Received: 30 November 2018 | Revised: 14 November 2019 | Accepted: 4 February 2020
DOI: 10.1111/desc.12952

PAPER

Developmental Science WILEY

Orthographic consistency influences morphological processing in reading aloud: Evidence from a cross-linguistic study

Petroula Mousikou[1,2] | Elisabeth Beyersmann[3,4] | Maria Ktori[5] | Ludivine Javourey-Drevet[4,6] | Davide Crepaldi[5] | Johannes C. Ziegler[4] | Jonathan Grainger[4] | Sascha Schroeder[1,2]

[1]Max Planck Institute for Human Development (MPIB), Berlin, Germany
[2]Department of Educational Psychology, University of Göttingen, Göttingen, Germany
[3]Department of Cognitive Science and Macquarie University Centre for Reading, Macquarie University, Sydney, NSW, Australia
[4]Laboratoire de Psychologie Cognitive, Centre National de la Recherche Scientifique (CNRS), Aix-Marseille University, Marseille, France
[5]International School for Advanced Studies (SISSA), Trieste, Italy
[6]Laboratoire Apprentissage, Didactique, Évaluation, Formation, Aix-Marseille University, Marseille, France

Correspondence
Petroula Mousikou, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany.
Email: mousikou@mpib-berlin.mpg.de

Funding information
European Research Council (ERC) Starting Grant, Grant/Award Number: 679010 (STATLEARN); Deutsche Forschungsgemeinschaft, Grant/Award Number: SCHR-1268/2–1; Agence Nationale de la Recherche, Grant/Award Number: ANR-11-IDEX-0001-02, ANR-11-LABX-0036 (BLRI) and ANR-15-FRAL-0003-01; Australian Research Council, Grant/Award Number: DE190100850

Abstract
The present study investigated whether morphological processing in reading is influenced by the orthographic consistency of a language or its morphological complexity. Developing readers in Grade 3 and skilled adult readers participated in a reading aloud task in four alphabetic orthographies (English, French, German, Italian), which differ in terms of both orthographic consistency and morphological complexity. English is the least consistent, in terms of its spelling-to-sound relationships, as well as the most morphologically sparse, compared to the other three. Two opposing hypotheses were formulated. If orthographic consistency modulated the use of morphology in reading, readers of English should show more robust morphological processing than readers of the other three languages, because morphological units increase the reliability of spelling-to-sound mappings in the English language. In contrast, if the use of morphology in reading depended on the morphological complexity of a language, readers of French, German, and Italian should process morphological units in printed letter strings more efficiently than readers of English. Both developing and skilled readers of English showed greater morphological processing than readers of the other three languages. These results support the idea that the orthographic consistency of a language, rather than its morphological complexity, influences the extent to which morphology is used during reading. We explain our findings within the remit of extant theories of reading acquisition and outline their theoretical and educational implications.

KEYWORDS
cross-linguistic, morphology, orthographic consistency, reading acquisition

1 | INTRODUCTION

Forming links between spoken and written language provides the foundation for learning to read. However, learning the print-to-sound relationships in a given orthography is not sufficient to become a skilled reader. Children also need to learn to map print onto meaning in order to recognize words quickly, reliably, and efficiently (Nation, 2009). How might this process occur? Morphemes, the minimal linguistic units with a lexical or a grammatical meaning (Booij, 2012), are thought to play an important role in reading acquisition. Critically, morpheme identification facilitates word recognition (see Amenta & Crepaldi, 2012), thus enabling skilled reading.

Despite its importance for the development of skilled reading, morphology has been neglected even in the most recent and prominent theoretical conceptualizations of reading acquisition (e.g. Perry, Zorzi, & Ziegler, 2019; Ziegler, Perry, & Zorzi, 2014). The empirical evidence shows that as children move beyond the first stages of learning to read, their ability to reflect on and manipulate the morphological structure of words, known as morphological awareness (Carlisle, 1995), starts to influence their reading (e.g. Carlisle & Stone, 2005; Mann, 2000; Nagy, Carlisle, & Goodwin, 2014; Verhoeven & Perfetti, 2003). However, less is known about the extent to which children use morphological knowledge during *online* reading, which is critical for advancing theories of reading development.

Studies in this domain show that children between the ages of 7 and 11 tend to analyse multi-morphemic words based on their constituent morphemes. This has been demonstrated in several alphabetic orthographies, including English (e.g. Beyersmann, Grainger, & Castles, 2019), French (e.g. Quémart, Casalis, & Duncan, 2012), German (e.g. Hasenäcker, Schröter, & Schroeder, 2017), Dutch (e.g. Perdijk, Schreuder, Baayen, & Verhoeven, 2012), Italian (e.g. Burani, Marcolini, De Luca, & Zoccolotti, 2008), and Spanish (e.g. Lázaro, Acha, de la Rosa, García, & Sainz, 2017). However, qualitative differences in the processing of morphological information have been observed across studies and languages. Such differences could be due to variations in the grade and age of the participants involved in the various studies, given that the influence of morphological knowledge on reading seems to be modulated by reading proficiency and experience (Andrews & Lo, 2013; Beyersmann, Grainger, Casalis, & Ziegler, 2015); or to the use of substantially different materials and tasks. Alternatively, it is likely that language characteristics modulate the extent to which morphology is used during online reading (Plaut & Gonnerman, 2000). In the present study, we tested this idea using four alphabetic orthographies with different characteristics, namely, English, French, German, and Italian. Investigating this issue is critical for understanding the universal and language-specific processes involved in reading acquisition (Frost, 2012; Share, 2008).

Research Highlights

- We investigated whether the orthographic consistency of a language or its morphological complexity influences morphological processing in developmental and skilled reading.
- A reading aloud task was used in four alphabetic orthographies that differ in orthographic consistency and morphological complexity (i.e. English, French, German, Italian).
- Developing and skilled readers of English, the least consistent and most morphologically sparse language, showed greater morphological processing than readers of the other three languages.
- Our findings suggest that the orthographic consistency of a language, and not its morphological complexity, influences the extent to which morphology is used in reading.

1.1 | Orthographic depth

Alphabetic orthographies differ with regard to how consistently written orthographic symbols (e.g. graphemes or letters) map onto speech units (e.g. phonemes), a factor known as orthographic depth (Liberman, Liberman, Mattingly, & Shankweiler, 1980). In transparent or shallow orthographies (e.g. German, Greek, Finnish, Italian), a given grapheme/letter is almost always pronounced in the same way across different contexts and words. In opaque or deep orthographies (e.g. English, French), the same grapheme/letter may receive alternative pronunciations depending on its context and position within a word (e.g. the English grapheme *ou* in *tough*, *though*, *through*, *bough*, *thorough*).

As far as the languages examined in the present study are concerned, English is deep, French is thought to be intermediate, German is shallow, and Italian is very shallow (see Goswami, Gombert, & de Barrera, 1998; Paulesu et al., 2001; Seymour, Aro, & Erskine, 2003; Sprenger-Charolles, Siegel, Jiménez, & Ziegler, 2011; Tabossi & Laghi, 1992). This classification is further supported by a study that sought to quantify the consistency of spelling-to-sound relations at the word onset level in seven alphabetic orthographies, including those of the present study (see Figure 1 in the study by Borgwaldt, Hellwig, & De Groot, 2005). Importantly, word-onset consistency is considered a valid index of a language's overall orthographic transparency (Ziegler et al., 2010). It is worth mentioning, nevertheless, that French is a special case, because while orthographic consistency estimates denote inconsistent phoneme-to-grapheme correspondences in this language (0.60), grapheme-to-phoneme correspondences (GPCs) are quite consistent (0.89), with values closer to 1 indicating more consistency (Caravolas & Kessler, 2016).[1] Accordingly, French is thought to be a predictable orthography (Schmalz, Marinus, Coltheart, & Castles, 2015).

Two theories of reading, the Orthographic Depth Hypothesis (Katz & Frost, 1992) and Psycholinguistic Grain Size Theory (Ziegler & Goswami, 2005) have been put forward to explain how orthographic depth may affect reading processes.

1.2 | Orthographic Depth Hypothesis and Psycholinguistic Grain Size Theory

The Orthographic Depth Hypothesis postulates that the use of phonology should be more prevalent when reading in a shallow orthography than when reading in a deep orthography, because the consistency of GPCs in the former makes the phonological representation of a printed word available to the reader at less cost when its phonology is assembled. In contrast, the inconsistency of grapheme-to-phoneme relationships in deep orthographies encourages the reader to focus on the visual-orthographic structure of printed words, which could be effectively done by referring to their morphology. Critically, according to the Orthographic Depth Hypothesis, during the course of reading acquisition, readers of deep orthographies would shift their reliance from phonological codes to orthographic lexico-semantic codes. From this viewpoint, only when readers have well-established lexical representations can they use a visual-orthographic semantic reading mechanism. The Orthographic Depth Hypothesis is consistent with the idea that different alphabetic orthographies afford different reading mechanisms or strategies (Seidenberg, 2011).

The Psycholinguistic Grain Size Theory was developed to explain cross-language variation in reading acquisition, namely, that children learning to read in a deep orthography lag behind children learning to read in a shallow orthography (e.g. Seymour et al., 2003). According to this theory, while readers of shallow orthographies can reliably use GPCs to pronounce words correctly, readers of deep orthographies need to rely on larger orthographic units, such as syllables, rimes, or even whole words to assign correct pronunciations. This is because smaller grain sizes tend to be more inconsistent than larger grain sizes in deep orthographies (Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995). There are many more orthographic units to learn when the grain size is large than when it is small, thus slowing down the rate of reading acquisition in deep orthographies compared to shallow orthographies. Even though morphemes are not mentioned in the original description of the Psycholinguistic Grain Size Theory, morphological units are thought to bring an important degree of consistency to orthographies that are characterized by inconsistency in the mapping between spelling and sound (Ulicheva, Harvey, Aronoff, & Rastle, 2018). It follows then from this theory, that readers of deep orthographies are likely to rely on morphemes to the same extent they rely on other sublexical units such as syllables and rimes when reading aloud (see also Goswami & Ziegler, 2006, who acknowledge that morphology should be given a greater role in Psycholinguistic Grain Size Theory). Critically, according to the Psycholinguistic Grain Size Theory, there is no shift during the reading development from phonological to lexico-semantic processing as a function of the consistency of the writing system. All readers have to go through an orthography-phonology mapping (phonological decoding), but they do so using different grain sizes.

With regard to our study, we predicted that based on the two theories outlined above, English readers should show overall more robust morphological processing than readers of French, German, and Italian. However, according to the Orthographic Depth Hypothesis, greater reliance on morphemes via a visual-orthographic reading mechanism should be apparent only in skilled, and not in developing readers of English, who just like the developing readers of the other three languages should show a preference for a phonological reading mechanism. In contrast, according to the Psycholinguistic Grain Size Theory, both developing and skilled adult readers of English should rely more on morphemes than on smaller grain sizes in reading aloud.

1.3 | Morphological complexity

Languages vary with regard to their morphological complexity. English is thought to be morphologically sparse, whereas French, German, and Italian are typically classified as morphologically rich languages (Rey-Debove, 1984; Roelcke, 1997; Talamo & Celata, 2011). Accordingly, deep orthographies seem to have simple inflectional morphology (e.g. English), whereas shallow orthographies tend to have complex inflectional morphology (e.g. German, Finnish, Italian, Serbo-Croatian). French appears to fall in the middle in this case, as its inflectional morphology is not as complex as in most shallow orthographies, but also not as simple as in English (Seidenberg, 2011). Attempts to quantify morphological complexity across languages (for a review, see Borleffs, Maassen, Lyytinen, & Zwarts, 2017) reveal that according to the three main morphological complexity methods used in the literature, namely, Linguistica (Bane, 2008), Juola (1998, 2008), and type-token ratio (TTR; Kettunen, 2014), English is the least morphologically complex, followed in increasing order by German, French, and Italian (Linguistica), or Italian, French, and German (Juola), or French, Italian, and German (TTR). An empirical question that arises is whether the use of morphology during reading depends on the morphological complexity of a language. We would expect that readers might be more sensitive to the morphological structure of printed letter strings in morphologically rich languages (e.g. French, German, Italian) than in morphologically sparse languages (e.g. English). Such sensitivity might be more prominent in skilled adult readers than in developing readers, because of greater exposure of the former to the characteristics of their language.

