

Analysis of reading errors in Portuguese: Digraphs and complex syllabic structures

Edlia Alves Simões* / Margarida Alves Martins**

* University of Saint Joseph, Macau, China; ** Centro de Investigação em Educação, ISPA – Instituto Universitário, Lisboa, Portugal

The Portuguese language poses several challenges for children in the initial phase of learning how to read, particularly in the case of letters that may correspond to more than one phoneme, two letters that correspond to a single phoneme and in the case of words containing complex syllabic structures. The objective of this study was to perform a psycholinguistic analysis of the reading errors of children, attending the 1st ($n=175$) and 2nd year ($n=137$) of schooling, specifically in the case of words containing digraphs or complex syllabic structures and to analyse the differences between children's reading errors in these two years. An oral reading test was used for data collection. A quantitative and qualitative analysis of the type of reading errors was conducted using words with consonant digraphs (ch, nh, lh, gu, rr, ss), and words with complex syllables <CVC and CCV>. This analysis showed that children presented greater difficulties in some specific digraphs and tended to simplify complex syllables, either by adding or deleting phonemes. The quantity and quality of the reading errors of children attending both grades were discussed in light of reading acquisition theories and children's phonological development.

Key words: Reading errors, Primary school, Reading acquisition, Portuguese.

Introduction

The literature in reading acquisition indicates the importance of evaluating the oral reading, both for investigating reading skills and for diagnosing reading difficulties (Ehri & Snowling, 2004). This assessment allows not only to analyze the fluency and accuracy in reading, but also to characterize the reading errors in order to identify the strategies that underlie them (e.g., Goikoetxa, 2006; Salles, 2002). The decoding of words and consequent proficiency in the accuracy of word reading will contribute in large part to reading comprehension. Therefore, being able to read words and know the grapheme-phoneme correspondence rules is a skill that must be acquired since the early years of schooling, so that reading can become fluent and access to understanding possible (Hulme & Snowling, 2011).

Reading words in different orthographies does not have the same complexity. This depends on the grapheme-phoneme correspondence system of each one, that is, on the consistency of the language, and also depends on the respective syllabic structure (Seymour, 2005). In the context

This research was funded by Fundação para a Ciência e Tecnologia (Portugal), under the UID/CED/04853/2016 project.

Correspondence concerning this article should be addressed to: Edlia Alves Simões, University of Saint Joseph, Estrada Marginal da Ilha Verde, 14-17, Macau, China; E-mail: edlia.simoed@usj.edu.mo

of European languages, Finnish has the most consistent orthography, followed by Italian, Greek and Spanish. English, being opaque is at the opposite extreme. Among these extremes, there are languages considered to be semi-transparent or having intermediate spelling, such as Portuguese, French and Danish (Seymour, Aro, & Erskine, 2003).

Portuguese spelling may be considered as having highly predictable grapheme-phoneme correspondences (Borgwaldt, Hellwig, & De Groot, 2005; Serrano, Genard, Sucena, Defior, & Alegria, 2010). However, there is no direct match between phonemes and graphemes because a grapheme can represent more than a phoneme and a phoneme can be represented by more than one grapheme (Freitas, Rodrigues, Costa, & Castelo, 2012; Mateus, 2007) and can therefore be considered as an intermediate or semitransparent orthography (e.g., Lima & Castro, 2010; Seymour et al., 2003; Sucena, Castro, & Seymour, 2009).

This variability in the correspondence between phonemes and graphemes in different languages led to the formulation of the orthographic depth hypothesis (Frost, 2005; Katz & Frost, 1992) that predicts that readers will rely more on the lexical (global recognition) or phonological (recoding) pathways depending on the requirements of the orthography. In more transparent orthographies, readers rely more on phonological or non-lexical strategies, while in opaque orthographies readers are less dependent on phonological strategies and more dependent on the lexical or “orthographic” path. However, the phonological pathway is always used – it only varies the extent of its resource in reading.

This effect of orthographic depth was verified by Seymour et al. (2003), which also demonstrated that syllabic complexity has an effect on the decoding process. They carried out a cross-linguistic study in which the acquisition of reading in 13 European languages was analyzed through reading of letters, familiar words and non-words in children of the 1st and 2nd year of schooling. Although the methods of teaching were not equal, teaching had always taken a phonics approach. This study was based on the Seymour’s dual foundation model, in which it is hypothesized that if linguistic complexity affects the first stage of acquisition of reading (the foundation phase of acquisition), then reading will be more difficult to learn in opaque languages and languages with more complex syllabic structure.

With regard to the reading of familiar words, Portuguese, French and Danish, as well as English, were identified as being read with less accuracy and fluency than other languages as Spanish. Portuguese achieved 73% accuracy, compared to French with 79% or Spanish with 95%. This difference in relation to other European orthographies is due, according to this study, to the orthographic depth and not to syllabic complexity.

