Aus der Klinik und Poliklinik für Kinder- und Jugendpsychiatrie,

Psychosomatik und Psychotherapie

Der Ludwig-Maximilians Universität München

On the role of graphemes, syllables and morphemes in reading: Evidence from different tasks and languages

Dissertation

zum Erwerb des Doktorgrades der Humanbiologie

an der Medizinischen Fakultät der

Ludwig-Maximilians-Universität München

vorgelegt von

Elisabetta De Simone

aus

Sant'Agnello, Italien

Jahr

2024

Mit Genehmigung der Medizinischen Fakultät der

Ludwig-Maximilians-Universität zu München

Erster Gutachter:	*Prof. Dr. med. Gerd Schulte-Körne*
Zweiter Gutachter:	*Prof. Dr. Karin Landerl*
Dritter Gutachter:	*Prof. Dr. Silvia Brem*
Mitbetreuung durch den promovierten Mitarbeiter:	*Dr. Elisabeth Beyersmann* *Dr. Xenia Schmalz*
Dekan:	*Prof. Dr. med. Thomas Gudermann*
Tag der mündlichen Prüfung:	*14.11.2024*

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Promotionsbüro
Medizinische Fakultät

Affidavit

De Simone, Elisabetta

Surname, first name

New Canterbury Road

Street

2203, Dulwich Hill NSW, Australia

Zip code, town, country

I hereby declare, that the submitted thesis entitled:

On the role of graphemes, syllables, and morphemes in reading: Evidence from different tasks and languages

is my own work. I have only used the sources indicated and have not made unauthorized use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the dissertation presented here has not been submitted in the same or similar form to any other institution for the purpose of obtaining an academic degree, except as provided in the respective Cotutelle Agreement.

München, 14.11.2024

place, date

Elisabetta De Simone

Signature doctoral candidate

Table of contents

List of abbreviations

DRC	*Dual route cascaded model of visual word recognition and reading aloud*
GPCs	*Grapheme-phoneme correspondences*
MIE	*Morpheme Interference Effect*
ODH	*Orthographic Depth Hypothesis*
PGST	*Psycholinguistic Grain Size Theory*

List of publications

Published and non-published studies part of this thesis:

De Simone, E., Moll, K., & Beyersmann. E. (2023). The Role of Orthographic Transparency and Morphological Complexity when Reading Complex Nonwords: Evidence from English and Italian. Manuscript submitted for publication.

De Simone, E., Moll, K., Feldmann, L., Schmalz, X., & Beyersmann, E. (2023). The role of syllables and morphemes in silent reading: An eye-tracking study. Quarterly Journal of Experimental Psychology, https://doi.org/10.1177/17470218231160638.

De Simone, E., Beyersmann, E., Mulatti, C., Mirault, J., & Schmalz, X. (2021). Order among chaos: Cross-linguistic differences and developmental trajectories in pseudoword reading aloud using pronunciation Entropy. PloS one, 16(5), https://doi.org/10.1371/journal.pone.0251629

Other publications:

Schmalz, X., Schulte-Körne, G., De Simone, E., & Moll, K. (2021). What Do Artificial Orthography Learning Tasks Actually Measure? Correlations Within and Across Tasks. Journal of cognition, 4(1), 7. https://doi.org/10.5334/joc.144

Görgen, R., De Simone, E., Schulte-Körne, G., & Moll, K. (2021). Predictors of reading and spelling skills in German: The role of morphological awareness. Journal of Research in Reading, 44(1), 210-227. https://doi.org/10.1111/1467-9817.12343

Schmalz, X., De Simone, E., & Mulatti, C. (2019). Rules and statistics in Italian. Preprint available at: https://doi.org/10.17605/osf.io/V3J2F

Lancia, L., De Simone, E. (2019). The endogenous nature of coordinative patterns underlying speech rhythm. Unpublished.

Conference talks and poster presentations:

De Simone, E., Moll, K., & Beyersmann, E. (2023, July). The Role of Orthographic Transparency and Morphological Complexity when Reading Complex Nonwords: Evidence from English and Italian. Paper presented at 30th Annual Conference of the Society for the Scientific Study of Reading, Port Douglas, Australia.

De Simone, E., Mulatti, C., Uccula, A., Schmalz, X. (2019, Sep). A cross-linguistic study on the lexical process acquisition across two languages with a different orthographic depth: Italian and French. Poster presented at the 21st Conference of the European Society for Cognitive Psychology, Tenerife, Spain.

Contribution to the publications

Contribution to Study 1: Order among chaos: Cross-linguistic differences and developmental trajectories in pseudoword reading aloud using pronunciation Entropy.

The first study (De Simone et al., 2021), has been published in PLoS One, a journal with a Journal Impact Factor (JIF) of 3.752 in 2021 and ranked 61.49(%) in “Multidisciplinary Sciences - Scie” category. This paper is a result of several contributors: Dr. Beyersmann, Dr. Mulatti, Dr. Mirault, Dr. Schmalz, and me as a first author. Regarding my own contributions, I designed the stimuli and conducted Experiment 3, transcribed the reading aloud responses, performed the linguistic and statistical analysis of all four experiments. I also curated the research article from draft to publication, which includes writing the manuscript, responding to peer-reviewers, and applying the necessary modifications.

Contribution to Study 2: The role of syllables and morphemes in silent reading: An eye-tracking study

This study (De Simone et al., 2023), has been published in the Quarterly Journal of Experimental Psychology, with a Journal Impact Factor (JIF) of 2.138 in 2021 (recent ranking not available through inCites - Journal Citation Reports) and ranked 31.88(%) in “Psychology – Scie” category and 29.12(%) in “Experimental Psychology – Ssci” category. This chapter is a result of several contributors: Dr. Kristina Moll, Dr. Lisa Feldmann, Dr. Schmalz, Dr. Beyersmann, and me, as a first author. As such, I have curated the materials for the purpose of the experiments, collected and analysed the data of both experiments and curated the research article from draft to publication, which includes writing the manuscript, responding to peer-reviewers, and applying the necessary modifications.

Contribution to Study 3: The role of Orthographic Transparency and Morphological complexity when reading complex nonwords: evidence from English and Italian.

This study has been submitted for publication at Scientific Studies of Reading, a journal with a Journal Impact Factor (JIF) of 4.200 in 2021 (recent ranking not available through inCites – Journal Citation Reports) and ranked 85.00(%) in “Education % Educational Research – Ssci” category and 81.15(%) in “Psychology, Educational – Ssci” category. The study is a result of several contributors: Dr. Kristina Moll, Dr. Beyersmann, and me, as a first author. As such, I have conceptualised the study, analysed the data collected online, and written the resulting manuscript in a publishable form. As this study is under review, but yet to be accepted for publication at the moment of writing, it has to be considered as an additional contribution to my thesis.

Summary

One key issue in reading is to determine how printed words are recognised. For decades researchers have tried to understand which sub-lexical units are more useful in reading. Specifically, evidence accumulated around *graphemes* (letters or letter clusters associated with a phoneme), *syllables* (a unit of pronunciation including one or more phonemes), and *morphemes* (the minimal unit carrying meaning). However, it is not clear how reliance on sublexical units changes according to specific languages. I investigate this topic by using a variety of experimental procedures, which reveal that that three main aspects contribute to cross-linguistic differences in sublexical processing: orthographic depth, morphological complexity, and syllabic complexity.

In the first study, published in PlosOne (De Simone et al., 2021), I explore how orthographic depth and the knowledge of letters to sounds mapping influence the reading of nonsense words by introducing a relatively new mean to calculate pronunciation variability. The study investigates four European languages (English, German, French, Italian) and examines different age groups (adults, children in grades 2, 3, and 4) as well as linguistic backgrounds (monolingual and bilingual children). Results indicated that pronunciation variability was greater in the language with the most opaque orthography, i.e., English.

In the second study, published in the Quarterly Journal of Experimental Psychology (De Simone, Moll, Feldmann, et al., 2023), I investigated the reliance on syllables and morphemes when reading words embedded in sentences. In this case, I measured participants' eye movements and restricted my focus on one language, German. The study's results suggested that syllables are the preferred units of analysis of native German speakers when silently reading for comprehension purposes.

In the third study, currently under review, I explored how morphological processing is affected by morphological complexity and orthographic depth. I did so by contrasting two languages that differ on both aspects: English, which has a scarce morphology but has an opaque

orthography, and Italian, which has a rich morphology but a transparent orthography. The findings of the study indicated that orthographic depth has a more profound impact on morphological processing than morphological complexity.

The findings of these three experimental chapters show that orthographic depth, and consequently, phonological processing, are the main cause of cross-linguistic differences in reading behaviour. Reliance on units larger than letters in reading aloud and silent reading is mostly driven by the specific orthography demands thus providing further evidence for the Orthographic Depth Hypothesis (Katz & Frost, 1992) and the Psycholinguistic Grain Size Theory (Ziegler & Goswami, 2005). Results are also discussed in terms of the Flexible-unit-size Hypothesis of Brown and Deavers (1999), and implications for theoretical and computational modelling are considered.

Zusammenfassung

Eine zentrale Frage beim Lesen besteht darin, festzustellen, wie gedruckte Wörter erkannt werden. Seit Jahrzehnten versuchen Forscher zu verstehen, welche sublexikalischen Einheiten beim Lesen nützlicher sind. Insbesondere sammelten sich Beweise rund um *Grapheme* (mit einem Phonem verbundene Buchstaben oder Buchstabencluster), *Silben* (eine Ausspracheeinheit, die ein oder mehrere Phoneme umfasst) und *Morpheme* (die minimale bedeutungstragende Einheit). Es ist jedoch nicht klar, wie sich die Abhängigkeit von sublexikalischen Einheiten je nach Sprache ändert. Ich untersuche dieses Thema mithilfe verschiedener experimenteller Verfahren, die zeigen, dass drei Hauptaspekte zu sprachübergreifenden Unterschieden in der sublexikalischen Verarbeitung beitragen: orthografische Tiefe, morphologische Komplexität und syllabische Komplexität.

Zunächst aber stellt sich die Frage, warum Leser größere Einheiten als Buchstaben zur Worterkennung nutzen? Ich behaupte, dass drei Faktoren, die eine Quelle sprachlicher Vielfalt sind, zur sublexikalischen Verarbeitung beitragen: orthographische Tiefe, morphologische Komplexität und Silbenkomplexität. Daher untersuche ich diese drei Aspekte in einer sprachvergleichenden Studie mit verschiedenen experimentellen Aufgaben (Pseudowortbenennung, Eye-Tracking und lexikalische Entscheidung).

In der ersten Studie, veröffentlicht in der Fachzeitschrift PlosOne (De Simone et al., 2021), untersuche ich, wie orthographische Tiefe und das Wissen über Graphem-Phonem-Korrespondenzen (GPCs) die Benennung von Pseudowörtern beeinflussen. Dazu bediene ich mich der Methode der Entropie, um die Aussprachevariabilität zwischen Teilnehmern zu berechnen, die dieselbe Sprache sprechen. Die Studie untersucht vier Sprachen (Englisch, Deutsch, Französisch, Italienisch) und unterschiedliche Altersgruppen (Erwachsene, Kinder der 2., 3. und 4. Klasse) sowie sprachliche Hintergründe (einsprachige und zweisprachige Kinder).

Die Ergebnisse zeigten, dass die Aussprachevariabilität (und die Entropiewerte) in der Sprache mit der tiefsten Orthografie, d. h. Englisch, am Größten waren.

In der zweiten Studie, veröffentlicht im Quarterly Journal of Experimental Psychology (De Simone et al., 2023), untersuche ich die Verarbeitung von Silben und Morphemen beim Lesen mehrsilbiger, multimorphemischer Wörter, die in Sätzen eingebettet sind. Hier habe ich die Augenbewegungen der Teilnehmer gemessen und meinen Fokus auf eine Sprache, Deutsch, beschränkt, da diese sowohl syllabisch als auch morphologisch komplex ist. Ich war daran interessiert, zu erforschen, wie phonologische und morphologische Verarbeitung die Augenbewegungen beim Lesen modulieren. Die Ergebnisse der Studie deuten darauf hin, dass Silben die bevorzugten Analyseeinheiten beim sinnerfassenden Lesen von deutschen Muttersprachlern sind.

In der dritten Studie, das zurzeit unter Begutachtung ist, untersuche ich, wie die morphologische Verarbeitung durch morphologische Komplexität und orthographische Tiefe beeinflusst wird. Dazu vergleiche ich zwei Sprachen, die sich in beiden Aspekten unterscheiden: Englisch, das eine limitierte Morphologie, aber eine intransparente Orthographie aufweist, und Italienisch, das eine umfangreiche Morphologie, aber eine transparente Orthographie aufweist. Die Ergebnisse der Studie zeigten, dass die orthografische Tiefe einen größeren Einfluss auf die morphologische Verarbeitung hat als die morphologische Komplexität.

Die Ergebnisse dieser drei experimentellen Kapitel zeigen, dass die orthographische Tiefe und folglich die phonologische Verarbeitung die Hauptursache für sprachübergreifende Unterschiede im Leseverhalten ist. Wie ich in der allgemeinen Diskussion darlegen werde, ist die Abhängigkeit von Einheiten, die größer als Buchstaben sind, beim lauten und leisen Lesen hauptsächlich auf die spezifischen orthografischen Anforderungen zurückzuführen. Die Ergebnisse stützen die Orthographic Depth Hypothesis (Katz & Frost, 1992) und die Psycholinguistic Grain Size Theory (Ziegler & Goswami, 2005) und werden hinsichtlich der

Flexible-Unit-Size-Hypothese von Brown and Deavers (1999) eingeordnet. Implikationen für theoretische Modelle und Computermodellierungen werden diskutiert.

Introduction

In the last decades, printed word recognition has received a great deal of attention by reading researchers. One key issue in the domain is to determine how printed words are recognised, and if readers rely on certain sublexical units (letters' groups smaller than words that have psychological saliency and function, such as syllables). These units are thought to convey orthographic, phonological or morphological information (Taft & Forster, 1975) – and therefore, might be helpful in mapping the written word form onto its meaning and pronunciation.

Sublexical units, such as graphemes, syllables and morphemes, were thus called *functional units of word analysis* (or more briefly, *reading units*), because of their supposed facilitation role in retrieving meaning and sound from written, unfamiliar words. After a number of studies in several alphabetic orthographies started reporting findings that these units were used in reading (Bowey, 1990; Brand et al., 2007; Hasenäcker et al., 2017; Healy, 1976; Prinzmetal et al., 1986; Rey et al., 2000), a debate started around which units are most important in visual word processing, and how reliance on such units might differ depending on languages' specific orthographies. These questions are of fundamental importance, as they inform about the language-specific factors that contribute to cross-linguistic differences in visual word recognition. While identifying language-specific reading mechanisms has its own merits, cross-linguistic research is pivotal in pinpointing both, language-specific factors as well as underlying universal reading behaviours (Bates et al., 2001). Therefore, further research comparing two or more languages is needed to advance our understanding of psycholinguistics universals and specific factors, especially considering the hegemony of Anglocentric findings in the field (Leminen et al., 2019).

The present thesis follows this line of research: its primary focus will be to investigate cognitive mechanisms involved in single written word recognition, either read in isolation (Study 1 and 3) or when embedded in sentences (Study 2), by examining sublexical processing differences across four languages: English, French, German and Italian.

In this regard I will argue that cross-linguistic differences in reading arise from three factors: (1) Orthographic Depth (Frost et al., 1987; Katz & Frost, 1992; Seymour et al., 2003; Ziegler & Goswami, 2005), (2) Morphological Complexity (Beyersmann et al., 2020; Casalis et al., 2015; Perfetti & Harris, 2013), and (3) Syllabic Complexity (Borleffs et al., 2017; Seymour et al., 2003).

Orthographic Depth

Alphabetical languages vary in their orthographic depth, which in turn is determined by language-specific phonological or morphological factors (Frost, 2005). On the surface, orthographic depth is determined by the relation between graphemes (singular letters or letter sequences, such as <ph> in the word "phone") and phonemes (language sounds: <ph> read as /f/) (Frost et al., 1987): the closer a given alphabetic orthography is to a one-to-one correspondence between graphemes and phonemes, the more "shallow" or "transparent" it is considered to be.

For example, graphemes are mostly associated with one phoneme in German, with few exceptions: <a> is always pronounced /a/. However, in a language like English, this correspondence is more intricate: the same grapheme can be read in several ways. For example, the grapheme <a> in English can be associated with the phoneme /æ/ (as in "banner"), /ɑː/ (as in "cart"), /eɪ/ as in (as in "status"), /ɔː/ (as in "award"), /ə/ (as in "woman"). Orthographies with this feature are said to be "opaque" or "deep". English is placed at the extreme end of the orthographic depth continuum, where alphabetical orthographies are aligned next to each other (Frost, 2005; Schmalz et al., 2015). For example, while Italian and German have mostly transparent orthographies, Welsh and Finnish grapheme-to-phoneme correspondences are even more consistent (Perfetti & Helder, 2022), and thus, are placed at the extreme transparent pole of this continuum. But note, that even the most transparent orthographies might have some opaque elements.

Orthographic depth is determined by how readily orthography changes when words' pronunciations do. Ideally, when pronunciations change, spelling should adapt to these new changes. However, orthography might be resistant to these changes when it tends to preserve morphological information (Kemp & Treiman, 2022), a phenomenon called morpheme consistency principle (Kargl & Landerl, 2018). Spellings that do not reflect phonology are useful to indicate relations between words: for example, the letter cluster <gn> in “sign” and “signature” is read respectively /n/ and /gn/, however, as the stem of the two words is the same, the orthography tends to preserve it. The relative tendency of a given orthography towards conveying synchronous pronunciations, or preserving morphological relations, determines its depth. English, for example, adheres to the morpheme consistency principle.

A Multi-dimensional Approach to Orthographic Depth

Historically, the orthographic depth continuum has been regarded as a single scale, but there are constructs that influence the orthographic depth of a language, such as incompleteness, complexity, and unpredictability (Schmalz et al., 2016; Schmalz et al., 2015). Incompleteness refers to the amount of phonological information not reported by the orthography: for example, in German, the stress placement is not indicated in writing. Complexity, refers to the set of rules needed, other than simple letter-to-sound rules, to correctly associate a phoneme to a grapheme. It can arise from multi-letter graphemes (in French, <ou> read as /u/ in “souvenir”, for example) or context-sensitive correspondences (in German, <d> is read /t/ at the end of the word, but /d/ in other positions), however, once a reader learns these complex correspondences, their pronunciation is entirely predictable. Unpredictability, instead, refers to the degree to which the knowledge of GPC rules can be used to identify the correct pronunciation of novel words. For example, while readers of English might be aware that <ch> is read /tʃ/ in front of some vowels (as in “chips”) and /k/ in front of consonants (as in “chrome”), they might be ill-prepared when they encounter the word “yacht” for the first time, where the <ch> grapheme is silent.

In some languages such as English, complexity and unpredictability might be difficult to disentangle, since these two constructs tend to co-occur. However, in other orthographies they are clearly separated. French represents a proto-typical case: while its orthography is highly complex, with many multi-letter and context-sensitive graphemes, a correct application of GPCs predicts correct word pronunciation, since the orthography itself is quite predictable. These concepts will be discussed in greater details in Chapter 2, where we identified pseudoword pronunciation variability, as a source of cross-linguistic differences.

The Orthographic Depth Hypothesis

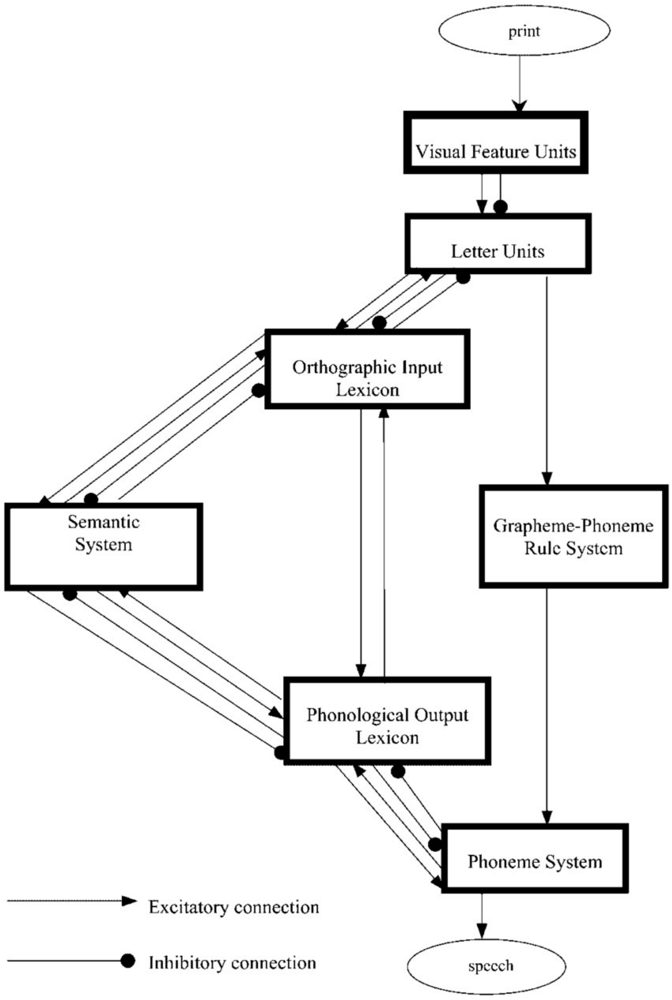
Orthographic Depth is believed to be responsible for cross-linguistic differences in written word recognition. According to the Orthographic Depth Hypothesis (Frost et al., 1987; Katz & Frost, 1992) readers of transparent orthographies tend to rely on their GPCs, which provide a path to retrieve the correct pronunciation through a simple process of phonological computation. For this reason, in a transparent orthography, like German and Italian, the phonology of words is already activated at the pre-lexical level (i.e., it is available before accessing the lexicon, Katz & Frost, 1992). Conversely, reading in opaque orthographies is primarily supported by the use of larger reading units, because the use of GPCs does not consistently lead to the correct pronunciation (Frost, 2005; Katz & Frost, 1992). In this case, the phonology of the target word is retrieved "*by referring to their morphology via the printed word's visual-orthographic structure*" (Katz & Frost, 1992, p. 71) or by using other sublexical units such as syllables or syllabic bodies (Marinus & De Jong, 2008). For example, readers might rely on the known pronunciation of the letter sequence <al> when encountering words with that ending, such as "practical" or "magical". Overall, the specific predictions from the ODH focus on how a printed word's phonology is produced during the reading process, but they also depend on the specific theory of skilled reading that is adopted, such as the dual-route model.

Predictions from the Dual Route Cascaded Model and the Orthographic Depth Hypothesis

One reading aloud model that has been extensively investigated to explore the nature of GPC mappings in reading is the Dual-Route Cascaded Model (Coltheart et al., 2001). The computational model focuses on the processes of skilled reading. The DRC architecture (see Figure 1) includes a non-lexical route, which applies GPC rules, and the lexical route, which involves whole-word recognition, allowing the reader to retrieve known words' pronunciation from memory through a direct correspondence between the orthographic lexicon (which contains the knowledge about the visual forms of the words, i.e., their spelling) and the phonological lexicon (holding the knowledge about words' pronunciations). Through the non-lexical route, readers can read aloud regular words and unfamiliar words. Through the lexical route, they can read regular words and irregular words that violate GPC rules, such as "blood" (which is read /blʌd/ and not /blud/ as the GPCs would predict, Coltheart, 2006, 2014).

Figure 1

Dual-Route Cascaded Model of Visual Word Recognition and Reading Aloud



Note. From "DRC: A Dual Route Cascaded Model of Visual Word Recognition and Reading Aloud", by Coltheart et al. (2001), Psychological Review, 108, p. 214. Copyright 2001 by the American Psychological Association

After readers have recognised letters, printed stimuli are then processed parallelly through the two routes. In the lexical pathway, the lexical entries are activated in the Orthographic Input Lexicon which is connected to the Phonological Output Lexicon. This latter contains the phonological codes of the known words and is connected to the Phoneme System, which contains the word’s phonemes. The non-lexical procedure applies the GPC rules to convert graphemes to phonemes serially (first letter in the string, then the first two together, and so forth).

The two routes will attempt to process any given stimuli, regardless of their nature, but with a different degree of efficiency: while the lexical route cannot decode correctly nonwords, it will still influence the reading process. For example, a nonword like SARE will produce some activation in the orthographic lexicon entries for similar words, such as CARE, whereas irregular words will be regularised in the nonlexical procedure following GPC rules, resulting in an incorrect pronunciation (such as /ˈbluːd/ for "blood", instead of the correct /ˈblʌd/(Coltheart, 2005, pp. 12–13).

Within the framework of the DRC model, the Orthographic Depth Hypothesis posits that readers of transparent orthographies can recover most words' pronunciations by simply relying on the non-lexical route of reading, thanks to the consistent grapheme-to-phoneme correspondences of their orthography. In contrast, readers of opaque orthographies generate words' phonological structure more reliably by counting on the phonological output lexicon following the activation of the visual lexicon, through the lexical route (Frost, 2005).

In an earlier version of the ODH (referred as the *strong version*), it was hypothesised that in transparent orthographies the phonological representations of words are derived exclusively through the analytic process carried out by the non-lexical route of reading, meaning that phonological representations, stored in memory, were not activated. Similarly, this version suggested that readers of opaque orthographies only employ the lexical route (Katz & Frost, 1992; Schmalz et al., 2015). However, this earlier hypothesis was later replaced with a more flexible view (the *weak version*), suggesting that lexical processing is required even in the most transparent orthographies, to retrieve syllable stress for example (Frost, 2005, p. 282). Let us consider the case of Italian. In this language stress patterns are scarcely predictable, and stress placement might even semantically distinguish two homographs (see "principi", which means "princes" if read as /ˈprintʃipi/ or "principles" if read as /prinˈtʃipi/). However, the orthography does not convey cues about where to place the stress, because of its *incompleteness* (exception made for oxytone words, where a diacritic is used to indicate that the stress should fall on the last syllable, as in “papà”). As such, to retrieve the correct pronunciation of a word, Italians need to access their phonological representations.

Although the non-lexical and lexical routes are predicted to work simultaneously and are both involved in the process of retrieving a word’s pronunciation, orthographic depth determines the relative pace of the two routes to retrieve the correct pronunciation. In opaque orthographies, for example, the sublexical route is slowed down by the complexity and unpredictability of the orthography, which makes the assembly of phonology a complicated process; thus, the product of the lexical route comes to activation faster (Katz & Frost, 1992).

Over the past two decades, the Orthographic Depth Hypothesis (ODH) has been supported by a range of behavioural studies (e.g., Ellis & Hooper, 2001) as well as brain-imaging studies (Frost, 2005). For example, Ellis and Hooper (2001) found that readers of Welsh, a language with a transparent orthography, showed larger length effects than readers of English, which indicates a bigger reliance on the sublexical route. Similarly, in Paulesu et al.’s (2000) study, Italian and English readers were administered word and nonword reading tasks, while being monitored using positron emission tomography (PET). The data indicate that Italian readers showed major activation in the left superior temporal regions, associated with phonemic processing, while the English readers showed higher activation in the left posterior inferior temporal and anterior inferior frontal gyri, associated with whole-word retrieval (see also Chyl et al., 2021 for similar results in fMRI).

The Psycholinguistic Grain Size Theory

Building on the Orthographic Depth Hypothesis, the Psycholinguistic Grain Size Theory (PSGT) of Ziegler and Goswami (2005), posits that the kinds of internal representations (the size of psycholinguistic units) that will develop in a child exposed to a consistent orthography will differ from those developing in a child exposed to an inconsistent (Goswami, 2010b). Thus, it is the very nature of the phonological process that changes across languages. Preliminary evidence for this theory came from Ziegler et al. (2001), who tested the hypothesis that smaller units (such as graphemes and phonemes) play a more dominant role in written word recognition in consistent orthographies, compared to larger units (such as bodies and rhymes), utilised instead by readers of inconsistent orthographies. Naming performances to identical words and nonwords (*zoo-Zoo, sand-Sand*, etc) revealed that native readers of the consistent German orthography were affected by the number of letters, whereas native readers of the inconsistent English orthography were affected more by words' bodies orthographic neighbourhood (Body-N). The authors interpreted these findings by suggesting that identical items were processed differently according to the orthography, and that orthographic consistency determined the preferred grain size of functional units.

Ziegler and Goswami (2005) argue that children learning to read in opaque orthographies rely on larger units, since the relationship between these and the corresponding phonemes is more reliable than GPCs (Treiman et al., 1995). Furthermore, having to use units of variable size is more demanding than being able to rely on phonological computation, because children need to learn more orthographic patterns and their mapping with phonology. Together, these two aspects slow the acquisition of reading fluency in children learning to read opaque orthographies (Ziegler & Goswami, 2005). It is worth noticing that in developing the Psycholinguistic Grain Size Theory the authors emphasised that their theory is not compatible with a dual-route framework, as they view the reliance on grain-sized units to go from a single continuum from fine-grained size units (such as letters) to larger ones (such as morphemes) until the largest one (that is, whole words), instead of a labour division between lexical and sublexical routes.

While empirical studies have found abundant evidence supporting the Psycholinguistic Grain Size Theory (Egan et al., 2019; Gottardo et al., 2016; Mousikou et al., 2020; Rau et al., 2015), challenging findings have also been gathered. In a series of new experiments and a re-analysis of Ziegler et al.'s (2001) data, Schmalz et al. (2014) found no reliable evidence for cross-linguistics differences in preferred grain size units, with weak body-N effects across tasks, languages, and conditions. Moreover, the use of larger units has also been attested in readers of transparent orthographies (Barca et al., 2007; Burani et al., 2002; Paizi et al., 2013).

It has also been suggested that units of analysis utilised in nonword reading mostly depend on the task at hand, therefore, readers might be more strategic and flexible in the use of small units or large units than the Psycholinguistic Grain Size Theory predicts. According to the flexible-unit-size hypothesis (Brown & Deavers, 1999) English speakers read nonwords through both small units (GPCs) and bigger units (body-level or morpheme correspondences). Across four experiments, the authors have found that both children and adults were adaptive in the usage of spelling-to-sound correspondences, and the strategy depends on the specific task participants are asked to perform and to what items they are responding to. For example, if the items presented consistent graphemes or if the nonwords were presented in isolation, then participants read nonword by using GPCs. Comparatively, if the items presented consistent bodies or clue-words are given prior to the nonwords, participants were biased into adopting an analogy strategy with bigger-size units. Overall, this hypothesis contradicts the idea that readers of opaque orthographies will automatically rely on larger units (Ziegler & Goswami, 2005).

Cross-Linguistic Differences in Orthographic Depth

Both the Orthographic Depth Hypothesis and the Psycholinguistic Grain Size Theory see orthographic depth as being one of the major factors in how easily children can learn to read, and which reading strategies are developed. In a large cross-linguistic study which included 13 European languages, Seymour et al. (2003) investigated how orthographic depth (and syllabic complexity) influenced reading acquisition by testing Grade 1 and 2 children in three reading tasks which tested letter-sound knowledge, familiar word reading and nonwords decoding. Reaction times and accuracy scores were calculated for each language. While their results did not show cross-linguistic differences in letter-sound knowledge in reading speed or accuracy, their data suggest that familiar word reading fluency is achieved much slower in the more opaque orthographies (such as English and Danish), than in the more transparent ones (Seymour et al., 2003, p. 152), and that nonword decoding performances were worse in terms of reading speed and accuracy as well (Seymour et al., 2003, p. 159). Seymour et al.'s (2003) study became a seminal in the field, and their findings have been replicated in several smaller-scale studies that compared reading skills of opaque and transparent orthographies readers (Frith et al., 1998; Goswami, 2010a; Landerl et al., 1997).

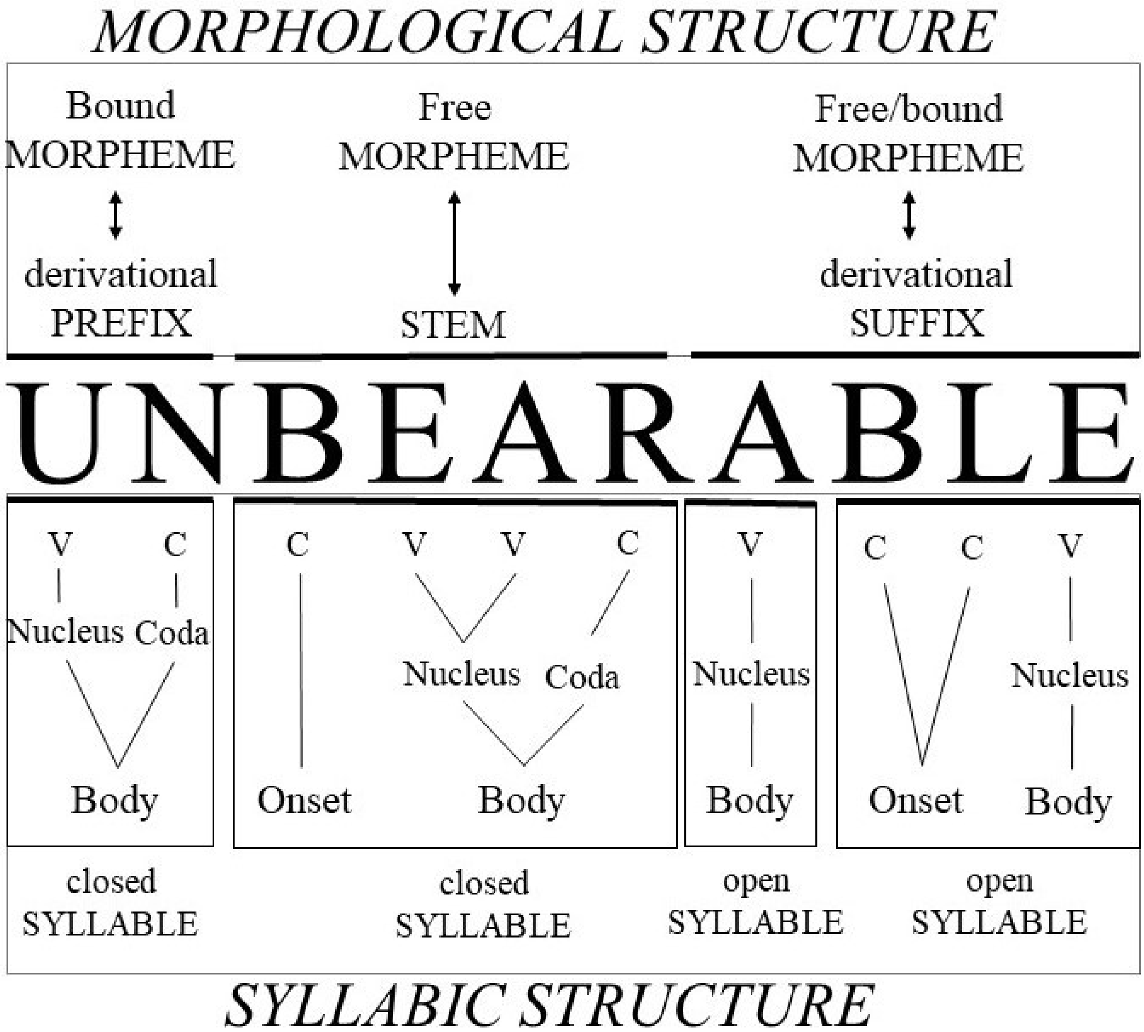
Nevertheless, cross-linguistic differences in skilled reading paint a different portrait: starting from late primary school reading accuracy and fluency differences between orthographies tend to flatten, but qualitative differences resulting from developing different reading strategies remain evident. For example, English-speaking readers seem to be more sensitive to body-rhyme and frequency effects (Marinelli et al., 2016), while German readers are more sensitive to length effects (Ziegler et al., 2001), a finding that has been replicated in other regular orthographies, such as Welsh (Ellis & Hooper, 2001) and Italian (Barca et al., 2002; Bates, Burani, et al., 2001). When reading non-words, readers of regular orthographies apply more thorough letter-by-letter decoding than English-readers (Landerl, 2000). In Chapter 2 we expand on this more, by proposing a new method to measure pseudoword pronunciation variability across four languages, English, Italian, French, and German.

Morphological Complexity

A further key issue in the field concerns how morphologically complex words are identified in reading, and whether or not their recognition varies across languages. Morphologically complex words consist of two or more "morphemes", usually defined as the smallest units that carry semantic or syntactic information (Bloomfield, 1933; Rastle, 2022). Morphemes can be categorised into *free* and *bound* morphemes. Free morphemes can be further divided in lexical and grammatical morphemes, with the first having semantic meaning (see *Apple*) and the latter being grammatical function words (such as *the*). Contrarily to free morphemes, bound morphemes are not independent and must be attached to a stem. These morphemes are also called affixes and can be divided according to their function (derivational or inflectional) or their position (prefixes, infixes, or suffixes). Derivational affixes are employed to create new words from existing ones (e.g., *immature* from *mature*), thus also sometimes changing the words' category (see how the derivational suffix "er" in "play_er" changes the word from the verb "*to play*" to the noun "*player*"). Inflectional affixes indicate grammatical relations between words (see the conjugation to the third person in "*She speaks*" at the end of the verb). Prefixes are found in front of the stem (*unbearable*), infixes are found within the stem (see the plural formation of *foot*: *feet*), and suffixes are found at the end of the stem (*unbearable* - see Figure 2 for an exemplified division of a bisyllabic bimorphemic word). Another mean to form new words is through *compounding*, where two free morphemes combine (*earphones*). In Study 2, as the experimental design involve sentence reading, I will examine both inflectional suffixes and derivational prefixes and suffixes, whereas in Study 3, I will examine the processing of derivational suffixes in a single word reading task.

Figure 2

An Example for Syllabic and Morphemic Structure within a Derived Suffix Word



Morphological Complexity across Languages

The complexity of the internal structure of words varies significantly among languages For example, languages can be divided in prefix languages (like Thai or Swahili) where stems are preceded by derivational prefixes and then inflectional prefixes, or by the more common suffix languages (like Finnish), where stems are followed by derivational and inflectional suffixes (Pirkola, 2001). Historically, languages were classified according to the transparency of morphological boundaries between the stem and the affixes (Schlegel, 1808). This index of fusion identified three major types: *isolating, agglutinative* and *fusional*. Considering that languages rarely belong to only one type (Brown, 2010; Greenberg, 1954; Pirkola, 2001), *isolating* languages' words are mostly simple with no or few signs or morphological structure (like in Vietnamese). In *agglutinative* languages, affixes are always appended to a base form (Frost & Grainger, 2000), and the boundaries between stems and affixes are clear-cut, thus making the word easily decomposable (see the Turkish word *Italyanların*: Italyan [Italian], lar [plural], ın [of, possessive]). On the contrary, *fusional* languages' morphological boundaries tend to be less definable, with the phonemes at the boundaries merging (see the English word *joker*: does <e> belong to the stem *joke* or the suffix *er*?) or morphs conveying more than one semantical feature (*bought* means "to buy" in a past tense).

A further parameter to classify languages in morphological types was later introduced by Sapir (1921), who measured the amount of affixation in a language, thus dividing languages in *analytical*, *synthetic* and *polysynthetic.* Words in analytical languages, like Chinese, typically have no or few bound morphemes, but words in synthetic languages, like Italian, are constructed from a number of morphemes. A language is deemed to be polysynthetic (like Finnish) if its total number of morphemes is extremely high. Much like orthographic depth, the degree of synthesis can be illustrated by means of a continuum, at whose extremes we find isolation and synthesis. Languages might be placed at any point of this continuum (Pirkola, 2001):

Isolating ◄────────────────► Synthetic

A further element to take into consideration when discussing the morphological complexity and variety of languages is how productive (and transparent) its compounding system is, or the number and type of morphosyntactic features conveyed by the language: for example, grammatical gender (feminine, neutral, masculine, or none), number (singular, dual, plural, or none) or syntactic case (nominative, accusative, dative […], or use of word order and prepositions) (Pirkola, 2001; Stump, 2001).

Morphological Processing in Visual Word Recognition

This wide variety of morphological structures lends credence to the hypothesis that, while sensitivity to morphological structure is found across languages and writing systems (Stevens &

Plaut, 2022), there might be language-specific effects on morphological processing, especially since reading morphologically complex words "*reflects a learned sensitivity to the systematic relationships among the surface forms of words and their meanings*" (Verhoeven & Perfetti, 2011, p. 464). For example, Havas et al. (2015) examined the process of extracting morphological information through an artificial morphological learning paradigm, in order to explore cross-language differences in morphological acquisition in adults. The authors found that while both Finnish and Spanish participants were able to learn and apply new morphological patterns, they also found that the morphological complexity of participants' native tongues provided an advantage in morphological learning, as Finnish participants were better at identifying embedded suffixes in the artificial language.

How morphologically complex words are recognised is a topic still under debate, with a vast literature showcasing a variety of results. To investigate the issue researchers have employed a number of different techniques, such as masked priming experiments (Beauvillain, 1994; Grainger et al., 1991; Hasenäcker et al., 2016), masked transposed-letter priming (Beyersmann et al., 2013), visual disruption paradigms paired with lexical decision tasks (Hasenäcker & Schroeder, 2017), single lexical decision tasks (Hasenäcker et al., 2017), illusory conjunction paradigms (Prinzmetal et al., 1986), and letter search tasks (Antzaka et al., 2019; Beyersmann, Casalis, et al., 2015; Hasenäcker et al., 2021). While there is some consensus regarding the facilitation role of morphemes during visual word recognition, the extent to which morphological processing varies due to cross-linguistic differences is still an underdeveloped topic.

The important role of morphological processing has been evidenced in a wide variety of languages (see Amenta & Crepaldi, 2012 for a review on morphological effects). However, these studies primarily investigate morphological processing within-languages, with few direct cross-linguistic investigations. For example, a series of lexical decision tasks with English and Finnish speakers were conducted by Vannest et al. (2002). These two languages differ in their morphological complexity, with English being less complex than Finnish (Bölücü & Can, 2019;

Borleffs et al., 2017). They also fall on opposite ends of the orthographic depth continuum, with English being at its opaque end and Finnish being at its transparent end. The researchers discovered that participants who spoke English were more sensitive to stem frequencies (i.e., lexical decisions were faster to words with higher-frequency stems) than those who spoke Finnish. This study was one of the first to suggest that morphological processing was modulated by cross-linguistic differences in orthographic depth rather than morphological complexity, a finding that was later replicated in the reading-aloud study of Mousikou et al. (2020). Chapter 4 will address this point in greater depth.

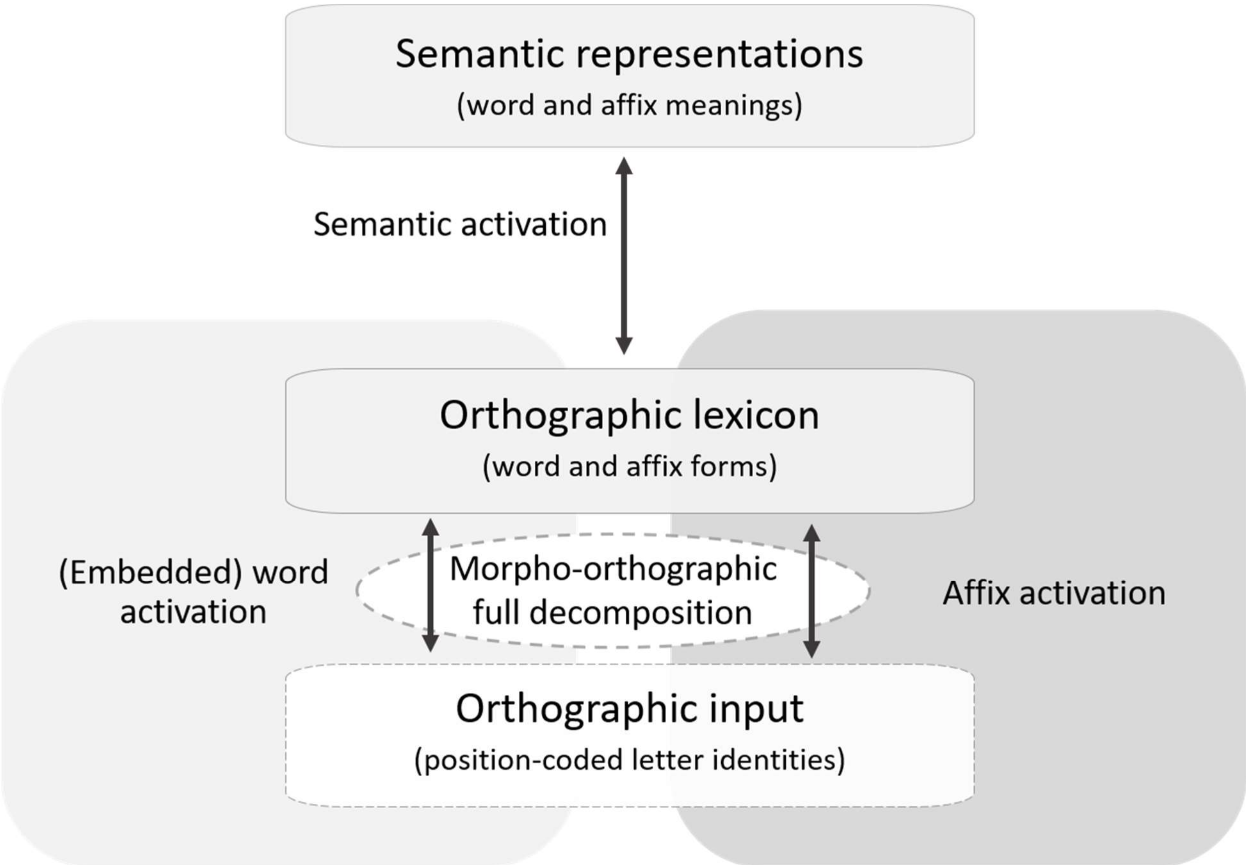
Morphological Processing Theories

The question whether morphologically complex known words are recognised at first glance or whether they are decomposed in morphological components first (or after) has received much attention over the past three decades (Giraudo & Grainger, 2003; Marslen-Wilson et al., 1994; Pollatsek et al., 2000; Rastle & Davis, 2008; Schreuder & Baayen, 1995; Marcus Taft, 2003). The form-then-meaning account posits that words are first decomposed in morpho-orthographic representations, and then morpho-semantic analysis occurs (Rastle & Davis, 2008; Marcus Taft, 2003). In contrast, supralexical decomposition accounts postulate that complex words recognition only happens through morpho-semantic decomposition (Giraudo & Grainger, 2003; Marslen-Wilson et al., 1994). A third theoretical framework sees complex words processed via two routes: a *holistic* route, which helps identify the word without segmenting it, and a *morphological decomposition* route (Pollatsek et al., 2000; Schreuder & Baayen, 1995). This has been suggested for compound word processing as well, which may benefit from the parallel, simultaneous work of both routes (Frost & Grainger, 2000; Marcus Taft, 1994). For example, in the Morphological Race Model (Frauenfelder & Schreuder, 1992) or the Parallel Dual Route Model (Schreuder & Baayen, 1995) complex word recognition is achieved via parallel processing via the holistic, direct route, and the decompositional, parsing route. This assumption seems to be, nowadays, the most supported one, with word length being a significant predictor (holistic route for short words, decomposition for longer words) of the relative dominance of the two routes (Hyönä, 2015). Empirical evidence coming from eye-tracking studies seem to corroborate this view: Niswander-Klement and Pollatsek (2006), for example, found dominant word frequency effects on gaze durations for short, prefixed words, but more dominant stem frequency in longer words.

Based on the growing body of evidence from morphological processing that supports a decompositional approach, researchers have theorised different ways in how this decomposition takes place. One example is the Word and Affix model (Beyersmann & Grainger, 2023; Grainger & Beyersmann, 2017), which builds on the assumption that all known words and affixes are stored and represented in the mental lexicon. The model implemented a morpho-orthographic full decomposition process operating through two routes: the embedded word activation route, and the affix activation route. The first mechanism performs a match between the input letter string and the orthographic lexicon, thus activating whole words, but also embedded words, as the match does not need to be exact (f-a-r-m-e-r would activate both farmer and farm alike). The second mechanism performs instead an exact orthographic match and activates affixes that are edge-aligned. In the study presented in the appendix A use this model to discuss my findings.

Figure 3

The Word and Affix model



Note. From "The role of embedded words and morphemes in reading", by Beyersmann and Grainger (2023), in D. Crepaldi (Eds.), *Morphology in the Mind and Brain* (p. 28). Copyright 2023 by Routledge.

Syllabic complexity

In alphabetic scripts, syllables can be defined as the phonological building blocks of words. Syllables are not only sublexical units, but they have a hierarchical structure of their own (Levitt et al., 1991; Treiman, 1983, 1986). Phonologically speaking, they might consist of an onset (when present, equal to initial consonant or consonant cluster) and a body, further divided in nucleus and coda (when present, see Figure 2).

Syllabic Complexity across Languages

Aside from few invariants, such as the ubiquity of the Consonant-Vowel syllable type (Zec, 2007), languages vary considerably in syllabic structure. Reasons of this variation can be found in the phoneme inventory size (Easterday, 2019; Maddieson, 2006, 2013) and population size (Fenk-Oczlon & Pilz, 2021) which have been found to be positively correlated with syllabic complexity. Seymour et al. (2003) positioned European languages on a simple to complex syllabic structure continuum. Measuring syllabic complexity by the average number of constituents of a syllable (Coupé et al., 2014), Seymour et al. judged to be simple the languages having mostly open CV syllables with few consonant clusters (typically the case of romance languages, such as Italian), and complex those languages whose syllables are predominantly closed, as in CVC syllables. This is typically the case of Germanic languages, which have clusters in both the onset and coda position (such as German or English). For example, De Cara and Goswami (2002) found 88.8% closed syllables in the English CELEX corpus (Baayen et al., 1993), with structures spanning from CVC to CCVCCC. Comparatively, in Italian the open syllable CV is the most frequent syllable type, accounting for 56% of the occurrences in written corpora (Burani et al., 2014).

It has been suggested that syllabic complexity influences reading acquisition. For example, Marinelli et al. (2016) suggests that the lower syllabic complexity of Italian facilitates segmentation in phonemes and syllables, and thus, accelerate the acquisition of grapheme-phoneme correspondences. This interpretation is consistent with Seymour et al. (2003), who found that nonword reading was more accurate in syllabically simple languages than in syllabically complex languages (Exp. 3), and that reading nonwords was slower in these latter (Seymour et al., 2003, p. 160). The authors interpreted these results as a consequence of the increased difficulty in acquiring GPCs due to the fact that they are embedded in consonant clusters.

Syllabic complexity has also been found to affect reading rate in sentences with real words. A direct impact of syllabic complexity has not only been found on the production of speech, with complex syllables taking longer to articulate (Naveh-Benjamin & Ayres, 1986; Pellegrino et al., 2011), but also on silent reading. In a cross-linguistic study of eight languages varying in syllabic complexity and script type, Coupé et al. (2014) found that both oral and silent reading were affected by syllabic complexity, regardless of script type, with reading rates being slower in the more syllabically complex languages.

Syllables as Reading Units

Research on visual word recognition carried out on different languages, such as French, Spanish and English, suggested that syllables could be relevant processing units, as they facilitate naming in masked priming paradigms (Carreiras et al., 1993; Ferrand et al., 1996; Ferrand et al., 1997). Slower lexical decisions have been evidenced when the first syllable of a stimuli is a high-frequent one (Alvarez et al., 2001), and faster lexical decisions when target words are primed by the word's initial syllable (Ashby & Martin, 2008). However, contradictory research using Illusory Conjunctions techniques (Doignon & Zagar, 2005; Doignon-Camus et al., 2009) showed that low-frequency bigrams in monosyllabic elements produce the same syllabic effects shown in polysyllabic words (thus providing evidence for the bigram trough hypothesis of Seidenberg, 1987, which posits that syllabic effects are due to orthographic redundancy, rather than syllabification). Similarly, other research using masked priming paradigms failed to replicate syllable priming effects (Brand et al., 2003; Schiller, 2000). Somewhat problematic is also the fact that syllables do not have specific boundaries, at least in English (Alvarez et al., 2001), and thus, are not recognizable orthographically as a grapho-syllable (the orthographic counterpart of the phonological syllable - Chetail & Mathey, 2010).

The syllabic processing literature is highly controversial (Yap & Balota, 2009), but studies conducted with eye-tracking methodologies have provided further insights, in particular in sentence reading contexts. Ashby and Rayner (2004) examined the effect of syllable primes in

English silent reading. They found that readers' first fixation times on the target word were shorter when the preview was syllabically congruent (target: de-vice, preview: de_πxw) than when it was incongruent (dev_πx), even though the orthographic overlap between the preview and the target was higher in the incongruent condition. More recently, eye tracking was used by Hawelka et al. (2013) to determine whether the inhibitory effects of first syllable frequency discovered in lexical decision tasks are transferrable to natural reading. The authors observed inhibitory effects that led to a longer first fixation on multi-syllabic words beginning with high-frequent first syllables using the multi-syllabic items of the German Potsdam Sentence Corpus as target words (Kliegl et al., 2004). This result was interpreted as an indication of phonological processing occurring prelexically, with the authors concluding that syllabic representations do act as "access units" to the mental lexicon, and are activated during visual word recognition (see also Stenneken et al., 2007).

