



# An Analysis of the Orthographic Conversion of the Phoneme /S/

Marciano Renato Ribeiro<sup>1</sup>, Raquel Márcia Fontes Martins<sup>2</sup>

Department of Language Studies, UFLA, Brazil. Email: [profmarciano7@gmail.com](mailto:profmarciano7@gmail.com)

Department of Language Studies, UFLA, Brazil. Email: [raquelmartins@ufla.br](mailto:raquelmartins@ufla.br)

Received: 23 May 2022; Received in revised form: 14 Jun 2022; Accepted: 21 Jun 2022; Available online: 26 Jun 2022

©2022 The Author(s). Published by Infogain Publication. This is an open access article under the CC BY license

(<https://creativecommons.org/licenses/by/4.0/>).

**Abstract**— This article describes a study on the learning of spelling, in particular, the multiplicity of graphic representations of the phoneme /s/ by students in the 5<sup>th</sup> year of primary education. Spelling acquisition related to the phoneme /s/ is highly complex for those learning to write in Portuguese, as it involves different relationships—some of them irregular—between the phoneme and the graphemes that represent it (LEMLE, 1994; MORAIS, 2008). This study takes usage-based phonology as its theoretical foundation (BYBEE, 2001; PIERREHUMBERT, 2001). The theory posits that the experience of language use shapes the grammar and lexicon, and that frequency effects are important. To conduct the present study, the written data of students in the 5<sup>th</sup> year of primary education at a public school in the municipality of Bom Sucesso, MG, (Brazil) was collected. The results of this work indicate that writing acquisition is based on usage and experience.

**Keywords**— writing acquisition; phoneme /s/; spelling; usage-based phonology.

## I. INTRODUCTION

The learning of spelling in the Portuguese language is accomplished by understanding the alphabetic writing system (AWS), which requires a complex cognitive process in which the learner perceives that letters represent or indicate the sound of words—the grapheme-phoneme relationship (LEMLE, 1994).

According to Oliveira (2005), the learning of writing is a knowledge construction process mediated by orality. Based on that reflection, the present article investigates the multiplicity of graphic representations of the phoneme /s/ by students in the 5<sup>th</sup> year of primary education, who are on average 10 years old. The orthographic representation of that phoneme is one of the most complex for those learning to write, as it involves the greatest number of relationships between a phoneme and graphemes (LEMLE, 1994; CRISTÓFARO-SILVA, 2010), which will be discussed in the following section. The theoretical apparatus utilized by this study, which was usage-based phonology, will also be presented (BYBEE,

2001; PIERREHUMBERT, 2001). Next, the methodology is presented along with an analysis and discussion of the data, followed by the final considerations.

## II. THE PHONEME /S/ IN WRITING

Different studies (LEMLE, 1994; OLIVEIRA, 2005; CRISTÓFARO-SILVA, 2010) indicate that the graphic representation of the phoneme /s/ is one of the most complex for the learner; in addition to having the highest number of graphemes, there are situations in which the spelling of /s/ is regular and situations in which it is irregular.

Given its complexity, the phoneme /s/ is considered to fall within two types of possible correspondences between letters and sounds in the Portuguese language: 1) a *one-to-more-than-one relationship*, in which the spelling of the phoneme is determined by its position (in a given context, each letter corresponds to a sound and each sound corresponds to a

letter) or 2) a *competitive relationship*, in which several letters compete to represent a single sound in the same position (Cf. LEMLE, 1994).

The graphic representation of the phoneme /s/ is a typical example of multiple relationships between phoneme and grapheme<sup>1</sup> in the language's orthographic system (LEMLE, 1994), which can be observed in the analysis in Figure 1:

The examples demonstrate the complexity of the graphic representation of the sound unit /s/ for students in the spelling acquisition stage and even for those in more advanced academic stages.

Seliar-Cabral (2003, p. 151) argues that the competing alternatives pose a great orthographic difficulty. This is the case with the graphemes “SS” and “Ç”, which represent the phoneme /s/ in “sessão” and “seção,” respectively. Those graphemes are competing in that situation, and there is no rule that determines the use of one grapheme or another. This is an example in which graphemes represent the same phoneme in an identical context (in this case, between the vowels “e” and “ã” in “*sessão*” and “*seção*”).

