

Getting to the bottom of orthographic depth

Xenia Schmalz^{1,2} · Eva Marinus¹ · Max Coltheart¹ · Anne Castles¹

Published online: 17 April 2015
© Psychonomic Society, Inc. 2015

Abstract Orthographic depth has been studied intensively as one of the sources of cross-linguistic differences in reading, and yet there has been little detailed analysis of what is meant by orthographic depth. Here we propose that orthographic depth is a conglomerate of two separate constructs: the complexity of print-to-speech correspondences and the unpredictability of the derivation of the pronunciations of words on the basis of their orthography. We show that on a linguistic level, these two concepts can be dissociated. Furthermore, we make different predictions about how the two concepts would affect skilled reading and reading acquisition. We argue that refining the definition of orthographic depth opens up new research questions. Addressing these can provide insights into the specific mechanisms by which language-level orthographic properties affect cognitive processes underlying reading.

Keywords Reading · Orthographic depth · Cross-linguistic differences

What is orthographic depth?

In the study of reading, it is important to establish to what extent findings from reading in one language can be generalized to another, and what particular experimental results are specific to the particular orthography used in the studies (Frost, 2012; Share, 2008). In recent decades, cross-linguistic research in reading has focussed particularly on

the concept called *orthographic depth* as a source of cross-linguistic orthographic differences in reading behavior. Broadly speaking, orthographic depth refers to the reliability of print-to-speech correspondences. English is considered to be a deep orthography, as there are often different pronunciations for the same spelling patterns (e.g., “tough” – “though” – “through” – “bough” – “cough” – “thorough” – “hiccough”); Ziegler, Stone, & Jacobs, 1997). Hence, it has often been contrasted with “shallow” orthographies with more reliable correspondences, such as Serbo-Croatian (Frost, Katz, & Bentin, 1987; Turvey, Feldman, & Lukatela, 1984), German (Frith, Wimmer, & Landerl, 1998; Landerl, Wimmer, & Frith, 1997; Wimmer & Goswami, 1994; Ziegler, Perry, Jacobs, & Braun, 2001), and many others (see Katz & Frost, 1992; Ziegler & Goswami, 2005 for reviews).

Orthographic depth is relevant for a broad range of issues, including reading development, developmental and acquired reading disorders, and theoretical accounts of reading. All aspects of reading are intrinsically linked to the characteristics of the orthography, therefore establishing what orthographic characteristics affect reading processes, and the cognitive mechanisms via which this occurs, is important for practical and theoretical reasons. For example, research on reading acquisition has consistently shown that achieving reading accuracy is a slower process for children learning to read in deep compared to shallow orthographies (e.g., Frith et al., 1998; Landerl, 2000; Seymour, Aro, & Erskine, 2003; Wimmer & Goswami, 1994). To account for these findings, theories of reading acquisition often consider the role of orthographic depth, and the challenges that it poses for young readers (Goswami, 1999; Liberman, Liberman, Mattingly, & Shankweiler, 1980; Ziegler & Goswami, 2005). The very mechanisms that underlie reading acquisition might be important to different degrees depending on orthographic depth: numerous studies have shown differences between orthographies in the strength of various predictors of reading ability.

✉ Xenia Schmalz
xenia.schmalz@gmail.com

¹ Cognitive Science, Macquarie University, Sydney, NSW, Australia

² DPSS, Università degli Studi di Padova, Padova, Italy

Specifically, phonological awareness appears to be a stronger predictor of reading ability for deep orthographies, as it is needed to make sense of the complicated print-to-speech conversion system. Conversely, there is some evidence that rapid automatized naming is a stronger predictor of reading abilities in shallow orthographies, as it is important for developing fluency, an aspect with which poor readers of shallow languages tend to struggle (e.g., Caravolas et al., 2012; Moll et al., 2014; Vaessen et al., 2010; Ziegler et al., 2010).

Furthermore, behavioral studies suggest that the symptoms associated with developmental dyslexia differ as a function of orthographic depth (Landerl et al., 1997; Landerl et al., 2013; Wimmer, 1996; but see Ziegler, Perry, Ma-Wyatt, Ladner, & Schulte-Körne, 2003). The phenotype of dyslexia has been shown to depend on the depth of the orthography: in deep orthographies, dyslexia is characterized by inaccurate reading, while in shallow orthographies high accuracy can be achieved, but a slowness in reading persists (Wimmer, 1993). Although work in English has established the presence of various subtypes of dyslexia (Castles & Coltheart, 1993), it has been questioned whether these can be applied to more shallow languages, where the hurdles associated with developing sound reading skills are different (Bergmann & Wimmer, 2008; Wimmer, Mayringer, & Landerl, 2000). These behavioral findings on dyslexia and orthographic depth are supplemented by neuroimaging data, which have shown cross-linguistic differences in the brain activation patterns during reading in dyslexic compared to control readers (for a recent review, see Richlan, 2014).

In addition, the concept of orthographic depth touches on issues that are central to debates in the reading literature in general, such as the extent to which reading processes are universal or language-specific (Dehaene, 2009; Frost, 2012; Share, 2008). Previous research suggests that the cognitive processes underlying skilled reading are dependent on orthographic depth (Frost, 1994; Frost et al., 1987; Schmalz et al., 2014; Ziegler et al., 2001). Determining whether any aspects of the reading process are universal, and which aspects depend on the characteristics of the orthography, has been recently argued to be an essential and inevitable step in creating models of reading (Frost, 2012). More specifically, and relating directly to the concept of orthographic depth, the majority of reading research is based on English. As has been argued elsewhere, this poses a threat to the generalizability of this research, especially since English is considered to be an outlier on the orthographic depth scale compared to other orthographies (Share, 2008). Although orthographic depth is not the only source of variability across orthographies, it has probably received the most attention in the past decades. Therefore, understanding what it is and how it affects reading processes is of theoretical importance.

It is clear that orthographic depth is an important concept, and understanding how it relates to reading is pivotal, as it is a strong source of linguistic variability between alphabetic orthographies. Here, we argue that it is currently unclear what precise mechanisms drive these cross-orthographic differences, both on a linguistic and behavioral level. We propose that a more precise definition of orthographic depth is needed for future research. In particular, answering the question, “what is orthographic depth,” involves determining, on a linguistic level, what different aspects underlie this concept, and how these can be quantified. Once a clear definition of orthographic depth is formulated, current theories and models of reading can be used to make specific predictions about how each aspect of orthographic depth might affect skilled reading and reading acquisition. In the current paper, we discuss the concept of orthographic depth in three sections. First, we provide an overview of the previous theoretical work on this concept (Definitions to date). Then, we propose quantification methods of the cross-linguistic variability which can be linked to theoretically important concepts (Quantifications of orthographic depth). Finally, we outline testable predictions that can be drawn from our proposed framework (Predictions of the new orthographic depth framework for theories of reading).

Definitions to date

Existing definitions of orthographic depth

As orthographic depth has been explored for decades, a number of definitions have been proposed. Originally, the concept was formulated in terms of a compromise between morphological and phonological transparency (Chomsky & Halle, 1968). In orthographies such as English or Dutch, such compromises are necessary, because the languages are morphologically deep, in that the same morphemes can have different pronunciations in different contexts. Therefore, the orthography needs to convey either the morphology or the phonology of the word: it cannot convey both. For example, in English, the words “heal” and “health” have the same spelling pattern because they are semantically related, even though they have different pronunciations. Thus, English often sacrifices phonological transparency for morphological transparency. In Dutch, conversely, the words “lezen” (to read) and “[ik] lees” ([I] read) have different spellings, despite being forms of the same verb. This is because the “z” in “lezen” is pronounced as /z/, whereas consonants in the final position of Dutch words are devoiced; therefore, the pronunciation of the final phoneme of “lees” is /s/, which is represented by the grapheme “s”. Here, the Dutch orthography sacrifices morphological for phonological transparency (Landerl & Reitsma, 2005).

Originally, the term “depth” had two levels, relating either to morphological or phonological transparency. In the context of the reading literature, the concept of phonological transparency has received the most attention (Feldman & Turvey, 1983; Frost, 1994; Frost et al., 1987). Katz and Frost (1992), in a review of the Orthographic Depth Hypothesis (ODH), provide an overview of the origins of the term, and its relationship to both morphological and phonological transparency. Their predictions about how depth would affect reading processes, however, focused exclusively on the relationship between orthography and phonology – as we will discuss in detail in a later section.

The relationship between orthography and phonology is considered to vary as a continuum (e.g., Frost et al., 1987; Goswami, Gombert, & de Barrera, 1998; Seymour et al., 2003; Sprenger-Charolles, Siegel, Jiménez, & Ziegler, 2011). This implies that a given orthography can be classified along a single scale. However, this is only possible if this concept has an explicit and agreed-on definition, which would allow for the development of a linguistic quantification scheme. Arguably, this is currently lacking in the available literature to date.

There is agreement that orthographic depth refers to the reliability of the print-to-speech correspondences, but what exactly differs across orthographies and how this should be quantified is less clear. Katz and Frost (1992) list three different aspects of letter-sound correspondences that could help to flesh out this definition of orthographic depth: “Because shallow orthographies have relatively *simple*, *consistent*, and *complete* connections between letter and phoneme, it is easier for readers to recover more of a printed word’s phonology prelexically by assembling it from letter-phoneme correspondences.” (pp. 71-72). Similarly, in a more recent paper, Richlan (2014) concurs by describing orthographic depth as “the *complexity*, *consistency*, or *transparency* of grapheme-phoneme correspondences in written alphabetic language” (p. 1). What is now needed are studies concerning how these different concepts work, whether they can be distinguished from each other, and how each might be quantified.