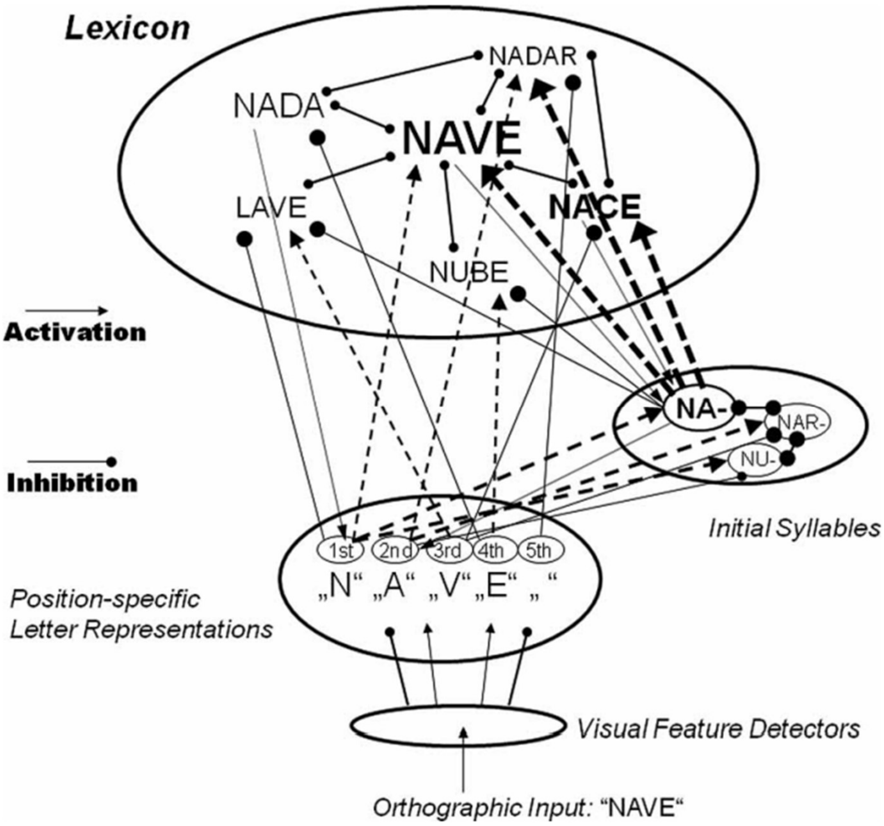
Syllabic Processing Theories

Historically, reading models only considered monosyllabic words (see the DRC of Coltheart et al., 2001) especially since early empirical studies have been conducted with thise kind of items (Yap & Balota, 2009). However, the idea that syllables might be placed at an intermediate stage between letter perception and whole-word recognition has led to recent attempts to accommodate the processing of polysyllabic words within reading models.

For instance, Conrad et al. (2010), adapted the Multiple Read-out Model (MROM) of Grainger & Jacobs (1996) with a separate route dealing with syllabic representations, together with a syllabary containing syllabification rules. The original model of Grainger and Jacobs (1996) could generate responses to lexical decision tasks through two processes: either global lexical activation reaches a threshold corresponding to a “fast guess”, or the activation of a single word unit reaches a threshold related to its identification. However, it was unable to account for syllable frequency effects in lexical decision tasks, which is the reason why in the MROM-s model (Figure 4) Conrad et al. (2010) designed an architecture where syllables activate words that contain them in the initial syllabic position. In the model, syllabic parsing is modulated both by the frequency of the letter cluster which composes the initial syllable and the syllabary. Any ambiguity is resolved by feedback from the word level. The MROM-S model accounts for the inhibitory effect of syllable frequency on lexical decision that has been seen in a number of languages (e.g., Conrad & Jacobs, 2004): that is, the processing of the target is impeded by lateral inhibition brought on by the coactivated syllabic neighbours, and the competition is worse when the syllable is frequent. Models of visual word recognition are discussed further in my second publication.

Figure 4

Spread activation in the MROM-S



Note. From "Simulating syllable frequency effects within an interactive activation framework" by Conrad et al. (2010), European Journal of Cognitive Psychology, 22(5), p. 872. Copyright 2010 by Psychology Press

Aims of the thesis

The above literature outlies a summary of evidence and theories focussing on the important role of orthographic depth, morphological and phonological processing in reading polysyllabic and polymorphemic words. However, these models all seem to concentrate on one particular reading unit (*graphemes*: Coltheart et al., 2001; *morphemes*: Beyersmann & Grainger, 2023; *syllables*: Conrad et al., 2010) rather than the interplay between different types of sublexical units. This leads to several outstanding questions. For example, syllabic and morphemic boundaries do not always coincide (see "eating", which can be divided in two syllables: ea + ting; and two morphemes: eat + ing; Alvarez et al., 2001): what is the parsing mechanism taking the lead in the reading of multi-morphemic multi-syllabic words? Do languages have a preferred unit of reading by which they access the lexicon, and if so, what (cross-)linguistic features dictate this preference? The overarching goal of this thesis was to address these questions by exploring the processing of sublexical units in visual word recognition; specifically, the reliance on graphemes, morphemes, and syllables. I will argue that cross-linguistic differences in sublexical processing arise from three major aspects of linguistic variation: orthographic depth, morphological complexity, and syllabic complexity. I will investigate the topic across a series of different methodologies and designs.

Studies summary

Study 1. Order among chaos: Cross-linguistic differences and developmental trajectories in pseudoword reading aloud using pronunciation Entropy.

In this study, a new measure, pseudoword pronunciation Entropy, was introduced to calculate the variability of pronunciations to monosyllabic and multisyllabic pseudowords, in four European languages (English, German, French, Italian), across ages (adults, and children in grade 2, 3, and 4), and linguistic background (in monolingual and bilingual children). Specifically, we investigated the reliance on graphemes, and how the knowledge of grapheme-phoneme correspondences affects pseudoword reading in a naming task.

To do so, we transcribed and analysed the audio recordings of each participant's individual reading aloud responses, and then calculated entropy for each pseudoword. Across four experiments, results consistently showed that pronunciation variability decreased with age, as knowledge of GPCs increases, and that in the language with the highest degree of orthographic unpredictability (English), entropy values were greater compared to the other languages in both the adult and children population, thus demonstrating a higher pronunciation variability to pseudowords. While our French and German-speaking participants also showed some degree of pseudoword pronunciation variability, we found that the entropy values associated to their response was significantly lower than the ones found in our English-speaking sample (Exp. 4).

Critically, we showed that these differences could only arise when items are truly representative of the language under investigation in terms of orthographic patterns and syllabic complexity, thereby providing support to the argument that researchers should be careful when employing cognate or stimuli that are too comparable when investigating cross-linguistic differences (Ellis et al., 2004; Marinelli et al., 2016). In particular, the fact that English speakers entropy values increased between Experiment 1 ($M = 0.39$) and Experiment 4 (pseudowords matched on base-word frequency, Set 1, $M = 1.03$; equally dissimilar pseudowords, Set 2, $M =$

2.27), points to the fact that in the first experiment their responses were heavily impacted by the fact that the monosyllabic items bodies were consistent, something that influenced the degree of agreement in pseudoword pronunciation among English-speaking participants. However, once deprived of this regularity cue coming from the consistent bodies (Treiman, 1986; Treiman et al., 1995; Ziegler & Goswami, 2006), pronunciation variability and hence, entropy values increased, as English-speaking participants had to rely on the often-unpredictable letter-to-sound correspondences. These results were in line with previous research which found entropy values to be influenced in English by spelling-to-sound consistency and orthographic neighborhood (Mousikou et al., 2017).

In sum, these findings seem to indicate that sublexical processing, in reading aloud, is impacted by cross-linguistic differences in orthographic depth. We come to this conclusion by observing that the speakers of the language with the most opaque orthography (English) were more reliant on large sub-units (in this case, bodies) to read pseudowords, largely because the correspondence between graphemes and phonemes is unpredictable, and thus, unreliable. These results are in line with the seminal study of Ziegler et al. (2001), where the authors found a stronger body-N effect in English as compared to German, indicating a bigger reliance on bodies as compared to graphemes. In fact, the length effect (that posits that longer words take more time to read) was more pronounced in German speaker. This study was pivotal for the development of the Psycholinguistic Grain Size Theory (Ziegler & Goswami, 2005), which predicts that readers of opaque orthographies rely on larger sublexical units than readers of transparent orthographies. As our results seem to be in line with both (Ziegler & Goswami, 2005; Ziegler et al., 2001), they provide further support for the Psycholinguistic Grain Size Theory.

Study 2. The role of syllables and morphemes in silent reading: An eye-tracking study

The focus of the second study was to investigate syllabic and morphological processing in a sample of German readers. The aim of this study was to investigate the relative reliance on syllables and morphemes when reading multi-syllabic and multi-morphemic words embedded in sentences. As German is a language which is both syllabically and morphologically complex, it represented the ideal candidate. To unveil which of these sublexical units are processed to a greater extent, we conducted two eye-tracking experiments, with two different manipulations that either disrupted or highlighted syllabic and morphological boundaries (with colours in Exp. 1 and hyphens in Exp. 2).

Eye-tracking data revealed that when disrupting syllable boundaries with hyphens (Exp. 2), participants showed significantly longer fixation times compared to when morphological boundaries are disrupted. This result lends credence to the hypothesis that German speakers rely more heavily on syllable-based than morpheme-based reading. Hence, the findings of Experiment 2 suggest that in silent reading native German speakers might fall into a syllable-based rhythm, an interpretation that is supported by previous studies revealing the impact that syllables have in silent reading (Alvarez et al., 2001; Conrad & Jacobs, 2004; Conrad et al., 2011; Hawelka et al., 2013; Hutzler et al., 2005). We also speculated that this reading behaviour could have been amplified by the German-specific literacy instructions guided by the Silbenmethode and other forms of syllable-based teaching strategies that have a long history in Germany (Velten, 2012). Several empirical studies have indicated that specific teaching methods have a long-lasting impact in reading development (Lyster, 2002; Lyster et al., 2016; Segers & Verhoeven, 2005), and therefore, likely persist in adulthood (see also Ziegler & Goswami, 2006).

Ultimately, this study provided some insights into the kind of sublexical analysis that skilled readers of orthographically transparent languages employ in a natural reading setting, where the purpose of the task is reading comprehension. Our findings were dissimilar to those of Hasenäcker and Schroeder (2017) whose visual disruptions (i.e., SPIN:AT) at the level of syllables and morphemes did not impair German skilled readers in a lexical decision task. The authors offered many explanations for why this was not the case. First, the stimuli, taken from the childLex corpus (Schroeder et al., 2015), might have been too simple for an adult audience.

Second, the visual disruption employed (:) might have been too subtle for a skilled reading system (p.748).

However, a third interpretation would be that for the task and items at hand skilled readers did not need to process sublexical units, which was not the case in our sentence reading study. Therefore, while our results do not align, this dissimilarity seems to be a good first indication for the flexible-unit-size hypothesis (Brown & Deavers, 1999) which posits that the choice of units of analysis change according to the task demands, as adult readers did not need to process sublexical information in a lexical decision task (Hasenäcker & Schroeder, 2017), but did in sentence reading (De Simone et al., 2023).

Interestingly, Hasenäcker and Schroeder (2017) did find evidence of switching between units in their child participants: in fact, they found syllables to be prominent in word reading, and morphemes in pseudoword reading. This suggests that beginner readers of transparent orthographies do strategise in the choice of reading units in a task that is still somewhat challenging at that stage of reading development (thus calling for a more pronounced need for sublexical processing than in skilled readers). That would be a first support of Brown and Deavers (1999) flexible-unit-size hypothesis in transparent languages. An interesting avenue for future research would be to gather empirical data in other transparent languages that are both syllabically and morphologically complex, such as Icelandic, which is a transparent, syllabically complex (Seymour et al., 2003), and also morphologically complex language with a rich inflectional and compounding system (Bjarnadóttir et al., 2019).

Study 3 (Appendix). The role of Orthographic Transparency and Morphological complexity when reading complex nonwords: evidence from English and Italian.

The third study investigated the extent to which morphological processing is affected by orthographic depth and morphological complexity. In a direct contrast of English, a language with an opaque orthography and a scarce morphology, and Italian, a language with a transparent orthography and a rich morphology, we compared reaction times and accuracy scores to lexical decisions with complex nonword stimuli. It is hypothesised that if cross-linguistic differences in orthographic depth had a greater impact on morphological processing than cross-linguistic differences in morphological complexity, then English-speaking participants reaction times would be slower, and accuracy performance be worse, than Italian-speaking participants. The opposite pattern should be found if cross-linguistic differences in morphological complexity were the driving process of morphological processing.

Results showed that morpheme interference effects, while present in the response of both groups, were significantly larger in English compared to Italian, thus further supporting the idea that morpheme-based reading is more dominant in orthographically opaque languages compared to orthographically transparent languages, regardless of morphological complexity. This is a further indication suggesting that inconsistent orthographies require readers to process printed words by referring to bigger grain-size units, as anticipated by the Psycholinguistic Grain Size Theory of Ziegler and Goswami (2005). These units include morphemes, and in fact, one of the key concepts of the Orthographic Depth Hypothesis of Katz and Frost (1992) is that readers of orthographically opaque languages tend to process printed words by referring to their morphology via the written word's visual-orthographic structure. In addition, the findings of Study 1 and 3 further indicate that readers of opaque orthographies rely on several sublexical units according to the task and stimulus at hand, as we have seen how they relied on consistent bodies of the monosyllabic items in Experiment 1 of Study 1 (as compared to the polysyllabic items of Experiment 4), and morphemes in Study 3 in complex pseudowords.

While sublexical units provide pronunciation cues, it is important to note that larger grain size units are not always reliable in their pronunciation either. Bodies in English are not always consistent (see the infamous example of "ough", which elicits different pronunciations in 'cough' /kɒf/, 'tough' /tʌf/, and 'bough' /baʊ/, Grainger, 2018) and morphemes are prompts to pronunciation variability too (see the high entropy values associated to prefixes such as 'pre', 'sur', 'com', 'ex', 'for', 'ar', and 'ad' in Mousikou et al., 2017). This inconsistency even at large grain size units in English provide further evidence for the necessity of readers to develop parallel strategies and to switch between reading units on an item-to-item basis, which is at the core of the flexible-unit-size hypothesis of Brown and Deavers (1999).

General Discussion

Overall, these findings suggest that orthographic depth has a profound impact on sublexical processing: readers of orthographically opaque languages seem to scan the stimuli for pronunciation cues, including the bodies of monosyllabic stimuli (Study 1), and morphemes (Study 3). Moreover, readers of transparent orthographies seem to easily rely on GPCs when reading aloud nonwords (Study 1), but also on larger size units in silent reading (Study 2 and 3).

This pattern of results provides support for the flexible-unit-size hypothesis of Brown and Deavers (1999), which claims that readers tend to use multiple strategies in reading, relying on both small and large units, depending on task and items. For example, English readers could use information about orthographic bodies if the stimulus has a high body neighbourhood (a high density of words that share the same body), and morphemic information when words are multimorphemic, while reverting to GPCs for those graphemes that are consistent.

The flexible-unit-size hypothesis has found support in the field (Metzger, 2017; Perry & Ziegler, 2000; Wyse & Goswami, 2012; Ziegler & Goswami, 2006). In particular, strong support come from Goswami et al. (2003), who found evidence of a switching cost between units of analysis in children. In their study, the authors developed two sets of nonwords where only one possible unit of analysis was available (through phonological recoding via *graphemes* or through orthographic analogies via *bodies*), and a mixed set. Results indicated that English-speaking children's performances were better in terms of speed and accuracy to the first two sets as compared to the mixed set, because they blocked these participants' necessity to strategise.

Modelling cross-linguistic differences and similarities of sublexical units processing

The findings of the current thesis point to the importance of sublexical units when reading in both transparent and opaque orthographies. Several theoretical and computational models of reading have proposed that account for sublexical processing mechanisms within an intermediate level between orthographic input and the orthographic lexicon, thus moving beyond monosyllabic and monomorphemic computational models such as the Dual-Route Cascaded Model of Reading Aloud (Coltheart et al., 2001).

However, very few theories have attempted to account for the combined processing of syllables and morphemes. One step forward in this sense is the Dual-Route Approach to Orthographic Processing (Grainger et al., 2012; Grainger & Ziegler, 2011), a theoretical framework in which both syllables and affixes can make a contribution to the recognition of multisyllabic and multimorphemic words. In this model, two routes exist: the phonology-based route and an orthography route. This latter consists in two sub pathways: the coarse-grained route and the fine-grained route.

The coarse-grained route provides a fast access to semantics. The route marks the presence of instructive letter combinations regardless of letters contiguity (both adjacent and non-adjacent bigrams are incorporated). Comparatively, in the fine-grained route letters are chunked in high-level orthographic representations, which arise as a form of frequently co-occurring adjacent letter combinations (such as multi-letter graphemes and morphemes). These orthographic representations then activate the corresponding phonemes, whose computation eventually activates the whole-word phonological representation, and finally the semantic representation.

In this way, the route accounts for the morpho-orthographic segmentation because the detection of affixes requires precise letter position coding (see the example with "*farmer*" in Figure 5). This approach is consistent with the morphological interference effects found in Study 3. The observed effects arose because both Italian and English participants were sensitive (although to different degrees) to morphemes: their fine-grained route then chunked the nonword strings into affixes, when present. Furthermore, this theory matches our data well because it does not predict variations in the processing of stems and affixes.

In a later account of the model (2012), Grainger et al. predicted that syllabic effects would be accounted for by the fine-grained route, and therefore, syllables could assume the role of functional units of analysis (see also Häikiö et al., 2015). In fact, evidence suggests that phonological processing for polysyllabic words is sequential for both silent reading and reading aloud, as syllabic priming facilitation effects seem to rely on the first syllable only (Carreiras et al., 2005; Chetail & Mathey, 2009, 2012). At the same time, syllabic bodies would also be activated in the fine-grained route, as they are frequent word endings in English.

Our findings from Study 2 fit well in the parallel work of the coarse-grained route and the fine-grained route. Our visual manipulation with hyphenation might have disrupted the usage of the coarse-grained route by skilled readers, which would be the reason why hyphenated conditions were read slower than the control condition. Moreover, the incongruent conditions also had the further disadvantage to disrupt the functioning of the fine-grained route: while the congruent condition was in line with the correct syllabic chunking, the incongruent condition disagreed with it. The chunking conflict resolution would explain the higher reaction times associated with the latter condition.

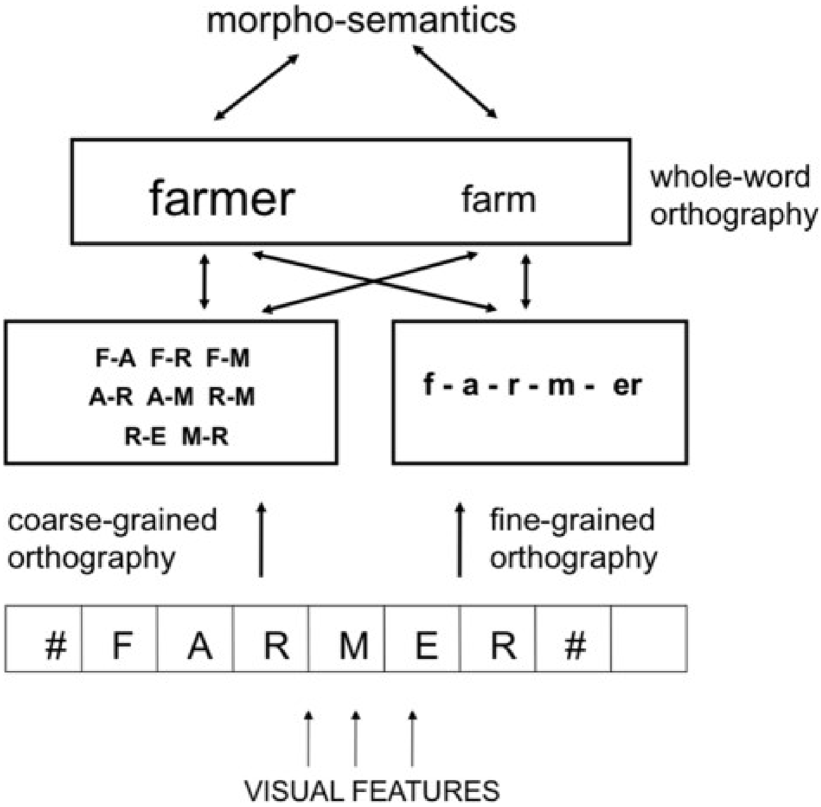
Similar conclusions were also reached by Häikiö et al. (2015), who observed slower reaction times in children reading polysyllabic words with hyphenated syllabic boundaries. The authors speculated that the slowdown occurred because the visual disruption impeded participants to process syllables simultaneously through the coarse-grained route, which forced participants to process words through the first syllable. Moreover, they suggest that the fine-grained route might be phonologically mediated at the early stages of reading, so that children would use syllables as frequently co-occurring letter clusters and process them in a sequential manner (see Figure 6).

Overall, we think that model is apt to predict reading behaviour across languages with different linguistic features and the emergence of cross-linguistic differences in sublexical processing, as the sublexical units that are chunked highly depend on the language at hand. During the development of the reading system, the fine-grained route will become sensitive to specific letter combinations, that could be multi-letter graphemes, affixes, syllables and bodies (p. 9). Moreover, the authors suggest that the associations that are established between sublexical orthographic units and phonological representations arise via supervised learning (which includes teaching strategies). This would explain why we have observed syllabic effects in German adult participants, as German reading instruction emphasise those units. Although the model predicts that phonology gradually becomes less important with age, it is possible that the syllabic effects that we witnessed in Study 2 were both orthographical and phonological in nature, as it has been suggested that the nature of the syllabic effect might reflect both in orthographically transparent orthographies (Conrad et al., 2009).

While this theoretical model seems to be effective in predicting sublexical processing of several units, it is important to note that it has not been implemented computationally yet. Moreover, another challenge would consist in predicting the switch between units according to stimulus nature and tasks (at the very least, in opaque orthographies), as foreseen by the flexible-unit-size hypothesis of Brown and Deavers (1999).

Figure 5

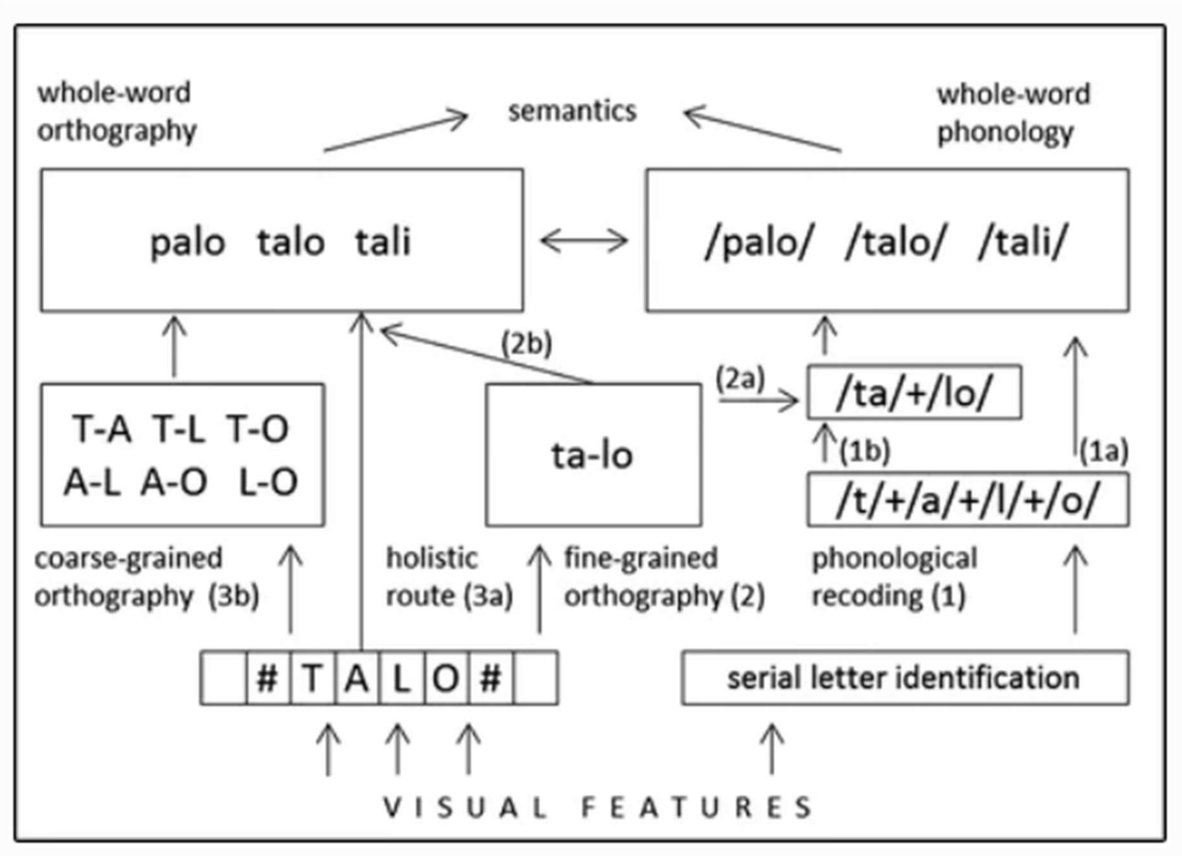
The Dual-Route Approach to Orthographic Processing.



*Note***.** Example with a morphological complex word. From "A dual-route approach to orthographic processing" by Grainger and Ziegler (2011). Frontiers in Psychology, 2, 54, p. 7.

Figure 6

The Dual-Route Approach to Orthographic Processing of Grainger and Ziegler (2011), adapted to Finnish disyllabic reading in Häikiö et al. (2015).



*Note***.** Example with a bisyllabic Finnish Word. From "The role of syllables in word recognition among beginning Finnish readers: Evidence from eye movements during reading" by Häikiö et al. (2015). Journal of Cognitive Psychology, 27(5), p. 574. Copyright 2015 by Routledge

Limitations and Future Directions

This thesis has several limitations, that I summarise below. First of all, its crosslinguistic research only relied on Indo-European languages that are alphabetically transcribed. The results do not necessarily generalise to non-alphabetic languages. For example, while Mousikou et al. (2020) consistently found greater morphological processing in readers of orthographically opaque languages in both children and adults, Barouch et al. (2022), found that morphological effects in Hebrew were only visible in the transparent version of the script in young children, while they become evident in the opaque version of the Hebrew script in older ones. Although readers of Hebrew might be adapting to this highly specific feature of their writing system, this would be a first indication of cross-script developmental differences in morphological processing.

Additionally, while the results from Study 3 seems to point to greater morphological processing in languages that are orthographically opaque, regardless of their morphological complexity, more research needs to be conducted in this area. So far, the handful of studies that have been conducted on this topic, including my own, broadly mirror the same pattern (Mousikou et al., 2020; Simone, Moll, & Beyersmann, 2023; Vannest et al., 2002). However, the existing investigations relied on items built around a combination of existing and non-existing stems and suffixes, which is not representative of the heterogeneity of languages' morphological complexity.

It is important to note that morphological complexity is an umbrella term, and the summation of the means that languages use to derive new words from existing ones and expressing relations among words constituting a sentence. These means are separate: it is customary to divide morphology in the three morphological operations of derivation, inflection, and composition (Leminen et al., 2019), which are also subdivided in their own mechanisms (such as prefixation or suffixation for derivation, for example). Hence, languages not only vary in their morphological richness, but those languages that we consider morphologically complex also vary in their means of achieving this complexity.

Therefore, future cross-linguistic studies aiming at isolating the impact that morphological complexity has on morphological processing, and more generally, sublexical processing, should also considerate the domain in which these languages are complex, especially when selecting affix type. For example, evidence points to differences in the processing of prefixes and suffixes (Beyersmann, Ziegler, & Grainger, 2015) as well as differences between the processing of derived and inflected words (Leminen et al., 2013; Niswander et al., 2000), derived words and compounds (Hasenäcker et al., 2017), and between free and bound morphemes (Coch et al., 2020). In order to have a more clearly defined picture of how morphological complexity and orthographic depth interact during complex word reading, future research need to move forward from utilising suffixed nonwords and include different types of morphemes in their materials. A way to do this would be, for instance, comparing pseudo-compounds in two languages that differ in orthographic depth, such as English and German.

Finally, the results presented in Study 2 did not support the hypothesis that colour alternation facilitates skilled reading. However, the Silbenmethode could affect children in a different way compared to adults: If reading is facilitated in the colour segmented conditions, it would provide important evidence-based support for the Silbenmetode in reading instruction in syllabic complex languages. From there, new doors will open for research in investigating the method in other languages.

Conclusions

In conclusion, this thesis has investigated cross-linguistic differences in sublexical processing in English, French, Italian and German, and how they are mediated by sources of linguistic diversity such as orthographic depth, morphological complexity and syllabic complexity. In Study 1, we have evaluated reliance on graphemes in the four above-mentioned languages, and how orthographic depth impacted pronunciation variability to nonwords. Our findings clearly indicate that complexity and unpredictability are two separate constructs within the orthographic depth conceptual space, and it is the unpredictability of an orthography that impacts pronunciation variability to a greater extent, as seen in Experiment 4. The difference in entropy values in the native English speakers between Experiments 1 and 4 further indicates that English speakers will rely on consistent bodies when present in reading aloud.

Study 3 builds on these findings, suggesting that native speakers of orthographically opaque languages process morphology to a greater extent as compared to speakers of orthographically transparent, but more morphologically complex languages, in lexical decision tasks. This is a further indication that readers of opaque orthographies are attuned to look for islands of regularity in reading, and are able to identify useful sublexical units across stimuli (from consistent bodies in monosyllabic items to morphemes in multisyllabic ones), even when a spoken output is not required by the task. Together these results support the flexible-unit-size hypothesis in English readers of Brown and Deavers (1999). Study 2 further shows that when in an orthographically transparent language like German that is both syllabic and morphologically complex, syllables tend to be the preferred unit of choice in a natural sentence reading context.

To conclude, the findings of the current thesis have provide new insights into how sublexical processing changes according to language specificities in four different languages (English, French, German and Italian) through a wide variety of experimental designs and tasks, thus opening new venues for cross-linguistic research and informing upcoming reading models seeking to integrate sublexical processing in their framework.

References

Alvarez, C. J., Carreiras, M., & Taft, M. (2001). Syllables and morphemes: Contrasting frequency effects in Spanish. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(2), 545–555. https://doi.org/10.1037/0278-7393.27.2.545

Amenta, S., & Crepaldi, D. (2012). Morphological processing as we know it: An analytical review of morphological effects in visual word identification. *Frontiers in Psychology*, *3*, 232. https://doi.org/10.3389/fpsyg.2012.00232

Antzaka, A., Acha, J., Carreiras, M., & Lallier, M. (2019). Does the visual attention span play a role in the morphological processing of orthographic stimuli? *Quarterly Journal of Experimental Psychology (2006)*, *72*(7), 1704–1716. https://doi.org/10.1177/1747021818806470

Ashby, J., & Martin, A. E. (2008). Prosodic phonological representations early in visual word recognition. *Journal of Experimental Psychology. Human Perception and Performance*, *34*(1), 224–236. https://doi.org/10.1037/0096-1523.34.1.224

Ashby, J., & Rayner, K. (2004). Representing syllable information during silent reading: Evidence from eye movements. *Language and Cognitive Processes*, *19*(3), 391–426. https://doi.org/10.1080/01690960344000233

Baayen, H. R., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database* [Computer software]. Linguistic Data Consortium. Philadelphia, PA: University of Pennsylvania.

Barca, L., Burani, C., & Arduino, L. S. (2002). Word naming times and psycholinguistic norms for Italian nouns. *Behavior Research Methods, Instruments, & Computers : A Journal of the Psychonomic Society, Inc*, *34*(3), 424–434. https://doi.org/10.3758/BF03195471

Barca, L., Ellis, A. W., & Burani, C. (2007). Context-sensitive rules and word naming in Italian children. *Reading and Writing*, *20*(5), 495–509. https://doi.org/10.1007/s11145-006-9040-z

Barouch, B., Weiss, Y., Katzir, T., & Bitan, T. (2022). Neural Processing of Morphology During Reading in Children. *Neuroscience*, *485*, 37–52. https://doi.org/10.1016/j.neuroscience.2021.12.025

Bates, E., Burani, C., D'Amico, S., & Barca, L. (2001). Word reading and picture naming in Italian. *Memory & Cognition*, *29*(7), 986–999. https://doi.org/10.3758/BF03195761

Bates, E., Devescovi, A., & Wulfeck, B. (2001). Psycholinguistics: A cross-language perspective. *Annual Review of Psychology*, *52*, 369–396. https://doi.org/10.1146/annurev.psych.52.1.369

Beauvillain, C. (1994). Morphological structure in visual word recognition: Evidence from prefixed and suffixed words. *Language and Cognitive Processes*, *9*(3), 317–339. https://doi.org/10.1080/01690969408402122

Beyersmann, E., Casalis, S., Ziegler, J. C., & Grainger, J. (2015). Language proficiency and morpho-orthographic segmentation. *Psychonomic Bulletin & Review*, *22*(4), 1054–1061. https://doi.org/10.3758/s13423-014-0752-9

Beyersmann, E., & Grainger, J. (2023). The Role of Embedded Words and Morphemes in Reading. In D. Crepaldi (Ed.), *Current issues in the psychology of language. Linguistic morphology in the mind and brain* (pp. 26–49). Routledge.

Beyersmann, E., McCormick, S. F., & Rastle, K. (2013). Letter transpositions within morphemes and across morpheme boundaries. *Quarterly Journal of Experimental Psychology (2006)*, *66*(12), 2389–2410. https://doi.org/10.1080/17470218.2013.782326

Beyersmann, E., Mousikou, P., Javourey-Drevet, L., Schroeder, S., Ziegler, J. C., & Grainger, J. (2020). Morphological Processing across Modalities and Languages. *Scientific Studies of Reading*, *24*(6), 500–519. https://doi.org/10.1080/10888438.2020.1730847

Beyersmann, E., Ziegler, J. C., & Grainger, J. (2015). Differences in the Processing of Prefixes and Suffixes Revealed by a Letter-Search Task. *Scientific Studies of Reading*, *19*(5), 360–373. https://doi.org/10.1080/10888438.2015.1057824

Bjarnadóttir, K., Hlynsdóttir, K. I., & Steingrímsson, S. (2019). DIM: The database of Icelandic morphology. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics* (pp. 146–154).

Bloomfield, L. (1933). *Language and Linguistics.* New York Holt.

Bölücü, N., & Can, B. (2019). Unsupervised Joint PoS Tagging and Stemming for Agglutinative Languages. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *18*(3), 1–21. https://doi.org/10.1145/3292398

Borleffs, E., Maassen, B. A. M., Lyytinen, H., & Zwarts, F. (2017). Measuring orthographic transparency and morphological-syllabic complexity in alphabetic orthographies: A narrative review. *Reading and Writing*, *30*(8), 1617–1638. https://doi.org/10.1007/s11145-017-9741-5

Bowey, J. A. (1990). Orthographic onsets and rimes as functional units of reading. *Memory & Cognition*, *18*(4), 419–427. https://doi.org/10.3758/BF03197130

Brand, M., Giroux, I., Puijalon, C., & Rey, A. (2007). Syllable onsets are perceptual reading units. *Memory & Cognition*, *35*(5), 966–973. https://doi.org/10.3758/BF03193469

Brand, M., Rey, A., & Peereman, R. (2003). Where is the syllable priming effect in visual word recognition? *Journal of Memory and Language*, *48*(2), 435–443.

Brown, D. (2010). Morphological Typology. In Jae Jung Song (Ed.), *The Oxford Handbook of Linguistic Typology* (pp. 487–503). Oxford University Press.

Brown, D., & Deavers, R. P. (1999). Units of Analysis in Nonword Reading: Evidence from Children and Adults. *Journal of Experimental Child Psychology*, *73*(3), 208–242. https://doi.org/10.1006/jecp.1999.2502

Burani, C., Marcolini, S., & Stella, G. (2002). How early does morpholexical reading develop in readers of a shallow orthography? *Brain and Language*, *81*(1-3), 568–586. https://doi.org/10.1006/brln.2001.2548

Burani, C., Thornton, A. M., & Zoccolotti, P. (2014). Literacy acquistion in Italian. In L. Verhoeven & C. Perfetti (Eds.), *Reading acquisition across languages and writing systems: An international handbook.* Cambridge University Press.

Cara, B. de, & Goswami, U. (2002). Similarity relations among spoken words: The special status of rimes in English. *Behavior Research Methods, Instruments, & Computers : A Journal of the Psychonomic Society, Inc*, *34*(3), 416–423. https://doi.org/10.3758/BF03195470

Carreiras, M., Álvarez, C. J., & Devega, M. (1993). Syllable frequency and visual word recognition in Spanish. *Journal of Memory and Language*, *32*(6), 766–780.

Carreiras, M., Ferrand, L., Grainger, J., & Perea, M. (2005). Sequential effects of phonological priming in visual word recognition. *Psychological Science*, *16*(8), 585–589. https://doi.org/10.1111/j.1467-9280.2005.01579.x

Casalis, S., Quémart, P., & Duncan, L. G. (2015). How language affects children's use of derivational morphology in visual word and pseudoword processing: Evidence from a cross-language study. *Frontiers in Psychology*, *6*, 452. https://doi.org/10.3389/fpsyg.2015.00452

Chetail, F., & Mathey, S. (2009). Syllabic priming in lexical decision and naming tasks: The syllable congruency effect re-examined in French. *Canadian Journal of Experimental Psychology = Revue Canadienne De Psychologie Experimentale*, *63*(1), 40–48. https://doi.org/10.1037/a0012944

Chetail, F., & Mathey, S. (2010). InfoSyll: A syllabary providing statistical information on phonological and orthographic syllables. *Journal of Psycholinguistic Research*, *39*(6), 485–504.

Chetail, F., & Mathey, S. (2012). Effect of Syllable Congruency in Sixth Graders in the Lexical Decision Task With Masked Priming. *Scientific Studies of Reading*, *16*(6), 537–549. https://doi.org/10.1080/10888438.2011.607480

Chyl, K., Kossowski, B., Wang, S., Dębska, A., Łuniewska, M., Marchewka, A., Wypych, M., van Bunt, M. den, Mencl, W., Pugh, K., & Jednoróg, K. (2021). The brain signature of emerging reading in two contrasting languages. *NeuroImage*, *225*, 117503. https://doi.org/10.1016/j.neuroimage.2020.117503

Coch, D., Hua, J., & Landers-Nelson, A. (2020). All morphemes are not the same: accuracy and response times in a lexical decision task differentiate types of morphemes. *Journal of Research in Reading*, *43*(3), 329–346. https://doi.org/10.1111/1467-9817.12306

Coltheart, M. (2005). Modeling Reading: The Dual-Route Approach. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 6–23). Blackwell Publishing Ltd.

Coltheart, M. (2006). Dual route and connectionist models of reading: an overview. *London Review of Education*, *4*(1). https://doi.org/10.1080/13603110600574322

Coltheart, M. (2014). Dual-Route Model of Reading Development. In P. Brooks & V. Kempe (Eds.), *Encyclopedia of language development* (pp. 167–168). SAGE Reference.

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. https://doi.org/10.1037/0033-295X.108.1.204

Conrad, M., Carreiras, M., Tamm, S., & Jacobs, A. M. (2009). Syllables and bigrams: Orthographic redundancy and syllabic units affect visual word recognition at different processing levels. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 461–479. https://doi.org/10.1037/a0013480

Conrad, M., & Jacobs, A. (2004). Replicating syllable frequency effects in Spanish in German: One more challenge to computational models of visual word recognition. *Language and Cognitive Processes*, *19*(3), 369–390. https://doi.org/10.1080/01690960344000224

Conrad, M., Tamm, S., Carreiras, M., & Jacobs, A. M. (2010). Simulating syllable frequency effects within an interactive activation framework. *European Journal of Cognitive Psychology*, *22*(5), 861–893. https://doi.org/10.1080/09541440903356777

Conrad, M., Vo, M., Schneider, D., & Jacobs, A. M. (2011). Syllable structure is modulating the optimal viewing position in visual word recognition. *Revista De Logopedia, Foniatría Y Audiología*, *31*(1), 14–21. https://doi.org/10.1016/S0214-4603(11)70166-8

Coupé, C., Oh, Y. M., Pellegrino, F., & Marsico, E. (2014). Cross-linguistic investigations of oral and silent reading. In *Interspeech 2014* (pp. 514–518). ISCA. https://doi.org/10.21437/interspeech.2014-127

De Simone, E., Beyersmann, E., Mulatti, C., Mirault, J., & Schmalz, X. (2021). Order among chaos: Cross-linguistic differences and developmental trajectories in pseudoword reading aloud using pronunciation Entropy. *PloS One*, *16*(5), e0251629. https://doi.org/10.1371/journal.pone.0251629

De Simone, E., Moll, K., & Beyersmann, E. (2023). The Role of Orthographic Transparency and Morphological Complexity when Reading Complex Nonwords: Evidence from English and Italian. *Manuscript Submitted for Publication.*

De Simone, E., Moll, K., Feldmann, L., Schmalz, X., & Beyersmann, E. (2023). The role of syllables and morphemes in silent reading: An eye-tracking study. *Quarterly Journal of Experimental Psychology (2006)*, 17470218231160638. https://doi.org/10.1177/17470218231160638

Doignon, N., & Zagar, D. (2005). Illusory conjunctions in French: The nature of sublexical units in visual word recognition. *Language and Cognitive Processes*, *20*(3), 443–464.

Doignon-Camus, N., Zagar, D., & Mathey, S. (2009). Can we see syllables in monosyllabic words? A study with illusory conjunctions. *European Journal of Cognitive Psychology*, *21*(4), 599–614.

Easterday, S. (2019). *Highly complex syllable structure*. Language Science Press.

Egan, C., Oppenheim, G. M., Saville, C., Moll, K., & Jones, M. W. (2019). Bilinguals apply language-specific grain sizes during sentence reading. *Cognition*, *193*, 104018. https://doi.org/10.1016/j.cognition.2019.104018

Ellis, N. C., & Hooper, A. M. (2001). It is easier to learn to read in Welsh than in English: Effects of orthographic transparency demonstrated using frequency-matched cross-linguistic reading tests. *Applied Psycholinguistics*, *22*(4), 571–599.

Ellis, N. C., Natsume, M., Stavropoulou, K., Hoxhallari, L., Daal, V. H., Polyzoe, N., Tsipa, M.-L., & Petalas, M. (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading Research Quarterly*, *39*(4), 438–468. https://doi.org/10.1598/RRQ.39.4.5

Fenk-Oczlon, G., & Pilz, J. (2021). Linguistic Complexity: Relationships Between Phoneme Inventory Size, Syllable Complexity, Word and Clause Length, and Population Size. *Frontiers in Communication*, *6*, Article 626032. https://doi.org/10.3389/fcomm.2021.626032

Ferrand, L., Segui, J., & Grainger, J. (1996). Masked priming of word and picture naming: The role of syllabic units. *Journal of Memory and Language*, *35*(5), 708–723.

Ferrand, L., Segui, J., & Humphreys, G. W. (1997). The syllable's role in word naming. *Memory & Cognition*, *25*(4), 458–470.

Frauenfelder, U. H., & Schreuder, R. (1992). Constraining psycholinguistic models of morphological processing and representation: The role of productivity. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology. Yearbook of Morphology 1991* (pp. 165–183). Springer Netherlands. https://doi.org/10.1007/978-94-011-2516-1_10

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. https://doi.org/10.1207/s1532799xssr0201_2

Frost, R. (2005). Orthographic Systems and Skilled Word Recognition Processes in Reading. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 272–295). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470757642.ch15

Frost, R., & Grainger, J. (2000). Cross-linguistic perspectives on morphological processing: An introduction. *Language and Cognitive Processes*, *15*(4-5), 321–328.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology. Human Perception and Performance*, *13*(1), 104–115. https://doi.org/10.1037/0096-1523.13.1.104

Giraudo, H., & Grainger, J. (2003). A Supralexical Model for French Derivational Morphology. In R. M. Joshi, E. Grigorenko, E. M. H. Assink, & D. Sandra (Eds.), *Neuropsychology and Cognition. Reading Complex Words* (Vol. 22, pp. 139–157). Springer US. https://doi.org/10.1007/978-1-4757-3720-2_7

Goswami, U. (2010a). Phonological development across different languages. In R. Andrews, D. Wyse, & J. V. Hoffman (Eds.), *Routledge international handbook of English, language and literacy teaching* (pp. 98–109). Routledge.

Goswami, U. (2010b). A psycholinguistic grain size view of reading acquisition across languages. In N. Brunswick, S. McDougall, & P. de Mornay Davies (Eds.), *Reading and dyslexia in different orthographies.* Psychology Press.

Goswami, U., Ziegler, J. C., Dalton, L., & Schneider, W. (2003). Nonword reading across orthographies: How flexible is the choice of reading units? *Applied Psycholinguistics*, *24*(2), 235–247. https://doi.org/10.1017/S0142716403000134

Gottardo, A., Pasquarella, A., Chen, X. I., & Ramirez, G. (2016). The impact of language on the relationships between phonological awareness and word reading in different orthographies: A test of the psycholinguistic grain size theory in bilinguals. *Applied Psycholinguistics*, *37*(5), 1083–1115. https://doi.org/10.1017/S0142716415000508

Grainger, J. (2018). Orthographic processing: A 'mid-level' vision of reading: The 44th Sir Frederic Bartlett Lecture. *Quarterly Journal of Experimental Psychology (2006)*, *71*(2), 335–359. https://doi.org/10.1080/17470218.2017.1314515

Grainger, J., & Beyersmann, E. (2017). Edge-Aligned Embedded Word Activation Initiates Morpho-orthographic Segmentation, *67*, 285–317. https://doi.org/10.1016/bs.plm.2017.03.009

Grainger, J., Colé, P., & Segui, J. (1991). Masked morphological priming in visual word recognition. *Journal of Memory and Language*, *30*(3), 370–384. https://doi.org/10.1016/0749-596X(91)90042-I

Grainger, J., Lété, B., Bertand, D., Dufau, S., & Ziegler, J. C. (2012). Evidence for multiple routes in learning to read. *Cognition*, *123*(2), 280–292. https://doi.org/10.1016/j.cognition.2012.01.003

Grainger, J., & Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology*, *2*, 54. https://doi.org/10.3389/fpsyg.2011.00054

Greenberg, J. H. (1954). A Quantitative Approach to the Morphological Typology of Language. In R. F. Spencer (Ed.), *Method and Perspective in Anthropology: Papers in Honor of Wilson D. Wallis* (pp. 192–220). University of Minnesota Press.

Häikiö, T., Hyönä, J., & Bertram, R. (2015). The role of syllables in word recognition among beginning Finnish readers: Evidence from eye movements during reading. *Journal of Cognitive Psychology*, *27*(5), 562–577. https://doi.org/10.1080/20445911.2014.982126

Hasenäcker, J., Beyersmann, E., & Schroeder, S. (2016). Masked Morphological Priming in German-Speaking Adults and Children: Evidence from Response Time Distributions. *Frontiers in Psychology*, *7*, 929. https://doi.org/10.3389/fpsyg.2016.00929

Hasenäcker, J., Ktori, M., & Crepaldi, D. (2021). Morpheme Position Coding in Reading Development as Explored With a Letter Search Task. *Journal of Cognition*, *4*(1), 16. https://doi.org/10.5334/joc.153.

Hasenäcker, J., & Schroeder, S. (2017). Syllables and morphemes in German reading development: Evidence from second graders, fourth graders, and adults. *Applied Psycholinguistics*, *38*(3), 733–753. https://doi.org/10.1017/S0142716416000412

Hasenäcker, J., Schröter, P., & Schroeder, S. (2017). Investigating developmental trajectories of morphemes as reading units in German. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(7), 1093–1108. https://doi.org/10.1037/xlm0000353

Havas, V., Waris, O., Vaquero, L., Rodríguez-Fornells, A., & Laine, M. (2015). Morphological learning in a novel language: A cross-language comparison. *Quarterly Journal of Experimental Psychology (2006)*, *68*(7), 1426–1441. https://doi.org/10.1080/17470218.2014.983531

Hawelka, S., Schuster, S., Gagl, B., & Hutzler, F. (2013). Beyond single syllables: The effect of first syllable frequency and orthographic similarity on eye movements during silent reading. *Language and Cognitive Processes*, *28*(8), 1134–1153. https://doi.org/10.1080/01690965.2012.696665

Healy, A. F. (1976). Detection errors on the word the: Evidence for reading units larger than letters. *Journal of Experimental Psychology. Human Perception and Performance*, *2*(2), 235–242. https://doi.org/10.1037/0096-1523.2.2.235

Hutzler, F., Conrad, M., & Jacobs, A. M. (2005). Effects of syllable-frequency in lexical decision and naming: An eye-movement study. *Brain and Language*, *92*(2), 138–152. https://doi.org/10.1016/j.bandl.2004.06.001

Hyönä, J. (2015). Are Polymorphemic Words Processed Differently From Other Words During Reading? In A. Pollatsek & R. Treiman (Eds.), *The Oxford Handbook of Reading* (pp. 114–128). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199324576.013.13

Kargl, R., & Landerl, K. (2018). Beyond Phonology: The Role of Morphological and Orthographic Spelling Skills in German. *Topics in Language Disorders*, *38*(4), 272–285. https://doi.org/10.1097/TLD.0000000000000165

Katz, L., & Frost, R. (1992). Chapter 4 The Reading Process is Different for Different Orthographies: The Orthographic Depth Hypothesis. In *Advances in Psychology. Orthography, Phonology, Morphology, and Meaning* (Vol. 94, pp. 67–84). Elsevier. https://doi.org/10.1016/S0166-4115(08)62789-2

Kemp, N., & Treiman, R. (2022). Learning to Spell Words. In M. J. Snowling, C. Hulme, & K. Nation (Eds.), *The Science of Reading* (Vol. 12, pp. 165–185). Wiley. https://doi.org/10.1002/9781119705116.ch8

Kliegl, R., Nuthmann, A., & Engbert, R. (2004). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology. General*, *135*(1), 12–35. https://doi.org/10.1037/0096-3445.135.1.12

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., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German-English comparison. *Cognition*, *63*(3), 315–334. https://doi.org/10.1016/S0010-0277(97)00005-X

Leminen, A., Leminen, M., Kujala, T., & Shtyrov, Y. (2013). Neural dynamics of inflectional and derivational morphology processing in the human brain. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *49*(10), 2758–2771. https://doi.org/10.1016/j.cortex.2013.08.007

Leminen, A., Smolka, E., Duñabeitia, J. A., & Pliatsikas, C. (2019). Morphological processing in the brain: The good (inflection), the bad (derivation) and the ugly (compounding). *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *116*, 4–44. https://doi.org/10.1016/j.cortex.2018.08.016

Levitt, A., Healy, A. F., & Fendrich, D. W. (1991). Syllable-internal structure and the sonority hierarchy: Differential evidence from lexical decision, naming, and reading. *Journal of Psycholinguistic Research*, *20*(4), 337–363.

Lyster, S.-A. H. (2002). The effects of morphological versus phonological awareness training in kindergarten on reading development. *Reading and Writing*, *15*(3/4), 261–294. https://doi.org/10.1023/A:1015272516220

Lyster, S.-A. H., Lervåg, A. O., & Hulme, C. (2016). Preschool morphological training produces long-term improvements in reading comprehension. *Reading and Writing*, *29*, 1269–1288. https://doi.org/10.1007/s11145-016-9636-x

Maddieson, I. (2006). Correlating Phonological Complexity: Data and Validation. *UC Berkeley Phonology Lab Annual Reports*, *1*. https://doi.org/10.5070/P795m171v6

Maddieson, I. (2013). Syllable structure. In Matthew Dryer, Martin Haspelmath, & Robert Forkel (Eds.), *The World Atlas of Language Structures Online*. Zenodo.

Marinelli, C. V., Romani, C., Burani, C., McGowan, V. A., & Zoccolotti, P. (2016). Costs and Benefits of Orthographic Inconsistency in Reading: Evidence from a Cross-Linguistic Comparison. *PloS One*, *11*(6), e0157457. https://doi.org/10.1371/journal.pone.0157457

Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, *101*(1), 3–33. https://doi.org/10.1037/0033-295X.101.1.3

Metzger, J. (2017). The Effects of Real Words on the Pronunciation of Nonwords: Honours Scholar Thesis(535). https://opencommons.uconn.edu/srhonors_theses

Mousikou, P., Beyersmann, E., Ktori, M., Javourey-Drevet, L., Crepaldi, D., Ziegler, J. C., Grainger, J., & Schroeder, S. (2020). Orthographic consistency influences morphological processing in reading aloud: Evidence from a cross-linguistic study. *Developmental Science*, *23*(6), e12952. https://doi.org/10.1111/desc.12952

Mousikou, P., Sadat, J., Lucas, R., & Rastle, K. (2017). Moving beyond the monosyllable in models of skilled reading: Mega-study of disyllabic nonword reading. *Journal of Memory and Language*, *93*, 169–192. https://doi.org/10.1016/j.jml.2016.09.003

Naveh-Benjamin, M., & Ayres, T. J. (1986). Digit span, reading rate, and linguistic relativity. *The Quarterly Journal of Experimental Psychology Section a*, *38*(4), 739–751. https://doi.org/10.1080/14640748608401623

Niswander, E., Pollatsek, A., & Rayner, K. (2000). The processing of derived and inflected suffixed words during reading. *Language and Cognitive Processes*, *15*(4-5), 389–420. https://doi.org/10.1080/016909600501119643

Niswander-Klement, E., & Pollatsek, A. (2006). The effects of root frequency, word frequency, and length on the processing of prefixed English words during reading. *Memory & Cognition*, *34*(3), 685–702. https://doi.org/10.3758/BF03193588

Paizi, D., Luca, M. de, Zoccolotti, P., & Burani, C. (2013). A comprehensive evaluation of lexical reading in Italian developmental dyslexics. *Journal of Research in Reading*, *36*(3), 303–329. https://doi.org/10.1111/j.1467-9817.2011.01504.x

Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S. F., Cotelli, M., Cossu, G., Corte, F., Lorusso, M., Pesenti, S., Gallagher, A., Perani, D., Price, C., Frith, C. D., & Frith, U. (2000). A cultural effect on brain function. *Nature Neuroscience*, *3*(1), 91–96. https://doi.org/10.1038/71163

Pellegrino, F., Coupé, C., & Marsico, E. (2011). A cross-language perspective on speech information rate. *Language*, *87*(3), 539–558.

