Different orthographies represent spoken language in different ways. An important aspect that differs across orthographies is how consistently letters map onto sounds. In relatively consistent orthographies, such as those of Serbo-Croatian, Italian, and German, letters or letter groups map relatively consistently onto sounds. Conversely, in a relatively inconsistent orthography, such as English orthography, the relation between letters and sounds is often equivocal: Some letters or letter clusters can be pronounced in more than one way, and some sounds can be spelled in more than one way (Ziegler, Stone, & Jacobs, 1997). Whether these kinds of cross-language differences influence the way languages are processed and represented is a central question in psycholinguistics (e.g., Frost, Katz, & Bentin, 1987; Gleitman, 1985; Katz & Feldman, 1983; Katz & Frost, 1992) and more recently in brain-imaging studies (Fiez, 2000; Paulesu et al., 2000). Are all written languages essentially processed the same way (universal hypothesis) or are there orthography-specific aspects that need to be taken into consideration?

Historically, this issue was first investigated with regard to the orthographic depth hypothesis (ODH; e.g., Katz & Feldman, 1983). The ODH suggests that readers adapt their processing strategies to the demands of the orthography they are reading. It assumes that there are at least two pathways to the mental lexicon, a phonological pathway that relies on the assembly of letters into sounds and a direct orthographic pathway. In a consistent language (shallow orthography), readers are encouraged to use the phonological pathway because the mapping between letters and sounds is relatively direct and unambiguous. In contrast, in an inconsistent language (deep orthography), readers should be reluctant to use the phonological pathway because of the less systematic mapping between spelling and sound. Instead, they should rely to a greater extent on the direct orthographic pathway. Thus, according to the ODH, the overall consistency of a language is thought to determine the specific mix of phonological and orthographic codes (Katz & Frost, 1992).

Although this hypothesis was supported by a wealth of intriguing cross-language research (e.g., Frost & Katz, 1989; Frost et al., 1987; Katz & Feldman, 1983; Katz & Frost, 1992), it has not remained unchallenged (for reviews, see Besner & Smith, 1992, and Seidenberg, 1992). Most important, it has been argued that the effects of orthographic depth on processing are not as simple as the original hypothesis predicted. For example, although the ODH predicted that phonology effects should be reduced in a relatively inconsistent orthography such as English orthography, a large number of studies found strong phonology effects in English (e.g., Rayner, Sereno, Lesch, & Pollatsek, 1995; Van Orden, 1987). Although Katz and Frost (1992) correctly pointed out that such data challenge only the strong version of the ODH, according to which people who read deep orthographies never use phonological information, it nevertheless seems that phonological processes play a role in both consistent and inconsistent orthographies. As original advocates of the ODH recently concluded: “We no longer believe that the difference (between shallow and deep orthographies) is one of whether or not phonology is routinely involved in visual word recognition” (Lukatela & Turvey, 1999, p. 1069). We agree with this statement and suggest that orthographic consistency may affect not so much the relative contribution of phonology (i.e., the specific mix between orthographic and phonological pathways), but rather the very nature of the phonological processes themselves.

Such a view is supported by a number of developmental studies that all agree that phonological recoding, that is, the ability to translate printed words into their spoken equivalents, underlies successful reading acquisition in all orthographies (e.g., Frith, Wimmer, & Landerl, 1998; Goswami, Gombert, & De Barrera, 1998; Share, 1995). What seems to differ between orthographies is not the amount of phonological recoding required but rather the nature of the phonological recoding process. That is, children who are learning to read more orthographically consistent languages, such as Greek, German, and Spanish, appear to rely heavily on grapheme-phoneme decoding strategies. In these languages, phonological recoding can reliably operate at the smallest grain size because the mapping of graphemes onto phonemes is relatively unambiguous. In contrast, children who are learning to read less orthographically consistent languages, like English, use a variety of decoding strategies, supplementing grapheme-phoneme conversion strategies with both the recognition of letter patterns for rhymes and attempts at whole-word recognition (Frith et al., 1998; Goswami et al., 1998; Goswami, Porpdas, & Wheelwright, 1997; Goswami, Ziegler, Dalton, & Schneider, in press; Landerl, Wimmer, & Frith, 1997; Wimmer & Goswami, 1994).