We argue that a more specific definition is needed to create an explicit theoretical framework that accounts for the way in which orthographic depth influences reading. In order to conduct meaningful behavioral cross-linguistic studies, the degree of orthographic depth of the orthographies which are being studied needs to be defined *a priori*, preferably using an objective linguistic quantification method. This is particularly important because orthographies differ from each other in many aspects apart from orthographic depth, such as syllabic complexity, morphological complexity, orthographic density, or the proportion of mono- versus polysyllabic words in the language. Unless the concept of orthographic depth is formally defined, it is easy to fall into circular reasoning, where any behavioral differences across orthographies are attributed to

orthographic depth *post-hoc*, even when there is a possibility that they are caused by other, uncontrolled language-level differences.

Devising a meaningful quantification method bears further challenges, because the quantification scheme needs to retain a link to theoretically and practically meaningful constructs; if it does not, it becomes unclear what the quantification method is actually measuring. Therefore, we need to first understand what constructs underlie orthographic depth and whether these are theoretically important. For a particular linguistic construct, there also needs to be enough variability across orthographies to make across-language studies meaningful. Then, on a behavioral level, we need to be able to show a noticeable effect that is directly associated with the concept of study. In the following section, we provide an overview of how previous theoretical work on orthographic depth has used this concept.

Orthographic depth in theories and models of reading

Two theories of reading across languages that are primarily concerned with orthographic depth are the Orthographic Depth Hypothesis (ODH; Katz & Frost, 1992) and the Psycholinguistic Grain Size Theory (PGST; Ziegler & Goswami, 2005). Both postulate how orthographic depth would affect reading processes, the ODH with a focus on skilled reading, and the PGST with a focus on reading acquisition, and provide some definition of what is meant by orthographic depth. As mentioned in the previous section, Katz and Frost (1992) distinguish between three concepts underlying orthographic depth: in a deep language, the print-to-speech correspondences are *complex*, *inconsistent*, and *incomplete*. It is unclear, however, precisely how each of these three aspects relates to each other, and whether each of them influences reading in different ways. Katz and Frost (1992) say the following about the specific mechanism that affects reading processes:

We would like to make two points, each independent of the other. The first states that, because shallow orthographies are optimised for assembling phonology from a word’s component letters, phonology is more easily available to the reader prelexically than is the case for a deep orthography. The second states that the easier it is to obtain prelexical phonology, the more likely it will be used for both pronunciation and lexical access. Both statements together suggest that the use of assembled phonology should be more prevalent when reading a shallow than when reading a deep orthography (p. 71, Katz & Frost, 1992).

It is easy to understand this quote in terms of the complexity of print-to-speech correspondences. Behavioral evidence has shown that the presence of complex multiletter rules (e.g.,

“th”, “sh”) slows down reading aloud latencies of words and non-words (Rastle & Coltheart, 1998; Rey, Jacobs, Schmidt-Weigand, & Ziegler, 1998). According to Katz and Frost, this is what gives more time for the lexical route to access the lexical information, before the sublexical computation of the pronunciation is complete.

This brings us to the question of how this mechanism would function in a case where the sublexical information is either inconsistent or incomplete. In English, an example of an *inconsistent* sublexical unit is the letter string “ough,” which can be pronounced in six different ways for monosyllabic words alone. For an inconsistent word, sublexical information is not sufficient to determine the pronunciation – instead, the orthographic lexicon must be consulted in order to determine how to pronounce a word containing the inconsistent correspondence (e.g., “though” and “through,” which contain nearly identical sublexical information, but have different body pronunciations).

The third concept introduced by Katz and Frost (1992) is *incompleteness* of the sublexical correspondences. In English, examples of words with incomplete sublexical information are heterophonic homographs. A heterophonic homograph, such as the word “wind”, has two different pronunciations, each of which is linked to a different meaning. The sublexical information is incomplete, as sentence context is needed to activate both the correct phonology of the word, and the correct semantic representation. In an orthography such as Hebrew, this presents a routine computational problem: here, vowels are mostly not represented in written texts. Many words have identical consonant constellations, and as a result vowel information is needed to tell them apart: for example, the consonant string “DVR” can be pronounced, among other alternatives, as “davar,” meaning “thing” or “dever,” meaning “pestilence” (Frost & Bentin, 1992). Here, the sublexical, but also an orthographic-lexical, procedure is insufficient to retrieve a single pronunciation. Instead, lexical-semantic information needs to be consulted.

Complexity may lead to a quantitative change in the reading processes by slowing down the sublexical route relative to the lexical route. In contrast, inconsistency and incompleteness may force reliance on lexical strategies, as access to a single phonemic representation cannot occur without lexical information. Thus, the distinction drawn by the ODH between different aspects underlying orthographic depth requires further consideration and empirical work. In particular, it is of interest – and so far, to our knowledge, unexplored – whether these three components would affect reading processes in different ways. This would question the utility of the concept of orthographic depth as a unified construct, and instead support the view that it consists of different sub-components.

The PGST emphasizes the role of complex correspondences in driving cross-linguistic differences in the difficulty of acquiring a given orthography. According to the

PGST, children learning to read in a deep orthography attempt to minimize the unreliability of their sublexical correspondences by relying on larger units because these tend to be more predictive of a word’s correct pronunciation (at least in English; Peereman & Content, 1998; Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995). As a result, children learning to read in a deep orthography need to learn a greater number of correspondences: children in a hypothetical perfectly shallow orthography can simply learn the letters and their corresponding sounds, and decode all words with perfect accuracy using only those small units. According to this view, the necessity to learn many print-to-speech correspondences in deep orthographies slows down the process of reading acquisition, leading to the well-established behavioral pattern where children learning to read in a deep orthography lag behind children learning to read in a shallow orthography based on word and non-word reading tasks (Frith et al., 1998; Landerl, 2000; Seymour et al., 2003; Wimmer & Goswami, 1994).

According to the PGST, it then seems that orthographic depth can be described as the existence of complex correspondence rules that are needed in order to decode new words in a given orthography. However, we argue that this is not the whole picture: as Katz and Frost (1992), point out, other properties of print-to-speech correspondences associated with orthographic depth relate to their inconsistency and incompleteness. In the following section, we focus on understanding how inconsistency can be defined, different theories of skilled reading have different conceptualisations of what it is and how it affects reading.

Dissociating inconsistency and complexity

Generally, consistency relates to the presence of more than one pronunciation for a given letter string. It can be defined either on the level of a grapheme¹ (e.g., “ea” is an inconsistent grapheme, because it can be pronounced as in “bread” or “leak”), or of a body (e.g., “-eak” is an inconsistent body, because it can be pronounced as in “break” or “leak”).² The consistency terminology is generally associated with connectionist models (Harm & Seidenberg, 1999; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989); dual-route models (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) are more concerned with the concept of regularity, which is defined as compliance to a set of predetermined grapheme-phoneme correspondence rules.

¹ By “grapheme” we mean a letter or group of letters that is the written representation of a phoneme.

² Although consistency can be defined at these two different levels, virtually all the research on effects of consistency on reading (e.g. Jared, 2002) has focused just on body-level consistency.

As we explain below, this distinction is important because the two classes of models make different assumptions about how speech is computed from print. Broadly speaking, both classes of models agree on two points: (1) That there is a non-lexical procedure which uses knowledge of print-to-speech regularities to assemble a word's pronunciation. This procedure is particularly important for reading non-words (in an experimental setting) and unfamiliar words (in a real-life setting). And (2) that there are some words for which lexical knowledge needs to be recruited, because the sublexical routine will not provide a correct pronunciation (in the dual-route framework, this includes all words with irregular print-to-speech correspondences, whereas for connectionist, it is limited to low-frequency words with exceptional correspondences, e.g., “meringue”, “colonel”). In order to adopt a theoretically neutral framework, we use the term *unpredictability* to refer to the degree to which this non-lexical reading route, essential for reading non-words aloud, correctly translates the words of the orthography from orthography to phonology.

The dissociation between *complexity* and *unpredictability* on a linguistic level is not straightforward in English. This can be illustrated with the example of the minimal word pair: “gist” and “gift.” Arguably, the pronunciation of the word “gist” is transparent, because it can be determined using the context-sensitive correspondence that a “g” followed by an “i” is pronounced as /d/. Alternatively, the pronunciation of the word “gift,” could also be argued to be transparent, if we instead apply the simpler rule that the letter “g” is pronounced as /g/. Therefore, the pronunciation of the word “gist” can be resolved by the use of a complex (context-sensitive) correspondence, but in the English orthography it is also, to some degree, *unpredictable* whether this complex rule will apply or not.

The “gift” – “gist” example shows that in English, the complexity and unpredictability of sublexical correspondences are related and confounded, and indeed it is difficult to dissociate the two. This is not always the case in other orthographies, however. Both Italian and French contain the “g[i]” context-sensitive correspondence. In both these orthographies there is no unpredictability regarding this rule, as it always applies, meaning there are no words with the pattern “gi” where the “g” would be pronounced as /g/. As we will show later, this is important: an orthography that contains many complex rules which are entirely predictable is different from an orthography that contains many complex rules but also a great deal of unpredictability. We propose that complexity and unpredictability are two related but linguistically and theoretically dissociable concepts. Thus, we argue that orthographic depth, in the context of European orthographies, is a conglomerate of two separate concepts, namely the complexity of sublexical correspondences and the unpredictability of words' pronunciations given these correspondences.

