EFEITOS ORTOGRÁFICOS NA PERCEPÇÃO DA FALA: EVIDÊNCIAS DE UMA TAREFA DE DECISÃO LEXICAL AUDITIVA COM FALANTES BRASILEIROS DE INGLÊS

ORTHOGRAPHIC EFFECTS ON SPEECH PERCEPTION: EVIDENCE FROM AN AUDITORY LEXICAL DECISION TASK WITH BRAZILIAN SPEAKERS OF ENGLISH

Alison Roberto Gonçalves
Universidade Federal do Paraná, Curitiba, Paraná, Brasil
arg@ufpr.br

Rosane Silveira
Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, Brasil
rosanesilveira@hotmail.com

Resumo: O presente estudo investigou efeitos ortográficos em uma tarefa de percepção da fala realizada por falantes brasileiros de inglês. O estudo empregou um léxico artificial que simulava relações grafo-fônicas opacas e transparentes do inglês em posição nuclear (*deit*, *toud*). Participantes aprenderam esse novo conjunto de palavras através de um paradigma de treinamento de exposição repetida, no qual foram inicialmente introduzidas formas fonológicas associadas aos seus pares visuais, seguidas de associações às suas representações ortográficas. Uma tarefa de decisão lexical auditiva foi administrada após o treinamento. Resultados indicaram que a consistência ortográfica não afetou o tempo de reação dos sujeitos com o léxico que haviam aprendido, embora o tempo de reação com palavras opacas tenha sido maior. No entanto, a ortografia influenciou o tempo de reação registrado para palavras com as quais participantes não haviam recebido treinamento. Entretemos que ter que realizar análise lexical de palavras desconhecidas levou os participantes a recrutarem a ortografia como um mecanismo que auxilia na análise lexical. O recrutamento ortográfico foi concebido, então, como um processo estratégico que auxilia a decisão lexical em tarefas auditivas temporalizadas.
Palavras-chave: Percepção da Fala; Ortografia; Tarefa de Decisão Lexical

Abstract: The present study investigated whether orthographic effects arise in a speech perception task performed by Brazilian speakers of English. The study employed an artificial lexicon that simulated opaque and transparent grapho-phonic English relations in nuclear position (e.g., *deit*, *toud*). Subjects were compelled to learn this new set of words through a repeated-exposure training paradigm in which they were initially introduced to phonological forms associated with their visual pairings, followed by associations to their orthographic representations. An auditory lexical decision task was taken after training. Results indicated that orthographic consistency did not affect subjects' latencies with the trained lexicon, although their reaction times were relatively longer with opaque items. However, orthography influenced latencies registered for untrained items in the task. We entertained that having to conduct lexical analysis with incoming unfamiliar auditory items compelled subjects to recruit

orthography as a mechanism to aid lexical analysis. Orthographic recruitment was thus conceived as a strategic process that assists lexical decision in timed auditory tasks.

Keywords: Speech Perception; Orthography; Lexical Decision Task

1 INTRODUCTION

A recent endeavor in psycholinguistic studies has demonstrated the signature that orthography assigns to speech perception and production (Escudero, 2011; Qu & Damian, 2019). Such a scientific orientation has led orthographic input to be treated as an empirical variable rendering cross-linguistic influences in speech acquisition (Bassetti, Escudero & Hayes-Harb, 2015). Hence, our objective was to investigate whether Brazilian speakers of English recruit orthography when performing an on-line speech perception task with an artificial lexicon. The theoretical underpinnings that buttress the current enterprise are presented below: an account for how the development of phonological representations unfolds with language acquisition is presented first, followed by a review of studies that investigated orthographic effects in speech perception with speakers of alphabetic languages. Finally, the method is described, with results discussed in the subsequent section of the article[1].

2 SPEECH PERCEPTION AND LANGUAGE ACQUISITION

A potential account for speech perception to be placed at the forefront of language acquisition has been motivated by infants' keen sensitivity to phonetic contrasts (Kuhl, 2000; Werker, 1995; Werker & Curtin, 2005; Werker & Gervain, 2013). From an early age, infants show preference for speech sounds over similarly complex nonspeech sounds (Vouloumanos & Werker, 2007), and are able to discriminate any phonetic contrasts extremely well (Maye, Werker, Gerken, 2002). Yet, by 10 to 12 months of age, they no longer maintain sensitivity to contrasts other than those in the native language, showing that these become language-specific with the establishment of native phonetic categories (Werker & Gervain, 2013).