Silva (2007) presents different conversions of the phoneme /s/ in competitive contexts, which are shown in Table 1:

Given the multiplicity of relationships involving the phoneme /s/, it is possible to confirm the difficulties that learners in the spelling acquisition phase face when selecting the correct form, particularly in competitive contexts, but also in cases of irregularity. Authors such as Morais (2001) argue irregular spellings are learned through memorization using cognitive processes. Along the same lines, Author(s) (2020) assert that spelling is learned through language use: “By using language, learners would, naturally, memorize the spellings of words” (AUTHOR(S), 2020, p. 68).

Considering that use practices directly affect the learning of spelling, the following section addresses the theory of usage-based phonology, which is adopted in this work.

### III. USAGE-BASED PHONOLOGY

<sup>1</sup> According to Luft (1986, p. 238), the phoneme /s/ can also be represented by “xs.” There are only three words in the Portuguese language with this grapheme: ‘exsicar,’ ‘exsolver’ and ‘exsu(d)ar.’

It is also important to mention words such as ‘expectativa’ and ‘atriz,’ in which the phoneme /s/ respectively presented by the letters *x* and *z*, with *z* occupying the place of the coda, i.e., the end of the syllable. In phonological studies, they are treated in Portuguese as an archiphoneme /S/.

As mentioned above, this study adopted usage-based phonology (BYBEE, 2001; BYBEE; HOPPER, 2010), a multirepresentational usage-based model. Based on the assumption that language is an aspect of human behavior, usage-based phonology (BYBEE, 2001) proposes that the cognitive and psychological processes and principles that govern language are not specific to language. They are the same processes and principles that govern other aspects of human cognitive and social behavior. This theory postulates certain basic principles, of which the following are particularly noteworthy (BYBEE, 2001, p. 6):

a) Experience affects representation. The use of forms and patterns both in production and perception affects their representation in memory;

b) Mental representations of linguistic objects have the same properties as mental representations of other objects. Of course, this is the simplest assumption we can make—that the brain operates in the same way in different domains;

[...] (BYBEE, 2001, p. 6).

In usage-based phonology, the word frequency factor proposed by Bybee (2001) is relevant. This factor is considered to affect the mental representations of the language's users. Bybee (2001) proposes two types of frequency effects:

- Token frequency: This refers to how many times a unit, usually a word (for example, “house”), occurs in an oral or written corpus.
- Type frequency: This refers to the dictionary frequency of a particular pattern, i.e., the number of times that a form, such as a sound sequence, occurs in a language corpus - for example, how many times the segmental pattern of the consonant cluster “pr” occurs in a corpus, i.e., the number of different words in which that consonant cluster occurs.

Type frequency is evaluated here with regard to the different grapheme patterns for the phoneme /s/. However, token frequency is also considered because, in line with the study of Huback (2010), the best analysis of a

linguistic phenomenon, such as the one investigated in this work, is the interaction between type frequency and token frequency. According to that author (p. 26), lexical items (such as syllable patterns or types) with a low frequency of occurrence are the first to be affected by changes because the lexical memory of them is weaker, a phenomenon that will be investigated here.

One consequence of frequency effects is that more frequently occurring items or patterns tend to be more accessible in lexical memory and to have a more robust mental representation (PIERREHUBERT, 2001; BYBEE; HOPPER, 2010). Conversely, less frequently occurring items or patterns tend to be more difficult to access in lexical memory and to have a weaker mental representation. This phenomenon will be evaluated in this work. The methodology used to investigate the graphic representation of the phoneme /s/ in writing acquisition will be presented below.

#### IV. METHODOLOGY

This study was conducted at a municipal school in the city of Bom Sucesso/ Minas Gerais (Brazil). The school under study is located in the central area of the municipality and serves students from that area, as well as those from outlying neighborhoods. The institution teaches students from the 1<sup>st</sup> to the 5<sup>th</sup> year of Primary Education I in morning and afternoon sessions and has a total of 530 students. The present work evaluated a class of 15 students in the 5<sup>th</sup> year of education. The students had an average age of 10 years and were already literate.

To collect the data, this work used methodological procedures based on Rocha *et al.* (2010) and presented three activities that will now be described. The first was a dictation activity involving high-frequency words containing multiple graphic representations of /s/. The words in the dictation activity were selected from the *Corpus Brasileiro*<sup>2</sup>, considering lexical items containing the phoneme /s/.

The words in the dictation activity are presented below:

- **Activity 1 – Word dictation**

Listen carefully to the dictated words and write them down.

