33(4), 375-396.
- Cassidy, S., & Harrington, J. (2001). Multi-level annotation in the Emu speech
 database management system. *Speech Communication*, 33(1), 61-77.
- 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*

- 955 *Review*, 108(1), 204.
- 956 Cortese, M. J., & Simpson, G. B. (2000). Regularity effects in word naming: What are
957 they? *Memory & Cognition*, 28(8), 1269–1276.
- 958 Cox, F., & Palethorpe, S. (2007). Australian English. *Journal of the International*
959 *Phonetic Association*, 37(03), 341–350.
- 960 Duncan, L. G., Castro, S. L., Defior, S., Seymour, P. H., Baillie, S., Leybaert, J., et al.
961 (2013). Phonological development in relation to native language and literacy:
962 Variations on a theme in six alphabetic orthographies. *Cognition*, 127(3),
963 398–419.
- 964 Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with
965 millisecond accuracy. *Behavior Research Methods*, 35(1), 116–124.
- 966 Frost, R., Behme, C., Beveridge, M. E., Bak, T. H., Bowers, J. S., Coltheart, M., et al.
967 (2012). Towards a universal model of reading. *Behavioral and Brain Sciences*,
968 35(5), 263.
- 969 Glushko, R. J. (1979). The organization and activation of orthographic knowledge in
970 reading aloud. *Journal of Experimental Psychology: Human Perception and*
971 *Performance*, 5(4), 674.
- 972 Goswami, U. (2002). In the beginning was the rhyme? A reflection on Hulme, Hatcher,
973 Nation, Brown, Adams, and Stuart (2002). *Journal of Experimental Child*
974 *Psychology*, 82(1), 47–57.
- 975 Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read*. London:
976 Wiley Online Library.
- 977 Goswami, U., Gombert, J. E., & De Barrera, L. F. (1998). Children's orthographic
978 representations and linguistic transparency: Nonsense word reading in English,
979 French, and Spanish. *Applied Psycholinguistics*, 19, 19–52.

- 980 Goswami, U., Porpodas, C., & Wheelwright, S. (1997). Childrens orthographic
 981 representations in English and Greek. *European Journal of Psychology of*
 982 *Education*, 12(3), 273–292.
- 983 Goswami, U., Ziegler, J. C., Dalton, L., & Schneider, W. (2003). Nonword reading
 984 across orthographies: How flexible is the choice of reading units? *Applied*
 985 *Psycholinguistics*, 24(2), 235–247.
- 986 Grömping, U. (2010). Inference with linear equality and inequality constraints using
 987 R: The package ic.infer. *Journal of Statistical Software*. (Forthcoming)
- 988 Hulme, C., Hatcher, P. J., Nation, K., Brown, A., Adams, J., & Stuart, G. (2002).
 989 Phoneme awareness is a better predictor of early reading skill than onset-rime
 990 awareness. *Journal of Experimental Child Psychology*, 82(1), 2–28.
- 991 Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming.
 992 *Journal of Memory and Language*, 46(4), 723–750.
- 993 Kerek, E., & Niemi, P. (2012). Grain-size units of phonological awareness among
 994 Russian first graders. *Written Language & Literacy*, 15(1), 80–113.
- 995 LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information
 996 processing in reading. *Cognitive Psychology*, 6(2), 293–323.
- 997 Monfort, A. (1995). *Statistics and econometric models* (Vol. 2). Cambridge, UK:
 998 Cambridge University Press.
- 999 Patterson, K., & Behrmann, M. (1997). Frequency and consistency effects in a pure
 1000 surface dyslexic patient. *Journal of Experimental Psychology: Human Perception*
 1001 *and Performance*, 23(4), 1217.
- 1002 Patterson, K., & Morton, J. (1985). From orthography to phonology: A new attempt
 1003 at an old interpretation. In K. Patterson, J. Morton, & M. Coltheart (Eds.),
 1004 *Surface dyslexia* (pp. 1217–1231). Hillsdale, NJ: Erlbaum.