PRIMIR[2], a conceptual framework put together by Werker and Curtin (2005), proposes that phoneme-level representations emerge gradually from the phoneme plane as statistical regularities that are extracted from word-level input. Initially, acoustic variability is used to discern phonetic organization (Maye et al., 2002), and as the phoneme plane emerges and experience with language is gained, infants are able to detect acoustic dimensions that are most informative due to perceptual salience (Werker & Gervain, 2013), which can contribute to the modifiability of phonetic categories (Werker & Curtin, 2005).

The myriad of linguistic processes that are guided by the perceptual properties of language unfold with exposure and bootstrap early language acquisition (Werker, 2018). Yet, mechanisms involved in language acquisition are changed considerably once subjects start being schooled and become literate (Kolinsky, 2015; Reis & Castro Caldas, 1997), hence altering the phonological representations developed in the lexicon (Werker & Gervain, 2013).

---

[2]Processing Rich Information from Multidimensional Interactive Representations.

Orthography is thus rendered as another factor worth of appraisal in the development of phonological representations.

When it comes to the representation of orthographic knowledge in the lexicon, Katz and Frost (2001) elucidate that

> the process of forming an orthographic representation is not completely based on perceiving, coding, and storing visual orthographic information alone. The claim is that the internal orthographic representation is not formed simply as a passive reflection of the visuo-spatial characteristics of the print but, rather, the reader's knowledge of the relations between orthography and phonology shapes the internal representation (Katz & Frost, 2001, p. 299).

It is noteworthy to say that orthographic forms will hold greater influence in the shaping of phonological forms when subjects become literate and therefore hold knowledge of graphophonic conversion relations, and when orthography is a robust stimulus in input available in instructional settings, with learners having their attention constantly driven to it, as is the case of acquiring an L2 (Veivo & Järvikivi, 2013; Yoncheva et al., 2013).

In concert with this, Veivo and Järvikivi's (2013) elucidation for orthographic and phonological information in word processing is worth consideration. These authors submit that there have been two main explanations to account for the conjecture of orthography and phonology in the lexicon. One is regarded as the on-line *co-activation account*, which posits that orthographic and phonological representations co-exist and are strongly linked at both pre-lexical and lexical levels. As representations are linked, they can be activated automatically in linguistic processing. The other account is the *restructuration account*, which claims that there are no separate representations for each of the systems. Instead, phonological representations that are pre-existing fundamentally change when one learns to read an alphabetic script. Thus, these representations, in nature, are abstract and include both orthographic and phonological information. Once orthographic effects arise during spoken word processing, these are interpreted as resulting from the abstract phonological representations influenced by orthography (Veivo & Järvikivi, 2013, p. 865).

However, Veivo and Järvikivi (2013) argue that a more plausible account is the *co-structuration account*, in which orthographic information contributes in parallel to the formation of lexical categories, along with phonological information. Therefore, the initial plane of representations would be phonetic in infancy, but with vocabulary growth, these representations are sharpened up into phonological representations, which are stable and encompass articulatory information as well (Bertelson, Vroomen, & de Gelder, 2003; Werker & Curtin, 2005). These representations then become co-structured with orthographic information because of a functional link that is established between orthographic and phonological representations with the attainment of literacy (Kolinsky, 2015).

Veivo and Järvikivi's (2013) claim can also be extended to the case of learning an L2. Their postulate allows one system to dominate over the other in specific cases, as with early learners in instructional settings, when orthography is believed to be more robust due to great amounts of written input, leading orthography to be regulatory over phonological encoding. Especially with an L2, the orthographic forms are learned either before or simultaneously to phonological forms, hence both of these systems are able to contribute to the formation of lexical entries, even if one is less autonomous than the other.