Defining print-to-speech correspondences

As discussed in the previous section, all computational models of reading include some kind of mechanism that uses knowledge of the statistical regularities between print and speech in a given orthography to assemble a word's pronunciation. The implementation thereof, however, varies considerably, and is a source of debate between computational modellers (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart et al., 2001; Perry, Ziegler, & Zorzi, 2007; Perry, Ziegler, & Zorzi, 2010; Plaut et al., 1996; Seidenberg & McClelland, 1989). The Dual Route Cascaded (DRC) model of reading (Coltheart et al., 2001) contains sublexical rules which are defined as the phoneme which corresponds most frequently to a given grapheme. The rules are position-specific: each rule is either valid for all positions (“t” → /t/ - at least for monosyllabic words), or for the beginning, middle, or end positions (e.g., “y” → /j/ in the beginning of a word, and “y” → /ai/ in the end of a word). All sublexical correspondences are grapheme-phoneme correspondence (GPC) rules, meaning that they describe the pronunciation of a single phoneme. GPC rules can also be context-sensitive: for example, “g” is pronounced as /d/ when followed by an “i” (“g[i]” → /d/). Even those rules are GPCs, because they relate to a single phoneme (in this case, /d/).

In contrast to the DRC and its GPC rules, triangle models develop sensitivity to units that are larger than graphemes, thereby also showing sensitivity to marker effects that are associated with “large” units, such as bodies (Harm & Seidenberg, 1999; Plaut et al., 1996; Seidenberg & McClelland, 1989). Connectionist Dual Process (CDP)+-type models (Perry et al., 2007; Perry, Ziegler, & Zorzi, 2010) represent a compromise between the DRC and triangle models: although the sublexical route is based on graphemes, it also develops sensitivity to the letters surrounding a given correspondence due to learning in a two-layer associative network.

The reliance on letter clusters that are larger than graphemes in the triangle models blurs our proposed distinction between complexity and unpredictability: Given sufficient training, the orthography-to-phonology mapping process will establish orthography-phonology connections between different types of units (e.g., Plaut et al., 1996). This will make even a word like “ache” (cf. “cache”) predictable (via the whole-word correspondence that “ache” maps onto /æ k/, and “cache” onto /kæ /). In practice, the amount of training which the triangle models undergo is not sufficient to establish whole-word representations of low-frequency words, such that the orthography-to-phonology pathway will struggle with low-frequency irregular words. Due to the shared-labour nature of the two routes, semantic processing will be activated during any reading task, but in the case of low-frequency irregular words the semantic route is required for a correct output.

In summary, the DRC model has a sublexical route which operates on a set of pre-determined print-to-speech correspondences. As long as a word complies to the print-to-speech correspondences, the sublexical route will be able to provide a correct output. If the word is irregular (i.e., it does not comply to the rules), lexical knowledge needs to be consulted. The triangle models, conversely, develop reliance on units that are larger than graphemes, meaning that even high-frequency irregular words can be read aloud by the orthography-to-phonology route. While a DRC-like model could be modified to contain larger-than-grapheme units, these would remain completely independent of any lexical process (e.g., see footnote 1 of Coltheart et al., 1993). This distinction between larger sublexical units and whole-word orthography-phonology correspondences is not explicit in the triangle framework. However, there is a qualitative difference between sublexical clusters and whole-word units, namely that whole-word units map onto semantic information while sublexical units, by definition, do not. Therefore, we propose that within any framework, resolving print-to-speech ambiguities by relying on larger-than-grapheme sublexical (i.e., complex) units, and resolving ambiguities by relying on whole-word units (for unpredictable words) are qualitatively different strategies, in that the latter inevitably involves excitatory activation of lexical-semantic processes.

For the sake of simplicity, we provide definitions of complexity and unpredictability which are in line with the DRC terminology. As explained above, this does not mean that our framework is incompatible with other models of reading, as all models can accommodate both the presence of complex correspondences and words with unpredictable pronunciations. We refer to a correspondence between orthography and phonology as complex if either the orthographic element involved consists of more than one letter (e.g., “th” → /θ/), or if the correspondence is context-sensitive (“g[i]” → /d /), or if both are true (“ch[r]” → /k/).

Unpredictable words can be defined as words where the sublexical route provides an incorrect pronunciation. In the DRC framework, such words are termed *irregular* words (e.g., Andrews, 1982; Content, 1991; Rastle & Coltheart, 1999; Schmalz, Marinus, & Castles, 2013). Within the connectionist models the concept of consistency is stressed. Although consistency differs from regularity, in the context of the current review it reflects the predictability of a word: a consistent word is defined as one where “its pronunciation agrees with those of similarly spelt words” (Plaut et al., 1996, p. 59). This reflects the mechanisms by which the pronunciation of a word is assembled in a connectionist model: as the sublexical route operates based on statistical regularities which are derived from the print-to-speech

correspondences of real words, unpredictable words in this framework are those which have a different pronunciation to similarly spelled words.

In summary, as a working definition, we refer to complex correspondences as those that are multi-letter (“th” → /θ/) and/or context-sensitive (“g[i]” → /d /), and to unpredictable words as irregular words given the set of GPCs that are implemented in the DRC. Given that the definitions are arguably biased towards the DRC framework, we seek for convergence in alternative approaches for all our findings in the following section.

Quantifications of orthographic depth

Existing measures of orthographic depth, and their relation to complexity and unpredictability

Having distinguished between complex versus unpredictable correspondences, we can attempt to devise a quantification method for each of these on a linguistic level. If these represent two separate concepts underlying orthographic depth, the first step is to demonstrate that they vary independently across orthographies. This will firstly show whether there is enough independent variation of the two concepts to warrant practically meaningful investigation on a behavioral level, and secondly will provide insights as to how orthographic depth may be quantified. Although large-scale linguistic corpus analyses are outside the scope of the current review, we provide some suggestions that can be expanded on by future work. We discuss and expand on previous quantification methods, and consider their advantages and disadvantages. In terms of demonstrating that there is a dissociation between complexity and unpredictability, we refer to a computational-model-driven approach (Ziegler, Perry, & Coltheart, 2000), and a linguistic-corpus-analysis approach (van den Bosch, Content, Daelemans, & de Gelder, 1994). We also discuss two commonly taken approaches to determine the relative depth of a given orthography, namely subjective consensus among experts (e.g., Frost et al., 1987; Seymour et al., 2003), and the onset entropy measure (Borgwaldt, Hellwig, & de Groot, 2004, 2005).

Given our DRC-based definitions of complexity and unpredictability, it is intuitive to start by using the existing versions of the DRC across orthographies as an attempt to illustrate cross-linguistic differences given the GPC rules. Specifically, we can simply take the numbers and proportions of complex rules, and the proportion of irregular words in the DRCs of the orthographies in which it has been implemented. The number of complex rules (those which are multi-letter

and/or context-sensitive) is a measure of complexity as per our working definition.³

This approach of comparing the number and types of GPC rules across orthographies, and the degree to which they are sufficient to read aloud words in a given orthography, has been also taken by Ziegler et al. (2000) when they implemented the DRC in German. They found that both the number of rules – and especially the number of complex rules – and the percentage of irregular words was higher in English than in German. This is in line with the general consensus that German is a shallow orthography, and English is deep. The DRC has also been implemented French (Ziegler, Perry, & Coltheart, 2003), Dutch, and Italian (C. Mulatti, personal communication, 25 May 2014), which allows us to list the numbers, proportions and types of rules in these five DRCs.⁴ The results of this analysis are presented in Table 1.

Table 1 shows, as expected, that English is a “deep” orthography, in that it has many rules, and a particularly high percentage of irregular words, while Dutch and German are “shallow”, in that they have few rules and a small proportion of irregular words. Interestingly, the DRC approach places the French orthography at one end of the continuum for the number/percentage of complex rules (complexity) – according to which French appears to be even more complex than the English orthography – and at the other end of the continuum for the percentage of irregular words (unpredictability) – where French appears to be even more predictable than German and Dutch. This shows that the distinction between the two concepts is meaningful, as they are not perfectly correlated between orthographies. The Italian DRC shows an even smaller number of rules, and a larger proportion of single-letter rules, consistent with the notion that it is an extremely shallow orthography (e.g., Tabossi & Laghi, 1992).

³ The working definition ignores two additional source of GPC complexity. The first stems from the position-specificity of rules. As position-specific rules are implemented as separate GPC rules, as opposed to rules which apply to all positions, the presence of position-specific GPC rules inflates the number of rules overall, and therefore this source of complexity is reflected in the total number of rules (see Table 1). A second source of GPC complexity is the presence of phonotactic rules, which are context-dependent rules where the influence of a certain phoneme which precedes or succeeds a letter influences its pronunciation. Such rules are particularly common in some orthographies, such as Russian, but have not been researched to a great extent in other orthographies. For example, neither the current implementation of the German, Dutch, or Italian DRC contain any phonotactic rules, even though these might be more suitable to describe numerous aspects of the print-to-speech conversion. We therefore excluded phonotactic GPC rules in Table 1.

⁴ Table 1 lists the *number* as well as the *percentages* of each type of GPC rule. From a theoretical standpoint, saying that a language has a large number of rules is different from saying that the rules of a given language are complex. In the set of orthographies that we present here, the number of rules and percentage of complex rules are highly correlated, therefore in practice they cannot be used to dissociate between the two different measures.