- 1005 Perry, C., Ziegler, J. C., Braun, M., & Zorzi, M. (2010). Rules versus statistics in
 1006 reading aloud: New evidence on an old debate. *European Journal of Cognitive*
 1007 *Psychology*, 22(5), 798–812.
- 1008 Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the
 1009 development of computational theories: the CDP+ model of reading aloud.
 1010 *Psychological Review*, 114(2), 273.
- 1011 Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale
 1012 modeling of reading aloud with the connectionist dual process (CDP++) model.
 1013 *Cognitive Psychology*, 61(2), 106–151.
- 1014 Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996).
 1015 Understanding normal and impaired word reading: computational principles in
 1016 quasi-regular domains. *Psychological Review*, 103(1), 56.
- 1017 Pritchard, S. C., Coltheart, M., Palethorpe, S., & Castles, A. (2012). Nonword
 1018 reading: Comparing dual-route cascaded and connectionist dual-process models
 1019 with human data. *Journal of Experimental Psychology: Human Perception and*
 1020 *Performance*, 38(5), 1268–1288.
- 1021 Protopapas, A., Gerakaki, S., & Alexandri, S. (2006). Lexical and default stress
 1022 assignment in reading Greek. *Journal of Research in Reading*, 29(4), 418–432.
- 1023 R Core Team. (2013). R: A language and environment for statistical computing
 1024 [Computer software manual]. Vienna, Austria. Available from
 1025 <http://www.R-project.org/>
- 1026 Robidoux, S., & Pritchard, S. C. (2014). Hierarchical clustering analysis of reading
 1027 aloud data: a new technique for evaluating the performance of computational
 1028 models. *Frontiers in Psychology*, 5.
- 1029 Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of

- word recognition and naming. *Psychological Review*, 96(4), 523.
- Ševa, N., Monaghan, P., & Arciuli, J. (2009). Stressing what is important: Orthographic cues and lexical stress assignment. *Journal of Neurolinguistics*, 22(3), 237–249.
- Share, D. L. (2008). On the anglocentricities of current reading research and practice: the perils of overreliance on an “outlier” orthography. *Psychological Bulletin*, 134(4), 584.
- Taft, M. (1979). Lexical access-via an orthographic code: The basic orthographic syllabic structure (BOSS). *Journal of Verbal Learning and Verbal Behavior*, 18(1), 21–39.
- Taft, M. (1991). *Reading and the mental lexicon*. Hillsdale, NJ: Psychology Press.
- 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(4), 1004.
- Taft, M. (1994). Interactive-activation as a framework for understanding morphological processing. *Language and Cognitive Processes*, 9(3), 271–294.
- Taft, M., & Radeau, M. (1995). The influence of the phonological characteristics of a language on the functional units of reading: A study in French. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 49(3), 330.
- Treiman, R., Kessler, B., & Bick, S. (2003). Influence of consonantal context on the pronunciation of vowels: A comparison of human readers and computational models. *Cognition*, 88(1), 49–78.
- Treiman, R., Kessler, B., Zevin, J. D., Bick, S., & Davis, M. (2006). Influence of consonantal context on the reading of vowels: Evidence from children. *Journal of*

- 1055 *Experimental Child Psychology*, 93(1), 1–24.
- 1056 Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The
 1057 special role of rimes in the description, use, and acquisition of English
 1058 orthography. *Journal of Experimental Psychology: General*, 124(2), 107.
- 1059 Venezky, R. L. (1970). *The structure of English orthography* (Vol. 82). The Hague:
 1060 Mouton.
- 1061 Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia,
 1062 and skilled reading across languages: a psycholinguistic grain size theory.
 1063 *Psychological Bulletin*, 131(1), 3.
- 1064 Ziegler, J. C., Perry, C., & Coltheart, M. (2000). The DRC model of visual word
 1065 recognition and reading aloud: An extension to German. *European Journal of*
 1066 *Cognitive Psychology*, 12(3), 413–430.
- 1067 Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical words are read
 1068 differently in different languages. *Psychological Science*, 12(5), 379–384.
- 1069 Ziegler, J. C., Perry, C., Ma-Wyatt, A., Ladner, D., & Schulte-Körne, G. (2003).
 1070 Developmental dyslexia in different languages: Language-specific or universal?
 1071 *Journal of Experimental Child Psychology*, 86(3), 169–193.
- 1072 Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). What is the pronunciation for
 1073 -ough and the spelling for/u/? a database for computing feedforward and
 1074 feedback consistency in English. *Behavior Research Methods*, 29(4), 600–618.
- 1075 Zorzi, M. (2010). The connectionist dual process (CDP) approach to modelling reading
 1076 aloud. *European Journal of Cognitive Psychology*, 22(5), 836–860.