After having laid out the underpinnings of speech perception in language acquisition and explained how orthography becomes functionally linked to phonological representations, the next section presents a review of studies that investigated orthographic effects on speech

perception. The scope of this review sheds light on the signature of orthography on phonology in studies in which spoken word processing is the underlying cognitive operation, within a variety of experimental conditions, both on-line, as processing unfolds (e.g., auditory lexical decisions), and off-line, as the outcome of processing (e.g., discriminatory perception tests). Only studies with cross-linguistic influences, in which subjects were speakers and learners of alphabetic languages, have been included.

3 ORTHOGRAPHIC INFLUENCES ON SPEECH PERCEPTION

Escudero, Hayes-Harb, and Mitterer (2008) investigated the effects that the phonemic mappings of /æ/ and /ɛ/ posed to the learning of novel words (e.g., t**a**ndek, t**e**nzer) to native Dutch speakers. Ten words were created for the experiment, five for each target phoneme, all of which adhered to English phonotatics and were paired with a control word that was identical to the target, except for the vowel in the first syllable (/u/, e.g., tenzer – tunzer). This was done in order to make subjects pay attention to the stimuli and also to have balanced vowel contrasts that later on could be compared based on their level of difficulty - /ɛ/ and /u/ are relatively easy, and the target contrast, /æ/ and /ɛ/, is not only difficult given the acoustic characteristics of the phones, but also do not exist in Dutch, as /æ/ is not found among Dutch vowels. Two learning conditions were created: Auditory + Spelled forms and Auditory Forms only. Subjects would participate in 10 learning blocks in which they would be required to click on the object that they though represented the new word and then on a geometric form, thus they learned to associate each word with a visual object.

Next, tested with an eye tracker, participants had to identify the picture that represented the target word they had heard. Among the visual items displayed, the picture that represented the target's competitor was also included. The authors discovered an asymmetric pattern for the recognition of words that were learned through the presentation of both aural and orthographic inputs. Words containing /æ/ would trigger participants to look at words that contained both /æ/ and /ɛ/, whereas /ɛ/ triggered participants to look at words with /ɛ/. For the Auditory Only condition, participants fixated equally on words containing /æ/ or /ɛ/. Therefore, the investigators concluded that lexical knowledge of spelled forms can trigger asymmetric lexical activation, that is, a lexical contrast is established on the basis of metalinguistic knowledge, but participants were not able to successfully map this contrast to the phonetic forms in the time given. Even though each target phoneme investigated was consistently mapped onto a graphemic representation (/æ / - <a>, /ɛ/ - <e>), participants succeeded only in learning the new words in 30 minutes, but were not able to tell a distinction between the contrasts.

Escudero and Wanrooij (2010) investigated whether orthography influenced the perception of Dutch vowels by Spanish learners of the language by predicting that Spanish learners would treat Dutch as a transparent orthography, thus transferring the decoding skills employed in their L1. They would, therefore, link graphemes such as <ie> to the diphthong /jɐ/ instead of the vowel /i/, for instance. The stimuli used in the task consisted of 20 natural speech tokens for each of the vowels /ɑ/, /a/, /ʏ/, /y/, /ɪ/ and /i/, which were produced by male and female Northern Dutch speakers. Participants took part in two tasks: Audio only, and Audio plus orthography. In the Audio task, they were required to decide whether the first sound was more similar to the second or more similar to the third. Results yielded that learners tended to associate Dutch /ɪ/ to the front unrounded Spanish vowel that was represented by <i>, while Dutch /y/ was linked to rounded Spanish vowel represented by <u>, /ɪ – y/ being the auditorily

less confused pair, whereas /a–ɑ/ had a significantly lower rate of correct responses than the other contrasts. In the orthography condition, subjects were asked to choose from the orthographic representations of the 12 Dutch monophthong vowels (<aa>, <a>, <ie>,<i>, <uu> and <u>), visually presented on a computer screen. Contrarily, the /a–ɑ/ contrast was the easiest, as the doubling of the letters <aa> versus <a> led listeners to pay attention to the durational cue. Thus, the orthographic cues enhanced temporal cues and helped learners identify this vowel. On the other hand, /y/ was the most difficult vowel for learners in this task, being identified as <u> instead of <uu>. Overall, the study showed that the presence of orthographic information was of facilitative nature for some percepts, but represented a hindrance for others.