1.4 | Previous studies

To our knowledge, only one study has implicitly tested the above hypotheses cross-linguistically (Casalis, Quémart, & Duncan, 2015). In that study, greater morphological processing was observed in French than in English developing readers, suggesting that the use of morphology in reading depends on the morphological complexity of a language. One limitation of that study was that the items in the different conditions were not matched on psycholinguistic variables that are known to influence reading processes. Also, the stems in the nonword items were often modified within and across languages inconsistently. This is problematic, because children seem to process morphologically complex words with modified stems differently than words with preserved stems (Lázaro, García, & Burani, 2015). We took these issues into consideration when constructing the stimuli for the present study.

1.5 | Present study

Conducting cross-linguistic research is challenging, insofar as both within- and across-language factors need to be taken into account. One common strategy is to use materials that are as similar as possible across the languages under examination (Frith, Wimmer, & Landerl, 1998; Ziegler, Perry, Jacobs, & Braun, 2001). Thus, we chose translation-equivalent nouns, which often happened to be cognates, either in some or all of the languages. These were used for the construction of morphologically structured and non-morphologically structured nonwords, which were the focus of the present study. Four conditions were created: Stem + Suffix (e.g. nightness), Stem + Non-Suffix (e.g. nightlude), Non-Stem + Suffix (e.g. nishtness), and Non-Stem + Non-Suffix (e.g. nishtlude). The advantage of this design is that it allowed us to investigate how the presence of a stem or a suffix in printed letter strings may independently influence reading aloud processes, as well as how these may interact during reading aloud. To avoid the use of strategic reading processes, such as focusing exclusively on sublexical units during nonword reading, morphologically simple, and morphologically complex words were also included in the study.

Our aim was to investigate the processes that are at play when developing readers encounter new words with familiar units (i.e. morphemes). To simulate the situation that children face in natural reading we presented the nonwords intermixed with words. We used the reading aloud task and focused on morphologically structured and non-morphologically structured nonwords, because nonword reading aloud provides an index of children's decoding skills independently of their word knowledge (Castles, Rastle, & Nation, 2018). The study was carried out with typically developing readers from Australia, France, Germany, and Italy, who attended Grade 3. We chose children in this grade, because compared to French, German, and Italian children, who typically reach 80%–90% nonword reading accuracy by the end of Grade 1, English-speaking children only start to reach similar levels of accuracy by Grade 3 (see Cossu, Gugliotta, & Marshall, 1995; Frith et al., 1998; Goswami et al., 1998; Landerl, 2000; Sprenger-Charolles, Siegel, & Bonnet, 1998; Wimmer & Goswami, 1994). Also, most words (60%–80%) that children encounter in third grade English texts tend to be morphologically complex (Anglin, 1993; Nagy & Anderson, 1984). Such proportions are likely to be higher in more morphologically productive languages. Critically, third graders are thought to be sensitive to the morphological characteristics of their language (Mann & Singson, 2003). Children in all four countries had roughly the same age. To test the predictions of the opposing theoretical accounts with regard to potential morphological processing differences as a function of reading experience, we also tested skilled adult readers on the same task in all four languages.

2 | EXPERIMENT

2.1 | Method

2.1.1 | Participants

A total of 126 children (30 Australian, 32 French, 32 German, and 32 Italian) in Grade 3 participated in the study for a small gift. French and German children were randomly selected from a larger sample that participated in an independent longitudinal project on the role of morphology in reading development. The selection criteria for these children were that (a) they were tested between February and March (to ensure that testing times were comparable across all languages—see below), and (b) their reading aloud accuracy was above 50%. Australian and Italian children were recruited for the purposes of the present study. Six Australian children achieved below 50% accuracy and were excluded, leaving a total of 24 to be included in the analyses. German, French, and Italian children were tested between February and May. Australian children were tested between September and October of the same year (given that the start of the school year in Australia is in February). Therefore, data collection in all countries took place after the first half of the third school year. Children in Australia started to receive formal reading instruction in the second half of the first school year, known as kindergarten (between ages five and six). In France, some reading instruction starts in the last year of école maternelle (at the age of five), which corresponds to kindergarten. In Germany and Italy, children start to receive reading instruction in the first grade (at the age of six).

A total of 128 adults (32 Australian, 32 French, 32 German, and 32 Italian) participated in the study for monetary compensation. Although studying at the university was not a requirement for participating in the study, most adult participants were university students. Both children and adult participants were native speakers of their respective languages, had normal or corrected-to-normal vision, and reported no hearing, reading, or language difficulties. Participants' age and gender, as well as other demographic information, are shown in Table 1. The study was approved by the ethics committees of the participating universities and research institutions, as well as the relevant school authorities. Prior to participating in the study children gave oral consent, while written consent was obtained from their parents. Adult participants gave written consent.

2.1.2 | Materials

Sixty morphologically simple frequent nouns were selected from each language (e.g. night, nuit, Nacht, notte) for the construction of morphologically structured and non-morphologically structured nonword targets. The selected nouns served as stems and were combined with a frequent suffix, forming nonwords in the Stem + Suffix condition (e.g. night**ness**, nuit**eur**, Nacht**er**, nott**enza**), or a letter sequence that did not correspond to a suffix, forming nonwords in the Stem + Non-Suffix condition (e.g. night**lude**, nuit**erge**, Nacht**atz**, nott**erto**). After a letter was replaced in the stems, the resulting non-stems were combined with the suffixes, forming nonwords in the Non-Stem + Suffix condition (e.g. nisht**ness**, nait**eur**, Necht**er**, nutt**enza**), or the letter sequences (e.g. nisht**lude**, nait**erge**, Necht**atz**, nutt**erto**), forming nonwords in the Non-Stem + Non-Suffix condition. Translation-equivalent stems and whenever possible, translation-equivalent suffixes were used across languages. Also, in all stem + suffix and stem + non-suffix combinations, we ensured that the stem remained intact.[2] The word items were also frequent nouns. Thirty of them were suffixed, hence morphologically complex (e.g. baker, boulanger, Bäcker, panettiere), and thirty were non-affixed (e.g. diamond, diamant, Diamant, diamante), hence morphologically simple.[3]

English nouns were chosen from the Celex database (Baayen, Piepenbrock, & Gulikers, 1995), French nouns from Manulex (Lété, Sprenger-Charolles, & Colé, 2004), German nouns from the childLex corpus (Version 0.16.03; Schroeder, Wurzner, Heister, Geyken, & Kliegl, 2015), and Italian nouns from SUBTLEX-IT (Crepaldi, Amenta, Mandera, Keuleers, & Brysbaert, 2015). Item frequencies and orthographic neighborhood metrics were obtained from SUBTLEX-UK (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014) for English, Lexique (New, Brysbaert, Veronis, & Pallier, 2007) for French, SUBTLEX-DE (Brysbaert et al., 2011) for German, and SUBTLEX-IT (Crepaldi et al., 2015) for Italian. As a lexical density index, we used OLD20 (Orthographic Levenshtein distance; Yarkoni, Balota, & Yap, 2008). To control for potential effects of syllabic orthographic/phonological frequency on nonword reading aloud we calculated the absolute type frequency for each biphone within an item, then we summed them up and log-transformed them. All items are shown in the Appendix and their psycholinguistic properties are shown in Table 2. OLD20 and Phoneme Length differed significantly across languages ($F = 11.466$, $p < .001$ and $F = 3.150$, $p = .024$, respectively), so both variables were included as covariates in the analyses.

2.1.3 | Procedure

Three hundred items (60 words and 240 nonwords) were used in each language. Nonword items belonged to four conditions: Stem + Suffix, Stem + Non-Suffix, Non-Stem + Suffix, Non-Stem + Non-Suffix. Word items belonged to two conditions: Suffix and Non-Suffix. Four lists were created with each target nonword appearing once across the four lists and each target word appearing once in every list. Thus, each list comprised 120 items, 60 nonwords (15 with stem + suffix, 15 with stem + non-suffix, 15 with non-stem + suffix, and 15 with non-stem + non-suffix) and 60 words (30 suffixed and 30 non-suffixed), with all conditions being represented in every list. An equal number of participants were assigned to each list.[4] The order of trial presentation within each list was randomized across participants. Six practice trials were presented prior to the experimental trials.

Participants were tested individually, seated approximately 60 cm in front of a laptop or a PC monitor in a quiet room. Stimulus presentation and data recordings were controlled by DMDX software (Forster & Forster, 2003). Participants were instructed to read aloud the items quickly and carefully. Each item was presented in lowercase letters, except for German, where the first letter of nouns is always uppercase. For consistency, all German items were presented in the same format. The stimuli appeared in white on a black background (20-point Arial font) and remained on the screen for 4,000 ms (children) or 3,000 ms (adults). The task lasted 15 min for children and 10 min for adults.

2.2 | Measures

2.2.1 | Reading fluency

Children's reading ability was assessed to ensure they had no reading impairments that could affect their performance on the task. The tests used were the TOWRE (Torgesen, Wagner, & Rashotte, 1999) in English, the 1-min-reading test (Gentaz, Sprenger-Charolles, & Theurel, 2015) and the TIME3 word-reading test (Écalle, 2006) in French, the SLRT II (Moll & Landerl, 2010) in German, and the MT Reading Test for Primary School (Cornoldi & Colpo, 1998) in Italian. The TOWRE, 1-min-reading, and SLRT reading tests involved reading aloud of words and nonwords. Based on each sample, we calculated a *z*-score for correctly read words and a *z*-score for correctly read nonwords. The average of the two was used as a reading ability score in the analyses. The Italian MT Reading Test involved reading aloud of a text passage that contained words. A measure of reading speed of correctly read words, which was expressed in seconds per syllable, was extracted. This meant that higher scores on this test corresponded to slower children. Hence, *z*-scores based on the sample were first calculated, and then multiplied by −1.

TABLE 1 Participant demographic characteristics

		Age			
	Language	*N*	*M (SD)*	Range	Recruitment area
Children	English	24 (16 boys)	9.2 (0.3)	8.7–9.8	Sydney, Australia
	French	32 (16 boys)	8.6 (0.4)	7.8–9.9	Côte d'Azur, France
	German	32 (16 boys)	9.0 (0.5)	8.0–10.4	Berlin, Germany
	Italian	32 (18 boys)	8.8 (0.3)	8.3–9.8	Trieste, Italy
Adults	English	32 (1 male)	22.0 (7.5)	18–58	Sydney, Australia
	French	32 (3 males)	19.2 (1.3)	17–22	Marseille, France
	German	32 (8 males)	24.7 (3.0)	20–29	Berlin, Germany
	Italian	32 (15 males)	24.9 (3.3)	20–32	Trieste, Italy

TABLE 2 Psycholinguistic properties of items in all languages (SDs in parentheses)