Appendix A: German and English nonwords used in Experiments 1 and 2

German

V[C] Reg. blaf blen (*blem in Exp. 1B*) blod breg brel brul flom flüb fryp grät
grem grom grul klid klur knul krel kril krön pflyp pid plät plön prod schmün schraf
schwüb speg zwül zwun

V[C]/[C] Reg. bamt birt blaft bling boft brals chrolf falb flarg flerk gärm ginn
gralb gunt kall kaxt kerv kluns knell pals peld pfern pulk purf schern spalf stelf sturg
zeng zwurt

V[C] Irreg. bax blex blig bres flim flis git glef glip krex krin krip pfis spic stef
zwix zwok

V[C]/[C] Irreg. bags blags füst gleks kagd kagt kets pagt pard peks poks schagd
stard

English

CS+BR+. hangst kazz mact phadge phamb phangst phants plact sangst slangs
slazz stract stramb tamb tazz tradge trazz zants

CS+BR-. clatt hald halse kalk kalse kalt phalk phaltz slaltz strald stralk stralse
straltz tald taltz tralse tralt tratt

CS-BR+. quadge quamb quangst quapse quazz squact squazz swact swangst
swants swazz twadge twangst twants twazz wact wamb wangst

CS-BR-. qualk qualse qualtz squald squalk squalse squaltz swalk swaltz twald
twalk twalse twalt twaltz wald walse walt whald

Appendix B: Implementing the fitting process in R

While fitting the models described in the text has a certain flavour of regression to it, there are some important differences. Most critically are the two constraints that we have placed on the parameters: $\beta_j \in [0, 1]$, and $\sum \beta_j = 1$. Considerable work has been done to develop and implement estimation methods for models with inequality constraints such as $\beta_j \in [0, 1]$ (Grömping, 2010). However, we know of no such work that has solved the problems presented by the $\sum \beta_j = 1$ constraint. To address this problem, we turned to the `optim` function that is part of the base statistical analysis package in R (R Core Team, 2013). `optim` is a very general optimisation package that allows the user to minimise any specified function, while also placing bounds on the returned values. That is, we can define a function, place upper and lower bounds on the returned weights, and `optim` will efficiently search the allowed parameter space to minimise our function. To satisfy (2), we defined the minimising function to be the residual sum of squares, and restricted the β weights to fall between 0 and 1. This ensures that $\hat{\beta}_j \in [0, 1]$ is satisfied.

In all of the optimisation analyses, we used the following command in R:

```
optim(par=runif(3, .2, .8), fn=..., ..., method='L-BFGS-B',
      control=list(factr=1e5), lower=0, upper=1)
```

The parameters for `optim` operate as follows: "`par=runif(3, .2, .8)`" initialises the β_i s to random values between .2 and .8. "`fn=..., ...`" specifies the function to be minimised along with any parameters it requires. In our case we used a simple function that calculates the residual sum of squares. "`method='L-BFGS-B'`" instructs `optim` to use an optimisation algorithm that allows for upper and lower bounds on the returned values (Byrd, Lu, Nocedal, & Zhu, 1995). "`factr=1e5`" sets the convergence tolerance, and "`lower=0, upper=1`" set the bounds on the returned

1124 values.