<sup>2</sup> The *Corpus Brasileiro* project makes one billion contemporary Brazilian Portuguese words—representative of different language types—available digitally. Available at: <[www.http://corpusbrasileiro.pucsp.br/cb/Inicial.html](http://corpusbrasileiro.pucsp.br/cb/Inicial.html)>. Accessed in November 2019. Note that this adult language corpus was selected predominantly due to its abundance of Portuguese language data.

1 – tecido (39,316) <sup>3</sup>	7 – texto (42,106)	13 – professora (10,043)	19 – sábado (36,870)
2 – moça (9,862)	8 – excelente (19,874)	14 – sol (21,766)	20 – passagem (52,535)
3 – aniversário (4,033)	9 – exceto (21,992)	15 – poço (5,720)	21 – nascimento (34,564)
4 – assado (1,295)	10 – extra (9,144)	16 – vacina (15,485)	22 – desça (1,268)
5 – piscina (6,139)	11 – floresça (67)	17 – cebola (3,593)	23 – tóxico (3,397)
6 – cresça (12,725)	12 – crescer (21,523)	18 – palhaço (605)	24 – exceção (36,863)

The second activity involved the dictation of five sentences<sup>4</sup> containing words (bolded below) with the /s/ sound, some of which were included in the word dictation activity presented above.

- **Activity 2 – Sentence dictation**

Listen carefully to the dictated sentences and write them down.

- |  |
|--|
| 1- O nome do <b>palhaço</b> é <b>Paçoca</b> .                |
| 2- A <b> festa</b> de <b>aniversário</b> estava animada.     |
| 3- O menino <b> escreveu</b> um lindo <b> texto</b> .        |
| 4- Perdi minha <b> certidão</b> de <b> nascimento</b> .      |
| 5- <b> Sábado</b> <b> será</b> um lindo dia de <b> sol</b> . |

The decision was made to include some of the words in Activity 1 in the sentences given the potential for different spelling errors to occur in different contexts, i.e., in word writing versus sentence writing. Author and Rocha (2013) indicate that a student who spells a certain word correctly in isolation may spell it incorrectly within the context of a sentence.

It is important to note that although the words in the first activity and the sentences in the second activity of

<sup>3</sup> The numbers in parentheses are equivalent to the number of occurrences of each dictated word in the *Corpus Brasileiro*. In line with YYYY and Oliveira Guimarães (2010), words with fewer than 10,000 occurrences are considered low frequency, while words with more than 10,000 occurrences are considered high frequency.

the instrument are presented in writing here, they were dictated to the students.

The third and final activity, shown below, was the production of a text based on a prompt suggested by this study's researchers:

- **Activity 3 – Text production**

**Text production**

What was the most enjoyable experience of your life? What happened/what was it like? Who was with you? What did you do? Where did it happen? Recount or invent this adventure.

The analysis and discussion of the data collected in this study are presented below.

## V. ANALYSIS OF THE GRAPHIC REPRESENTATION OF THE PHONEME /S/

The data collected in this study, pertaining to students in the 5<sup>th</sup> year of primary education, were analyzed in relation to the written representation of the phoneme /s/ with and without spelling errors<sup>5</sup>.

The data from the word dictation and sentence dictation activities are presented in Tables 1 and 2, while the data from the text production activity are reported and compared with the data from the other two dictation activities.

The tables for the word dictation and sentence dictation activities each contain seven columns. In those tables, the first column presents the words that were targeted in the study. The second column shows the absolute number of correct spellings (as defined by the spelling standard) for each word. In the third column, the percentage of correct spellings is presented. The fourth column shows the absolute number of occurrences with errors (related to /s/) for each word. In the fifth column, the percentage of occurrences with errors is presented. The sixth column shows the types of errors that occurred. Finally, in the seventh column, the total number of errors and correct spellings for each word in the activities is presented, along with overall totals for all the data obtained and their percentages. The words in the tables are presented in descending order according to the percentage of occurrences with errors. The data for the word dictation activity are presented below.

<sup>5</sup> In this work, the term “misspellings” is used to refer to spelling errors, which, according to Moraes (2008), are often the cause of censure and discrimination against learners, both in and outside of school.

### 5.1 Word dictation

The data related to the word dictation activity are shown in Table 2:

In the word dictation activity, there were a total of 137 occurrences with misspellings (38%) and 223 correctly written occurrences (62%), for a total of 360 misspellings and correct spellings.