Perfetti, C., & Harris, L. N. (2013). Universal Reading Processes Are Modulated by Language and Writing System. *Language Learning and Development*, *9*(4), 296–316. https://doi.org/10.1080/15475441.2013.813828

Perfetti, C., & Helder, A. (2022). Progress in Reading Science. In M. J. Snowling, C. Hulme, & K. Nation (Eds.), *The Science of Reading* (Vol. 6, pp. 5–35). Wiley. https://doi.org/10.1002/9781119705116.ch1

Perry, C., & Ziegler, J. C. (2000). Linguistic difficulties in language and reading development constrain skilled adult reading. *Memory & Cognition*, *28*(5), 739–745.

Pirkola, A. (2001). Morphological typology of languages for IR. *Journal of Documentation*, *57*(3), 330–348. https://doi.org/10.1108/EUM0000000007085

Pollatsek, A., Hyönä, J [J.], & Bertram, R [R.] (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology. Human*

Perception and Performance, *26*(2), 820–833. https://doi.org/10.1037/0096-1523.26.2.820

Prinzmetal, W., Treiman, R., & Rho, S. H. (1986). How to see a reading unit. *Journal of Memory and Language*, *25*(4), 461–475. https://doi.org/10.1016/0749-596X(86)90038-0

Rastle, K. (2022). Word Recognition III: Morphological Processing. In M. J. Snowling, C. Hulme, & K. Nation (Eds.), *The Science of Reading* (pp. 102–119). Wiley.

Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, *23*(7-8), 942–971. https://doi.org/10.1080/01690960802069730

Rau, A. K., Moll, K., Snowling, M. J., & Landerl, K. (2015). Effects of orthographic consistency on eye movement behavior: German and English children and adults process the same words differently. *Journal of Experimental Child Psychology*, *130*, 92–105. https://doi.org/10.1016/j.jecp.2014.09.012

Rey, A., Ziegler, J. C., & Jacobs, A. M. (2000). Graphemes are perceptual reading units. *Cognition*, *75*(1), B1-12. https://doi.org/10.1016/S0010-0277(99)00078-5

Sapir, E. (1921). *Language: An Introduction to the Study of Speech*. Harcourt, Brace & World.

Schiller, N. O. (2000). Single word production in english: The role of subsyllabic units during phonological encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(2), 512–528. https://doi.org/10.1037/0278-7393.26.2.512

Schlegel, F. (1808). *Ueber die Sprache und Weisheit der Ein Beitrag zur Begru¨ndung der Altertumskunde*. Mohr and Zimmer.

Schmalz, X., Beyersmann, E., Cavalli, E., & Marinus, E. (2016). Unpredictability and complexity of print-to-speech correspondences increase reliance on lexical processes: more evidence for the orthographic depth hypothesis. *Journal of Cognitive Psychology*, *28*(6), 658–672. https://doi.org/10.1080/20445911.2016.1182172

Schmalz, X., Marinus, E., Coltheart, M., & Castles, A. (2015). Getting to the bottom of orthographic depth. *Psychonomic Bulletin & Review*, *22*(6), 1614–1629. https://doi.org/10.3758/s13423-015-0835-2

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. https://doi.org/10.1080/20445911.2014.968161

Schreuder, R., & Baayen, H. R. (1995). Modeling Morphological Processing. In L. B. Feldman (Ed.), *Morphological Aspects of Language Processing* (pp. 257–294). Psychology Press.

Schroeder, S., Würzner, K.-M., Heister, J., Geyken, A., & Kliegl, R. (2015). Childlex: A lexical database of German read by children. *Behavior Research Methods*, *47*(4), 1085–1094. https://doi.org/10.3758/s13428-014-0528-1

Segers, E., & Verhoeven, L. (2005). Long-term effects of computer training of phonological awareness in kindergarten. *Journal of Computer Assisted Learning*, *21*(1), 17–27. https://doi.org/10.1111/j.1365-2729.2005.00107.x

Seidenberg, M. S. (1987). Sublexical structures in visual word recognition: Access units or orthographic redundancy?. In *Attention and performance XII* (pp. 245-263). Routledge.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology (London, England: 1953)*, *94*(Pt 2), 143–174. https://doi.org/10.1348/000712603321661859'

Stenneken, P., Conrad, M., & Jacobs, A. (2007). Processing of syllables in production and recognition tasks. *Journal of Psycholinguistic Research*, *36*(1), 65–78. https://doi.org/10.1007/s10936-006-9033-8

Stevens, P., & Plaut, D. C. (2022). From decomposition to distributed theories of morphological processing in reading. *Psychonomic Bulletin & Review*, *29*(5), 1673–1702. https://doi.org/10.3758/s13423-022-02086-0

Stump, G. T. (2001). Inflection. In A. Spencer & A. Zwicky (Eds.), *The handbook ofmorphology* (pp. 13–43). Wiley.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 638–647. https://doi.org/10.1016/S0022-5371(75)80051-X

Taft, M. (1994). Interactive-activation as a framework for understanding morphological processing. *Language and Cognitive Processes*, *9*(3), 271–294. https://doi.org/10.1080/01690969408402120

Taft, M. (2003). Morphological Representation as a Correlation Between form and Meaning. In R. M. Joshi, E. Grigorenko, E. M. H. Assink, & D. Sandra (Eds.), *Neuropsychology and Cognition. Reading Complex Words* (Vol. 22, pp. 113–137). Springer US. https://doi.org/10.1007/978-1-4757-3720-2_6

Treiman, R. (1983). The structure of spoken syllables: Evidence from novel word games. *Cognition*, *15*(1-3), 49–74.

Treiman, R. (1986). The division between onsets and rimes in English syllables. *Journal of Memory and Language*, *25*(4), 476–491.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology. General*, *124*(2), 107–136. https://doi.org/10.1037/0096-3445.124.2.107

Vannest, J., Bertram, R., Järvikivi, J., & Niemi, J. (2002). Counterintuitive cross-linguistic differences: More morphological computation in English than in Finnish. *Journal of Psycholinguistic Research*, *31*(2), 83–106. https://doi.org/10.1023/A:1014934915952

Velten, H. R. (2012). Frühe Lese- und Schreiblernbücher des 16. Jahrhunderts. Zu Valentin Ickelsamers Die rechte weis, aufs kürtzist lesen zu lernen (1527) und Teütsche Grammatica (1532?). *Zeitschrift Für Erziehungswissenschaft*, *15*(S2), 31–48. https://doi.org/10.1007/s11618-012-0309-5

Verhoeven, L., & Perfetti, C. A. (2011). Morphological processing in reading acquisition: A cross-linguistic perspective. *Applied Psycholinguistics*, *32*(3), 457–466. https://doi.org/10.1017/S0142716411000154

Wyse, D., & Goswami, U. (2012). Early reading development. *Handbook of Early Childhood Literacy*, 379–394.

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, *60*(4), 502–529. https://doi.org/10.1016/j.jml.2009.02.001

Zec, D. (2007). The syllable. In P. de Lacy (Ed.), *The Cambridge Handbook of Phonology.* Cambridge University Press.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3

Ziegler, J. C., & Goswami, U. (2006). Becoming literate in different languages: Similar problems, different solutions. *Developmental Science*, *9*(5), 429–436. https://doi.org/10.1111/j.1467-7687.2006.00509.x

Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science*, *12*(5), 379–384. https://doi.org/10.1111/1467-9280.00370

Study 1

Order among Chaos: Cross-linguistic Differences and Developmental Trajectories in Pseudoword Reading Aloud using Pronunciation Entropy.

Elisabetta De Simone
Elisabeth Beyersmann
Claudio Mulatti
Jonathan Mirault
Xenia Schmalz

Study published in *PloS One*:

De Simone, E., Beyersmann, E., Mulatti, C., Mirault, J., & Schmalz, X. (2021). Order among chaos: Cross-linguistic differences and developmental trajectories in pseudoword reading aloud using pronunciation Entropy. PloS one, 16(5), e0251629. https://doi.org/10.1371/journal.pone.0251629

RESEARCH ARTICLE

Order among chaos: Cross-linguistic differences and developmental trajectories in pseudoword reading aloud using pronunciation Entropy

Elisabetta De Simone[1☯]*, **Elisabeth Beyersmann**[2☯], **Claudio Mulatti**[3☯], **Jonathan Mirault**[4☯], **Xenia Schmalz**[1☯]

1 Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital, LMU Munich, Munich, Germany, **2** Department of Cognitive Science, Macquarie University Centre of Reading, Sydney, Australia, **3** Department of Psychology and Cognitive Science, Università degli studi di Trento, Trento, Italy, **4** Aix-Marseille University and Centre National de la Recherche Scientifique, Marseille, France

☯ These authors contributed equally to this work.
* Elisabetta.De@med.uni-muenchen.de

Check for updates

OPEN ACCESS

Citation: De Simone E, Beyersmann E, Mulatti C, Mirault J, Schmalz X (2021) Order among chaos: Cross-linguistic differences and developmental trajectories in pseudoword reading aloud using pronunciation Entropy. PLoS ONE 16(5): e0251629. https://doi.org/10.1371/journal.pone.0251629

Editor: Madelon van den Boer, Universiteit van Amsterdam, NETHERLANDS

Received: May 20, 2020

Accepted: April 30, 2021

Published: May 19, 2021

Data Availability Statement: All data and script files are available at this link https://osf.io/94wjx/.

Funding: EB was supported by a Discovery Early Career Researcher Award (DECRA) by the Australian Research Council (DE190100850) https://www.arc.gov.au/grants/discovery-program/discovery-early-career-researcher-award-decra The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

In this work we propose the use of Entropy to measure variability in pronunciations in pseudowords reading aloud: pseudowords where participants give many different pronunciations receive higher Entropy values. Monolingual adults, monolingual children, and bilingual children proficient in different European languages varying in orthographic depth were tested. We predicted that Entropy values will increase with increasing orthographic depth. Moreover, higher Entropy was expected for younger than older children, as reading experience improves the knowledge of grapheme-phoneme correspondences (GPCs). We also tested if interference from a second language would lead to higher Entropy. Results show that orthographic depth affects Entropy, but only when the items are not strictly matched across languages. We also found that Entropy decreases across age, suggesting that GPC knowledge becomes refined throughout grades 2-4. We found no differences between bilingual and monolingual children. Our results indicate that item characteristics play a fundamental role in pseudoword pronunciation variability, that reading experience is associated with reduced variability in responses, and that in bilinguals' knowledge of a second orthography does not seem to interfere with pseudoword reading aloud.

Introduction

It is common practice in reading research to use pseudowords in order test participants' ability to use grapheme-phoneme correspondences (GPCs) to correctly retrieve sound from print [1]. This ability is considered fundamental to learning to read: since children at the beginning of reading acquisition do not have a large sight vocabulary, they need to more heavily on their

Competing interests: The authors have declared that no competing interests exist.

knowledge of letter-sound correspondences to assemble the correct pronunciation, a process known as phonological decoding [2].

Pseudowords have received a great deal of attention in this field. Pseudowords are graphotactically legal stimuli with plausible pronunciations [3]. Their importance lies in their helpfulness in predicting poor reading skills: studies have shown that dyslexic readers perform worse than their non-impaired peers on pseudowords reading aloud tasks [4, 5]. Pseudowords are usually assessed by calculating reaction times (the time between stimulus onset and voice onset) and reading accuracy (the number of errors that participants make while reading). Concerning reaction times, two assumptions underlie its use for inference: Firstly, they have to assume that if a participant is taking more time in naming a particular item, it means that item is more difficult than others. Secondly, the researcher has to hypothesise about features of that particular item that make it difficult to name. For example, when 100 participants read aloud two pseudowords, "rop" and "wap", they might have faster reaction times to the former than to the latter. With this finding, we can calculate differences, on the linguistic level, between these two pseudowords (e.g. in terms of vowel consistency, orthographic neighborhood or letter bigram frequency). This would allow for indirect inferences about which linguistic characteristics affect reading aloud processes, which would, in turn, allow us to hypothesise a cognitive structure that would explain why this particular characteristic should affect reading processes. The transcribed responses of the participants give more direct information about the cognitive processes [6–8]. For example, for the two pseudowords above, participants might pronounce the former consistently as /ɹɔp/, and for the latter, some participants might pronounce the pseudoword as /wæp/ or as /wɔp/. This is more direct evidence that consistency (i.e., the presence of more than one possible pronunciation for the letter cluster *wa*, "in wasp" versus "wax") affects reading aloud processes.

As for accuracy, since pseudowords do not have conventional pronunciations, it is difficult to decide whether they are pronounced correctly or not [7, 9]. Often, faced with the variety of responses participants give, researchers need to arbitrarily decide whether a pseudoword is correctly read by analysing all the plausible pronunciations that they think it could have [9–11]. Even if a given software is used to score accuracy, decisions need to be made concerning response accuracy. For example, if we accept any pronunciation as correct whenever there is at least one instance of the grapheme-phoneme correspondence in the language, we would consider, the pronunciation /jɔn/ for the English pseudoword <yan> as correct, although, intuitively, most English native speakers would consider this pronunciation incorrect, because it corresponds to the vowel pronunciation of the word <yacht>.

With this in mind, we aim to investigate the number and kind of different pronunciations participants give, an information that is not captured by only scoring the answers as correct and incorrect [6–8]: The quantification of response variability to a given pseudoword may be a more sensitive measure of pseudoword reading aloud performance, since it does not involve any kind of arbitrary decisions from the researchers. Of course, the variability of responses and accuracy may be correlated: If participants give many different pronunciations to a given pseudoword, by definition, the variability will be high for this item. This also implies that any scoring scheme would likely mark more responses as incorrect.

Considering this, our study's goal was to test an alternative variable, namely pseudoword reading aloud Entropy, as a way to quantifying participants' pseudoword reading aloud performances [12]. This approach has the advantage that rather than making decisions about whether a given pronunciation is incorrect, we can include and analyse all responses.

Pseudowords pronunciation Entropy

Entropy is a concept first introduced by Shannon's Information Theory [13], which can be defined as the degree of chaos within a closed system. Earlier studies in psycholinguistic research used Entropy as a measure to investigate processing difficulty in sentence comprehension [14], quantify orthographic transparency in different orthographies (using word onsets: [15–17], using mono-syllabic words: [18], using whole words: [19]), and to assess variability in responses to disyllabic English pseudowords [20] as well as diversity in vowel pronunciation in German and English children reading aloud pseudowords [12].

In the present study, we use Entropy to calculate the variability of responses to both monosyllabic and multisyllabic pseudowords. This considering, we focus in this study on the following three aspects:

1. Orthographic depth, by investigating orthographies varying in depth (English, German, French, Italian);
2. Age (adults and children) and grade (2, 3, 4, for monolingual German children):
3. Bilingualism (comparing bilingual English-German children, reading German items, with monolingual German children)

Entropy values are calculated as follows: the more alternative pronunciations a given pseudoword has, the bigger its Entropy value is. Since Entropy focuses on the whole pseudoword pronunciation, Entropy values are not affected by the readers' strategy to retrieve sound from larger (morphemes, bodies) or smaller embedded reading units (letters, graphemes). For each pseudoword, we have the transcription of each participant reading this particular item. Entropy is calculated, for each item, by taking the percentage of each type of response, multiplying it by its logarithm, and summing the resulting value for all possible pronunciations of this item. This process is described in the formula:

$$H_j = -\sum_{i=1}^{N} p(i|j) \cdot \log_2 p(i|j)$$

where $p(i|j)$ refers to the percentage of responses i for item j, where N is the number of different pronunciations provided across the participants. Negative numbers were converted into positive numbers (because the logarithm of a proportion, i.e., a number between 0 and 1, is always negative) for easier interpretability, by multiplying the summed Entropy value for each item j by -1. An example of how Entropy is calculated for a specific item can be found in Table 1.

When participants provided the same pronunciation for a given pseudoword, the Entropy value of that item was zero, because log1 = 0. Higher Entropy values ($H > 0$) instead resulted

Table 1. How to calculate Entropy from participants pronunciations for the pseudoword <wap>.

Pronunciations	Proportion	Proportion * Log
(7) wæp	7/(7 + 9 + 1) = 0.41	0.41 * $\log_2$(0.41) = 0.53
(9) wɔp	9/(7 + 9 + 1) = 0.53	0.53 * $\log_2$(0.53) = 0.49
(1) wælp	1/(7 + 9 + 1) = 0.06	0.06 * $\log_2$(0.06) = 0.24
		Entropy
		1.26

Note: In the Proportion * Log column we multiplied the numbers by -1 for easier interpretability

https://doi.org/10.1371/journal.pone.0251629.t001

from participants giving different pronunciations, and as the distribution of multiple pronunciations approaches equiprobability. This formula allows us to focus on item-level differences, that is, to calculate Entropy per item, while for subject-level performances, we average across participants.

To summarize, Entropy is defined here as the number of different pronunciations that participants give to the same pseudoword (pseudoword pronunciation variability). For example, in a sample of five participants, Participants 1 and 3 could read the pseudoword <wap> as /wæp/; Participant 2, instead, would read the item as /wɔp/, while Participants 4 and 5 would agree on a yet different pronunciation: /wælp/. These different choices would increase the Entropy value associated with the pseudoword <wap>, calculated as seen in Table 1. However, the same five participants could agree on the pronunciation of another pseudoword: for example, all of them could read <drell> as /drel/. In this case, the Entropy value of <drell> would be equal to zero. As we will discuss below, there are reasons to think that pseudoword pronunciation variability (Entropy) may vary according to Language, Bilingualism and Age.

Orthographic depth

As for orthography, the relationship between letters and sounds can affect Entropy. The closeness of this relationship is referred to as orthographic depth, and is traditionally described as a continuum [21]. For example, on the shallow end of the continuum are orthographies like Finnish or Italian, where one letter typically corresponds to one sound (i.e, <i> only maps to /i/), while on the deep end are orthographies with a high degree of inconsistency between its letters and sounds (i.e. in the word "gist" <g> is read /dʒ/, but the grapheme itself could be read /g/ as well), like English [22].

Shallow orthographies are easier to read and learn [21, 23–26] because of the straightforward mapping between graphemes and phonemes. Italian and German, for instance, are considered to have shallow orthographies [21], therefore we expected that the Entropy value of pseudowords read by our Italian and German participants will be very low, because the consistent correspondences between graphemes are phonemes will lead to none or very few possible alternative pronunciations (e.g., in Italian, <fulm> can be only read /fulm/ because all the letters in that pseudoword have only one phoneme corresponding to them, leading to only one possible pronunciation). Consequently, since the pseudowords do not have many different pronunciations, their Entropy was also expected to be low.

On the opposite end of the continuum are deep orthographies like English. Children learning to read in deep orthographies have been found to take longer to learn the correspondences between letters and sounds, because of their inconsistent and unpredictable relationships (the same grapheme <i>, found in words like <kit> and <pint> will be read /ɪ/ in the first case and /aɪ/ in the second). As a result, it takes longer to acquire the ability to read accurately [21, 23–25]. We expect that the pseudoword Entropy value for English-speaking children and adults will be the highest, because letters are normally associated with more than one sound, leading to multiple alternative pronunciations (for the pseudoword <sind> can be read /s nd/ or /sa nd/). For this reason, we predicted higher respoonse variability in English-speaking children than in adults (because of their scarcer knowledge of GPCs); and higher response variability in English-speaking participants than in French-, Italian- and German-speaking children, who are learning to associate graphemes to phonemes in more consistent and transparent orthographies.

Complexity and unpredictability

More recent work suggests that orthographic depth should not be seen as a single continuum, but rather as a multidimensional space [27–29]. Even within Europe, orthographies differ on many aspects which are difficult to condense into a single construct. While inconsistency of the print-to-speech correspondences has always been central to the concept of orthographic depth, the study from Schmalz, Marinus, Coltheart and Castles [28] showed that, across orthographies, inconsistency can result either from "complexity" or "unpredictability" which, according to models of reading, should have differential effects on cognitive processes underlying reading and reading acquisition.

Complexity, on the one hand, can lead to inconsistency on the level of letters or graphemes due to the presence of multiletter-correspondences (<aw> → /ɔ:/; this is a complex correspondence because the reading of the individual letters will not give the exact pronunciation), or due to the presence of context-sensitive correspondences (<g[i]> -> /ʤ/; <g[a]> → /g/) or from both (<ch[r]> → /k/; <ch[i]> -> /tʃ/). The French word "ciseaux", for example, contains three complex correspondences: the context-sensitive rule dictates that <c[i]> is read /s/ , while the multiletter grapheme <au> corresponds to /o/ and a position correspondence dictates that the plural morpheme <x> is silent because of its position at the end of the word. Nonetheless, even if there are three different context correspondences, the pronunciation is entirely predictable. Unpredictability, on the other hand, refers to the degree to which the reading system is capable of correctly translating written words into their phonological equivalents [28]. The pronunciation of the word "yacht", for example, is unpredictable, because this word cannot be read correctly without the reader having encountered it before.

Within languages, complexity and unpredictability are correlated. This makes it difficult to dissociate between them. For example, in the English orthography it can be hard to dissociate complexity from unpredictability, as for example in the words "range" and "flange". English phonotactics correspondences state that if an <a> is to be found before the ending <nge> then it should be read as /eɪ/, as in "range" (/reɪnʤ/). However, "flange" is not read /fleɪnʤ/, but /flænʤ/. In this case there is a grapheme which is read differently while being in the same context: in "range" a complex correspondence is applied (a + nge), while in "flange" a simple grapheme-phoneme correspondence is used (<a> is read /æ/). Thus, complex context-sensitive correspondence alone cannot predict how we should read <a>, and readers are often unsure about which strategy is to be applied (context-sensitive or simple GPCs?). Instances like the case we described are not rare, and they make English orthography both highly complex and unpredictable.

The French orthography, on the contrary, is high in complexity, but low in unpredictability. On the one hand. it presents many complex correspondences, caused by multiletter and context-sensitive graphemes (respectively like <au> and <c>). On the other hand, these correspondences are mostly predictable (<au> will be always only read as /o/, while <c> will always be read /s/ before <i, e> and /k/ before <a, o, u>).

Considering the relation between complexity and unpredictability, in the current study we will look at languages that are simple and predictable (Italian and, to a lesser degree, German), complex and predictable (French) and complex and unpredictable (English), in order to investigate the possibility that these features may differentially affect Entropy.

Bilingualism

Another factor that may influence pseudowords pronunciation Entropy is bilingualism. Two scenarios are possible: when told to read pseudowords in Language A, individuals could show interference from Language B, by associating phonemes of Language B to graphemes of

Language A. For example, English/German bilingual may read a German pseudoword like "moch" as /motʃ/ instead of /moχ/, because the grapheme <ch> is read differently in English. Similarly Treiman, Kessler and Evans [30] found interferences from French to English- <c>and<g>pronunciation in English-speaking students who just started learning French. Thus, a grapheme-phoneme correspondence from Language B that interferes with reading Language A, may increase Entropy for bilingual individuals compared to monolingual individuals.

The second scenario goes in the opposite direction. Studies have shown that bilingualism improves metalinguistic awareness, that is the ability "to think about and reflect upon the nature and functions of language" [31]. Metalinguistic awareness refers to different aspects of language, as for example word awareness and phonological awareness. Moreover, results from Yelland, Pollard and Mercuri [43] show that this improved metalinguistic awareness in bilingual children also enhances reading skills, at least in regards to word recognition. Consequently, there are reasons to believe that bilingual children's metalinguistic awareness could improve the overall understanding and sensitivity to GPCs, especially if one of the languages is more transparent than the other. For example, the prior learning of one consistent orthography could help understand the mechanisms underlying the GPCs in the other language, because children already have experience with the dynamics of associating letters to sounds, thus producing a facilitatory effect on the other language.

Aim and hypothesis

Our study's goal was to evaluate the use of Entropy (H) in participants' pseudoword reading aloud responses. Although Entropy has already been used to measure the diversity of vowel pronunciations in German and English children reading aloud pseudowords across grades [12], alternative pronunciations of disyllabic pseudowords in English [20], we are the first, to our knowledge, to use it to compare individual responses to both mono-syllabic and multi-syllabic pseudowords across age (primary school children and adults) and languages (shallow and deep orthographies), including a consideration for bilingualism (in children).

In Experiment 1, we re-analyse novel and published pseudoword reading aloud data from different languages (Italian, German and English) which are on different points along the orthographic depth continuum. In Experiment 2, 3 and 4 we report new data from different age groups. According to the Orthographic Depth Hypothesis [32], we expect that readers of shallow orthographies (like Italian, and, to a lesser degree, German) will be associated overall with low Entropy values, because the very predictable and consistent GPC of their orthography should prevent the possibility of many different alternative pronunciations for pseudowords.

Readers of deep orthographies (like English) will be more likely to be associated with higher Entropy values: this is because in deep orthographies different phonemes can be assigned to one grapheme, which translates to the higher probability that the same pseudoword will be read differently, depending on which phonemes the individual will decide to assign to the graphemes contained in the given pseudoword. A second prediction concerns age.

Adults, as well as children from different grades (2, 3 and 4), participated in this study. We expect that overall children would show a greater variability in responses in all language groups compared to adults (exception made for Italians, for which we only have data from children), because their reading skills development is still on-going, that is, their knowledge of graphemes-phonemes mapping is still incomplete. Hence, children may assign a greater number of phonemes to a given grapheme, because of a greater uncertainty regarding GPCs. A direct comparison will be made among monolingual German children in grade 2, 3 and 4 to investigate whether younger children show greater response variability in responses compared to older children. Overall, we expected that grade 2 children's responses to show higher Entropy values compared to grade 3 and grade 4 children, and grade 3 children to show higher Entropy values compared to grade 4 children.

With respect to bilingualism, as discussed earlier in the introduction, we believe that two outcomes may be possible: If it is true that grapheme-phoneme correspondences from one language interfere with the reading of the other language, we would expect that higher Entropy values will be reported in bilingual children's responses. However, if it is true that enhanced metalinguistic awareness in bilinguals lead to enhanced reading skills compared to monolinguals, we would expect that, on the contrary, bilingual children responses will be associated with lower Entropy values compared to monolinguals.

Experiment 1: Entropy in German and English adults reading matched pseudowords

In the first experiment, we re-analysed pseudoword reading aloud data from a previously published study [33]. This study aimed to compare the nature of sublexical processing in English and German. The items were chosen such that they were matched on orthographic characteristics, such as the number of letters and orthographic neighborhood. In the published study, only RT data were analysed. Here, we are extending the published data by providing new insights into the role of Entropy on pseudoword reading in German and English.

Methods

Participants. German (n = 19) and Australian (n = 48) adults participated in this study. All were staff or students at universities in Germany and Australia, respectively, and received course credit or a small monetary compensation for their participation. The procedure was approved by the ethics committees of both Macquarie's University, Australia (Macquarie University Faculty of Human Sciences (FHS) Ethics Committee) and Ludwig-Maximilian University, Germany (Ethikkommission bei der Medizinischen Fakultät der LMU München).

Materials. Participants read aloud pseudowords in their respective language, which were chosen in respect to the size of their body-neighborhood (see [34]). The size of the body-neighborhood (body-N) for all items was measured thanks to the CELEX database, which is available for both German and English. In the original experiment, participants read aloud both words and pseudowords (in their respective languages) varying in body-N while being matched across body-N condition on length and orthographic neighborhood. Here, we analyse only the pseudoword data. The pseudowords were monosyllabic and matched on the number of letters and orthographic neighborhood [35], as well as on body-neighborhood [34]. Moreover, all items had consistent bodies (i.e., while the number of body-neighbors was manipulated, all body-neighbors had the same pronunciation). Altogether, there were 90 English and 90 German pseudowords, half of which contain high-frequency bodies and the other half contain low-frequency bodies.

Procedure. Each participant was tested individually in a dimly lit, sound-proof testing booth. Each item was shown on the screen for 5 seconds or until the voicekey was triggered, in random order. The items were presented, one at a time, using the software DMDX [55], which created audio recordings for each participants and each item. Here, we analyse only the pseudoword reading aloud responses. A native speaker of each language transcribed the participants' responses from the audiofiles previously recorded and a scorers who had received training in the phonology of the respective language scored the pronunciation accuracy. Both scorers were told to follow a lenient marking criterion, that is, all legally possible grapheme-phoneme relations (including context-inappropriate relations) were considered correct [23, 36, 37]. We then calculated the Entropy, for each pseudoword, using the formula described in the introduction and analysed the data using the statistical environment software R [38]. Afterwards, as an additional analysis, we accounted for non-plausible pronunciations and random noise (meaningless misreadings, such as "dolt" read as /bolt/) by calculating Levenshtein distance [39] from the most common reading to a given pseudowords and all other alternative readings. We did a normalization of the distances obtained (by dividing the distance by the number of phonemes) so that it could be compared one to another. Since our shorter items counted three letters, we decided to exclude all pronunciations whose Levenshtein distance was higher than 0.334. With the resulting, diminished datasets, we then re-calculated Entropy and statistical tests (this re-analysis will be referred from now on as "pronunciation plausibility analysis"). The Python scripts which we used to calculate the Entropy values, as well as supplementary files, can be found here: https://osf.io/94wjx/.

Results and discussion

Non-responses (1 trial from the German data, 6 trials from the English data) were excluded before calculating the Entropy. For German, the median of the Entropy value, across all items, was 0.48 (min = 0, max = 2.21), and for English, the median Entropy was 0.39 (min = 0, max = 1.96).

As the Entropy measure is still relatively new to the field of pseudoword reading, the first question we asked was whether Entropy for each item depends on random or systematic factors. As the English sample was larger than the German sample, we randomly split the English sample 25 times into two groups of 24 participants each, and calculated the item-level Entropy for each item for the two different sub-samples. The mean of the correlations between the fifty sub-samples was 0.89, with a standard deviation of 0.02. All of the correlations were significant $r(90) = p < 0.001$.

The second question was if and how Entropy correlated with accuracy. Two scorers scored English pronunciation accuracy, while one scorer scored German pronunciation accuracy. We then calculated a correlation matrix between Entropy, accuracy, number of answers and percentage of the most common responses for both groups. Table 2 shows the results for English speaking participants, while Table 3 shows the results for German speaking participants. The agreement between scorers was calculated with Cohen's kapp to measure inter-rater reliability [40]. Results show that, for the English data, the scorers were in a moderate agreement ($k = 0.57$).

Entropy was weakly correlated with accuracy, in a significant fashion for scorer 2: $r = 0.26$, $p < 0.05$ but not for scorer 1: $r = 0.04$, $p = 0.70$. This result was unexpected: Entropy was

Table 2. Intercorrelations for English-speaking participants (Exp 1).

Measure	1	2	3	4	5
1. Entropy	-	.26*	.04	.73*	-.86*
2. acc_s2	.26*	-	.29*	.17	-.21*
3. acc_s1	.04	.29*	-	.02	.03
4. n_sw	.73*	.17	.02	-	-.75*
5. perc	-.86*	-.21*	.03	-.75*	-
n = 90					

Note: n_asw = number of different pronunciations per pseudowords, acc_s1 and acc_s2 = accuracy scored by scorer 1 and 2, perc = percentage of the most common response,
* = significant result.

https://doi.org/10.1371/journal.pone.0251629.t002

Table 3. Intercorrelations for German-speaking participants (Exp 1).

Measure	1	2	3	4
1. Entropy	-	-.34*	.92*	-.94*
2. acc	-.34*	-	-.47*	.20
3. n_asw	.92*	-.47*	-	-.76*
4. perc	-.94*	.20	-.76*	-
n = 90				

Note: n_asw = number of different pronunciations per pseudowords, acc = scored accuracy, perc = percentage of the most common response,
* = significant result.

https://doi.org/10.1371/journal.pone.0251629.t003

expecteed to be correlated negatively with accuracy, because it was calculated based on the number of pronunciations. This means that scorers were more likely to accept several alternative pronunciations as correct for English than for German, with the latter showing a negative correlation ($r = -0.34$, $p < 0.05$).

As expected, we found a significant positive correlation with the number of pronunciations per English pseudowords: $r = 0.73$, $p < 0.001$, showing that items with a high Entropy received more different pronunciations than items with a low Entropy, and a significant negative correlation with the percentage of the most common pronunciation ($r = -0.86$, $p < 0.001$). In German participants, Entropy negatively correlated with the accuracy scoring ($r = -0.34$, $p < 0.05$). This is more in line with what we would expect: as accuracy is high, Entropy is naturally low. However, since we could not recruit a second scorer for the German data, the reliability of this correlation remains to be seen. For the other measures, Entropy correlated positively, with the number of pronunciations ($r = 0.92$, $p < 0.001$) and negatively with the percentage of the most common response ($r = -0.94$, $p < 0.001$).

The third, theoretically relevant question, was whether or not the observed Entropy differed between the English and German readers. To visualise the distribution of the Entropy values, we generated a density plot of the English and German Entropy values (see Fig 1). Fig 1 shows that the distribution is right-skewed, with many items having an Entropy value close to zero. Therefore, we performed a Mann-Whitney test, with language as a predictor of Entropy. The difference in Entropy between English and German was not significant, $W = 3710$, $p = 0.33$,

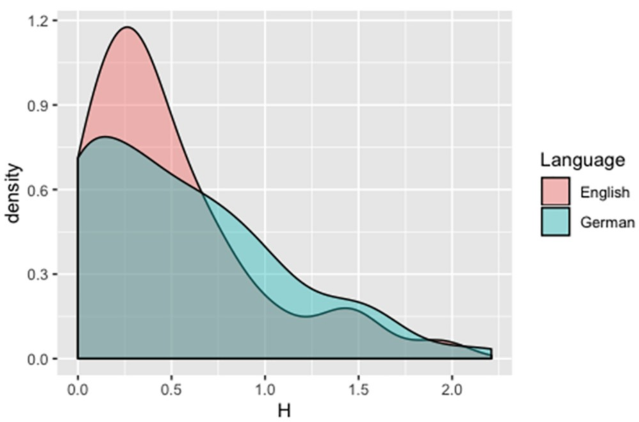


Fig 1. Distribution of Entropy values for German and English adults.

https://doi.org/10.1371/journal.pone.0251629.g001

95%CI = [−0.15, 0.10]. The pronunciation plausibility analysis confirmed the non significance of the original analysis: $W = 3689$, $p = 0.29$, 95%CI = [−0.15, 0.09].

Tables 2 and 3 in S1 Appendix show the participants pronunciations to the ten items with the highest Entropy values. Participants mistakenly read some pseudowords as real words, but there was no significant difference in number of real words pronunciation between German ($m = 0.05$, $sd = 0.22$) and English adults($m = 0.02$, $sd = 0.16$): $p = 0.18$. A list can be found in the Table 1 in S1 Appendix.

Both in English and in German, we found a non-normal distribution of Entropy values, with many Entropy values being close to zero (suggesting consistent pronunciations across participants). Thus, even in the English orthography, despite a number of items which result in a high degree of variability of responses, there is often a consensus about how to pronounce a given item (see also [20] for a similar conclusion). Mousikou, Sadat, Lucas and Rastle [20] argue that this agreement in English pseudowords pronunciation, despite the inconsistency of its orthography, can be explained by the influence that a pseudoword's orthographic neighbors have on its pronunciation (for example *key* could interfere with the pronunciation of *kuy*), and by the fact that, even if a grapheme maps into several phonemes (<i> can be read as /ai/, /ɪ/ or /ɜ:/), participants will tend to pronounce it with the phoneme that is most frequently associated with it. For example, participants read the pseudoword "dize" mostly as /daiz/ (14 participants) and less likely as /dɪze/ (5 participants).

In German, the analysis of the ten items with the highest Entropy values revealed that there were few phonotactic properties that were not systematically applied to pseudowords. For example, the final consonant devoicing phenomenon, which normally makes the voiced final consonant voiceless in words (*Rad—bike* being read as */rat/*) was not always applied: the pseudoword *gund* was read only half of the time *gunt*. Two context correspondences also triggered higher Entropy values: the first concerns the pronunciation of the grapheme <s> in front of the grapheme <p>. Normally, in words like *Sport*, the <s> would be read as /ʃ/. However, in our data, participants read pseuwords like *sprau* either /ʃprau/ or /sprau/. Similarly, the grapheme <n> before the final grapheme <g> should give the phoneme /ɳ/, but participants productions in pseudowords like *quang* varied from /ɳ/, /ɳg/ to final /n/.

The present cross-linguistic comparison did not reveal differences in Entropy between English and German. Previous studies have found differences in accuracy as a function of orthographic depth (e.g., [21]). Since a low accuracy should be evident with high Entropy, we expected to find higher Entropy values in English compared to German. However, most previous reading aloud studies were conducted with children [23–25]. Adult studies have often used lenient marking criteria, and accuracy tends to reach ceiling. Thus, there is little evidence to suggest that cross-linguistic differences in accuracy or pronunciation variability persist into adulthood. The current analysis overcomes this limitation by using Entropy instead of a lenient marking criterion and suggests that, in adulthood, orthographic depth has a minimal influence on the heterogeneity of pseudoword reading aloud responses.

Experiment 2: Entropy in German monolingual children and German/English bilingual children

The aim of the second experiment was to test whether there were differences in Entropy in a younger population: that is, in primary school children. Although the results of Experiment 1 demonstrate that the Entropy of pseudoword reading aloud responses did not differ across German and English-speaking adults, this does not rule out that Entropy differences may exist between German and English-speaking primary school children who are still in the process of learning to read. Entropy differences in adults may be washed out by the fact that the skilled reading system has already established an optimal prediction system for letter-sound correspondences, which may not yet have developed to the same level of precision in developing readers. Experiment 2 put this hypothesis to test by acquiring data from monolingual German children and German/English bilingual children in grades 2, 3, and 4 reading matched pseudowords both in German and in English. This allowed us to compare Entropy within the same items and participants across grade (in German monolingual children) and across orthographies within the same participants.

Overall, we predicted higher Entropy in younger than in older children, because the knowledge of the GPCs may not be full developed, which could lead to a greater level of noisiness in their decision about how to pronounce a given GPC [12]. Moreover, Entropy was expected to be higher for the English than German items, because the depth of English may make it more difficult for children to learn the GPCs. Such a finding would be in line with previous studies, suggesting that pseudoword reading aloud accuracy is lower in English than in shallower orthographies (e.g., [21]). Finally, we hypothesised that Entropy may be higher in bilingual children than monolingual children, because the knowledge of GPCs within one language may interfere with the pseudowords reading aloud responses in the other language [30].

Methods

Participants. Six groups of children participated in this experiment: Three groups of monolingual German children, enrolled in in grade 2 (N = 22), grade 3 (N = 19), and grade 4 (N = 22) (for a more detailed description of this sample, see [12]) were recruited in German primary schools, as well as, three groups of German/English bilingual children attending grade 2 (N = 12), grade 3 (N = 5), and grade 4 (N = 5) of a bilingual primary school in Australia. Prior to testing informed consent was obtained from children's parents. The data reported here were not analysed or reported in Schmalz et al (2020) study. Participants' German proficiency of both bilingual and monolingual children was tested with the standardised reading test SLRT (Salzburg Reading and Spelling Test [41]). The median percentile for monolingual children was 50.50 (min = 8, max = 94;sd = 28.36) and for bilingual children 45.50 (min = 5, max = 88;sd = 21.87). A t-test comparing Monolingual German proficiency and Bilingual German proficiency in grade 2 (the comparison between bilingual and monolingual participants is done for grade 2 children only) revealed no significant difference between the two groups: $t = 0.73$, $p = 0.46$.

Materials. The same items as in Experiment 1 were used.

Procedure. The experimental procedure, as well as the transcription of the audio files and calculation of Entropy for each pseudoword, was identical to Experiment 1. The German monolingual children read the German pseudowords, and the bilingual children read both German and English pseudowords. The bilingual children were tested on separate days. On one day, to avoid any external, cross-linguistic influences on the children's reading behavior, the experimenter spoke only German to them and they performed a number of additional German reading tasks, and on the other day, the experimenter spoke only English and they performed a number of additional English reading tasks (which are not reported here). The order of session was counterbalanced across participants, so half the children started with the German session and the other half of the children started with the English session.

Results and discussion

We excluded non-responses before calculating Entropy. This resulted in a loss of 179 trials (3.16% of all trials) for the monolingual sample, and 77 trials (3.10% of all trials) for the

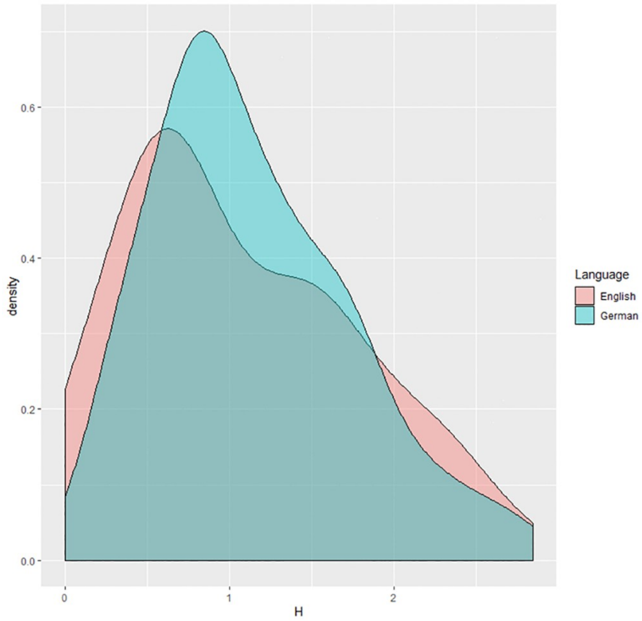


Fig 2. Distribution of Entropy for German/English bilingual children reading German vs English items. The dashed lines are the medians for each orthography.

https://doi.org/10.1371/journal.pone.0251629.g002

bilingual sample. The data from this experiment were used to compare Entropy across three dimensions: language, bilingualism and age.

Language. Firstly, we used language as a predictor (bilingual children reading German vs. the same bilingual children reading English) and performed a Mann-Whitney test: there was no significant difference in Entropy between languages: $W = 3660$, $p = 0.26$, 95%$CI = [-0.28, 0.10]$. The pronunciation plausibility re-analysis confirmed the results from the original analysis $W = 4238$, $p = 0.59$, 95%$CI = [-0.10, 0.26]$. In German, the median of the Entropy value, across all items, was 0.99 (min = 0.27, max = 2.77), whereas in English it was 0.91 (min = 0, max = 2.83). See Fig 2 for the distribution of Entropy.

Tables 4–6 in S1 Appendix show the pronunciations of bilingual children reading the ten English-like pseudowords with the highest Entropy values, while Tables 7–9 in S1 Appendix show the same children reading the ten German-like pseudowords.

A direct observation of children's responses to German and English items showed that both adults and children produced the same alternative pronunciations to certain units. For example, in German, both groups were not uniform regarding the final consonant devoicing phenomenon that is, on the contrary, systematically applied on words (see Tables 2, 6–12 in S1 Appendix). The pseudoword "fold" was read either /folt/ or /fold/. Similarly, when in English pseudowords the letter <r> was preceded by a vowel, both adults and children were divided whether to read it or not. Note that the participants were native Australian speakers: in Australian English, for monosyllabic words, vowels followed by the letter *r* always form a multi-letter rule (but not in multisyllabic words:"kangaroo"is read, for example, /kægəru:/). Tables 3, 5–7 in S1 Appendix show such similar instances. As for the number of lexicalization errors, participants did not significantly read German items as real words ($m = 0.035$, $sd = 0.89$) more than English items ($m = 0.038$, $sd = 0.19$): *p-value* = 0.74.

Furthermore, we calculated a correlation matrix for bilingual children reading English items in grade 2, and for bilingual children reading German items in grade 2, similarly to Experiment 1.

Table 4. Intercorrelations for bilingual children reading German items (Exp 2).

Measure	1	2	3	4
1. Entropy	-	-.75*	.94*	-.93*
2. acc	-.75*	-	-.73*	.64*
3. n_asw	.94*	-.73*	-	-.76*
4. perc	-.93*	.64*	-.76*	-
n = 90				

Note: n_asw = number of different pronunciations per pseudowords, acc = scored accuracy, perc = percentage of the most common response,
* = significant result.

https://doi.org/10.1371/journal.pone.0251629.t004

For bilingual children reading English items, we calculated the agreement between the scorers using Cohen's Kappa. In this case scorers were in a strong agreement ($k = 0.70$). Entropy correlated negatively with both accuracy scoring (s2, $r = -0.62$, $p < 0.001$, s1, $r = -0.61$, $p < 0.001$). In fact, higher accuracy means lower Entropy. Naturally we also found significant correlations between Entropy and number of pronunciations ($r = 0.96$, $p < 0.001$) and Entropy and percentage of the most common response ($r = -0.96$, $p < 0.001$). As for the German items, Entropy correlated negatively with accuracy scoring (s1, $r = -0.75$, $p < 0.001$) and with the percentage of the most common response ($r = -0.93$, $p < 0.001$), while positively correlating with number of different pronunciations ($r = 0.94$, $p < 0.001$). Tables 4 and 5 show the correlation matrix.

Bilingualism. Secondly, we investigated whether bilingualism affected Entropy (German monolingual vs German/English bilingual children reading the same items in German). Only participants from grade 2 were included in this comparison, since the number of participants from those groups was rather similar (N = 22 monolingual, N = 13 bilingual). The median of the Entropy value for the bilingual group was 1.14 (min = 0, max = 2.81), while it was 1.32 (min = 0, max = 3.22) for the monolingual group (see Fig 2). We also performed a Mann-Whitney test to see whether there was a difference in Entropy values for the second contrast, but again we found no significant difference: $W = 4144$, $p = 0.79$, $95\%CI = [-0.10, 0.27]$. The pronunciation plausibility analysis confirmed the marginal significance of the result $W = 3452$, $p = 0.08$, $95\%CI = [-0.36, 0.26]$. This result indicates that, although there was a marginally significant difference between the two groups ($p < 0.1$), bilingualism did not increase answer variability (see Fig 3). This is in line with studies which show that learning an orthography that is

Table 5. Intercorrelations for bilingual children reading English items (Exp 2).

Measure	1	2	3	4	5
1. Entropy	-	-.62*	-.61*	.96*	-.96*
2. acc_s2	-.62*	-	.51*	-.59*	.64*
3. acc_s1	-.61*	.51*	-	-.60*	.53*
4. n_asw	.96*	-.59*	-.60*	-	-.89*
5. perc	-.96*	.64*	.53*	-.89*	-
n = 90					

Note: n_asw = number of different pronunciations per pseudowords, acc_s1 and acc_s2 = accuracy scored by scorer 1 and 2, perc = percentage of the most common response,
* = significant result.

https://doi.org/10.1371/journal.pone.0251629.t005

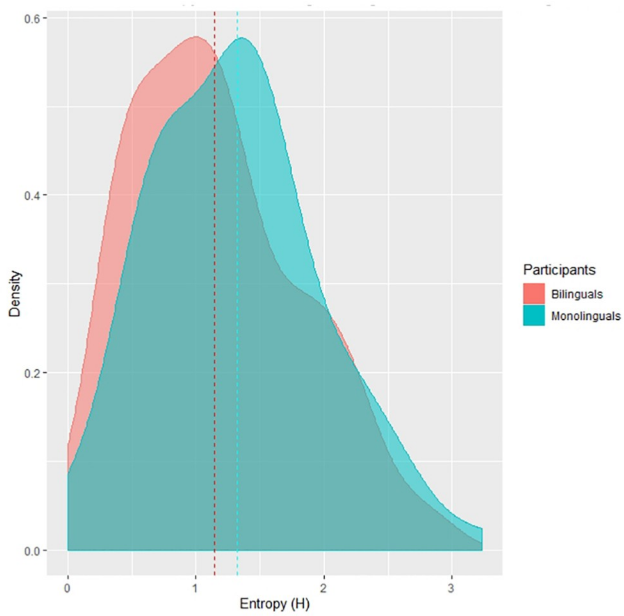


Fig 3. Distribution of Entropy for German/English bilingual and German monolingual children. The dashed lines are the medians for each orthography.

https://doi.org/10.1371/journal.pone.0251629.g003

more transparent than the other, if anything, improves the understanding of the deeper orthography GPCs [42, 43].

The current result diverges to some degree from Treiman, Kessler and Evans [30] who found that exposure to a second language affects graphemes-phonemes correspondences of the first language. In this study, native English speakers learning French applied French fronting context rule while reading<c>and<g>graphemes in English word and pseudowords. Those students took into account the following vowel to determine pronunciation, and so much more than students who were not studying a second language. Translated to Entropy, pronunciation variability in students learning a second language was lower compared to students who did not undertake a second (romance) language class. The difference with our results might be due to the fact that our participants were bilingual English-German (two Germanic languages) children (and not university students), and we could only find limited occurrences of GPC interference from English to German (for example, the German item "loo" was read /lu/ instead of /lo:/). Therefore the findings of the present study and those from Treiman, Kessler & Evans are not in direct contradiction, given the nature of participants (bilingual pupils being proficient in two languages, compared to monolingual English-speaking students, who just started to learn French as a second language) and nature of direction (interference between equally mastered languages, compared to interferences from L2 to L1) even though our result was only marginally significant. In the real word data, there was no significant difference in real words reading between bilinguals ($m = 0.035$, $sd = 0.18$) and monolingual German children reading in German($m = 0.42$, $sd = 0.19$): $p\text{-}value = 0.40$. A list can be found in Table 13 in S1 Appendix.

Grade. Finally we used grade as a predictor of Entropy (we compared German monolingual children from grade 2, 3 and 4 across grades). We performed a Mann-Whitney test (between grades 2 and 3; 2 and 4; 3 and 4) and calculated Entropy medians. In grade 2 the median of Entropy values was of 1.31 (min = 0; max = 3.22), 0.58 in grade 3 (min = 0; max = 1.51) and 0.39 in grade 4 (min = 0; max = 1.25) (Fig 4). The Mann-Whitney test showed

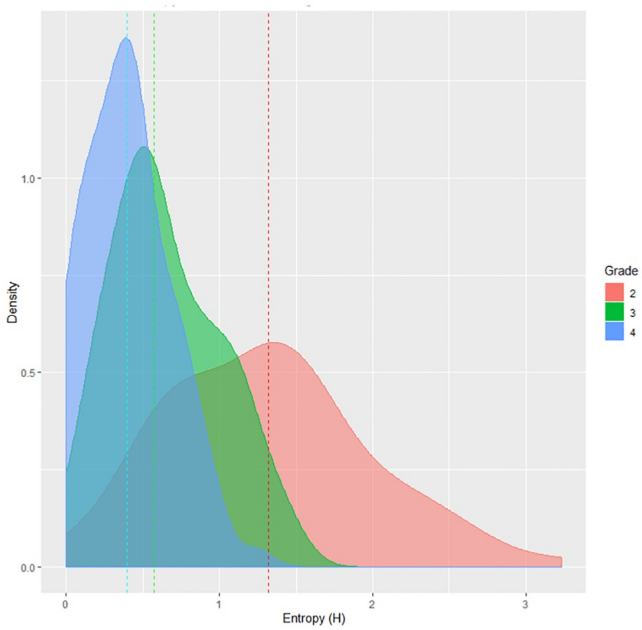


Fig 4. Distribution of Entropy for German monolingual children across grades. The dashed lines are the medians for each orthography.

https://doi.org/10.1371/journal.pone.0251629.g004

significant differences between grade 2 and 3: $p < 0.001$, 95%CI = [0.45, 0.79]; grade 2 and 4: $p < 0.001$, 95%CI = [0.72, 1.23] and grade 3 and 4: $p < 0.001$, 95%CI = [0.14, 0.35]. The pronunciation plausibility analysis confirmed the significance of all comparisons: grade 2 and 3: $p < 0.001$, 95%CI = [−0.57, −0.30]; grade 2 and 4: $p < 0.001$, 95%CI = [−0.79, −0.54] and grade 3 and 4: $p < 0.001$, 95%CI = [0.14, 0.31]. As we can see from the Entropy values medians changing across grades (see Fig 4), by developing and practicing their reading skills children gradually became more acquainted with the GPCs of their language, and their answer variability decreased. This result is in line with the findings of [12], who found decreasing Entropy in vowel pronunciation variability (but not in consonant pronunciation variability, which was not investigated) as a function of grade. In real words, there was no significant difference in lexicalizations between grades(grade 2: m = 0.41, sd = 0.19; grade 3: m = 0.04, sd = 0.21; grade 4: m = 0.03, sd = 0.17): p = 0.594 for the comparison between grade 2 and grade 3; p = 0.524 between grade 3 and grade 4; and p = 0.274 between grade 2 and grade 4).

We calculated correlations between Entropy and other measures for all grades (Tables 6–8). In grade 2 Entropy correlated significantly with accuracy (r = −0.71, $p < 0.001$), the number of different pronunciations (r = 0.92, $p < 0.001$) and the percentage of the most common response (r = −0.91, $p < 0.001$). In grade 3, correlations remained significant for the number of different pronunciations (r = 0.90, $p < 0.001$) and for the percentage of the most common response (r = −0.93, $p < 0.001$), but the correlation with accuracy was not significant (r = −0.05, p = 0.62). The same scenario from grade 2 repeated in grade 4: Entropy significantly correlated with accuracy (r = −0.70, $p < 0.001$), the number of different pronunciations (r = 0.91, $p < 0.001$) and the percentage to the most common response (r = −0.93, $p < 0.001$).