In nonword reading, there was a poor lexical effect (i.e., read nonwords as words) in the orthographies of less complex syllabic structures in relation to those of more complex structures. Portuguese and French were the orthographies with the greatest difficulty in the decoding of nonwords, among orthographies with less syllabic complexity. Portuguese reached 77% accuracy while French obtained 85% and Spanish 89% accuracy. Both in the reading of words and in the reading of nonwords Portuguese had less accuracy in relation to other languages with a simple syllabic structure. Seymour et al. (2003) have argued that there is a threshold of orthographic complexity from which, instead of using only a base process for reading acquisition (transparent orthographies), it is necessary to use two base processes with distinct word decoding and recognition of words. Portuguese has been described as above this limit, that is, the acquisition of reading would be processed similar to more opaque orthographies.

In this way, being a complex process, the acquisition of written language depends on several factors, such as phonology and orthography. Knowing in depth how the sounds of the language relate to the spelling of the language can help to account for the cognitive processes associated with learning to read and for the difficulties that arise.

According to Mateus (2007), “phonology, studies the sounds that have a function in the language and that allow the speakers to distinguish meanings” (p. 3). In speech, a word can have a variety of phonetic forms, whereas in writing, it can only have a graphic form, which is taught in school and which is socially accepted: its spelling.

The phonological system of the mother tongue, which corresponds to an exercise in abstraction elaborated from the phonetic reality, is progressively constructed. The elements of the phonological system are called phonemes, distinguishing themselves in three large groups: vowels, consonants, and semivowels.

The Portuguese alphabet has 26 letters (according to the new orthographic agreement) and the Portuguese language consists of 35 phonemes (“abstract entities of the grammatical system of the language” p. 41, Freitas et al., 2012) for 67 graphemes (Gomes, 2001): 19 consonant sounds; 14 vowel sounds and 2 glides.

The Portuguese language is formed mostly by biunivocal relations between phonemes and graphemes and when there is no such relationship there are contextual and positional rules that facilitate reading. However, vowels are particularly challenging in Portuguese because there are only five vowels and 14 vowel phonemes. The consonants also present difficulties in reading and writing due to a consonant being able to correspond to different sounds. In addition, there are sounds that can be represented by several consonants. For example, the sound [s], can be represented by <s>; <ss>; <ç> and <c> when this letter is followed by the vowels <i> or <e>. On the other hand, the grapheme <s> can also have the value of [ʃ] or [z], depending on the consonant that follows it, for example in <disco>/[d'isʃku] and <musgo>/[m'uzgu] (Freitas et al., 2012).

There may also be vowel encounters, that is, in what are known as diphthongs – the encounter of a semivowel and vowel or vowel and semivowel, and also digraphs that serve to represent the nasal vowels: <am>, <an>, , <im>, <in>, <om>, <on>, <um>, <un> as for example in <tampo>/[t'ẽpu]; <lindo>/[l'ĩdu]; <pombo> [p'õbu] e <mundo>/[m'ũdu]. On the other hand, <m> only marks the nasality of the vowel when it precedes a <p> or (Cunha & Sintra, 2013).

As with vowels, we may also have cases where there is a grouping of two consonants but forming a single phoneme, such as <nh>; <ch> and <lh>, which are called consonant digraphs. Even in oral language, the digraphs present a certain order of acquisition, and the digraph <nh>, nasal consonant [ɲ] is the first for children to acquire, followed by the <ch>, fricative consonant [ʃ] and the <lh> lateral consonant [ʎ] (Freitas et al., 2012).

Still in the group of consonant digraphs we include the double consonants: <rr> and <ss>. In the case of double consonants they serve to maintain the sound [R] and [s] in the middle of vowels in the word. These graphemes can never appear at the beginning or end of a word. Among the digraphs we can also consider the digraphs <gu> and <qu> which are read as [g] and [k] when followed by the vowels <e> and <i>, making the letter <u> in a mute phoneme, as for example, in <guerra>/[g'ẽrẽ]; <aquilo>/[ẽk'ĩlu].

There are also consonant encounters, which refer to the grouping of consonants inside a word. It is worth highlighting the consonants <l> and <r> which are often grouped following the consonants at the beginning or middle of the words, as is the case, for example, of <bl> in <blusa>/[bl'uzẽ]; <gl> in <glutão>/[glut'ẽw]; <pl> in <triplo> [tr'iplu]; <dr> in <vidro>/[v'idru] and <vr> in <palavra>/[pẽl' avrẽ]. These consonant groupings form the CCV (consonant / consonant / vowel) syllables.

Portuguese orthography has a domain of CV syllables, and the CV syllabic structure corresponds to 46,36% in Portuguese words. The syllable CVC (11,01%) and CCV (2,18%) occurs less frequently in Portuguese, and its frequency is predominantly in the initial position of the word (Vigário, Martins, & Frota, 2006). These syllables are particularly difficult because they are later acquired phonological structures (Afonso & Freitas, 2010; Freitas, Frota, Vigário, & Martins, 2005).

Studies in a number of languages have shown that the syllabic structure, as well as the graphemic complexity affect the acquisition of reading (Barca, Ellis, & Burani, 2007; Nunes & Byrant, 2014; Sprenger-Charolles & Siegel, 1997).