	English	French	German	Italian
Words				
Suffixed				
OLD20	2.0 (0.6)	2.3 (0.5)	2.3 (0.5)	1.9 (0.4)
N letters	7.3 (1.5)	8.1 (1.2)	7.7 (1.2)	8.9 (1.5)
N phonemes	6.2 (1.9)	6.4 (1.4)	6.3 (1.1)	8.3 (1.6)
N syllables	2.4 (0.8)	2.6 (0.7)	2.1 (0.4)	3.9 (0.7)
Frequency (*Zipf*)	4.3 (0.6)	4.0 (0.6)	4.1 (0.5)	3.9 (0.6)
Non-Suffixed				
OLD20	1.9 (0.7)	2.0 (0.5)	2.3 (0.4)	1.6 (0.4)
N letters	6.2 (1.5)	6.7 (0.8)	6.7 (0.8)	7.3 (1.0)
N phonemes	5.2 (1.4)	5.0 (0.9)	6.2 (1.3)	6.6 (1.2)
N syllables	1.9 (0.6)	1.8 (0.5)	2.4 (0.7)	3.0 (0.6)
Frequency (*Zipf*)	4.0 (0.7)	3.7 (0.6)	3.8 (0.6)	3.9 (0.7)
Nonword stems				
OLD20	1.2 (0.3)	1.5 (0.4)	1.4 (0.4)	1.1 (0.2)
N letters	4.4 (1.1)	4.6 (1.0)	4.3 (1.0)	5.4 (1.2)
N phonemes	3.5 (0.9)	3.2 (1.0)	4.0 (0.9)	5.3 (1.1)
N syllables	1.1 (0.3)	1.2 (0.4)	1.1 (0.3)	2.2 (0.5)
Frequency (*Zipf*)	4.9 (0.5)	4.4 (0.9)	4.7 (0.6)	4.7 (0.5)
Nonwords				
Stem + Suffix				
OLD20	2.5 (0.5)	2.5 (0.6)	2.8 (0.6)	2.3 (0.6)
N letters	8.0 (1.3)	8.1 (1.3)	7.7 (1.2)	8.3 (1.7)
N phonemes	6.8 (1.2)	5.7 (1.3)	6.8 (1.1)	7.9 (1.7)
N syllables	2.1 (0.3)	2.2 (0.5)	2.1 (0.3)	3.5 (0.6)
Biphone frequency	10.3 (0.6)	9.9 (0.6)	11.6 (0.5)	11.1 (0.3)
Stem + Non-Suffix				
OLD20	3.0 (0.7)	2.9 (0.5)	2.7 (0.5)	2.6 (0.7)
N letters	8.0 (1.2)	8.2 (1.1)	7.3 (1.2)	8.3 (1.7)
N phonemes	6.6 (1.1)	6.2 (1.1)	6.8 (1.1)	7.9 (1.7)
N syllables	2.1 (0.3)	2.3 (0.5)	2.1 (0.3)	3.5 (0.6)
Biphone frequency	10.2 (0.5)	10.0 (0.5)	11.6 (0.6)	11.1 (0.3)
Non-stem + Suffix				
OLD20	2.7 (0.6)	2.7 (0.6)	3.0 (0.7)	2.5 (0.6)
N letters	8.0 (1.3)	8.1 (1.3)	7.7 (1.2)	8.3 (1.7)
N phonemes	6.8 (1.2)	5.9 (1.3)	6.9 (1.1)	7.9 (1.7)
N syllables	2.1 (0.3)	2.3 (0.5)	2.1 (0.3)	3.5 (0.6)
Biphone frequency	10.3 (0.6)	9.9 (0.6)	11.6 (0.5)	11.0 (0.3)
Non-stem + Non-Suffix				
OLD20	3.2 (0.7)	3.2 (0.6)	2.9 (0.5)	2.9 (0.8)
N letters	7.9 (1.2)	8.1 (1.2)	7.3 (1.2)	8.3 (1.7)
N phonemes	6.6 (1.1)	6.3 (1.2)	6.9 (1.1)	7.9 (1.7)
N syllables	2.1 (0.3)	2.3 (0.5)	2.1 (0.3)	3.5 (0.6)
Biphone frequency	10.1 (0.5)	10.0 (0.5)	11.5 (0.6)	11.0 (0.3)

2.2.2 | Vocabulary

Children's vocabulary size was assessed to obtain an estimate of the general level of lexical knowledge. The tests used were the vocabulary subtest of the Wechsler Abbreviated Scale of Intelligence (Wechsler, 2011) in English, the Test de vocabulaire actif et passif pour enfants de 5 à 8 ans (TVAP 5–8) in French (Deltour & Hupkens, 1980), and the vocabulary subtest of the CFT-20R (Weiß, 2006) in German. Based on each sample, *z*-scores were calculated. Due to testing time limitations, Italian children could not be administered a vocabulary test.

2.3 | Results

Naming latencies were determined by the acoustic onsets of participants' reading aloud responses. Acoustic onsets were hand-marked with CheckVocal (Protopapas, 2007) following the criteria specified by Rastle, Croot, Harrington, and Coltheart (2005). A response was deemed correct or incorrect using the same rules in all languages. In the case of words, incorrect responses were considered those where the word was read incorrectly. In the case of nonwords, incorrect responses corresponded to utterances containing mispronounced, deleted, or additional phonemes. The vast majority of nonwords yielded a single pronunciation. In those cases where nonwords could be pronounced in more than one way, all plausible pronunciations were considered correct. Generally, only pronunciations of nonwords that native speakers of the corresponding language considered illegitimate were marked as incorrect. In each language, trained research assistants who were naïve to the purposes of the study labelled the acoustic onsets and determined the accuracy of the reading aloud responses.

Analyses were performed using (generalized) linear mixed-effects (LME) models (Baayen, Davidson, & Bates, 2008) as implemented in the *lme4* package (Version 1.1-21; Bates, Maechler, Bolker, & Walker, 2015) in the statistical software R (Version 3.6.1, 2019-07-05, 'Action of the Toes', R Core Team, 2018). Naming latencies were log-transformed to normalize residuals and analysed using a LME model. For the error analysis, a generalized linear mixed-effects (GLME) model was created using logit transformation and a binomial link function. The significance of the fixed effects was determined with type III model comparisons using the *Anova* function in the *car* package (Version 3.0-4, Fox & Weisberg, 2011). Post hoc comparisons were carried out using cell means coding and single *df* contrasts with the *glht* function of the *multcomp* package (Version 1.4-10;

Hothorn, Bretz, & Westfall, 2008) using the normal distribution to evaluate significance.

Data from all four languages were analysed together. The analyses of the children's data are reported first, followed by the analyses of the adult data. Nonwords and words were analysed separately. Nonwords were the focus of the present study, so only the nonword analyses are reported in the paper. All data and the R code corresponding to the present analyses are available via the Open Science Framework (OSF; https://osf.io/byqp9/).

2.3.1 | Nonword naming latencies in children

Incorrect responses to words and nonwords (17.8% of the data) were removed. For the nonword analyses, latencies below 300 or above 3,000 ms (2.2% of the data) were considered as extreme values and were also removed. Outliers were identified following the procedure outlined by Baayen and Milin (2010). A base model that included only participants and items as random intercepts was fitted to the data and data points with residuals exceeding 2.5 *SD*s were removed (1.8% of the data).

The LME model included the effect-coded fixed effects of Language (English vs. French vs. German vs. Italian), Stem (Stem vs. Non-Stem), Suffix (Suffix vs. Non-Suffix), and Reading Ability, as well as their interactions. Reading Ability scores from the corresponding reading tests in each language were standardized for each language separately.[5] OLD20 and Phoneme Length (both standardized) were included in the model as covariates to control for cross-linguistic differences between the items. Random intercepts and random slopes for the effects of Stem and Suffix, and their interaction were used for both subjects and items. Results are shown in Table 3 and mean model naming latencies are shown in Figure 1.

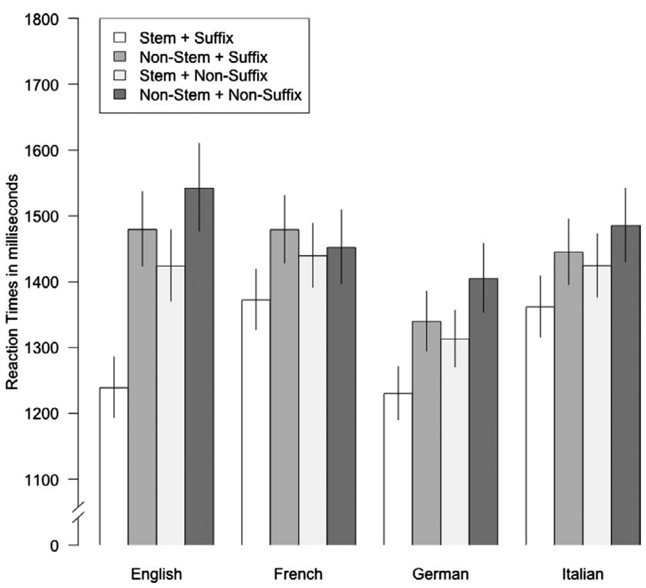

FIGURE 1 Children nonword naming latencies (in milliseconds) and standard errors

Main effects

Results showed a significant main effect of Stem. Nonwords with stems (M = 1,348 ms, SE = 22) were read aloud significantly faster (Δ = 104 ms, z = 8.557, p < .001) than nonwords without stems (M = 1,452 ms, SE = 26). The main effect of Suffix was significant. Nonwords with suffixes (M = 1,365 ms, SE = 22) were read aloud significantly faster (Δ = 69 ms, z = 5.093, p < .001) than nonwords without suffixes (M = 1,434 ms, SE = 25). Reading Ability was significant, with higher reading ability scores associated with faster naming latencies (b = −0.098, t = −5.976, p < .05), and so was OLD20 (b = 0.013, t = 2.040, p < .05).

Interactions

Critically for the present study, the interaction between Language and Stem was significant. The Stem effect (i.e. Non-Stem minus Stem condition) in English (Δ = 182 ms, z = 7.216, p < .001) was significantly larger (z = 3.491, p < .001) than the Stem effect in French (Δ = 60 ms, z = 2.391, p = .017), significantly larger (z = 2.174, p = .030) than the Stem effect in German (Δ = 101 ms, z = 4.694, p < .001), and significantly larger (z = 3.193, p = .001) than the Stem effect in Italian (Δ = 72 ms, z = 3.006, p = .003). Stem effects in French, German, and Italian did not differ from each other (z = 1.453, p = .146, for French vs. German; z = 0.369, p = .712, for French vs. Italian; z = 1.101, p = .271, for German vs. Italian). Furthermore, the interaction between Language and Suffix was significant. The Suffix effect (i.e. Non-Suffix minus Suffix condition) in English (Δ = 128 ms, z = 4.452, p < .001) was significantly larger (z = 2.776, p = .006) than the Suffix effect in French (Δ = 21 ms, z = 0.772, p = .440), and significantly larger (z = 2.027, p = .043) than the Suffix effect in Italian (Δ = 52 ms, z = 1.962, p = .050). The Suffix effect in English was also much larger than the Suffix effect in German (Δ = 74 ms, z = 3.192, p = .001), yet not significantly so (z = 1.256, p = .209). Suffix effects in French, German, and Italian were not significantly different (z = 1.593, p = .111, for French vs. German; z = 0.820, p = .412, for French vs. Italian; z = 0.789, p = .430, for German vs. Italian). The Stem by Suffix interaction was significant. The Stem effect for suffixed nonwords (Δ = 135 ms, z = 8.221, p < .001) was larger than the Stem effect for non-suffixed nonwords (Δ = 71 ms, z = 4.200, p < .001). Taken together, the observed results suggest that developing readers of English use morphology in reading aloud to a greater extent than developing readers of French, German, and Italian.[6]

Morphological processing as a function of Reading Ability

The interaction between Suffix and Reading Ability was significant. The Suffix effect for good readers (Δ = 94 ms, z = 6.030, p < .001) was much larger than the Suffix effect for poor readers (Δ = 38 ms, z = 1.842, p = .066). This is shown in Figure 2, where the main effect of Suffix is displayed for poor (−1 *SD*), average (*M*), and good (+1 *SD*) readers. As the figure shows, the Suffix effect increased with reading skill. Also, the Language by Stem by Reading Ability interaction reached significance. The Stem effect was modulated by reading

TABLE 3 Analysis of variance table for nonword naming latencies for children and adults

	Children		Adults	
	χ^2	*p*	χ^2	*p*
Fixed effects (*df*)				
Intercept (1)	199,750.000	<.001	194,170.000	<.001
Language (3)	4.669	=.198	27.343	<.001
Stem (1)	73.224	<.001	119.920	<.001
Suffix (1)	25.934	<.001	87.435	<.001
Language × Stem (3)	14.810	=.002	16.975	=.001
Language × Suffix (3)	8.435	=.038	8.030	=.045
Stem × Suffix (1)	9.217	=.002	12.511	<.001
Language × Reading Ability (3)	4.545	=.208		
Stem × Reading Ability (1)	1.510	=.219		
Suffix × Reading Ability (1)	8.592	=.003		
Language × Stem × Suffix (3)	4.282	=.233	0.517	=.915
Language × Stem × Reading Ability (3)	7.876	=.049		
Language × Suffix × Reading Ability (3)	3.985	=.263		
Stem × Suffix × Reading Ability (1)	0.196	=.658		
Language × Stem × Suffix × Reading Ability (3)	5.842	=.120		
Reading Ability (1)	35.710	<.001		
OLD20 (1)	4.163	=.041	73.513	<.001
Phoneme Length (1)	2.890	=.089	11.915	=.001