Simon, Chambless, and Alves (2010) trained American participants to learn the French vowels /u/ and /y/ in two conditions: auditory information only and auditory information linked to spelling. They were trained with words that contained the target vowels, forming minimal triplets (e.g., dûge, douge, dige). Training consisted of displaying a picture along with its orthographic form on a computer screen, followed by its corresponding audio form, or just the picture and the corresponding audio form. Participants were tested on their ability to match the picture to its audio form. Moreover, a perception task was designed to test for the ability of generalizing the novel stimuli. In trials with triplets, participants had to identify if the second word they heard was either the same as the first word or as the third word. No significant results were found for the word learning experiment, which presented a great deal of variation. As concerns the perception task, no significant results were found between groups and participants tended to perform very well. Simon et al. (2010) argued that the lack of significant results in their tasks was due to the great stimuli load in declarative memory: subjects had to learn the meaning of 36 new words and of 12 distractor words in a 25-minute training section.

Next, Simon et al. (2010) revised the experiment, presenting participants with a longer training phase and fewer items to learn. They also hypothesized that the American listeners did not have single-category assimilation for French /y/ and /u/ and that orthographic information would assist learning only in cases of single-category assimilation. The objective of this experiment was to identify to which native categories participants mapped the French vowels. The stimuli was presented over headphones while five words were displayed on the screen (peek, pick, booth, book and poke), containing the English vowels /i:/, /ɪ/, /u:/, /ʊ/, and /oʊ/. Participants had to choose the vowel that most resembled the vowel aurally presented. Much variation was found in the categories which participants assigned French /y/ tokens to, and no single-category assimilation was found for both of the target categories, /u/ and /y/, to English /u/ (e.g. 'booth'). Interestingly, the authors discovered that the consonantal context helped participants to distinguish between French /u/ and /y/, especially in a bilabial context. This might have been the reason why they did not have to recruit orthography to better learn the vowels of the task, as they already could distinguish between the two.

The next step taken by Simon et al. (2010) was to adapt the first two experiments, having only one native speaker of French record the stimuli and all the vowels inserted in a constant alveolar context, for these factors might have influenced the previous results. Again, in the word learning task, no differences were found between groups, who once more performed considerably well. As for the perception task, by keeping the coda consonant constant, /u/ and /y/ were categorized in similar ways to English /u/, attesting for the presence of single-category assimilation. As for the effects of orthography, the authors surmise that these listeners may not rely on spelling to create distinct phonological categories as would speakers of more transparent orthographic systems.

Pytlyk (2011) inquired whether Canadian participants who learned Mandarin via a familiar orthographic script (Pinyin, the Romanized transcription) differed from participants

who learned it via a non-familiar script (Zhuyin, the syllabary system), in terms of perception of English-Mandarin consonant-pairs. Some of the targets tested, followed by their phonic mappings in English and Mandarin are: <c>→ [s], [tsʰ]; <z>→ [z], [ts]; <r>→ [ɹ], [ʐ], <h>→ [h], [x]. The author's design included a pre-test, an instruction phase that lasted 4.5 hours distributed over three meetings, and a posttest. While receiving training, participants were not allowed to write alphabetic symbols to help them remember any Mandarin sounds. The researcher created the stimuli by using CV syllables, in which the target consonant appeared in onset position, followed by an [a] vowel. The perception test was a discrimination task in which participants had to choose the odd item out.

Pytlyk's (2011) study revealed no significant differences in the responses among the experimental groups (familiar orthography, unfamiliar orthography, and control group with no orthography). The scholar tentatively offers the explanation that this may reflect an inability to disassociate the L2 forms from the L1 orthography, given the constant reference to the latter in language classrooms. Thus, no interference in perception occurred given how constant those targets mapped among the L1 forms. Anecdotal evidence from the study suggested that the participants found it difficult to disassociate themselves from thinking in terms of letters, which was "virtually impossible" (p. 552). Pytlyk (2011) also acknowledged that the 4.5 hours of training received by the participants were not "able to 'out-influence' a lifetime of associations made in the L1 orthographic code" (p. 554). Overall, her study showed how difficult it might be to set up an experimental condition that resembles a more ecological learning experiment.