There are studies that focus on an analysis of reading errors in certain grapho-phonological correspondences, such as digraphs or complex syllabic structures (Ehri & Soffer, 1999; Treiman, 1991). However, there is a scarcity of research in the Portuguese language. We will now refer to some studies.

Nunes and Aldinis (1999) found that children in their first years of schooling made more errors in words that contained digraphs as opposed to those that did not, just as weaker readers had greater difficulties with digraphs. They hypothesized that these readers do not understand the need for the consonantal digraph to represent a sound due to phonological difficulties (e.g., difference between <n>/[n] and <nh>/[ɲ]), or because of the difficulty of using graphemes involving more than one grapheme (Nunes & Byrant, 2014). It also showed that reading words that contain digraphs is done more slowly, particularly if they are the first syllable (Capovilla, Capovilla, & Macedo, 2001).

Rego and Buarque (1999) found that children in the first years of schooling choose to substitute the <rr> for <r>, perhaps based on the alphabetical principle that each sound corresponds to a letter.

As for syllabic structure, although most words have a simple structure, there are some with complex syllabic structures in which children make more errors, as found in Simões and Alves Martins (2015, 2018) and Gomes (2001). The latter author verified that most of the errors made by children from 1st to 3rd year were in words with complex syllables. These syllables often cause hesitant reading and writing errors, and children tend to simplify the structure of the syllable (Afonso & Freitas, 2010; Gomes, 2001; Gonçalves, Guerreiro, Freitas, & Sousa, 2011).

Monteiro (2007) and Monteiro and Soares (2014), in Portuguese, as well as Goikoetxea (2006), in Spanish, also verified that children having difficulties in reading tended to transform complex syllables into simple ones, that is, in the canonical syllable CV. According to Monteiro (2007) and Monteiro and Soares (2014), this mechanism causes children to return to the decoding strategy typical of the beginning of learning how to read. These results may be related to Portuguese orthography, since it has a domain of CV syllables.

Given the scarcity of studies in relation to difficulties in reading digraphs and complex syllables in Portuguese, this study aimed to analyze what are the reading errors that occur when children read these specific cases, as well as to find whether they are different in quantity and quality in the first two years of schooling. Thus, we stated the following research questions:

- 1) Will there be differences between grade 1 and grade 2 in the reading of digraphs? What are the main reading errors that occur in each grade?
- 2) Will there be differences between grade 1 and grade 2 in the reading of complex syllables? What are the main reading errors that occur in each grade?

Method

Research design

This study is part of a broader investigation in which the reading performance in the first two years of schooling and the most frequent errors were analyzed.

Participants

Participants were 175 children attending the 1st year of schooling and 137 children attending the 2nd year from 6 Primary schools (4 public schools and 2 private) from the District of Lisbon. In the first year there were 75 female participants and 100 male participants; in the second year there were 63 females and 74 males. The mean age of the 1st year children was 84.01 months ($SD=4.06$) and of the 2nd year was 95.63 months ($SD=4.51$).

With regard to the level of schooling of the parents the majority completed the equivalent of high school (fathers: 27%, mothers: 25,3%) or attended higher education (fathers: 32,1%, mothers: 36%). The public schools were in the medium / medium-low socioeconomic strata, while private schools were located in the medium-high / high socioeconomic strata.

Authorization to participate in this study was requested from the schools, as well as from the parents of all the children involved.

A questionnaire was given to teachers with regard to the reading teaching method they used. Most teachers claimed to use a mixed method, that is, a phonics (teaching of grapho-phonological correspondences) and global (working on words and short texts) approach.

In order to reduce the probability of the effects of differences between teachers and type of teaching, the students were selected from several classes, from each year of schooling in the six schools.

Children who had special educational needs or who had not Portuguese as their first language were excluded.

Tasks

An oral word reading test (Alves Martins & Simões, 2008) was given to grades 1 and 2. This test consists of a list of 32 words in random order. The words vary in regularity, frequency, length and syllabic structure. Of the 32 words, 16 are regular and 16 are irregular; 21 are infrequent and 11 are frequent. The number of letters of the different words varies between 4 and 9; 19 were classified as short words and 13 as long words.

In terms of the syllabic structure, words with different syllabic formats were used. Thus, there were syllables CV (e.g., as in <lavrador>[lɐvɾəd'or]), syllables V (e.g., <a> as in <arroz>[ɐɾ'oz]), syllables VC (e.g., <ir> as in <irmão>[iɾm'ɐw]), CVC (e.g., <dis> as in <disco>[d'isku]), CCV syllables (e.g., <cla> as in <clarão>[klɐr'ɐw]) and CCVC syllables (e.g., <cris> as in <cristal>[kɾist'at]).

With regard to the internal consistency of the test, the value of the Cronbach coefficient α was .97. The test can thus be considered as having a high reliability.

The test is of individual administration. The test is presented in paper form.

The following instructions were given:

“Read the words that follow on this list aloud, as best as you can and as fast as you can”.

How each word read was recorded, which allowed the assessment of reading accuracy and the analysis of error type. Each correct answer is scored 1 point, with the results ranging from 0 to 32 points.