Although the DRC approach offers insights into the relative positions of the four orthographies on the two continuums, there are three reasons why this approach is limited. Firstly, current versions of the DRC are based on monosyllabic words only. In some languages, the proportion of monosyllabic words is relatively high; for others, polysyllabic words form the majority of all words. This is problematic for across-language comparisons. Furthermore, even in languages where monosyllabic words are frequent, structural properties vary between monosyllabic and polysyllabic words. Therefore, monosyllabic words are not a perfectly representative sample of any orthography (for a review, see Protopapas & Vlahou, 2009). Although the DRC approach may still be useful to determine the relative position of each orthography in terms of orthographic complexity and unpredictability, it would be valuable to replicate these findings with an approach which is not limited to monosyllabic words.

Secondly, the cross-linguistic versions of the DRC were implemented independently of each other, without the aim of comparing them directly to each other. For example, the number of words in the DRC’s lexicons varies extensively, with 4583 words for Dutch, 8027 words for English, 2245 words for French, and 1448 words for German. The varying number of words in the DRCs may also reflect the relative percentage of monosyllabic words in each language.

Thirdly, it is not established that the GPC rules that are implemented in the DRC have full psychologic reality. Indeed, there is evidence that other sublexical units are used during reading in English (Glushko, 1979; Treiman, Kessler, & Bick, 2003), German (Perry, Ziegler, Braun, & Zorzi, 2010; Schmalz et al., 2014) and French (Perry, Ziegler, & Zorzi, 2014). Although this does not mean that the DRCs cannot be used as a tool to capture linguistic variability in the complexity and predictability of print-to-speech correspondences by using GPCs and irregular words as a proxy, it is, again, desirable to find converging evidence from a different approach.

Such converging evidence can be found from a computational study of a linguistic corpus of English, Dutch, and French (van den Bosch et al., 1994). The corpuses used in this study included polysyllabic words as well as monosyllabic words. In addition, this paper predates the DRCs, and has not been conducted within the framework of the any particular theory or model. The approach of this paper was data-driven, and the authors made no *a priori* predictions about the results.

Van den Bosch et al. (1994) conclude that orthographic depth can be dissociated into two separate measures: the difficulty of parsing letter strings into graphemes on the one hand, and the degree of redundancy in the print-to-speech correspondences on the other hand. The former – which we equate with our concept of complexity – was measured by applying a computationally obtained parsing mechanism to a set of test words. In an orthography with simple correspondences, parsing is easier, because in many cases parsing a

Table 1 Measures of complexity and unpredictability for Dutch, English, French, German, and Italian based on the DRCs, and from van den Bosch et al., (1994) for comparison

Measure	Dutch	English	French	German	Italian
Total number of rules (DRC)	104	226	340	130	59
Single-letter rules (DRC)	51 (49.0 %)	38 (16.9 %)	46 (13.5 %)	44 (33.8 %)	19 (32.2 %)
Multi-letter rules (DRC)	42 (40.4 %)	161 (71.2 %)	218 (64.1 %)	55 (42.3 %)	8 (13.6 %)
Context-sensitive rules (DRC)	11 (10.6 %)	27 (11.9 %)	76 (22.4 %)	31 (23.8 %)	32 (54.2 %)
Irregular words (%)	6.3	16.9	5.6	10.5	NA
Parsing accuracy (%)	21.3	24.5	12.9	NA	NA
Generalization accuracy (%)	81.4	54.3	89.1	NA	NA

Note: For the DRC, the numbers represent the number of rules of each type, and the percentage out of the total number of rules in brackets. Results for parsing accuracy and generalisation accuracy (defined below) are taken from van den Bosch et al. (1994)

word into letters would enable the correct mapping of graphemes to phonemes: an English word with simple correspondences, like “cat”, would be parsed into “c”, “a” and “t,” which can be mapped correctly onto the phonemes of /k/, /æ/, and /t/; a word with complex correspondences, such as “chair” would instead need to be parsed as “ch” and “air,” because the constituent letters (“c”, “h”, etc.) do not map onto the correct phonemes. Since for each of the three orthographies the same amount of training was used, differences in parsing accuracy of untrained test words reflect the difficulty of parsing. Parsing accuracy, overall, was low, indicating that all three orthographies are characterised by high complexity. French showed the lowest level of accuracy, while Dutch and English were at approximately the same level (see Table 1).

For quantifying the degree of redundancy, van den Bosch et al. (1994) report the generalization performance, or the number of test words pronounced correctly by a computationally obtained set of print-to-speech correspondences in the three orthographies. In order to obtain these correspondences, they first derived all possible print-to-speech correspondences of all sizes (ranging from single letters to whole words). Then they compressed the set of correspondences for each orthography to reduce the redundancy among these rules (e.g., knowing the correspondences “a” → /æ/ and “t” → /t/, as well as “-at” → /æt/, is redundant; knowing the correspondences “a” → /æ/, “l” → /l/, and “m” → /m/, as well as “-alm” → / m/ is not).

The results showed that both Dutch and French outperformed English, meaning that there are many English words that do not comply with these rules. The generalization measure is reflective of unpredictability: given the set of correspondences that were defined through the compression algorithm, a large number of words in the English orthography were still unpredictable. The predictability, according to this measure, was higher in Dutch and French than in English. The summary of both the variables is presented in Table 1.

To our knowledge, the quantification scheme of van den Bosch et al. (1994) has not been used to study behavioral differences in the effects of orthographic depth, nor has it been

applied to other orthographies. This is an important direction for future research. For our current purposes, it is particularly interesting that the two concepts which van den Bosch et al. (1994) suggest as underlying orthographic depth based on their linguistic-computational analysis are consistent with the results of the DRC, and our distinction between complexity and unpredictability. It is worth noting that the results of the DRC approach and the analysis of van den Bosch et al. (1994) converge despite the DRCs’ limitation to monosyllabic words only. This suggests that for the orthographies studied, general findings about complexity and unpredictability show broadly the same patterns for monosyllabic words compared to a wider sample of words.

The case of French is particularly interesting: Both the DRC approach and the analysis of van den Bosch et al. (1994) classified French as a relatively complex orthography (many complex GPC rules, low generalization performance) – even compared to English. Conversely, both approaches classified French as a predictable orthography – even compared to Dutch and German. In previous work on orthographic depth, French has often been described as an intermediate orthography (Goswami et al., 1998; Paulesu et al., 2001; Seymour et al., 2003; Sprenger-Charolles et al., 2011). The French orthography, therefore, shows the importance of distinguishing between the two concepts, as a failure to do so provides a different picture.

This intuitive classification of French as an orthography of intermediate depth has been supported by some of the previous quantification schemes, which did not make the distinction between complexity and unpredictability. For example, Seymour et al. (2003) classified 13 European orthographies based on their degree of depth.⁵ They consulted researchers in

⁵ Seymour et al. (2003) draw a distinction between orthographic depth and complexity, but what is meant by complexity is not what we mean by the same term: they refer to the presence of consonant clusters (i.e., syllabic complexity), that do not necessarily map onto the same phoneme (e.g., “str”). This does not relate to orthographic depth, but constitutes a different dimension.

participating countries and ranked the orthographies in terms of their depth based on a more intuitive approach. This landed French in an “intermediate” position. It seems, therefore, that this intuitive approach “averages out” potentially theoretically relevant distinctions between separate concepts underlying orthographic depth.

A more objective approach, which has been picked up by cross-linguistic researchers, has been introduced by a measure called *onset entropy* (Borgwaldt et al., 2004, 2005). This quantification scheme reflects the number of different ways in which the initial letter of a word, on average, can be pronounced in a given orthography. Initial letters which consistently map onto the same phoneme involve no ambiguity, so they are assigned a value of 0. The greater the number of possible pronunciations of the letter, the higher the entropy value. Borgwaldt et al. (2005) calculated the entropy values for initial letters across orthographies. The average onset entropy for each orthography was then considered to reflect its relative degree of orthographic depth.

This measure has intuitive appeal, and has been used in large-scale behavioral studies of cross-linguistic differences (Landerl et al., 2013; Moll et al., 2014; Vaessen et al., 2010; Ziegler et al., 2010). One of its advantages is the focus on the first letter only. Firstly, this eliminates the bias towards monosyllabic words that is present in the DRC and some other approaches. Secondly, it also increases the comparability across orthographies, because words in all orthographies have initial letters (Borgwaldt et al., 2005; Ziegler et al., 2010). Still, neglecting additional information in a word provides other problems. In English, for example, it is often the vowel pronunciation that is unpredictable, and vowels occur more frequently in the middle of a word (Treiman et al., 1995). In French, print-to-speech irregularities occur mostly in the final consonants, which are often silent (Lete, Peereman, & Fayol, 2008; Perry et al., 2014; Ziegler, Jacobs, & Stone, 1996).

We provide two examples that show that although the onset entropy measure is a useful first step in quantifying orthographic depth, it confounds orthographic complexity with unpredictability, meaning that it does not provide the whole picture. According to the onset entropy measure, French (with a value of 0.46) is about half-way between English (0.83) and “shallow” orthographies such as Finnish (0.0) and Hungarian (0.17) (Ziegler et al., 2010). Like the subjective rankings described above, it seems therefore that this approach to quantifying orthographic depth averages out two different sources that underlie this construct.