The pronunciations of the ten items with the highest Entropy values are listed in Tables 10–12 in S1 Appendix, andfor a list of pseudowords read as realwords are listed in Table 14 in S1 Appendix (grade 2: m = 0.04, sd = 0.17; grade 3: m = 0.04, sd = 0.21; grade 4: m = 0.03, sd = 0.17).

Table 6. Monolingual German children in grade 2 (exp 2).

Measure	1	2	3	4
1. Entropy	-	-.71*	.91*	-.91*
2. acc	-.71*	-	-.70*	.58*
3. n_sw	.91*	-.70	-	-.73*
4. perc	-.91*	.58	-.73*	-
n = 90				

Note: n_asw = number of different pronunciations per pseudowords, acc = scored accuracy, perc = percentage of the most common response,
* = significant result.

https://doi.org/10.1371/journal.pone.0251629.t006

Table 7. Monolingual German children in grade 3 (exp 2).

Measure	1	2	3	4
1. Entropy	-	-.05	.90*	-.93*
2. acc	-.05	-	-.13	-.03
3. n_asw	.90*	-.13	-	-.72*
4. perc	-.93*	-.03	-.72*	-
n = 90				

Note: n_asw = number of different pronunciations per pseudowords, acc = scored accuracy, perc = percentage of the most common response,
* = significant result.

https://doi.org/10.1371/journal.pone.0251629.t007

Experiment 3: Entropy in French and Italian children

The cross-linguistic contrast in Experiments 1 and 2 relied on a comparison of German and English pseudowords and did not reveal any cross-linguistic differences in English and German speaking adults and children.

However, since we used a bilingual sample to search for cross-linguistics differences in children, it may be the case that the knowledge of one shallow orthography (German) had a facilitatory effect on the knowledge of the deeper language (English). One possible explanation is that the children's knowledge of two different orthographies enhanced their understanding of GPCs. Many studies on bilingualism, in fact, suggest that bilingual children possess greater

Table 8. Monolingual German children in grade 4 (exp 2).

Measure	1	2	3	4
1. Entropy	-	-.70*	.91*	-.93*
2. acc	-.70*	-	-.79*	.53*
3. n_asw	.91*	-.79*	-	-.71*
4. perc	-.93*	.53*	-.71*	-
n = 90				

Note: n_asw = number of different pronunciations per pseudowords, acc = scored accuracy, perc = percentage of the most common response,
* = significant result.

https://doi.org/10.1371/journal.pone.0251629.t008

metalinguistic awareness [30, 44–46], defined as "the explicit knowledge of the structural components of their orthography" [43].

At the same time, we wanted to test whether complexity, rather than unpredictability affected Entropy. Since English orthography is considered both unpredictable and complex, and could not serve for this purpose, we chose to collect data from two more groups of children, French and Italian fourth graders. By comparing them, we were able to also assess the effect of complexity on Entropy: French, compared to other European orthographies, has many complex correspondences, while Italian has relatively few, with unpredictability being relatively low in both orthographies [28].

Methods

Participants. A group of Italian fourth graders (n = 33) and a group of French fourth graders (n = 29) were recruited for this experiment. Children's parents agreed to the participation by signing an informed consent.

Materials. The children read aloud a list of 40 pseudowords, generated from a list of cognate words (with similar orthograpny and the same meaning in both languages, like "maternité" and "maternità"- maternity). Pseudowords were matched in number of syllables, number of letters, orthographic neighborhood entity and base-word frequency.

Procedure. First, during a preliminary phase, we ensured that no children had learning disorders. One French child who was already diagnosed with dyslexia was excluded. Second, we administered the pseudoword reading aloud task to each participant. The procedure was identical to Experiments 1 and 2.

Results and discussion

Entropy was calculated using the same script and formula of the other experiments. For French speaking children the median of the Entropy value was 0.99 (min = 0, max = 2.53), while for Italian speaking children it was 1.38 (min = 0, max = 3.24). We then performed a Mann-Whitney test between French and Italian items, which showed a significant effect $W = 460$, $p < 0.05$, $95\%CI = [-0.98, 0.22]$, reflecting higher Entropy in Italian than French children (see Fig 5). Once again, the pronunciation plausibility analysis confirmed this result: $W = 468.5$, $p < 0.05$, $95\%CI = [-0.81, 0.19]$.

Cohen's kappa calculation revealed that scorers were in a moderate agreement for French data ($k = 0.57$) and in a nearly perfect agreement for Italian data ($k = 0.92$). The fact that for Italian data the scorers were in a nearly perfect agreement does not come as a surprise: since it is a shallow orthography, and has an almost perfect isometric mapping between graphemes and phonemes, it is easier and more straightforward to determine which pronunciation can be considered correct or wrong.

In the French data we found a significant, negative correlations between accuracy and Entropy (scorer 1 $r = -0.49$, scorer 2 $r = -0.58$): $p < 0.001$. Again, this was expected, as higher accuracy implies lower Entropy. A significant, positive correlation was found with number of pronunciations ($r = 0.75$, $p < 0.001$) and a negative correlation was found with the percentage of the most common response ($r = -0.81$.$p < 0.001$). In Italian, Entropy was significantly correlated with number of different pronunciations ($r = 0.94$, $p < 0.001$) and percentage of most common response ($r = -0.93$, $p < 0.001$), but surprisingly not with the accuracy judgements (s1 $r = 0.08$, $p = 0.63$, s2 $r = 0$, $p = 0.99$).

Tables 9 and 10 show the correlation matrix.

We then analysed the responses to the ten items with the highest Entropy values, in order to qualitatively assess which factors may lead to higher Entropy value (see Tables 16 and 17 in

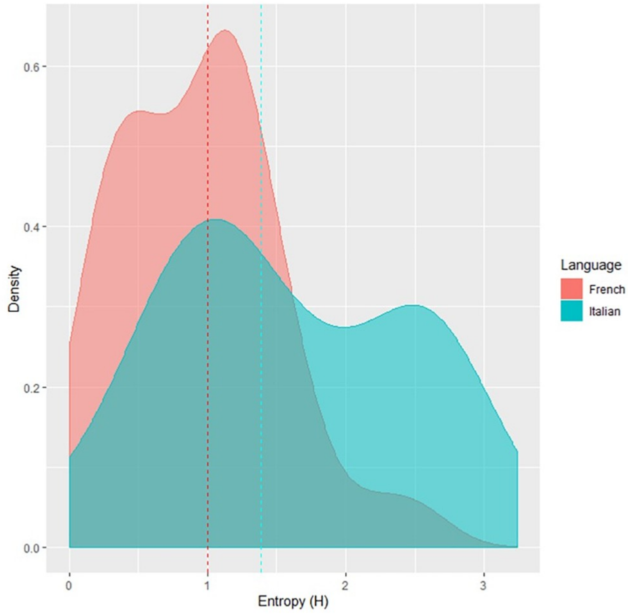


Fig 5. Distribution of Entropy values in French and Italian children. The dashed lines are the medians for each orthography.

https://doi.org/10.1371/journal.pone.0251629.g005

Table 9. Intercorrelations for French children (Exp 3).

Measure	1	2	3	4	5
1. Entropy	-	-.58*	-.49*	.75*	-.81*
2. acc_s2	-.58*	-	.74*	-.48*	.71*
3. acc_s1	-.49*	.74*	-	-.51*	.62*
4. n_asw	.75*	-.48*	-.51*	-	-.63*
5. perc	-.81*	.71*	.62*	-.63*	-
n = 40					

Note: n_asw = number of different pronunciations per pseudowords, acc_s1 and acc_s2 = accuracy scored by scorer 1 and 2, perc = percentage of the most common response,
* = significant result.

https://doi.org/10.1371/journal.pone.0251629.t009

Table 10. Intercorrelations for Italian children (Exp 3).

Measure	1	2	3	4	5
1. Entropy	-	.00	.08	.94*	-.93*
2. acc_s2	.00	-	.97*	.05	-.01
3. acc_s1	.08	.97*	-	.15	-.08
4. n_asw	.94*	.05	.15	-	-.79*
5. perc	-.93*	-.01	-.08	-.79*	-
n = 40					

Note: n_asw = number of different pronunciations per pseudowords, acc_s1 and acc_s2 = accuracy scored by scorer 1 and 2, perc = percentage of the most common response,
* = significant result.

https://doi.org/10.1371/journal.pone.0251629.t010

S1 Appendix). A comparison between Italian and French children revealed that Italian participants misread items as real words more often than French participants ($p < 0.05$). A list of real words readings can be found in Table 15 in S1 Appendix.

The qualitative analysis performed on French children's answers showed that alternative answers were given especially when pseudowords contained inconsistent graphemes (as the sibilant <s>) or nasal sounds (e.g. <am>, <en>, <aim>). As for the <r> grapheme, it has been previously shown that its corresponding phoneme /ʁ/ is challenging for children to acquire, and its acquisition occurs very late [47]. Our results suggest that grade 3 children's GPCs are not fully developed yet. Another element that created alternative readings was the pronunciation of final consonants that are not normally read in real words, such as <t>, <r>, <d> and <s> in pseudowords as *stort, fratis, buffat, antobus, gord and cosputer*. Our participants were very divided regarding this issue, and since we got the same alternative answers from French adults (Exp. 4), we concluded that age reading skills are not possible causes of these answers. These results are consistent with [48] who found that French-speaking participants pronounced letters in nonwords that are typically silent in words.

Italian children's responses were affected by the pseudowords' orthographic neighbors or by the recognition of the base word itself; this is the case for the grapheme <g> read as the phoneme /ʒ/ for the pseudowords "benge" (baseword: "beige") and "darage" (baseword: "garage"). Children who produced this phoneme (which is not in the Italian phoneme inventory) recognised the French loanwords and had knowledge of their irregular reading. In regard to the other occurrences, Italian children did not apply phonotactic cues that normally indicate which phoneme must be pronounced. For example, when <s> and <z> are surrounded by vowels, their voiced alternative (/z/ and /dz/, respectively) should be produced. Therefore, <anisale> should be read as /anizale/ and <vazionalità> as /vadzionalita/. This voicing assimilation phenomenon, which is the norm in the central and northern areas of Italy, is however not common in the southern regions of Italy, and specifically not in Sardinia (the native region of our participants). Moreover, these phonemes are often considered allophones by Italian speakers, depending on their geographical origin. Consequently,the alternative pronunciations of some inconsistent graphemes are not considered wrong or not fitting, and individual responses can vary even within the same participant (who will produce the alternative readings of that grapheme in a non-systematic fashion).

Against our predictions, the median Entropy value was lower in French than Italian children, suggesting that complexity may not increase pseudoword reading aloud Entropy. One explanation for the higher Entropy values in Italian could lie in the characteristics of the items themselves. In order to create a set of cognate items, we chose similar words in Italian and French, and then generate pseudowords by changing letters. Since Italian syllabic structure is simpler than in French [21], it is possible that French children had to read items that were not representative of their structural complexity. For example, "pizza", a common word in both languages that reflects the Italian syllabic structure [CVCV], has a simpler syllabic structure than "fauteuil" [CVCVC], a typical French word, and also fewer diphtongs and inconsistent graphemes). The goal of Experiment 4 was to rule out this possibility by providing sets of items that are truly representative of the participants' languages.

Experiment 4: Entropy in English, French, and German adults reading non-matched pseudowords

The results of Experiments 1-3 did not reveal any cross-linguistic differences that would have suggested that deeper orthographies lead to greater variability in pseudoword pronunciations. However, in these experiments, items were strictly matched on various psycholinguistic properties across orthographies (syllable structure, number of letters, frequency, orthographic neighborhood and so forth). The advantage of this setup is that researchers can control for a number of psycholinguistic properties that can influence participants' reading behavior. However, one disadvantage is that cognate pseudowords may not always be representative of the types of words that the readers typically encounter in their native orthography, and therefore had the potential to reduce variability in participants' responses across languages.

This considered, in the last experiment, we created two different sets of items. For the first set, we created pseudowords that only matched on frequency, and not, for example, on syllable structure or number of letters. We based this design on the idea of a frequency-matched reading aloud study [49]. In this study Ellis et al. argued that matching items on all possible characteristics creates item sets which are unnatural for most of the orthographies. Note that this problem persists in the cognate design: for example, the German/English cognate "Zeitgeist", the spelling is typical, regular and predictable in German, but strange and irregular in English. The reverse is true for the cognate "steak". The solution proposed by Ellis et al. [49] was to allow words to vary across languages on all dimensions except frequency. All words from a corpus are divided into frequency bands, and an equal amount of words is randomly chosen from each frequency band from each language. Word frequency is a measure of the frequency with which participants are expected to have encountered this word. Thus, if frequency is matched across languages, participants' familiarity with a given word is kept constant. All other item-level characteristics vary, but this variation is systematic, as it reflects the orthographies' characteristics. In the current study, we were interested in pseudoword rather than word reading. Therefore, we first chose a series of words, using the frequency-matched design, and then created a set of pseudowords from these words using the same procedure across orthographies. The advantage of this approach is that pseudowords will inherit properties that are characteristic of the orthography, such as length and bigram frequency.

In the second set of items, we took the opposite approach. We created pseudowords which were identical across orthographies, and which were equally untypical of real words in all orthographies in question (orthographic neighborhood of 0). These were pseudwords with a CVCVCV structure, containing only letters which occur frequently in all three orthographies in question.

In Experiments 1-2, furthermore, the items were all monosyllabic. In Experiment 4, we relaxed this constraint. In general, pronunciations become less consistent when polysyllabic words are considered [51]. Therefore, the presence of polysyllabic pseudowords gives more scope for readers of deep orthographies to provide variable pronunciations.

Methods

Participants. Participants were 16 students from universities in southern Germany, 28 students from a university in southern France, and 39 students from a university in Australia. They participated in exchange for course credit or payment.

Materials. As outlined in the previous section, we chose two subsets of items. For the first subset, we selected a number of words from each language using a frequency-matched design, following the same procedure as [49]. We randomly selected words from different frequency bands: 10 words each with a log-frequency between 0 and 0.5, between 0.5 and 1, between 1 and 1.5, and between 1.5 and 2 [51–53]. We then created pseudowords for each item in each language, using the software Wuggy [54]. Wuggy's algorithm generates pseudowords that are similar to the input words in terms of subsyllabic structure, bigram frequency, and orthographic neighborhood. Thus, we obtained 40 pseudowords, based on real words varying in frequency, for each orthography.

In the second set, the items were equally easy to pronounce, we assembled 20 pseudowords from simple CV syllables. Each pseudoword had three syllables, and an orthographic neighborhood of zero. Thus, in all languages, items were equally dissimilar to real words.

Procedure. The two sets of pseudowords were presented to the participants in a mixed random order, using the software DMDX [55]. The participants saw each item on the screen for 2.5 seconds or until the voice key was triggered, and were instructed to read aloud the items as fast as possible, while being as accurate as they could be. The data was then transcribed, for each orthography, by a native speaker.

Results and discussion

We excluded all non-responses (12 trials for English, no trials for French, and 3 trials for German, <1% of the data) before further processing. Entropy was calculated, for each language separately, as in the previous experiments (see Fig 6).

Since Fig 6 showed a non-normal Entropy distribution was compared across languages in a pairwise manner, we used the Mann-Whitney test.

Word-like pseudowords. First, in the comparison of the word-like pseudowords, which were derived from the frequency-matched words, the median Entropy was 1.03 for English (min = 0, max = 3.91), 0.31 for French (min = 0, max = 2.98), and 0.36 for German (min = 0, max = 2.74). The Mann-Whitney tests showed a significant difference between English and French, $W = 1160$, $p < 0.001$, 95%CI = [0.30, 0.88], between English and German, $W = 1177$, $p < 0.001$, 95%CI = [0.22, 0.83], but no significant difference between French and German, $W = 800$, $p > 0.9$, 95%CI[−0.24, 0.23]. Again the results were confirmed by the pronunciation plausibility analysis: comparisons between English and French ($W = 1161.5$, $p < 0.001$, 95%CI = [0.32, 0.73]) and English and German remained significant ($W = 1089$, $p = 0.005$, 95%CI = [0.16, 0.64]), while the non-significance of French and German comparison was corroborated ($W = 705$, $p = 0.34$, 95%CI = [−3.53, 8.76]).

Three multiple linear regressions were calculated including language, length, number of syllables, baseword and bigram frequency, orthographic neighborhood, phonological neighborhood and BodyN. Baseword frequency and Body Neighborhood were calculated using Leipzig Corpora Collection [56–58], while Bigram frequency, orthographic and phonological neighborhood were calculated thanks to the Clearpond database [59].

Results indicated that none of these variables, apart from Language (in the comparison between English and French, and English and German) and Length (in the comparison between English and Grench) were significant predictors in the model (see Tables 11–13), which confirmed the results of the Mann-Whitney tests.

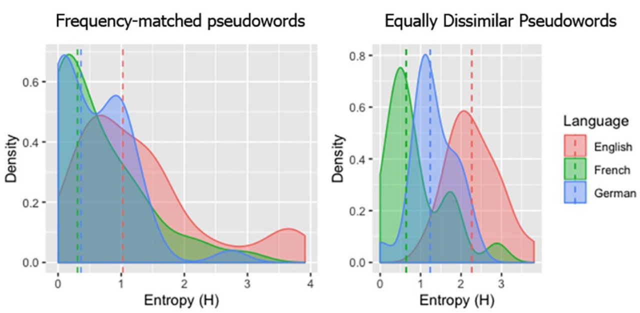


Fig 6. Distribution of Entropy values for French, German and English adults. The dashed lines are the medians for each orthography.

https://doi.org/10.1371/journal.pone.0251629.g006

Table 11. Multiple regression, English-French, word-like pseudoword.

	Model 1	Model 2
(Intercept)	0.39, (0.15)*	0.03, (0.42)
Language	−0.77, (0.21)***	−0.79, (0.24)**
Length		0.28, (0.13)*
Syllables count		0.22, (0.25)
Baseword Frequency		0.07, (0.11)
Orthographic N		−0.00, (0.10)
Phononological N		−0.08, (0.11)
Body N		−0.15, (0.13)
Bigram Frequency		−0.08, (0.10)
R^2	0.15	0.30
Adj. R^2	0.14	0.22
Num. obs.	80	80

	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	0.0345	0.4183	0.08	0.9346
Language	-0.7878	0.2395	-3.29	0.0016**
Length	0.2813	0.1322	2.13	0.0368*
Syllables count	0.2212	0.2530	0.87	0.3849
Baseword Frequency	0.0687	0.1103	0.62	0.5352
Orthographic N	-0.0008	0.1025	-0.01	0.9937
Phononological N	-0.0800	0.1083	-0.74	0.4627
Body N	-0.1469	0.1291	-1.14	0.2589
Bigram Frequency	-0.0785	0.1048	-0.75	0.4562

***$p < 0.001$;

**$p < 0.01$;

*$p < 0.05$

https://doi.org/10.1371/journal.pone.0251629.t011

Since German and French have predictable mappings between graphemes and phonemes and are easy orthographies to read, this result is in line with our expectations. In contrast, English orthography is both complex and unpredictable, a characteristic that led to higher variability in responses (compared to orthographies which are complex, but predictable).

We then calculated agreement between English scorers. Cohen's kappa measure revealed a moderate agreement ($k = 0.54$). Entropy correlated significantly with scorer 1's accuracy judgment (s1 $r = -0.44$, $p < 0.05$), but not with scorer 2's ($r = -0.21$, $p = 0.19$), with number of different pronunciations ($r = 0.95$, $p < 0.001$) and percentage of the most common response ($r = -0.97$, $p < 0.001$) As for the French data, the Cohen's kappa for our scorers was $k = 0.77$, revealing a strong agreement. Entropy correlated significantly with both accuracy judgements (s1 and s2 $r = -0.75$, $p < 0.001$), number of different pronunciations ($r = 0.92$, $p < 0.001$) and percentage of the most common response ($r = -0.97$, $p < 0.001$). Similarly, in German data Entropy correlated with accuracy ($r = -0.58$, $p < 0.001$), number of different answers ($r = -0.77$, $p < 0.001$) and percentage of the most common response ($r = -0.92$, $p < 0.001$). Tables 14–16 show the correlation matrix.

Dissimilar pseudowords. For the equally dissimilar pseudowords, the median Entropy values were 2.27 (min = 1.18, max = 3.80) for English, 0.64 for French (min = 0, max = 2.89), and 1.23 (min = 0, max = 2.29) for German. The Mann-Whitney tests showed a significant difference between English and French, $W = 368$, $p < 0.001$, 95%CI[1.11, 1.92], between English and German, $W = 356$, $p < 0.001$, 95%CI[0.58, 1.34], and between French and German,

Table 12. Multiple regression, English-German, word-like pseudoword.

	Model 1		Model 2	
(Intercept)	−0.40(0.15)**		−0.86(0.40)*	
Language	0.79(0.21)***		0.73(0.23)**	
Length			0.18(0.14)	
Syllables count			0.34(0.26)	
Baseword Frequency			−0.01(0.11)	
Orthographic N			−0.04(0.11)	
Phonological N			0.03(0.12)	
Body N			−0.16(0.12)	
Bigram Frequency			−0.10(0.11)	
R^2	0.16		0.26	
Adj. R^2	0.15		0.17	
Num. obs.	80		80	
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.8563	0.4048	-2.12	0.0379*
Language	0.7289	0.2316	3.15	0.0024**
Length	0.1756	0.1373	1.28	0.2051
Syllables count	0.3392	0.2578	1.32	0.1924
Frequency	-0.0116	0.1133	-0.10	0.9189
Orthographic N	-0.0414	0.1051	-0.39	0.6947
Phonological N	0.0319	0.1219	0.26	0.7943
Body N	-0.1575	0.1236	-1.27	0.2068
Bigram Frequency	-0.0993	0.1053	-0.94	0.3489

***$p < 0.001$;
**$p < 0.01$;
*$p < 0.05$

https://doi.org/10.1371/journal.pone.0251629.t012

$W = 102$, $p = 0.008$, 95%$CI[-0.91, -0.21]$ (the pronunciation plausibility analysis confirmed the significance of all comparisons. Between English and French: $W = 324$, $p < 0.001$, 95%CI [0.57, 1.43], English and German: $W = 301.5$, $p = 0.006$, 95%CI[0.22, 0.88] and French and German: $W = 115$, $p = 0.02$, 95%$CI[-0.83, -0.05]$). These findings indicate that reading aloud Entropy was higher in English than in either French or German, and higher in German compared to French, which is in contrast with the results of the previous experiment (please refer to the General Discussion, where we discuss this point in more detail).

As in our previous experiments, the observation of the participants' responses to pseudowords confirmed that high Entropy values were associated with those items that contained non-consistent graphemes as <s> or <z>, context or position correspondences (like terminal devoicing in German or silent final consonants in French), different vowel lengths (especially in German) and different kind of phoneme manipulations (especially in English, like syllable manipulations [zulumu -> zumulu]). These phenomena can be seen in Tables 19–24 in S1 Appendix.

We then performed Cohen's kappa between our scorers and correlations matrix. In English data, scorers were in a strong agreement ($k = 0.74$). Entropy correlated significantly with number of different pronunciations ($r = 0.86$, $p < 0.001$), percentage of the most common response ($r = 0.85$, $p < 0.001$) and scorer 1's accuracy judgement ($r = -0.55$, $p < 0.05$), but not scorer 2's ($r = -0.35$, $p = 0.13$).

Table 13. Multiple regression, French-German, word-like pseudowords.

	Model 1		Model 2	
(Intercept)	0.01(0.16)		−0.11(0.47)	
Language	−0.01(0.23)		−0.08(0.28)	
Length			−0.04(0.15)	
Syllables count			0.09(0.29)	
Baseword Frequency			−0.01(0.13)	
Orthographic N			−0.07(0.12)	
Phonological N			−0.09(0.13)	
Body N			−0.24(0.14)	
Bigram Frequency			−0.04(0.14)	
R^2	0.00		0.07	
Adj. R^2	−0.01		−0.03	
Num. obs.	80		80	
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.1128	0.4660	-0.24	0.8093
Language	-0.0764	0.2780	-0.28	0.7841
Length	-0.0421	0.1541	-0.27	0.7856
Syllables count	0.0916	0.2885	0.32	0.7519
Baseword Frequency	-0.0069	0.1280	-0.05	0.9571
Orthographic N	-0.0665	0.1183	-0.56	0.5758
Phonological N	-0.0949	0.1306	-0.73	0.4698
Body N	-0.2436	0.1444	-1.69	0.0960
Bigram Frequency	-0.0442	0.1386	-0.32	0.7507

$^{***}p < 0.001$;
$^{**}p < 0.01$;
$^{*}p < 0.05$

https://doi.org/10.1371/journal.pone.0251629.t013

In French, our scorers were in a moderate agreement ($k = 0.45$). Entropy correlated significantly with both scorers' accuracy judgements (s1 $r = -0.87$, $p < 0.001$; s2 $r = -0.76$, $p < 0.001$), number of different pronunciations ($r = 0.92$, $p < 0.001$), and percentage of the most common response ($r = -0.85$, $p < 0.001$).

Table 14. Intercorrelations for English-speaking adults reading word-like pseudowords (Exp 4).

Measure	1	2	3	4	5
1. Entropy	-	-.21	-.44*	.95*	-.97*
2. acc_s2	-.21	-	.44*	-.21	.19
3. acc_s1	-.44*	.44*	-	-.37*	.45*
4. n_asw	.95*	-.21	-.37*	-	-.86*
5. perc	-.97*	.19	.45*	-.86*	-
n = 40					

Note: n_asw = number of different pronunciations per pseudowords, acc_s1 and acc_s2 = accuracy scored by scorer 1 and 2, perc = percentage of the most common response,
* = significant result.

https://doi.org/10.1371/journal.pone.0251629.t014

Table 15. Intercorrelations for French adults reading word-like pseudowords (Exp 4).

Measure	1	2	3	4	5
1. Entropy	-	-.75*	-.75*	.92*	-.97*
2. acc_s2	-.75*	-	.75*	-.62*	.80*
3. acc_s1	-.75*	.75*	-	-.61*	.82*
4. n_asw	.92*	-.62*	-.61*	-	-.86*
5. perc	-.97*	.80*	.82*	-.86*	-
n = 40					

Note: n_asw = number of different pronunciations per pseudowords, acc_s1 and acc_s2 = accuracy scored by scorer 1 and 2, perc = percentage of the most common response,
* = significant result.

https://doi.org/10.1371/journal.pone.0251629.t015

In German, Entropy correlated significantly with number of different pronunciations ($r = 0.89$, $p < 0.001$), percentage of the most common response ($r = -0.92$, $p < 0.001$) but not with accuracy judgement ($r = 0.08$, $p = 0.74$). Intercorrelations can be seen in Tables 17–19.

An analysis of real word misreadings revealed no significance difference between the three groups (p = 0.10 for the comparison between English and German, p = 0.12 for the comparison between English and French; p = 0.96 for the comparison between German and French). A list of pseudowords read as real words can be seen in Table 18 in S1 Appendix.

Table 16. Intercorrelations for German adults reading word-like pseudowords (Exp 4).

Measure	1	2	3	4
1. Entropy	-	-.58*	.77*	-.92*
2. acc	-.58*	-	-.81*	.38*
3. n_asw	.77*	-.81*	-	-.58*
4. perc	-.92*	.38*	-.58*	-
n = 40				

Note: n_asw = number of different pronunciations per pseudowords, acc = scored accuracy, perc = percentage of the most common response, * = significant result,
* = significant result.

https://doi.org/10.1371/journal.pone.0251629.t016

Table 17. Intercorrelations for English adults reading dissimilar pseudowords (Exp 4).

Measure	1	2	3	4	5
1. Entropy	-	-.34	-.55*	.86*	-.85*
2. acc_s2	-.34	-	.54*	-.55*	.12
3. acc_s1	-.55*	.54*	-	-.75*	.22
4. n_asw	.86*	-.55*	-.75*	-	-.51*
5. perc	-.85	.12	.22	-.51*	-
n = 20					

Note: n_asw = number of different pronunciations per pseudowords, acc_s1 and acc_s2 = accuracy scored by scorer 1 and 2, perc = percentage of the most common response,
* = significant result.

https://doi.org/10.1371/journal.pone.0251629.t017

Table 18. Intercorrelations for French adults reading dissimilar pseudowords (Exp 4).

Measure	1	2	3	4	5
1. Entropy	-	-.76*	-.87*	.92*	-.96*
2. acc_s2	-.76*	-	.58*	-.63*	.82*
3. acc_s1	-.87*	.58*	-	-.78*	.86*
4. n_asw	.92*	-.63*	-.78*	-	-.80*
5. perc	-.96*	.82*	.86*	-.80*	-
n = 20					

Note: n_asw = number of different pronunciations per pseudowords, acc_s1 and acc_s2 = accuracy scored by scorer 1 and 2, perc = percentage of the most common response,
* = significant result.

https://doi.org/10.1371/journal.pone.0251629.t018

General discussion

The present study used Entropy as a measure to assess participants' reading aloud responses to pseudowords in English, French, Italian and German adults and children. Our main aim was to assess the impact of age, orthographic depth and bilingualism on Entropy, defined as the number of alternative pronunciations that participants give to a given pseudoword.

The role of children's development in Entropy

Experiment 2 clearly showed a significant decrease in Entropy (H) from grade 2 to 4, with a great fall between grade 2 and 3. This finding is in line with similar results reported in English by [12]), who show that by the end of grade 2, children already start to develop sensitivity to context-sensitive correspondences, which is probably the cause of the reduction in response variability, as Treiman and Kessler [60] suggest for spelling. In fact, children may use the surrounding context of a grapheme to derive pronunciation. This progressive diminution also explains why the majority of pronunciations by adult participants had an Entropy value of zero or very close to zero. However, data from Experiments 1, 2, 3 and 4 show that specific alternative readings did not disappear from childhood into adulthood.

Our results suggest that the pronunciation of some sublexical units is intrinsically ambiguous and variability thus does not depend on reading skills. For example, both French adults and children were divided in whether or not to pronounce final consonants that are normally silent in real words. For example, the pronunciation of a real word like "mot" (word) would uniformly be read as /mo/, while our participants read the pseudoword <stort> as /stər/ or

Table 19. Intercorrelations for German adults reading dissimilar Pseudowords (exp 4).

Measure	1	2	3	4
1. Entropy	-	.08	.89*	-.92*
2. acc	.08	-	-.14	-.19
3. n_sw	.89*	-.14	-	-.70*
4. perc	-.92*	-.19	-.70*	-
n = 20				

Note: n_asw = number of different pronunciations per pseudowords, acc = scored accuracy, perc = percentage of the most common response,
* = significant result.

https://doi.org/10.1371/journal.pone.0251629.t019

/stɔrt/. Similarly, both German adults and children devoiced pseudoword codas half of the time, although final consonant devoicing is the norm in real word reading: <Bad> (bath) will always be read as /ba:t/, while the pseudoword <gund> was read as /gund/ or /gunt/. This phenomenon is not only restricted to position-sensitive correspondences, but also to context-sensitive correspondences. In German, for example, the letter <s>, followed by the letter <p>, should give the phoneme /ʃ/ as in the word "Sport" /ʃpɔrt/. Nonetheless, the pseudoword <sprau> is read by children and adults as /sprau/ or /ʃprau/. Similar instances were found in all languages, and can be found in the tables of the S1 Appendix.

Entropy differences across languages with varying levels of orthographic transparency

To assess how the response variability to pseudowords changed in a deep compared to a shallow orthography, we tested participants in four languages that are on different points of the orthographic depth space: English, French, German and Italian.

In Experiment 1, we firstly compared English and German adults reading monosyllabic pseudowords matched on the number of letters, orthographic neighborhood and body consistency, but against our hypothesis, we did not obtain significant differences. We hypothesised that cross-linguistics differences may be manifested in childhood but would disappear into adulthood. Hence, in the subsequent experiments, we assessed cross-linguistics differences in children.

In Experiment 2 bilingual English/German children read items in both languages, but we did not find an effect of orthographic depth within the same participants. We reasoned that bilingualism itself could have caused this result, because the knowledge of one shallow orthography could have had a facilitatory effect on the deeper orthography by providing a better understanding of the systematicity of GPCs [30].

In Experiment 3 we compared Italian and French children reading a set of cognate pseudowords, against our predictions, we found that Italian children showed significant higher Entropy values than French children. However, we suspected that the reasons behind this result were to be found in the nature of the languages itself and in the items characteristics. In fact, French is considered to be asymmetric in its orthographic depth: while spelling is considered to be hard (/mɛʀ/ can be spelled as "maire" [mayor], "mère"[mother] or "mer" [sea]), reading, in spite of the presence of complex correspondences, is considered predictable [61]. Given that in a complex but predictable orthography, the pronunciation is not ambiguous, there should be a consensus in the responses.

As we did not find higher Entropy in French (a complex, predictable orthography) than Italian (a less complex, predictable orthography), this could, in theory, suggest that Entropy may not be affected by complexity. This would be a first behavioral finding suggesting that complexity and unpredictability have different effects on reading processing, thus providing further weight to the proposal of treating orthographic depth as a multidimensional construct [28]. However, the present study as it is cannot exclude with certainty that other confounding factors are not in action.

In fact, another possibility is that the items that we used in Experiment 3 may have not been representative enough of French orthography. Since we derived pseudowords from a set of cognate items which are, by definition, similar in both orthographies, they could have been lacking the presence of those complex correspondences that French and Italian do not share, but that are common in the respective languages. The CV structure of the items in French was easy, relative to the CV structure of French words in general, which may have facilitated sublexical processing in French relative to Italian and lead to the counter-intuitive finding of higher Entropy in Italian than French.

To rule out this hypothesis we administered a fourth reading aloud experiment in three groups of adults (French, German and English) using two different item sets: in the first set, items were truly representative of the three different orthographies and matched on base-word frequency, while the second set was consisted of items that were equally dissimilar in all three orthographies and had no orthographic neighbors. In the first condition, significant differences were found between English and German and English and French, but not between French and German. In the second condition significant differences were found between all three groups, with English having significant higher Entropy values than both German and French, and German having significantly higher Entropy values than French.

The findings from Experiment 4 suggest that cross-linguistic differences in Entropy seem to be a response to item characteristics [62], and in cross-linguistics research, matching pseudowords on several aspects may hide significance differences in reading behaviour. This would explains why the comparison between English and German was significant in Experiment 4, but not in Experiment 1. In Experiment 1, German and English adults read monosyllabic pseudowords matched on number of letters, orthographic neighborhood, body-neighborhood and, importantly, body consistency. The lack of a difference suggests that participants reading in English did not have overall greater uncertainty when items are made of consistent bodies. However, when English-speaking participants are confronted with a set of items truly representative of their language, with both consistent and inconsistent bodies, uncertainty arises significantly, compared to other languages.

Results from Experiment 2 seemed to follow the same direction. The use of bilingual children had the advantage that the same children read the same items in two different orthographies, thus reducing between-subject variance. While knowledge of a second orthography may have affected the results (knowing a shallow orthography—i.e. German—may have reduced the pronunciation variability of English pseudowords), we found no differences between bilinguals and monolinguals, suggesting that cross-language contamination is an unlikely explanation of the lack of a cross-linguistic difference.

Relations among Entropy and other measures

Throughout the study we compared Entropy with the number of pronunciations per pseudowords, the percentage of the most common response and the accuracy measure. For the first two, we found, as we expected, significant positive correlations between Entropy and number of pronunciations. Clearly, as the number of pronunciations increase, Entropy values also increases. At the same time, the higher the percentage of the most common response to a pseudoword is, the lower the Entropy value for that particular item is, since a high percentage of the most common response means that participant strongly agreed on the pronunciation. Therefore, Entropy and percentage of the most common response was always in a negative, significant correlation.

The most interesting relationship was found between Entropy and accuracy. For three of the four language groups (Italian, German and French), we asked two different scorers who had received training in the phonology of the respective language to evaluate the accuracy of pseudoword readings. We calculated Cohen's kappa to determine scores' agreement. Strong agreement was found in bilingual children reading English items (Exp 2), French children (Exp 3) and English-speaking participants reading dissimilar pseudowords (Exp 4), although we found a nearly perfect agreement only in Italian scorers. Strong agreements were found across all children: a possible explanation would be that judging children's response accuracy was easier for scorers, as some readings were clearly not plausible. Regrettably, we could not hire a second scorer for the German data to provide further evidence to this hypothesis. Since our grade comparison focused on German data, it is possible we could find a negative correlation between scorers' agreement and grade.

Strong agreement was found for English-speaking adults in Experiment 4. It seems that accuracy agreement on pronunciations in which phoneme-grapheme correspondences were clearly unlikely ("gist" read as "gust, <i> − > /u/) is higher compared to the accuracy agreement on pronunciations where readers have to decide which phonemes, virtually associated to a particular grapheme, are to choose ("gid" read as /gid/ or /dʒist/). Since dissimilar English pseudowords were associated with the greatest number of different pronunciations in all three groups, it is not surprising that scorers found a strong agreement in this group as well, even if the participants were adults.

The only nearly perfect agreement was found in Italian data: this may be due to age, because participants were children, and to the fact that Italian is the most transparent language in the pool. This suggests that accuracy and Entropy were not correlated in Italian, because the number of implausible readings was very low, and Entropy was driven by the presence of two or more plausible pronunciations. For example, in Italian there are two phonemes mapping to <g>: but scorers marked both pronunciations as correct based on a lenient marking criterion.

All in all, while by definition Entropy should be significantly, and negatively correlated with accuracy, in practice accuracy judgement itself, for pseudowords, is not a straightforward and error-free process. Even though we gave the same instructions to all scorers, and even though all scorers were trained in phonology, there is variability in judgement both between scorers and within scorers (as the results for the Monolingual German children sample seem to show).

We interpret this finding as a further evidence that accuracy scoring is subject to arbitrariness and its reliability is low. Accuracy, as a measure to evaluate pseudoword reading behavior, is less than ideal. This finding points toward the need to find a different, subjectivity-free measure to investigate pseudoword reading aloud behavior. Since Entropy calculation does not involve any type of human intervention and is a complete, mathematical process, we propose here that Shannon's Entropy, when investigating item-level behavior, could represent, in this regard, a good candidate.

Limitations and future directions

In this study we isolated the effects of orthographic depth, age and bilingualism on a new measure for pseudowords reading aloud performance: Entropy. This measure opens possibilities for future research. The relationship between this measure and a more traditional one, reaction time requires further clarification. For example, pseudowords associated with high Entropy values can take more time to read, because readers have to scan all plausible phonological representations and decide which one is more fitting given a particular context. This would shed light about the cognitive processes that correlate with pseudoword reading aloud Entropy. It is possible that readers have a set of context-sensitive GPCs which they always apply when they encounter a particular orthographic cluster. It might depend on their reading experience, and in particular the frequency with which they encounter a given cluster in real words. Activation of other possible pronunciations is suppressed at an early processing stage, such that Entropy is not reflected in participants' response latencies. Alternatively, it is possible that participants generate possible pronunciations at a late processing stage, before articulation is initiated. This would lead to a closer link between item-level Entropy and RT.

An advantage of the Entropy measure is that its calculation is theory-neutral. While we used the terminology of the Dual Route Cascaded model throughout the paper (e.g.

grapheme-phoneme correspondences), the results also fit within alternative models of reading, such as Connectionist models [63, 64]. The current analysis do not allow us to provide evidence for one model over another.

However, this could be a direction for future research. For example, language-level Entropy at different unit sizes (as described by [18]) could be used as a predictor of pseudoword reading aloud Entropy. This would allow us to assess whether GPCs, as currently implemented in the DRC, are the best predictors of Entropy, or if participants rely on larger units such as bodies. Our multiregression analysis seem to suggest, already, that variables such as orthographic neighborhood, phonological neighborhood, body neighborhood, baseword and bigram frequency and number of syllables do not seem to be good predictors. A further possibility would be to investigate the role of sublexical units in Entropy: using participant-level Entropy by pre-senting the same participants repeatedly with the same orthographic units (as was done by [12]), would allow us to assess whether participants use the same type of units across time, or if there is intra-participant variability in which type of correspondence is applied, which would speak against the notion of an all-or-none rule.

Methodologically, future research on the application of the Entropy measure in pseudowords reading behavior may want to directly assess how and if the present findings change given a different sample size. As with all measures, Entropy is likely to be sensitive to sample size: smaller samples are more likely to be affected by random noise. Furthermore, the number of possible pronunciations, which is a major determiner of the Entropy measure, depends on the number of participants, because the maximum number of possible pronunciations is capped by the number of participants. In practice, the number of different pronunciations is likely to be lower than the number of participants in our experiments. For example, Pritchard, Coltheart, Palethorpe and Castles [7] found, on average, 8 different responses among 45 English-speaking participants. Nevertheless, future research is required to establish at what sample sizes and under what circumstances pseudoword reading aloud Entropy yields stable estimate.

Finally, in our fourth experiment, we randomly picked base words from which we derived pseudowords, without any systematic control (exception made for frequency) to linguistic properties such as body neighborhood, orthographic neighborhood, or number of syllables and length (for the first subset). Our choice was driven by the consideration that we could not find significant cross-linguistics differences in Entropy using systematically chosen items that matched across languages in Experiment 1-3, and we suspected that the item characteristics themselves could be the cause. Although choosing items whose orthographic characteristics were controlled for has the advantage of having potentially psycholinguistic relevant factors contained (for example, the word length), in cross-linguistic research using items that were forced to be similar across orthographies may prevent an adequate representation of the different orthographies features of the languages in question [49]. This shortcoming could be avoided by choosing random base words from which we can derive pseudowords, while using only frequency as a control variable (since frequency is not orthography related, contrary to the above-mentioned characteristics). However, a random selection of base words can result in an unbalanced list, merely due to chance rather than as a reflection of the systematic features of the language. To account for both deficiencies, we decided to try both approaches.

Another interesting application of the Entropy measure may be to investigate subject-level performances. In our study we used Entropy to assess item-level variability while averaging across participants. However, more could be done, for example, by using Entropy to calculate intra-participant variability (whether the same participant was consistent in pronunciation when asked to read a given pseudoword more than once). Lastly, in the current study we investigated how Entropy correlated with accuracy and discussed the short-comings of the latter in pseudoword reading behavior investigation. However, some questions are still unresolved. It remains to ascertain if and how Entropy interacts with other measures, such as reaction times: while pseudoword accuracy scoring is subject to human arbitrary decisions, RT measures are not. At the same time, in our study we did not focus on what specifically could be a predictor of Entropy.

As an exploratory analysis, we ran some multiple regressions addying body neighborhood, orthographic neighborhood, baseword frequency, bigram frequency, number of syllables and length as predictors (forin Experiment 1, 2 and 4), but none of those turned out to be reliable predictors (length only affected Entropy in the comparison between English and French). Future studies could look into this specifically, for example by running a model using different measures such as body-rime consistency or vowel consistency, like in [18].

Conclusion

The present study contributes to the literature using Entropy as a measure to quantify the variability in pseudowords pronunciation. We investigated whether Entropy changes in relationship to orthographic depth, age and bilingualism.

The results indicate that deeper orthographies lead to higher Entropy values, provided that items are truly representative of the orthographies under scrutiny. Furthermore, our preliminary results suggest that the effect of Entropy is driven by the degree of unpredictability, but not by the complexity of an orthography. This is a first demonstration of a differential effect of complexity and unpredictability as dissociable constructs underlying orthographic depth, which stresses the need to consider the multidimensional nature of orthographic depth in cross-linguistic reading research.

The present study demonstrates that Entropy decreases across age, indicating that the agreement on pseudoword pronunciations increases in relation to the development of reading skills. Finally, we did not find significant differences in Entropy values between monolingual and bilingual children. All this considered, this study can be regarded as a starting point to evaluate the use of alternative measures, and specifically Entropy, to investigate cross-linguistics differences in pseudoword reading and reading development.

Supporting information

S1 Appendix.
(PDF)

Acknowledgments

We wish to thank Tara Arrow for scoring the accuracy of the English data, Caterina Pace for scoring the accuracy of the Italian data, Erik Henzler for collecting the German data used in Experiment 4 and Roland Trouville for his technical advice.

Author Contributions

Conceptualization: Xenia Schmalz.

Data curation: Elisabetta De Simone, Elisabeth Beyersmann, Claudio Mulatti, Jonathan Mirault, Xenia Schmalz.

Formal analysis: Elisabetta De Simone, Xenia Schmalz.

Funding acquisition: Claudio Mulatti, Xenia Schmalz.

Investigation: Elisabetta De Simone, Jonathan Mirault.

Methodology: Elisabetta De Simone, Xenia Schmalz.

Project administration: Elisabetta De Simone, Xenia Schmalz.

Resources: Elisabetta De Simone, Elisabeth Beyersmann, Xenia Schmalz.

Software: Xenia Schmalz.

Supervision: Elisabeth Beyersmann, Claudio Mulatti, Xenia Schmalz.

Validation: Elisabetta De Simone, Xenia Schmalz.

Visualization: Elisabetta De Simone, Xenia Schmalz.

Writing – original draft: Elisabetta De Simone.

Writing – review & editing: Elisabetta De Simone, Elisabeth Beyersmann, Xenia Schmalz.

References

1. Castles A, Polito V, Pritchard S, Anandakumar T, Coltheart M. Do nonword reading tests for children measure what we want them to? An analysis of year 2 error responses. Australian Journal of Learning Difficulties. 2018; 23(2):153–165. https://doi.org/10.1080/19404158.2018.1549088
2. Coltheart M, Rastle K, Perry C, Langdon R, Ziegler J. DRC: a dual route cascaded model of visual word recognition and reading aloud. Psychological review. 2001; 108(1):204–256. https://doi.org/10.1037/0033-295X.108.1.204
3. Jackson N, Coltheart M. Routes to reading success and failure: Toward an integrated cognitive psychology of atypical reading Psychology Press. 2013.
4. Holligan C, Johnston RS. The use of phonological information by good and poor readers in memory and reading tasks. Memory & Cognition. 1988; 16(6):522–532. https://doi.org/10.3758/BF03197053
5. Olson RK, Wise BW, Rack JP. Dyslexia: Deficits, genetic aetiology and computer-based remediation. The Irish Journal of Psychology. 1989; 10(4):494–508. https://doi.org/10.1080/03033910.1989.10557765
6. Andrews S, Scarratt DR. Rule and analogy mechanisms in reading nonwords: Hough dou peapel rede gnew wirds? Journal of Experimental Psychology: Human Perception and Performance. 1998; 24(4):1052–1086.
7. Pritchard SC, Coltheart M, Palethorpe S, Castles A. Nonword reading: Comparing dual-route cascaded and connectionist dual-process models with human data. Journal of Experimental Psychology: Human Perception and Performance. 2012; 38(5):1268–1288.
8. Schmalz X, Marinus E, Robidoux S, Palethorpe S, Castles A, Coltheart M. Quantifying the reliance on different sublexical correspondences in German and English. Journal of Cognitive Psychology. 2014; 26(8):831–852. https://doi.org/10.1080/20445911.2014.968161
9. Seidenberg MS, Plaut DC, Petersen AS, McClelland JL, McRae K. Nonword pronunciation and models of word recognition. Journal of Experimental Psychology: Human Perception and Performance. 1994; 20(6):1177–1196.
10. Glushko RJ. The organization and activation of orthographic knowledge in reading aloud. Journal of experimental psychology: Human perception and performance. 1979; 5(4):674–691.
11. McCann RS, Besner D. Reading pseudohomophones: Implications for models of pronunciation assembly and the locus of word-frequency effects in naming. Journal of Experimental Psychology: Human Perception and Performance. 1987; 13(1):14–24.
12. Schmalz X, Robidoux S, Castles A, Marinus E. Variations in the use of simple and context-sensitive grapheme-phoneme correspondences in English and German developing readers. Annals of Dyslexia, 2020;1–20.
13. Shannon CE. A mathematical theory of communication. Bell system technical journal. 1948; 27(3):379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
14. Levy R. Expectation-based syntactic comprehension. Cognition. 2008; 106(3):1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006 PMID: 17662975
15. Borgwaldt SR, Hellwig FM, de Groot AM. Word-initial Entropy in five languages: Letter to sound, and sound to letter. Written Language and Literacy. 2004; 7(2):165–184.

16. Borgwaldt SR, Hellwig FM, De Groot AM. Onset Entropy matters–Letter-to-phoneme mappings in seven languages. Reading and Writing. 2005; 18(3):211–229
17. Ziegler JC, Bertrand D Tóth D, Csépe V, Reis A, Faísca L, et al. Orthographic depth and its impact on universal predictors of reading: a cross-language investigation. Psychological science. 2010; 21(4): 551–559. https://doi.org/10.1177/0956797610363406 PMID: 20424101
18. Siegelman N, Kearns DM, Rueckl JG. Using information-theoretic measures to characterize the structure of the writing system: the case of orthographic-phonological regularities in English. Behavior Research Methods. 2020; 52:1–21.
19. Protopapas A, Vlahou EL. A comparative quantitative analysis of Greek orthographic transparency. Behavior research methods. 2009; 41(4):991–1008. https://doi.org/10.3758/BRM.41.4.991
20. Mousikou P, Sadat J, Lucas R, Rastle K. Moving beyond the monosyllable in models of skilled reading: Mega-study of disyllabic nonword reading. Journal of Memory and Language. 2017; 93:169–192. https://doi.org/10.1016/j.jml.2016.09.003
21. Seymour PH, Aro M, Erskine JM, with COST Action A8 Network C. Foundation literacy acquisition in European orthographies. British Journal of psychology. 2003; 94(2):143–174. https://doi.org/10.1348/000712603321661859
22. Mousikou P, Beyersmann E, Ktori M, Javourey Drevet L, Crepald D, Ziegle JC, et al. Orthographic consistency influences morphological processing in reading aloud: Evidence from a cross linguistic study. Developmental Science. 2020, e12952. PMID: 32061144
23. Frith U, Wimmer H, Landerl K. Differences in phonological recoding in German-and English-speaking children. Scientific Studies of reading. 1998; 2(1):31–54. https://doi.org/10.1207/s1532799xssr0201_2
24. Landerl K. Influences of orthographic consistency and reading instruction on the development of nonword reading skills. European journal of psychology of education. 2000; 15(3):239–257. https://doi.org/10.1007/BF03173177
25. Wimmer H, Goswami U. The influence of orthographic consistency on reading development: Word recognition in English and German children. Cognition. 1994; 51(1):91–103. https://doi.org/10.1016/0010-0277(94)90010-8
26. Goswami U. Orthography, phonology, and reading development: A cross-linguistic perspective. Handbook of orthography and literacy. 2006; p. 463–480.
27. Share DL. Alphabetism in reading science Frontiers in psychology. 2014; 5: 752. https://doi.org/10.3389/fpsyg.2014.00752 PMID: 25101024
28. Schmalz X, Marinus E, Coltheart M, Castles A. Getting to the bottom of orthographic depth. Psychonomic Bulletin & Review. 2015; 22(6):1614–1629. https://doi.org/10.3758/s13423-015-0835-2
29. Share DL. Orthographic learning, phonological recoding, and self-teaching. Advances in child development and behavior. 2008; 36:31–82. https://doi.org/10.1016/S0065-2407(08)00002-5
30. Treiman R, Kessler B, Evans R. Anticipatory conditioning of spelling-to-sound translation. Journal of Memory and Language. 2007; 56(2):29–245.
31. Pratt C, Grieve R. The development of metalinguistic awareness: An introduction. Metalinguistic awareness in children. Springer Series in Language and Communication. 1984(15):2–11.
32. Katz L, Frost R. The reading process is different for different orthographies: The orthographic depth hypothesis. Advances in psychology. 1992; 94:67–84. https://doi.org/10.1016/S0166-4115(08)62789-2
33. Schmalz X, Robidoux S, Castles A, Coltheart M, Marinus E. German and English bodies: No evidence for cross-linguistic differences in preferred orthographic grain size. Collabra: Psychology. 2017; 3(1): 5. https://doi.org/10.1525/collabra.72
34. Ziegler JC, Perry C, Jacobs AM, Braun M. Identical words are read differently in different languages. Psychological science. 2001; 12(5):379–384. https://doi.org/10.1111/1467-9280.00370
35. Coltheart M. Access to the internal lexicon. The psychology of reading. 1977:535–555.
36. Landerl K, Wimmer H, Frith U. The impact of orthographic consistency on dyslexia: A German-English comparison Cognition. 1997a; 63:315–334.
37. Ziegler JC, Perry C, Ma-Wyatt CA, Ladner D, Schulte-Körne G. Developmental dyslexia in different languages: Language-specific or universal? Journal of experimental child psychology. 2003; 86(3):169–193 https://doi.org/10.1016/S0022-0965(03)00139-5 PMID: 14559203
38. The R Development Core Team R: A language and environment for statistical computing. 2013
39. Levenshtein VIBinary codes capable of correcting deletions, insertions, and reversals. Soviet physics doklady. 1966; 10(8):707–710.
40. Cohen JA coefficient of agreement for nominal scales. Educational and psychological measurement. 1960; 20(1):37–46 https://doi.org/10.1177/001316446002000104

41. Pritchard SC, Coltheart M, Palethorpe S, Castles A. Nonword reading: Comparing dual-route cascaded and connectionist dual-process models with human data. Journal of Experimental Psychology: Human Perception and Performance. 2012; 38(5):1268.
42. Bassetti B. Bilingualism and Writing Systems. The handbook of bilingualism and multilingualism. 2013; 649–670.
43. Yelland GW, Pollard J, Mercuri A. The metalinguistic benefits of limited contact with a second language. Applied psycholinguistics. 1993; 14(4):423–444. https://doi.org/10.1017/S0142716400010687
44. Diaz RM. Bilingual cognitive development: Addressing three gaps in current research. Child development. 1985; 5(6):1376–1388.
45. Bialystok E. Factors in the growth of linguistic awareness. Child development. 1986; 57(2):498–510. https://doi.org/10.2307/1130604
46. Galambos SJ, Goldin-Meadow S. The effects of learning two languages on levels of metalinguistic awareness. Cognition. 1990; 34(1):1–56. https://doi.org/10.1016/0010-0277(90)90030-N
47. Rose Y, Wauquier-Gravelines S. French speech acquisition. The international guide to speech acquisition. 2007; p. 364–384.
48. Perry C, Ziegler JC, Zorzi M. When silent letters say more than a thousand words: An implementation and evaluation of CDP++ in French. Journal of Memory and Language. 2014; 72:98–115. https://doi.org/10.1016/j.jml.2014.01.003
49. Ellis NC, Natsume M, Stavropoulou K, Hoxhallari L, Van Daal VH, Polyzoe N, et al. The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. Reading research quarterly. 2004; 39(4):438–468. https://doi.org/10.1598/RRQ.39.4.5
50. Venezky RL. The structure of English orthography. Walter de Gruyter. Janua Linguarum: Series Minor. 2011; 82.
51. Brysbaert M, Buchmeier M, Conrad M, Jacobs AM, Bölte J, Böhl A. The word frequency effect. Experimental psychology. 2011; 58(5):412–424. https://doi.org/10.1027/1618-3169/a000123
52. Ferrand L, New B, Brysbaert M, Keuleers E, Bonin P, Méot A, et al. The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. Behavior Research Methods. 2010; 42(2):488–496. https://doi.org/10.3758/BRM.42.2.488 PMID: 20479180
53. Van Heuven WJ, Mandera P, Keuleers E, Brysbaert M. SUBTLEX-UK: A new and improved word frequency database for British English. The Quarterly Journal of Experimental Psychology. 2014; 67 (6):1176–1190. https://doi.org/10.1080/17470218.2013.850521
54. Keuleers E, Brysbaert M. Wuggy: A multilingual pseudoword generator. Behavior research methods. 2010; 42(3):627–633. https://doi.org/10.3758/BRM.42.3.627
55. Forster KI, Forster JC. DMDX: A Windows display program with millisecond accuracy. Behavior Research Methods, Instruments, & Computers. 2003; 35(1):116–124. https://doi.org/10.3758/BF03195503
56. Leipzig Corpora Collection German news corpus based on material from 2020. Dataset. https://corpora.uni-leipzig.de?corpusId=deu_news_2020
57. Leipzig Corpora Collection English news corpus based on material from 2020. Dataset. https://corpora.uni-leipzig.de?corpusId=eng_news_2020
58. Leipzig Corpora Collection French news corpus based on material from 2020. Leipzig Corpora Collection. Dataset. https://corpora.uni-leipzig.de?corpusId=fra_news_2020
59. Marian V, Bartolotti J, Chabal S, Shook A. CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. PloS one, 7(8). 2012;e43230 https://doi.org/10.1371/journal.pone.0043230 PMID: 22916227
60. Treiman R, Kessler B. Spelling as statistical learning: Using consonantal context to spell vowels. Journal of Educational Psychology. 2006; 98(3):642 https://doi.org/10.1037/0022-0663.98.3.642
61. Sprenger-Charolles L, Messaoud-Galusi S. Review of research on reading acquisition and analyzes of the main international reading assessment tools. 2009;pp 41. hal-00733607v1.
62. Egan C, Oppenheim GM, Saville C, Moll K, Jones MW. Bilinguals apply language-specific grain sizes during sentence reading. Cognition. 2019; 193:104018. https://doi.org/10.1016/j.cognition.2019.104018
63. Plaut DC, McClelland JL, Seidenberg MS, Patterson K Understanding normal and impaired word reading: computational principles in quasi-regular domains. Psychological review 1996; 103(1):56–115. https://doi.org/10.1037/0033-295X.103.1.56
64. Perry C, Ziegler JC, Zorzi M. Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud. Psychological review, 2007; 114(2), 273–315. https://doi.org/10.1037/0033-295X.114.2.273

S1 Appendix

Table 1. List of pseudowords read as real words (Exp 1).