1125 $\Sigma\beta_j = 1$ *Constraint*. There is no way to explicitly tell **optim** to meet the
 1126 constraint that the β s must sum to 1 ($\Sigma\beta_j = 1$). One way to ensure that the constraint
 1127 is met is to simply scale the weights returned by **optim** using the formula:

$$\beta'_j = \beta_j / \Sigma\beta_j \quad (6)$$

1128 where β'_j are the new scaled weights, and are guaranteed to sum to 1. However, since
 1129 this adjustment *follows* the optimisation process, there is little reason to believe that
 1130 the resulting β'_j s would remain an optimal solution to (2).

An alternative to simply scaling the β_j s, is to make use of the influence of outliers on parameter estimation. For example, according to (2) **optim** is trying to satisfy the following 180 equations (two per item) simultaneously, by minimising the residual sum of squares (while also meeting the $\beta_j \in [0, 1]$ constraint):

$$P_1(Short) = \beta_{gpc} \times GPC_{short,1} + \beta_{csc} \times CSC_{short,1} + \beta_{brc} \times BRC_{short,1}$$

$$P_1(Long) = \beta_{gpc} \times GPC_{long,1} + \beta_{csc} \times CSC_{long,1} + \beta_{brc} \times BRC_{long,1}$$

...

$$P_{90}(Short) = \beta_{gpc} \times GPC_{short,90} + \beta_{csc} \times CSC_{short,90} + \beta_{brc} \times BRC_{short,90}$$

$$P_{90}(Long) = \beta_{gpc} \times GPC_{long,90} + \beta_{csc} \times CSC_{long,90} + \beta_{brc} \times BRC_{long,90}$$

1131 The introduction of a new data point that can only be met by satisfying the constraint
 1132 that the $\Sigma\beta_j = 1$ will put some pressure on **optim** to select appropriate parameters.
 1133 For example,

$$1 = \beta_{gpc} \times 1 + \beta_{csc} \times 1 + \beta_{brc} \times 1 \quad (7)$$

1134 (7) is equivalent to creating an artificial data point where all of the dependent and
 1135 independent variables [P(Short), GPC, CSC, and BRC] are set to 1. Though (7)

1136 provides some pressure to satisfy $\Sigma \hat{\beta}_j = 1$, it is unlikely to have a very large influence
 1137 since it is only a single equation with roughly equal weight to the other 180. However,
 1138 dramatically increasing the weight of this data point will exert a much stronger
 1139 influence on the final parameter selection. For example,

$$10000 = \beta_{gpc} \times 10000 + \beta_{csc} \times 10000 + \beta_{brc} \times 10000 \quad (8)$$

1140 would put enormous pressure on `optim` to arrive at a set of weights that satisfy
 1141 $\Sigma \hat{\beta}_j = 1$ without putting any further constraints on how the weights are apportioned to
 1142 the strategies. Though (8) does not guarantee $\Sigma \hat{\beta}_j = 1$ precisely, it is sufficiently strong
 1143 for the present purposes. Other applications may require a larger multiplier.

Finally, because the number of items is not equal across all conditions in our studies, the sums of squares were weighted by item to ensure each condition contributed equally. For example in Experiment 1, items in the V[C] Irregular and V[C][C] Irregular conditions received relatively more weight than items in the V[C] Regular and V[C][C] Regular conditions. If this isn't done, there is a tendency for the Regular items to have a stronger influence on the eventual parameters. The weights applied to each item were determined as follows:

$$\omega_{type} = \frac{.25}{n_{type}}$$

where *type* is one of the four item types (e.g., V[C][C] Irregular in Experiment 1), ω_{type} is the weight assigned to items of that type, and n_{type} is the total number of items of that type. As this formula implies, each item contributes equally to the influence of its category, but items in smaller categories have more influence than items in larger categories. These weights are then used in the usual weighted sum of squares formula

that optim is trying to minimise:

$$SS_{resid} = \sum_i (\hat{Y}_i - \bar{Y}_i)^2 \omega_{type_i}$$

Author Note

1144

1145 We are grateful to Angela Heine and Tila Brink for collecting the Berlin data for
1146 Experiment 1B. We also thank Petra Schienmann and Reinhold Kliegl for their help
1147 with organising data collection at Potsdam University for Experiment 1B. We thank
1148 Johannes Ziegler for providing a list of German consistent words. Further thanks are
1149 due to Stephen Lupker, James Adelman, and three anonymous reviewers for their
1150 helpful comments on earlier versions of this paper.