The words that had high rates of errors and high variation in the types of misspellings were *exceção*, *excelente*, *exceto* and *floresça*, each of which had 15 types of misspellings, corresponding to an error percentage of 100%. The occurrences were the following: *exceção* - *exessão* (5x), *exeção* (1x), *eceção* (2x), *ecesão* (1x), *esseção* (1x), *esxesão* (1x), *exeção* (1x), *esçessão* (1x), *esseção*, (1x), *exceção* (1x). *Excelente* – *exelente* (8x), *ecelente* (3x), *eselente* (3x) *eçelente* (1x). *Exceto* – *exeto* (5x), *eceto* (3x), *eçeto* (2x), *eseto* (2x), *exsseto* (1x), *inseto* (1x). *Floresça* – *floreça* (5x), *floressa* (3x), *floresa* (3x), *floresta* (2x), *floreca* (1x), *floreca* (1x). The next most frequently misspelled word was *desça*, with an error rate of 93.5% and 14 different misspellings: *desa* (6x), *desa* (4x), *deça* (3x), *desca* (1x).

The variation in the type of errors and the number of misspellings in the words containing the digraphs SÇ, XC and SC, which are less frequently occurring<sup>6</sup> syllable patterns/types in Portuguese writing (Cf. REIS, 2020), indicate a high percentage of difficulties related to the phoneme /s/ and its spellings among the three classes of students that participated in the study. When analyzed within the context of the network model, this indicates that the maintenance of a word with irregular inflection and a less-frequent syllable pattern or type depends on its frequency of occurrence. Consequently, according to Huback (2010), regular and frequently occurring words have a stronger lexical representation and thus usually resist analog changes. Conversely, irregular and infrequent words are not sufficiently reinforced in the mental lexicon to allow their irregularity to be maintained, and it is therefore common for users to apply the most frequently occurring paradigms in the language when spelling them (HUBACK, 2010, p. 11).

The next most frequently misspelled words were *creça*, with 11 types of errors (73.5%): *creça* (4x), *cresa*

<sup>6</sup> According to studies conducted by Reis (2020), the graphemes SÇ, XC and SC are digraphs with a low frequency of occurrence in the syllable structure of Portuguese words, while the SS grapheme has a higher frequency rate (related to the phoneme /s/). The mean frequencies of these digraphs per 1000 characters in Portuguese are as follows: SC = 0.03; XC/SÇ = 0.013; SS = 6.33. Reis clarifies that these data were obtained by analyzing a current corpus of journalistic text containing more than 180 million words: CTEMPúblico.

(2x), *cresca* (2x), *cressa* (1x), *greça* (1x), *gresa* (1x). *Piscina*, with 10 types of errors (*picina* (6x), *piçina* (1x), *pisina* (1x), *picisna* (1x)), had an error percentage of 66.5%, and with nine occurrences, *crescer* had a spelling error percentage of 60% (*craser* (4x), *crecer* (1x), *greçer* (1x), *greser* (1x), *crece* (1x), *cereser* (1x)).

In descending order, the words *nascimento*, *tóxico*, *poço*, *extra*, *vacina*, *moça*, *assado*, *tecido*, *aniversário*, *cebola* each had between five and one occurrences of misspelling, representing a spelling error percentage between 33.5% and 6.5%. Based on this analysis, it is evident that among 5<sup>th</sup>-year students, despite the difficulties in the spelling of the phoneme /s/, particularly it is represented by the digraphs XC, SÇ and SC, there is less variation in the types of errors than among younger students. As such, it is important to mention that according to Morais (2008), as the literacy process advances, the student becomes an active learner of spelling through the opportunities for linguistic experience they obtain from their environment.

Furthermore, for the words *palhaço*, *passagem*, *sábado*, *texto*, *professora*, *texto* and *sol*, there were no misspellings; i.e., they were written correctly.

To obtain statistical evidence<sup>7</sup> of the students' spelling performance, the written data from the 5<sup>th</sup>-year class was adjusted to the quasi-binomial model. The results are shown in Table 3.

According to the presented data, the grapheme covariable is considered statistically significant to explain the variable of interest: misspellings among 5<sup>th</sup>-year students. The number of misspellings of the graphemes SÇ and XC were similar, and there was no significant difference. However, the other graphemes tended to have a smaller number of misspellings than the grapheme SÇ. The relative decrease in misspellings between the graphemes SÇ and S is noteworthy as it these graphemes showed the greatest difference.