Procedure

The oral word reading test was always administered, in a room near the classroom, free of distractions and with as little noise as possible in the vicinity. The test was conducted individually with the child and the researcher.

The reading was recorded in audio format, with the program *Audacity*, and the phonetic form of words read was then transcribed.

Data analysis

The words with complex syllables and digraphs were analyzed in order to assess the reading strategies used.

In this way we proceeded to the collection and analysis of the errors for the set of words in which there was a consonant digraph or a complex syllable. The grapho-phonological correspondences selected correspond to relations between graphemes and phonemes that are difficult to acquire, as in the case with digraphs. Six words were analyzed with consonant digraphs (*trincha*; *unhas*; *zarolho*; *águias*; *arroz*; *girassol*).

Complex syllables are characterized by three graphemes and can be positioned at the beginning, middle or end of words. We analyzed 7 words from each of the two syllabic structures CVC (*serpente*; *disco*; *crystal*; *pombal*; *oval*; *fritar*; *lavrador*) and CCV(*blusa*; *glutões*; *clarão*; *próximo*; *fritar*; *lavrador*; *quadros*).

In each word, only the errors in the syllables and digraphs under study were scored and analyzed.

Chi-square tests were performed to answer the research questions.

Results

We started by analyzing the consonant digraphs (<ch>, <nh>, <lh>, <gu>, <rr>, <ss>).

Table 1 shows the frequency and percentage of children who committed errors in reading these consonant digraphs by grade.

Table 1

Frequency and percentage of reading errors on digraphs in Grade 1 and 2

Digraphs	Grade 1		Grade 2	
<ch>	38	21,7%	7	5,1%
<lh>	17	9,7%	10	7,3%
<nh>	10	5,7%	8	5,8%
<ss>	9	5,1%	1	0,7%
<rr>	13	7,4%	1	0,7%
<gu>	17	9,7%	9	6,6%

We can see in Table 1 that in Grade 1 the digraph in which a greater number of errors occurs is <ch> followed, in descending order by <lh> and <gu>, <rr>, <nh> and <ss>. In Grade 2, the <rr> and <ss> digraphs have already been acquired by most children. The number of errors in the digraph <ch> decreases considerably in Grade 2, although there are still some children who continue to experience difficulties in reading it. The percentage of errors in reading this digraph is quite similar to that of the digraph <nh>, which is slightly higher. The digraphs <lh> and <gu> are those in which some problems still seem to persist, with a higher percentage of children failing at the <lh> digraph.

Chi-square tests were performed to analyse whether there were differences between grade 1 and grade 2 in reading words that contained the above digraphs. Statistically significant differences between the two grades in what concerns the digraphs <ch> ($\chi^2=17.17$; $p<.001$), <ss> ($\chi^2=4.82$ $p<.05$) and <rr> ($\chi^2=8.05$ $p<.005$) were found. There is a decrease in the frequency of errors in grade 2. No differences were found between the two grades for the digraphs <lh> ($\chi^2=.57$; $p=.451$), <nh> ($\chi^2=.002$, $p=.962$) and <gu> ($\chi^2=.10$; $p=.319$).

What are the main errors that occur in each grade in what concerns the reading of these digraphs?

In relation to the reading of the digraph <ch>/ [ʃ] in the word <trincha>[tr'ĩʃ], the most common error in grade 1 is the omission of the first consonant of this digraph, the word <trincha> being read as (<trinha>/ [tr'ijnɐ]) ($N=13$). This error is almost inexistent in grade 2 ($N=1$). It is also common in grade 1 to read only the initial consonant of the digraph which results on the word (<trinca>/ [tr'ikɐ]) ($N=11$). In grade 2 less children commit this error ($N=3$). In other cases, the strategy is also the incorrect reading of the initial consonant of the digraph <c> read as [s] which is a sound that may correspond to this consonant in other contexts (when followed by the vowels <e> or <i>) ($N=8$ in grade 1; $N=3$ in grade 2). In grade 1 there are also some other errors that disappear in grade 2 as the omission of the digraph ($N=3$) or its substitution by [ʒ] or by [z] may be because of the confusion of similar sounds from a phonetic point of view ($N=3$).

Regarding the errors that occur in the reading of the digraph <lh>/ [λ] in the word <zarolho> [zɐr'olɥ], in grade 1 the digraph <lh> is read as it was the digraph <nh> children reading the word as zaronho/[zɐr'ɔnu] ($N=9$), an error that persists in grade 2 ($N=8$). In grade 1, the word is also read as if the digraph did not exist, but only its initial consonant <l> the word being read as zarolo/[zɐr'olu] ($N=3$). There is also the substitution of the digraph by the digraph <ch>, the word being read as zarocho/[zɐr'ɔfu] ($N=2$). These two strategies are rare in grade 2 ($N=1$). Finally, some children in grade 1 ($N=3$) skip the reading of this digraph reading the word as zarioio/[zɐr'ɔju]. In grade 2 this error disappears.