1151 This article was written as part of XS's doctoral dissertation under supervision
1152 of EM, AC, and MC. SR conducted the data analyses and contributed to the write-up
1153 and revision of the manuscript, and SP scored the English data.

1154 Correspondence concerning this article should be addressed to Xenia Schmalz,
1155 Department of Cognitive Science, Macquarie University. Email:
1156 xenia.schmalz@mq.edu.au; phone: +61 (0)2 9850 2992; fax: +61 (0)2 9850 6059

Footnotes

1157

1159 ¹It is not always true that graphemes are smaller (i.e., contain fewer letters than)
 1160 bodies, e.g., the grapheme "igh" is larger than the body of the word "cat" ("-at"). For
 1161 the sake of clarity, we follow the terminology of Ziegler and Goswami (2005) and refer
 1162 to graphemes as small units, and bodies as large units.

1163 ²There are some differences associated with dialects. Here, we use the
 1164 pronunciations given by the DRC's vocabulary and the Macquarie Essential Dictionary
 1165 (5th Edition) as representative of Australian English, and the IPA as illustrated by
 1166 Cox and Palethorpe (2007).

1167 ³Though we refer to the reliance on different types of correspondences as a
 1168 "strategy", we do not mean to imply that readers *consciously* choose the type of
 1169 correspondence that maximises the chance of correctly reading an unfamiliar word.

1170 ⁴In standard linear regression, only one of these two formulae would be required,
 1171 since they are entirely dependent (i.e., $P_i(Long) = 1 - P_i(Short)$, etc...). In traditional
 1172 regression, the only difference between the first and second equations would be the
 1173 location of the estimated intercept and the sign of the slope. However, by removing the
 1174 intercept term, our modelling strategy undermines this interdependence. Since the
 1175 intercept is not free to vary (it is forced to be 0) the parameter estimates for P(short)
 1176 would not match those for P(long). As a result, we must simultaneously fit both vowel
 1177 pronunciations. While it is useful to use the language of regression to describe some of
 1178 the procedures, it is very important to remember that the β s here do not represent
 1179 regression slopes, but weights. Also, if this were a regression problem, it would be more
 1180 properly treated as a *logistic* regression problem. However, this would be incompatible
 1181 with our interpretation of the weights as "the probability that a certain strategy is
 1182 adopted."

1183 ⁵It is noteworthy that Perry, Ziegler, Braun, and Zorzi (2010) report data with a
 1184 similar set of nonwords to the current study (though the study was conducted with
 1185 different aims): the authors manipulated the number of consonants in the coda, but
 1186 rather than controlling for the consistency of the base-word, their nonwords differed in
 1187 terms of the existence of the body in real words: the body either occurred in real
 1188 German words, or it did not. In other words, they did not independently manipulate
 1189 the predictions of body-rime correspondences and context-sensitive correspondences,
 1190 and predictions of super-rules and body analogy were heavily correlated,
 1191 $r(39) = 0.78, p < 0.001$, as were the predictions of super-rules and GPCs,
 1192 $r(39) = 0.51, p < 0.001$. This means that the Perry et al. data is unsuitable for our
 1193 purposes: the analysis would be unreliable, as it is impossible to disentangle reliance
 1194 on bodies versus super-rules, and super-rules versus GPCs.