Another example is the German orthography. Table 1 shows that according to the DRC measure, German has a high degree of predictability: although some context-sensitive rules are required, these allow the sublexical route to read aloud 90 % of all monosyllabic words correctly. This contrasts with the results from the onset entropy measure, which classifies it as relatively deep: the onset entropy value for German is

higher (reflecting higher degree of depth) than that of Dutch, Hungarian, Italian, and even Portuguese, and only slightly lower than French (Borgwaldt et al., 2005). This goes against both the results of the DRC, and the intuitive notion that German is close to the shallow end of the orthographic depth continuum. (Frith et al., 1998; Goswami, Ziegler, Dalton, & Schneider, 2003; Landerl, 2000; Landerl et al., 1997; Seymour et al., 2003; Wimmer & Goswami, 1994; Wimmer et al., 2000; Ziegler et al., 2001; Ziegler, Perry, Ma-Wyatt, et al., 2003).

Similar to the example of French, this counter-intuitive finding can be explained by the distinction between complexity and unpredictability: the *relative complexity* of the German orthographic system inflates the onset entropy value, despite German’s relatively high degree of predictability. For example, German words starting with the letter “s” can have the first phoneme /z/, /l/, or /s/. The pronunciation is, however, *predictable*: in the onset position, when “s” is succeeded by a vowel, it is pronounced as /z/, when it is succeeded by “p” or “t” or is part of the grapheme “sch”, it is pronounced as /l/, and in all other cases it is pronounced as /s/. The two examples of French and German show that onset entropy thus has no way of distinguishing between correspondence complexity and unpredictability, and instead “averages out” the two dimensions, thus making French and German appear to be “intermediate” orthographies despite their relatively high predictability.

In summary, we have described two separate approaches that both suggest that orthographic depth is not a single concept, but can be dissociated into complexity and unpredictability of the print-to-speech correspondences. One was introduced two decades ago (van den Bosch et al., 1994), but to our knowledge it has not been extended to other orthographies or formed the basis of behavioral research. For the purpose of the current paper, the study is valuable because the data-driven computational-linguistic study by van den Bosch et al. (1994) led them to the same conclusions as our theory-based DRC approach. This strengthens the position that on a linguistic level, orthographic depth can be dissociated into two separate constructs.

Limitations and open questions for further research

Our definition of orthographic depth was conceptualized with the aim of being specific, as this is essential for an objective classification measure and for precise predictions about behavior, based on theories and models of reading. The specificity of our definitions comes with the trade-off that it does not capture all sources of cross-linguistic variability. For example, unpredictability, when defined at the level of the print-to-speech correspondences, ignores two further sources of unpredictability that exist in alphabetic orthographies, namely

incompleteness and irregularities associated with lexical stress assignment.

As discussed earlier, a previous definition of orthographic depth also included the concept of incompleteness (Katz & Frost, 1992). Incomplete sublexical information, within our framework, makes the pronunciation of a word unpredictable for the sublexical route, but also requires the use of contextual semantic information to access a single phonological and semantic entry. Given the need to rely on semantic context to resolve this type of unpredictability, it is possible that the incompleteness of the sublexical correspondences presents a qualitatively different problem compared to complexity and consistency. If this is the case, placing an orthography which is characterised by incomplete correspondences, such as Hebrew, on the same continuum as the European orthographies, might not be particularly meaningful.

Another source of unpredictability that varies across orthographies, but is not captured by any of the previous quantification schemes, is lexical stress assignment. Some orthographies, such as French, have entirely predictable stress assignment, but others, such as English (Rastle & Coltheart, 2000; Seva, Monaghan, & Arciuli, 2009), Greek (Protopapas, Gerakaki, & Alexandri, 2006), Russian (Jouravlev & Lupker, 2014, 2015), and Italian (Burani & Arduino, 2004; Colombo, 1992), have some ambiguity when it comes to determining the position of the stressed syllable, and lexical-semantic knowledge needs to be recruited to resolve these conflicts. In English, for example, the word “entrance” a different meaning depending on whether the first or the second syllable is stressed. It is still, to some extent, unclear via what mechanisms stress irregularity affects reading (Sulpizio, Arduino, Paizi, & Burani, 2013; Sulpizio, Burani & Colombo, 2015), and how it relates to GPC irregularity. Therefore, this leaves open questions for future research: for example, to what extent can stress assignment be predicted, how this differs across orthographies, and what cognitive mechanisms are used to resolve ambiguities underlying stress assignment.

Defining language-level differences across orthographies becomes even more complicated when we consider non-alphabetic orthographies: both the languages and the orthographic systems of Chinese or Japanese, for example, are so different to the alphabetic orthographies that we consider here, that classifying and comparing them along the same continuum is not possible. In addition to differences in the nature of the process by which speech is computed from print, non-alphabetic orthographies may differ in terms of the visual complexity, morphological principles, and even definitions of word boundaries (Chang, Maries, & Perfetti, 2014; Cui et al., 2012; Huang & Hanley, 1995; McBride-Chang et al., 2012).

Therefore, we believe that the most valuable studies in terms of getting to the bottom of orthographic depth would involve the following: (1) Cross-linguistic comparisons, where two orthographies which are similar on as many aspects

as possible, but different on the particular issue of interest. Given the difficulty in doing this, a single comparison involving only two orthographies should not be taken at face value, and needs to be replicated with other orthographies. (2) Within-language studies can be conducted to isolate the particular aspect of orthographic depth that is proposed to drive cross-linguistic differences. For example, Frost (1994) compared marker effects of lexical-semantic processing in pointed Hebrew, where the sublexical information is complete, to unpointed Hebrew, where the sublexical information for the same words is incomplete. In line with the Orthographic Depth Hypothesis, the unpointed script showed stronger lexical-semantic marker effects than pointed Hebrew, in a design that controlled for any cross-linguistic differences that may exist in an across-language design.

Predictions of the new orthographic depth framework for theories of reading

Some key studies within the new framework

Previous research on orthographic depth has been conducted without bearing in mind the distinction of complexity and predictability. Therefore, this work is often subject to more than one interpretation, depending on whether behavioral differences are proposed to arise as a function of complexity, or as a function of unpredictability. We review two previous key studies on orthographic depth and illustrate how different conclusions may be drawn depending on how orthographic depth is defined.

A key finding supporting the Orthographic Depth Hypothesis (ODH) comes from a study of the frequency and lexicality effects, and a semantic priming manipulation (Frost, Katz, & Bentin, 1987). The orthographies explored in this study were Hebrew (deep), English (medium-deep), and Serbo-Croatian (shallow). Indeed, there was an increase in the size of the lexical-semantic effects associated with increasing orthographic depth, suggesting stronger involvement of the lexical-semantic route.

Upon closer inspection, it is not clear whether this can be attributed to orthographic complexity or unpredictability. The Hebrew orthography is characterized by incompleteness, therefore these results show that incompleteness of print-to-speech correspondences leads to increased reliance on the lexical-semantic route, but remain silent about complexity and unpredictability. English and Serbo-Croatian differ to each other on both complexity and unpredictability, therefore the difference between these two orthographies can be attributed to either. If unpredictability drives the increased involvement of the lexical-semantic route, this means that lexical knowledge is recruited because sublexical information is not sufficient to assemble a fully-specified phonological representation. If complexity drives these behavioral differences, this

means that the presence of complex information slows down the processing of the sublexical route, which allows for more involvement of the lexical route. Thus, although the outcomes of the two scenarios are identical, the mechanisms that lead to this end state are different. As a result, we do not know whether indeed complexity has the effect of increased reliance on lexical-semantic processing, or whether this is specific to incompleteness and unpredictability. In future research, this question could be addressed by comparing complex but predictable orthographies, such as French, to simple and predictable orthographies, such as German or Dutch. If increased lexical-semantic processing is associated with unpredictability, but not complexity, we would expect to find similar lexical-semantic marker effects when we hold predictability constant.

Key evidence for the Psycholinguistic Grain Size Theory (Ziegler & Goswami, 2005) stems from a study comparing the size of the length and body-N effects in English and German (Ziegler et al., 2001). The length effect was stronger in German compared to English, and the body-N effect was weaker in German compared to English, suggesting differences in the nature of the sublexical processing underlying reading in the two orthographies. German and English differ to each other on both complexity and unpredictability, so it is unclear which aspects of these writing systems drive the behavioral differences. An increased body-N effect in English compared to German may reflect a difference in the nature of the sublexical processing (as suggested by Ziegler et al., 2001). If the dominant functional sublexical units of English are bodies, this would mean that the sublexical units are more complex. According to this interpretation, the results of the body-N effect reflect a difference in the complexity of sublexical correspondences. An alternative explanation is that the unpredictability of English encourages a qualitatively different reading strategy compared to German, namely an increased reliance on lexical analogy. In this case, a German reader might tend towards reading words and non-words via the sublexical correspondences, whereas an English readers relies to a greater extent on similar lexical entries. Thus, English readers might show a stronger body-N effect compared to German readers, because they are facilitated by the presence of orthographically similar words. Again, studies with orthographies which are matched on complexity but differ in terms of predictability (e.g., English – French) or vice versa (e.g., German – French) might be used by future research to dissociate between effects that are associated with each of these two constructs.

In summary, the next step for future research will be to conduct behavioral studies to establish the extent to which the two dimensions affect reading processes. This opens up a plethora of new research questions about the mechanisms via which the variables underlying orthographic depth independently affect reading, or learning to read. We can show that the distinction between complexity and unpredictability of sublexical correspondences is theoretically meaningful if,

based on existing models of reading, there are different predictions about how the two constructs affect cognitive processing. To this end, we provide an overview of specific predictions within existing models of reading about how both complexity and unpredictability, as defined above, might affect both skilled reading processes and reading acquisition.