If the reading system develops differently as a function of the consistency of the orthography, there should be traces of such differential development even in skilled adult reading. That is, phonological recoding in an inconsistent orthography should require processing at a
Reading Across Languages

A variety of grain sizes, including large grain sizes such as bodies and rhymes. The resulting lexical representations should be shaped to represent orthography and phonology at multiple levels. In contrast, in a consistent orthography, phonological recoding should be based mainly on the processing of small grain sizes, such as graphemes and phonemes, and the resulting lexical representations should show less variability. Thus, our working hypothesis was that consistency of an orthography should have a measurable effect on the grain size of units that are likely to play a role during reading and reading development.

We tested the grain-size hypothesis by comparing word and nonword reading in German and English. The German-English comparison is ideal for this purpose because the two languages have similar orthographies and phonologies but differ quite dramatically in the consistency of orthography-phonology correspondences (Frith et al., 1998; Jacobs, in press; Landerl, 1997; Landerl et al., 1997; Ziegler, Perry, & Coltheart, 2000). For example, the words ball, park, and hand exist in both languages in identical form. However, although the grapheme “a” receives the same pronunciation in all three words in German, it receives a different pronunciation in each word in English. In theory, this is exactly the kind of inconsistency at the grapheme level that leads English readers to consider larger units that help reduce inconsistencies at the grapheme level (Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995). Thus, English readers should show stronger large-unit effects, whereas German readers should show stronger small-unit effects. As a marker for large-unit processing, we used body-rhyme effects because in order to show these effects readers have to process units that are bigger than individual graphemes. As a marker for small-unit processing, we used length effects. When the processing unit is small, more chunks need to be processed, and hence reading times should increase with increasing word or nonword length.

In the present study, we tested this prediction in a strong way by manipulating word length and body neighborhood (body-N) using almost-identical word pairs (i.e., cognates) in English and German (De Groot & Nas, 1991). Cognates are words that look similar, sound similar, and have the same meaning in two languages (e.g., sand in English vs. Sand in German). The body-N consists of words that share the same orthographic rhyme (e.g., late, date, fate are body-Ns of hate). In addition to the cognates, we used nonwords in the two languages that were mostly identical (e.g., lank vs. Lamp). The nonwords varied with respect to stimulus length and body-N in the same way as the words. This design is illustrated in Table 1.

The hypothesis of interest was whether stimulus length and body-N would produce different effects across languages when literally identical items were used in the two languages. Finding such cross-language differences would support the hypothesis that very basic word recognition processes can differ across languages. The present cognate design reduces potential confounds that may occur when different items are used in different languages. An additional strength of the present design is that words like zoo, sand, and start not only look, sound, and mean the same in the two languages, but also are probably acquired at approximately the same time in reading development. Finally, we employed only words that are regular in both languages. This is important because English has a higher proportion of irregular words than German, and it would not have been a fair comparison if irregular words were used in only one language.

In sum, the goal of the present study was to investigate whether orthographic consistency in different languages affects the preferred grain size of processing units that are involved in reading. The specific prediction was that for the same type of words (e.g., sand-Sand) and nonwords (e.g., lank-Lank), naming times of skilled adult German readers would show stronger length effects than naming times of skilled adult English readers. At the same time, we expected that English readers would show stronger body-N effects than their German counterparts.

### METHOD

#### Participants

Fifty-three undergraduate psychology students participated in the study. Thirty were native German speakers recruited at the Catholic University of Eichstätt, Germany. Twenty-three were native Australian-English speakers recruited at Macquarie University in Sydney, Australia.

#### Design and Stimuli

As illustrated in Table 1, the experimental conditions resulted from a factorial combination of language (English vs. German), stimulus length (three, four, five, and six letters), body-N (large vs. small), and lexicality (words vs. nonwords). Stimulus selection was based on the CELEX database, which exists for both German and English (Baayen, Piepenbrock, & Rijn, 1993). In a first run, all German-English cognates were selected from the CELEX database. Pairs were then considered for further selection if (a) both words were monosyllabic, (b) both words had regular grapheme-phoneme correspondences, (c) the

<table>
<thead>
<tr>
<th>Table 1. Basic design and examples of items used in each category</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Body neighborhood (body-N) and length</strong></td>
</tr>
<tr>
<td>English</td>
</tr>
<tr>
<td>Large body-N</td>
</tr>
<tr>
<td>3 letters</td>
</tr>
<tr>
<td>4 letters</td>
</tr>
<tr>
<td>5 letters</td>
</tr>
<tr>
<td>6 letters*</td>
</tr>
<tr>
<td>Small body-N</td>
</tr>
<tr>
<td>3 letters</td>
</tr>
<tr>
<td>4 letters</td>
</tr>
<tr>
<td>5 letters</td>
</tr>
<tr>
<td>6 letters*</td>
</tr>
</tbody>
</table>

*Six-letter words were not real cognates (i.e., they had different meanings).
two words had the same number of letters, and (d) the two words belonged to the same body-N class (either large or small).