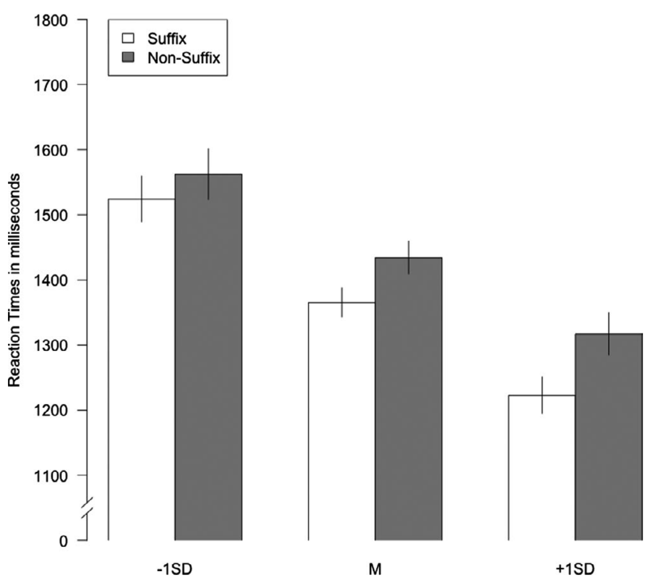

FIGURE 2 Suffix by Reading Ability interaction for children nonword naming latencies

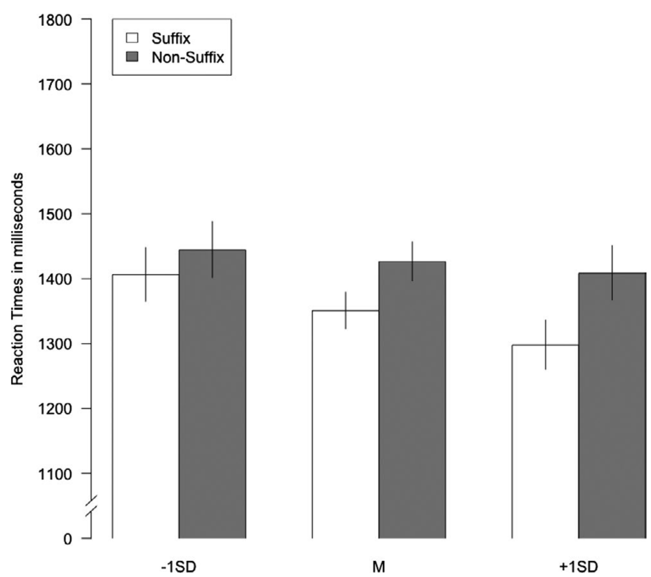

FIGURE 3 Suffix by Vocabulary Knowledge interaction for children nonword naming latencies

ability in French (z = 2.633, p = .008), but not in English (z = −0.727, p = .467), German (z = 1.319, p = .187), or Italian (z = −0.664, p = .507). In particular, French good readers yielded a significant Stem effect (Δ = 108 ms, z = 3.783, p < .001), whereas French poor readers yielded no Stem effect (Δ = 2 ms, z = 0.058, p = .954).

Morphological processing as a function of Vocabulary Knowledge

An additional prediction derived from the Psycholinguistic Grain Size Theory is that vocabulary knowledge might facilitate reading in all languages, because the phonological decoding network can only be reinforced when children know the words they decode from

TABLE 4 Analysis of variance table for nonword accuracy for children and adults

	Children		Adults	
	χ^2	*p*	χ^2	*p*
Fixed effects (*df*)				
Intercept (1)	417.327	<.001	1,082.111	<.001
Language (3)	47.889	<.001	24.547	<.001
Stem (1)	30.358	<.001	0.577	=.448
Suffix (1)	29.920	<.001	3.790	=.052
Language × Stem (3)	4.258	=.235	0.162	=.984
Language × Suffix (3)	16.428	=.001	1.358	=.715
Stem × Suffix (1)	3.953	=.047	0.155	=.694
Language × Stem × Suffix (3)	3.200	=.362	5.070	=.167
Reading Ability (1)	78.447	<.001		
OLD20 (1)	3.714	=.054	10.721	=.001
Phoneme Length (1)	9.985	=.002	6.683	=.010

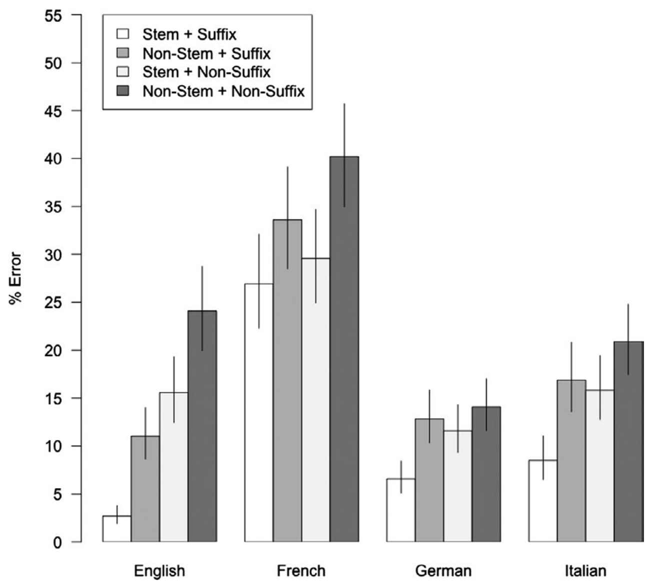

FIGURE 4 Children nonword accuracy (%) and standard errors

their spoken language (see Ziegler et al., 2014), and this is true in all languages no matter the grain size they might use for the computation. In contrast, the Orthographic Depth Hypothesis would predict that vocabulary knowledge is particularly important for reading in English, because it would further boost lexico-semantic processing in this language (Harm & Seidenberg, 2004). We tested these predictions in English, French, and German.

The analyses were conducted in the same way as the analyses on naming latencies, except that Vocabulary, instead of Reading Ability, was entered in the model. Vocabulary scores were standardized for each language separately. Results were practically identical to those reported earlier as far as the main effects and critical interactions are concerned. Furthermore, the interaction between Suffix and Vocabulary was significant. The Suffix effect for children with good vocabulary knowledge (Δ = 111 ms, z = 5.996, p < .001) was much larger than the Suffix effect for children with poor vocabulary knowledge (Δ = 38 ms, z = 1.849, p = .064), indicating that children with better vocabulary knowledge were more sensitive to morphological structure. This is shown in Figure 3, where the effect of Suffix is displayed for children with poor (−1 *SD*), average (*M*), and good (+1 *SD*) vocabulary knowledge. No other interactions were significant. Importantly, no differences across languages were observed.

2.3.2 | Nonword accuracy in children

Accuracy was analysed in the same way as naming latencies. The GLME model included the same fixed effects and interactions as the LME model. Results are shown in Table 4 and mean model errors are shown in Figure 4.

Main effects

The main effect of Stem was significant. Nonwords with stems (M = 12.1, SE = 1.1) yielded significantly fewer errors (Δ = 8.2, z = 5.510, p < .001) than nonwords without stems (M = 20.3, SE = 1.5). The main effect of Suffix was significant. Nonwords with suffixes (M = 12.1, SE = 1.1) yielded significantly fewer errors (Δ = 8.2, z = 5.470, p < .001) than nonwords without suffixes (M = 20.3, SE = 1.5). The main effect of Language was also significant. Errors in French (M = 32.4, SE = 3.4) were significantly more (Δ = 21.7, z = 5.753, p < .001) than errors in English (M = 10.6, SE = 1.6), significantly more (Δ = 21.5, z = 6.055, p < .001) than errors in German (M = 10.9, SE = 1.5), and significantly more (Δ = 17.5, z = 3.916, p < .001) than errors in Italian (M = 14.9, SE = 2.1). There were no significant differences between English and German, English and Italian, and German and Italian (z = 0.100, p = .921; z = 1.472, p = .141; and z = 1.484, p = .138, respectively).[7] Reading Ability was significant, with higher scores associated with fewer errors (b = 0.631, z = 8.857, p < .001). As Reading Ability did not interact with the other factors, it was modelled as a main effect. Phoneme Length was significant (b = −0.280, t = −3.160, p = .002).

Interactions

Language interacted with Suffix. Post hoc contrasts showed that the Suffix effect in English (Δ = 13.9, z = 5.645, p < .001) was significantly larger (z = 3.870, p < .001) than the Suffix effect in French (Δ = 4.5, z = 1.044, p = .296), significantly larger (z = 3.198, p = .001) than the Suffix effect in German (Δ = 3.6, z = 1.744, p = .081), and significantly larger (z = 2.923, p = .003) than the Suffix effect in Italian (Δ = 6.1, z = 2.273, p = .023). No significant differences were observed between the other three languages (z = 0.544, p = .586 for French vs. German; z = 0.969, p = .333, for French vs. Italian; z = 0.394, p = .694, for German vs. Italian). The absence of suffixes in nonwords was thus detrimental to children's reading accuracy in English. Last, the interaction between Stem and Suffix was significant. The Stem effect for suffixed nonwords (Δ = 8.9, z = −4.832, p < .001) was larger than the Stem effect for non-suffixed nonwords (Δ = 6.4, z = −2.820, p = .005).

2.3.3 | Nonword naming latencies in adults

Incorrect responses to words and nonwords (2.3% of the data) were removed. For the nonword analyses, latencies below 200 or above 2,000 ms (0.4% of the data) were considered as extreme values and were also removed. Outliers (1.9% of the data) were removed following the same procedure as for the children. The same LME model as for the analyses of the children data was created except that Reading Ability was not included in the model. Results are shown in Table 3 and mean model naming latencies are shown in Figure 5.

Main effects

The main effect of Stem was significant. Nonwords with stems (M = 796 ms, SE = 12) were read aloud significantly faster (Δ = 66 ms, z = 10.950, p < .001) than nonwords without stems (M = 862 ms, SE = 14). The main effect of Suffix was significant. Nonwords with suffixes (M = 799 ms, SE = 12) were read aloud significantly faster (Δ = 61 ms, z = 9.351, p < .001) than nonwords without suffixes (M = 860 ms, SE = 14). The Language effect was also significant. German nonwords (M = 725 ms, SE = 22) were read aloud significantly faster (Δ = 113 ms, z = 3.346, p = .001) than English nonwords (M = 838 ms, SE = 26), significantly faster (Δ = 160 ms, z = 4.604, p < .001) than French nonwords (M = 885 ms, SE = 28), and significantly faster (Δ = 152 ms, z = 4.365, p < .001) than Italian nonwords (M = 877 ms, SE = 28). OLD20 and Phoneme Length were significant (b = 0.045, t = 8.574, p < .05, and b = 0.019, t = 3.452, p < .05, respectively).

Interactions

Similarly to the children, the interaction between Language and Stem was significant. The Stem effect in English (Δ = 106 ms, z = 8.934, p < .001) was significantly larger (z = 2.844, p = .004) than the Stem effect in French (Δ = 62 ms, z = 4.898, p < .001), significantly larger (z = 2.617, p = .009) than the Stem effect in German (Δ = 54 ms, z = 5.235, p < .001), and significantly larger (z = 3.972, p < .001) than the Stem effect in Italian (Δ = 41 ms, z = 3.234, p = .001). Stem effects did not differ in the other languages (z = −0.229, p = .819, for French vs. German; z = 1.142, p = .254, for French vs. Italian; z = 1.370, p = .171, for German vs. Italian). Moreover, the Language by Suffix interaction was significant. The Suffix effect in English (Δ = 82 ms, z = 6.276, p < .001) was significantly larger (z = 2.746, p = .006) than the Suffix effect in French (Δ = 35 ms, z = 2.599, p = .009). The Suffix effect in English was also larger than the Suffix effect in German (Δ = 55 ms, z = 5.028, p < .001) and Italian (Δ = 71 ms, z = 5.307, p < .001); however, these differences were not significant (z = 1.057, p = .291, and z = 0.788, p = .431, respectively). Differences between French and Italian just reached significance (z = 1.960, p = .050), while the other comparisons were not significant (z = −1.656, p = .098, for French vs. German; z = 0.288, p = .773, for German vs. Italian). The Stem by Suffix interaction was significant. The Stem effect for suffixed nonwords (Δ = 83 ms, z = 9.831, p < .001) was significantly larger than the Stem effect for non-suffixed nonwords (Δ = 47 ms, z = 5.768, p < .001). These results indicate that skilled readers of English use morphology in reading aloud to a greater extent than skilled readers of French, German, and Italian.