Predictions for complexity and unpredictability in adults

In skilled adult readers, the ODH proposes that complexity slows down the sublexical assembly process, which gives more time for the lexical route to access the relevant word information. Based on computational models of reading, we would indeed expect that the complexity of print-to-speech correspondences should affect the speed of sublexical assembly. Simulations with the DRC, as well as behavioral data, have shown that non-words which contain multi-letter GPCs (“boace”) are processed more slowly than non-words of equal length, but containing only simple correspondences (Marinus & de Jong, 2011; Rastle & Coltheart, 1998; Rey et al., 1998; Rey, Ziegler, & Jacobs, 2000). This is postulated to occur because the sublexical route in the DRC operates in a serial fashion. When reading an item containing a multi-letter rule, it activates the first letter of the digraph and its equivalent pronunciation. This pronunciation needs to be inhibited once the second letter starts being processed, because the two letters are then parsed into a two-letter grapheme which has a different pronunciation.

To our knowledge, it has not yet been explicitly shown that the slowing-down of sublexical assembly leads to an increased reliance on the lexical procedure, as stated by Katz and Frost (1992), but this is a question that can be easily addressed by future empirical research. One set of predictions that would follow is that, in words containing complex correspondences, lexical and semantic markers such as frequency or imageability effects should be stronger compared to words containing simple correspondences only.

The concept of unpredictability, and its effect on single-word reading, has already been addressed in detail by computational modellers in the form of a debate between regularity and consistency. As explained in an earlier section, this debate reflects different mechanisms that are used by the sublexical routes of computational models to derive the pronunciation of a letter string. Importantly, most studies have shown an inhibitory effect of unpredictability, based on their measure of choice (Andrews, 1982; Hino & Lupker, 2000; Jared, 1997, 2002; Jared, McRae, & Seidenberg, 1990; Metsala, Stanovich, & Brown, 1998; Parkin, McMullen, & Graystone, 1986; Rastle & Coltheart, 1999; Waters & Seidenberg, 1985; Waters, Seidenberg, & Bruck, 1984).

Open questions remain about the relationship between consistency and completeness. While scenarios of inconsistency can be resolved by relying on a non-semantic lexical route,

access of phonological information for incomplete words relies on semantics. Thus, it is of interest whether semantic effects may be stronger for words with incomplete versus inconsistent correspondences. This has theoretical implications for models of reading as well as defining orthographic depth: a difference between triangle models and the dual-route models (both the DRC and CDP+-type models) is that the latter contain a purely orthographic lexical route that bypasses the activation of semantics. In triangle models, such a route is non-existent: all lexical activation passes through a semantic route. Therefore, dual-route models may predict a difference in the strength of semantic marker effects between unpredictable and incomplete words. Conversely, triangle models require semantics for both types of items, and predict equally strong semantic effects for both inconsistent and incomplete words.

In this section, we have listed several testable predictions, based on models of skilled reading, which can be explored by future research using either within- or across-language designs. This will contribute to the understanding of the precise cognitive mechanisms which drive the cross-orthographic differences that have been previously attributed to the broad concept of orthographic depth.

Theories of reading acquisition and orthographic depth

There are fewer specified models of reading acquisition than models of adult reading. In the case of exploring the effect of orthographic depth on reading acquisition and making specific predictions, computational models could be particularly useful. Connectionist-type models, for example, use a learning algorithm that extracts the regularities in the correspondences between print and speech. Thus, they are faced with a similar problem as a child learning to read (Hutzler, Ziegler, Perry, Wimmer, & Zorzi, 2004). If, for the sake of simplicity, we focus purely on the acquisition of sublexical skills, we can make clear predictions about complexity and unpredictability. As stated by the Psycholinguistic Grain Size Theory (PGST), complexity of the sublexical correspondences should make it more difficult to acquire these for children learning to read (Ziegler & Goswami, 2005). This means that becoming proficient at using sublexical decoding should take longer in an orthography with complex correspondences than in an orthography with simple correspondences.

In terms of learning the sublexical correspondences, we can also make clear predictions about unpredictability that should be testable both with a connectionist-type model and in children learning to read. If we were to pick two orthographies that are comparable in terms of complexity, but different in terms of predictability (e.g., French and English), we would expect that learning these correspondences would take the same amount of time. However, after the correspondences are learnt, we would expect that the accuracy in applying these

correspondences to new words would be higher in the more predictable orthography.

Both behavioral and computational data from English and German provide some support for this set of predictions. For example, behavioral data has shown that non-word reading accuracy is higher for German than English children (Frith et al., 1998; Landerl, 2000; Ziegler, Perry, Ma-Wyatt, et al., 2003). This holds true even when a lenient marking criterion is used for English, whereby any plausible pronunciation of the non-word is scored as correct. Similar data has also been obtained from comparisons of English with other shallow orthographies, most notably from a large-scale study which included children from 13 European countries (Seymour et al., 2003). A computational study has compared the performance of a sublexical learning algorithm in German and English (Hutzler et al., 2004). In these simulations, the model's non-word reading accuracy of German exceeded that of English, even after a large number of training cycles when the models had reached a plateau.⁶

A limitation of both the computational and behavioral comparisons is that the orthographies differ to English both in terms of complexity and unpredictability. Therefore, although the existing data is suggestive, we cannot unequivocally attribute the differences in non-word reading accuracy to unpredictability.

It is also important to bear in mind that the sub-skills underlying reading do not develop in isolation. In particular, there is a bidirectional relationship between the acquisition of the lexical and the sublexical route (Share, 1995; Ziegler, Perry, & Zorzi, 2014): lexical entries are predominantly established by a self-teaching mechanism which uses knowledge of the sublexical correspondences to decode unfamiliar words, but lexical entries are also used to refine the knowledge of sublexical correspondences. Due to this bidirectional relationship, we expect that high complexity will not only delay the acquisition of sublexical skills, but also the build-up of orthographic entries. As a result, complexity should lead to a quantitative difference where, for example, French children should lag behind German children, but eventually reach the same level of decoding accuracy for both words and non-words.

In the case of unpredictability, there could be some *qualitative* differences in the mechanisms that are used for self-teaching, as lexical involvement is necessary to resolve ambiguous pronunciations. Recent within-language studies of orthographic learning provide some support for this notion, and in particular for the use of semantics (Taylor, Plunkett,

⁶ A simulation study like that by Hutzler et al. (2004) could also be used to test the claims we make about the effect of complexity on learning to read. However, as Hutzler et al. used a lenient marking criterion to evaluate the models' non-word reading performance, it is hard to make direct comparisons based on their reported data about the speed of acquisition of the sublexical correspondences.

& Nation, 2011; Wang, Castles, & Nickels, 2012). In an orthographic learning study, participants are asked to learn new words. These can be assigned either a predictable or an unpredictable pronunciation. Both studies found that when the pronunciation was unpredictable, semantic context facilitated learning. This was not the case for predictable words, where phonological decoding appeared sufficient for orthographic learning. These findings raise questions about cross-linguistic differences in learning to read as a function of unpredictability of sublexical correspondences. It is possible that children learning to read in a relatively unpredictable orthography routinely rely to a greater extent on contextual cues compared to children learning to read in a relatively predictable orthography. This would result in a qualitative shift in the types of cognitive strategies that are used across orthographies differing in predictability to establish orthographic representations.⁷

As our framework proposes that orthographic depth does not represent a single continuum, it contrasts with the view that the ease of learning to read depends on the relative position of one's orthography on this continuum (Seymour et al., 2003). Our suggestion is consistent with some previous studies, which have found a non-linear relationship between reading achievement and orthographic depth: specifically, once English is removed, the correlation between reading outcomes and orthographic depth disappears (Aro & Wimmer, 2003; Whetton & Twist, 2003). According to our framework, these results indicate that reading acquisition in English is impeded both by the high degree of complexity and unpredictability. The cross-linguistic comparisons of French and English consistently show that French children fare better than English children on word and non-word reading tasks (e.g., Goswami et al., 1998; Seymour et al., 2003), which suggests that complexity is not sufficient to account for the well-established lag of English-speaking children. Future research is needed to establish whether English children learning to read are predominantly hampered by the unpredictability of their orthography, or whether complexity and unpredictability interact to create a particularly high hurdle for children learning to read in English.

In summary, given the theories of reading acquisition, we can assume that the two concepts of complexity and unpredictability should affect cognitive processes during learning to read in different ways. Looking purely at the development of sublexical decoding skills, we expect that complexity would slow down the speed of reading acquisition, whereas unpredictability would reduce decoding accuracy, even after all the correspondences have been learned. Furthermore, if children

are routinely faced with unpredictable words, it is possible that they need to develop compensatory strategies to achieve high reading accuracy and comprehension.

Conclusions

Behavioral studies of orthographic depth have been conducted for decades, and have shown that it affects the cognitive processes underlying skilled adult reading (Frost et al., 1987; Schmalz et al., 2014; Ziegler et al., 2001), the rate of reading acquisition (Frith et al., 1998; Landerl, 2000; Seymour et al., 2003; Wimmer & Goswami, 1994), the prevalence and symptoms of developmental dyslexia (Landerl et al., 1997; Paulesu et al., 2001; Wimmer, 1993, 1996), brain activation (Paulesu et al., 2000; Richlan, 2014), and the strength of cognitive predictors of reading ability (Caravolas et al., 2012; Caravolas, Lervag, Defior, Malkova, & Hulme, 2013; Landerl et al., 2013; Moll et al., 2014; Vaessen et al., 2010; Ziegler et al., 2010). Clearly, orthographic depth is an important and relevant factor, both for practical and theoretical reasons. Based on the existing evidence, we can be confident in concluding that orthographic depth affects reading, but in order to learn more about why and how this happens, a more precise definition of orthographic depth is required. Such a definition is needed to (1) devise a quantification method of linguistic characteristics of the orthographies that is theoretically meaningful, and (2) to use this quantification method for future cross-linguistic research to isolate specific cognitive mechanisms that are affected by the linguistic constructs.