Veivo and Järvikivi (2013) investigated orthographic influences in French spoken word recognition by Finnish learners. To observe whether the activation of the orthographic form facilitated the processing of the phonological form, these authors used masked-cross modal priming in a lexical decision task. Experiment 1 presented real word repetitions of the target in both auditory and orthographic forms ([staʒ] <stage>) and nonword pseudohomophones that could be pronounced like the target words ([staʒ] <staje>). In experiment 2, Finnish-based primes preceded French auditory words with (1) orthographic overlap (Finish <huivi> "scarf", French <huile> "oil"), which were semantically unrelated and presented no phonological overlap; (2) Finnish pseudohomophones that could be pronounced like the target words (phonological overlap), e.g., <yil> ([yil]) to prime <huile> ([ɥil]). Both experiments also presented a third condition that used words with no semantic, phonological, or orthographic overlap with the target. Participants were instructed to decide as quick and accurate as possible if the spoken word was a French word or not. In experiment 1, the visual primes facilitated the processing of the auditory targets, thus showing that the participants were able to map the orthographic forms into phonological forms. Repetition also reduced the number of errors. Such a condition appeared to be stronger than the pseudohomophone effect. Familiarity was a good predictor of latencies and error data, whereas high proficiency participants showed stronger effects for repetition. With this experiment, Veivo and Järvikivi (2013) were able to establish cross-modal influence from orthography to phonology. Experiment 2 showed a facilitative effect due to orthography, but this effect was proficiency-dependent. More proficient learners did not profit from the availability of orthography, as the facilitative effect on the latencies decreased for familiar words. As concerns the lower proficiency group, stronger facilitative effects were induced by activating the phonology via L1 grapheme-to-phoneme mappings. Overall, the authors were able to conclude that orthography offers sublexical facilitation for L2 processing when the lexical representations are not yet fully stable, confirming the suitability of the *co-structuration account* that allows for both orthographic and phonological influences in lexical representation (even if one of overrides the other) (Veivo & Järvikivi, 2013).

Escudero, Simon, and Mulak (2014) investigated whether orthographic congruence would influence participants' accuracy on a word recognition task, whose words contained either perceptually easy or difficult minimal and non-minimal pairs. Escudero et al. (2014) had 73 Spanish listeners learn novel Dutch words. Forty-three of these listeners had varying levels of Dutch proficiency and 30 were unfamiliar to Dutch. The words (ex., "paag", "pag", "pieg" etc) adhered to Dutch phonotatics and were recorded by a native speaker of Dutch for the stimuli. During training, participants were exposed to either one of two conditions in order to examine the effects of exposure to orthography. They were taught word-object associations through the visual presentation of a picture and its corresponding aural form, or its corresponding aural form along with its orthographic form. During testing phase, participants were required to identify the picture from a pair of images that corresponded to the spoken pseudoword heard.

The experimenters predicted that the effect of orthographic congruence would be stronger for listeners exposed to the orthographic representation of non-words, regardless of their proficiency level, because orthography is always activated during word learning and therefore it can inhibit the learning process. Moreover, the authors claimed that a higher proficiency may deactivate the L1 orthography, leading to a decrease between congruent and incongruent orthographic mappings of both languages involved in the task (Spanish and Dutch). The results demonstrated an absence of effects concerning the Word-learning condition, as participants who were exposed to auditory forms only or both auditory and orthographic forms performed similarly without significant differences. The exposure to orthography during training was not entirely beneficial or hindering. There was a benefit for congruent word pairs, but participants performed worse on incongruent word pairs. Therefore, the authors concluded that the influence of orthography on speech processing relies greatly on the congruence of grapheme to phoneme conversions, rather than simply the addition of a visual referent. As regards participants' proficiency, Escudero et al. (2014) showed that Spanish listeners who were naïve to Dutch did better in identifying members of non-minimal pairs compared to minimal pairs and were more accurate at perceptually easy contrasts. Dutch learners, on the other hand, were more accurate at distinguishing perceptually difficult minimal pairs and more accurate with non-minimal pairs.