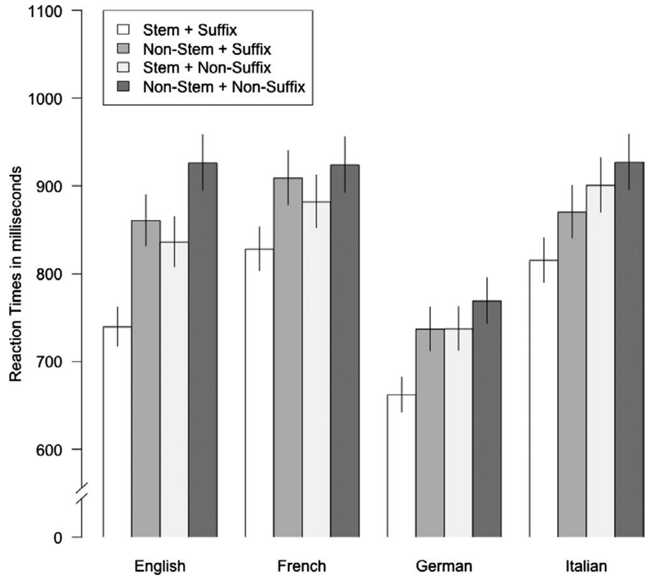

FIGURE 5 Adult nonword naming latencies (in milliseconds) and standard errors

2.3.4 | Nonword accuracy in adults

The GLME model included the same fixed effects and interactions as the LME model. Results are shown in Table 4 and mean model errors are shown in Figure 6.

Main effects

There was a significant main effect of Language. Errors in English (M = 0.3, SE = 0.1) were significantly fewer (Δ = 0.9, z = 3.123, p = .002) than errors in French (M = 1.2, SE = 0.2), significantly fewer

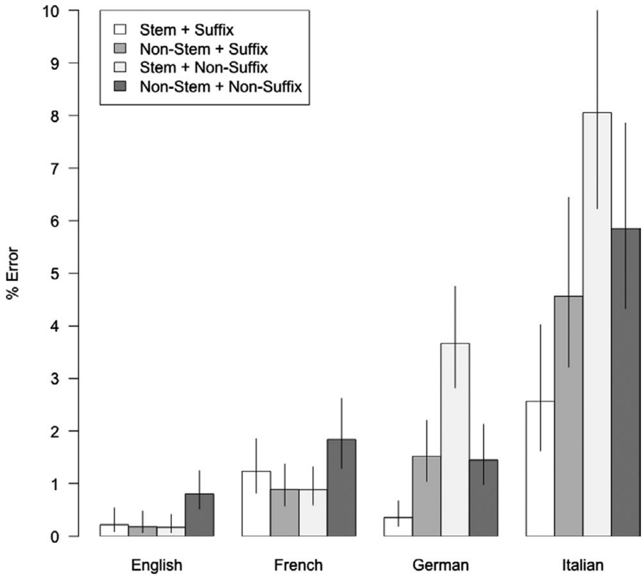

FIGURE 6 Adult nonword accuracy (%) and standard errors

(Δ = 1.0, z = 3.318, p = .001) than errors in German (M = 1.3, SE = 0.3), and significantly fewer (Δ = 4.6, z = 5.978, p < .001) than errors in Italian (M = 4.9, SE = 0.9). Differences between French and Italian, and German and Italian, were also significant (z = 4.110, p < .001, and z = 4.046, p < .001, respectively). The main effect of Suffix approached significance. Nonwords with suffixes (M = 0.9, SE = 0.2) yielded significantly fewer errors (Δ = 0.8, z = 2.509, p = .012) than nonwords without suffixes (M = 1.7, SE = 0.3). OLD20 and Phoneme Length were significant (b = −0.572, z = −3.274, p = .001, and b = 0.433, z = 2.585, p = .010, respectively).

3 | GENERAL DISCUSSION

The present study is the first that uses a tightly controlled cross-linguistic experimental design to examine whether readers of deep orthographies use morphemes to compute pronunciations (rather than other large grain sizes such as syllables, rimes, or whole words, which have been extensively investigated in the literature). We observed that morphological processing is, indeed, more robust in English than in more consistent orthographies such as French, German, and Italian. Our findings provide support for the Orthographic Depth Hypothesis and the Psycholinguistic Grain Size Theory, showing that the orthographic consistency of a language, and not its morphological complexity, modulates the extent to which morphology is used in reading (see Vannest, Bertram, Järvikivi, & Niemi, 2002, who also found more morphological computation in English than in Finnish, even though Finnish is renowned for its morphological richness).[8] It is worth noting that cross-linguistic differences were even greater for stems than for suffixes, perhaps because of the serial left-to-right nature of the reading aloud task, which requires stem recognition prior to suffix recognition, thus placing more emphasis on the stem. Also, stem morphemes are thought to be highly salient units contributing the largest amount of meaning to morphologically complex words (Grainger & Beyersmann, 2017).

Another important finding of the present study is that the observed cross-linguistic differences in morphological processing were astonishingly similar for developing and skilled readers. This result is predicted by the Psycholinguistic Grain Size Theory, according to which readers of all alphabetic writing systems have to go through an orthography-to-phonology mapping (decoding) to acquire reading, but they use different grain sizes to do so. English spelling prioritizes the consistency of morphemes over the consistency of phonemes (Bowers & Bowers, 2018), which is why morphological units might be used by English-speaking children right from the start to achieve an efficient orthography-to-phonology mapping. The Orthographic Depth Hypothesis makes a somewhat different prediction by stating that readers of deep orthographies, such as English, shift their reliance from the phonological to the orthographic lexico-semantic route (Katz & Frost, 1992). Given that lexico-semantic processing takes time to develop, the Orthographic Depth Hypothesis would predict that cross-linguistic differences in morphological processing may only emerge with sufficient reading experience. This specific finding also challenges connectionist reading models (e.g. Seidenberg & McClelland, 1989), which require a huge amount of training before they exhibit any cross-language differences in reading aloud (e.g. Hutzler, Ziegler, Perry, Wimmer, & Zorzi, 2004). Such models make the strong prediction that morphological effects would only occur late during the learning-to-read process, when the division of labour shifts the focus from spelling-to-sound to spelling-to-meaning mappings.

3.1 | Morphological processing as a function of Reading Ability and Vocabulary Knowledge

The analyses on naming latencies showed that good readers yielded a 56-ms larger suffix effect than poor readers (see Figure 2), indicating that reading skill modulates sensitivity to suffixes. Also, French good readers yielded a large stem effect (i.e. 108 ms), whereas French poor readers yielded no stem effect (i.e. 2 ms), indicating that reading ability may modulate sensitivity to stem morphemes too. Moreover, we observed that children with good vocabulary knowledge yielded a 73-ms larger suffix effect than children with poor vocabulary knowledge (see Figure 3), indicating that vocabulary knowledge also modulates sensitivity to suffixes. Taken together, these results suggest that children who read fluently and children who have a rich vocabulary make more extensive use of morphology during reading. This finding is consistent with the idea that individuals with good reading and language skills are better at mapping letters onto large grain sizes (Andrews & Lo, 2013; Beyersmann, Casalis, Ziegler, & Grainger, 2015; Beyersmann, Grainger, et al., 2015), thus promoting more efficient reading. As per the Psycholinguistic Grain Size Theory, and in contrast to the Orthographic Depth Hypothesis, the extensive use of morphology as an index of efficient reading by children with good vocabulary knowledge was not modulated by the orthographic consistency of the language.

3.2 | Limitations

One limitation of our study is that we used only suffixed items, so it might be that our results do not generalize to prefixed items. It is worth noting though that a few recent studies that specifically sought to investigate differences in the processing of prefixed and suffixed items during reading found no differences between the two affix types. For example, in a study conducted in French, equivalent priming was observed when targets (e.g. AMOUR) were preceded by prefixed (e.g. preamour) and suffixed (e.g. amouresse) nonword primes, or similarly constructed non-affixed nonword primes (e.g. brosamour, amourugne), compared to an unrelated condition (Beyersmann, Cavalli, Casalis, & Colé, 2016). Similar findings have been reported in English (Heathcote, Nation, Castles, & Beyersmann, 2018) and German (Mousikou & Schroeder, 2019). Moreover, the results from the German study were replicated in three single-word reading experiments and one sentence reading experiment. Therefore, there is no reason to think that the cross-linguistic differences observed in the present study would not also arise with prefixed items.

3.3 | Educational implications

There is a general consensus that systematic phonics, that is, explicit instruction of the relationship between letters and sounds, is best practice for early reading instruction in English (see Castles et al., 2018). However, as it has been recently pointed out by Bowers and Bowers (2018), English is a morphophonemic system that evolved to jointly represent units of meaning (morphemes) and phonology (phonemes). In fact, English prioritizes the consistent spelling of morphemes over the consistent spelling of phonemes. Accordingly, it has been suggested that reading instruction in English should be guided by the logic of the English writing system (Bowers & Bowers, 2017). Thus, it should be organized around morphology and phonology rather than just phonology. Our findings support this idea. We found that developing readers of English made extensive use of morphology in reading aloud. Furthermore, we observed that good readers were overall more sensitive to morphological structure than poor readers. Importantly, poor readers of English often exhibit phonological processing deficits, so these children might benefit even more by teaching methods that focus on optimal grain sizes of their writing system (i.e. morphemes), which would allow a more straightforward mapping between print and sound, in addition to an easy mapping between print and meaning.

To conclude, cross-linguistic studies can help us gain an insight into both universal and language-specific processes involved in reading acquisition, which is critical for addressing theoretical and applied issues that are relevant for a universal science of reading.

ACKNOWLEDGEMENTS

This research was supported by grants ANR-11-LABX-0036 (BLRI), ANR-11-IDEX-0001-02, and ANR-15-FRAL-0003-01 from the French National Agency for Research (ANR), and grant SCHR-1268/2–1 from the German Research Foundation (DFG). Elisabeth Beyersmann was supported by a Discovery Early Career Researcher Award (DECRA) by the Australian Research Council (DE190100850). Davide Crepaldi was supported by ERC Starting Grant 679010 (STATLEARN). The authors thank Jérémy Labille and Morgane Mounié-Moser (CNRS), Pia Linscheid and Katharina Pittrich (MPIB), Andrea Nadalini, Valentina Pescuma, Mara De Rosa, and the Brains at Work 2018 team (SISSA), for research assistance.

CONFLICT OF INTERESTS

The authors declare that they have no conflicts of interest with respect to their authorship or the publication of this article.

DATA AVAILABILITY STATEMENT

The datasets analysed during the current study are available in the Open Science Framework repository, https://osf.io/byqp9/

ORCID

Petroula Mousikou https://orcid.org/0000-0003-0611-9718
Elisabeth Beyersmann https://orcid.org/0000-0001-9653-6106
Davide Crepaldi https://orcid.org/0000-0002-6765-1516
Johannes C. Ziegler https://orcid.org/0000-0002-2061-5729
Jonathan Grainger https://orcid.org/0000-0002-6737-5646
Sascha Schroeder https://orcid.org/0000-0001-7001-4588

ENDNOTES

[1] Note that for English, these values are 0.68 for GPCs and 0.65 for PGCs.

[2] For the Italian nonwords *busazione* and *busalorte*, a completive thematic vowel was added to the stem *bus* (*-a*, resulting in *busa*), so that the nonwords were phonologically legal when combined with the suffix *-zione* and the letter sequence *-lorte*. Due to an oversight, the counterpart “non-stem” nonwords *basazione* and *basalorte* contained the real stem *bas-*. Therefore, neither the psycholinguistic properties of this quadruplet nor the naming latencies corresponding to it were included in the respective calculations and analyses. Similarly, the nonwords *bisfuitful* (English) and *tanneloso* (Italian) were accidentally used twice, correctly in the Non-Stem + Suffix condition but incorrectly in the Non-Stem + Non-Suffix condition. The psycholinguistic properties of these nonwords and the naming latencies corresponding to them were excluded from the respective calculations and analyses.