Table 1

Percentage of Short Vowel Responses for Each Condition in Experiment 1, and the average predictions from each of the three types of correspondences

Responses	V[C] Regular	V[C][C] Regular	V[C] Irregular	V[C][C] Irregular
<i>Example</i>	"Wal" → "bral"	"Wald" → "brald"	"Bus" → "brus"	"Magd" → "bragd"
% Short 1A	47.25	83.69	84.63	61.04
% Short 1B	37.28	86.79	72.95	62.84
<i>Correspondence Predictions</i>				
P(Short GPC)	70.21	79.53	90.59	78.77
P(Short CSC)	26.2	92.57	62.82	91.38
P(Short BRC)	2.76	100.00	100.00	0.00
<i>Model Predictions</i>				
% Short 1A	44.68	87.04	87.83	60.53
% Short 1B	36.58	89.78	83.67	61.70

Note: GPC = context-insensitive GPC; CSC = Context-sensitive correspondences; BRC = Body-rime correspondence

Table 2

Weightings for the three types of correspondences in Experiments 1A, 1B, 2A and 2B

Correspondence type	1A (German bil.)	1B (German mon.)	2A (English mon.)	2B (English bil.)
GPC	0.56	0.38	0.05	0.03
CSC	0.19	0.35	0.69	0.61
BRC	0.26	0.27	0.26	0.36

Table 3

Summary of the fits between the models and the observed response proportions. Each value is the correlation between the predictions from the GPC, CSC, BRC or model and the observed response pattern

Sample	GPC	CSC	BRC	Optimal (95%CI)
1A (German bil.)	.714	.681	.540	.844 (.830, .847)
1B (German mon.)	.578	.730	.659	.827 (.812, .832)
2A (English mon.)	.522	.630	.385	.729 (.719, .731)
2B (English bil.)	.514	.573	.568	.792 (.785, .793)

Table 4

Summary of vowel responses of English monolinguals (2A) and German/English bilinguals (2B), predictions from the three types of correspondences (context-independent GPCs; context-sensitive correspondences; body rhyme correspondences), and predictions from the model using the weights in Table 2.

Experiment	Responses	CS-BR-	CS+BR+	CS-BR+	CS+BR-
Example		"qualk"	"hangst"	"quadge"	"hald"
<i>Participant Responses</i>					
2A - Monolinguals	%æ	8.12	96.20	76.04	39.18
	%ɔ	60.25	0.00	17.19	27.19
	%o:	10.63	0.00	0.88	8.77
2B - Bilinguals	%æ	8.07	83.33	62.50	41.20
	%ɔ	38.24	0.93	19.91	27.31
	%o:	52.19	0.00	0.00	7.87
<i>Correspondence Predictions</i>					
GPC	P(æ GPC)	72.00	72.00	72.00	72.00
	P(ɔ GPC)	5.00	5.00	5.00	5.00
	P(o: GPC)	6.00	6.00	6.00	6.00
CSC	P(æ CSC)	29.00	77.00	29.00	77.00
	P(ɔ CSC)	47.00	0.00	47.00	0.00
	P(o: CSC)	0.00	0.00	0.00	0.00
BRC	P(æ BRC)	0.00	100.00	100.00	0.00
	P(ɔ BRC)	0.00	0.00	0.00	0.00
	P(o: BRC)	100.00	0.00	0.00	100.00
<i>Model Predictions</i>					
2A	%æ	22.46	82.98	49.07	56.37
	%ɔ	33.35	0.14	33.35	0.14
	%o:	26.77	0.16	0.16	26.77
2B	%æ	18.88	85.34	55.39	48.84
	%ɔ	29.38	0.05	29.38	0.05
	%o:	36.57	0.07	0.07	36.57

Table 5

Percentages of "Regular" responses (/æ/ in English, short vowels in German) given by the DRC and CDP+/CDP++

Model	CS+BR+	CS-BR+	CS+BR-	CS-BR-
English Behavioural Data	100	81	51	18
English DRC Simulation 1	100	100	100	100
English DRC Simulation 2	100	67	11	0
English CDP +	100	35	57	0
English CDP++.	100	43	44	0
German Behavioural Data	86	73	63	37
German DRC	100	0	100	0
German CDP+	93	94	8	24