In what concerns the reading of the digraph <nh> in the word <unha>['unɐ] some children in grade 1 ($N=2$) read this digraph as it was the digraph <lh>, the word being read as ulha/['u(ɥ)] ($N=2$). This error persists in grade 2, being the most frequent one ($N=7$). In grade 1 some children also replace this digraph by the digraph <ch>, the word being read as ucha/['uʃ]. In grade 2 this substitution is rare ($N=1$). As it was previously described regarding the digraphs <ch> and <lh>, some children only read the initial consonant of the digraph <n> ($N=3$). In other cases the digraph was also replaced by the phonemes [s] or [z] ($N=3$). These last two types of error no longer happen in grade 2.

As for the digraph <ss> in the word <girassol> [zĩrɐs'ɔɫ], several children in grade 1 ($N=7$) read it as [z], girazol/[zĩrɐz'ɔɫ]. These children read the digraph as if it was a single letter, in this case <s>, which is read as [z] when occurring between two vowels. In grade 2 this error is rare ($N=1$). Also in grade 1, some children replaced the digraph <ss> by the digraph <ch> reading girachol/[zĩrɐʃ'ɔɫ]. In grade 2, the reading of this digraph seems to be acquired by nearly all the children.

As for the digraph <rr>/[R] in the word <arroz>[ɐr'ɔʃ], the majority of children who failed to read this digraph in grade 1 (13) used the strategy previously described, reading only one of the letters which leads to its replacement by the phoneme [r] the word being read as [ɐr'ɔʃ]. In grade 2 the reading of this digraph seems to be acquired.

In relation to the reading of the digraph <gu>/[g] in the word <águas> ['agjɐ], several children in grade 1 read the two letters separately, reading águ-ias/['agu-ieʃ] ($N=9$ in grade 1; $N=2$ in grade 2), or read only the first letter of the digraph reading the word as águas/['aziɐ] ($N=6$). This last error also exists in grade 2 ($N=6$). These errors are due to the contravention of an orthographic rule in which the letter <g> followed by the vowels <e> or <i> is read [ʒ]; to read [g] the digraph [gu] must be used. There were also 2 children who replaced the phoneme [g] by the phoneme [k] that is close phonetically, whereas this does not occur in grade 2.

Are there differences between grade 1 and grade 2 in the reading of complex syllables? What are the main reading errors that occur in each grade?

Syllables CCV

We analyzed the groups of words with the complex syllables: CCV (<blu>, <glu>, <cla>, <pro>, <fri>, <vra> and <dro>) and CVC (<ser>, <dis>, <tal>, <bal>, <val>, <tar>, <dor>).

Table 2 shows the frequency and percentage of errors in the syllables CCV in grades 1 and 2.

Table 2

Frequency and percentage of reading errors on CCV syllables in Grade 1 and 2

CCV	Grade 1		Grade 2	
<blu>	43	24,6%	16	11,7%
<glu >	23	13,1%	11	8,0%
<cla>	21	12,0%	10	7,3%
<pro>	24	13,7%	8	5,8%
<fri>	18	10,3%	4	2,9%
<vra>	23	13,1%	6	4,4%
<dro>	15	8,6%	0	0,0%

As we can see in Table 2, children in grade 2 had better results than children in grade 1. Chi-square tests were performed to analyse whether there were differences between grade 1 and grade 2 in reading words that contained CCV syllables. There are statistically significant differences between the two grades in what concerns all syllables except two: <blu> $\chi^2=8.33$; $p=.004$; <pro> $\chi^2=5.18$; $p=.023$; <fri> $\chi^2=6.36$; $p=.012$); <vra> $\chi^2=7.00$; $p=.008$; <dro> $\chi^2=12.34$; $p<.001$. No differences were found for the syllables <glu> ($\chi^2=2.07$; $p=.150$) and <cla> ($\chi^2=1.90$; $p=.168$).

Children commit different types of errors when they read words that contain CCV syllables. In grade 1 they frequently add a vowel to the initial consonant (e.g., <blusa>[bl'uzɐ] read as [bil'uzɐ]; <fritar>[frit'ar] read as [firit'ar]). Generally, the added vowel is the phoneme [i]. In grade 2 this strategy is rarely used.

The second most common error in grade 1 is the inversion of the last two letters of the CCV syllable, transforming it into a CVC one (e.g., <blusa>[bl'uzɐ] read as *bulsã*, [ˈbulsɐ] or <clarão>[klar'ẽw] read as *calrãõ*/[kɛlr'ẽw]). In grade 2 this is the most common strategy. Another common error in both the 1st and 2nd grade is not reading the second consonant in the syllable (e.g., <lavrador> [lɛvrɛd'or] read as *lavador*/ [lɛvɛd'or]).

Syllables CVC

Table 3 shows the frequency and percentage of errors in the syllables CVC in grades 1 and 2.