In general, these studies demonstrate that orthographic influences to speech perception are still largely undetermined. For some it might be a hindrance (Escudero et al., 2008), or a redundant factor (Escudero & Wanrooij, 2010), while for others it more clearly exerts influence (Escudero et al., 2014; Simon et al., 2010; Veivo & Järvikivi, 2013). The next section outlines the method of the study and describes in detail the Auditory Lexical Decision task.

4 METHOD

This study aims at answering the following research question: How does orthography affect speech perception in an auditory lexical decision task? To do so, the study employed an exposure-based training paradigm through which subjects were compelled to learn picture-aural-orthographic forms associations of words belonging to an artificial lexicon. Exposure-based training has been found to boost learning of both simple and complex grammar (Antoniou, Ettlinger & Wong, 2016), and has proven advantageous for lexical processing and word production (Van Assche, Duyck & Gollan, 2016). It is paramount to say that the method of the present study is a conceptual replication of previous studies which employed training with an artificial lexicon (Escudero et al., 2014; Tamminen, Davis & Rastle, 2015; Taylor, Davis, & Rastle, 2017; Rastle et al., 2011), but specially the study of Rastle et al. (2011), given that the design of their training tasks was replicated here. For details concerning the factors

considered for the development of the pseudowords used in the artificial lexicon, namely, phonotatics, the target percepts and their spellings, and neighborhood density, refer to Gonçalves (2017). All pseudowords used in this study are available in Appendix A. Moreover, details regarding the recording session of the auditory stimuli used in the training phase and in the Auditory Lexical Decision task are provided elsewhere (Gonçalves, 2017).

4.1 TRAINING

In the training phase, participants took part in study and verification blocks in which they were introduced to the study stimuli. Stimuli presentation was controlled with DMDX (Forster & Forster, 2003), version 5.1.3.6. (April 2016). Participants took the study and verification blocks in a quiet room, while sitting in front of a computer with a headset on. The training session consisted of eight study and eight verification blocks. Each study block presented the stimuli three times, in three different sets. Among each set, participants were offered a short break. A verification block presented the stimuli twice, in two different sets, between which participants were offered a short break.

In study blocks, participants were shown a picture of a novel object while listening to its spoken form over headphones. They needed to repeat the object's name after each trial. In order to familiarize the participant with the procedure, three trials were provided as a familiarization block. The stimuli consisted of the 22 new words which were presented twelve times each, adding up to a total of 264 trials split into eight different study blocks during training.

A participant firstly took part in three training sets in one study block, with a total of 66 trials, which were then followed by a verification block with two testing sets. Each trial presentation in a study block lasted 2000ms to allow for object recognition and phonological encoding. This duration is comparable to previous research involving training on new lexical items (2000ms: Simon et al*.,* 2010; Bartolotti & Marian, 2017; Escudero, 2015). The participants were explicitly instructed to repeat each spoken form while paying attention to the visual form that was presented simultaneously on the computer screen. No response was registered from study blocks.

After each study block, each subject took part on two testing sets in a verification block. Verification blocks consisted of a Picture Identification Task in which participants needed to choose, from two pictures displayed on the computer screen, the one that matched the stimuli heard. Feedback was given immediately for wrong responses with the message "*Wrong response! Try harder!*". Each trial was available for 5000ms before time out occurred in case the participant did not respond. In such a case, the message "No response" was displayed on the screen before the next trial came up. Four practice trials were provided to familiarize the participant with the experiment before the presentation of the verification block started. Each verification block in the Picture Identification Task contained 44 trials, divided into two testing sets.

Beginning with the fifth study block, participants were exposed to the lexicon written forms in conjunction with the spoken forms and the picture in study blocks. The procedure was very similar to the protocol followed with the first four study blocks. Each trial lasted 2200ms to allow for picture recognition, and orthographic and phonological encoding. These extra 200ms were allowed for orthographic input was presented, thus entailing in one extra process that was not present in the first four parts of training. After three study sets in a study block, participants were required to take the Picture Identification Tasks with two testing sets in a verification block in which they needed to select the target, from two pictures displayed on the screen, which matched the stimulus heard. Feedback was given immediately for wrong

responses with the message "*Wrong response! Try harder!*". Each trial was available for 5000ms, before time out occurred in case the participant did not respond. In such a case, the message "No response" was displayed on the screen before the next trial came up.