We propose that orthographic depth is a conglomerate of two separate concepts, namely the degree of complexity and unpredictability of print-to-speech correspondences in a given orthography. We have shown that, on a linguistic level, the two concepts can be dissociated. Furthermore, given the currently available models and theories of reading, we also expect that each of the two concepts would influence skilled reading and reading acquisition in different ways. Thus, we argue that there are many unanswered questions in the area of cross-linguistic research relating to orthographic depth. These can be pursued more effectively in the context of a systematic framework for orthographic depth.

Acknowledgments We thank Claudio Mulatti for providing information about the Italian DRC. Furthermore, we are grateful to Naama Friedmann and Anastasia Ulicheva for valuable discussions about orthographic depth and reading across languages, to Becky Treiman for organising a talk at WUSTL, where parts of this paper were presented, and to the anonymous reviewers for providing insightful comments on earlier versions of the manuscript.

⁷ Wang et al. were working within a DRC framework and manipulated GPC *regularity*. Taylor et al. were working within a more triangle-like framework, and manipulated *body consistency*. Their results and conclusions were strikingly similar.

References

- Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? *Memory & Cognition*, *10*(6), 565–575.
- Aro, M., & Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics*, *24*, 621–635.
- Bergmann, J., & Wimmer, H. (2008). A dual-route perspective on poor reading in a regular orthography: Evidence from phonological and orthographic lexical decisions. *Cognitive Neuropsychology*, *25*(5), 653–676. doi:10.1080/02643290802221404
- Borgwaldt, S., Hellwig, F., & de Groot, A. (2004). Word-initial entropy in five languages: Letter to sound, and sound to letter. *Written Language and Literacy*, *7*(2), 165–184.
- Borgwaldt, S., Hellwig, F., & de Groot, A. (2005). Onset entropy matters: Letter-to-phoneme mappings in seven languages. *Reading and Writing*, *18*, 211–229.
- Burani, C., & Arduino, L. S. (2004). Stress regularity or consistency? Reading aloud Italian polysyllables with different stress patterns. *Brain and Language*, *90*, 318–325. doi:10.1016/s0093-934x(03)00444-9
- Caravolas, M., Leråg, A., Mousikou, P., Efrim, C., Litavsky, M., Onochie-Quintanilla, E., . . . Hulme, C. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological Science*, *23*(6), 678–686. doi:10.1177/0956797611434536
- Caravolas, M., Lervag, A., Defior, S., Malkova, G. S., & Hulme, C. (2013). Different patterns, but equivalent predictors, of growth in reading in consistent and inconsistent orthographies. *Psychological Science*, *24*(8), 1398–1407. doi:10.1177/0956797612473122
- Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition*, *47*, 149–180.
- Chang, W., Maries, A., & Perfetti, C. (2014, 17/7/2014). *Visual Orthographic Variation across Writing Systems*. Paper presented at the Scientific Studies of Reading SSSR Santa Fe.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.
- Colombo, L. (1992). Lexical stress effect and its interaction with frequency in word pronunciation. *Journal of Experimental Psychology-Human Perception and Performance*, *18*(4), 987–1003. doi:10.1037//0096-1523.18.4.987
- Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud - dual-route and parallel-distributed-processing approaches. *Psychological Review*, *100*(4), 589–608.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. doi:10.1037//0033-295x.108.1.204
- Content, A. (1991). The effect of spelling-to-sound regularity on naming in French. *Psychological Research*, *53*(3), 3–12.
- Cui, L., Drieghe, D., Yan, G., Bai, X., Chi, H., & Liversedge, S. (2012). Parafoveal processing across different lexical constituents in Chinese reading. *The Quarterly Journal of Experimental Psychology*, 1-14. doi:10.1080/17470218.2012.720265
- Dehaene, S. (2009). *Reading in the brain: The new science of how we read*. London: Penguin.
- Feldman, L. B., & Turvey, M. T. (1983). Word recognition in Serbo-Croatian is phonologically analytic. *Journal of Experimental Psychology-Human Perception and Performance*, *9*(2), 288–298. doi:10.1037//0096-1523.9.2.288
- Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in German- and english-speaking children. *Scientific Studies of Reading*, *2*(1), 31–54.
- Frost, R. (1994). Prelexical and postlexical strategies in reading: Evidence from a deep and shallow orthography. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*(1), 116–129.
- Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences*, *35*(5), 263–279. doi:10.1017/S0140525x11001841
- Frost, R., & Bentin, S. (1992). Reading consonants and guessing vowels: Visual word recognition in Hebrew orthography. *Advances in Psychology*, *94*, 27–44.
- Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographic depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(1), 104–115.
- Glushko, R. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology-Human Perception and Performance*, *5*(4), 674–691.
- Goswami, U. (1999). The relationship between phonological awareness and orthographic representations in different orthographies. In M. Harris & G. Hatano (Eds.), *Learning to read and write: A cross-linguistic perspective* (pp. 134–156). Cambridge: Cambridge Press.
- 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.
- Goswami, U., Ziegler, J., Dalton, L., & Schneider, W. (2003). Nonword reading across orthographies: How flexible is the choice of reading units? *Applied Psycholinguistics*, *24*, 235–247. doi:10.1017/S0142716403000134
- Harm, M., & Seidenberg, M. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106*(3), 491–528.
- Hino, Y., & Lupker, S. J. (2000). Effects of word frequency and spelling-to-sound regularity in naming with and without preceding lexical decision. *Journal of Experimental Psychology-Human Perception and Performance*, *26*(1), 166–183. doi:10.1037//0096-1523.26.1.166
- Huang, H. S., & Hanley, J. R. (1995). Phonological awareness and visual skills in learning to read Chinese and English. *Cognition*, *54*(1), 73–98. doi:10.1016/0010-0277(94)00641-W
- Hutzler, F., Ziegler, J., Perry, C., Wimmer, H., & Zorzi, M. (2004). Do current connectionist learning models account for reading development in different languages? *Cognition*, *91*, 273–296.
- Jared, D. (1997). Spelling-sound consistency affects the naming of high-frequency words. *Journal of Memory and Language*, *36*(4), 505–529. doi:10.1006/jmla.1997.2496
- Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming. *Journal of Memory and Language*, *46*(4), 723–750. doi:10.1006/jmla.2001.2827
- Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language*, *29*(6), 687–715. doi:10.1016/0749-596X(90)90044-Z
- Jouravlev, O., & Lupker, S. J. (2014). Stress consistency and stress regularity effects in Russian. *Language Cognition and Neuroscience*, *29*(5), 605–619. doi:10.1080/01690965.2013.813562
- Jouravlev, O., & Lupker, S. J. (2015). Lexical stress assignment as a problem of probabilistic inference. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-015-0802-y
- Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 67–84). Amsterdam: Elsevier Science Publishers.
- Landerl, K. (2000). Influences of orthographic consistency and reading instruction on the development of nonword reading skills. *European Journal of Psychology of Education*, *15*(3), 239–257.
- Landerl, K., Ramus, F., Moll, K., Lyytinen, H., Leppänen, P. H. T., Lohvansuu, K., . . . Schulte-Körne, G. (2013). Predictors of developmental dyslexia in European orthographies with varying complexity. *Journal of Child Psychology and Psychiatry*, *54*(6), 686–694. doi:10.1111/Jcpp.12029