Table 3

Frequency and percentage of reading errors on CVC syllables in Grade 1 and 2

CVC	Grade 1		Grade 2	
<ser>	22	12,6%	7	5,1%
<dis>	13	7,4%	1	0,7%
<tal>	15	8,6%	4	2,9%
<bal>	22	12,6%	5	3,6%
<val>	30	17,1%	5	3,6%
<tar>	25	14,3%	7	5,1%
<dor>	11	6,3%	4	2,9%

As we can see in Table 3, children in grade 2 had better results than children in grade 1. Chi-square tests were performed to analyse whether there were differences between grade 1 and grade 2 in reading words that contained CVC syllables. There are statistically significant differences between the two grades in what concerns all syllables except one: <ser> $\chi^2=5.08$; $p=.024$; <dis> $\chi^2=8.05$; $p=.005$; <tal> $\chi^2=4.29$; $p=.038$; <bal> $\chi^2=7.74$; $p=.008$; <val> $\chi^2=14.05$; $p<.001$; <tar> $\chi^2=7.03$; $p=.008$. No differences were found for the syllable <dor> ($\chi^2=1.90$; $p=.168$).

The most frequent error in grade 1 is the addition of a vowel following the last consonant of the CVC syllable (e.g., <crystal> [kriʃt' at] read as *cristala* [kriʃt' alɐ], or <pombal> read as *pombale* [põb' ali]). In grade 2, this error practically ceases to occur.

Two other common errors in grade 1, as well as in grade 2, are the deletion of the last consonant of the syllable (e.g., <serpentes> [sɨp' êtiʃ] read as *sepentes* [sɨp' êtiʃ], <fritar> read as *frita*).

Less frequent in both grade 1 and 2 are the inversions (e.g., <crystal> read as *cristla* [kr' iʃtlɐ]).

Discussion

The first research question that we have tried to answer in this research concerns the differences that may occur between grade 1 and grade 2 in the reading of digraphs. Our data shows that there were statistically significant differences between the two grades in the reading of the digraphs <ch>; <ss>; <rr>, the errors decreasing from first to second grade. Nevertheless no differences were found regarding the remaining digraphs <lh>, <nh> and <gu>. Concerning the main reading errors that occur in both grades, we can see that the substitution by another digraph in the cases of <ch> <lh> or <nh> is the main error in both grades. When reading the digraph <ch> and <lh> children tend to replace them by the digraph <nh> [ɲ]. This substitution may be explained by the fact that the digraph [ɲ] is the earliest acquired in oral language, therefore the most easiest for children to use (Freitas et al., 2012).

As for the remaining digraphs <ss> <rr> and <gu>, children often read the word as if the digraph did not exist, suppressing the second letter and reading only the first one.

The second research question concerning the differences between the two grades regarding the number of reading errors for complex syllables, the results show that there are statistically significant differences between the two grades, the frequency of errors in grade 2 being lower than in grade 1 for almost the CCV and the CVC syllables.

Regarding the types of errors that occur in the words with CCV syllables, there is a higher frequency of three types of error: addition of a vowel after the first consonant, inversion and suppression of the second consonant.

The addition of a vowel stands out as the main type of error in children in grade 1. This error may reveal a strategy of simplifying the syllabic structure of the word into a simpler syllabic structure CVCV, which is the most frequent in the Portuguese orthography.

With regard to the inversion error, the CCV syllable is transformed into a CVC one, which also stands out as very common in grade 1 and often prevails in grade 2. This may be explained by the fact that CVC syllables are acquired earlier than CCV ones from a phonological development perspective. According to Freitas et al. (2005) and Afonso and Freitas (2010), the ramified onset (CC) is the last syllabic structure to be acquired in Portuguese. Also the CVC syllable is more common in the Portuguese language than the syllable CCV (Vigário et al., 2006).

The suppression of the second consonant is also frequent in grade 2, as in the studies by Sprenger-Charolles and Siegel (1997) and Treiman (1991), in which the simplification of CCV branching onsets occurred with greater omission of liquid (<r>, <l>) than of obstructive consonants. Monteiro (2007) also observed this strategy in word reading with this type of syllables.

As for the most frequent reading errors in CVC syllables, children from grade 1 tend to transform these syllables into CVCV ones by adding a vowel or into CV syllables by suppressing the last consonant. This was also the case, in the studies by Sprenger-Charolles and Siegel (1997), in French, as well as Monteiro (2007), in Portuguese.

If phonological mediation is one of the most important skills in reading acquisition, the phonological characteristics of word structures will have influence on reading performance. One of these phonological features is the syllabic structure, in particular the most basic division of the syllable in onset and rhyme. When children face difficulties in reading complex syllables they tend to return to a decoding model characterized by CV syllables, which is characteristic of the onset of their literacy acquisition.

The syllable CV in Portuguese is the most common syllabic structure, so it would be expected that there would be superior performance in items with syllabic structure CV, compared to CCV syllables or CVC ones. Gomes (2001) found that Portuguese children from 1st to 3rd grades committed more errors in the more complex syllables. Goikoetxea (2006), Monteiro (2007) and Monteiro and Soares (2014) also found similar results in that when children face difficulties in reading complex syllables. In future research it would be interesting to explore this issue.