[3] Due to an oversight, three English (*power*, *fisherman*, *hairstyle*), three German (*Bescheid*, *Existenz*, *Frisur*), and four Italian words (*potere*, *posizione*, *libertà*, *gioventù*) were incorrectly classified as suffixed. We took such oversights into account when calculating the psycholinguistic properties of the items in the different conditions and when analyzing the data.

[4] Due to the specific selection criteria for the children, one list ended up containing four children less than the other three lists.

[5] Population norms were available for the German SLRT, the French TIME3, and the English TOWRE. A one-sample *t* test revealed that German children performed significantly below the population mean for words, $t(31) = -3.436$, $p = .002$, and nonwords, $t(31) = -3.173$, $p = .003$. French children did not differ significantly from the population mean on word reading, $t(31) = 1.145$, $p = .261$. English-speaking children did not differ significantly from the population mean for words, $t(23) = 0.637$, $p = .530$, yet they scored slightly above the population mean for nonwords, $t(23) = 2.934$, $p = .007$. Because of these differences, we computed a population-based Reading Ability score for each child in these three languages, where population norms for the corresponding reading tests were available. We then analyzed the data in the same way, except that population-based reading scores were included in the model. Results did not differ from those reported in the paper (see relevant analyses at https://osf.io/byqp9/).

[6] One possibility is that the stronger morphological effects observed in English are due to the higher overall exposure to print of the Australian sample, given that formal reading instruction in Australia begins in kindergarten, so before Grade 1. To exclude this possibility, we carried out additional analyses. Given that our French and German data came from a sample of children who participated in an independent longitudinal project (see Section 2.1), we had reading aloud data (from the same task that includes the same stimuli) from the same French and German children in Grade 4. Our additional analyses included thus the reading aloud data from 24 English-speaking third-graders, 24 French fourth-graders, and 31 German fourth-graders (testing in Grade 4 occurred exactly a year later than testing in Grade 3 in both France and Germany, so there were a few dropouts). Critically, results from these analyses were similar to those reported in the paper, showing more robust morphological effects in English-speaking third-graders than in French and German fourth-graders (see relevant analyses at https://osf.io/byqp9/).

[7] Six English-speaking children were originally excluded from the analyses due to an error rate of over 50%. Hence, French and English children yielded more errors than German and Italian children. The high error rate in French could be due to the substantial number of silent letters in the French nonwords, which would likely increase pronunciation uncertainty.

[8] One possibility is that English readers' sensitivity to morphology is due to the "high visibility" of morphological information in English spelling (Rastle, 2018). English past tense forms, for example, are always spelled with 'ed' even when their ending is pronounced /əd/, /d/, or /t/, thus making morphological relationships in print particularly prominent. However, morphological information is not less visible in the other three languages. In German, for example, a similar morphological principle applies: the written form of morphologically related words (e.g., Sand-sandig 'sand-sandy') is preserved even when the spoken form slightly varies (/zant/-/zandık/), where the 'd' in 'Sand' is pronounced /t/ due to devoicing.

REFERENCES

Amenta, S., & Crepaldi, D. (2012). Morphological processing as we know it: An analytical review of morphological effects in visual word identification. *Frontiers in Psychology*, *3*, 1–12. https://doi.org/10.3389/fpsyg.2012.00232

Andrews, S., & Lo, S. (2013). Is morphological priming stronger for transparent than opaque words? It depends on individual differences in spelling and vocabulary. *Journal of Memory and Language*, *68*, 279–296. https://doi.org/10.1016/j.jml.2012.12.001

Anglin, J. M. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development*, *58*, 1–166. https://doi.org/10.2307/1166112

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

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*, 12–28. https://doi.org/10.21500/20112084.807

Bane, M. (2008). Quantifying and measuring morphological complexity. In C. B. Chang & H. J. Haynie (Eds.), *Proceedings of the 26th west coast conference on formal linguistics* (pp. 69–76). Somerville, MA: Cascadilla Proceedings Project.

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

Beyersmann, E., Casalis, S., Ziegler, J., & Grainger, J. (2015). Language proficiency and morpho-orthographic segmentation. *Psychonomic Bulletin and Review*, *22*, 1054–1061. https://doi.org/10.3758/s13423-014-0752-9

Beyersmann, E., Cavalli, E., Casalis, S., & Colé, P. (2016). Embedded stem priming effects in prefixed and suffixed pseudowords. *Scientific Studies of Reading*, *20*, 220–230. https://doi.org/10.1080/10888438.2016.1140769

Beyersmann, E., Grainger, J., Casalis, S., & Ziegler, J. C. (2015). Effects of reading proficiency on embedded stem priming in primary school children. *Journal of Experimental Child Psychology*, *139*, 115–126. https://doi.org/10.1016/j.jecp.2015.06.001

Beyersmann, E., Grainger, J., & Castles, A. (2019). Embedded stems as a bootstrapping mechanism for morphological parsing during reading development. *Journal of Experimental Child Psychology*, *182*, 196–210. https://doi.org/10.1016/j.jecp.2019.01.010

Booij, G. (2012). *The grammar of words. An introduction to linguistic morphology*. Oxford, UK: Oxford University Press.

Borgwaldt, S. R., Hellwig, F. M., & De Groot, A. M. B. (2005). Onset entropy matters—Letter-to-phoneme mappings in seven languages. *Reading and Writing*, *18*, 211–229. https://doi.org/10.1007/s11145-005-3001-9

Borleffs, E., Maassen, B. A. M., Lyytinen, H., & Zwarts, F. (2017). Measuring orthographic transparency and morphological-syllabic complexity in alphabetic orthographies: A narrative review. *Reading and Writing*, *30*, 1617–1638. https://doi.org/10.1007/s11145-017-9741-5

Bowers, J. S., & Bowers, P. N. (2017). Beyond phonics: The case for teaching children the logic of the English spelling system. *Educational Psychologist*, *52*, 124–141. https://doi.org/10.1080/00461520.2017.1288571

Bowers, J. S., & Bowers, P. N. (2018). Progress in reading instruction requires a better understanding of the English spelling system. *Current Directions in Psychological Science*, *27*, 407–412. https://doi.org/10.1177/0963721418773749

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, *58*, 412–424. https://doi.org/10.1027/1618-3169/a000123

Burani, C., Marcolini, S., De Luca, M., & Zoccolotti, P. (2008). Morpheme-based reading aloud: Evidence from dyslexic and skilled Italian readers. *Cognition*, *108*, 243–262. https://doi.org/10.1016/j.cognition.2007.12.010

Caravolas, M., & Kessler, B. (2016). Spelling development in alphabetic orthographies of varying consistency: Measurement matters. Paper presented at the meeting of the Society for the Scientific Study of Reading (SSSR), Porto, 2016 July 13–16. Retrieved form http://spell.psychology.wustl.edu/CaravolasSSSR2016/CaravolasSSSR2016.pdf

Carlisle, J. F. (1995). Morphological awareness and early reading achievement. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 189–210). Hillsdale, NJ: Lawrence Erlbaum Associates.

Carlisle, J. F., & Stone, C. A. (2005). Exploring the role of morphemes in word reading. *Reading Research Quarterly*, *40*, 428–449. https://doi.org/10.1598/RRQ.40.4.3

Casalis, S., Quémart, P., & Duncan, L. G. (2015). How language affects children's use of derivational morphology in visual word and pseudoword processing: Evidence from a cross-language study. *Frontiers in Psychology*, *6*, 1–10. https://doi.org/10.3389/fpsyg.2015.00452

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, *19*, 5–51. https://doi.org/10.1177/1529100618772271

Cornoldi, C., & Colpo, G. (1998). *Prove di Lettura MT per la Scuola Elementare – 2 [The MT Reading Test for Primary School – 2]*. Firenze: Giunti O.S.

Cossu, G., Gugliotta, M., & Marshall, J. C. (1995). Acquisition of reading and written spelling in a transparent orthography: Two non-parallel processes? *Reading and Writing*, *7*, 9–22. https://doi.org/10.1007/BF01026945.

Crepaldi, D., Amenta, S., Mandera, P., Keuleers, E., & Brysbaert, M. (2015). SUBTLEX-IT. Subtitle-based word frequency estimates for Italian. *Proceedings of the Annual Meeting of the Italian Association for Experimental Psychology*. Rovereto. 10-12 September 2015.

Deltour, J.-J., & Hupkens, D. (1980). *Test de vocabulaire actif et passif pour enfants de 5 à 8 ans (TVAP 5–8)*. Braine-le-Château: Editions de l'Application des Techniques Modernes (ATM). Nouvel étalonnage en 1998.

Écalle, J. (2006). *Timé3: Test d'Identification de Mots Écrits pour enfants de 7 à 15 ans*. Paris, France: Mot-à-Mot.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods Instruments and Computers*, *35*, 116–124. https://doi.org/10.3758/BF03195503

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (2nd ed.). Thousand Oaks, CA: Sage. Retrieved from http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in German- and English-speaking children. *Scientific Studies of Reading*, *2*, 31–54. https://doi.org/10.1207/s1532799xssr0201_2

Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences*, *35*, 263–279. https://doi.org/10.1017/S0140525X11001841

Gentaz, E., Sprenger-Charolles, L., & Theurel, A. (2015). Differences in the predictors of reading comprehension in first graders from low socio-economic status families with either good or poor decoding skills. *PLoS ONE*, *10*, e0119581. https://doi.org/10.1371/journal.pone.0119581

Goswami, U., Gombert, J., & de Barrera, L. (1998). Children's orthographic representations and linguistic transparency: Nonsense word reading in English, French, and Spanish. *Applied Psycholinguistics*, *19*, 19–52. https://doi.org/10.1017/S0142716400010560

Goswami, U., & Ziegler, J. C. (2006). Fluency, phonology and morphology: A response to the commentaries on becoming literate in different languages. *Developmental Science*, *9*, 451–453. https://doi.org/10.1111/j.1467-7687.2006.00511.x

Grainger, J., & Beyersmann, E. (2017). Edge-aligned embedded word activation initiates morpho-orthographic segmentation. *Psychology of Learning and Motivation*, *67*, 285–317. https://doi.org/10.1016/bs.plm.2017.03.009

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*, 662–720. https://doi.org/10.1037/0033-295X.111.3.662

Hasenäcker, J., Schröter, P., & Schroeder, S. (2017). Investigating developmental trajectories of morphemes as reading units in German. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 1093–1108. https://doi.org/10.1037/xlm0000353

Heathcote, L., Nation, K., Castles, A., & Beyersmann, E. (2018). Do 'blacheap' and 'subcheap' both prime 'cheap'? An investigation of morphemic status and position in early visual word processing. *Quarterly Journal of Experimental Psychology*, *71*, 1645–1654. https://doi.org/10.1080/17470218.2017.1362704

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*, 346–363. https://doi.org/10.1002/bimj.200810425

Hutzler, F., Ziegler, J. C., Perry, C., Wimmer, H., & Zorzi, M. (2004). Do current connectionist learning models account for reading development in different languages? *Cognition*, *91*, 273–296. https://doi.org/10.1016/j.cognition.2003.09.006

Juola, P. (1998). Measuring linguistic complexity: The morphological tier. *Journal of Quantitative Linguistics*, *5*, 206–213. https://doi.org/10.1080/09296179808590128

Juola, P. (2008). Assessing linguistic complexity. In M. Miestamo, K. Sinnemäki, & F. Karlsson (Eds.), *Language complexity: Typology, contact, change* (pp. 89–108). Amsterdam: John Benjamins Publishing Company.

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), *Advances in psychology, Vol. 94. Orthography, phonology, morphology, and meaning* (pp. 67–84). Oxford, UK: North-Holland.