- Landerl, K., & Reitsma, P. (2005). Phonological and morphological consistency in the acquisition of vowel duration spelling in Dutch and German. *Journal of Experimental Child Psychology*, 92(4), 322–344. doi:10.1016/J.jecp.2005.04.005
- Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German-English comparison. *Cognition*, 63, 315–334.
- Lete, B., Peereman, R., & Fayol, M. (2008). Consistency and word-frequency effects on spelling among first- to fifth-grade French children: A regression-based study. *Journal of Memory and Language*, 58(4), 952–977. doi:10.1016/J.jml.2008.01.001
- Lieberman, I., Lieberman, A., Mattingly, I., & Shankweiler, D. (1980). Orthography and the beginning reader. In J. Kavenagh & R. L. Venezky (Eds.), *Orthography, reading, and dyslexia*. Baltimore: University Park Press.
- Marinus, E., & de Jong, P. F. (2011). Dyslexic and typical-reading children use vowel digraphs as perceptual units in reading. *Quarterly Journal of Experimental Psychology*, 64(3), 504–516. doi:10.1080/17470218.2010.509803
- McBride-Chang, C., Chen, H. C., Kasisopa, B., Burnham, D., Reilly, R., & Leppanen, P. (2012). What and where is the word? *Behavioral and Brain Sciences*, 35(5), 295–296. doi:10.1017/S0140525x1200009x
- Metsala, J. L., Stanovich, K. E., & Brown, G. D. A. (1998). Regularity effects and the phonological deficit model of reading disabilities: A meta-analytic review. *Journal of Educational Psychology*, 90(2), 279–293. doi:10.1037/0022-0663.90.2.279
- Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., . . . Landerl, K. (2014). Cognitive mechanisms underlying reading and spelling development in five European orthographies. *Learning and Instruction*, 29, 65–77. doi:10.1016/J.learninstruc.2013.09.003
- Parkin, A., McMullen, M., & Graystone, D. (1986). Spelling-to-sound regularity affects pronunciation latency but not lexical decision. *Psychological Research*, 48, 87–92.
- 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. doi:10.1126/science.1057179
- Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S. F., . . . Frith, U. (2000). A cultural effect on brain function. *Nature Neuroscience*, 3(1), 91–96.
- Peereman, R., & Content, A. (1998). Quantitative analyses of orthography to phonology mapping in English and French. 2014, from <http://homepages.vub.ac.be/accontent/OPMapping.html>
- Perry, C., Ziegler, J., Braun, M., & Zorzi, M. (2010a). Rules versus statistics in reading aloud: New evidence on an old debate. *European Journal of Cognitive Psychology*, 22(5), 798–812.
- Perry, C., Ziegler, J., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, 114(2), 273–315. doi:10.1037/0033-295X.114.2.273
- Perry, C., Ziegler, J., & Zorzi, M. (2010b). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology*, 61(2), 106–151.
- Perry, C., Ziegler, J., & Zorzi, M. (2014). When silent letters say more than a thousand words: An implementation and evaluation of CDP plus plus in French. *Journal of Memory and Language*, 72, 98–115. doi:10.1016/J.jml.2014.01.003
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103(1), 56–115. doi:10.1037/0033-295x.103.1.56
- Protopapas, A., Gerakaki, S., & Alexandri, S. (2006). Lexical and default stress assignment in reading Greek. *Journal of Research in Reading*, 29(4), 418–432. doi:10.1111/J.1467-9817.2006.00316.X
- Protopapas, A., & Vlahou, E. L. (2009). A comparative quantitative analysis of Greek orthographic transparency. *Behavior Research Methods*, 41(4), 991–1008. doi:10.3758/Brm.41.4.991
- Rastle, K., & Coltheart, M. (1998). Whammies and double whammies: The effect of length on nonword reading. *Psychonomic Bulletin & Review*, 5(2), 277–282.
- Rastle, K., & Coltheart, M. (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology-Human Perception and Performance*, 25(2), 482–503.
- Rastle, K., & Coltheart, M. (2000). Lexical and nonlexical print-to-sound translation of disyllabic words and nonwords. *Journal of Memory and Language*, 42(3), 342–364. doi:10.1006/Jmla.1999.2687
- Rey, A., Jacobs, A. M., Schmidt-Weigand, F., & Ziegler, J. C. (1998). A phoneme effect in visual word recognition. *Cognition*, 68(3), B71–B80. doi:10.1016/S0010-0277(98)00051-1
- Rey, A., Ziegler, J., & Jacobs, A. (2000). Graphemes are perceptual reading units. *Cognition*, 74, 1–12.
- Richlan, F. (2014). Functional neuroanatomy of developmental dyslexia: The role of orthographic depth. *Frontiers in Human Neuroscience*, 8, 1–13. doi:10.3389/fnhum.2014.00347
- Schmalz, X., Marinus, E., Robidoux, S., Palethorpe, S., Castles, A., & Coltheart, M. (2014). Quantifying the reliance on different sublexical correspondences in German and English. *Journal of Cognitive Psychology*, 26(8), 831–852.
- Schmalz, X., Marinus, E., & Castles, A. (2013). Phonological decoding or direct access? Regularity effects in lexical decisions of Grade 3 and 4 children. *Quarterly Journal of Experimental Psychology*, 66(2), 338–346. doi:10.1080/17470218.2012.711843
- Seidenberg, M., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96(4), 523–568.
- Seva, N., Monaghan, P., & Arciuli, J. (2009). Stressing what is important: Orthographic cues and lexical stress assignment. *Journal of Neurolinguistics*, 22(3), 237–249. doi:10.1016/J.jneuroling.2008.09.002
- Seymour, P., Aro, M., & Erskine, J. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94, 143–174.
- Share, D. (1995). Phonological recoding and self-teaching: *Sine qua non* of reading acquisition. *Cognition*, 55, 151–218.
- Share, D. (2008). On the anglocentricities of current reading research and practice: The perils of overreliance on an “Outlier” orthography. *Psychological Bulletin*, 134(4), 584–615.
- 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(6), 498–521.
- Sulpizio, S., Arduino, L. S., Paizi, D., & Burani, C. (2013). Stress assignment in reading Italian polysyllabic pseudowords. *Journal of Experimental Psychology-Learning Memory and Cognition*, 39(1), 51–68. doi:10.1037/A0028472
- Sulpizio, S., Burani, C., & Colombo, L. (2015). The process of stress assignment in reading aloud: Critical issues from studies on Italian. *Scientific Studies of Reading*, 19(1), 5–20.
- Tabossi, P., & Laghi, L. (1992). Semantic priming in the pronunciation of words in two writing systems: Italian and English. *Memory & Cognition*, 20(3), 303–313.
- Taylor, J. S. H., Plunkett, K., & Nation, K. (2011). The influence of consistency, frequency, and semantics on learning to read: An artificial orthography paradigm. *Journal of Experimental Psychology: Learning*, 37(1), 60–76. doi:10.1037/A0020126
- Treiman, R., Kessler, B., & Bick, S. (2003). Influence of consonantal context on the pronunciation of vowels: A comparison of human readers and computational models. *Cognition*, 88(1), 49–78. doi:10.1016/S0010-0277(03)00003-9

- Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, *124*(2), 107–136.
- Turvey, M., Feldman, L., & Lukatela, G. (1984). The Serbo-Croatian orthography constrains the reader to a phonologically analytic strategy. In L. Henderson (Ed.), *Orthographies and reading: Perspectives from cognitive psychology, neuropsychology, and linguistics* (pp. 81–89). Hillsdale, NJ: Lawrence Erlbaum.
- Vaessen, A., Bertrand, D., Tóth, D., Csépe, V., Faisca, L., Reis, A., & Blomert, L. (2010). Cognitive development of fluent word reading does not qualitatively differ between transparent and opaque orthographies. *Journal of Educational Psychology*, *102*(4), 827–842. doi:10.1037/a0019465
- van den Bosch, A., Content, A., Daelemans, W., & de Gelder, B. (1994). Measuring the complexity of writing systems. *Journal of Quantitative Linguistics*, *1*(3), 178–188.
- Wang, H. C., Castles, A., & Nickels, L. (2012). Word regularity affects orthographic learning. *Quarterly Journal of Experimental Psychology*, *65*(5), 856–864. doi:10.1080/17470218.2012.672996
- Waters, G., & Seidenberg, M. (1985). Spelling-sound effects in reading: Time-course and decision criteria. *Memory & Cognition*, *13*(6), 557–572.
- Waters, G., Seidenberg, M., & Bruck, M. (1984). Children's and adults' use of spelling-sound information in three reading tasks. *Memory & Cognition*, *12*(3), 293–305. doi:10.3758/bf03197678
- Whetton, C., & Twist, L. (2003). *What Determines the Range of Reading Attainment in a Country?* Paper presented at the International Association for Educational Assessment Conference, Manchester, UK.
- Wimmer, H. (1993). Characteristics of developmental dyslexia in a regular writing system. *Applied Psycholinguistics*, *14*(1), 1–33. doi:10.1017/S0142716400010122
- Wimmer, H. (1996). The nonword reading deficit in developmental dyslexia: Evidence from children learning to read German. *Journal of Experimental Child Psychology*, *61*, 80–90.
- Wimmer, H., & Goswami, U. (1994). The influence of orthographic consistency on reading development: Word recognition in English and German children. *Cognition*, *51*(1), 91–103.
- Wimmer, H., Mayringer, H., & Landerl, K. (2000). The double-deficit hypothesis and difficulties in learning to read a regular orthography. *Journal of Educational Psychology*, *92*(4), 668–680. doi:10.1037//0022-0663.92.4.668
- Ziegler, J., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faisca, L., . . . Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science*, *21*(4), 551–559. doi:10.1177/0956797610363406
- Ziegler, J., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29. doi:10.1037/0033-2909.131.1.3
- Ziegler, J., Jacobs, A., & Stone, G. (1996). Statistical analysis of the bidirectional inconsistency of spelling and sound in French. *Behaviour Research Methods, Instruments and Computers*, *28*(4), 504–515.
- Ziegler, J., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science*, *12*(5), 379–384. doi:10.1111/1467-9280.00370
- Ziegler, J., Perry, C., Ma-Wyatt, A., Ladner, D., & Schulte-Körne, G. (2003a). Developmental dyslexia in different languages: Language-specific or universal? *Journal of Experimental Child Psychology*, *86*, 169–193.
- Ziegler, J., Perry, C., & Coltheart, M. (2000). The DRC model of visual word recognition and reading aloud: An extension to German. *European Journal of Cognitive Psychology*, *12*(3), 413–430.
- Ziegler, J., Perry, C., & Coltheart, M. (2003b). Speed of lexical and nonlexical processing in French: The case of the regularity effect. *Psychological Bulletin and Review*, *10*(4), 947–953.
- Ziegler, J., Perry, C., & Zorzi, M. (2014). Modelling reading development through phonological decoding and self-teaching: Implications for dyslexia. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *369*(1634). doi:10.1098/Rstb.2012.0397
- Ziegler, J., Stone, G. O., & Jacobs, A. M. (1997). What is the pronunciation for -ough and the spelling for /u/? A database for computing feedforward and feedback consistency in English. *Behavior Research Methods, Instruments, & Computers*, *29*(4), 600–618. doi:10.3758/bf03210615