Pseudoword	Word	Nread	Translation	Percentage	Participants
stide	tag	1/24		4.17	English
floud	flood	1/24		4.17	English
dinn	dünn	1/19	thin	5.25	German
reuz	kreuz	2/19	cross	10.53	German
stork	stock	1/19	floor	5.25	German
wolz	wolf	1/19	wolf	5.25	German

Note: Nread indicates the number of participants in the group that read the pseudoword as real words.

Table 2. German Adults reading matched monosyllabic pseudowords (Exp. 1)

Items	Pronunciations	Comments
quang	(6) kvaŋ, kwaŋg (5) kwaŋ (1) kwan, kvaŋg	Different readings for the grapheme <u>: /v/ or /w/ Different reading for the grapheme <n[g]>: /ŋ/ or /ng/
splur	(10) ʃplur, (4) splur, (2) ʃplu:r, (1) splu:r, ʃplul, plu:r	Different readings for the grapheme <s[p]>: /ʃ/ or /s/ Different vowel lengths + Phoneme replacement or deletion
frur	(11) fru:r, (5) frur, (1) flu:r, flu:rur, fur	Different vowel lengths + Phoneme deletion and replacement Addition of syllable (flu:rur)
gund	(9) gunt, (7) gund, (2) gu:nd, (1) grund	Real word reading (grund - reason) The final consonant devoicing rule is either applied or not Different Vowel lengths
mang	(9) maŋ, (9) maŋg, (1) maŋk	Different readings for the grapheme <n[g]>: /ŋ/, /ŋg/ or /ŋk/
schweck	(11) ʃvek, (5) svek, (2) ʃveŋk, (1) ʃve:k	Different vowel lengths + phoneme insertion (ʃveŋ k) Different readings for the grapheme <s>: /ʃ/ or /s/
zein	(9) zain, (8) tsain, (1) tse:n, tsaim	Different readings for the following graphemes: <z> as /ts/ or /z/ and <ei> as /ai/ or /e:/
zwau	(9) tsvau, (8) zvau, (1) tsva:u, zvau:	Different readings for the grapheme <z>: /ts/ or /z/ Different vowel lengths
sprau	(13) ʃprau, (3) sprau, (2) ʃpau, (1) ʃprau:	Different readings for the grapheme <s[p]>: /ʃ/ or /s/ Different vowel lengths + Phoneme deletion (ʃpau)
fold	(14) folt, (2) fold, fo:lt (1) foalt	The final consonant devoicing rule is either applied or not Different vowel lengths

Note: The numbers in parentheses indicate how many participants read the item with the following pronunciations

Table 3. English Adults reading matched monosyllabic pseudowords (Exp. 1)

Items	Pronunciations	Comments
wurn	(11) wɛ:n, (2) wɛrn (1) wuʌn, wʌn, wun, (1) wo:n, woʌn, w3:nt	Different readings for the following graphemes: <u[r]> as /ɛ/, /u/, /a/, /o/ and <[u]r> as /ʌ/ or ø
dize	(14) daiz, (2) dizi, (1) di:z, di:ze, dizi	Different readings for the following graphemes <i>: as /ai/ or /i/ and final <e> as /i/, /e/ or ø
gule	(14) gu:l, (2) gju:l, g ʌl, (1) goʊl	Different readings for the grapheme <u>: /u/, /oʊ/, /ʌ/
kuy	(10) kai, (6) kui, (2) ki:, (1) ku:	Different readings for the grapheme <uy>: /ai/, /i:/, /u:/ or /ui/
luice	(10) luis, (6) lu:s, (3) lus	Different readings for the grapheme <ui>: as /u:/, /u/ or /ui/
whun	(14) wʌn, (2) wun (1) huan, hwʌn, wu:n	Different readings for the following graphemes: <u> as /u/, /u:/ or /ʌ/ and <w> as /h/ or /w/
pluit	(13) pluit, (5) plu:t, (1) plait	Different readings for the grapheme <ui>: /ui/, /u/, /ai/
sirt	(13) sɛ:t, (2) sɛʌt, sɛt, (1) siʌt, sɛ:ʌt	Different readings for the following graphemes: <[i]r> as /ʌ/ or øand <i[r]> as /ɛ/, /ɛ:/ or /i/
fice	(12) fais, (4) fi:s	Different readings for the grapheme <i>: /ai/ or /i:/
yorch	(16) jo:ʧ, (1) jorʧ, jortn, jorʧt	Different readings for the grapheme <[o]r>: /ʌ/ or ø Phoneme addition (jorʧt)

Note: The numbers in parentheses indicate how many participants read the item with the following pronunciations

Table 4. Bilingual German/English children (grade 2) reading English-like pseudowords (Exp. 2)

Item	Pronunciations	Comments
derge	(3) dɜ:ʤ (1) deɹʤ, dik, dɜɹg, dɜ:ge, dɜɹʤ, dɜ:rg, dɹeg, dɜ:g, deɹg	Different readings for the following graphemes: <e[r]> as /ɜ:/ or /e/ and <g> as /ʤ/ or /g/
gurt	(4) gɜ:t (2) gʌ:t (1) gɜʌt, got, goɹt, gæɔt, goʊɹt	Different readings for the following graphemes: u[r] as /ɜ:/, /ʌ:/, /ɜr/, /oʊ/, /oɹ/ <u> as /æɔ/, /ʌ/
gule	(4) gu:l, (2) gʌl (1) gul, glu:, glai, goʊl, gju:l	Different readings for the grapheme <u>: /u/, /oʊ/, /ju/ Different vowel lengths Real word reading (glue)
murse	(5) mɜ:s (2) moɹs (1) muɹze, mɜʌs, mɜ:si, maɹs	Different readings for the following graphemes: <u[r]> as /ɜ:/, /ʌ/, /ɜɹ/, /o/ and <s> as /z/ or /s/
luice	(5) luis (1) lauis, lu:s, luls, lak, laŋk	Different readings for the grapheme <u[i]>: /u:/ or /ui/ Phoneme insertions (luls, laŋk)
pluit	(5) plu:t (2) pluit, plinkt, plait, plot, plont	Different readings for the grapheme <u[i]>: /u:/ or /ui/ Phoneme insertions (plinkt, plont)
suzz	(6) sʌz (1) su:z, suts, sʌts, zʌz, suz	Different readings for the following graphemes: <z> as /z/ or /ts/, <s> as /z/ or /s/ and <u> as /ʌ/ or /u/ Different vowel lengths
sirt	(5) sɜ:t (4) siɹt (1) zɜ:t, ʃɜ:t, sɜɹt	Different readings for the following graphemes: <s> as /ʃ/, /z/ or /s/ and <i[r]> as /ɜɹ/ or /iɹ/
tirm	(5) tɜ:m (3) tɜɹm (1) tɹim, tim, tiɹm	Different readings for the following graphemes: <i[r]> as /ɜ/, /i/, /ɜ:/ and <[i]r> as /ɹ/ or /ø/
roud	(6) ɹæ (2) ɹu:d, ɹæɔnd (1) ɹoʊd, ɹod	Different readings for the following grapheme: <ou> as /æɔ/, /u:/, /oʊ/, /o/ Real word reading (round)

Note: The numbers in parentheses indicate how many participants read the item with the following pronunciations

Table 5. Bilingual German/English children (grade 3) reading English-like pseudowords (Exp. 2)

Items	Pronunciations	Comments
derge	(2) dɛ:ʤ, (1) dɛrʤ, dɹag, dɛ:g	Different readings for the following graphemes: <[e]r> as ɹor øand <g> as /g/ or /ʤ/ Phoneme inversion (drag)
fich	(2) fiʧ, (1) fiʃ, fɹiʃ, fit	Different reading for the grapheme <ch>: /ʃ/ or /ʧ/ Real word reading (fish) Phoneme insertion (friʃ)
gule	(2) gʌl, (1) gu:l, gju:l, glu:	Different reading for the grapheme: <u> as /u:/, /ʌ/ or /ju/ Real word reading (glue)
krilk	(2) kɹilk, (1) kɹik, klɛ:k, kɹai	Different readings for the grapheme <i>: /i/, /ai/ or /ɛ:/ Phoneme deletion (kik)
pliz	(3) plits, (1) plis, pliz	Different readings for the grapheme <z>: /ts/, /z/ or /s/
pluit	(2) plu:t, (1) plu:ʧ, pluit, palt	Different readings for the grapheme <ui>: /u:/ or /ui/ Phoneme inversion (plat)
stum	(2) stam, (1) stʌŋ, stu:m, stum	Different readings for the following graphemes: <u> as /ʌ/, /u:/ or /u/ and <m> as /ŋ/ or /m/
whun	(2) wʌn, (1) win, wu:m, wu:n	Different readings for the following graphemes: <u> as /ʌ/, or /u:/ and <n> as /m/ or /m/
wrum	(2) ɹam, (1) wan, wɛ:n, wo:m	Different readings for the following graphemes: <u> as /ʌ/, /ɛ:/ or /o:/ and <m> as /n/ or /m/
chyle	(2) kai, ʧail, (1) cycle	Different readings for the grapheme <ch>: /ʧ/ or /k/

Note: The numbers in parentheses indicate how many participants read the item with the following pronunciations

Table 6. Bilingual German/English children (grade 4) reading English-like pseudowords (Exp. 2)

Items	Pronunciations	Comments
gule	(1) ʤu:l, glu:, gu:l, gʌl, gju:l	Different readings for the following graphemes: <u> as /u:/, /ʌ/ or /ju:/ and <g> as /g/ or /ʤ/
whun	(2) wʌn, (1) wu:n, vʌn, wun	Different readings for the following graphemes: <u> as /u:/, /u/ or /ʌ/ and <w> as /w/ or /v/
yorch	(2) yo:ʧ, (1) zoɹʧ, youɹʧ, ʧoɹʧ	Different readings for the grapheme <[o]r>: /ɹ/ or ø Phoneme replacement (ʧoɹʧ)
barsh	(2) ba:ʃ, baɹʃ, (1) bɹaʃ	Different readings for the grapheme <[a]r>: /ɹ/ or ø Phoneme inversion (bɹaʃ)
chycle	(2) ʧail, cycle (1) kju:li	Different readings for the grapheme <ch>: /ʧ/ or /k/
splaw	(2) splo:, splæɔ, (1) spo:	Different readings for the grapheme <aw>: /o:/ or /æɔ/
swuff	(2) swʌf, swuf, (1) stuf	Different readings for the grapheme <u>: /u/ or /ʌ/ Phoneme replacement (stuf)
wrum	(2) ɹam, wɛ:m, (1) woɹm	Different readings for the following graphemes: <u> as /ʌ/ or /ɛ:/ and <w> as /w/ or ø
chy	(3) ʧi, (1) ʧi:, ʃai	Different readings for the following graphemes: <y> as /i:/, /i/ or /ai/ and <ch> as /ʧ/ or /ʃ/
frict	(3) fɹikt, (1) fikt, fɹiʧ	Phoneme deletion (fikt)

Note: The numbers in parentheses indicate how many participants read the item with the following pronunciations

Table 7. Bilingual German/English children (grade 2) reading German-like pseudowords (Exp. 2)

Items	Pronunciations	Comments
splur	(3) ʃplu:r, ʃplur (2) ʃplʊ:r (1) splu:r, splʊr, ʃplu:ur, ʃpu:ur, ʃpau	Different readings for the grapheme <s>: /ʃ/ or /s/, even if <s>before <p>should always be ʃ Different readings for the grapheme <u>: /u/ or /ʊ/ Different vowel lengths, Phonemes deletion
reuz	(6) roits, (1) roiz, raits, rʊts, kroits, krets, raiz, ru:z, røz	Different readings for the diphtong <eu>: /oi/, /ai/, /e/ and /ø/ Different readings for the grapheme <z>: /z/ or /ts/ Real word reading (kreuz - cross)
klund	(5) klunt, (2) kluŋk (1) kolt, klaunt, klu:nt, klun, klund	The final consonant devoicing rule is either applied or not Real word reading (colt) Different readings for the grapheme <n>: /ŋ/ or /n/ Different vowel lengths, phoneme deletions
goos	(4) gu:s, (3) gus, gos, (1) gʊ:s, gʊs, bu:s	Different readings for the grapheme <oo>: /u/, /o/ or /ʊ/ Real word reading (bus) Different vowel lengths
frur	(5) fru:r, (2) frør, frur, (1) fu:r, jur, frau:ç	Different readings for the grapheme <u>: /u/ or /ø/ Different vowel lengths, phonemes deletions and replacements
lonch	(5) lœn:ç, (2) loinç, loç, (1) lʊnç, lʊnf, lʊk	Different readings for the grapheme <o>: /o/, or /ʊ/ Different readings for the grapheme <ch>: /ç/ or /k/ Phoneme deletions
seng	(5) zeŋ, zeŋk, (1) zain, ze:ŋk, ziŋ, zin	Different readings for the grapheme <n>: /n/ or /ŋ/ Different readings for the grapheme <e>: /e/, /ai/ or /i/ Final consonant devoicing or deletion
pang	(5) paŋk, (2) beŋ, paŋ, (1) praŋ, paŋg, pa:ŋg	Different readings for the grapheme <n>: /n/ or /ŋ/ The final consonant devoicing rule is either applied or not Different vowel lengths Phonemes replacement, deletions or insertion
truck	(6) bruk, (3) truk, (4) brʊke, bru:k, gral, brʊk	Different readings for the consonant cluster <tr>: /br/ or /tr/ Different readings for the grapheme <u>: /u/ or /ʊ/ Different vowel lengths + Phonemes replacement
spand	(5) ʃpant, (3) spant, (2) spand, (1) swant, ʃpand	Different readings for the grapheme <s>: /ʃ/ or /s/, even if <s>before <p>should always be ʃ The final consonant devoicing rule is either applied or not

Note: The numbers in parentheses indicate how many participants read the item with the following pronunciations

Table 8. Bilingual German/English children reading German-like pseudowords in grade 3 (Exp. 2)

Items	Pronunciations	Comments
beld	(1) belt, be:lt, pelt, blent	Final consonant devoicing Real word reading (blend - it dazzles) Different vowel lenghts + Phoneme inversion and insertion
frur	(2) frur, (1) fu:r, frʊr, fru:r	Different readings for the grapheme <u>: /u/ or /ʊ/ Different vowel lengths + Phonemes deletion
pies	(2) pi:s, (1) pa:is, pis, bi:s	Different readings for the grapheme <ie>: /i/ or /ai/ Probable recognition of the English word "pies" and subsequent reading. Different vowel lenghts + Phoneme replacement
poot	(2) po:t, (1) pu:t, plut, pot	Different readings for the grapheme <oo>: /o/ or /u/ Different vowel lengths + Phoneme insertion
reil	(2) rail, (1) ra:ail, pail, prail	Phoneme insertions
splur	(2) ʃplur, (1) ʃplu:r, ʃlur, splur	Different readings for the grapheme <s>: /ʃ/ or /s/, even if <s> before <p> should always be ʃ Different vowel lengths + Phonemes deletions
lusch	(2) luʃ, lu:ʃ, (1) loʃ	Different readings for the grapheme <u>: /u/ or /o/ Different vowel legths
melz	(2) melz, melts, (1) molz	Different readings for the grapheme <z>: /z/ or /ts/ Phoneme replacement
kreck	(3) krek, (1) frek, ʃrek	Real word readings (frech - rebellious & schreck - fright)
quang	(3) kwaŋ (1) kwaŋk, kuwaŋ	Final consonant devoicing Phoneme deletion and insertion

Note: The numbers in parentheses indicate how many participants read the item with the following pronunciations

Table 9. Bilingual German/English children reading German-like pseudowords in grade 4 (Exp. 2)

Items	Pronunciations	Comments
zwau	(2) tsvau, (1) zvau, tsau	Different readings for the grapheme <z>: /z/ or /ts/
spand	(2) ʃpant, (1) ʃpaŋ, ʃpand	Different readings for the grapheme <s>: /ʃ/ or /s/, even if <s> before <p> should always be ʃ The final consonant devoicing rule is either applied or not Different readings for the grapheme <n>: /n/ or /ŋ/
retz	(2) retst, (1) pets, rets	Phoneme insertion and replacement
poot	(2) pu:t, (1) put, prot	Different readings for the grapheme <oo>: /u/ or /o/ Possible influences from the knowledge of English Different vowel lengths + Phoneme insertion
pies	(2) pi:s, (1) pis, pais	Different readings for the grapheme <ie>: /i/ or /ai/ Probable recognition of the English word "pies" and subsequent reading. Different vowel lengths
nech	(2) neç, (1) neʧ, heç	Different readings for the grapheme <ch>: /ç/ or /ʧ/ Phoneme replacement
jenf	(2) ʧenf, jenf	Different readings for the grapheme <j>: /j/ or /ʧ/
laat	(2) lat, la:t	Different vowel lengths
mohl	(2) mol, mo:l	Different vowel lengths
silm	(2) zilm, zelm	Different readings for the grapheme <i>: /i/ or /e/

Note: The numbers in parentheses indicate how many participants read the item with the following pronunciations

Table 10. Monolingual German children (grade 2) reading monosyllabic pseudowords (Exp. 2)

Item	Pronunciations	Comments
dels	(6) delts (3) de:lts, belts (2) delt, (1) dels, delz, de:lt, te:ls, de:ls, de:s, dele:ts, dalts	Phoneme replacements, insertions or deletion different vowel lengthening different readings for the grapheme <s> (/ts/, /s/, or /z/)
silm	(9) zilm, (3) zil, (2) zelm, silm (1) selm, tsil, zelf, zi:l, fi:lm, tsilm	Phoneme replacements and deletions, due to mispronunciations Different readings for the grapheme <s> (/ts/, /s/, or /z/), and <e> (/i/ or /e/) Real word readings (film)
keiz	(9) kraits (2) keits, keis, kreits (1) kaints, kreis, veiz, kalts, kraits, kaiz, ke:ts	Different readings for the grapheme <ei> (/ai/,/ei/ or /e:/), although the first one is the correct one. Different readings for the grapheme <s> (/ts/, /s/, or /z/) Phoneme insertions or replacements due to mispronunciations
grein	(11) grain, (5) grai:n (3) krain, (2) gren (1) grai:, kre:, gre, gre:n	Different readings for the grapheme <ei> (/ai/,/ei/ or /e:/) Devoicing of the first consonant <g> -> k Different vowels length
gund	(10) grain, (4) grai:n, (3) krain (2) gren, (1) gre:n, gre, kre:, grai:	Application or no of the final consonant devoicing phonotactic rule Different vowels length, first consonant devoicing Participants read similar real words instead of the item (grund - reason) Phoneme deletion or insertion
quang	(7) kuaŋk, (4) kuaŋg, (3) kwaŋk, kuwaŋk (1) kuaŋ, kwaŋ, kuwaŋg	Different readings for the grapheme <u> (/u/ or /w/) The final consonant devoicing phonotactic rule is either applied or not Phoneme insertion or deletion
reuz	(8) roits, (5) kreuts, (1) reuts, keuts, raits, roi:z, roi:, roi:ts	Different readings for the grapheme <eu> (/oi/, /eu/ or /ai/) Different readings for the grapheme <z> (/ts/ or /z/) Real word readings (kreuz - cross) Different vowel length
melz	(1) melts, (4) me:lts, (1) mol, nelts, ne:z, ma:lts, malts	Different readings for the grapheme <z> (/ts/ or /z/) Phonemes deletion or replacements Different vowel lengths
seng	(8) zeŋk, (4) tseŋk, tseŋg (2) tsiŋk (1) seŋk, ziŋk	Different readings for the grapheme <s> (/ts/, /z/ or /s/) The final consonant devoicing phonotactic rule is either applied or not Final consonant deletion
sinks	(13) ziŋk, (2) tsiŋks (1) siŋkts, tsiŋkts, stiŋkz, tsiŋ, tsiŋk, ziŋkz, ziŋk	Different readings for the grapheme <s> (/ts/, /z/ or /s/) Phonemes insertion or deletion

Note: The numbers in parentheses indicate how many participants read the item with the following pronunciations

Table 11. Monolingual German children (grade 3) reading monosyllabic pseudowords (Exp. 2)

Items	Pronunciations	Comments
quang	(7) kwaŋg, (5) kwaŋk, (2) kuwaŋk, kwan, kwaŋ (1) gwaŋk, kuwaŋg	The final consonant devoicing rule is either applied or not Different readings for the grapheme <n>: /n/ or /ŋ/ even if it should always be ŋ before g Different readings for the initial grapheme <g>: /g/ or /k/ Phoneme insertions and deletions
pang	(9) paŋk, (3) paŋg, (2) peŋk, paŋ, (1) pan, pa:ŋg, praŋg	The final consonant devoicing rule is either applied or not Different readings for the graphene <a>: /a/ or /e/ Different vowel lengths + Phoneme insertions and deletions
dels	(8) delts,(5) delt, (2) de:lz, delz, (1) belts	Different readings for the grapheme <s>: /z/ or /ts/ Different vowel lengths + Phoneme replacements
teins	(11) taints, (2) tainz, paints, (1) painz, taits, tain, taint, taiŋ ts	Different readings for the grapheme <s>: /s/ or /ts/ Phoneme replacements
goos	(8) gos, (5) go:s (2) gu:s, (1) gous, gu:ʃ, gus	Different readings for the grapheme <oo>: /o/, /ou/ or /u/ Different radings for the grapheme <s>: /s/ or /ʃ/ Different vowel lengths
femd	(10) femt (4) fent (1) fe:nt, fem, fremt	The final consonant devoicing rule is either applied or not Different readings for the grapheme <m>: /n/ or /m/ Different vowel lengths + Phoneme insertions and deletions
sinks	(11) ziŋks (2) ziŋk (1) zints, tsiŋks, ziŋkst, skiŋkz, zinks	Different readings for the grapheme <s>: /z/, /ts/ or /s/ Different readings for the grapheme <n>: /n/ or /ŋ/ Phoneme insertions and deletions
gund	(11) gunt (3) gund (2) gu:nt (1) kult, krunt, bunt	The final consonant devoicing rule is either applied or not Real word reading (bund - confederation) First consonant devoicing /g/ ->/k/ Different vowel lengths + Phoneme replacement
reil	(8) rail, (7) krail, (2) prail, (1) grail, frail	Consonant insertion before the first consonant
grein	(11) grain, (4) krain (1) grai:n, gain, graint, gwain	First consonant devoicing Different vowel lengths + Phoneme insertions and deletions

Note: The numbers in parentheses indicate how many participants read the item with the following pronunciations

Table 12. Monolingual German children (grade 4) reading monosyllabic pseudowords (Exp. 2)

Items	Pronunciations	Comments
quang	(6) kwaŋk	The final consonant devoicing rule is either applied or not
	(5) kwaŋ	Different readings for the grapheme <n>: /n/ or /ŋ/
	(4) kwaŋg	even if it should always be ŋ before g
	(3) kuwaŋk	Different readings for the grapheme <a>: /a/ or /e/
	(1) kwant, praŋk, kweŋk	Phoneme insertions, deletions and replacements
pehl	(11) pe:l, (4) pel	Different vowel lengths + Phonemes insertions and deletions
	(2) perl, fe:l, (1) pe, pfe:l	Real word readings (Perl & Fehl - flaw)
beld	(13) belt, (3) be:lt	The final consonant devoicing rule is either applied or not
	(2) delt	First consonant devoicing /b/ ->/p/
	(1) beld, pelt, berlt, telt	Different vowel lengths + Phoenemes insertion ands replacements
pang	(11) paŋ, (5) paŋk	The final consonant devoicing rule is either applied or not
	(4) pan	Different readings of the grapheme <n>: /n/ or /ŋ/
	(1) klak, paŋg	Phonemes deletions and replacements
seng	(9) zeŋ, zeŋk	The final consonant is either devoiced or not read
	(1) zen, reŋ	Different readings for the grapheme <e>: /e/ or /i/
	(2) ziŋ	Different readings of the grapheme <n>: /n/ or /ŋ/
goos	(10)go:s, (8) gos	Different readings for the grapheme <oo>: /o/ or /u/
	(2) gus, (1) bos, fo:s	Different vowel lengths + Phonemes replacements
drast	(12) drast	First consonant devoicing /d/ ->/t/
	(5) dra:st	Different readings for the grapheme <a>: /a/ or /e/
	(2) trast, (1) tra:st, drest	Different vowel lengths
reuz	(11)roits	Different readings for the diphtong <eu>: /oi/, /au/
	(6) kroits	Different readings for the grapheme <z>: /s/ or /ts/
	(1) kraus, poits, rois	Real word reading (kreuz - cross)
		Phonemes replacements and insertions
kust	(15) kust, (3) ku:st	Real word reading (kunst - art)
	(1) kus, gust, kunst, kuts	Different vowel lengths + Phonemes deletions and replacement
mang	(10) maŋ, (9) maŋk	The final consonant devoicing rule is either applied or not
	(2) maŋg, (1) moŋ	Different readings for the grapheme <a>: /a/ or /o/

Note: The numbers in parentheses indicate how many participants read the item with the following pronunciations

Table 13. List of pseudowords read as real words in Experiment 2 (Bilingual children).

Pseudoword	Word	Nread	Translation	Percentage	Items	grade
roud	round	2/12		16.67	en	two
gule	glue	1/5		20	en	two
roud	round	2/5		40	en	two
traw	straw	1/4		25	en	three
roud	round	1/5		20	en	four
waus	raus	1/14	outside	7.13	de	two
mauch	maus	1/12	mouse	8.32	de	two
gund	grund	1/12	reason	8.32	de	two
krein	klein	1/12	small	8.32	de	two
reuz	kreuz	1/13	cross	7.7	de	two
truck	brücke	1/13	bridge	7.7	de	two
wolz	wolf	1/14	wolf	7.13	de	two
wolz	volt	2/14	voltage	14.29	de	two
polf	pol	2/11	pole	18.17	de	two
gund	grund	1/10	reason	10	de	two
plur	pur	1/12	pure	8.32	de	two
laat	laut	1/13	loud	7.7	de	two
pies	pies	3/11	cakes	27.26	de	two
truck	brücke	1/13	bridge	7.7	de	two
wolz	volt	1/13	voltage	7.7	de	two
wolz	wolf	1/13	wolf	7.7	de	two
laft	lauf	1/5	run	20	de	three
pies	pies	1/4	cakes	25	de	four

Note: Nread indicates the number of participants in the group that read the pseudoword as real words.

Table 14. List of pseudowords read as real words in Experiment 2 (Monolingual children).

Pseudoword	Word	Nread	Translation	Percentage	grade
frur	Frau	1/16	madame	6.25	two
gund	gut	1/19	good	5.25	two
gund	Grund	2/19	reason	10.53	two
jaus	Haus	1/19	house	5.25	two
kast	Gast	1/19	guest	5.25	two
femd	fremd	1/17	foreign	5.89	two
kast	Gast	1/19	guest	5.25	three
kast	krass	1/19	great	5.25	three
krau	grau	1/19	grey	5.25	three
dinn	dünn	1/19	thin	5.25	three
femd	fremd	1/19	foreign	5.25	three
kust	Kunst	2/19	art	10.53	three
reuz	Kreuz	3/17	cross	17.65	three
polf	Golf	1/22	golf	4.54	four
polf	Wolf	1/22	wolf	4.54	four
frur	fur	1/20	for	5	four
gund	Grund	1/22	reason	4.54	four
kast	Gast	1/22	guest	4.54	four
krau	Kraut	1/22	herb	4.54	four
femd	fremd	1/21	foreign	4.75	four
kust	Kunst	1/22	art	4.54	four
reuz	Kreuz	6/22	cross	27.26	four

Note: Nread indicates the number of participants in the group that read the pseudoword as real words.

Table 15. List of pseudowords read as real words in Experiment 3

Pseudoword	Word	Nread	Translation	Percentage	Participants
orrivo	arrivo	4/32	arrival	12.5	Italian
amdio	amido	6/32	starch	18.75	Italian
antobus	autobus	5/32	autobus	15.62	Italian
benge	bende	1/32	bandage	3.13	Italian
calion	camion	1/32	truck	3.13	Italian
clampagne	campagne	1/32	champagne	3.13	Italian
cosputer	computer	1/32	computer	3.13	Italian
fulm	film	1/32	film	3.13	Italian
darage	garage	1/32	garage	3.13	Italian
geseralità	generalità	1/32	generality	3.13	Italian
dapa	papà	1/32	dad	3.13	Italian
restonsabilità	responsabilità	1/32	responsability	3.13	Italian
schepa	schema	1/32	scheme	3.13	Italian
srog	strong	3/32	strong	9.38	Italian
stort	sport	1/32	sport	3.13	Italian
ubiversità	università	1/32	university	3.13	Italian
betro	berto	1/32	male name	3.13	Italian
antobus	autobus	3/29	autobus	10.33	French
clampagne	champagne	1/29	champagne	3.45	French
corfetti	confetti	2/29	confetti	6.9	French

Note: Nread indicates the number of participants in the group that read the pseudoword as real words.

Table 16. French children reading cognate pseudowords (Exp. 3)

Items	Readings	Comments
srog	(15) sʁog, (3) skʁog, (2) strog, stʁog (1) stoʁg, ʃʁog, ʃrog, stʁoʒ, sʁoj, sʁoʒ, sʁo	Different readings of the following graphemes <g> as /ʒ/, /j/ or /g/ and <s> as /s/ or /ʃ/ Consonant insertions between the first two letters
tigamisu	(8) tigamizy, (7) tigamisy, tigamisu (2) tiʒamisu, (1) tiʒamisy, tigami	Different readings of the following graphemes: <g> as /ʒ/ or /g/ and <u> as /u/ or /y/ Syllable removal (tigami)
stort	(18) stoʁ, (4) stort, sto (1) stʁo, stɔrt, stoʁʃ	Different vowel openness in <o>: /o/ or /ɔ/ The final consonant is either read or not
restonsabilité	(10) ʁɜstɔ̃sabilite, (4) ʁɔ̃sabilite (1) ʁəstɔ̃sabili, rɜstɔ̃sabilite, ʁɜstɔ̃nasibilite	Base word reading (responsabilité - responsibility) Syllable removal and addition (ʁəstɔ̃sabili - ʁɜstɔ̃nasibilite)
fratis	(13) fʁatis, fʁati (4) fratis, (1) fatis	Silent final consonant pronounciation Phoneme deletion (fatis). Different readings of grapheme <r>: /r/ or /ʁ/
benge	(20) bɜnʒə, (5) bɜnʤ (1) banʤ, bɑ̃g, bɜnʤɜ, ʒɑ̃g	Different readings of the following graphemes: <g> as /ʒ/,/g/ or /ʤ/ and <en> as /ɛ̃/, /ɑ̃/ or /an/
imcunité	(11) ɛ̃kynite, (4) ɑ̃kynite, (1) ɑ̃komynite, impynite	Different readings for the grapheme <im>: /ɛ̃/ or /ɑ̃/ Base word reading (impunité - impunity)
fulm	(20) fylm, (4) fəlm, fulm (1) flym	Different readings of the grapheme <u>: /y/, /u/ and /ə/ Phoneme inversion (flym)
corfetti	(21) kɔʁfeti, (2) kɔʁf3ʁti, kɔ̃feti, kɔʁfeʁti	Base word reading (confetti) Different vowel openness for <e> :/e/ or /ɜ/

Note: The numbers in parentheses indicate how many participants read the item with the following pronunciations

Table 17. Italian children reading cognate pseudowords (Exp. 3)

Items	Pronunciations	Comments
geseralità	(7) ʤɛsɛralita, (2) ʤɛnɛsalità (1) ʤɛresiralita, ʤɛsarilita, ʤɛsarilita, ʤɛsalalita, ʤɛsɛrabilita, ʤɛsɛlilita, kosɛralita, ʤɛseaita, ʤɛsɛrita, ʤɛnɛ:ralita, ʤjusali:ta	Pronunciations difficulties Real word reading (generalità - generality)
amdio	(11) amdjo, (4) ami:do, (3) ami:dɔ(2) amdi:o, ambi:o, ami:djo (1) umi:dio, amɛ:djo, amo:dio, ami:djɔ, ami:djo, ambjo, ama:djo	Consonant replacements Real word reading (amido - starch) Different wowel lengths and openness
srog	(13) srɔg, (7) strɔg, (2) strɔng (1) srɔng, trɔd, krɔ:g, snɔrd, sgrɔg (1) sɔrʤ, stɔr, grɔ:g, sɔrgrɔg, sgrɔf	Pronounciation difficulties due to uncommon consonant combination (s+r) Real word reading (strong) Phoneme replacements, insertions, inversions, deletions.
actista	(11) akti:sta, (3) akista, (2) akti:ta (1) ak:u:lista, ak:li:sta, aktista, aksi:star, aʧti:sta, aksti:sta, askrita, akʧi:sta	Pronounciation difficulties due to uncommon consonant combinatio (k+t) Phoneme replacements, insertions, inversions, deletions
raternità	(14) ratɛrnita, (2) ratɛrni:ta (1) rateita, raternita, retarnita,reti:mrma, rantɛmita, ratɛnalita, ratɛrmi:ta, ratɛminita, ratɛmita, redat:ernita, ratɛrita	Some syllables are skept Tendency to retourn to a CVCV syllabic structure Different vowel openness: <e>->e / textepsilon Phoneme replacements or inversions
antobus	(11) antɔbus, (7) antɔ:bus (5) autobus (2) anto:bus, (1) autɔbus, antɔsbus	Real word reading (autobus) Different vowel lengths and openness <o>as /o/ or /ɔ/ Phoneme insertion
orrivo	(11) or:i:vɔ, (6) ɔr:i:vɔ, (4) ɔri:vɔ, ar:i:vɔ(3) ori:vɔ, (1) o:r:ivo, aor:i:vɔ	Real word reading (arrivo - arrival) Different vowel lengths and openness <o>as /o/ or /ɔ/
imcunità	(13) iŋkunita, (4) iŋkomunita (1) iŋknunita, iŋkuita, im:unkita, imsunita, iŋgunita, iŋku:nita, komunita, iŋkumita, im:unita	Real word reading (immunità - immunity) Real word reading (comunità - community) Phoneme insertions, deletions and replacements
restonsabilità	(11) rɛstɔnsabilita, (3) rɛspɔnsabilita, (2) rɛstɔnabilita, (1) rɛstɔrnsi:ba, rɛzostambilita, rɛstaubalita, rɛstɔbilita, rɛstɔnsibilita, rɛstɔnansabilita, rɛstɔlisabilita	Real word reading (responsabilità - responsability) Addition or removal of syllables and phonemes Different readings for the grapheme <s>: /s/ or /z/
alsterità	(11) alstɛrita, (6) alsɛrita (1) altrɛri:sta, altɛsenita, alstɛrilita, alstɛralita, altɛri:sta, alastralita, alstɛrnita, alsɛita	Removal of syllables and phonemes Consonant cluster simplification by addying vowels

Note: The numbers in parentheses indicate how many participants read the item with the following pronunciations

Table 18. List of pseudowords read as real words in Experiment 4

nonword	word	nread	percentage	participants
asiet	aside	1/38	2.62	English
deorly	dearly	3/38	7.9	English
dise	dice	4/38	10.53	English

Note: Nread indicates the number of participants in the group that read the pseudoword as real words.

Table 19. German adults reading frequency matched pseudowords (Exp. 4)

Items	Pronunciations	Comments
quaw	(6) kva:f, (2) kua:, (1) ka:, kuab, kuaf, kva:, kva:v	Different readings for the following graphemes: <u>as /v/ or /u/ and <w>as /v/, /f/, /b/ or ø Different vowel lenths + Phoneme deletion (ka:)
spafe	(9) ʃpa:fe, (4) spa:fe, (1) ʃpafe, spa:ve	Different readings for the following graphemes: <s[p]>as /ʃ/ or /s/ and <f> as /f/ or /v/ Different vowel lengths
jotzt	(11) jotst, (1) jøtst, jots, jo:tst, jost	Different readings for the grapheme <o>: /ø/ or /o/ Different vowel lengths + Phoneme inversions (jost)
hog	(10) ho:k, (3) hok, (2) hog	The final consonant devoicing rule is either applied or not Different vowel lengths
mab	(9) ma:p, (5) map, (1) mab	The final consonant devoicing rule is either applied or not Different vowel lengths
stenn	(11) ʃten, (2) sten, (1) ʃte:n	Different readings for the grapheme <s>: /ʃ/ or /s/ Different vowel lengths
peren	(13) pe:ren, (1) pe:re:n, peren	Different vowel lengths
mat	(7) ma:t, (6) mat	Different vowel lengths
üfrer	(8) y:frer, (7) yfrer	Different vowel lengths
gerielst	(10) geri:lst (2) ge:ri:lst, (1) ge:rilt	Different vowel lenghts Phoneme deletion (ge:rilt)

Note: The numbers in parentheses indicate how many participants read the item with the following pronunciations

Table 20. German adults reading dissimilar pseudowords (Exp. 4)

Items	Pronunciations	Comments
vitifu	(6) viti:fu, (3) vi:tifu, (2) fitifu, fi:ti:fu, (1) fiti:fu	Different readings for the grapheme <v>: /v/ or /f/ (in real word reading it would be /f/) Different vowel lengths
zigidu	(8) tsigidu:, (5) tsigi:du, (1) zigidu:, zi:gidu	Different readings for the grapheme <z>: /ts/ or /z/ Different vowel lengths
hevimi	(10) he:vi:mi, (3) he:vi:mi:, (1) he:vimi, he:fi:mi	Different readings for the grapheme <v>: /v/ or /f/ (in real word reading it would be /f/) Different vowel lengths
zulumu	(8) tsulu:mu, (2) tsulumu: (3) tsu:lumu, (1) zulumu:, zu:lu:mu	Different readings for the grapheme <z>: /ts/ or /z/ Different vowel lengths
ledigi	(7) le:digi, (5) le:di:gi, (1) le:digi:, ledigi, legi:gi	Different vowel lengths Phoneme replacement (legi:gi)
lopove	(9) lo:po:ve, (2) lopo:fe, lo:po:fe, (1) lopofe, lo:po:ve:	Different readings for the grapheme <v>: /v/ or /f/ (in real word reading it would be /f/) Different vowel lengths
luxeto	(10) lukseto, (4) lukse:to, (1) luksedo	Different readings for the grapheme <t>: /t/ or /d/ Different vowel lengths
tizafe	(10) ti:tsa:fe, (4) ti:tsafe, (1) ti:za:fe	Different readings for the grapheme <z>: /ts/ or /z/ Different vowel lengths
rimuze	(7) ri:mu:tse, (7) rimu:tse, (1) ri:mu:tse:	Different vowel lengths
galido	(11) gali:do, (2) ga:li:do, (1) ga:lido:, galido:	Different vowel lengths

Note: The numbers in parentheses indicate how many participants read the item with the following pronunciations

Table 21. French adults reading frequency matched pseudowords (Exp. 4)

Items	Pronunciations	Comments
regils	(11) ʁəʒil, (2) regils, ʁəglis (1) ʁegil, ʁeʒis, ʁəglils, ʁəʒəli, ʁə	Different readings for the graphemes: <g>: /g/ or /ʒ/, <e>: /e/ or /ə/ and <r>: /r/ or /ʁ/ The final consonant is either read or not
fremd	(11) fʁɑ̃, (8) fʁɛ̃d, (3) fʁɑ̃d,(1) bʁɛ̃d	Different readings for the grapheme <em>: /ɑ̃/ or /ɛ̃/ Phoneme replacement f -> b and deletion (fʁɑ̃)
brend	(12) bʁɛ̃d, (9) bʁɑ̃ (4) bʁɑ̃d, (2) bʁɛ:n, (1) bʁɛ̃	Different readings for the grapheme <en>: /ɑ̃/, /ɛ̃/ or /ɛ:n/ The final consonant is either read or not
etcere	(18) ɛtsɛʁ, (5) ɛsɛʁ (1) etsɛʁ, ɛtsœʁ, ɛtsɛʁə, ɛts, ɛtsɛtʁ	Different readings for the grapheme <e>: /œ/, /ɛ/, /e/ or /ə/ The final <e> is either read or not Phonemes deletions (ɛts, ɛtsɛʁ) or insertion (ɛtsɛtʁ)
degias	(19) deʒja, (5) deʒjas (1) degias, deja	Different readings for the grapheme <g>: /ʒ/, /g/ or // The final consonant is either read or not
panbing	(21) pɑ̃biŋ, (3) pɑ̃bɛ̃ŋ (2) pɑ̃bɛ̃, (1) pɑ̃bin, pɛ̃bɛ̃	Different readings for the graphemes: <an>: /ɑ̃/ or /ɛ̃/, <in>: /in/, /iŋ/ or /ɛ̃/ and <n>: /n/ or /ŋ/
casment	(21) kasmɑ̃, (4) kazmɑ̃, (1) kasmɛ̃t, kasmɑ̃t, ka	Different readings for the graphemes: <en>: /ɑ̃/ or /ɛ̃/ and <s>: /s/ or /z/ The final consonant is either read or not
fleuiller	(20) flœje, (6) fœje (1) flœle, fyøle	Real words readings: feuiller - come into leaf & fiole - phial Different readings of the grapheme <ill>: /j/ or /l/
tausait	(19) tozɛ, (7) tosɛ (2) tose	Different readings for the graphemes: <e>: /ɛ/ or /e/ and <s>: /s/ or /z/
duntre	(23) dœ̃tʁ, (2) dyntʁ (1) kuntʁ, dœ̃, dɑ̃tʁ	Different readings for the grapheme <un> : /œ̃/, /yn/, /un/. /ɑ̃/ Phoneme replacement /d/ -> /k/

Note: The numbers in parentheses indicate how many participants read the item with the following pronunciations

Table 22. French Adults reading dissimilar pseudowords (Exp. 4)

Items	Pronunciations	Comments
tesusa	(8) tesusa, (7) tezyza, (3) təzyza, (1) tesyza, tesysa, tysusa, tezuza, tezyz	Different readings for the graphemes: <s>: /s or /z/; <u>: /u/ or /y/ and <e>: /e/ or /ə/ Phoneme deletion (tezyz)
ledigi	(10) lədiʒi, (11) lediʒi, (5) ledigi, (1) legiʒi, lədigi	Different readings for the graphemes: <g>: /g/ or /ʒ/ and <e>: /e/ or /ə/ Phoneme replacement
buleba	(14) byleba, (7) byləba, (3) bylba, (2) buleba, (1) belyla, pyleba	Different readings for the graphemes: <u>: /u/ or /y/ and <e>: /e/ or /ə/ Phoneme deletion (bylba) and devoicing /b/ ->/p/
zigidu	(13) zigidy, (9) ziʒidy, (4) zigidu, (1) ziʒidu, ziʒibyty	Different readings for the graphemes: <g>: /g/ or /ʒ/ and <u>/u/ or /y/ Addition of syllable (ziʤibyty)
gobujo	(21) gɔbyʒo, (1) gɔbuʤo, gɔbyʒœ, gɔbyʒy, bɔʒyʒy, gɔbyjʤo, gogɔbuʒo, go	Different readings for the graphemes: <g>: /g/, /ʒ/ or /ʤ/, <u>/u/ or /y/ and <o>: /o/ or /ɔ/ Addition of syllable (gogɔbuʒo) + Mispronunciation (go)
luxeto	(21) lykseto, (5) lyksəto (1) lyksito, lykzeto	Different readings for the graphemes: <x>: /ks/ or /kz/ and <e>: /e/, /i/ or /ə/
fuduja	(24) fydyʒa (1) fuduʤa, fedyʤa, fydja, fydyʤa	Different readings for the graphemes: <g>: /j/, /ʒ/ or /ʤ/ and <u>/u/ or /y/ Phoneme replacement /u/ ->/e/ and deletion (fydja)
zulumu	(24), zylymy, (3) zulumu (1) zybymy	Different readings for the grapheme <u>: /u/ or /y/ Phoneme replacement /l/ ->/b/
mumade	(25) mymad, (1) mumad, mumade, myman	Different readings for the grapheme <u>: /u/ or /y/ The final <e>is either read or not Phoneme replacement /d/ ->/n/
vitifu	(25) vitify, (2) vitivy, vitifu (1) fitiflu	Different readings for the grapheme <u>: /u/ or /y/ Consonant assimilation and Phoneme insertion (fitiflu)

Note: The numbers in parentheses indicate how many participants read the item with the following pronunciations

Table 23. English adults reading frequency matched pseudowords (Exp. 4)

Items	Pronunciations	Comments
asiet	(13) asiet, (1) ast, asit, asite, assist, aset, iset	Different readings for the grapheme <ie>: /i/ or /e/ Phonemes deletion (ast), replacement (a -> i) Real word reading (assist)
aecrer	(4) aesrer, (2) aser, asra, (3) asrer, (1) ases, aseri, aeksrer, asresser, asekrer, aserer, askrer	Phonemes and syllables deletions (ases), and phoneme insertions (asekrer, asresser)
strylture	(10) strailture, (4) strailiture, (1) strai, strailetto, straili, straiturle	Phonemes and syllables addition (strailiture), deletions (strai) and inversions (straiturle)
watheet	(9) wai:t, (4) wati:t, (1) wahti:st, waet, wati:, atpi:t	Different readings for the following graphemes: <ee> as /i:/ or /e/ and <th> as /t/ or //
dousse	(13) dousse, (4) dosse, (1) doussi:s, deuse	Different readings for the following graphemes: <ou> as /ou/, /o/ or /eu/ and <e> as /e/ or /i/
deorly	(17) deorli, (1) dori, dearli	Different readings for the grapheme <eo>: /eo/, /ea/ or /o/ Phoneme deletions (dori)
rebube	(16) rebube, (2) rebubi:, (1) rubab	Different readings for the grapheme <e>: /e/ or /i:/ Phoneme deletions (rubab) and relacements (e -> u; u -> a)
dise	(10) dise, (4) diese, (3) diss (1) daese	Different readings for the grapheme <i>: /i/, /ie/ and /ae/ Phoneme deletion and consonant doubling (diss)
speached	(18) spi:ched, (1) spleched	Different readings for the grapheme <ea>: /i/ or /e/ Phoneme addition (spleched)
vuing	(15) vuing, (1) voning, vling, vigged	Different readings for the grapheme <u>: /u/ or /o/ Phonemes addition (voning, vling, vigged)

Note: The numbers in parentheses indicate how many participants read the item with the following pronunciations

Table 24. English adults reading dissimilar pseudowords (Exp. 4)

Items	Pronunciations	Comments
buleba	(10) buleba, (1) buleaba, bubala, blubabd, bulba, bulejub, bubela, buliba, beleba	Phonemes deletions (bulba), insertions (blubabd) and inversions (bubela, bubala)
tizafe	(10) tizafe, (5) tizaf, (1) tizaife, zafafe, tizave	Different readings for the grapheme <f>: /f/ or /v/ Phoneme deletion (tizaf, zafafe)
ledigi	(18) ledigi, (1) ledgili, ledgi	Phonemes deletion (ledgi) and addition (ledgili)
gobujo	(14) gobujo, (1) goboju, gobuju, goguba	Different readings for the grapheme <o>: /o/ or /u/ Syllable reduplication (goguba)
luxeto	(12) lukseto, (4) luksetto, (1) leksuto, luzejto	Consonants doubling (luksetto) Phonemes replacements (e ->u; ks ->z)
rimuze	(7) rimuse, (4) rimusi, (1) rimuzu, remuse, rismuse	Different readings for the following graphemes: <e> as /i/ or /e/ and <z> as /z/ or /s/ Phoneme addition (rismuse) and replacement (u -> e)
pazile	(9) pazille, (7) pazile, (1) pazit	Consonant doubling (pazille) Phoneme deletion and replacement (pazit)
zigidu	(13) zigidu, (1) zigido, zizidu, zigudu, zigidiu, zikidi	Different readings for the following graphemes: <u> as /u/ or /o/ and <g> as /g/ or /k/ Syllable reduplication (zizidu) and vowels assimilation (zikidi)
zulumu	(15) zulumu, (1) zumulu, zlumu, zulumi	Syllables inversion (zumulu) Phoneme deletion (zlumu) and replacement (u ->i)
hevimi	(14) hevimi, (1) hemivi, hevim, hevilimi, hevini	Syllables inversion (hemivi) and addition (hevilimi) Phoneme deletion (hevim) and replacement (m ->n)

Note: The numbers in parentheses indicate how many participants read the item with the following pronunciations

Study 2

The role of Syllables and Morphemes in Silent Reading: an Eye-Tracking Study

Elisabetta De Simone

Kristina Moll

Lisa Feldmann

Xenia Schmalz

Elisabeth Beyersmann

Study published in the *Quarterly Journal of Experimental Psychology*:

De Simone, E., Moll, K., Feldmann, L., Schmalz, X., & Beyersmann, E. (2023). The role of syllables and morphemes in silent reading: An eye-tracking study. Quarterly Journal of Experimental Psychology, 17470218231160638. https://doi.org/10.1177/17470218231160638

QJEP

Original Article

Quarterly Journal of Experimental Psychology
1–21

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/17470218231160638
qjep.sagepub.com
SAGE

The role of syllables and morphemes in silent reading: An eye-tracking study

Elisabetta De Simone[1,2], Kristina Moll[2], Lisa Feldmann[2], Xenia Schmalz[2] and Elisabeth Beyersmann[1]

Abstract

German skilled readers have been found to engage in morphological and syllable-based processing in visual word recognition. However, the relative reliance on syllables and morphemes in reading multi-syllabic complex words is still unresolved. This study aimed to unveil which of these sublexical units are the preferred units of reading by employing eye-tracking technology. Participants silently read sentences while their eye-movements were recorded. Words were visually marked using colour alternation (Experiment 1) or hyphenation (Experiment 2)—at syllable boundary (e.g., *Kir-schen*), at morpheme boundary (e.g., *Kirsch-en*), or within the units themselves (e.g., *Ki-rschen*). A control condition without disruptions was used as a baseline (e.g., *Kirschen*). The results of Experiment 1 showed that eye-movements were not modulated by colour alternations. The results of Experiment 2 indicated that hyphens disrupting syllables had a larger inhibitory effect on reading times than hyphens disrupting morphemes, suggesting that eye-movements in German skilled readers are more influenced by syllabic than morphological structure.