These constraints resulted in the selection of 80 monosyllabic German-English word pairs. Half (n = 40) had a small body-N (3.0 and 3.7 for German and English, respectively), and half had a large body-N (9.6 and 12.9 for German and English, respectively). Each word pair belonged to one of four orthographic-length groups (three, four, five, and six letters), and each group had the same number of items (n = 20). Three-letter words had on average 2.5 phonemes, four-letter words had 3.5 phonemes, five-letter words had 4.2 phonemes, and six-letter words had 4.4 phonemes. The stimuli in the two languages were matched in terms of number of phonemes (3.7 vs. 3.6 phonemes), and hence number of grapheme-phoneme correspondences. For the six-letter words, it was not possible to find monosyllabic German-English cognates. Thus, for the six-letter group, we used words that had comparable word frequencies and were orthographically and phonologically similar in the two languages (e.g., flight-Frucht).

Mean frequency for the German-English word pairs ranged between 2 and 1,035 per million, with an average frequency of 100 per million according to CELEX frequency counts that are available for both English and German. To make sure that the word-length manipulation was not contaminated by differences in word frequency, we matched word frequency across the different word lengths (112, 92, 87, and 108 per million, for three-, four-, five-, and six-letter words, respectively). Similarly, to ensure that potential body-N effects could not result from differences in word frequency, we matched large body-N words and small body-N words with respect to word frequency (112 vs. 88 per million, respectively).

The nonword manipulation paralleled the word manipulation. That is, 80 nonwords were used in each language. Most of the nonwords were identical across the two languages (e.g., fot-Fot, lank-Lank, plock-Plock, etc.). When this was not possible because of body-N constraints, we made sure that the nonwords were orthographically and phonologically similar as possible (e.g., ler-Lir, sibe-Seib, meast-Miest). Half of the nonwords (n = 40) had a small body-N (2.5 and 2.9 for German and English, respectively), and half had a large body-N (8.5 and 10.9 for German and English, respectively). One quarter of the nonwords were in each orthographic-length group (three, four, five, or six letters).

### Procedure

In both Australia and Germany, a standard reading-aloud procedure was used. That is, items were presented visually in the center of a computer monitor. The same typography and font size were used in the two locations. Participants were asked to name the stimuli as quickly as possible. Naming latency was calculated from the onset of each stimulus until the triggering of the voice key. In Australia, naming responses were recorded using a Gateway Pentium-333 computer and a Realistic microphone. In Germany, a Umax Pulsar computer and a Vivanco EM 32 condensor microphone were used.

### RESULTS

Mean naming latencies (reaction times, or RTs) for both German and English are presented in Table 2. Because error rates were fairly low (1.8% and 3.7% in English and German, respectively) and evenly distributed across conditions, they are not discussed further. In the German sample, the data from 3 participants were excluded because of a large number of voice-key errors (above 20%). An additional participant was excluded because of a large overall error rate (16%). For the remaining participants, voice-key errors (e.g., anticipations, mouth clicks) were excluded from the analysis (1.94% and 1.71% of the German and English data, respectively). In addition, RTs that were three standard deviations above or below the individual grand mean were excluded from the analysis (less than 1.1% of the data).

The RT data were submitted to a $2 \times 4 \times 2 \times 2$ analysis of variance (ANOVA) that resulted from the factorial combination of language (English vs. German), length (three vs. four vs. five vs. six letters), body-N (many vs. few body-Ns), and lexicality (word vs. nonword). In the subject analysis ($F_S$), language was a between-subjects factor whereas length, body-N, and lexicality were within-subjects factors. In the item analysis ($F_I$), language was a within-items factor whereas length, lexicality, and body-N were between-items factors.

The ANOVA exhibited significant effects of length, $F_S(3, 141) = 87.19, p < .0001$, and $F_I(3, 143) = 33.72, p < .0001$: of body-N, $F_S(1, 47) = 21.41, p < .0001$, and $F_I(1, 143) = 4.61, p < .05$; and of lexicality, $F_S(1, 47) = 95.13, p < .0001$, and $F_I(3, 143) = 256.83, p < .0001$. The language effect was not significant by subjects but was significant by items, $F_S(1, 47) = 0.129, p > .70$, and $F_S(3, 143) = 4.20, p < .05$. This pattern of main effects shows that in both languages, shorter items were named faster than longer items (length effect), items with small body-Ns were named more slowly than items with large body-Ns (body-N effect), and words were named faster than nonwords (lexicality effect). The German participants were slightly slower than the English participants (576 vs. 565 ms, respectively), although this difference failed to reach significance in the subject analysis.