The analysis of the words in the sentence dictation activity is shown below.

### 5.2 Sentence dictation

<sup>7</sup> A statistical analysis was performed with the quasi-binomial model (WEDDERBURN (1974); NELDER and WEDDERBURN (1972)), with the assumption that the variable of interest followed a binomial distribution. The analysis addressed the number of successes and failures, in which success was regarded as an error in spelling a word with the /s/ phoneme. The variable of interest was the number of misspellings per word, with eight categories of graphemes as covariables: SÇ, XC, SC, Ç, X, SS, S and C. The significance level was 5%. For the analysis of *Activity 1 – word dictation*, the variable of interest (Y) was classified as the number of misspellings per word for the 5<sup>th</sup>-year students.

The 5<sup>th</sup>-year students' data from the sentence dictation activity is shown in Table 4.

The data obtained for the sentence dictation activity indicated 18 misspelled words (10%) out of a total of 180 words.

The word with the highest number of errors (in the sentence dictation activity) was *certidão*, which had six occurrences – *sertidão* (5x), *cerdidão* (1x) – corresponding to an error percentage of 40%. Next was *nascimento*, which had five occurrences in both the word and sentence dictation activities, equivalent to a spelling error percentage of 33.5%. The word *escreveu* also had five occurrences, corresponding to a spelling error percentage of 33.5%.

Lastly, the words *aniversário*, *palhaço*, *sábado*, *texto*, *sol* (selected for both activities), *feira*, *será* and *estava* were not misspelled. Compared to younger students, the 5<sup>th</sup>-year students had even greater tendency towards less variation in the types of errors. This is significant for learning considering that students in this age group are in the consolidation stage of literacy development. According to Morais (2008), students at this stage have begun to understand that their writing cannot be merely a representation of speech, i.e., they must go well beyond that by, for example, understanding the specific characteristics (rules and irregularities) of spelling and being mindful of the need to use one or more graphemes to write different words in the language.

Based on these considerations, the data from the 5<sup>th</sup>-year students was adjusted to a binomial generalized linear model<sup>8</sup>, and the results are presented in Table 5.

Based on the data obtained, it is possible to affirm that the grapheme covariable offered a statistically significant explanation for the number of misspellings by the 5<sup>th</sup>-year students. Between the grapheme SC and the graphemes Ç, X and C, no significant differences in the number of errors were expected. Between the grapheme SC and the grapheme S, a relative decrease of 96% in the number of misspellings was expected.

The analysis of the text production activity is presented below

### 5.3 Text production

<sup>8</sup> For the analysis of Activity 2, the data was adjusted to a Binomial Generalized Linear Model (assuming the data do not present overdispersion; Nelder and Wedderburn (1972)) and a Quasi-Binomial Model (assuming the data present overdispersion; Wedderburn (1974)). The covariates used to explain the response of interest (misspelling) are the following: Grapheme (SC, Ç, X, S, C).

The analysis of the 5<sup>th</sup>-year students' text production found a total of 373 words containing the phoneme /s/; of these, 18 words exhibited errors (5%), and 355 words were spelled correctly (95%).

The misspelled words were *passarela* (*pasarela*), *criança* (*crianca*), *ansiosos* (*anciosos*), *piscina* (*picina*), *recompensa* (*recompença*), *vassoura* (*vassoura*), *funcionários* (*funsonarios*), *assim* (*asim*), *história* (*hitoria*), *vocês* (*vosês*), *disse* (*dise*) and *passeio* (*paseio*). The examples with errors reflect the complexity of the orthographic principles of Portuguese, in which the relationships between sounds and letters can be regular regardless of context, regular according to context or arbitrary. The students still have not learned to spell according to the basic rules and irregularities of the language's orthography, in which phoneme-letter relationships are not always the same (SOARES; 2020, p. 143).

Lastly, in the analysis of Activity 3, text production, the exact binomial proportions test was performed (Clopper and Pearson (1934); Conover (1971); Hollander and Wolfe (1973)) to test the alternative hypothesis that the probability of the occurrence of a spelling error is different from 0.5. In the three tests, a  $p$ -value =  $2.2 \times 10^{-10}$  was obtained.