Keywords

Syllabic processing; morphological processing; eye-tracking; visual segmentation

Received: 29 June 2022; revised: 8 November 2022; accepted: 7 December 2022

Many decades of reading research have shown that reading involves more than just processing orthographic whole-word representations. Rather than simply relying on letter-by-letter decoding, expert readers reliably and automatically read words by forming associations in memory between embedded sublexical units (i.e., syllables and morphemes) and their corresponding lexical representations (Ehri, 1995). In fact, this trend is already found in children's first years of reading instruction (Colé et al., 2012; Häikiö et al., 2011) and spills over to adulthood, with evidence of both syllable and morpheme processing gathered in several languages across tasks (Colé et al., 1999; Conrad et al., 2009; Dawson et al., 2018). The goal of the present eye-tracking study was to shed further light on the mechanisms involved in sublexical reading in a shallow orthography, namely, German. In particular, we sought to directly compare the relative relevance of syllable and morpheme processing by monitoring participants' eye-movements. Prior research has typically focused either on bi-syllabic, monomorphemic words (Conrad et al., 2011) or on multi-morphemic words (Bertram et al., 2004), but studies that have directly compared morpheme and syllable processing are more scarce (Alvarez et al., 2001; Colé et al., 2012; Domínguez et al., 2006; Häikiö & Vainio, 2018; Hasenäcker & Schroeder, 2017). As such, the interplay between syllable and morpheme processing within multi-syllabic, multi-morphemic words (e.g., *formation*, a word with three syllables and two morphemes) is not well understood.

Reading multi-syllabic words

Research carried out on different languages suggests that syllables are not only relevant in speech production and comprehension (Cholin et al., 2004) but also represent

[1]School of Psychological Sciences and Macquarie Centre for Reading, Macquarie University, Sydney, NSW, Australia
[2]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Hospital of the Ludwig-Maximilians-University (LMU) Munich, Munich, Germany

Corresponding author:
Elisabetta De Simone, School of Psychological Sciences and Macquarie Centre for Reading, Macquarie University, Australian Hearing Hub, 16 University Avenue, Sydney, NSW 2109, Australia.
Email: desimone.elisabetta@hdr.mq.edu.au

relevant processing units in visual word recognition (e.g., Conrad & Jacobs, 2004). For example, Ferrand et al. (1996, 1997) found that primes with the same syllabic structure as the target word (cv—CV) produced facilitatory effects in French in a naming task when the prime shared the first syllable with the target, and the syllable had clear boundaries (bal%%%%—BAL.CON), than when preceded by primes containing one letter more (bal%%%%—BA.LADE) or less (ba%%%%—BAL.CON) than the first syllable. Syllable effects have been reported in French (e.g., Chetail & Mathey, 2009a, 2009b), Spanish (e.g., Carreiras & Perea, 2002; Perea & Carreiras, 1998), and German (Conrad & Jacobs, 2004; Stenneken et al., 2007).

Converging evidence of the early involvement of phonological processing in visual word recognition also comes from eye-tracking studies using silent reading tasks (Fitzsimmons & Drieghe, 2011; Inhoff & Topolski, 1994; Pollatsek et al., 1992; Rayner et al., 1998; Sparrow & Miellet, 2002). For example, Pollatsek et al. (1992) found a parafoveal preview benefit for homophones (*sent* as preview for *cent*) but not for visually similar words (*rent* as a preview for *cent*). Ashby and Rayner (2004) investigated whether syllabically congruent primes could aid visual word recognition during silent English reading. They discovered that when the preview was syllabically congruent (target: *de-vice*, preview: *de_πxw*), readers' first fixation durations on the target word were shorter than when it was incongruent (*dev_πx*), although the proportion of orthographic overlap between the preview and the target was greater in the incongruent condition. More recently, Hawelka et al. (2013) used eye-tracking to assess whether inhibitory effects of first syllable frequency found in lexical decision tasks are generalisable to natural reading. Using the multi-syllabic items of the German Potsdam Sentence Corpus (Kliegl et al., 2004) as target words, the authors found inhibitory effects that resulted in longer first fixation on multi-syllabic words starting with high-frequent first syllables. This effect was interpreted as evidence of prelexical phonological processing. The authors argued that syllabic representations served as "access units" to the mental lexicon, which also tend to be activated in visual word recognition (see also Stenneken et al., 2007).

The prior evidence for the important role of syllables in silent reading has led to the assumption that syllables are represented as prelexical units of reading at an intermediate stage between letter perception and word recognition. For example, Mathey et al. (2006) extended the interaction activation model (IA model) of McClelland and Rumelhart (1981) by including syllables (the IAS model). As a result, the model accounts for both syllable effects and the influence of orthographic information (see also Ans et al., 1998; Grainger & Ziegler, 2011). Similarly, Conrad et al. (2010) extended the MROM model (MROM-S) by adding syllabic representations units in a separate route, which are connected to both letter and word representations. Syllables activate words that contain them in the initial syllabic position, so that the processing of the target is hampered by lateral inhibition caused by the coactivated syllabic neighbours, with the competition being more severe when the syllable is of high frequency. This way, the model explains the inhibitory effect of syllable frequency on lexical decision reported several languages including German (e.g., Conrad & Jacobs, 2004).

Reading multi-morphemic words

Evidence for morphological processing in skilled reading is abundant (for reviews, see Amenta & Crepaldi, 2012; Rastle & Davis, 2008). For example, in a lexical decision study in English, Ji et al. (2011) reported that transparent and opaque compound words are processed faster than frequency-matched monomorphemic words. The authors argued that morphological decomposition is beneficial because the activation of the individual constituents facilitates the recognition of the whole word. Masked priming studies have also unveiled facilitatory priming effects for truly suffixed primes (*player—PLAY*) and pseudo-suffixed primes (*mother—MOTH*), relative to non-morphological controls (*cashew-CASH*), with prime displays as brief as 42 ms (e.g., Rastle et al., 2004). This shows that at the early stages of visual word recognition, both suffixed (*play + er*) and pseudo-suffixed (*moth + er*) words are swiftly decomposed into their morpho-orthographic constituents (Beyersmann et al., 2016; Diependaele et al., 2009; Longtin et al., 2003; Rastle et al., 2004). These results were consistently found in several orthographies, such as French (e.g., Grainger et al., 1991), English (e.g., Beyersmann et al., 2012), and Russian (e.g., Kazanina et al., 2008).

The scope of morphological effects is not only limited to priming studies and lexical decision tasks but also extends to more natural reading settings. Several eye-tracking studies have found that morphemes are recognised and rapidly processed during sentence reading. For instance, Deutsch et al. (2003), using a boundary-contingent paradigm, found a morphological preview benefit effect for Hebrew speakers. When readers parafoveally processed a morphologically related word, target words were processed faster compared with an orthographic control condition, as reflected by early processing measures such as first fixation durations and gaze duration.

Other eye-tracking studies have instead turned to frequency as a diagnostic tool to examine the influence of morphological processing in complex word processing. These studies typically manipulate the frequency of the entire word as well as specific morphological components, and their effects on early and late measures like first, second, or third fixation durations and gaze duration. This methodology is based on the premise that if the frequency of a single morpheme influences fixation durations, then this will demonstrate morphological decomposition in reading complex words (Kuperman et al., 2010; Pollatsek et al., 2000). For instance, Pollatsek et al. (2000), using Finnish compound words, reported a second constituent frequency effect on second fixation duration and gaze duration (as did Juhasz et al., 2003 in English), thus building on Hyönä and Pollatsek's (1998) findings of first constituent frequency affecting first fixation duration, second fixation duration and gaze duration. In addition, the same group of authors reported a whole-word frequency effect on gaze duration (Experiment 2), indicating that the identification of compound words involves simultaneous processing of morphological constituents and whole-word representations. Further evidence for the important role of constituent frequency on eye-movements comes from a study by Bertram and Hyönä (2003), showing that long Finnish compound words with a high-frequency first constituent elicited shorter gaze durations, shorter first fixation durations, and fewer third fixations than long compounds with a low-frequency first constituent. In addition, the first constituent frequency effect was more pronounced in long compounds than in short ones (Experiment 1), suggesting that short compounds are more likely to be processed as wholes (Experiment 2). This finding was also confirmed by Kuperman et al. (2010) using derived complex words in Dutch. They found that words with shorter suffixes exhibited a stronger whole-word frequency effect on reading times compared with suffix frequency, thus pointing to the important role of suffix length in morphological processing.

An appreciation of the influence of morphological information in reading also comes from eye-tracking studies examining landing positions. Normally, the eyes land in the middle of the word or slightly to the left of it, a phenomenon called *optimal viewing position* (O'Regan, 1992; Rayner, 1979), from which a word can be processed the fastest. However, it has been demonstrated that when words are morphologically complex, the eye fixations land closer to the beginning of the words (Hyönä et al., 2018; Yan et al., 2014), suggesting that readers are able to pick up words' morphological structure during parafoveal processing, which helps them adjusting their saccade programming accordingly.

Over the past decades, many morphological processing theories have emerged from the field of visual word recognition (Diependaele et al., 2009; Duñabeitia et al., 2007; Grainger et al., 1991; Grainger & Ziegler, 2011; Longtin et al., 2003; Rastle, 2019; Rastle & Davis, 2008; Rastle et al., 2004). These theories make different assumptions with respect to the time-course of morphological processing during reading, with some predicting that the early stages of morphological processing are semantically "*blind*" (Beyersmann et al., 2016; Longtin et al., 2003; Rastle et al., 2004), whereas others assume that semantics do already assert an influence on morphological processing during the initial stages of complex word recognition (Feldman et al., 2009, 2015). However, they all agree that skilled readers are experts at rapidly extracting morphological information from print. More recently, it has been proposed that the activation of edge-aligned embedded words posits one of the key ingredients in the analysis of multi-morphemic words as well as in children's reading acquisition (for more detail, see the *word and affix* model by Beyersmann & Grainger, 2022 and Grainger & Beyersmann, 2017).

Syllables versus morphemes

The above summary shows that silent reading is clearly modulated by the syllabic and morphemic structure of words; however, only few studies have directly compared the salience of syllables and morphemes in this process. In many Indo-European languages, syllabic and morphemic structure do not always overlap (e.g., *far-mer* vs. *farm-er*: see Alvarez et al., 2001; Domínguez et al., 2006), but how do readers solve this challenging conflict between reading units? Are syllables and morphemes equally salient during silent reading?

Fracasso et al. (2016) showed that while phonological and morphological awareness are predictors of reading comprehension in adults, the latter was also a unique predictor of vocabulary skills, as the "ability to break up morphologically complex words into their morphemic constituents enables a reader to use their knowledge of the meanings of the base morpheme and suffix to infer meanings of unfamiliar, morphologically complex words" (Fracasso et al., 2016, p. 147). For example, via access to the individual morphemes in *paleo-geo-graph-er*, readers may be able to partially derive word meaning, that is, a person (*-er*) working in the field of ancient (*paleo-*) geography. Thus, compared with syllable-based reading, morpheme-based reading has the advantage of breaking down the word's meaning into meaningful chunks (Bhattacharya, 2020; Goodwin et al., 2013; Kearns & Whaley, 2019).

Studies comparing morphological processing in children, adolescents, and adults have suggested that morphological decomposition becomes more important and automatised throughout reading acquisition, such that morphological effects are more pronounced in adults compared with the younger age groups (e.g., Beyersmann et al., 2012; Dawson et al., 2018; Schiff et al., 2012). Although this indicates that skilled readers become increasingly skilled at parsing complex words into morphemes as they become more fluent readers, it does not undermine the possibility that syllable-based parsing is equally important.

Indeed, syllable-based reading has its own advantages of drawing from preexistent oral knowledge (Perfetti et al., 1992) and retaining identified words in the phonological loop of short-term memory (Besner, 1987; Bruck et al., 1995), which is useful for sentence-processing. In this regard, the few studies that directly compared syllable and morpheme processing focused on developing readers. One such study was conducted by Colé et al. (2012) with French second and third graders. The stimuli employed in their naming task (Experiment 2) were segmented using a space at the syllable boundary (*den tiste*) or morpheme boundary (*dent iste*), with the assumption that recognition will be faster if the visual manipulation conformed to the units activated during written word identification. Results showed that word recognition times were comparable between the morphemes and syllables-spaced conditions, suggesting that both morpheme and syllable-based reading affected reading fluency to the same degree.

In contrast to Colé et al.'s (2012) findings from French primary schoolers, Häikiö and Vainio (2018) reported differences between syllable and morpheme processing in Finnish first- and second-graders. The authors examined the processing of bimorphemic targets words embedded in sentences using eye-tracking. In half of the target words the last syllable boundary coincided with the (inflectional) morpheme boundary. Using hyphens, words were divided into syllable-congruent and syllable-incongruent (which were also morpheme-congruent in half of the cases) conditions. Second graders spent significantly more time fixating hyphenated than non-hyphenated words, an effect that was less noticeable in first graders. However, when the syllable-incongruent condition overlapped with the morpheme-congruent condition (MCC), neither age group's gaze durations increased, suggesting that Finnish beginning readers were not slowed down by the broken syllables or the hyphenation, as long as hyphens segmented words into morphemes. The authors interpreted this finding as an indication that participants relied more on morphemes than on syllables in their reading.

A similar paradigm, using lexical decision, was employed by Hasenäcker and Schroeder (2017) in German second and fourth graders, and adults. Multi-syllabic monomorphemic and multi-morphemic words and multi-syllabic pseudo-affixed nonwords were segmented into syllable-congruent and incongruent items using a colon (:). The syllable-incongruent items overlapped with morphemes boundaries. Results showed that second graders were faster in both word identification and pseudoword rejection when the disruption was syllable congruent. However, in fourth graders the syllable-incongruent/morpheme-congruent manipulation impeded the rejection of multi-morphemic pseudowords, suggesting that syllable-based reading was more pronounced in Grade 2, but morpheme-based reading was predominant in Grade 4. Lexical decision responses in adults were not hindered or facilitated by any of the manipulations, which may however have been ascribable to the excessive simplicity of their items (taken from the childLex corpus of Schroeder et al., 2015).

In sum, although the prior evidence shows that both children and adults use syllables and morphemes in their reading, it remains less clear whether the reading system gives preference to the analysis of syllabic or morphemic structure and how potential confounds between syllable- and morpheme-congruency are resolved.

The present study

To address the question of how skilled readers process words with deviating syllable and morpheme-structures, the current eye-tracking study used a method to highlight syllable and morpheme structure in silent reading. The goal was to explore syllable- and morpheme-congruency effects in sentence reading while monitoring participants' eye-movements, to test if readers find it easier to read text where syllables and morphemes are visually marked. The benefits of eye-tracking are not limited to study online cognitive processes in an ecological reading setting (without the constraint of having participants perform an unfamiliar task; Rayner et al., 1998) but can also be used to tease apart the early versus late processes involved in reading. The goal was to build on prior findings by directly comparing the processing of syllables and morphemes in skilled readers of German, a morphologically rich (Juola, 1998; Kettunen, 2014; Mousikou et al., 2020), syllabically complex (Adsett & Marchand, 2010; Seymour et al., 2003; Stenneken et al., 2007), and orthographically transparent language (Borleffs et al., 2017; De Simone et al., 2021). In the first experiment, syllables and morphemes were highlighted using colours, whereas the second experiment used hyphenation to segment words into their respective reading units.

Experiment 1

In Experiment 1, we presented words in which morphemes or syllables were highlighted using colours, based on a method that is commonly used in German reading instruction ("Silbenmethode" [syllable method], where reading books for children in early school grades mark each alternate syllable in a different colour). The syllable method relies on the assumptions that colour information helps to identify objects and better remember information (Gegenfurtner & Rieger, 2000; Tanaka et al., 2001). Indeed, colour similarities make it easier to aggregate an item formed by many elements, while colour disparities help separate stimuli into multiple things (see Goldfarb & Treisman, 2011). However, only few studies have applied the method to examine the cognitive mechanisms of reading in adults. As a result, the cognitive underpinnings of the "Silbenmethode" are still little understood.

Perhaps, one of the most informative studies in this regard is one by Carreiras et al. (2005), in which the authors examined syllabic effects in Spanish-speaking adults using event-related potentials, while participants performed a lexical decision task. The authors used colours to segment words which varied in frequency, as well as pseudowords such that the colours either did or did not coincide with the syllable structure. In a baseline condition, only one colour was used. No congruency effects were found in the behavioural measures (for related evidence from the transposed letter similarity effect, see Marcet et al., 2019). However, the colour manipulation led to a temporal and spatial dissociation in the ERPs in the P200 time window for pseudowords and low-frequency words, with an amplitude increase for the colour-syllable incongruent condition compared with the colour-syllable congruent and baseline conditions. This suggests that the syllabic structure of low-frequency words and pseudowords is processed during the early stages of visual word recognition (see also Carreiras et al., 2009 for similar conclusions). Crucially, Carreiras et al. (2005) reported no facilitatory effect of syllable colouring (i.e., no differences were found between the baseline and the congruent stimuli in the ERPs or RTs) and therefore does not provide direct support of the hypothesis that the use of syllable colouring supports reading. However, given that Spanish has a simpler syllable structure compared with German (Seymour et al., 2003), it is unclear whether these results are generalisable across languages.

Indeed, preliminary eye-tracking evidence from Chinese (Zhou et al., 2018, 2019), a typically unspaced script, shows that Chinese silent reading is facilitated when words are alternately coloured, such that between-word boundaries are explicitly signalled. Colouring influenced landing position, showing how this method helped L1 Chinese speakers (Zhou et al., 2018) and L2 Chinese learners (Zhou et al., 2019) to optimise their eye fixations. However, colouring segmentation might not have the same impact on reading speed. While alternating colours at word boundaries increased the reading fluency of skilled Chinese readers when reading aloud difficult, technical texts with unfamiliar words (Perea & Wang, 2017), the same effect was not found under normal circumstances, that is, when reading more common texts with familiar words (Perea & Wang, 2017). In fact, it seems that the positive impact of colour segmentation on reading speed decreases with age: in an eye-tracking study, Song et al. (2021) found colour facilitation effects in multi-chromatic compared with mono-chromatic sentence-processing in Grades 2–3 children (as did Perea & Wang, 2017, Experiment 3), but not in Grades 4–5 children.

Similar findings with Grade 2 children have been reported also in alphabetic scripts, with within-word colour segmentation. For example, Lopes and Barrera (2019) investigated syllable colouring in Grade 2 Brazilian-Portuguese speaking children performing an isolated word reading task. They found that highlighting syllables through the use of colours had a positive effect on good readers when reading irregular words, and on poor readers when reading regular and irregular words. Similar findings in French speaking children of the same age were also found by Chetail and Mathey (2009b) using a colour lexical decision task. However, in an eye-tracking study with Finnish beginning readers (Grade 1–2), Häikiö et al. (2015) found no evidence that alternating colours at the syllable boundary affects reading speed, compared with a control condition where no visual cues where given.

Research on using the colouring method with morphemes is scarcer. Ji et al.'s (2011) lexical decision study (Experiment 5) employed a colour contrast (red/black) to encourage the decomposition of English compound words into morphemes, with the assumption that the display manipulation would facilitate morphological parsing, and hence support the retrieval of the compound words' meaning. However, the colour manipulation did not influence response time or accuracy scores, and if anything inhibited the processing advantage of opaque words compared with monomorphemic words, potentially because in the case of opaque words, this colouring supported a computed meaning that was inconsistent with its stored, conventional meaning.

Other colour segmentation studies have used illusory conjunctions (ICs) first described by Treisman and Schmidt (1982). ICs have been defined as a type of errors that happen in the perceptual binding of proximal elements in a given stimulus, when attention is deviated or diverted (Henderson & McClelland, 2020). In reading research, ICs have been used to "determine the nature of the sublexical units that are automatically perceived at a perceptual level of word processing" (Doignon-Camus & Zagar, 2005), first employed in English by Prinzmetal et al. (1986). The procedure is normally paired with a letter detection task, which serves to divert the participant's attention. In a stimulus string divided in two colours, participants are asked to (a) detect the target letter; and (b) report in which colour the target letter was presented. For example, participants are asked to determine if the target letter was present in a given word and what colour it was. The target letter (e.g., letter "v" in *anvil*) is presented either in a unit-congruent condition (syllables, in the example): anvil, or in a unit-incongruent condition: anvil. Prinzmetal et al. (1986) reported syllable preservation errors, where participants reported that a given target letter (v) appeared in the same colour as the rest of its syllable unit (*vil*), including within the incongruent condition (*anvil*), suggesting that syllables were automatically activated during visual word processing. The authors reported similar preservation errors in polymorphemic words, indicating that both syllables and morphemes represent functional units in the visual analysis of words (for similar findings from French, see Doignon-Camus et al., 2009; Doignon-Camus & Zagar, 2005).

In sum, the review of the literature shows that it is still unclear whether or not segmentation by colouring does indeed expedite word processing. Although several studies have failed to provide evidence for a facilitatory role of syllable colouring on reading fluency (e.g., Carreiras et al., 2005, 2009; Häikiö et al., 2015), the Silbenmethode still continues to be used extensively in German reading instruction and therefore calls for a more thorough investigation within the German language in particular. This study used eye-tracking to investigate the effectiveness of the colouring method in German by directly comparing syllable and morpheme processing in an ecologically valid way. Participants read sentences where colours were either congruent or incongruent with the embedded syllabic and morphemic structures, relative to a black-coloured control. We hypothesised that if any facilitation effects were to be found (due to highlighting relevant reading units), then the syllable-congruent and the MCCs would be read faster than the control condition. Moreover, if morphological information is more relevant in reading than syllabic information because of its direct link with semantics (as argued by Kearns & Whaley, 2019), participants would spend less time fixating and make fewer regressions to words in the MCC compared with the syllable-congruent condition. As facilitation could also occur through an optimised landing position variability across conditions, we predicted that coloured units may change the landing position of the eye, compared with where no unit is evident (a preregistration of these hypothesis can be found at https://osf.io/csja8).

Method

Participants. The desired sample size (40) was preregistered and determined a priori using the package SimR (Green & MacLeod, 2016) in the R computing environment (R Core Team, 2021) to calculate power for linear mixed model. Using pilot data collected from two volunteers, we ran simulations for each hypothesis. Our simulations predicted that to obtain a power of 80% (d = 0.5), we needed a minimum sample of 30 participants. Based on previous studies in the reading and eye-movements literature (Rayner, 1998), it was decided to increase the calculated sample size to 40 typically reading adults.

In total, 42 native German speakers (35 females, 7 males) participated for monetary reimbursement. Two participants who scored poorly on the Standardised Reading Fluency Test II (SLRT-II) word reading test (Moll & Landerl, 2010) were excluded (i.e., <16 percentile, corresponding to at least one standard deviation below the mean of the adult population in the standardised reading test), following the German clinical diagnostic guidelines for dyslexia (Galuschka & Schulte-Körne, 2016). The remaining 40 participants were between 20 and 53 years old (M = 29.3; SD = 7.59) and had normal or corrected-to-normal vision. The study was approved by the ethics committee of the Hospital of the Ludwig–Maximilians–University (LMU) and conforms with the ethical principles of the Declaration of Helsinki. Prior to participating in the study, participants provided written, informed consent.

Table 1. Target word structure. Percentage (and number of words) of mono-syllabic and poly-syllabic words; and mono-morphemic and poly-morphemic words in the Potsdam sentence Corpus.

Syllables	%	Morphemes	%
Mono-syllabic	47.02 (535)	Mono-morphemic	61.86 (704)
Poly-syllabic	52.98 (603)	Poly-morphemic	38.14 (434)
-2 syllables	71.47 (431)	-2 morphemes	82.02 (356)
-3 syllables	19.40 (117)	-3 morphemes	15.89 (69)
-4 syllables	7.96 (48)	-4 morphemes	2.03 (9)
-5 syllables	0.82 (5)		
-6 syllables	0.33 (2)		

Materials. Adults read 140 sentences from the Potsdam Sentence Corpus (Kliegl et al., 2004, 2006). Sentences included 5–11 words (M = 7.9, SD = 1.4), with logarithmic word frequencies averaging to M = 2.1 (SD = 1.3). Overall, there were a total of 1,138 words in the corpus (see Table 1 for words' statistics).

The sentences were randomly divided into five conditions of 28 sentences each, which were counterbalanced across participants. Each condition corresponded to a different colour manipulation: syllables-congruent; syllables-incongruent; morphemes-congruent; morphemes-incongruent condition; black-coloured. Moreover, we randomised the stimuli presentation in two additional ways: the order of blocks (conditions were not presented in a fixed order), and the order of sentences.

In the syllables-congruent condition (SCC), syllables were alternately coloured in blue and green, and the colour changed at the syllable boundary or between monosyllabic words. Conversely, in the syllables-incongruent condition (SIC), the syllable was disrupted by moving the colour alternation either to the left or to the right of the syllable boundary. The reason for this manipulation was twofold: (a) to rule out possible facilitation or inhibition effects due to the mere colour alternation and (b) to check whether reading was impaired when the integrity of the unit was broken. The same number of coloured units of the syllable-congruent condition was also kept, to avoid a possible increase in saccades, which could have resulted in higher reading times. Moreover, we made sure not to break any multi-letter graphemes (*Schüssel*, but not *Schüssel*, where *sch* is a multi-letter grapheme in German corresponding to the phoneme /ʃ/).

For the next condition, morphemes were alternately coloured in blue and green (MCC, see Table 2). Affixes and morphemic stem constituents (within both affixed and compound words) were coloured in this condition. We separated inflectional morphemes: past tense (*stellte*, "put": verb root + tense and person), gender (*eine*, "a": base indefinite article + gender), and number (*Kirschen*, "cherries": stem + number). Derivational morphemes were separated from their stems, as in *Häuschen* (little home) and compound words were divided in their

Table 2. Experimental design and stimuli examples (Experiment 1).

Condition	Acronym	Stimuli Example
Syllable-Congruent	SCC	*Laura stellte eine Schüssel Kirschen auf den Tisch*
Syllable-Incongruent	SIC	*Laura stellte eine Schüssel Kirschen auf den Tisch*
Morpheme-Congruent	MCC	*Laura stellte eine Schüssel Kirschen auf den Tisch*
Morpheme-Incongruent	MIC	*Laura stellte eine Schüssel Kirschen auf den Tisch*
Control	CTRL	*Laura stellte eine Schüssel Kirschen auf den Tisch*

constituents, as in *Groβvater* (grandfather). Similar to syllables, we created a group of sentences where the morpheme unit was disrupted (morpheme-incongruent condition [MIC]) by moving the morpheme boundary either to the left or the right, without separating multi-letter graphemes. We applied the same principles used in the syllables-incongruent condition.

Finally, a fifth group of sentences was composed of black-coloured sentences (CTRL), without any colour alternation. This condition served as a baseline condition. Stimuli examples for each condition are provided in Table 2.

To make sure participants read the sentences carefully, we created a comprehension test in the form of multiple-choice questions. Questions would appear after a random interval of sentences, and they would always refer to the previously shown sentence. In total, participants answered 18 questions. All participants scored more than 80% of correct answers in the multiple-choice questions. The full list of materials is available in the following online repository: https://osf.io/w4rsm/.

Apparatus. Eye-movements were recorded using an Eye-Link 1000 Plus Desktop Mount eye-tracker (SR Research, Toronto, Canada) in head-stabilised mode. Participants were seated in front of a 15.6-in. monitor (120 Hz refresh rate, 1280 × 960 resolution) at a viewing distance of 65 cm. Stimuli were presented with an uppercase letter height of about 0.62° of visual angle. A 9-point calibration cycle at the beginning and after each break was used to ensure a spatial resolution of less than 0.5° of visual angle.

The experiment was controlled with Experiment Builder software (SR Research, version 1.10.1630). Sentences were presented in Courier New Bold, 30 pt. font, and projected in full window. One sentence was presented per trial, vertically centred on the screen, on a white background.

Procedure. The sessions took place individually in a silent room. Prior to the eye-tracking experiment, an SLRT II (Moll & Landerl, 2010) was administered where participants had to read aloud a list of words and pseudowords as quickly and as accurately as possible in 1-min time.

Participants were then seated in front of a computer. Reading was binocular, but only the movements of the dominant eye were monitored. Eye dominance was determined using a Miles Test (Rice et al., 2008): participants were asked to extend their arms forward and make a triangle-shape like window with their hands. Then, they positioned the window such that a target point hanging on the wall appeared in the centre while both eyes were open. Next, they were told to close one eye at a time and note the position of the target point. The dominant eye was the eye in which the target stayed centred in the frame when the eye was open.

A 9-point calibration procedure was performed, followed by six practice trials, and then the experimental sentences followed. Participants fixated a drift correct target prior to each trial and recalibration was performed as needed. The participant clicked on the mouse to terminate each trial when they had finished reading.

Participants were instructed to read for comprehension at their own pace. It was emphasised that the task was to comprehend the sentences, and not to memorise the content. They were further told that after varying intervals they would get multiple-choice questions about the content of the previously presented sentence.

Results

Data preparation. Practice trials (0.5%), sentences' first and last words (28.1%), and skipped words (19%) were excluded from the analyses, as is customary in eye-tracking research (Kliegl et al., 2006; Yan et al., 2014; Zhou et al., 2018). To detect outliers, we used Q-Q plots for total reading time, which revealed 12 data points exceeding 2,400 ms, that we excluded (0.04%). We also excluded fixations shorter than 100 ms (3,5%), as it has been argued that they do not reflect cognitive processes, but instead the outcome of micro-saccades performed to adjust eyes' position (Rayner, 1998) or blinks during the neighbouring fixation (Bertram, 2011). After these exclusions, 25,821 observations were available for analysis.

Subsequently, we divided the data set in three subsets: one for the comparison between syllable-congruent, syllable-incongruent and control conditions (a), one for the comparison between morpheme-congruent, morpheme-incongruent and control conditions (b), and one for the comparison between morpheme-congruent and syllable-congruent conditions. This was done to maximise the number of items in each comparison.

For the comparison between the syllable-congruent, syllable-incongruent, and control conditions (a), we excluded all monosyllabic items (39.8%). Moreover, we excluded words where the morpheme and syllable boundaries fully overlapped (7.45%) to maximise the strength of the manipulation (e.g., in the word *Künstler* [artist], the colour alternation happens both at the morpheme and at the syllable boundary). 14,378 observations were available for analysis.

For the comparison between the morpheme-congruent, morpheme-incongruent, and control conditions (b), we excluded monomorphemic items (58.6%) and items where the morpheme and syllable boundaries fully overlapped (9.5%). 9,670 observations were available for analysis.

For the comparison between the syllable-congruent and MCCs (c), to ensure comparability, we restricted the data set to items that had only two syllables and two morphemes. Therefore, we excluded monosyllabic and monomorphemic items and items that had more than two syllables/morphemes (76.9%), as well as items where the morpheme and syllable boundaries fully overlapped (14%). 5,134 items were available for analysis.

Linear mixed effect models were run for each of the following dependent variables: first fixation duration (duration of initial fixation on the target during the first pass through the text), gaze duration (sum of all first-pass fixations made on the target), total reading time (sum of all fixations on the target, including any regressions back to it), regressions (probability of making a regression back to the target from a later portion in the sentence), total number of saccades, and landing position. We had preregistered the analyses on total reading time as an indicator of overall ease-of-processing, and performed the additional analyses to explore the time-course of the effects. Data were analysed in the R computing environment (R Core Team, 2021). Linear mixed effects models were constructed using the *lme4* package (Bates et al., 2015) with four fixed effects (condition, and centred word length, frequency and predictability), their interactions, and two random effects (participants and items, with correlated random intercepts for both, and random slopes for participants). *P* values were obtained using the *lmerTest* package (Kuznetsova et al., 2017). Factor condition was coded using sum-to-zero contrasts to carry out five pair-wise comparisons between the congruent (1), incongruent (–1), and control conditions (0).

Analysis. First fixation duration, gaze duration, total reading time, number of regressions, number of saccades, and landing position were analysed separately. Response time distributions were checked using the Box-Cox system of the powerTransform function in the CAR package (Fox & Weisberg, 2019), showing that response time transformations were not necessary. Moreover, we applied a Holm–Bonferroni correction (Holm, 1979), a sequential approach with the advantage of maintaining the power of the statistical tests (compared with the more common Sidàk/Bonferroni corrections) while controlling for familywise Type 1 errors (Abdi, 2010). The method compares each observed *p*-value to an adjusted α-threshold. The original *p* values are listed from the smallest to the largest within each of comparison, across variables. We performed the correction using two different Microsoft Excel tools made available by researchers (Boustani, 2020; Gaetano, 2018) to double-check the correction. For the sake of clarity, we will report adjusted *p* values (adj. *p*), instead of adjusted α thresholds, and consider significant any adjusted *p* value $< .05$. Observed power calculations indicated that all models had above 80% chance to find an effect.

Table 3. Descriptive statistics of first fixation durations, gaze durations, and total reading times, in milliseconds, for all conditions (Experiment 1).

Comp	*M*	*SE*	*Comp*	*M*	*SE*	*Comp*	*M*	*SE*
First fixation duration								
SCC	187	4.47	*MCC*	186	4.72	*MCC*	186	5.00
SIC	185	4.66	*MIC*	188	4.52	*SCC*	199	5.09
CTRL	185	4.29	*CTRL*	185	4.61			
Gaze duration								
SCC	208	6.58	*MCC*	207	6.99	*MCC*	208	6.73
SIC	208	6.58	*MIC*	215	7.39	*SCC*	211	6.79
CTRL	208	7.16	*CTRL*	207	7.65			
Total reading time								
SCC	299	16.87	*MCC*	301	17.78	*MCC*	304	16.12
SIC	298	16.37	*MIC*	306	16.41	*SCC*	311	15.20
CTRL	300	17.23	*CTRL*	306	19.06			

Comp: comparison; SE: standard error; SCC: syllable-congruent; MCC: morpheme-congruent; SIC: syllable-incongruent; MIC: morpheme-incongruent; CTRL: control.

Reading times. Mean first fixation duration, gaze duration, total reading time, and corresponding standard errors are reported in Table 3. There were no significant differences between any of the conditions, in any of the comparisons (adj. $p > .05$).

Number of regressions. This analysis was conducted to test the possibility that participants would make fewer regressions in the MCC compared with the marked syllable condition, since highlighting morphological information could make word recognition faster because of its link with semantics. However, the data did not support this hypothesis ($b = 0.001$; $SE = 0.03$; $t = 0.05$; $p = .95$).

Number of saccades. There was no difference between the morpheme-congruent and the syllable-congruent conditions ($b = –0.29$; $SE = 0.34$; $t = –0.84$; $p = .39$).

Landing position. We predicted that highlighting relevant units could change the eyes' landing position within words if participants early recognised those units. However, using log-transformed data (powerTransform = 0.59; 0.54), we found no evidence that this was the case in any of the conditions (adj. $p > .05$).

Discussion

We investigated whether highlighting syllables or morphemes with the "Silbenmethode" enhanced reading fluency in German skilled readers, and whether marking morphemes yielded larger facilitation effects compared with marking syllables. The results revealed no such evidence, neither in reading times, number of regressions, number of saccades, nor landing position.

Despite the absence of a syllable or morpheme effect in these data, the results of Experiment 1 do not rule out that German skilled readers rely on syllables and morphemes in their reading. Instead, the findings suggest that the colouring of embedded reading unit, as used in the "Silbenmethode," does not modulate the eye-movements of German skilled readers.

Earlier ERP results by Carreiras et al. (2005) showed a temporal and spatial dissociation of colour-syllable congruency effects for Spanish low-frequency words and pseudowords, with a larger amplitude for the colour-syllable incongruency condition in the P200 time window compared with the colour-syllable congruency and baseline conditions. However, in the current experiment, we found that the colour congruency effects were not modulated by the difference in word frequency. Cross-linguistically, there might be little reason for assuming differences in syllable processing between German and Spanish, as both languages are orthographically transparent (Seymour et al., 2003). If anything, German is characterised by a higher degree of syllabic complexity than Spanish, with its close CVC syllables and consonant clusters in onset and coda positions (Borleffs et al., 2018; Stenneken et al., 2007). In fact, like us, Carreiras and colleagues did not find any significant differences in the behavioural data. Therefore, it is likely that the ERP signal has a greater sensitivity to reflect changes in the colour of the text compared with eye-tracking.

This might be also the reason why in their eye-tracking study Häikiö et al. (2015) reported an absence of congruency effects using alternated colours (black/red) as syllable boundary cues. However, it is worth noting that the same group of authors conducted a second experiment where they found significant differences in reading speed when colour alternations were replaced with hyphens as segmentation cues. Finnish beginning readers' gaze durations and sentence reading time were significantly longer when reading hyphenated items compared with nonhyphenated control condition, especially if the hyphen position did not match the syllable boundary. Building on the critical findings by Häikiö and colleagues, we designed a second experiment to test the use of hyphens (-) as an alternative segmentation cue and to directly examine its impact on syllable and morpheme processing in German.

Experiment 2

Results from Experiment 1 provided no evidence that highlighting relevant subword units such as syllables or morphemes via colour alternations modulated eye-movements. Therefore, in Experiment 2 we opted for a more obvious visual disruption using hyphenation while segmenting the stimuli in the same way as in Experiment 1 (e.g., *Laura stell-te ein-e Schüssel Kirsch-en auf den Tisch*). Hyphenation has been widely used to study syllabic and morphological processing in an orthographically transparent, morphologically rich language such as Finnish (Häikiö et al., 2011, 2015, 2016).

The predictions slightly differed from those of Experiment 1. We theorised that segmenting words in syllables or morphemes would result in longer fixation times in the hyphenated conditions compared with the control conditions. In other words, hyphenation cues were expected to hinder rather than facilitate word processing (Deilen et al., 2022; Häikiö et al., 2011, 2015, 2016; Häikiö & Luotojärvi, 2022). Following the rationale of Experiment 1, we hypothesised that the MCC would result in shorter fixation times than the syllable-congruent condition. Furthermore, if eye-movements are modulated by syllabic and morphemic structure, hyphens placed within units would be expected to disrupt reading, thus resulting in longer fixation times in the incongruent than congruent conditions.

Method

Participants. We recruited 36 participants (27 females, 9 males) with the same characteristics of Experiment 1. Eighteen participants already participated in Experiment 1 and a minimum of 2 months passed from participating in the first experiment. Due to restrictions in COVID-19 mobility at the time of data collection, we terminated recruitment early, thus not reaching the targeted sample size of 40 participants. Furthermore, data of four participants who scored poorly on the SLRT II word reading test ($<$16 percentile) were not included in the analysis. The remaining 32 participants were between 18 and 52 years old ($M = 31.3$; $SD = 7.7$). All had normal or corrected-to-normal vision. All participants provided written, informed consent.

Materials. We used the same stimuli as in Experiment 1. In this experiment, syllables and morphemes were separated using a hyphen in the morpheme-congruent, morpheme-incongruent, syllable-congruent, syllable-incongruent conditions (see Table 4). No hyphens were used in the control condition. As inserting hyphens to separate relevant units resulted in longer words, 7.63% of sentences in morpheme conditions, and 11.11% in syllable conditions, did not fit into single lines and fell into double lines. All sentences were presented in black against a white background.

Procedure and apparatus. We used the same procedure and eye-tracker in Experiment 2 as we did in Experiment 1.

Table 4. Hyphenated stimuli examples (Experiment 2).

Condition	Acronym	Stimuli example
Syllable-Congruent	SCC	*Lau-ra stell-te ei-ne Schüs-sel Kir-schen auf den Tisch*
Syllable-Incongruent	SIC	*L-aura ste-llte ein-e Sch-üssel Kirsche-n auf den Tisch*
Morpheme-Congruent	MCC	*Laura stell-te ein-e Schüssel Kirsch-en auf den Tisch*
Morpheme-Incongruent	MIC	*Laura ste-llte ei-ne Schüssel Ki-rschen auf den Tisch*
Control	CTRL	*Laura stellte eine Schüssel Kirschen auf den Tisch*

Results

Data preparation. Data preparation followed the same principles as in Experiment 1. We excluded practice sentences' data (0.5%), sentences' first and last words (27.8%), and skipped words (18.8%). Fixation duration outliers were also excluded if they exceeded 2,400 ms for total reading time (0.03%) or were shorter than 100 ms (3.4%). All participants scored more than 80% of correct answers in the multiple-choice questions.

For the comparison between syllable-congruent, syllable-incongruent, and control conditions (a), we excluded monosyllabic items (39.6%) and items were the morpheme and syllable boundaries fully overlapped (6.3%). In addition, the number of hyphens naturally varied between the syllable-congruent and syllable-incongruent conditions (e.g., Po-li-ti-ker/Polit-iker). Since item length represents an important predictor of eye-movements (Hyönä, 2012), item pairs with varying number of hyphens were excluded (–47.6%). 6,132 observations remained available for analysis.

For the comparison between MCC, morpheme-incongruent and control conditions (b), we excluded monomorphemic items (58.2%) and items where the morpheme and syllable boundaries fully overlapped (9.2%). 7,849 observations were available for analysis.

For the comparison between syllable-congruent versus MCCs (c), we restricted the data set to items that were bi-morphemic and bi-syllabic (excluded items were 76% of the original data set). We also excluded items were the morpheme and syllable boundaries fully overlapped (16%). 4,197 observations were available for analysis.

We extracted first fixation duration, gaze duration, total reading time, and regressions.[1] As in Experiment 1, data were analysed using linear mixed effect models.

Analysis. Similarly to Experiment 1, we corrected *p* values using the Holm–Bonferroni correction, and response time distributions were checked using the powerTransform. Except for first fixation duration, response time transformations were not necessary.[2] Observed power calculations indicated that all models had above 78% chance to find an effect.

First fixation duration. First fixation durations were log-transformed for the analysis. For clarity, we report raw mean first fixation durations and standard errors in Figures 1 and 4.

We found a significant difference in the syllable-congruent versus syllable-incongruent comparison (b=–0.01; SE=0.007; t=–2.31; p=.02; adj. p =.04), with participants reading the latter condition significantly more slowly than syllable-congruent stimuli. All the other comparisons were not significant (control vs. MCCs, control vs. syllable-congruent conditions, morpheme-congruent vs. morpheme-incongruent conditions, and morpheme-congruent vs. syllable-congruent conditions, adj. $p > .05$).

Gaze duration. Mean gaze durations and standard errors are reported in Figures 2 and 4. We found significant differences between the control and MCCs (b=–30.06; SE=5.33; t=–5.63; p < .001; adj. p < .001), and between the control and syllable-congruent conditions (b=–32.47; SE=5.68; t=–5.70; p < .001; adj. p < .001), with the syllable-congruent and MCCs being read more slowly than the control condition. The difference between the syllable-congruent and MCCs was also significant (b=28.55; SE=9.16; t=3.11; p=.01; adj. p =.03), with morpheme-congruent stimuli being read more slowly. Finally, we found a significant difference between the syllable-congruent and syllable-incongruent conditions (b=–18.17; SE=6.99; t=–2.59; p=.009; adj. p=.02), with syllable-incongruent stimuli being read more slowly than syllable-congruent stimuli. The comparison between the morpheme-congruent and morpheme-incongruent conditions was not significant (adj. $p > .05$).

Total reading time. Mean gaze durations and standard errors are reported in Figures 3 and 4. We found significant differences between the control and MCCs (b=–56.59; SE=8.9; t=–6.30; p=< 0.001; adj. p < .001), and between the control and syllable-congruent conditions (b=–52.49; SE=9.46; t=–5.54; p < .001; adj. p < .001), with the syllable-congruent and MCCs being read more slowly than the control condition. The comparison between the morpheme-congruent and syllable-congruent conditions was also significant (b=28.5; SE=9.16; t=3.11; p=.001; adj. p=.007), with morpheme-congruent stimuli read more slowly than syllable-congruent stimuli. The comparison between the syllable-congruent and syllable-incongruent conditions was significant (b=–24.66; SE=12.074; t=–2.04; p=.04; adj. p =.04), with syllable-incongruent stimuli read more slowly than syllable-congruent ones. Again, we found no significant difference in the morpheme-congruent and morpheme-incongruent comparison (adj. $p > .05$).

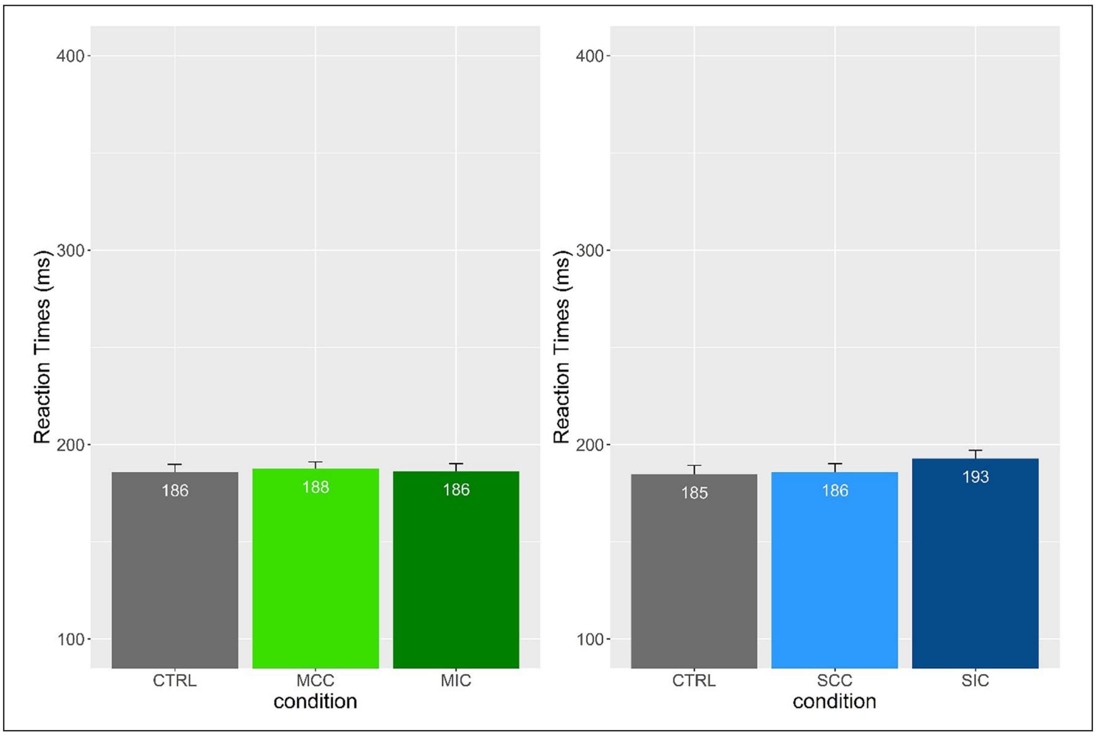


Figure 1. Mean first fixation duration (FFD) and standard errors for the CTRL-MCC-MIC comparison and CTRL-SCC-SIC comparisons.
The SIC condition was read significantly more slowly than the SCC condition. No other significant differences were found. CTRL: control condition; MCC: morpheme-congruent condition; MIC: morpheme-incongruent condition; SCC: syllable-congruent condition; SIC: syllable-incongruent condition.

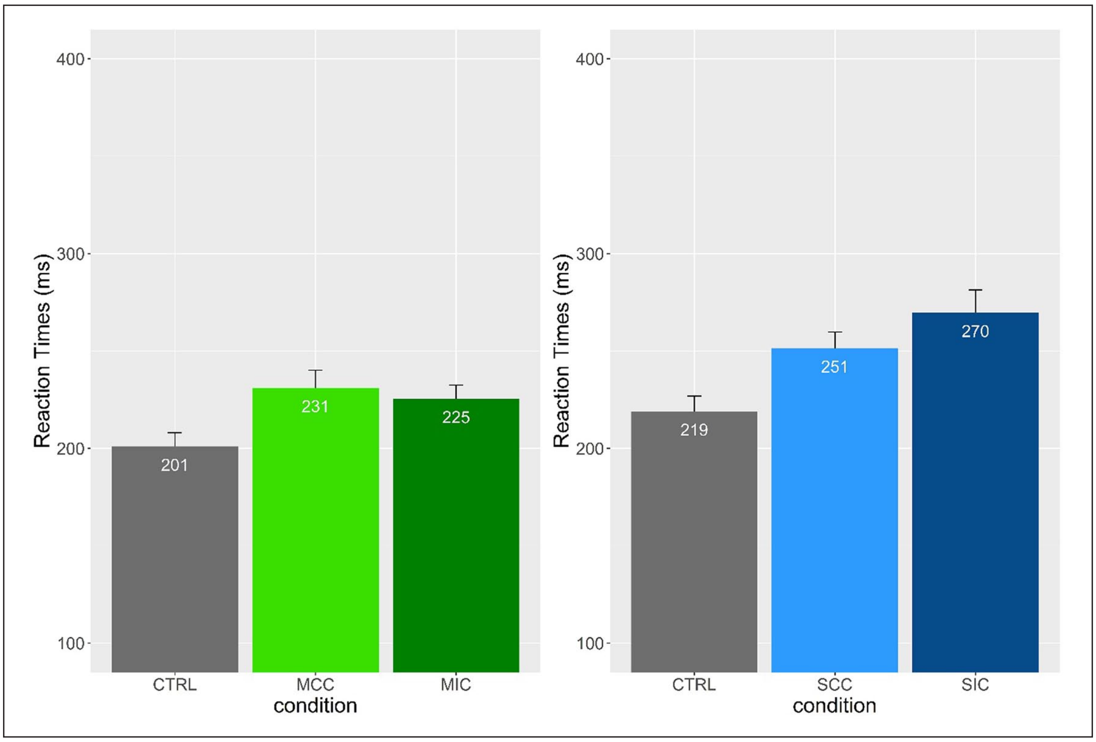


Figure 2. Mean gaze duration (GD) and standard errors for the CTRL-MCC-MIC comparison and CTRL-SCC-SIC comparisons.
Congruent hyphenated conditions (MCC, SCC) were read significantly more slowly than the control condition (CTRL). The SIC condition was read significantly more slowly than the SCC condition. No significant differences were found between MCC versus MIC. CTRL: control condition; MCC: morpheme-congruent condition; MIC: morpheme-incongruent condition; SCC: syllable-congruent condition; SIC: syllable-incongruent condition.

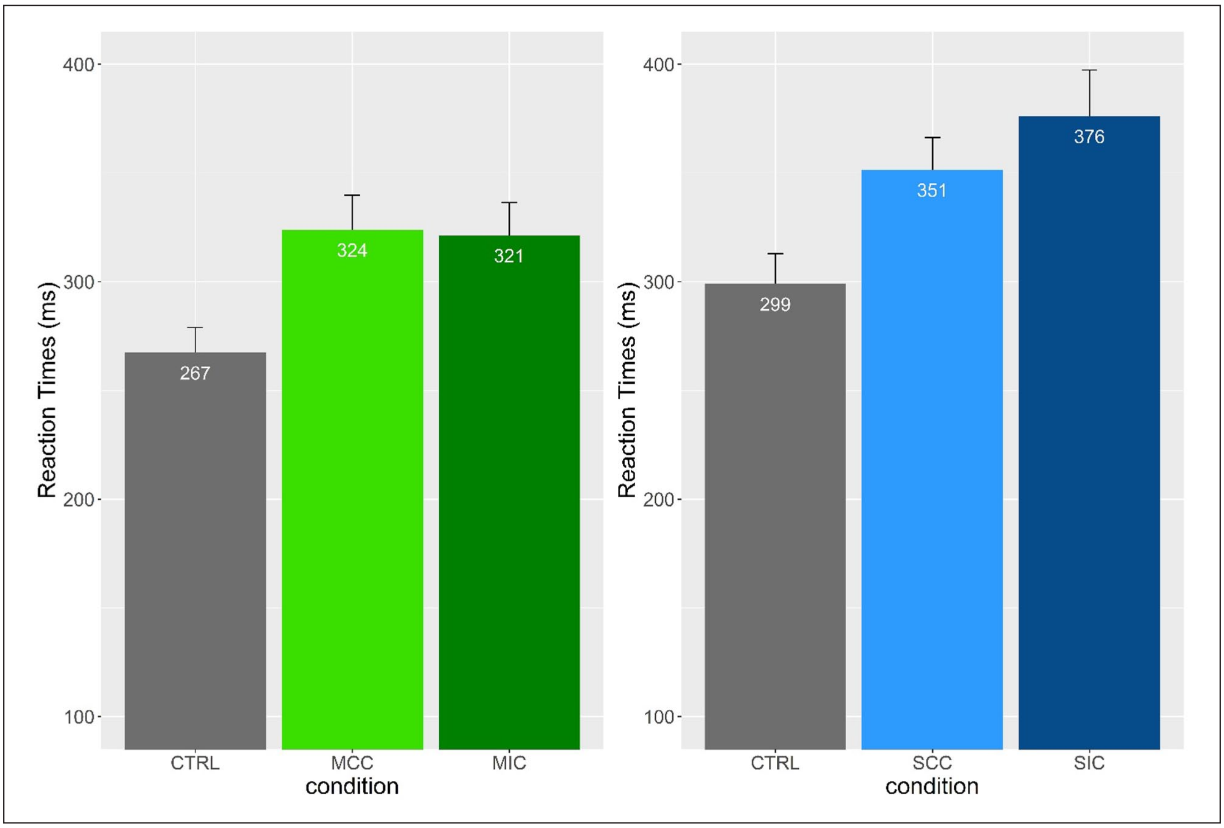


Figure 3. Mean total reading time (TRT) and standard errors for the CTRL-MCC-MIC comparison and CTRL-SCC-SIC comparisons.
Congruent hyphenated conditions (MCC, SCC) were read significantly more slowly than the control condition (CTRL). The SIC condition was read significantly more slowly than the SCC condition. No significant differences were found between MCC versus MIC. CTRL: control condition; MCC: morpheme-congruent condition; MIC: morpheme-incongruent condition; SCC: syllable-congruent condition; SIC: syllable-incongruent condition.