Kettunen, K. (2014). Can type-token ratio be used to show morphological complexity of languages? *Journal of Quarterly Linguistics*, *21*, 223–224. https://doi.org/10.1080/09296174.2014.911506

Landerl, K. (2000). Influences of orthographic consistency and reading instruction on the development of nonword reading skills. *European Journal of Psychology of Education*, *15*, 239–257. https://doi.org/10.1007/BF03173177

Lázaro, M., Acha, J., de la Rosa, S., García, S., & Sainz, J. (2017). Exploring the derivative suffix frequency effect in Spanish speaking children. *Reading and Writing*, *30*, 163–185. https://doi.org/10.1007/s11145-016-9668-2

Lázaro, M., García, L., & Burani, C. (2015). How orthographic transparency affects morphological processing in young readers with and without reading disability. *Scandinavian Journal of Psychology*, *56*, 498–507. https://doi.org/10.1111/sjop.12213

Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX: A grade-level lexical database from French elementary school readers. *Behavior Research Methods, Instruments, & Computers*, *36*, 156–166. https://doi.org/10.3758/BF03195560

Liberman, I. Y., Liberman, A. M., Mattingly, I. G., & Shankweiler, D. (1980). Orthography and the beginning reader. In J. Kavanaugh & R. Venezky (Eds.), *Orthography, reading and dyslexia* (pp. 137–153). Baltimore: University Park Press.

Mann, V. (2000). Introduction to special issue on morphology and the acquisition of alphabetic writing systems. *Reading and Writing: An Interdisciplinary Journal*, *12*, 143–147.

Mann, V., & Singson, M. (2003). Linking morphological knowledge to English decoding ability: Large effects of little suffixes. In E. M. H. Assink & D. Sandra (Eds.), *Reading complex words: Cross-language studies* (pp. 1–25). New York, NY: Kluwer Academic.

Moll, K., & Landerl, K. (2010). *SLRT-II: Lese- und Rechtschreibtest: Weiterentwicklung der Salzburger Lese- und Rechtschreibtests (SLRT): Manual*. Bern, Switzerland: Verlag Hans Huber.

Mousikou, P., & Schroeder, S. (2019). Morphological processing in single-word and sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*, 881–903. https://doi.org/10.1037/xlm0000619

Nagy, W. E., & Anderson, R. C. (1984). The number of words in printed school English. *Reading Research Quarterly*, *19*, 304–330. https://doi.org/10.2307/747823

Nagy, W. E., Carlisle, J. F., & Goodwin, A. P. (2014). Morphological knowledge and literacy acquisition. *Journal of Learning Disabilities*, *47*, 3–12. https://doi.org/10.1177/0022219413509967

Nation, K. (2009). Form-meaning links in the development of visual word recognition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1536), 3665e3674. https://doi.org/10.1098/rstb.2009.0119

New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics*, *28*, 661–677. https://doi.org/10.1017/S014271640707035X

Paulesu, E., Demonet, J. F., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., … Frith, U. (2001). Dyslexia: Cultural diversity and biological unity. *Science*, *291*(5511), 2165–2167. https://doi.org/10.1126/science.1057179

Perdijk, K., Schreuder, R., Baayen, R. H., & Verhoeven, L. (2012). Effects of morphological family size for young readers. *British Journal of Developmental Psychology*, *30*, 432–445. https://doi.org/10.1111/j.2044-835X.2011.02053.x

Perry, C., Zorzi, M., & Ziegler, J. C. (2019). Understanding dyslexia through personalized large-scale computational models. *Psychological Science*, *30*, 386–395. https://doi.org/10.1177/0956797618823540

Plaut, D. C., & Gonnerman, L. M. (2000). Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes*, *15*(4–5), 445–485. https://doi.org/10.1080/01690960050119661

Protopapas, A. (2007). CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, *39*, 859–862. https://doi.org/10.3758/BF03192979

Quémart, P., Casalis, S., & Duncan, L. G. (2012). Exploring the role of bases and suffixes when reading familiar and unfamiliar words: Evidence from French young readers. *Scientific Studies of Reading*, *16*, 424–442. https://doi.org/10.1080/10888438.2011.584333

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rastle, K. (2018). The place of morphology in learning to read in English. *Cortex*, *116*, 45–54. https://doi.org/10.1016/j.cortex.2018.02.008

Rastle, K., Croot, K. P., Harrington, J. M., & Coltheart, M. (2005). Characterizing the motor execution stage of speech production: Consonantal effects on delayed naming latency and onset duration. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 1083–1095. https://doi.org/10.1037/0096-1523.31.5.1083

Rey-Debove, J. (1984). Le domaine de la morphologie lexicale [The domain of lexical morphology]. *Cahiers de Lexicologie*, *45*, 3–19.

Roelcke, T. (1997). *Sprachtypologie des Deutschen: Historische, regionale und funktionale Variation [Language typology of German: Historical, regional, and functional variation]*. Berlin, Germany: de Gruyter.

Schmalz, X., Marinus, E., Coltheart, M., & Castles, A. (2015). Getting to the bottom of orthographic depth. *Psychonomic Bulletin and Review*, *22*, 1614–1629. https://doi.org/10.3758/s13423-015-0835-2

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

Seidenberg, M. S. (2011). Reading in different writing systems: One architecture, multiple solutions. In P. McCardle, J. Ren, & O. Tzeng (Eds.), *Dyslexia across languages: Orthography and the gene-brain-behavior link* (pp. 146–168). Baltimore, MD: Paul Brooke Publishing.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523–568. https://doi.org/10.1037/0033-295X.96.4.523

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*, 143–174. https://doi.org/10.1348/000712603321661859

Share, D. L. (2008). On the anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological Bulletin*, *134*, 584–615. https://doi.org/10.1037/0033-2909.134.4.584

Sprenger-Charolles, L., Siegel, L. S., & Bonnet, P. (1998). Reading and spelling acquisition in French: The role of phonological mediation and orthographic factors. *Journal of Experimental Child Psychology*, *68*, 134–165. https://doi.org/10.1006/jecp.1997.2422

Sprenger-Charolles, L., Siegel, L., Jiménez, J., & Ziegler, J. (2011). Prevalence and reliability of phonological, surface, and mixed profiles in dyslexia: A review of studies conducted in languages varying in orthographic depth. *Scientific Studies of Reading*, *15*, 498–521. https://doi.org/10.1080/10888438.2010.524463

Tabossi, P., & Laghi, L. (1992). Semantic priming in the pronunciation of words in two writing systems: Italian and English. *Memory & Cognition*, *20*, 303–313. https://doi.org/10.3758/BF03199667

Talamo, L., & Celata, C. (2011). Toward a morphological analysis of the Italian lexicon: developing tools for a corpus-based approach. In *Quaderni del Laboratorio di Linguistica 1.10*. Retrieved from http://linguistica.sns.it/QLL/QLL11.htm

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of word reading efficiency*. Austin, TX: PRO-ED Publishing Inc.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, *124*, 107–136. https://doi.org/10.1037/0096-3445.124.2.107

Ulicheva, A., Harvey, H., Aronoff, M., & Rastle, K. (2018). Skilled readers' sensitivity to meaningful regularities in English writing. *Cognition*, *195*, 103810. https://doi.org/10.1016/j.cognition.2018.09.013

Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*, 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Vannest, J., Bertram, R., Järvikivi, J., & Niemi, J. (2002). Counterintuitive cross-linguistic differences: More morphological computation in English than in Finnish. *Journal of Psycholinguistic Research*, *31*, 83–106.

Verhoeven, L., & Perfetti, C. A. (2003). The role of morphology in learning to read. *Scientific Studies of Reading*, *7*, 209–217. https://doi.org/10.1207/S1532799XSSR0703_1

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence-Second Edition (WASI-II)*. San Antonio, TX: NCS.

Weiß, R. H. (2006). *Grundintelligenztest Skala 2 (CFT 20-R) mit Wortschatztest (WS) und Zahlenfolgentest (ZF) [Basic test of intelligence scale 2 (CFT 20-R) with vocabulary test (WS) and digit sequence test (ZF)]*. Göttingen, Germany: Hogrefe.

Wimmer, H., & Goswami, U. (1994). The influence of orthographic consistency on reading development: Word recognition in English and German children. *Cognition*, *51*, 91–103. https://doi.org/10.1016/0010-0277(94)90010-8

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

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., … Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science*, *21*, 551–559. https://doi.org/10.1177/0956797610363406

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic

grain size theory. *Psychological Bulettin*, *131*, 3–29. https://doi.org/10.1037/0033-2909.131.1.3
Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science*, *12*, 379–384. https://doi.org/10.1111/1467-9280.00370
Ziegler, J. C., Perry, C., & Zorzi, M. (2014). Modelling reading development through phonological decoding and self-teaching: Implications for dyslexia. *Philosophical Transactions of the Royal Society B*, *369*, 20120397. https://doi.org/10.1098/rstb.2012.0397

How to cite this article: Mousikou P, Beyersmann E, Ktori M, et al. Orthographic consistency influences morphological processing in reading aloud: Evidence from a cross-linguistic study. *Dev Sci*. 2020;00:e12952. https://doi.org/10.1111/desc.12952

APPENDIX

Items used in the study

Words							
Non-Suffixed				Suffixed			
English	French	German	Italian	English	French	German	Italian
ant	fourmi	Ameise	formica	baker	boulanger	Bäcker	panettiere
asphalt	asphalte	Asphalt	asfalto	treatment	traitement	Behandlung	trattamento
banana	banane	Banane	banana	decision	décision	Bescheid	decisione
basilisk	basilic	Basilisk	basilisco	servant	serviteur	Diener	servitore
custom	coutume	Brauch	costume	stupidity	stupidité	Dummheit	stupidità
brush	brosse	Bürste	pennello	existence	existence	Existenz	esistenza
diamond	diamant	Diamant	diamante	fortress	forteresse	Festung	fortezza
shower	douche	Dusche	doccia	fisherman	pêcheur	Fischer	pescatore
success	succès	Erfolg	successo	researcher	chercheur	Forscher	ricercatore
flame	flamme	Flamme	fiamma	freedom	liberté	Freiheit	libertà
giraffe	girafe	Giraffe	giraffa	hairxmlstyle	coiffure	Frisur	acconciatura
guitar	guitare	Gitarre	chitarra	youth	jeunesse	Jüngling	gioventù
hostel	auberge	Herberge	ostello	illness	maladie	Krankheit	malattia
colleague	collègue	Kollege	collega	artist	artiste	Künstler	artista
chest	commode	Kommode	comodino	power	puissance	Leistung	potere
contact	contact	Kontakt	contatto	clearing	clairière	Lichtung	radura
control	contrôle	Kontrolle	controllo	liar	menteur	Lügner	bugiardo
concert	concert	Konzert	concerto	manager	directeur	Manager	direttore
claw	griffe	Kralle	artiglio	humanity	humanité	Menschheit	umanità
wig	perruque	Perücke	parrucca	musician	musicien	Musiker	musicista
puddle	flaque	Pfütze	pozza	beauty	beauté	Schönheit	bellezza
plate	plaque	Platte	piatto	security	sécurité	Sicherheit	sicurezza
puzzle	puzzle	Puzzle	puzzle	settlement	règlement	Siedlung	insediamento
pyramid	pyramide	Pyramide	piramide	winner	gagnant	Sieger	vincitore
soldier	soldat	Soldat	soldato	player	joueur	Spieler	giocatore
stork	cigogne	Storch	cicogna	position	position	Stellung	posizione
talent	talent	Talent	talento	tracker	viseur	Sucher	mirino
tomato	tomate	Tomate	pomodoro	trainer	entraîneur	Trainer	allenatore
triumph	triomphe	Triumph	trionfo	training	formation	Training	formazione
cigar	cigare	Zigarre	sigaro	wisdom	sagesse	Weisheit	saggezza