These main effects were qualified by theoretically important interactions between language and length, $F_S(3, 141) = 15.77, p < .0001$, and $F_S(3, 143) = 4.38, p < .01$, and between language and body-N, $F_S(3, 47) = 7.74, p < .01$, and $F_S(1, 143) = 1.7, p < .20$. As illustrated in Figure 1, the cross-language interaction with length resulted from the fact that the length effect was stronger in German than in English. As illustrated in Figure 2, the cross-language interaction with body-N indicates that the body-N effect was stronger in English than in Ger-

### Table 2. Mean naming latencies (in milliseconds) for words and nonwords in German and English

<table>
<thead>
<tr>
<th>Body neighborhood (body-N) and length</th>
<th>English Words</th>
<th>Nonwords</th>
<th>German Words</th>
<th>Nonwords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large body-N</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 letters</td>
<td>508</td>
<td>568</td>
<td>512</td>
<td>563</td>
</tr>
<tr>
<td>4 letters</td>
<td>505</td>
<td>588</td>
<td>501</td>
<td>565</td>
</tr>
<tr>
<td>5 letters</td>
<td>515</td>
<td>622</td>
<td>568</td>
<td>621</td>
</tr>
<tr>
<td>6 letters</td>
<td>526</td>
<td>628</td>
<td>572</td>
<td>687</td>
</tr>
<tr>
<td>Mean</td>
<td>513</td>
<td>601</td>
<td>538</td>
<td>609</td>
</tr>
<tr>
<td>Small body-N</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 letters</td>
<td>526</td>
<td>578</td>
<td>524</td>
<td>568</td>
</tr>
<tr>
<td>4 letters</td>
<td>527</td>
<td>618</td>
<td>521</td>
<td>588</td>
</tr>
<tr>
<td>5 letters</td>
<td>528</td>
<td>623</td>
<td>566</td>
<td>637</td>
</tr>
<tr>
<td>6 letters</td>
<td>534</td>
<td>654</td>
<td>563</td>
<td>657</td>
</tr>
<tr>
<td>Mean</td>
<td>528</td>
<td>618</td>
<td>543</td>
<td>612</td>
</tr>
</tbody>
</table>

J.C. Ziegler et al.
Fig. 1. Naming times in English and German as a function of word and nonword length.

Fig. 2. Naming times in English and German as a function of the size of the body neighborhood (body-N).

Weickes (1997) showed that in English, length effects for words disappeared when N was partialed out. We thus repeated our item analyses treating N as a covariate. The results showed that the length effect and the critical interaction between language and length were still significant when N was partialed out, $F(3, 286) = 11.28, p < .0001,$ and $F(3, 286) = 4.45, p < .01,$ respectively. As in our previous analyses, the triple interaction among the effects of lexicality, language, and length was still not significant ($F < 1$), suggesting that German readers showed stronger length effects than English readers for both words and nonwords.

The second analysis investigated whether body-N effects persisted if N and length were partialed out. For this purpose, we conducted simple and partial regression analyses. In the simple regression analysis in English, body-N had a significant facilitatory effect on RTs ($r = .175, p < .05$). When N and length were partialed out, this correlation even increased ($r = -.221, p < .005$). In German, however, there was no significant correlation between body-N and RTs in either analysis (both $p_s > .52$). We performed a similar analysis for N. In English, N had a small facilitatory effect on RTs that failed to reach significance ($r = -.119, p > .10$). When body-N and length were partialed out, the facilitatory effect of N turned into an inhibitory effect ($r = .181, p < .05$). In German, the situation was quite different. N had a strong facilitatory effect on RTs in both simple and partial regression analyses ($r = -.383, p < .005,$ and $r = -.161, p < .05,$ for simple and partial correlations, respectively).