Number of regressions. As in Experiment 1, we found no evidence that participants made fewer regressions in the MCC compared with the syllable-congruent condition (adj. $p > .05$).

Discussion

In Experiment 2, we examined the effect of hyphenation as a syllabic / morphemic segmentation cue using eye-tracking. As expected, the use of hyphenation led to longer fixation times compared with the nonhyphenated control condition, even when hyphens segmented words into informative (i.e., congruent) units. Similar results have been previously reported in Finnish developing readers, using hyphens placed between syllables (Häikiö et al., 2015, 2016; Häikiö & Luotojärvi, 2022; but see Häikiö & Vainio, 2018) and between morphemes (Häikiö et al., 2011; Häikiö & Vainio, 2018). The authors found that first and second graders took longer to read words that were hyphenated at syllable and morpheme boundaries compared with the concatenated control, although hyphenation represents a common reading teaching strategy in Finland. This study extends this finding to adults, and to a language which has a more complex syllabic structure (Seymour et al., 2003), with a fairly complex and productive morphological system.

A second, important finding of Experiment 2 is that participants read syllable-incongruent and morpheme-congruent hyphenated words more slowly than syllable-congruent words; that is, the disruption of syllable boundaries significantly impaired reading, whereas participants were less affected by the disruption of morpheme boundaries. These results suggest that in German, syllable structure is more salient than morpheme structure during reading. After all, previous research has demonstrated that phonology plays a central role in silent reading (see Clifton, 2015 for a selective review in English). For example, Ashby and Clifton (2005) showed that participants read multi-syllabic words with two stressed syllables more slowly than those with one stressed syllable. Similarly, Fitzsimmons and Drieghe (2013) found that five-letter monosyllabic words were skipped more often than bi-syllabic words of the same length, even after predictability and frequency were accounted for, suggesting that skilled readers rely on syllabic processing in silent reading. Given that German words are on average longer

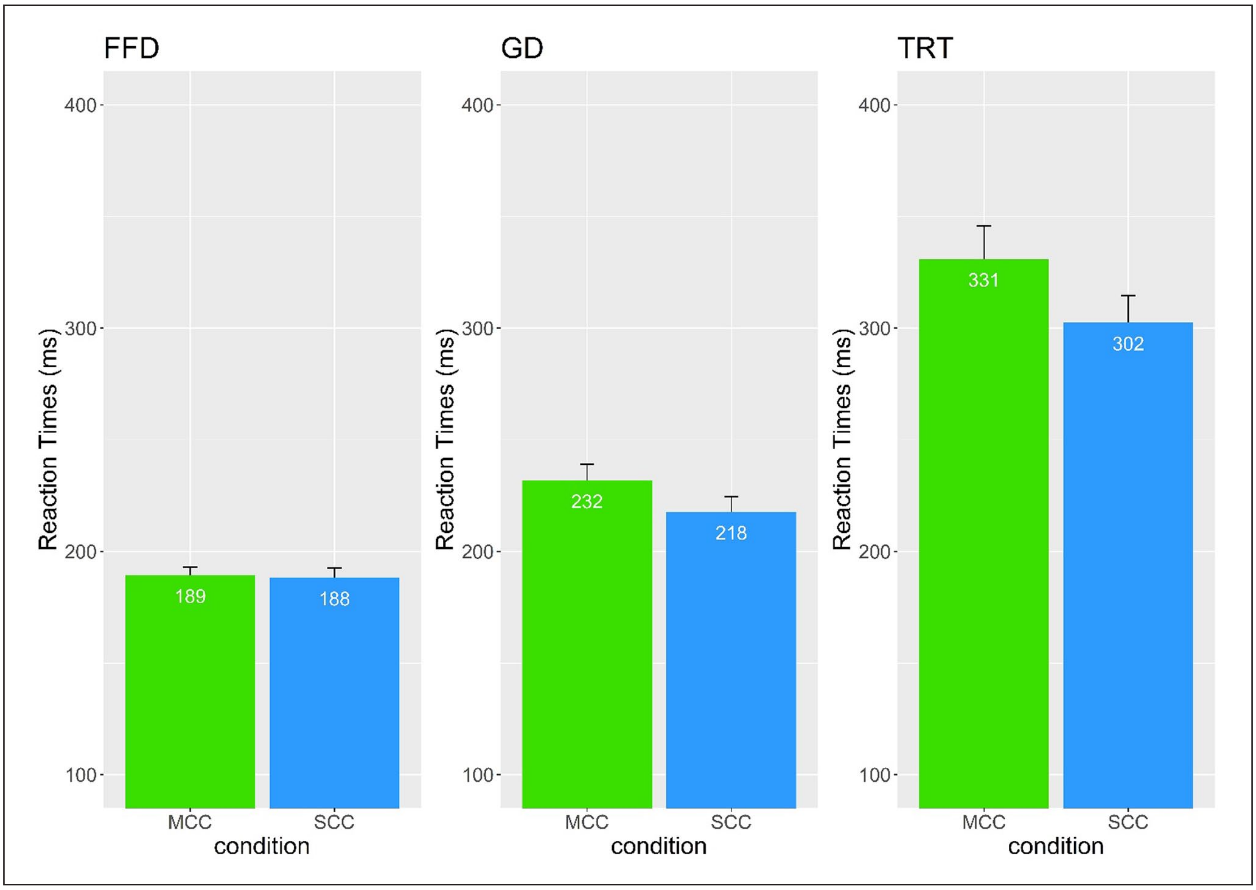


Figure 4. Mean first fixation duration (FFD), gaze duration (GD), and total reading time (TRT) with standard errors for the MCC-SCC comparison.
The comparison between MCC-SCC was significant in gaze duration and total reading time. This comparison was performed using items that only had two syllables and two morphemes. MCC: morpheme-congruent condition; SCC: syllable-congruent condition.

than English words, the role of syllable structure in silent reading might be more pronounced.

Once again, our results converge with those of Häikiö et al. (2015) who showed that hyphenation that did not coincide with syllable boundaries was more disruptive than hyphenation that matched syllable boundaries. A later study in Finnish (Häikiö & Vainio, 2018) further showed that hyphens that disrupted syllables (SIC) but not morphemes (MCC) did not impair reading, a finding that was not confirmed in the current German data. This difference between the Finnish and the German data might be due to typological considerations. Although both Finnish and German are known to be morphologically productive, German is a highly synthetic, whereas Finnish is a polysynthetic language. In polysynthetic languages, multiple morphemes can be combined into a single continuous word that can even constitute an entire clause at times. In synthetic languages like German, the number of concatenations is more limited and form sentence clauses. As a result, readers of polysynthetic languages like Finnish may be more sensitive to the morphological structure of words, thus leading to more robust morpheme segmentation effects in Finnish than in German. Häikiö and Vainio (2018) reached similar conclusions when comparing their study with the one of Hasenäcker and Schroeder (2017), arguing that the different results might be ascribable to either the fact that Finnish is a more morphologically rich language, or that the difference is to be found in the fact that they used inflectional morphemes, whereas Hasenäcker and Schroeder used derivational ones (Häikiö & Vainio, 2018, p. 1236), while we note that this study used a combination of inflectional and derivational morphemes.

Under this aspect, our results converge with those of Hasenäcker and Schroeder (2017), although their adult participants were not impaired by the syllable-incongruent condition. It must be noted that Hasenäcker and Schroeder used a lexical decision task, thus investigating single-word recognition. It is possible that the longer fixations on the syllable-incongruent stimuli in our study were due to the additional time needed to integrate the disrupted words in the sentence context: that is, not only word recognition was impaired, but sentence-processing was, too.

General discussion

The aim of this study was to directly compare the processing of syllables and morphemes in a language with a complex syllabic and morphological structure and transparent orthography like German. To address this aim, we employed two different types of segmentation cues: colouring in Experiment 1, hyphenation in Experiment 2. Hyphens and colour alternation positions matched or mismatched syllables boundaries (SCC-SIC) or morpheme boundaries (MCC-MIC). Black-coloured, nonhyphenated sentences served as control condition.

The results of Experiment 1 did not support the hypothesis that highlighting syllabic or morphological information using colours modulates eye-movements or general processing speed. However, Experiment 2 revealed that segmentation by hyphenation leads to longer eye fixations, compared with the non-hyphenated condition, both in the morpheme and the syllable-congruent condition. Critically, the results further showed that the morpheme-congruent and syllable-incongruent conditions were read significantly more slowly than the syllable-congruent condition. The evidence that hyphenation disrupted reading to a greater extent when hyphen position did not match with syllable boundary, can be interpreted as an indication that in silent reading, German skilled readers automatically recognise the underlying syllabic structure of words.

There are several explanations why participants' reading behaviour may have been affected by syllable structure. In the case of the syllable-congruent condition, participants' eye-movements may have more easily fallen into a rhythm of syllable-based reading while shifting their eyes through the target sentences, given the important role of syllables in silent reading (e.g., in German, Conrad et al., 2011; Conrad & Jacobs, 2004; Hawelka et al., 2013; Hutzler et al., 2005). Since readers automatically captured the underlying syllable structure, their eye-movements were disrupted when hyphenation did not occur at syllable boundaries. In contrast, the segmentation of words into morphemes would have required a more thorough analysis of letter chunks as units of meaning and therefore may have presented a level of complexity that was not as easily grasped during sentence reading. Of course, there is abundant evidence that German readers engage in morphological processing (e.g., Beyersmann et al., 2020; Smolka et al., 2009), which is not inconsistent with our current findings. In a large web-based study with German third and fourth graders, Görgen et al. (2021) found that morphological awareness is a better predictor for spelling than for reading fluency, suggesting that the use of morphological knowledge is modulated by task-specific requirements. The present data show that within a sentence reading paradigm, highlighting syllable structure via hyphenation has a larger impact on reading behaviour in German compared with highlighting morpheme structure.

An alternative explanation for the prominence of a syllable effect in the current data is that syllable structure is explicitly taught as part of the curriculum of the German schooling system (Bredel et al., 2013). Syllable-based reading instruction has a long tradition in Germany (Reh & Wilde, 2016), with syllable separators dating back to the 16th century (Velten, 2012). As such, skilled readers have a long history of applying syllable-based reading. Duncan et al. (1997) suggested that the style of reading instruction may affect the relative use of different sized units during reading acquisition. Furthermore, in a study looking into the benefits of teaching children orthographic analogies based on onset and rime units, Peterson and Haines (1992) found that the training boosted the children's phonemic awareness, promoting segmentation skills of these units. Hence, learning to read using syllable-based strategies is likely to boost the syllabic awareness in a similar way, with effects spanning throughout reading development into adulthood.

This would also explain the absence of a morphological effect in the current data, as formal morphological instruction is comparatively less common in the German schooling system.[3] Indeed, recent studies have shown that German readers are proficient at identifying embedded stems (Beyersmann et al., 2020, 2021) perhaps due to the abundant presence of compound words in the German language and have reported to be less reliant on morphological processing than French (Beyersmann et al., 2021) and English readers (Mousikou et al., 2020). Recent longitudinal data involving two large samples of German and French primary schoolers have shown that embedded stem priming effects are more pronounced in German third and fourth graders whereas morphological priming effects are more pronounced in French third and fourth graders (Beyersmann et al., 2021), suggesting that the development of morphological processing mechanisms is influenced by the intrinsic linguistic properties of the language to which children are exposed to. Thus, although German readers clearly process morphemes in their reading (Kempe, 1999), the recognition of syllabic structure appears to predominantly underpin the word processing in our task.

Previous studies argued that syllables, especially the first syllable of polysyllabic words, mediate lexical access (e.g., Carreiras et al., 1993; Prinzmetal et al., 1986; Spoehr & Smith, 1973; Taft & Forster, 1975). In some models of visual word recognition, syllables are represented at an intermediate level situated between the letter and the lexical levels (Jacobs et al., 1998; van Heuven et al., 2001) including the dual-route interactive-activation framework (the IAS model, see Figure 1 of Mathey et al., 2006, p. 389; or the MROM-S model, see Figure 1 of Conrad et al., 2010, p. 872). In these models, two routes allow the reader to access the lexicon, the orthographic (from letters to words) and the phonological (from syllables to words) routes. The latter is equipped with a level of syllabic representations which mediate between the levels of the letter and the word. In the first model (IAS), when letters are activated, this activity spreads to consistent positional syllables. For example, a word's first bigram, such as "co" in "comix" activates not just the syllable /kom/ but also additional consistent syllables like /kɒr/, /kʌl/, or /kon/. Syllable activation strength is determined not only by the degree of activation at the letter level, but also by the syllable resting level. In the second model (MROM-S), syllabic parsing is modulated both by the frequency of the letter cluster forming the initial syllable and the model's syllabary, which contains syllabification rules. Further ambiguity is resolved by feedback from the word level. In both models, syllable frequency is proportional to the level of resting activity. As a result, high-frequency syllables are engaged faster than low-frequency syllables. The syllables then activate the corresponding target word, as well as its syllabic neighbours.

Evidence for models implementing syllables at an intermediate level between orthographic input and the lexicon comes from lexical decision and naming tasks (Conrad et al., 2009; Conrad & Jacobs, 2004), showing that reading times in lexical decision and naming tasks tend to be longer if the first syllable of the words is highly frequent (i.e., syllables that are often found in first positions). This suggests that first syllables that are shared among many word candidates lead to lateral inhibition at the level of the orthographic lexicon (Conrad & Jacobs, 2004; Hawelka et al., 2013; Perea & Carreiras, 1998).

The important role of syllables in word processing has also been demonstrated by studies reporting a syllable congruency effect. This effect typically emerges in lexical decision tasks paired with masked priming (Chetail & Mathey, 2009a, 2012). In this paradigm, primes are quickly displayed so that readers can only process them subliminally, and then are replaced by target words to which participants must make a lexical decision. Facilitatory syllabic priming effects are found when the prime and the target share the first syllable as opposed to just the first letters (Chetail & Mathey, 2012). The present eye-tracking study extend these prior findings from single-word psycholinguistics tasks to a more ecologically valid sentence reading paradigm, suggesting that syllables mediate lexical access in a shallow orthography like German.

A further important point to note is that this study was not designed to directly tease apart the independent role of phonology in processing syllabic structures. While phonological and syllabic processing are naturally intertwined (Álvarez et al., 2004; Conrad et al., 2009), orthographic redundancy (i.e., low-frequency bigrams that can be found at the syllable boundary, Seidenberg, 1987) can increase the salience of syllabic units (Conrad et al., 2009; Doignon-Camus & Zagar, 2005). Moreover, participants might segment words in syllable-like orthographic units, which could activate phonological syllables, with orthographic processing preceding phonological processing. Hence, the here observed syllable effect might reflect either an orthographic or a phonological syllable effect, or a combined effect, since readers preprocess orthographic and phonological information already in the parafovea (see Schotter et al., 2012 for a review on parafoveal processing in reading).

As a final point, we do not consider our findings to be automatically generalisable across languages, as reliance on sublexical units may vary depending on morphological and syllabic complexity, orthographic depth, or linguistic typology. Studies specifically designed to reveal readers' preferred units according to these constructs are needed. Also, while Experiment 1 did not support the idea that colouring segmentation (i.e., the Silbenmethode) modulated sublexical processing in skilled adult readers, it is possible that children who are still in the process of learning to read would differently benefit from colour cues in their reading. An extension of this study to developing readers of German may thus provide fertile grounds for future research, particularly given its importance in German reading instruction.

Conclusion

The present eye-tracking study was designed to directly compare the processing of syllables and morphemes and investigate the use of different segmentation cues including colour highlighting and hyphenation in an orthographically transparent, morphologically rich, and syllabically complex language, namely, German. The results of the first experiment showed that eye-movements were not modulated by colour alternations. Critically, the results of the second experiment revealed that German skilled readers rely more heavily on syllable-based than morpheme-based reading when hyphenation was used as segmentation cue. We speculate that this preference might have either originated from the syllabic awareness boost resulting from the syllable-based reading instruction in the German schooling system; or be the product of an underlying reading mechanism relying on syllables, as has been shown for other transparent orthographies such as French and Spanish.

Acknowledgements

The authors thank Lior Weinreich and Björn Witzel for piloting the experiments, Serje Robidoux for providing statistical advice, and Roland Trouville for trouble-shooting coding scripts.

Data accessibility statement

The data and materials for the experiment are available at https://osf.io/w4rsm/. The preregistration of the first experiment is available at https://osf.io/csja8. Registered project files are available at https://osf.io/ya8kw.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Förderung für Forschung und Lehre grant to X.S. and E.d.S. (REG.NR. 1036), the Deutsche Forschungsgemeinschaft to X.S. (SCHM3460 1-1), and by a Discovery Early Career Researcher Award to E.B. from the Australian Research Council (DE190100850).

ORCID iD

Elisabetta De Simone https://orcid.org/0000-0002-9737-7626

Notes

1. As inserting hyphens within words modifies their spatial length, which in turn influences landing position and the number of saccades, the analysis of landing position in Experiment 2 is not reported within the article. Indeed, when analysed, the results showed that fixations landed more to the right in the hyphenated compared with the control condition (a detailed summary of these results is provided in the RMarkdown script on https://osf.io/dt9yf).
2. To control for any impact that the returning participants from Experiment 1 could have had on the results of Experiment 2, we conducted an additional set of non-preregistered, post hoc analyses by adding "Participant's status" (new or returning) as a fixed effect in all models. The analyses revealed that participants' status did not modulate the direction or significance of our findings in any of the eye-tracking analyses (a detailed summary of the results is provided in the RMarkdown within the corresponding OSF repository: https://osf.io/dt9yf).
3. In more recent years, German reading instruction methods are beginning to emphasise morpheme-based reading more explicitly. Morphological awareness training has been shown to improve poor readers' spelling, reading comprehension and fluency (Arnbak & Elbro, 2000; Good et al., 2015).

References

Abdi, H. (2010). Holm's sequential Bonferroni procedure. In N. J. Salkind (Ed.), *Encyclopedia of research design* (pp.1–8). SAGE. https://doi.org/10.4135/9781412961288

Adsett, C. R., & Marchand, Y. (2010). Syllabic complexity: A computational evaluation of nine European languages. *Journal of Quantitative Linguistics*, *17*(4), 269–290. https://doi.org/10.1080/09296174.2010.512161

Álvarez, C. J., Carreiras, M., & Perea, M. (2004). Are syllables phonological units in visual word recognition? *Language and Cognitive Processes*, *19*(3), 427–452. https://doi.org/10.1080/01690960344000242

Alvarez, C. J., Carreiras, M., & Taft, M. (2001). Syllables and morphemes: Contrasting frequency effects in Spanish. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(2), 545–555. https://doi.org/10.1037//0278-7393.27.2.545

Amenta, S., & Crepaldi, D. (2012). Morphological processing as we know it: An analytical review of morphological effects in visual word identification. *Frontiers in Psychology*, *3*, Article 232. https://doi.org/10.3389/fpsyg.2012.00232

Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review*, *105*(4), 678–723. https://doi.org/10.1037/0033-295X.105.4.678-723

Arnbak, E., & Elbro, C. (2000). The effects of morphological awareness training on the reading and spelling skills of young dyslexics. *Scandinavian Journal of Educational Research*, *44*(3), 229–251. https://doi.org/10.1080/00313830050154485

Ashby, J., & Clifton, C. (2005). The prosodic property of lexical stress affects eye movements during silent reading. *Cognition*, *96*(3), B89–B100. https://doi.org/10.1016/j.cognition.2004.12.006

Ashby, J., & Rayner, K. (2004). Representing syllable information during silent reading: Evidence from eye movements. *Language and Cognitive Processes*, *19*(3), 391–426. https://doi.org/10.1080/01690960344000233

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

Bertram, R. (2011). Eye-movements and morphological processing in reading. *The Mental Lexicon*, *6*(1), 83–109. https://doi.org/10.1075/ml.6.1.04ber

Bertram, R., & Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language*, *48*(3), 615–634. https://doi.org/10.1016/S0749-596X(02)00539-9

Bertram, R., Pollatsek, A., & Hyönä, J. (2004). Morphological parsing and the use of segmentation cues in reading Finnish compounds. *Journal of Memory and Language*, *51*(3), 325–345. https://doi.org/10.1016/j.jml.2004.06.005

Besner, D. (1987). Phonology, lexical access in reading, and articulatory suppression: A critical review. *The Quarterly Journal of Experimental Psychology Section A*, *39*(3), 467–478. https://doi.org/10.1080/14640748708401799

Beyersmann, E., Castles, A., & Coltheart, M. (2012). Morphological processing during visual word recognition in developing readers: Evidence from masked priming. *Quarterly Journal of Experimental Psychology*, *65*(7), 1306–1326. https://doi.org/10.1080/17470218.2012.656661

Beyersmann, E., & Grainger, J. (2022). *Current issues in the psychology of language. The role of embedded words and morphemes in reading*. Psychology Press.

Beyersmann, E., Mousikou, P., Javourey-Drevet, L., Schroeder, S., Ziegler, J. C., & Grainger, J. (2020). Morphological processing across modalities and languages. *Scientific Studies of Reading*, *24*(6), 500–519.

Beyersmann, E., Mousikou, P., Schroeder, S., Javourey-Drevet, L., Ziegler, J. C., & Grainger, J. (2021). The dynamics of morphological processing in developing readers: A cross-linguistic masked priming study. *Journal of Experimental Child Psychology*, *208*, Article 105140. https://doi.org/10.1016/j.jecp.2021.105140

Beyersmann, E., Ziegler, J. C., Castles, A., Coltheart, M., Kezilas, Y., & Grainger, J. (2016). Morpho-orthographic segmentation without semantics. *Psychonomic Bulletin &*

Review, *23*(2), 533–539. https://doi.org/10.3758/s13423-015-0927-z

Bhattacharya, A. (2020). Syllabic versus morphemic analyses: Teaching multisyllabic word reading to older struggling readers. *Journal of Adolescent & Adult Literacy*, *63*(5), 491–497. https://doi.org/10.1002/jaal.984

Borleffs, E., Maassen, B. A. M., Lyytinen, H., & Zwarts, F. (2017). Measuring orthographic transparency and morphological-syllabic complexity in alphabetic orthographies: A narrative review. *Reading and Writing*, *30*(8), 1617–1638. https://doi.org/10.1007/s11145-017-9741-5

Borleffs, E., Maassen, B. A. M., Lyytinen, H., & Zwarts, F. (2018). Cracking the code: The impact of orthographic transparency and morphological-syllabic complexity on reading and developmental dyslexia. *Frontiers in Psychology*, *9*, Article 2534. https://doi.org/10.3389/fpsyg.2018.02534

Boustani, S. (2020). *Holm-Bonferroni correction in Excel*. https://osf.io/yv5zr

Bredel, U., Becke, T., von der Cramm, D., von Krüßmann, M., Zepnik, S., Dummer-Smoch, L., Klicpera, R., Kuhn, K. P., Noack, C., & Vennemann, T. (2013). *Die Silbe im Anfangsunterricht Deutsch: Festschrift zum zehnjährigen Jubiläum des Lehrgangs « ABC der Tiere »—Silbenmethode mit Silbentrenner (4. Auflage) [The syllable in introduction classes, a commemorative volume for the tenth anniversary of the course ABC of animals – syllable method with hybentrenner®]*. Mildenberger Verlag.

Bruck, M., Treiman, R., & Caravolas, M. (1995). Role of the syllable in the processing of spoken English: Evidence from a nonword comparison task. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 469–479. https://doi.org/10.1037//0096-1523.21.3.469

Carreiras, M., Alvarez, C. J., & Devega, M. (1993). Syllable frequency and visual word recognition in Spanish. *Journal of Memory and Language*, *32*(6), 766–780. https://doi.org/10.1006/jmla.1993.1038

Carreiras, M., & Perea, M. (2002). Masked priming effects with syllabic neighbors in a lexical decision task. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(5), 1228–1242. https://doi.org/10.1037/0096-1523.28.5.1228

Carreiras, M., Riba, J., Vergara, M., Heldmann, M., & Münte, T. F. (2009). Syllable congruency and word frequency effects on brain activation. *Human Brain Mapping*, *30*(9), 3079–3088. https://doi.org/10.1002/hbm.20730

Carreiras, M., Vergara, M., & Barber, H. (2005). Early event-related potential effects of syllabic processing during visual word recognition. *Journal of Cognitive Neuroscience*, *17*(11), 1803–1817. https://doi.org/10.1162/089892905774589217

Chetail, F., & Mathey, S. (2009a). Syllabic priming in lexical decision and naming tasks: The syllable congruency effect re-examined in French. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale*, *63*(1), 40–48. https://doi.org/10.1037/a0012944

Chetail, F., & Mathey, S. (2009b). The syllable frequency effect in visual recognition of French words: A study in skilled and beginning readers. *Reading and Writing*, *22*(8), 955–973. https://doi.org/10.1007/s11145-008-9135-9

Chetail, F., & Mathey, S. (2012). Effect of syllable congruency in sixth graders in the lexical decision task with masked priming. *Scientific Studies of Reading*, *16*(6), 537–549. https://doi.org/10.1080/10888438.2011.607480

Cholin, J., Schiller, N. O., & Levelt, W. J. (2004). The preparation of syllables in speech production. *Journal of Memory and Language*, *50*(1), 47–61. https://doi.org/10.1016/j.jml.2003.08.003

Clifton, C. (2015). The roles of phonology in silent reading: A selective review. In L. Frazier & E. Gibson (Eds.), *Studies in theoretical psycholinguistics. Explicit and implicit prosody in sentence processing* (Vol. 46, pp. 161–176) Springer. https://doi.org/10.1007/978-3-319-12961-7_9

Colé, P., Bouton, S., Leuwers, C., Casalis, S., & Sprenger-Charolles, L. (2012). Stem and derivational-suffix processing during reading by French second and third graders. *Applied Psycholinguistics*, *33*(1), 97–120. https://doi.org/10.1017/S0142716411000282

Colé, P., Magnan, A., & Grainger, J. (1999). Syllable-sized units in visual word recognition: Evidence from skilled and beginning readers of French. *Applied Psycholinguistics*, *20*(4), 507–532. https://doi.org/10.1017/S0142716499004038

Conrad, M., Carreiras, M., Tamm, S., & Jacobs, A. M. (2009). Syllables and bigrams: Orthographic redundancy and syllabic units affect visual word recognition at different processing levels. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 461–479. https://doi.org/10.1037/a0013480

Conrad, M., & Jacobs, A. (2004). Replicating syllable frequency effects in Spanish in German: One more challenge to computational models of visual word recognition. *Language and Cognitive Processes*, *19*(3), 369–390. https://doi.org/10.1080/01690960344000224

Conrad, M., Tamm, S., Carreiras, M., & Jacobs, A. M. (2010). Simulating syllable frequency effects within an interactive activation framework. *European Journal of Cognitive Psychology*, *22*(5), 861–893. https://doi.org/10.1080/09541440903356777

Conrad, M., Vo, M., Schneider, D., & Jacobs, A. M. (2011). Syllable structure is modulating the optimal viewing position in visual word recognition. *Revista de Logopedia, Foniatría y Audiología*, *31*(1), 14–21. https://doi.org/10.1016/S0214-4603(11)70166-8

Dawson, N., Rastle, K., & Ricketts, J. (2018). Morphological effects in visual word recognition: Children, adolescents, and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(4), 645–654. https://doi.org/10.1037/xlm0000485

Deilen, S., Hansen-Schirra, S., & Nagels, A. (2022). The effect of segmentation signs in compounds on reading behavior: An eye-tracking study. *Lebende Sprachen*, *67*(1), 45–99. https://doi.org/10.1515/les-2022-0003

De Simone, E., Beyersmann, E., Mulatti, C., Mirault, J., & Schmalz, X. (2021). Order among chaos: Cross-linguistic differences and developmental trajectories in pseudoword reading aloud using pronunciation Entropy. *PLOS ONE*, *16*(5), Article e0251629. https://doi.org/10.1371/journal.pone.0251629

Deutsch, A., Frost, R., Pelleg, S., Pollatsek, A., & Rayner, K. (2003). Early morphological effects in reading: Evidence from parafoveal preview benefit in Hebrew. *Psychonomic Bulletin & Review*, *10*(2), 415–422. https://doi.org/10.3758/bf03196500

Diependaele, K., Sandra, D., & Grainger, J. (2009). Semantic transparency and masked morphological priming: The case of prefixed words. *Memory & Cognition*, *37*(6), 895–908. https://doi.org/10.3758/MC.37.6.895

Doignon-Camus, N., & Zagar, D. (2005). Illusory conjunctions in French: The nature of sublexical units in visual word recognition. *Language and Cognitive Processes*, *20*(3), 443–464. https://doi.org/10.1080/01690960444000269

Doignon-Camus, N., Zagar, D., & Mathey, S. (2009). Can we see syllables in monosyllabic words? A study with illusory conjunctions. *European Journal of Cognitive Psychology*, *21*(4), 599–614. https://doi.org/10.1080/09541440802052543

Domínguez, A., Alija, M., Cuetos, F., & de Vega, M. (2006). Event related potentials reveal differences between morphological (prefixes) and phonological (syllables) processing of words. *Neuroscience Letters*, *408*(1), 10–15. https://doi.org/10.1016/j.neulet.2006.06.048

Duñabeitia, J. A., Perea, M., & Carreiras, M. (2007). Do transposed-letter similarity effects occur at a morpheme level? Evidence for morpho-orthographic decomposition. *Cognition*, *105*(3), 691–703. https://doi.org/10.1016/j.cognition.2006.12.001

Duncan, L. G., Seymour, P. H., & Hill, S. (1997). How important are rhyme and analogy in beginning reading? *Cognition*, *63*(2), 171–208. https://doi.org/10.1016/s0010-0277(97)00001-2

Ehri, L. C. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading*, *18*(2), 116–125. https://doi.org/10.1111/j.1467-9817.1995.tb00077.x

Feldman, L. B., Milin, P., Cho, K. W., Del Moscoso Prado Martín, F., & O'Connor, P. A. (2015). Must analysis of meaning follow analysis of form? A time course analysis. *Frontiers in Human Neuroscience*, *9*, Article 111. https://doi.org/10.3389/fnhum.2015.00111

Feldman, L. B., O'Connor, P. A., & Del Prado Martín, F. M. (2009). Early morphological processing is morphosemantic and not simply morpho-orthographic: A violation of form-then-meaning accounts of word recognition. *Psychonomic Bulletin & Review*, *16*(4), 684–691. https://doi.org/10.3758/PBR.16.4.684

Ferrand, L., Segui, J., & Grainger, J. (1996). Masked priming of word and picture naming: The role of syllabic units. *Journal of Memory and Language*, *35*(5), 708–723. https://doi.org/10.1006/jmla.1996.0037

Ferrand, L., Segui, J., & Humphreys, G. W. (1997). The syllable's role in word naming. *Memory & Cognition*, *25*(4), 458–470. https://doi.org/10.3758/BF03201122

Fitzsimmons, G., & Drieghe, D. (2011). The influence of number of syllables on word skipping during reading. *Psychonomic Bulletin & Review*, *18*(4), 736–741. https://doi.org/10.3758/s13423-011-0105-x

Fitzsimmons, G., & Drieghe, D. (2013). How fast can predictability influence word skipping during reading? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *39*(4), 1054–1063. https://doi.org/10.1037/a0030909

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). SAGE.

Fracasso, L. E., Bangs, K., & Binder, K. S. (2016). The contributions of phonological and morphological awareness to literacy skills in the adult basic education population. *Journal of Learning Disabilities*, *49*(2), 140–151. https://doi.org/10.1177/0022219414538513

Gaetano, J. (2018). *Holm-Bonferroni sequential correction: An Excel calculator (1.3): Microsoft Excel workbook*. https://www.researchgate.net/publication/322568540_Holm-Bonferroni_sequential_correction_An_Excel_calculator_13

Galuschka, K., & Schulte-Körne, G. (2016). The diagnosis and treatment of reading and/or spelling disorders in children and adolescents. *Deutsches Arzteblatt International*, *113*(16), 279–286. https://doi.org/10.3238/arztebl.2016.0279

Gegenfurtner, K. R., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology*, *10*(13), 805–808. https://doi.org/10.1016/S0960-9822(00)00563-7

Goldfarb, L., & Treisman, A. (2011). Does a color difference between parts impair the perception of a whole? A similarity between simultanagnosia patients and healthy observers. *Psychonomic Bulletin & Review*, *18*(5), 877–882. https://doi.org/10.3758/s13423-011-0123-8

Good, J. E., Lance, D. M., & Rainey, J. (2015). The effects of morphological awareness training on reading, spelling, and vocabulary skills. *Communication Disorders Quarterly*, *36*(3), 142–151. https://doi.org/10.1177/1525740114548917

Goodwin, A. P., Gilbert, J. K., & Cho, S.-J. (2013). Morphological contributions to adolescent word reading: An item response approach. *Reading Research Quarterly*, *48*(1), 39–60. https://doi.org/10.1002/rrq.037

Görgen, R., De Simone, E., Schulte-Körne, G., & Moll, K. (2021). Predictors of reading and spelling skills in German: The role of morphological awareness. *Journal of Research in Reading*, *44*(1), 210–227. https://doi.org/10.1111/1467-9817.12343

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

Grainger, J., Colé, P., & Segui, J. (1991). Masked morphological priming in visual word recognition. *Journal of Memory and Language*, *30*(3), 370–384. https://doi.org/10.1016/0749-596X(91)90042-I

Grainger, J., & Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology*, *2*, Article 54. https://doi.org/10.3389/fpsyg.2011.00054

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Häikiö, T., Bertram, R., & Hyönä, J. (2011). The development of whole-word representations in compound word processing: Evidence from eye fixation patterns of elementary school children. *Applied Psycholinguistics*, *32*(3), 533–551. https://doi.org/10.1017/S0142716411000208

Häikiö, T., Bertram, R., & Hyönä, J. (2016). The hyphen as a syllabification cue in reading bisyllabic and multisyllabic words among Finnish 1st and 2nd graders. *Reading and Writing*, *29*(1), 159–182. https://doi.org/10.1007/s11145-015-9584-x

Häikiö, T., Hyönä, J., & Bertram, R. (2015). The role of syllables in word recognition among beginning Finnish readers: Evidence from eye movements during reading. *Journal of*

Cognitive Psychology, *27*(5), 562–577. https://doi.org/10.1080/20445911.2014.982126

Häikiö, T., & Luotojärvi, T. (2022). The effect of syllable-level hyphenation on novel word reading in early Finnish Readers: Evidence from eye movements. *Scientific Studies of Reading*, *26*(1), 38–46. https://doi.org/10.1080/10888438.2021.1874384

Häikiö, T., & Vainio, S. (2018). Syllables and inflectional morphemes in early Finnish readers: Evidence from eye-movements. *Journal of Child Language*, *45*(5), 1227–1245. https://doi.org/10.1017/S0305000918000132

Hasenäcker, J., & Schroeder, S. (2017). Syllables and morphemes in German reading development: Evidence from second graders, fourth graders, and adults. *Applied Psycholinguistics*, *38*(3), 733–753. https://doi.org/10.1017/S0142716416000412

Hawelka, S., Schuster, S., Gagl, B., & Hutzler, F. (2013). Beyond single syllables: The effect of first syllable frequency and orthographic similarity on eye movements during silent reading. *Language and Cognitive Processes*, *28*(8), 1134–1153. https://doi.org/10.1080/01690965.2012.696665

Henderson, C. M., & McClelland, J. L. (2020). Intrusions into the shadow of attention: A new take on illusory conjunctions. *Attention, Perception & Psychophysics*, *82*(2), 564–584. https://doi.org/10.3758/s13414-019-01893-3

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Hutzler, F., Conrad, M., & Jacobs, A. M. (2005). Effects of syllable-frequency in lexical decision and naming: An eye-movement study. *Brain and Language*, *92*(2), 138–152. https://doi.org/10.1016/j.bandl.2004.06.001

Hyönä, J. (2012). Foveal and parafoveal processing during reading. In S. P. Liversedge (Ed.), *The Oxford handbook of eye movements* (pp. 820–838). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199539789.013.0045

Hyönä, J., & Pollatsek, A. (1998). Reading Finnish compound words: Eye fixations are affected by component morphemes. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(6), 1612–1627. https://doi.org/10.1037//0096-1523.24.6.1612

Hyönä, J., Yan, M., & Vainio, S. (2018). Morphological structure influences the initial landing position in words during reading Finnish. *Quarterly Journal of Experimental Psychology*, *71*(1), 122–130. https://doi.org/10.1080/17470218.2016.1267233

Inhoff, A. W., & Topolski, R. (1994). Use of phonological codes during eye fixations in reading and in on-line and delayed naming tasks. *Journal of Memory and Language*, *33*(5), 689–713. https://doi.org/10.1006/jmla.1994.1033

Jacobs, A. M., Rey, A., Ziegler, J. C., & Grainger, J. (1998). *MROM-p: An interactive activation, multiple readout model of orthographic and phonological processes in visual word recognition: Localist connectionist approaches to human cognition* (Scientific Psychology Series). Lawrence Erlbaum Associates.

Ji, H., Gagné, C. L., & Spalding, T. L. (2011). Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language*, *65*(4), 406–430. https://doi.org/10.1016/j.jml.2011.07.003

Juhasz, B. J., Starr, M. S., Inhoff, A. W., & Placke, L. (2003). The effects of morphology on the processing of compound words: Evidence from naming, lexical decisions and eye fixations. *British Journal of Psychology*, *94*(Pt. 2), 223–244. https://doi.org/10.1348/000712603321661903

Juola, P. (1998). Measuring linguistic complexity: The morphological tier. *Journal of Quantitative Linguistics*, *5*(3), 206–213. https://doi.org/10.1080/09296179808590128

Kazanina, N., Dukova-Zheleva, G., Geber, D., Kharlamov, V., & Tonciulescu, K. (2008). Decomposition into multiple morphemes during lexical access: A masked priming study of Russian nouns. *Language and Cognitive Processes*, *23*(6), 800–823. https://doi.org/10.1080/01690960701799635

Kearns, D. M., & Whaley, V. M. (2019). Helping students with dyslexia read long words: Using syllables and morphemes. *TEACHING Exceptional Children*, *51*(3), 212–225. https://doi.org/10.1177/0040059918810010

Kempe, V. (1999). Processing of morphological and semantic cues in Russian and German. *Language and Cognitive Processes*, *14*(2), 129–171. https://doi.org/10.1080/016909699386329

Kettunen, K. (2014). Can type-token ratio be used to show morphological complexity of languages? *Journal of Quantitative Linguistics*, *21*(3), 223–245. https://doi.org/10.1080/09296174.2014.911506

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*(1–2), 262–284. https://doi.org/10.1080/09541440340000213

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology. General*, *135*(1), 12–35. https://doi.org/10.1037/0096-3445.135.1.12

Kuperman, V., Bertram, R., & Baayen, R. H. (2010). Processing trade-offs in the reading of Dutch derived words. *Journal of Memory and Language*, *62*(2), 83–97. https://doi.org/10.1016/j.jml.2009.10.001

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Longtin, C.-M., Segui, J., & Hallé, P. A. (2003). Morphological priming without morphological relationship. *Language and Cognitive Processes*, *18*(3), 313–334. https://doi.org/10.1080/01690960244000036

Lopes, W. N., & Barrera, S. D. (2019). Effects of a syllabic highlight technique in word reading of beginning readers. *Estudos de Psicologia* (Natal), *24*(4), 349–358. https://doi.org/10.22491/1678-4669.20190035

Marcet, A., Perea, M., Baciero, A., & Gomez, P. (2019). Can letter position encoding be modified by visual perceptual elements? *Quarterly Journal of Experimental Psychology*, *72*(6), 1344–1353. https://doi.org/10.1177/1747021818789876

Mathey, S., Zagar, D., Doignon, N., & Seigneuric, A. (2006). The nature of the syllabic neighbourhood effect in French. *Acta Psychologica*, *123*(3), 372–393. https://doi.org/10.1016/J.ACTPSY.2006.02.003

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I.

An account of basic findings. *Psychological Review*, *88*(5), 375–407. https://doi.org/10.1037/0033-295X.88.5.375

Moll, K., & Landerl, K. (2010). *Lese- und Rechtschreibtest (SLRT II) [= Reading and spelling test (SLRT II)]*. Huber.

Mousikou, P., Beyersmann, E., Ktori, M., Javourey-Drevet, L., Crepaldi, D., Ziegler, J. C., Grainger, J., & Schroeder, S. (2020). Orthographic consistency influences morphological processing in reading aloud: Evidence from a cross-linguistic study. *Developmental Science*, *23*(6), Article e12952. https://doi.org/10.1111/desc.12952

O'Regan, J. K. (1992). Optimal viewing position in words and the strategy-tactics theory of eye movements in reading. In H. A. Whitaker & K. Rayner (Eds.), *Springer Series in Neuropsychology. Eye movements and visual cognition* (pp. 333–354). Springer. https://doi.org/10.1007/978-1-4612-2852-3_20

Perea, M., & Carreiras, M. (1998). Effects of syllable frequency and syllable neighborhood frequency in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(1), 134–144. https://doi.org/10.1037/0096-1523.24.1.134

Perea, M., & Wang, X. (2017). Do alternating-color words facilitate reading aloud text in Chinese? Evidence with developing and adult readers. *Memory & Cognition*, *45*(7), 1160–1170. https://doi.org/10.3758/s13421-017-0717-0

Perfetti, C. A., Zhang, S., & Berent, I. (1992). Reading in English and Chinese: Evidence for a "universal" phonological principle. *Advances in Psychology*, *94*, 227–248. https://doi.org/10.1016/S0166-4115(08)62798-3

Peterson, M. E., & Haines, L. P. (1992). Orthographic analogy training with kindergarten children: Effects on analogy use, phonemic segmentation, and letter-sound knowledge. *Journal of Reading Behavior*, *24*(1), 109–127. https://doi.org/10.1080/10862969209547764

Pollatsek, A., Hyönä, J., & Bertram, R. (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 820–833. https://doi.org/10.1037/0096-1523.26.2.820

Pollatsek, A., Lesch, M., Morris, R. K., & Rayner, K. (1992). Phonological codes are used in integrating information across saccades in word identification and reading. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 148–162. https://doi.org/10.1037//0096-1523.18.1.148

Prinzmetal, W., Treiman, R., & Rho, S. H. (1986). How to see a reading unit. *Journal of Memory and Language*, *25*(4), 461–475. https://doi.org/10.1016/0749-596X(86)90038-0

R Core Team. (2021). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rastle, K. (2019). The place of morphology in learning to read in English. *Cortex*, *116*, 45–54. https://doi.org/10.1016/j.cortex.2018.02.008

Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, *23*(7–8), 942–971. https://doi.org/10.1080/01690960802069730

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, *11*(6), 1090–1098. https://doi.org/10.3758/bf03196742

Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception*, *8*(1), 21–30. https://doi.org/10.1068/p080021

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K., Pollatsek, A., & Binder, K. S. (1998). Phonological codes and eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(2), 476–497. https://doi.org/10.1037/0278-7393.24.2.476

Reh, S., & Wilde, D. (2016). *Bildungsgeschichte. Forschung—Akzente—Perspektiven. Die Materialität des Schreiben- und Lesenlernens: Zur Geschichte schulischer Unterweisungspraktiken seit der Mitte des 18. Jahrhunderts [History of Education. Research - Trends - Perspectives. The Materials of writing and reading acquisition: On the history of scohlastic Intructional practices since the mid-18th century]*. Verlag Julius Klinkhardt.

Rice, M. L., Leske, D. A., Smestad, C. E., & Holmes, J. M. (2008). Results of ocular dominance testing depend on assessment method. *Journal of AAPOS*, *12*(4), 365–369. https://doi.org/10.1016/j.jaapos.2008.01.017

Schiff, R., Raveh, M., & Fighel, A. (2012). The development of the Hebrew mental lexicon: When morphological representations become devoid of their meaning. *Scientific Studies of Reading*, *16*(5), 383–403. https://doi.org/10.1080/10888438.2011.571327

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception & Psychophysics*, *74*(1), 5–35. https://doi.org/10.3758/s13414-011-0219-2

Schroeder, S., -Würzner, K. M., Heister, J., Geyken, A., & Kliegl, R. (2015). Childlex: A lexical database of German read by children. *Behavior Research Methods*, *47*(4), 1085–1094. https://doi.org/10.3758/s13428-014-0528-1

Seidenberg, M. S. (1987). Sublexical structures in visual word recognition: Access units or orthographic redundancy? In M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading* (pp. 245–263). Lawrence Erlbaum Associates.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(Pt. 2), 143–174. https://doi.org/10.1348/000712603321661859

Smolka, E., Komlósi, S., & Rösler, F. (2009). When semantics means less than morphology: The processing of German prefixed verbs. *Language and Cognitive Processes*, *24*(3), 337–375. https://doi.org/10.1080/01690960802075497

Song, Z., Liang, X., Wang, Y., & Yan, G. (2021). Effect of alternating-color words on oral reading in grades 2–5 Chinese children: Evidence from eye movements. *Reading and Writing*, *34*(10), 2627–2643. https://doi.org/10.1007/s11145-021-10164-3

Sparrow, L., & Miellet, S. (2002). Activation of phonological codes during reading: Evidence from errors detection and eye movements. *Brain and Language*, *81*(1–3), 509–516. https://doi.org/10.1006/brln.2001.2543

Spoehr, K. T., & Smith, E. E. (1973). The role of syllables in perceptual processing. *Cognitive Psychology*, *5*(1), 71–89. https://doi.org/10.1016/0010-0285(73)90026-1

Stenneken, P., Conrad, M., & Jacobs, A. (2007). Processing of syllables in production and recognition tasks. *Journal of Psycholinguistic Research*, *36*(1), 65–78. https://doi.org/10.1007/s10936-006-9033-8

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 638–647. https://doi.org/10.1016/S0022-5371(75)80051-X

Tanaka, J., Weiskopf, D., & Williams, P. (2001). The role of color in high-level vision. *Trends in Cognitive Sciences*, *5*(5), 211–215. https://doi.org/10.1016/S1364-6613(00)01626-0

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*(1), 107–141. https://doi.org/10.1016/0010-0285(82)90006-8

van Heuven, W. J., Dijkstra, T., Grainger, J., & Schriefers, H. (2001). Shared neighborhood effects in masked orthographic priming. *Psychonomic Bulletin & Review*, *8*(1), 96–101. https://doi.org/10.3758/BF03196144

Velten, H. R. (2012). Frühe Lese- und Schreiblernbücher des 16. Jahrhunderts. Zu Valentin Ickelsamers Die rechte weis, aufs kürtzist lesen zu lernen (1527) und Teütsche Grammatica (1532?) [Early reading and writing books of the 16th century. On Valentin Ickelsamer's The wise knows how to learn to read in a nutshell (1527) and Teutsche Grammatica (1532?)]. *Zeitschrift für Erziehungswissenschaft*, *15*(S2), 31–48. https://doi.org/10.1007/s11618-012-0309-5

Yan, M., Zhou, W., Shu, H., Yusupu, R., Miao, D., Krügel, A., & Kliegl, R. (2014). Eye movements guided by morphological structure: Evidence from the Uighur language. *Cognition*, *132*(2), 181–215. https://doi.org/10.1016/j.cognition.2014.03.008

Zhou, W., Liu, Y., Su, M., Yan, M., & Shu, H. (2019). Alternating-color words influence Chinese sentence reading: Evidence from neural connectivity. *Brain and Language*, *197*, Article 104663. https://doi.org/10.1016/j.bandl.2019.104663

Zhou, W., Wang, A., Shu, H., Kliegl, R., & Yan, M. (2018). Word segmentation by alternating colors facilitates eye guidance in Chinese reading. *Memory & Cognition*, *46*(5), 729–740. https://doi.org/10.3758/s13421-018-0797-5

Appendix: Study 3

The Role of Orthographic Transparency and Morphological complexity when Reading Complex Nonwords: evidence from English and Italian.

Elisabetta De Simone

Kristina Moll

Elisabeth Beyersmann

Manuscript submitted for publication in the *Journal of Scientific Studies of Reading*

Abstract

Purpose: The present study examined cross-linguistic differences in orthographic transparency and morphological complexity during complex word recognition. If morphological processing is more important in morphemically rich languages, we would expect larger morpheme-effects in Italian. However, if morphological processing is more important in opaque orthographies morpheme-effects should be larger in English.

Method: 60 Italian and 60 English native-speakers completed an online lexical decision task, while reaction times and accuracy were measured. To tease apart the independent role of stems and suffixes, we employed four types of nonwords: Stem+Suffix: *night+ness*, Stem+NonSuffix: *night-lude*, NonStem+Suffix: *nisht+ness*, NonStem+NonSuffix: *nisht-lude*).

Results: The results revealed a significant morpheme interference effect in both languages: nonwords with stems were read slower and less accurately than those without stems and nonwords with suffixes slower and less accurately than those without suffixes. Crucially, this observed pattern was larger in English than Italian. Also, a significant stem-by-suffix interaction suggested that Stem+Suffix nonwords were harder to reject than all others.

Conclusion: The current findings suggest that morphological processing is more pronounced in opaque orthographies like English, possibly because the activation of morphemic chunks can be used to compensate for grapheme-to-phoneme inconsistencies, even in a silent reading task where phonological decoding is not a necessity.

Introduction

Many studies have investigated mechanisms of morphological processing in the reading of single languages, particularly English, but only few have investigated how morphological processing differs across languages. The goal of the current study was to examine cross-linguistic differences in orthographic depth and morphological complexity in the reading of morphologically complex words.

On the Role of Orthographic Depth in Morphological Processing

Orthographic depth refers to the consistency of grapheme-phoneme correspondences (GPCs) in alphabetic orthographies. In transparent orthographies the mapping between the phonemic and orthographic code is mostly isomorphic, whereas in opaque orthographies this correspondence is more intricate and unpredictable. This apparently simple orthographic construct has been consistently found to influence reading mechanisms across languages (e.g., Ziegler, 2010). According to the Orthographic Depth Hypothesis (Katz & Feldman, 1983; Katz & Frost, 1992) readers of transparent orthographies have an easier access to prelexical phonology, since they can use the consistent GPCs of their orthography to retrieve the correct pronunciation of novel and unfamiliar words. Conversely, readers of opaque orthographies have been found to read primarily via the lexical route by using groups of letters, such as bodies or morphemes, and other types of lexical information specific to each word (Miller, 2018). It has been found that orthographic depth affects both reading aloud (Schmalz et al., 2016; De Simone et al., 2021) and silent reading (Rau et al., 2015; Schroeder et al., 2022), reading in skilled and beginning readers (Marinelli et al., 2016; Rau et al., 2016; Rau et al., 2015), and also determines specific symptoms associated with developmental dyslexia (Marinelli et al., 2023; Provazza et al., 2022). Even more related to the current study, findings from Mousikou et al. (2020) and Vannest et al. (2002) further suggest that morphological processing is modulated by cross-linguistic differences in orthographic depth, indicating that readers of languages that are orthographically opaque are more likely to parse letter strings into morphemic subunits than speakers of languages that are orthographically transparent.

Relatedly, the Psycholinguistic Grain Size Theory (Ziegler & Goswami, 2005) predicts that the emergence of larger-than-letters reading units depends on the reader's needs, which in turn are dictated by the depth of their specific language orthography (Marinus & Jong, 2008). In English, for example, children learning to read would learn to rely on larger sublexical units such as bodies or morphemes, because English is considered to have a opaque orthography (Seymour et al., 2003), and larger sublexical units are more consistent than smaller ones (Treiman et al., 1995). For instance, in English GPCs are not isomorphic: several letters might represent a single phoneme (e.g., *might* - /maɪt/), the same grapheme might correspond to different phonemes (e.g., *river* - /ˈrɪvər/ vs *driver*- /ˈdraɪvər/) and pronunciation may depend on the context as well (e.g., *read* in present vs past tense). In contrast, in an orthographically transparent language like Italian, readers may need to rely less on larger sublexical units (Marinelli et al., 2016), which is consistent with Mousikou et al.'s (2020) and Vannest et al.'s (2002) findings suggesting that morphological processing is enhanced in opaque orthographies.

Cross-Linguistic Differences in Morphological Complexity

A further factor that has been found to influence cross-linguistic differences in morphological processing are language specific disparities in morphological complexity, i.e., the complexity of words' internal structure. Historically, morphological typologists have grouped languages based on the transparency of morphological boundaries between morphemes, classifying them as *agglutinative, isolating* or *fusional* (Iacobini, 2006b; Schlegel, 1808). For example, in agglutinating languages (such as Turkish and Finnish) the boundaries between morphemes are easily recognizable, and each morpheme is conveyed by one single morph (the concrete written or oral realization of the morpheme). Contrarywise, in fusional languages (such as Italian and English) the boundary between stem and affix tend to blend. Several morphemes may also correspond to the same morph (see the inflectional morph "*a*" in the Italian word "*bambin-a*", which indicates both the singular number and gender).

In the early 20th century, Sapir (1921) added a second typological parameter relying on the number of morphemes per word, further dividing languages in *analytic*, *synthetic* and *polysynthetic*. Words of analytic languages (such as Chinese) tend to have none or few bound morphemes, whereas synthetic languages (such as Italian) build words made of several morphemes. When the combined number of morphemes is large, the language is considered to be polysynthetic (like Finnish). It is important to point out that languages rarely fully belong to one category or another (see also Brown, 2010; Greenberg, 1954), and that Schlegel's and Sapir's parameters are to be considered independent from each other. For example, while English has fusional features, its abundant percentage of monomorphemic words showcases usage of analytic constructions (Aikhenvald, 2007).