In the analyzed phenomenon, the phoneme /s/ is represented by different graphemes according to its position in the word, i.e., it is regular according to the context. Furthermore, more than one letter can compete to represent it in the same position. This causes difficulties for learners in the literacy development stage and for those who are already literate. Consequently, Lemle (1994, p. 23) argues that cases such as these are the most difficult when learning the written language and that there is no phonetic principle that can guide the choice between competing letters.

#### 5.4 General evaluation of the three study activities completed by the 5<sup>th</sup> year students

The data from the 5<sup>th</sup> year students' word dictation, sentence dictation and text production activities correspond to a total of 913 words; of these, 173 words presented spelling errors (19%) and 740 words were spelled correctly (81%).

Table 6 shows the total occurrences of correct spellings and misspellings in words with the phoneme /s/ in the written activities. Based on these data, it is possible to verify the 5<sup>th</sup>-year students' total number of correct spellings and misspellings of words containing the phoneme /s/ for each activity. The last row shows the overall total (absolute number and percentage) of

occurrences; there was a significant percentage (19%) of errors in the spelling of the phoneme /s/ among the evaluated 5<sup>th</sup>-year students:

The overall result of 81% correct spelling and 19% misspellings in the representation of the phoneme /s/ points to the fact that the spelling of students in the consolidation stage of literacy (5<sup>th</sup> year of primary education) shows considerable development, but work and reflection on the spelling rules is needed in the coming school years.

Figure 2 refines the analysis, presenting the number of errors involving each grapheme representing the phoneme /s/ in the words analyzed in this study:

The data in Figure 2 show that the highest number of misspellings of the phoneme /s/ was related to the digraph patterns SÇ, XC and SC. A pertinent fact is that, in the students' writing, the grapheme Ç also had a high number of errors— more, in fact, than the number of errors for the digraph SÇ.

The high number of errors in the spelling of the grapheme Ç in the target words is related to the arbitrary nature of spelling, in which more than one letter can compete to represent the same phoneme in the same position - as, for example, in the words *nossa*, *moça* and *nasça*, in which different graphemes (SS, Ç and SÇ) are used to represent the phoneme /s/ in the intervocalic context between O and A. In this case, specifically, the phoneme /s/ occurs in competitive contexts involving different graphemes: in *nossa*, the graphemes SS or Ç (before a back vowel) are observed; in *moça*, the graphemes S or Ç (before a back vowel) are observed; and in *nasça*, the graphemes S or SÇ (before a front vowel) compete for the orthographic representation of the phoneme /s/.

The high rate of errors related to the spelling of digraphs supports the “variable occurrence frequency” hypothesis in which the small number of words with syllable types containing the digraphs (SÇ, XC and SC) representing the phoneme /s/ are more subject to spelling errors. According to this result, and guided by usage-based phonology, lexical items (syllable patterns or types, for example) that are used frequently are easily remembered; conversely, items that are rarely used are more difficult to recover from the mental lexicon and are more subject to doubt. Additionally, according to Huback (2010, p. 10), it is quite common for people to experience a memory lapse when trying to remember words that they rarely use.

## VI. FINAL CONSIDERATIONS

This study analyzed the graphic representation of the phoneme /s/ by learners of spelling. By analyzing the written production of 5<sup>th</sup> year students, it was possible to observe that the highest rate of errors involving the phoneme /s/ was related to syllable patterns containing the digraphs SÇ, XC and SC, which, according to Reis (2020), are rare graphemes in Portuguese writing. These items are thus less robust and likely have a low lexical frequency in the minds of the study subjects. Conversely, the graphemes C, S and SS, which are more frequent, had the lowest rates of spelling errors, possibly because their lexical access is easier. It was thus confirmed that type frequency was relevant in this study. It is suggested, based on usage-based phonology (BYBEE, 2001), that experience is a key factor in the organization and management of language knowledge and in writing acquisition.

Following Morais (2008), it is also possible, based on this study, to consider that errors are actually part of the spelling acquisition process; it is not possible to instantaneously spell correctly. Future studies could propose a didactic sequence to systematize the teaching of the different graphemes that represent /s/ and to focus on less-frequent graphemes (SÇ, XC and SC) rather than only the more frequently occurring ones (S, C and SS).