**DISCUSSION**

The present results showed theoretically important cross-language differences in the effects of word length and body-N. As predicted by the grain-size hypothesis, the length effect was stronger in German than in English, suggesting more small-unit processing in German. In contrast, the body-N effect was stronger in English than in German, suggesting more large-unit processing in English. The striking aspect of the present

---

3. Because it was not possible to find six-letter German-English word pairs that were true cognates (i.e., identical meaning across languages), an additional analysis was carried out using only the data from the three-, four-, and five-letter items, which were in fact true cognates. This analysis basically replicated the main pattern. Most important, the interaction between language and length was still significant by both participants and items, $F(1, 22) = 17.46, p < .001,$ and $F(1, 144) = 6.13, p < .05,$ although it was only marginally significant for subjects and not significant for items in German, $F(1, 25) = 3.14, p < .10,$ and $F(1, 143) = 0.17, p > .60.$

4. N is a neighborhood measure that stands for the number of words that can be obtained by changing a single letter at any position (see Coltheart, Davelaar, Jonasson, & Besner, 1977).
findings is that these cross-language dissociations were obtained although participants named literally identical words and nonwords (zoo-Zoo, sand-Sand, etc.). Thus, the data suggest that identical items are processed differently by readers of different orthographies. The difference seems to be one relating to the grain size of units involved in reading.

A variety of post hoc analyses confirmed that length effects in German and body-N effects in English were still significant even when N was partialled out. If we further assume that N is a good index for small-unit processing whereas body-N is a good index for large-unit processing, then the correlational analyses provide additional support for the grain-size hypothesis. That is, body-N had a facilitatory effect on reading latencies in English, which remained significant even when N and length were partialled out. In contrast, body-N had no significant effect on German. Interestingly, N had a strong facilitatory effect in German, which remained significant when body-N and length were partialled out, whereas the effects of N in English were weak and changed from facilitation to inhibition once the large-unit contribution of body-N was partialled out. This finding is consistent with previous work suggesting that the effects of N in English are weak or even inhibitory once large-unit effects of body-N are controlled (Ziegler & Perry, 1998).

The present results go beyond earlier cross-language investigations for two reasons. First, previous research often looked at absolute differences in reading speed or reading accuracy between readers of different orthographies. For example, recently, Paulesu et al. (2000) reported that Italian university students were faster at word and nonword reading than their English counterparts. However, such results are based on comparing absolute reading speed across different groups of participants and items. Although one can try to match subjects and items across languages, absolute differences in reading speed or reading accuracy between different groups of subjects and items can be due to a variety of factors, not only orthographic consistency. In interpreting the results of our study, we do not have to argue in terms of absolute differences in reading speed or accuracy because only the difference in the size of theoretically relevant marker effects is of interest.

Second, previous research was based mainly on the idea that reading different orthographies varies with respect to the weight that is given to orthographic versus phonological processes (Frost et al., 1987; Katz & Feldman, 1983; Landerl et al., 1997). This view was criticized because processing in different orthographies appeared to be more similar than initially suggested (Besner & Smith, 1992; Lukatela & Turvey, 1999; Seidenberg, 1992). Our results underscore the possibility that the amount of orthographic, phonological, and semantic processing might be fairly similar across orthographies. What seems to differ is the size of the dominant processing units, the number of different grain-size levels, and the reader’s flexibility to switch between these different levels. In our study, for example, larger units were preferred for German than in English whereas smaller units were preferred in German. Thus, orthographic consistency appears to determine the very nature of the orthographic and phonological processes and not only the relative contribution of orthographic and phonological codes.

The present cross-language dissociations might have implications for current computational models of reading. Models with a fixed grain size of orthography phonology correspondences might be forced into adapting more flexible coding schemes (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). Other models may need to be evaluated for their capacity to capture the present cross-language dissociations simply as a product of statistical learning in different linguistic environments (e.g., Zorzi, Houghton, & Butterworth, 1998). Finally, it is important to keep in mind that finding evidence for a preferential use of one particular grain size does not mean that processing at other grain sizes does not occur. In fact, readers show considerable flexibility in their use of spelling-to-sound correspondences at different grain sizes (Brown & Deavers, 1999), and small grapheme-size reading units can play a role even in a language like English (Rey, Ziegler, & Jacobs, 2000). However, when cross-language variance is reduced by comparing identical items across languages, the differential use of different grain sizes in different languages becomes clearly evident.

REFERENCES


Gleitman, L.R. (1985). Orthographic resources affect reading acquisition—if they are used. Remedial and Special Education, 6, 24–36.


5. This assumption seems justified given that N correlates quite strongly with position-specific letter frequency (Ziegler, Rey, & Jacobs, 1998).
Reading Across Languages


(Received 7/19/00; Revision accepted 12/12/00)