In conclusion, and although Portuguese orthography is considered to contain a simple syllabic structure (Seymour et al., 2003), CVC and CCV syllables exist in the Portuguese language and present obstacles in reading which are difficult for children at the beginning of their schooling.

We can also conclude that the consonant digraphs still present many challenges for children in the first and second year of schooling. This difficulty is related to the intermediate transparency of Portuguese spelling which, despite having many consistent grapho-phonological correspondences, presents some graphemes that do not have a biunivocal relation to their corresponding phonemes.

References

- Afonso, C., & Freitas, M (2010). Consciência fonológica e desenvolvimento fonológico: O caso do constituinte ataque em Português Europeu. In M. Freitas, I. Duarte, & A. Goncalves (Eds.), *Avaliação da consciência linguística: Aspetos fonológicos e sintáticos do Português* (pp. 45-68). Lisboa: Edições Colibri. Recuperado de http://www.clul.ul.pt/files/anagrama/Afonso_Freitas_2010.pdf
- Alves Martins, M., & Simões, E. (2008). Teste de reconhecimento de palavras para os dois primeiros anos de escolaridade. In A. P. Noronha et al. (Eds.), *Atas da XIII conferência internacional de avaliação psicológica: Formas e contextos*. Braga: Psiquilibrios Edições.
- Barca, L., Ellis, A., & Burani, C. (2007). Context-sensitive rules and word naming in Italian children. *Reading and Writing: An Interdisciplinary Journal*, 20, 495-509. Recuperado de <https://doi.org/10.1007/s11145-006-9040-z>
- Borgwaldt, S., Hellwig, F., & De Groot, A. (2005). Onset entropy matters: Letter-to-phoneme mappings in seven languages. *Reading and Writing*, 18, 211-229. Retrieved from <https://doi.org/10.1007/s11145-005-3001-9>
- Capovilla, F., Capovilla, A., & Macedo, E. (2001). Rota perilexical na leitura em voz alta: Tempo de reação, duração e segmentação na pronúncia. *Psicologia: Reflexão e Crítica*, 14, 409-427. Recuperado de <https://doi.org/10.1590/S0102-79722001000200015>
- Cunha, C., & Sintra, L. (2013). *Nova gramática do português contemporâneo* (21^a ed.). Lisboa: Editora Figueirinhas.

- Ehri, L. C., & Snowling M. J. (2004). Developmental variation in word recognition. In C. A. Stone, E. R. Silliman, B. J. Ehren, & K. Apel (Eds.), *Handbook of language and literacy: Development and disorders* (pp. 433-460). New York: Guilford.
- Ehri, L. C., & Soffer, A. G. (1999). Graphophonemic awareness: Development in elementary students. *Scientific Studies of Reading*, 3, 1-30.
- Freitas, M., Frota, S., Vigário, M., & Martins, F (2005). Efeitos prosódicos e efeitos de frequência no desenvolvimento silábico em Português Europeu. Seleção de Comunicações apresentadas no XXI Encontro Nacional da Associação Portuguesa de Linguística.
- Freitas, M., Rodrigues, C., Costa, T., & Castelo, A. (2012). *Os sons que estão dentro das palavras: Descrição e implicações para o ensino do português como língua materna*. Lisboa: Edições Colibri e Associação de Professores de Português.
- Frost, R. (2005). Orthographic systems and skilled word recognition processes in reading. In M. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 272-295). Oxford: Blackwell Publishing. Retrieved from <https://doi.org/10.1111/b.9781405114882.2005.x>
- Goikoetxea, E. (2006). Reading errors in first and second-grade readers of a shallow orthography: Evidence from Spanish. *British Journal of Educational Psychology*, 76, 333-350. Retrieved from <https://doi.org/10.1348/000709905x52490>
- Gomes, I. (2001). *Ler e escrever em português europeu*. Dissertação apresentada à Universidade do Porto.
- Gonçalves, F., Guerreiro, P., Freitas, M., & Sousa, O. (2011). *O conhecimento da língua: Percursos de desenvolvimento linguístico*. Lisboa: Ministério da Educação – DGIDC.
- Hulme, Ch., & Snowling, M. J. (2011). Children's reading comprehension difficulties: Nature, causes, and treatments. *Current Directions in Psychological Science* 20, 139-142.
- Katz, L., & Frost, R. (1992). Reading in different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (Eds), *Orthography, phonology, morphology, and meaning* (pp. 67-84). Amsterdam: North-Holland.
- Lima, C. F., & Castro, S. L. (2010). Reading strategies in orthographies of intermediate depth are flexible: Modulation of length effects in Portuguese. *European Journal of Cognitive Psychology*, 22, 190-215. Retrieved from <https://doi.org/10.1080/09541440902750145>
- Mateus, M. (2007). A contribuição do estudo dos sons para a aprendizagem da língua. *Atas do 7º Congresso da APP_Saber Ouvir / Saber Falar* (pp. 1-22). Coimbra.
- Monteiro, S. (2007). *O processo de aquisição da leitura no contexto escolar por alfabetizandos considerados portadores de dificuldades de aprendizagem*. Tese de doutoramento não publicada, Universidade Federal de Minas Gerais, Belo Horizonte.
- Monteiro, S., & Soares, M. (2014). Processos cognitivos na leitura inicial: Relação entre estratégias de reconhecimento de palavras e alfabetização. *Educação e Pesquisa*, 40, 449-465. Recuperado de <https://doi.org/10.1590/S1517-97022014005000006>
- Nunes, T., & AIdinis, A. (1999). A closer look at the spelling of children with reading problems. In T. Nunes (Ed.), *Learning to read: An integrated view from research and practice* (pp. 155-171). Dordrecht: Kluwer.
- Nunes, T., & Bryant, P. (2014). *Leitura e ortografia: Além dos primeiros passos*. Porto Alegre: Penso. [Tradução do original em língua inglesa *Children's reading and spelling: Beyond the first steps*. Wiley-Blackwell, 2009].
- Salles, J. (2002). Processos cognitivos na leitura de palavras em crianças: Relações com compreensão e tempo de leitura. *Psicologia: Reflexão e Crítica*, 15, 321-331. Recuperado de <https://doi.org/10.1590/s0102-79722002000200010>