## REFERENCES

- [1] BYBEE, J.; HOPPER, P. (ed.). *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins, 2010.
- [2] BYBEE, Joan. *Phonology and language use*. Cambridge: Cambridge University Press, 2001.
- [3] CLOPPER, C. J. & Pearson, E. S. *The use of confidence or fiducial limits illustrated in the case of the binomial*. *Biometrika*, 26, 404-413, 1934. doi: 10.2307 / 2331986.
- [4] CRISTÓFARO SILVA, Thaís. *Mental representations in the acquisition of oral and written language*. *ABRALIN Magazine*, v. 1, p. 153-178, 2011. Available at: <[http://www.lettras.ufmg.br/padrao\\_cms/documentos/profs/thaiscristofaro/2012\\_art\\_representa%20mentais%20na%20aquisi%20da%20linguagem%20oral%20e%20escrita.pdf](http://www.lettras.ufmg.br/padrao_cms/documentos/profs/thaiscristofaro/2012_art_representa%20mentais%20na%20aquisi%20da%20linguagem%20oral%20e%20escrita.pdf)>. Accessed in March 2020.
- [6] CRISTÓFARO SILVA, Thaís. *Phonetics and phonology of Portuguese: study guide and exercise guide*. 10. ed. São Paulo: Contexto, 2010.
- [7] FARACO, Carlos Alberto. *Writing and literacy*. 9. ed. São Paulo: Contexto, 1992.
- [8] FERRARI, L. *Introduction to Cognitive Linguistics*. São Paulo: Contexto, 2011.
- [9] GELC. *Brazilian Corpus*. São Paulo: GELC. Available at: <http://corpusbrasileiro.pucsp.br/cb/Inicial.html>. Accessed in: nov. 2019.
- [10] HUBACK, ANA PAULA. *Irregular Plurals of Brazilian Portuguese: Frequency Effects*. *ABRALIN Magazine*, v. 9, p. 11-40, 2010.
- [11] LEMLE, M. *Theoretical guide for the literacy teacher*. 8. ed. São Paulo: Ática, 1994.
- [12] AUTHOR; OLIVEIRA-GUIMARÃES, D. M. L. *Effects of frequency in the written production of consonant clusters*. *Linguistic Studies (São Paulo)*, v. 39, p. 440-451, 2010.
- [13] AUTHOR; ROCHA, G. A. S. *Research on correction of large-scale assessment writing items in the field of literacy*. In: I CONBALF - Brazilian Literacy Congress, 2013, Belo Horizonte. *Anais do I CONBALF - Brazilian Literacy Congress, 2013*. v. 1.
- [14] MORAIS, A. G. de. *Spelling: teach and learn*. 4th ed. São Paulo: Ática, 2008.
- [15] NELDER, John Ashworth & Wedderburn, Robert WM. *Generalized linear models*. *Journal of the Royal Statistical Society: Series A (General)*, v. 135, n. 3, p. 370-384, 1972.
- [16] OLIVEIRA, M. A. *Linguistic knowledge and appropriation of the writing system*. Belo Horizonte: Literacy, Reading and Writing Center - Ceale / FaE / UFMG, 2005. *Literacy and Literacy Collection (Trainer's Notebook)*.
- [17] PIERREHUMBERT, J. *Exemplar dynamics: Word frequency, lenition and contrast*. In: BYBEE, J. (Ed.); HOPPER, P., *Typological studies in language*, Amsterdam: John Benjamins.
- [18] REIS, Rogério. *Frequency Tables in Portuguese*. Rogério Reis. Faculty of Sciences, University of Porto, v.1, p. 1-2, 2020. Available at: <<https://www.dcc.fc.up.pt/~rvr/naulas/tabelasPT/>>. Accessed on: 14 nov. 2020.
- [19] AUTHOR(S). *Study about the graphic representation of the phoneme / s / in the writing of students of the 3rd year of elementary school*. *Caligrama Magazine (UFMG)*, v. 25, p. 63-84, 2020.
- [20] ROCHA, G. et al. *Literacy assessment: how to use it in its diagnostic function?* In: *Literacy Guide*. Belo Horizonte: Segmento, 2010, p. 78-90.
- [21] SCLiar-CABRAL, Leonor. *Principles of the Brazilian Portuguese alphabetical system*. São Paulo: Contexto, 2003.
- [22] SOARES, Magda. *Alphabets: Every child can learn to read and write*. São Paulo: Contexto, 2020.
- [23] WEDDERBURN, Robert WM. *Quasi-likelihood functions, generalized linear models, and the Gauss — Newton method*. *Biometrika*, vol. 61, n. 3, p. 439-447, 1974.