4.2 TESTING

Subjects were tested with an Auditory Lexical Decision task. This type of task entails processes of lexical access or lexical search as well as the analysis of the speech signal (Goldinger, 1996). The execution of the task required participants to decide whether the stimulus was learned in training or not, by pressing "yes" or "no" corresponding buttons on the computer keyboard. "Yes" responses should be given for the pseudowords that participants learned during training, whereas "no" words are items prepared just for this task. As participants were compelled to conduct lexical analysis of the items presented in the task, it is the aim of the study to observe whether orthography is recruited during this process to aid the spoken recognition of the item. Such a process can be evidenced by significantly different response times for "yes" pseudowords of differing orthographic depths, as well as accuracy of response. Response times for "no" responses were also looked at in comparison to the "yes" responses.

The word-initial cohort theory (Marslen-Wilson & Zwitserlood, 1989), used for assessing spoken words, was revisited for the preparation of the "no" words in the present study. Marslen-Wilson and Zwitserlood (1989) demonstrated that the decision space for the lexicality of words stands on their beginning. Their theory maintains that the speech input at the initial portion of a word maps onto all competing lexical items that phonologically share the same initial sequence, that is, words will compete for auditory recognition when they overlap in their initial structure. Hence, the new "no" pseudowords presented with a mismatch in relation to the "yes" pseudowords in either onset or coda positions. CLEARPOND[3] (Marian et al., 2012) was used to check for their lexicality status. For the "no" answers, the items[4] displayed in Table 1 below were used.

Table 1 - Items used for eliciting "no" answers in the Auditory Lexical Decision task

Target percept	Words used for "no" answers
i:	/miːʃ; liːb; diːv; biːb; miːg; giːm; kiːv/
ʌ	/gʌʃ; lʌp; sʌv; tʌv; gʌd; kʌg; gʌg; mʌd; pʌv/

Source: elaborated by the authors.

Each lexical item was presented twice in Auditory Lexical Decision task, which resulted in 76 trials. A trial consisted of the presentation of a fixation point, followed by an aural stimulus. The fixation point lasted for 5000ms, which was then followed by the presentation of the aural stimulus. The trial faded away as soon as the participant registered their response or was timed out 2000ms after the aural stimulus was presented. This time was deemed adequate as only 3,3% (*N*: 48; Total *N* of responses: 1,584) of responses for the "yes"

---

[3] https://clearpond.northwestern.edu/

[4] In order to test for their validity, a list containing both unidentified "yes" and "no" items was given to a Psycholinguist experienced with lexicality judgements, who was asked to spot words that stood out from the list for presenting any outstanding syllabic patterns.

words in this task were timed out during piloting. Moreover, presenting each item twice allowed for observing participants' reliability when taking the task.

To familiarize the participant with the procedure, this experiment consisted of a practice block, with four trials, each with a word from the study. In sequence, there were two different blocks, with 38 trials each (22 "yes" items plus 16 "no" items), presented in the automatic randomized order DMDX applies. All participants were allowed to take breaks between blocks and required to use their dominant right hand for the "yes" responses.

4.3 PARTICIPANTS AND DATA COLLECTION PROCEDURES

Thirty-six volunteering participants took part in the study. They were thirteen men and twenty-three women, whose ages varied from 18 to 47 (*M:* 26,1). All of them had normal or corrected-to-normal vision, and were right-handed. Moreover, participants received an Informed Consent Form that described in detail all requirements for taking part in the study, as well as their rights to anonymity.

As regards data collection, participants encountered individually with the experimenter in a quiet room, while sitting in a comfortable chair. The headset volume was adjusted to a comfortable listening level. A Microsoft LifeChat headset was used for auditory presentation and the recording of oral responses, and an Avell notebook was used to administer all the experiments. Firstly, participants were given the Consent Form and took part in training sessions. Next, participants were tested with the Auditory Lexical Decision. At the beginning of all encounters, the experimenter emphasized that answers should be given as quickly and as accurately as possible for when they were tested. The next section explains how the data analysis was conducted and presents the discussion of results.