More recent attempts at employing quantitative strategies to place European languages on a morphological complexity continuum are based on indicators such as the number of inflectional categories and the combination of morpheme types (Bane, 2008; Juola, 2008), number of cases and vocabulary size divided by text length (Kettunen, 2014), or expression of number (Stump, 2001). Although these measures disagree on the relative placement of the languages towards the center of the continuum, they have consistently found English to be the least morphologically complex language and Finnish to be the most morphologically complex one (see Borleffs et al., 2017 for a review). From a psycholinguistic perspective, research on cross-linguistic differences driven by different degrees of morphological complexity has been scarce. However, morphology has been found to impact critical areas of reading, such as reading comprehension (Carlisle et al., 2010; Frost et al., 2005) or spelling (e.g., Görgen et al., 2021), and most importantly, it has been suggested that morphological processing is modulated by morphological complexity (Beyersmann et al., 2020; Casalis et al., 2015; Haddad et al., 2017), showing that readers of languages that are more morphologically complex are more likely to parse letter strings in morphemic subunits than speakers of languages that are less complex.

Mousikou et al. (2020) conducted a comprehensive investigation of the matter by comparing the performances of speakers of four different alphabetic orthographies (English, French, German, Italian) with a different degree of orthographic consistency and morphological complexity in a reading aloud task. Participants had to name morphologically structured and non-morphologically structured nonwords, which derived from between-language cognate words. Nonwords belonged to one of these four conditions: Stem+Suffix (*night-ness*), Stem + NonSuffix (*night-lude*), NonStem+Suffix (*nisht-ness*) and NonStem+Nonsuffix (*nisht-lude*). Morphologically simple and morphologically complex words were also added to the task, to prevent participants from developing strategies relying on sublexical units only (Mousikou et al., 2020, p. 4). The authors found that morphological processing was more prominent in English, the language with the most opaque orthography, and the poorest morphology, of the four languages. The authors argued therefore that it is orthographic consistency, and not morphological complexity, that influenced morphological processing. While Mousikou et al. (2020) provide compelling evidence for the relative roles that orthographic depth and morphological complexity play in morphological processing during reading aloud, it is uncertain whether their results from spoken word production generalise to the modality of silent reading, where a phonological output is not required.

Indeed, the results obtained from the visual lexical decision task of Beyersmann et al. (2020), comparing two languages with varying degrees of morphological complexity (German vs. French), revealed that the speakers of the morphologically richer language (German) exhibited more robust morphological processing than the speakers of the morphologically poorer language (French), thus demonstrating the influence of morphological productivity on morphological processing. However, this study was not designed to test the role of orthographic transparency within visual word recognition, which is reflected by the authors' choice of comparing German and French. In fact, although the German orthography is considered to be more transparent than French (Seymour et al., 2003), French is entirely predictable in the reading direction (Schmalz et al., 2016; De Simone et al., 2021), thus these two languages do not differ much in terms of orthographic depth in the reading direction.

More relevant with respect to the cross-linguistic investigation of orthographic transparency is a study by Vannest et al. (2002), who carried out a series of lexical decision experiments with English and Finnish speakers. These two languages fall onto the extreme poles of the orthographic depth continuum, with English being at its opaque end and Finnish being at its transparent end, but also differ in their morphological complexity, with English being less morphologically complex than Finnish, which is a polysynthetic, agglutinating language (Bölücü & Can, 2019; Borleffs et al., 2017; Pirkola, 2001). The authors found that English-speaking participants were more sensitive to stem frequencies than Finnish-speaking participants. One possible explanation for this finding is that morphological processing is more important in orthographically opaque language like English, which would be line with Mousikou et al.'s earlier findings.

However, Stevens and Plaut (2022) argue that Vannest et al.'s findings may be ascribable to the fact that, in order to keep the items comparable between English and Finnish, Vannest et al. had to employ bi-morphemic items. While these items are representative enough of English, where bi-morphemic words are common (Vannest et al., 2002, p. 104), a polysynthetic language's vocabulary like Finnish encompass many multi-morphemic words. Vannest et al. do acknowledge this difference: for example, they calculated that the suffix -TÖN, one of the suffixes used by in the Finnish experiment, has a chance to be followed by an additional suffix in 87.5% of its occurrence, while the English suffix -ABLE is only followed by an additional suffix in 11.8% of the cases. Stevens and Plaut argue that because of this, words with a single derivational suffix, like those employed by Vannest et al., may have been processed by Finnish participants similarly to monomorphemic stems, something that could account for the poor stem frequency effects found in this group (p. 1691). Finally, it is worth noting that some of the suffixes used in the English stimuli were not only bound affixes, like in the Finnish items, but stand-alone morphemes, such as *-able, -hood, -ship* and *-less*, and therefore had a more of a compound-type status. In sum, existing cross-linguistic evidence around the role of orthographic transparency and morphological complexity in reading is not conclusive and therefore formed the focus of investigation in the current study.

Present Study

The present study was designed to provide more insights to the topic by investigating cross-linguistic differences in silent reading of complex nonwords. To address this aim, we compared English, which is orthographically opaque and morphologically less complex; and Italian, which is orthographically transparent and morphologically more complex (Borgwaldt et al., 2005; Borleffs et al., 2017; Kettunen, 2014; Pagliuca & Monaghan, 2010; Seymour et al., 2003). Typologically, we are contrasting two languages that are both fusional but with a different degree of synthesis, with English tending towards analytical constructs and Italian tending towards synthetical constructs. In both languages, however, bi-morphemic words are common, thus ensuring the representativeness of the experimental items for both languages, and overcoming one of the main criticism of Vannest et al.'s (2002) design (Stevens & Plaut, 2022). Participants performed an online visual lexical decision task, using Mousikou et al.’s (2020) items. The analysis focused on the four types of nonwords: Stem+Suffix (e.g., *night-ness*), NonStem+Suffix (e.g., *nisht-ness*), Stem+NonSuffix (e.g., *night-lude*) and NonStem+NonSuffix (e.g., *nisht-lude*).

The main effect of Stem was examined by comparing the two stem conditions (Stem+NonSuffix; Stem+Suffix) with the two non-stem conditions (NonStem+NonSuffix; NonStem+Suffix) where the presence of stems was expected to hinder the NO response in the lexical decision task (i.e., slower and less accurate responses in the stem vs. non-stem conditions, when rejecting nonwords). The main effect of Suffix was examined by comparing the two suffix conditions (NonStem+Suffix; Stem+Suffix) with the two non-suffix conditions (NonStem+NonSuffix; Stem+NonSuffix), where the presence of suffixes was expected to hinder the NO response in the lexical decision task (i.e., slower and less accurate responses in the suffix vs. non-suffix conditions, when rejecting nonwords). Additionally, assuming that visual word recognition is sensitive to the full decomposability of complex letter strings, we predicted a significant interaction between Stem and Suffix, showing a larger stem effect for suffixed than non-suffixed nonwords (*night-ness* vs *night-lude*), and a larger suffix effect for nonwords including stems vs. non-stems (*night-ness* vs *nisht-ness*).

We hypothesised that if morphological processing is modulated by language specific differences in orthographic transparency, there should be a significant Language-by-Stem and a significant Language-by-Suffix interaction with greater stem and suffix effects in English compared to Italian participants. However, if morphological processing is modulated by language specific differences in morphological complexity, there should be a significant Language-by-Stem and a significant Language-by-Suffix interaction with greater stem and suffix effects in Italian compared to English participants. A third possibility was that morphological processing will be comparable across both languages. This could either be because the effects of morphological complexity and orthographic transparency cancel each other out, or because neither has an effect on morphological processing.

These hypotheses and the corresponding data analysis plan were pre-registered: https://aspredicted.org/FRM_C8V

Methods

Participants

Using the package SimR (Green & MacLeod, 2016) in the R computing environment (R Core Team, 2021), based on previous data showing significant cross-linguistics differences in morphologic processing (Beyersmann et al., 2020; Mousikou et al., 2020), we predicted that to obtain a power of 80% for small effect size (d = 0.01), we needed a minimum sample of sixty individuals in the Italian and English participant sample.

In total, 66 monolingual Australian English speakers participated in exchange for course credit, and 73 monolingual Italian speakers participated for monetary reimbursement. Australian participants were Macquarie University undergraduate students, whereas Italian participants were recruited through Prolific (https://www.prolific.co). All participants completed the study online and confirmed to have normal or corrected-to-normal vision. Furthermore, as we wanted to recruit skilled readers, we included standardised reading tests in the online study. We excluded from the analysis six Australian English speakers on the basis of TOWRE-2 (Torgesen et al.,

2012) and thirteen Italian speakers on the basis of MT-16-19 (Cornoldi & Candela, 2014), thus achieving the pre-registered sample size for both groups (see Table 1 for demographic data, and further below for the results of the reading tests).

The study was approved by the Human Sciences Subcommittee of Macquarie University and conforms with the ethical principles of the Declaration of Helsinki. Prior to participating in the study, participants provided written, informed consent.

Table 1.

Demographic Characteristics

Group	English	Italian
Gender		
Females	51	33
Males	9	27
Age		
Mean	25.46	25.33
SD	10.17	3.83
Education		
HighSchool D.	54	24
College D.	6	36

Materials

The Italian and English items were adopted from Mousikou et al. (2020) including sixty words and sixty nonwords. The authors chose English nouns from the Celex database (Baayen et al., 1995) and Italian nouns from the SUBTLEX-IT (Crepaldi et al., 2015). Stems were translation-equivalent across languages, and so were Suffixes, when possible. Words were frequent nouns that could be either suffixed (Stem+Suffix, e.g., clearing) or not (Stem+NonSuffix, e.g., chest). Each list had thirty morphologically complex and thirty morphologically simple words. Nonwords were created from frequent nouns and could be built around a real stem with a real suffix (Stem+Suffix, e.g., armful), a real stem with a fake suffix

(Stem+NonSuffix, e.g., dognule), a fake stem with a real suffix (NonStem+Suffix, e.g., selseness) or without any apparent morphological structure (NonStem+NonSuffix, e.g., tervan). Item frequency for English were obtained from SUBTLEX-UK (van Heuven et al., 2014) and SUBTLEX-IT for Italian. Furthermore, we included Orthographic Levenshtein distance (OLD20: Yarkoni et al., 2008), word length, biphone frequency and number of syllables of each item as a covariate in the analysis, as the authors reported that they varied significantly across languages (Mousikou et al., 2020, p. 5). Four counterbalanced experimental lists were created to avoid participants seeing any stem/non-stem more than once. The psycholinguistic properties of items can be found in Tables 1 and 2.

The full list of materials is available in the following online repository: https://osf.io/4g2vs/?view_only=86e666b9fa9445ed987c5dfef41fa9b5

Table 2.

Psycholinguistic Properties of Nonwords.

	English		Italian	
	M	*SD*	*M*	*SD*
Nonwords				
Stem + Suffix				
OLD20	2.5	(0.5)	2.3	(0.6)
N letters	8.0	(1.3)	8.3	(1.7)
N phonemes	6.8	(1.2)	7.9	(1.7)
N syllables	2.1	(0.3)	3.5	(0.6)
Biphone frequency	10.3	(0.6)	11.1	(0.3)
Stem + Non-Suffix				
OLD20	3.0	(0.7)	2.6	(0.7)

N letters	8.0	(1.2)	8.3	(1.7)
N phonemes	6.6	(1.1)	7.9	(1.7)
N syllables	2.1	(0.3)	3.5	(0.6)
Biphone frequency	10.2	(0.5)	11.1	(0.3)
Non-Stem + Suffix				
OLD20	2.7	(0.6)	2.5	(0.6)
N letters	8.0	(1.3)	8.3	(1.7)
N phonemes	6.8	(1.2)	7.9	(1.7)
N syllables	2.1	(0.3)	3.5	(0.6)
Biphone frequency	10.3	(0.6)	11	(0.3)
Non-Stem + Non-Suffix				
OLD20	3.2	(0.7)	2.9	(0.8)
N letters	7.9	(1.2)	8.3	(1.7)
N phonemes	6.6	(1.1)	7.9	(1.7)
N syllables	2.1	(0.3)	11.0	(0.3)
Biphone frequency	10.1	(0.5)	11.0	(0.3)

Note. Properties taken from Mousikou et al. (2020).

Procedure

We used Gorilla Experiment Builder (www.gorilla.sc) to create and host online both the English and the Italian experiments. Participants were instructed to sit in a quiet environment, close other applications running on their device, and to deactivate notifications. Before starting the experiment, participants answered a short survey created to screen bilingual speakers and individuals with learning or cognitive impairments, as we wished to recruit monolingual, neurotypical readers. Then, demographic data such as age, gender and highest level of education were collected.

During the experiment, participants were instructed to indicate, as quickly and as accurately as possible, whether a string of letters was a word or a nonword. To perform the task, they were advised to hold the index finger of their left hand on the “Z” key for nonwords, and the index finger of their right hand on the “M” key for words. Before starting the experimental session, participants responded to ten practice items and were given feedbacks about the accuracy of their response. The experimental items were divided in two blocks of sixty items each. Each participant was randomly assigned to one of the four lists available (fifteen participants per list). The order of the items within the blocks was randomised, and a fixation cross appeared before each item. On average, participants took 4 minutes to complete this task. Reaction Times (RT) and accuracy were measured.

Finally, the last phase of the experiment consisted of a reading test. Participants were alerted that their answers would be recorded for later assessment. Reading skills of English-speaking participants were assessed through the standardised Test of Word Reading Efficiency 2 (Torgesen et al., 2012), Form A. Participants had to read aloud as quickly and as accurately as possible as many words/non-words as possible out of a list of 104 words (measuring sight word reading efficiency) and a list of 63 non-words (measuring phonetic decoding) in fourty-five seconds. We calculated the number of items that were read correctly within the time limits, and then converted the results into the age-based standard scores and percentiles. The mean standard score for the word reading was 89.3 (SD = 8.4, percentile = 26) and 104.8 (SD = 11.2, percentile = 59) for the non-word reading.

Italian speakers’ reading skills were assessed using the standardised MT-16-19 reading test (Cornoldi & Candela, 2014). Participants had to read aloud as quickly and as accurately as possible a list of 112 words and a list of 56 nonwords, while the time for reading each list was measured. We calculated the total number of syllables per second and number of errors per list.

In average, the total number of syllables read per second was 8 for words (SD = 1.4) and 2.8 for nonwords (SD = 0.5). The mean number of errors per list for words was 1.1 (SD = 1.2) and 2 for nonwords (SD = 2). For both measures, Italian participants fell at least in the PS category ("Performance is sufficiently good"), when compared to normative data collected on a sample of 1060 students for words, and 1063 students for nonwords.

Results

Practice trials were excluded from the analyses, as well as one item in the English list that was erroneously labelled as a nonword ("armful"). As the focus of this study is on nonwords, we also excluded words from the dataset, which were originally included only for the purpose of the lexical decision task. Outliers were detected using Q-Q plots for reaction times: data points smaller than 200 ms or exceeding 2,500 ms were excluded (1.6%). Additionally, we applied a 2.5 residual outlier trimming procedure (Baayen, 2008), further removing 3.4% of the data. Response time distributions were checked using the Box-Cox system of the powerTransform function in the CAR package (Box & Cox, 1964; Fox & Weisberg, 2019), which suggested to inverse reaction times to reduce skewedness ($\lambda = -0.94$). In total, 6,827 observations were available for analysis.

Reaction times and accuracy data were analyzed in the R computing environment (R Core Team, 2021). Linear mixed effects models were constructed using the lme4 package (Bates et al., 2015) with three fixed effects calculating the investigated interactions (Language by Stem, Language by Suffix and Stem by Suffix) plus four additional control measures (scaled orthographic levenshtein distance, word length, bigram frequency and number of syllables) as well as random slopes and random intercepts for participants and items. As per Barr and colleagues (2013), models were computed with the maximal random effects structure, but these were overfitted (Baayen, 2008). Next, the random intercepts model was computed and random slopes were added incrementally. The highest converging nonsingular models are reported (Matuschek et al., 2017). The final RT's model random structure included random intercepts for items and participants, and by-stem random slopes for participants, whereas only random intercepts were included in the accuracy model.

We used the Anova function (type III) within the car package (version 3.0-12, Fox & Weisberg, 2019) to calculate *p*-values and the effsize package (version 0.6.0.1, Ben-Shachar et al., 2020) to calculate Cohen's d. Language groups, presence of stem and presence of suffix were coded using sum-to-zero contrasts. The detailed analyses scripts and results are reported within the R Markdown file as part of the study's Open Science Framework repository:

https://osf.io/4g2vs/?view_only=86e666b9fa9445ed987c5dfef41fa9b5

Reaction times

Only correct answers were included in this analysis, amounting to 6,401 observation points. Mean reaction times are reported in Figure 1 while the model's results are reported in Table 3.

Table 3.

Summary of linear mixed-effects analyses for nonword RTs and accuracy.

Variables	RTs		Accuracy	
	χ^2	*p*	χ^2	*p*
(Intercept)	3523.7635	< .001	588.5686	< .001
Language	4.0066	= .045	8.0783	= .004
Stem	99.5021	< .001	81.1646	< 0.001
Suffix	89.2265	< .001	59.0154	< 0.001
Old20	13.3692	< .001	4.1567	= .041
Length	67.8215	< .001	5.6486	= .017
N° of Syllables	0.1859	= .666	2.345	= .125
Bigram Freq.	25.8415	< .001	6.1229	= .013
Language:Stem	9.5996	= .001	10.7662	= .001
Language:Suffix	14.9052	< .001	6.8034	= .009
Stem:Suffix	32.9238	< .001	70.4499	< .001

Stem and Suffix Effects.

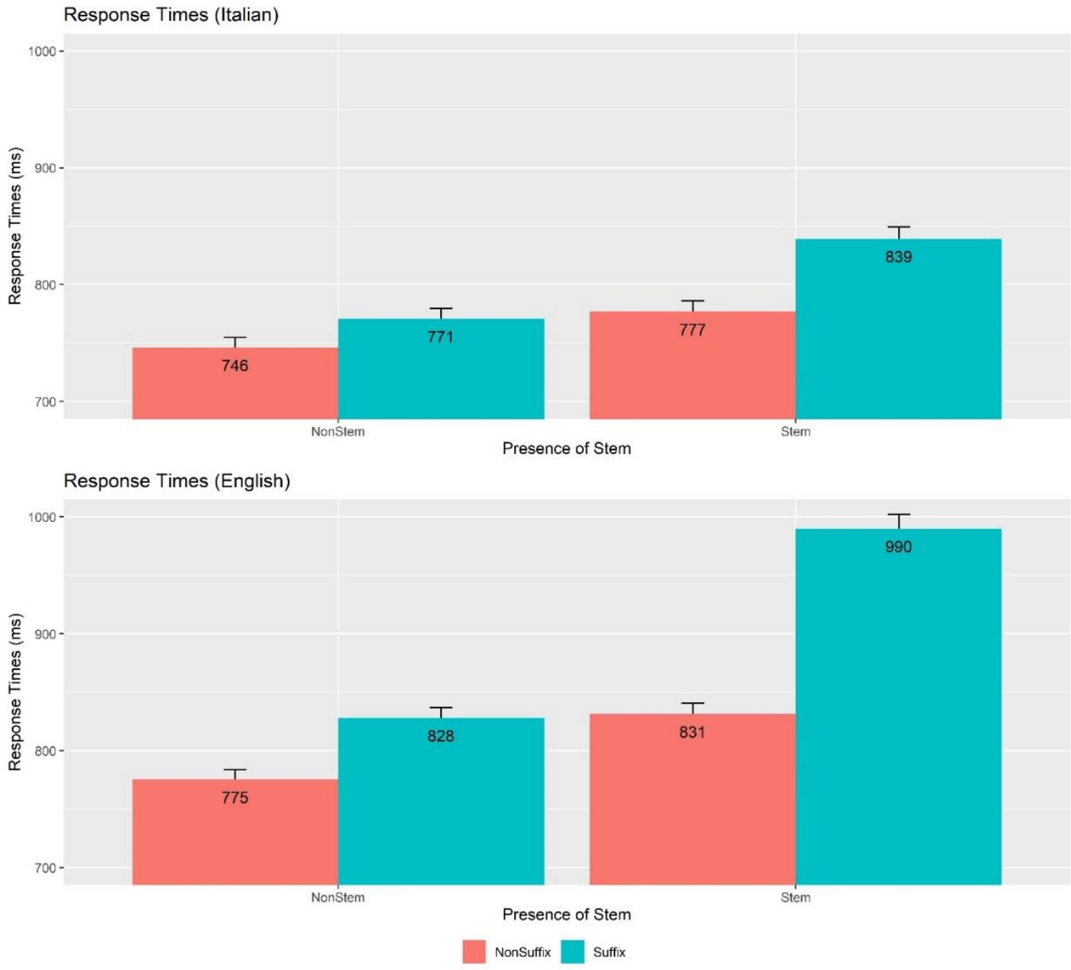
The main effect of Stem was significant ($\chi 2 = 99.50$; $p < 0.001$). Nonwords with stems were rejected significantly slower than nonwords without stems ($\Delta = 70$; $d = 0.26$, $t = -9.97$). The main effect of Suffix was also significant ($\chi 2 = 89.22$; $p < 0.001$), as nonwords with suffixes were rejected significant slower than nonwords without suffixes ($\Delta = 66$, $d = 0.24$, $t = -9.51$). Finally, the Stem by Suffix interaction was also significant ($\chi 2 = 32.92$; $p < 0.001$). In fact, we found a larger Stem effect for suffixed nonwords ($\Delta = 108$, $d = 0.39$, $t = 11.26$) compared to non-suffixed nonwords ($\Delta = 43$, $d = 0.16$, $t = -4.10$) and a larger Suffix effect for nonwords with stems ($\Delta = 103$, $d = 0.37$, $t = 10.90$) compared to nonwords without stems ($\Delta = 38.59$, $d = 0.15$, $t = 3.38$).

Language effects.

The main effect of Language was significant ($\chi2 = 4$; $p = 0.04$). English participants were significantly slower than Italian Participants ($\Delta = 64$, $d = -0.24$, $t = -2.02$). The Language by Stem interaction was significant ($\chi2 = 9.59$; $p = 0.001$), suggesting that the Stem effect was larger in English ($\Delta = 97$, $d = 0.36$, $t = -9.32$) than in Italian ($\Delta = 47$, $d = 0.18$, $t = 5.39$). There was also a significant Language by Suffix interaction ($\chi2 = 14.90$; $p < 0.001$), showing that the Suffix effect was larger in English ($\Delta = 97$, $t = 8.35$; $d = 0.35$) than in Italian ($\Delta = 41$, $t = 4.56$; $d = 0.15$; see Figure 1).

Figure 1

RTs in Italian and English per Item Type.



Note. From left to right. NonStem+NonSuffix (e.g., nisht-lude), NonStem+Suffix (e.g., nisht-ness), Stem+NonSuffix (e.g., night-lude), Stem+Suffix (e.g., night-ness)

Accuracy

Overall, accuracy results reflect response time results. Error rates per language group are reported in Figure 2, and the model's output is reported in Table 3.

Stem and Suffix Effects.

The main effect of Stem was significant ($\chi 2 = 81.16$; $p < 0.001$). In general, nonwords with stems were rejected with significantly lower accuracy than nonwords without stems ($\Delta = 0.09$, $d = 0.35$, $t= 9.00$). The main effect of Suffix was also significant ($\chi 2 = 59.01$; $p < 0.001$), as nonwords with suffixes were rejected with significant less accuracy than nonwords without suffixes ($\Delta = 0.08$, $d = 0.35$, $t= 7.68$).

The Stem by Suffix interaction was significant ($\chi 2 = 70.44$; $p < 0.001$), in fact, the presence of a stem had a bigger impact on accuracy on suffixed nonwords ($\Delta = 0.15$, $d = -0.52$, $t = -12.41$, $p < .001$) than on non-suffixed nonwords ($\Delta = 0.01$, $d = -0.09$, $t = 0.68$, $p = 0.49$). Conversely, the suffix effect was larger for nonwords with stems ($\Delta = 0.15$, $d = -0.52$, $t = 11.47$, $p < .001$) compared to nonwords without stems ($\Delta = 0.01$, $d = 0.09$, $t = 0.32$, $p = 0.74$).

Language Effects.

The Language effect was significant ($\chi 2 = 8.07$; $p = .004$), with Italians being more accurate than English participants ($\Delta = 0.03$, $d = 0.11$, $t = 2.84$). The Stem effect was significant in both languages ($p < .001$), with nonwords with stems being classified more inaccurately compared to nonwords without stems in both languages. The Language by Stem interaction was significant ($\chi 2 = 10.76$; $p = 0.001$), showing that the effect was larger in English compared to Italian ($\Delta = 0.11$ vs $\Delta = 0.06$, $t = 8.86$ vs 4.18; $d = 0.42$ vs 0.27).

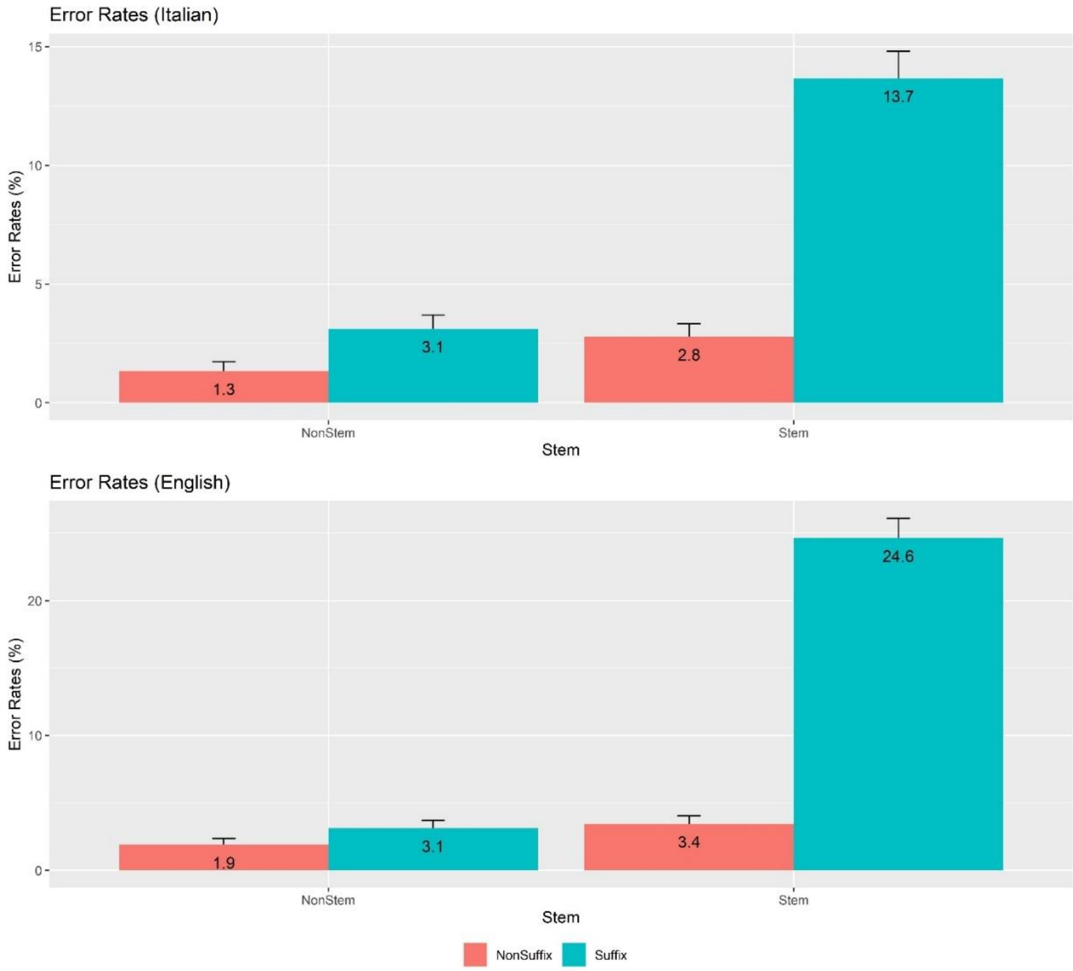
The Suffix effect was also significant In both languages ($p < .001$), with suffixed nonwords being incorrectly classified to a greater extent compared to non-suffixed. Similarly to the stem effect, the Language by suffix interaction was significant ($\chi 2 = 6.80$; $p = 0.009$), as the suffix effect was larger in English compared to Italian ($\Delta = 0.11$ vs $\Delta = 0.06$, $t= 7.32$ vs 4.20; $d = 0.41$ vs 0.28).

Post-hoc analysis.

Although the focus of the current study was on the analysis of the two-way interactions between Language * Stem, and Language * Suffix, an additional, non-preregistered analysis including the three-way interaction between Language, Stem, and Suffix was carried out to explore whether the size of the Stem*Suffix interaction differed between the two languages. The results revealed a significant three-way interaction in the accuracy analysis ($\chi2 = 9.87$; $p = .001$), but not in the RT analysis ($\chi2 = 3.10$; $p = .07$), suggesting that the Stem*Suffix interaction was larger in English ($\chi2 = 62.54$; $p < .001$) than in Italian ($\chi2 = 16.53$; $p < .001$).

Figure 2

Italian and English Error Rates per nonword type



Note. From left to right. NonStem+NonSuffix (e.g., nisht-lude), NonStem+Suffix (e.g., nisht-ness), Stem+NonSuffix (e.g., night-lude), Stem+Suffix (e.g., night-ness)

Discussion

The current study sought to examine cross-linguistic differences in morphological processing with a particular focus on the role of orthographic depth and morphological complexity. To address this aim, we recruited 60 Italian and 60 English native speakers who participated in a lexical decision task under the assumption that Italian has a transparent orthography but rich morphology, and English a opaque orthography but poor morphology. Participants responded to four types of nonwords, including Stem +Suffix (*night-ness*), NonStem + NonSuffix (*nisht-lude*), or Stem +NonSuffix and NonStem + Suffix items (*night-lude, nisht-ness*). In both reaction times and accuracy analyses, our findings revealed significant differences in morphological processing across languages, which are summarised below.

Our key finding is that, although stem and suffix effects were present in both languages, they were nearly twice as strong in English than in Italian, suggesting that readers of orthographically opaque languages engage in greater morphological processing than readers of orthographically transparent languages, regardless of morphological complexity. This finding, in concordance with both Vannest et al. (2002) and Mousikou et al.'s (2020) data, is of theoretical and practical importance, and lends credit to the suggestion that readers of opaque orthographies use morphemes as islands of regularity (Bowers & Bowers, 2018; Haddad et al., 2017; Mousikou et al., 2020; Plaut & Gonnerman, 2000). Crucially, the current study goes beyond earlier findings, not only by showcasing the importance of cross-linguistic differences in orthographic depth during complex word recognition, but also by directly comparing two languages that clearly differed on the orthographic depth and morphological complexity spectrum, while being matched in their morphological typology. One of the main criticisms regarding Vannest et al.'s (2002) earlier work was that the experimental items were not equally representative in both languages under examination (Stevens & Plaut, 2022; Vannest et al., 2002, p. 104). The authors employed bi-morphemic items for both languages, regardless of the highly synthetic nature of Finnish morphology. Additionally, the suffixes used in their English stimuli were both free and bound morphemes (-*able*, -*hood*, -*ship* and -*less* can occur both as stand-alone words and as suffixes), unlike those used for the Finnish items, and as such they had a more compound-type status. Our research addresses these typological and representativeness concerns by directly comparing two fusional languages (Italian: Ataman et al., 2019; Iacobini, 2006a; English: Choudhary et al., 2018; Mroczkowski et al., 2021; Zhang et al., 2020) with a more comparable incidence of bi-morphemic words through the employment of complex non-words with very limited semantic interpretability (a factor that affects lexical decisions, see for example Burani et al., 1999).

The current data support the idea that, in silent reading, orthographic depth has a bigger impact on morphological processing than morphological complexity. This is consistent with the Psycholinguistic Grain Size Theory, which contends that in languages where small orthographic units (such as graphemes) do not consistently map to speech sounds, readers will progress towards direct mapping of form to meaning through larger units, such as morphemes, that can be considered psychologically salient (Ziegler & Goswami, 2005). Our results lend credence to the hypothesis that readers of opaque orthographies rely on morphemes as a source of pronunciation consistency, even when speaking is not required by the task, thereby transferring skills learned to read aloud to silent reading. Several prominent theories of morphological processing exist that consider morphemes as primary units of word recognition and access to the orthographic lexicon (*Frequency Ordered Bin Search*: Taft, 2013, *Morphological Pathway Framework*: Levesque et al., 2021, *Word and Affix Model*: Beyersmann & Grainger, 2023; Grainger & Beyersmann, 2017).

One key example for a recent theoretical framework of morphological processing is the word and affix model, which builds on the idea that when readers encounter an orthographic input, they engage in three distinct processes, including embedded word activation, affix activation, and morpho-orthographic full decomposition. The first mechanism, embedded word activation, matches the orthographic input to the orthographic lexicon. As the activation of embedded words is an entirely non-morphological process, which also applies to morphologically simple words (e.g., *cash* in *cashew*), it explains why the here observed stem effects were evident even in nonwords that did not have a fully decomposable morphological structure (e.g., *night-lude*). The second mechanism, affix activation, matches the letter string onto the lexicon's pre-existing morpho-orthographic form representations. Provided that they are in the proper position (i.e., prefixed in string-initial and suffixes in string-final position), this mechanism activates affix representations (such as *-ness*), regardless of whether they are attached to stems (e.g., *night-ness*) or non-stems (e.g., *nisht-ness*), thus explaining the presence of a suffix effect in both the stem and non-stem conditions. Lastly, the word and affix model predict increased interference in the stem + suffix condition, where the combined activation of the stem and the suffix (e.g., *night + ness*) leads to greater interference as opposed to items where only a single morpheme is present (i.e., in the stem + non-suffix and non-stem + suffix conditions), thereby providing an explanation for the here reported stem by suffix interaction.

Limitations and Future Directions

An interesting extension of the current study would be the examination of cross-linguistic differences in morphological processing throughout children's reading development. While Mousikou et al. (2020) found that orthographic depth modulated morphological processing both in adults and children performing a naming task, research directly comparing adults and children in silent reading has not been conducted yet. Previous research comparing English-speaking and French-speaking children in Grade 4 pointed at stronger morphological effects in the morphologically richer language with the more transparent orthography (Casalis et al., 2015), which is in contrast with our findings with adult readers. Therefore, it would be fundamental to verify whether, in silent reading, the relative impact of morphological complexity and orthographic depth in morphological processing shifts during reading development by directly comparing skilled and beginning readers. As Mousikou et al. (2020) did not find such evidence in a reading aloud study when comparing adults and children data, that would point to a significant difference between reading aloud and silent reading.

Conclusions

The aim of this cross-linguistic study was to test the influence of orthographic depth and morphological complexity on visual word recognition by comparing two languages that clearly differ in these two key dimensions: English and Italian. Our results suggest that readers of orthographically opaque languages process morphological structure to a greater extent than readers of orthographically transparent languages, thus providing further support to the orthographic depth hypothesis and those models, such as the Word and Affix Model, that use morpheme units as proxies to lexicon access.

Declaration of Conflicting Interests

The authors declare that there is no conflict of interest.

Data availability statement

The data that support the findings of this study are openly available in OSF at https://osf.io/4g2vs/?view_only=86e666b9fa9445ed987c5dfef41fa9b5

References

Aikhenvald, A. Y. (2007). Typological distinctions in word-formation. In T. Shopen (Ed.), *Language Typology and Syntactic Description: Volume 3. Grammatical Categories and the Lexicon* (2nd ed., pp. 1–64). Cambridge University Press.

Ataman, D., Firat, O., Di Gangi, M. A., Federico, M., & Birch, A. (2019). On the Importance of Word Boundaries in Character-level Neural Machine Translation. In A. Birch, A. Finch, H. Hayashi, I. Konstas, T. Luong, G. Neubig, Y. Oda, & K. Sudoh (Eds.), *Proceedings of the 3rd Workshop on Neural Generation and Translation* (pp. 187–193). Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-5619

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

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

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

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3). https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Ben-Shachar, M., Lüdecke, D., & Makowski, D. (2020). effectsize: Estimation of Effect Size Indices and Standardized Parameters. *Journal of Open Source Software*, *5*(56), 2815. https://doi.org/10.21105/joss.02815

Beyersmann, E., & Grainger, J. (2023). The role of embedded words and morphemes in reading. In D. Crepaldi (Ed.), *Linguistic Morphology in the Mind and Brain.* Routledge.

Beyersmann, E., Mousikou, P., Javourey-Drevet, L., Schroeder, S., Ziegler, J. C., & Grainger, J. (2020). Morphological Processing across Modalities and Languages. *Scientific Studies of Reading*, *24*(6), 500–519. https://doi.org/10.1080/10888438.2020.1730847

Bölücü, N., & Can, B. (2019). Unsupervised Joint PoS Tagging and Stemming for Agglutinative Languages. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *18*(3), 1–21. https://doi.org/10.1145/3292398

Borgwaldt, S. R., Hellwig, F. M., & Groot, A. M. B. de (2005). Onset Entropy Matters – Letter-to-phoneme Mappings in Seven Languages. *Reading and Writing*, *18*(3), 211–229. https://doi.org/10.1007/s11145-005-3001-9

Borleffs, E., Maassen, B. A. M., Lyytinen, H., & Zwarts, F. (2017). Measuring orthographic transparency and morphological-syllabic complexity in alphabetic orthographies: A narrative review. *Reading and Writing*, *30*(8), 1617–1638. https://doi.org/10.1007/s11145-017-9741-5

Bowers, J. S., & Bowers, P. N. (2018). Progress in Reading Instruction Requires a Better Understanding of the English Spelling System. *Current Directions in Psychological Science*, *27*(6), 407–412. https://doi.org/10.1177/0963721418773749

Box, G., & Cox, D. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society*, *26*, 211–252.

Brown, D. (2010). Morphological Typology. In Jae Jung Song (Ed.), *The Oxford Handbook of Linguistic Typology* (pp. 487–503). Oxford University Press.

Burani, C., Dovetto, F. M., Spuntarelli, A., & Thornton, A. M. (1999). Morpholexical access and naming: The semantic interpretability of new root-suffix combinations. *Brain and Language*, *68*(1-2), 333–339. https://doi.org/10.1006/brln.1999.2073

Carlisle, J. F., McBride-Chang, C., Nagy, W., & Nunes, T. (2010). Effects of Instruction in Morphological Awareness on Literacy Achievement: An Integrative Review. *Reading Research Quarterly*, *45*(4), 464–487. https://doi.org/10.1598/RRQ.45.4.5

Casalis, S., Quémart, P., & Duncan, L. G. (2015). How language affects children's use of derivational morphology in visual word and pseudoword processing: Evidence from a cross-language study. *Frontiers in Psychology*, *6*, 452. https://doi.org/10.3389/fpsyg.2015.00452

Choudhary, H., Pathak, A. K., Saha, R. R., & Kumaraguru, P. (2018). Neural Machine Translation for English-Tamil. In O. Bojar, R. Chatterjee, C. Federmann, M. Fishel, Y. Graham, B. Haddow, M. Huck, A. J. Yepes, P. Koehn, C. Monz, M. Negri, A. Névéol, M. Neves, M. Post, L. Specia, M. Turchi, & K. Verspoor (Eds.), *Proceedings of the Third Conference on Machine Translation: Shared Task Papers* (pp. 770–775). Association for Computational Linguistics. https://doi.org/10.18653/v1/W18-6459

Cornoldi, C., & Candela, M. (2014). *Prove di lettura e scrittura MT-16-19: Batteria per la verifica degli apprendimenti e diagnosi di dislessia e disortografia*. Erickson.

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

De Simone, E., Beyersmann, E., Mulatti, C., Mirault, J., & Schmalz, X. (2021). Order among chaos: Cross-linguistic differences and developmental trajectories in pseudoword reading aloud using pronunciation Entropy. *PloS One*, *16*(5), e0251629. https://doi.org/10.1371/journal.pone.0251629

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3. ed.). Sage.

Frost, R., Grainger, J., & Rastle, K. (2005). Current issues in morphological processing: An introduction. *Language and Cognitive Processes*, *20*(1-2), 1–5. https://doi.org/10.1080/01690960444000287

Görgen, R., De Simone, E., Schulte-Körne, G., & Moll, K. (2021). Predictors of reading and spelling skills in German: the role of morphological awareness. *Journal of Research in Reading*, *44*(1), 210–227. https://doi.org/10.1111/1467-9817.12343

Grainger, J., & Beyersmann, E. (2017). Edge-Aligned Embedded Word Activation Initiates Morpho-orthographic Segmentation. In *Psychology of Learning and Motivation* (Vol. 67, pp. 285–317). Elsevier. https://doi.org/10.1016/BS.PLM.2017.03.009

Green, P., & MacLeod, C. J. (2016). SIMR : an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Greenberg, J. H. (1954). A Quantitative Approach to the Morphological Typology of Language. In R. F. Spencer (Ed.), *Method and Perspective in Anthropology: Papers in Honor of Wilson D. Wallis* (pp. 192–220). University of Minnesota Press.

Haddad, L., Weiss, Y., Katzir, T., & Bitan, T. (2017). Orthographic Transparency Enhances Morphological Segmentation in Children Reading Hebrew Words. *Frontiers in Psychology*, *8*, 2369. https://doi.org/10.3389/fpsyg.2017.02369

Iacobini, C. (2006a). Italian as a Fusional Language. In *Encyclopedia of Language & Linguistics* (pp. 64–69). Elsevier. https://doi.org/10.1016/B0-08-044854-2/04237-1

Iacobini, C. (2006b). Morphological Typology. In *Encyclopedia of Language & Linguistics* (pp. 278–282). Elsevier. https://doi.org/10.1016/B0-08-044854-2/00155-3

Juola, P. (2008). Assessing linguistic complexity. In M. Miestamo, K. Sinnemäki, & F. Karlsson (Eds.), *Studies in Language Companion Series. Language Complexity* (Vol. 94, pp. 89–108). John Benjamins Publishing Company. https://doi.org/10.1075/slcs.94.07juo

Katz, L., & Feldman, L. B. (1983). Relation between pronunciation and recognition of printed words in deep and shallow orthographies. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *9*(1), 157–166. https://doi.org/10.1037/0278-7393.9.1.157

Katz, L., & Frost, R. (1992). Chapter 4 The Reading Process is Different for Different Orthographies: The Orthographic Depth Hypothesis. In *Advances in Psychology. Orthography, Phonology, Morphology, and Meaning* (Vol. 94, pp. 67–84). Elsevier. https://doi.org/10.1016/S0166-4115(08)62789-2

Kettunen, K. (2014). Can Type-Token Ratio be Used to Show Morphological Complexity of Languages? *Journal of Quantitative Linguistics*, *21*(3), 223–245. https://doi.org/10.1080/09296174.2014.911506

Levesque, K. C., Breadmore, H. L., & Deacon, S. H. (2021). How morphology impacts reading and spelling: advancing the role of morphology in models of literacy development. *Journal of Research in Reading*, *44*(1), 10–26. https://doi.org/10.1111/1467-9817.12313

Marinelli, C. V., Romani, C., Burani, C., McGowan, V. A., & Zoccolotti, P. (2016). Costs and Benefits of Orthographic Inconsistency in Reading: Evidence from a Cross-Linguistic Comparison. *PloS One*, *11*(6), e0157457. https://doi.org/10.1371/journal.pone.0157457

Marinelli, C. V., Romani, C., McGowan, V. A., Giustizieri, S., & Zoccolotti, P. (2023). Characterization of reading errors in languages with different orthographic regularity: an Italian–English comparison. *Journal of Cultural Cognitive Science.* Advance online publication. https://doi.org/10.1007/s41809-023-00119-1

Marinus, E., & de Jong, P. F. (2008). The Use of Sublexical Clusters in Normal and Dyslexic Readers. *Scientific Studies of Reading*, *12*(3), 253–280. https://doi.org/10.1080/10888430802132246

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Miller, R. T. (2018). English orthography and reading. In J. I. Liontas & M. Delli Carpini (Eds.), *The TESOL Encyclopedia of English Language Teaching* (1(7)). Wiley.

Mousikou, P., Beyersmann, E., Ktori, M., Javourey-Drevet, L., Crepaldi, D., Ziegler, J. C., Grainger, J., & Schroeder, S. (2020). Orthographic consistency influences morphological processing in reading aloud: Evidence from a cross-linguistic study. *Developmental Science*, *23*(6), e12952. https://doi.org/10.1111/desc.12952

Mroczkowski, R., Rybak, P., Wróblewska, A., & Gawlik, I. (2021). *HerBERT: Efficiently Pretrained Transformer-based Language Model for Polish.* https://doi.org/10.48550/arXiv.2105.01735

Pagliuca, G., & Monaghan, P. (2010). Discovering large grain sizes in a transparent orthography: Insights from a connectionist model of Italian word naming. *European Journal of Cognitive Psychology*, *22*(5), 813–835. https://doi.org/10.1080/09541440903172158

Pirkola, A. (2001). Morphological typology of languages for IR. *Journal of Documentation*, *57*(3), 330–348. https://doi.org/10.1108/EUM0000000007085

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

Provazza, S., Carretti, B., Giofrè, D., Adams, A.-M., Montesano, L., & Roberts, D. (2022). Shallow or deep? The impact of orthographic depth on visual processing impairments in developmental dyslexia. *Annals of Dyslexia*, *72*(1), 171–196. https://doi.org/10.1007/s11881-021-00249-7

R Core Team. (2021). *R: A Language and Environment for Statistical Computing* [Computer software]. R Foundation for Statistical Computing. Vienna. https://www.R-project.org/

Rau, A. K., Moll, K., Moeller, K., Huber, S., Snowling, M. J., & Landerl, K. (2016). Same Same, but Different: Word and Sentence Reading in German and English. *Scientific Studies of Reading*, *20*(3), 203–219. https://doi.org/10.1080/10888438.2015.1136913

Rau, A. K., Moll, K., Snowling, M. J., & Landerl, K. (2015). Effects of orthographic consistency on eye movement behavior: German and English children and adults process the same words differently. *Journal of Experimental Child Psychology*, *130*, 92–105. https://doi.org/10.1016/j.jecp.2014.09.012

Sapir, E. (1921). *Language: An Introduction to the Study of Speech*. Harcourt, Brace & World.

Schlegel, F. (1808). *Ueber die Sprache und Weisheit der Ein Beitrag zur Begru¨ndung der Altertumskunde*. Mohr and Zimmer.

Schmalz, X., Beyersmann, E., Cavalli, E., & Marinus, E. (2016). Unpredictability and complexity of print-to-speech correspondences increase reliance on lexical processes: more evidence for the orthographic depth hypothesis. *Journal of Cognitive Psychology*, *28*(6), 658–672. https://doi.org/10.1080/20445911.2016.1182172

Schroeder, S., Häikiö, T., Pagán, A., Dickins, J. H., Hyönä, J., & Liversedge, S. P. (2022). Eye movements of children and adults reading in three different orthographies. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *48*(10), 1518–1541. https://doi.org/10.1037/xlm0001099

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology (London, England : 1953)*, *94*(Pt 2), 143–174. https://doi.org/10.1348/000712603321661859

Sona Systems. *Sona Systems: Cloud-based Participant Management Software [Computer software]* [Computer software]. Sona Systems, Ltd. https://www.sona-systems.com/

Stevens, P., & Plaut, D. C. (2022). From decomposition to distributed theories of morphological processing in reading. *Psychonomic Bulletin & Review*, *29*(5), 1673–1702. https://doi.org/10.3758/s13423-022-02086-0

Stump, G. T. (2001). Inflection. In A. Spencer & A. Zwicky (Eds.), *The handbook ofmorphology* (pp. 13–43). Wiley.

Taft, M. (2013). *Reading and the Mental Lexicon*. Psychology Press. https://doi.org/10.4324/9780203765357

Torgesen, J. K., Rashotte, C. A., & Wagner, R. K. (2012). *TOWRE 2: Test of Word Reading Efficiency* (2nd ed.). Pro-Ed.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology. General*, *124*(2), 107–136. https://doi.org/10.1037/0096-3445.124.2.107

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology (2006)*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Vannest, J., Bertram, R., Järvikivi, J., & Niemi, J. (2002). Counterintuitive cross-linguistic differences: More morphological computation in English than in Finnish. *Journal of Psycholinguistic Research*, *31*(2), 83–106. https://doi.org/10.1023/A:1014934915952

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*(5), 971–979. https://doi.org/10.3758/PBR.15.5.971

Zhang, F., Wang, Y., Zhang, X., Liu, C., Saraf, Y., & Zweig, G. (2020). Faster, Simpler and More Accurate Hybrid ASR Systems Using Wordpieces. In *Interspeech 2020* (pp. 976–980). ISCA. https://doi.org/10.21437/interspeech.2020-1995

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3

Acknowledgements

It is customary to compare a doctorate to a journey. In my case, the journey was not only metaphorical, but also physical. During my journey, not only was I given the opportunity to live in two countries, Germany and Australia, but also to visit the wonders of Europe and Oceania through training schools and conferences. Living an adventurous life and getting to know different cultures is what, in my view, makes a life worth living. For all the adventures and the happiness that I experienced in these four years, I have to thank first and foremost Xenia, who believed in me during my master degree studies and invited me to join her in Germany as her PhD student. Xenia taught me all the research fundamentals when I knew nothing, and put me on a pathway of becoming a rigorous researcher by imbuing me with statistical knowledge and good research practices. On my end, I am guilty of exhausting her good will, but still, that helped me significantly to grow as a person. My life at LMU was tremendously fun thanks to my colleagues and peers, especially Caro, Vroni, Björn and Lior, who assisted me all the way and for whose affection and friendship I could not be more grateful. A special mention goes to Kristina, my supervisor, who gracefully took me in along the way as one of her students, and keeps on giving me valuable insights and feedback on my work, without which the quality of my research would have been considerably lower.

I was also extremely fortunate to meet amazing people outside Academia: Paolo, whom I met as a Dungeons & Dragons master, became the best friend I always dreamed of, and in the company of Emanuel, Andrea, Matteo and Carolina, he gifted me with plenty of great DnD and board game sessions, which were always paired with good food, coffee and wine, something that should never be lacking during a PhD. I also need to mention my other Italian friends from the book club, with whom I toured Munich restaurants while having profound literary conversations: Andrea, Emmanuela, Giovanna, Enzo, Anna, Chiara, Ida and Lara. Together, you made Germany home, and I cannot wait to celebrate the defence of my thesis with all of us reunited.

Talking about home, I was blessed with parents who believed in my university studies. While deciding to pursue a doctorate so far from Italy was not always a resolution well received, I could only undertake this venture because my parents suffered through years of financial and emotional sacrifice to grant me (and my brother!) the possibility to attend the best universities in my field. One of these universities was the Sorbonne-Nouvelle in Paris, where I met Roland, my academic tutor and now dearest friend, who took his job so seriously in 2015 that in 2023 he is still around filling my days with banter, encouragement and aid. There were times when I thought that I could not do this PhD without his support. Gladly, those times are long gone, but only because he helped me build those skills that eventually made me an independent and confident student. I feel so much affection for Roland, and I can only hope that he will always be willing to be bothered.

Far from home was my life in Australia. When the border restrictions were lifted, I could not believe I was going to go for real: it was 2022 when I took my flight, although in our original plans I was scheduled to start at Macquarie University in 2020. Before I arrived in Australia, I was convinced that my life there was going to be riddled with work. For a variety of reasons, up until then my progress with my PhD had not been ideal, therefore, I left Europe with the idea that I was going to focus only on my studies and keep out all distractions. Luckily, things turned out to be quite different, and now I cherish my years in Sydney as the best of my life so far. As always, there were a few people responsible for this fortuitous change, the first of which is my supervisor, Lisi. I have no accurate words to describe how fundamental and impactful my collaboration with her has been on my professional and personal growth. Lisi guided me with kindness and encouragement when the winds of my PhD journey were making me shiver, and in line with the country we were in, showed me how sunny and warm academic life can be. She also gave me the space to thrive and pursue my own interests inside and outside Academia, with an enthusiastic and unconditional support that I have rarely experienced in my life. Lisi is one of

the most wholesome and radiant people I have ever met, and together with her (many) other students, I feel so extremely lucky that I ended up spending some time in her company.

Speaking of company, how could I not mention Esra. Esra was the first friend I made in Australia, and together we went on many bushwalks and toured many ramen restaurants. Sometimes we shared our sorrows about being so far from our loved ones, but our friendship made it all the more tolerable, and allowed us to enjoy the many wonders that Sydney has to offer. I am so glad to have met Ozan as well, thanks to whom I could finally find a worthy opponent in voluminous eating and office pranks, although he is not the only one to thank for my amazing office life at MQU. Amin, Xindong, Zixia, Katherine, Silvana, Hasibe, Alicia, Alessia, Demian, Arianna, Pauline, Ali and Yan: they are all deserving of their own specific thanks. I cannot believe how much support they showed me, and how fun they made the office life. I am incredibly fond of them all, and if there is anything that makes my submission feel bittersweet, it would be not getting to see them as much.

Stars, Divas and VIPs are always fashionably late, and Rara being all of them at once, I reserved the last thanks for them. You came late into my life, but believe me when I say that you are the person I longed for the most. Since I met you, I have felt constantly pulled towards you as if you had your own gravitational force, so much so that in fact I feared about being able to concentrate on the last tasks I had left to complete this thesis. Luckily, your wisdom supervised my attention, and your care handled my needs in this demanding time. At the same time, your presence constantly released me from the pressure that is usually associated with writing a PhD dissertation, and you made these days so entertaining in fact, that I did not feel like I was writing a thesis at all. I cannot fathom how random and yet perfect our coming together has been, and I am still in disbelief at how much happiness we have found in each other. You have brought so much joy to my life and to these last months of my PhD journey, something so rare that I could never stop expressing my gratitude, which I am sure you will always remind me of doing.