Nonwords							
Stem + Suffix				Stem + Non-Suffix			
English	French	German	Italian	English	French	German	Italian
armful	brasable	Armbar	bracciore	armase	brasaste	Armucht	bracciodo
treement	arbrement	Baumkeit	alberoso	treetege	arbrelot	Baumarf	alberome
legful	jambeable	Beinbar	gambabile	legose	jambelot	Beinatz	gambacilo
broomment	balaiment	Besenkeit	scopazione	broomlude	balailot	Besenau	scopalorte
bedness	liteur	Better	lettoso	bedmose	literge	Bettarf	lettome
flashment	foudrement	Blitzkeit	fulminenza	flashnule	foudrenule	Blitzpern	fulminempo
bloodful	sangeux	Bluthaft	sanguenza	blooduck	sangonne	Blutam	sanguerdo
letterment	lettrement	Briefkeit	letteramento	letternule	lettrenule	Briefmen	letteralerto
breadful	paineux	Brothaft	panenza	breadrel	painache	Brotarf	panerto
breaster	seineur	Bruster	senore	breastel	seinate	Brustekt	senoco
busness	busion	Busung	busazione	busnete	busuque	Busarf	busalorte
roofer	toiteur	Dacher	tettore	roofel	toitipe	Dachpfen	tettome
iceful	glacable	Eisbar	ghiacciore	icenep	glacenule	Eismen	ghiacciodo
fieldful	champeux	Feldbar	camposo	fieldane	champonne	Feldatz	campome
filmful	filmeux	Filmhaft	filmoso	filmose	filmuque	Filmarf	filmodo
flightment	volment	Flugkeit	voloso	flightmose	volige	Flugucht	volome
hallful	halleux	Flurbar	salabile	hallept	hallache	Flurpern	salacilo
facement	facement	Gesichtkeit	facciamento	facenure	facenure	Gesichtarf	faccialorte
ghostment	fantômement	Gespenstkeit	fantasmamento	ghostnule	fantômenule	Gespenstpern	fantasmalorte
stopment	arrêtment	Haltkeit	arrestore	stopnept	arrêtipe	Haltarf	arrestoco
woodness	boision	Holzung	legnore	woodnane	boisipe	Holzat	legnoco
henful	pouletable	Huhnbar	pollore	henude	pouletème	Huhnam	pollome
dogness	chienion	Hundung	canenza	dognule	chienaste	Hundat	canempo
biscuitful	biscuitable	Keksbar	biscottore	biscuitude	biscuitil	Keksmen	biscottoco
guyful	garsable	Kerlbar	tiposo	guybal	garsare	Kerlmen	tipome
headment	têtement	Kopfkeit	testario	headnure	têtelot	Kopfekt	testachio
holement	troument	Lochkeit	bucoso	holenept	trounure	Lochucht	bucodo
airment	airement	Luftkeit	ariamento	airnule	airenure	Luftucht	arialorte
mousement	sourisment	Mauskeit	topore	mouserund	sourisisse	Mauspern	topodo
milkment	laitment	Milchkeit	lattenza	milkrane	laitope	Milcharf	latterto
moonment	lunement	Mondkeit	lunamento	moonhoke	lunelot	Mondatz	lunalerto
nightness	nuiteur	Nachter	nottenza	nightlude	nuiterge	Nachtatz	notterto
nestness	nidion	Nestung	nidore	nestnane	nidil	Nestarf	nidoco
parkful	parcable	Parkbar	parcore	parkure	parcache	Parkarf	parcoco
horseness	chevalion	Pferdung	cavalloso	horsenure	chevalème	Pferdam	cavalloco
pointment	pointment	Punktkeit	puntoso	pointvose	pointerge	Punktam	puntome
wheelment	rouement	Radkeit	ruotamento	wheelhoke	rouenure	Radam	ruotalorte
lawable	droiteux	Rechtbar	destrabile	lawnept	droitate	Rechtmen	destratilo
juiceness	jusion	Saftung	succore	juicehoke	jusache	Saftmen	succoco
sandness	sablion	Sandung	sabbiazione	sandlude	sablenule	Sanducht	sabbialerto
treasureness	trésorion	Schatzung	tesorore	treasuremose	trésorisse	Schatzarf	tesorodo
senseness	sension	Sinnung	sensoso	senserane	sensare	Sinnau	sensome
trackment	pistement	Spurkeit	pistamento	tracklude	pistenure	Spurnauf	pistalerto
stonement	pierrement	Steinung	pietramento	stonelabe	pierrenule	Steinam	pietralerto

Nonwords							
Stem + Suffix				Stem + Non-Suffix			
English	French	German	Italian	English	French	German	Italian
frontful	fronteux	Stirnbar	frontenza	frontase	frontaste	Stirnatz	fronterto
dayful	jourable	Tagbar	giornoso	daytege	jourouse	Tagucht	giornoco
carpetful	tapisable	Teppichbar	tappetoso	carpetrel	tapisisse	Teppichatz	tappetome
tablement	tablement	Tischkeit	tavolore	tablenept	tablenure	Tischarf	tavolodo
potful	poteux	Topfbar	pentolabile	potaph	potare	Topfekt	pentolatilo
tunnelness	tunnelion	Tunnelung	tunneloso	tunnelmose	tunnelipe	Tunnelau	tunnelodo
wallful	mureux	Wandbar	muroso	wallund	muruque	Wandekt	murome
worldment	mondement	Weltkeit	mondore	worldnule	mondenure	Weltekt	mondoco
windful	ventable	Windbar	ventore	windane	venterge	Winducht	ventodo
jokement	blaguement	Witzkeit	scherzore	jokelabe	blaguipe	Witzarf	scherzoco
wolfful	loupeux	Wolfhaft	luposo	wolfrel	loupouse	Wolfat	lupome
wordful	motieux	Worthaft	parolabile	wordane	motige	Wortpern	parolatilo
toother	denteur	Zahner	dentenza	toothel	dentaste	Zahnarf	denterto
timeful	tempsable	Zeitbar	temposo	timenul	tempsouse	Zeitam	tempome
tentment	tentement	Zeltkeit	tendamento	tentlure	tentenure	Zeltat	tendalerto
trainful	traineux	Zughaft	trenoso	trainege	trainaste	Zugat	trenoco
Non-Stem + Suffix				Non-Stem + Non-Suffix			
English	French	German	Italian	English	French	German	Italian
arfful	brusable	Arfbar	brocciore	arfase	brusast	Arfucht	brocciodo
treiment	aubrement	Baufkeit	alburoso	treitege	aubrelot	Baufarf	alburome
ligful	jombeable	Beunbar	gumbabile	ligose	jombelot	Beunatz	gumbacilo
broosment	bavaiment	Belenkeit	scupazione	brooslude	bavailot	Belenau	scupalorte
berness	lateur	Botter	littoso	bermose	laterge	Bottarf	littome
flishment	foidrement	Blatzkeit	folminenza	flishnule	foidrenule	Blatzpern	folminempo
bloudful	sargeux	Blethaft	senguenza	blouduck	sargonne	Bletam	senguerdo
lotterment	lottrement	Bliefkeit	lotteramento	lotternule	lottrenule	Bliefmen	lotteralerto
brealful	paimeux	Bromhaft	ponenza	brealrel	paimache	Bromarf	ponerto
breister	seifeur	Bluster	sunore	breistel	seifate	Blustekt	sunoco
bulness	bumion	Bumung	basazione	bulnete	bumuque	Bumarf	basalorte
roifer	taiteur	Ducher	tittore	roifel	taitipe	Duchpfen	tittome
ifeful	glatable	Eusbar	ghiecciore	ifenep	glatenule	Eusmen	ghiecciodo
fierdful	chalpeux	Faldbar	cumposo	fierdane	chalponne	Faldatz	cumpome
filtful	falmeux	Filthaft	folmoso	filtose	falmuque	Filtarf	folmodo
flishtment	vosment	Fluskeit	vuloso	flishtmose	vosige	Flusucht	vulome
hollful	holleux	Flerbar	selabile	hollept	hollache	Flerpern	selacilo
ficement	ficement	Gosichtkeit	fucciamento	ficenure	ficenure	Gosichtarf	fuccialorte
ghistment	fastômement	Gestenstkeit	fentasmamento	ghistnule	fastômenule	Gestenstpern	fentasmalorte
stosment	arvêtment	Holtkeit	arrastore	stosnept	arvêtipe	Holtarf	arrastoco
woosness	boimion	Holmung	lugnore	woosnane	boimipe	Holmat	lugnoco
honful	pauletable	Hehnbar	pillore	honude	pauletème	Hehnam	pillome
domness	chionion	Hondung	cunenza	domnule	chionaste	Hondat	cunempo
bisfuitful	bisfuitable	Kelsbar	bisbottore	bisfuitful	bisfuitil	Kelsmen	bisbottoco
gueful	garpable	Kertbar	tuposo	guebal	garpare	Kertmen	tupome

Non-Stem + Suffix				Non-Stem + Non-Suffix			
English	French	German	Italian	English	French	German	Italian
heafment	têrement	Korfkeit	tistario	heafnure	têrelot	Korfekt	tistachio
hilement	traument	Lechkeit	becoso	hilenept	traunure	Lechucht	becodo
aisment	aipement	Luptkeit	aroamento	aisnule	aipenure	Luptucht	aroalorte
moufement	sourifment	Maunkeit	tipore	mouferund	sourifisse	Maunpern	tipodo
molkment	lautment	Mulchkeit	littenza	molkrane	lautope	Mulcharf	litterto
mootment	luvement	Moldkeit	lonamento	moothoke	luvelot	Moldatz	lonalerto
nishtness	naiteur	Nechter	nuttenza	nishtlude	naiterge	Nechtatz	nutterto
nistness	nedion	Nostung	nadore	nistnane	nedil	Nostarf	nadoco
parmful	parmable	Parmbar	pircore	parmure	parmache	Parmarf	pircoco
horpeness	chetalion	Pfeldung	cavelloso	horpenure	chetalème	Pfeldam	cavelloco
poiltment	poiltment	Pulktkeit	purtoso	poiltvose	poilterge	Pulktam	purtome
wheilment	rauement	Ridkeit	ruetamento	wheilhoke	rauenure	Ridam	ruetalorte
lewable	draiteux	Rachtbar	dostrabile	lewnept	draitate	Rachtmen	dostratilo
juileness	julion	Saktung	siccore	juilehoke	julache	Saktmen	siccoco
santness	satlion	Sardung	sebbiazione	santlude	satlenule	Sarducht	sebbialerto
treamureness	trisorion	Schetzung	tasorore	treamuremose	trisorisse	Schetzarf	tasorodo
selseness	selpion	Sintung	sansoso	selserane	selpare	Sintau	sansome
trockment	pisfement	Smurkeit	pirtamento	trocklude	pisfenure	Smurnauf	pirtalerto
stanement	piurrement	Steunung	puetramento	stanelabe	piurrenule	Steunam	puetralerto
frintful	fronseux	Stirmbar	frentenza	frintase	fronsaste	Stirmatz	frenterto
dauful	jaurable	Tafbar	giarnoso	dautege	jaurouse	Tafucht	giarnoco
carfetful	tupisable	Teplichbar	tippetoso	carfetrel	tupisisse	Teplichatz	tippetome
teblement	teblement	Teschkeit	tevolore	teblenept	teblenure	Tescharf	tevolodo
pomful	pomeux	Tolfbar	pertolabile	pomaph	pomare	Tolfekt	pertolatilo
tunfelness	tunfelion	Tunfelung	tanneloso	tunfelmose	tunfelipe	Tunfelau	tanneloso
walsful	muleux	Wardbar	muposo	walsund	muluque	Wardekt	mupome
worltment	monpement	Woltkeit	mindore	worltnule	monpenure	Woltekt	mindoco
wisdful	veltable	Wisdbar	vuntore	wisdane	velterge	Wisducht	vuntodo
jubement	bleguement	Wetzkeit	schirzore	jubelabe	bleguipe	Wetzarf	schirzoco
wolpful	loufeux	Wolphaft	leposo	wolprel	loufouse	Wolpat	lepome
werdful	mapieux	Wosthaft	porilabile	werdane	mapige	Wostpern	porilatilo
toither	dunteur	Zuhner	dontenza	toithel	duntaste	Zuhnarf	donterto
tiveful	telpsable	Zeilbar	tamposo	tivenul	telpsouse	Zeilam	tampome
tertment	tenfement	Zelpkeit	tundamento	tertlure	tenfenure	Zelpat	tundalerto
treinful	praineux	Zighaft	trunoso	treinege	prainaste	Zigat	trunoco

Note: Due to an oversight, five Italian stems (i.e. lettera, fantasma, testa, parola, and pentola), which were used for the construction of the nonwords, were combined with slightly different suffixes and non-suffixed letter sequences in the children and adult experiment. However, given that the criteria adopted for the construction of the nonwords in the four conditions were identical in both cases, we do not think that these minor differences in the stimuli would influence the children and adult results in a significant way. The items included in the Appendix correspond to the items that were presented to the children.