- Serrano, F., Genard, N., Sucena, A., Defior, S., & Alegria, J. (2010). Variations in reading and spelling acquisition in Portuguese, French and Spanish: A cross-linguistic comparison. *Journal of Portuguese Linguistics*, 9-10, 183-204.
- Seymour, P. (2005). Early reading development in European orthographies. In M. J. Snowling & Ch. Hulme (Eds.), *The science of reading: A handbook* (pp. 296-315). Oxford: Blackwell.
- Seymour, P., Aro, M., Erskine, J. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94, 143-174. Retrieved from <https://doi.org/10.1348/000712603321661859>
- Simões, E., & Alves Martins, M. (2015). Análise psicolinguística dos erros de leitura em crianças do 1º ao 4º ano de escolaridade. In L. Mata et al. (Eds.), *Diversidade e educação: Desafios atuais* (pp. 258-273). Lisboa: ISPA – Instituto Universitário. Retrieved from http://cie.ispa.pt/ficheiros/areas_utilizador/user23/simoes_e_alves_martins_m_2015_analise_psicolinguistica_dos_erros_de_leitura_em_crianças_do_1o_ao_4o_ano_de_escolaridade.pdf
- Simões, E., & Alves Martins, M. (2018). Reading acquisition in beginning readers: Typical errors in European Portuguese. *Educação e Pesquisa*, 44, e165734. Epub 16 de abril de 2018. Retrieved from <https://dx.doi.org/10.1590/S1678-4634201844165734>
- Sprenger-Charolles, L., & Siegel, L. (1997). A longitudinal study of the effects of syllabic structure on the development of reading and spelling skills in French. *Applied Psycholinguistic*, 18, 485-505. Retrieved from <https://dx.doi.org/10.1017/s014271640001095x>
- Sucena, A., Castro, S., & Seymour, P. (2009). Developmental dyslexia in an orthography of intermediate depth: The case of European Portuguese. *Reading and Writing*, 22, 791-810. Retrieved from <https://dx.doi.org/10.1007/s11145-008-9156-4>
- Treiman, R. (1991). Children's spelling errors on syllable-initial consonant cluster. *Journal of Educational Psychology*, 83, 346-360. Retrieved from <https://dx.doi.org/10.1037/0022-0663.83.3.346>
- Vigário, M., Martins, F., & Frota, S. (2006). A ferramenta FreP e a frequência de tipos silábicos e classes de segmentos no Português. In *Seleção de Comunicações apresentadas no XX Encontro Nacional da Associação Portuguesa de Linguística*.

Análise dos erros de leitura no Português: Dígrafos e estruturas silábicas complexas

A língua portuguesa apresenta vários desafios para as crianças na fase inicial de aprendizagem da leitura, dado que nem sempre as correspondências entre letras e sons são consistentes, havendo letras que podem corresponder a mais do que um som e duas letras que podem corresponder a um único som (dígrafos) e dado que há palavras que contêm sílabas complexas. O objetivo deste estudo foi realizar uma análise psicolinguística dos erros de leitura de crianças no 1º ano ($n=175$) e no 2º ano ($n=137$) de escolaridade e analisar as diferenças entre os erros de leitura nesses dois anos. Para a recolha dos dados foi usado um teste de leitura oral de palavras. Foi realizada uma análise quantitativa e qualitativa do tipo de erros de leitura em palavras contendo dígrafos (ch, nh, lh, gu, rr, ss) e palavras com sílabas complexas (CVC e CCV). Essa análise mostrou que as crianças apresentavam maiores dificuldades em alguns dígrafos específicos e tendiam a simplificar sílabas complexas, adicionando ou excluindo fonemas. A quantidade e qualidade dos erros de leitura de crianças que frequentam os dois anos foram discutidos à luz das teorias de aquisição de leitura e desenvolvimento fonológico das crianças.

Palavras-chave: Erros de leitura, Ensino básico, Aquisição da leitura, Português.