5 RESULTS AND DISCUSSION

Spurious and wrong responses were excluded from the data spreadsheet. For items that required a positive response (words belonging to the trained lexicon), 197 wrong answers were excluded out of 1584 total, thus 12% of data were lost. For the negative answers, 34% of data were eliminated (546 data points). Missing values[5] were unchanged, and the data were analyzed with multi-level statistical models (Lachaud & Renaud, 2011). Descriptive statistics were run in SPSS, which yielded the results displayed in Table 2 below. Note that the results are displayed separately for consistent (e.g., "seeg") and inconsistent items (e.g., "toud") to observe whether participants' performance differed according to orthographic depth.

Table 2 - Descriptive statistics for reaction times in Auditory lexical decision

	M	*SD*	Min.	Max.
Consistent Items				
RT1	945ms	238ms	387ms	1956ms
RT2	930ms	218ms	522ms	1893ms

---

[5]Missing values are sometimes replaced with the item or subject mean, but this may artificially reduce the variability in the data set (Sanz & Grey, 2015).

	M	*SD*	Min.	Max.
	Inconsistent Items			
RT1	941ms	252ms	502ms	1953ms
RT2	943ms	251ms	421ms	1990ms

Source: elaborated by the authors.

As stated before, two reaction times were registered for each word in each level of orthographic transparency to test for participants' reliability when dealing with recently learned pseudowords. Considering Table 2, it can be observed that the means for both consistent and inconsistent items are similar. However, participants' responses varied more with inconsistent items, as the SD means were higher than with consistent items. Overall, participants made correct judgments for the positive responses 88% of time, whereas for the "no" items, participants scored 66% of correct answers. Participants timed out on only 2% of trials.

Normality tests (Kolmogorov-Smirnov and Shapiro-Wilk) indicated that the latencies did not achieve normal distribution (*p:* .000). Therefore, a Wilcoxon Signed rank test was run to observe whether there were significant statistical differences between reaction times 01 and 02 for the items conceived as "yes" answers, which were the lexicon learned by the participants in the training sessions. Results indicated that there were no significant differences between participants' answers from the first to the second reaction time, indicating that they performed consistently across different testing times with the learned items in this task (*Z:* -,954; *p:* .340).

To observe whether orthographic consistency affected subjects' performance with the "yes" items, a new variable was computed, which consisted of the mean value of the reaction times 01 and 02, as there was no significant difference across both these measures. Orthographic consistency was used as the grouping variable, and a Mann-Whitney U test demonstrated that orthography did not influence participants' performance with the items that were learned during training (*Z:* -,291; *p:* .771). Therefore, we entertain the hypothesis that orthography is simply not necessary for the lexical search conducted with familiar words, what also justifies why consistent and inconsistent items had similar reaction times for the "yes" answers in this task. Any orthographic activation caused by these items had been bypassed so that lexical access was not conducted with reference to written codes for speech perception.

It is also relevant to address whether reactions times differed significantly between "yes" and "no" items. To do so, variables containing the mean reaction times for each orthographic level of transparency were used and their descriptive statistics are displayed in Table 3. A Wilcoxon Signed Rank U test unveiled that participants' latencies were statistically different across "yes" and "no" items (*Z:* -,868; *p:* .000). This confirms that participants' performance with the lexical items learned throughout the training sessions differed from the lexical items introduced only in the task to elicit "no" answers, which signals participants' consistent performance with the trained lexicon. Table 3 below demonstrates that participants' scored relatively higher latencies for "no" items in both levels of orthographic depth, showing that opaque items took longer to be recognized auditorily.

Table 3 - Descriptive statistics for "yes" and "no" items according to orthographic consistency

	Consistent orthography			
	M	SD	Min.	Max.
Yes items	971ms	194ms	572ms	1543ms
No items	1059ms	169ms	324ms	1527ms
	Inconsistent orthography			
	M	SD	Min.	Max.
Yes items	937ms	192ms	462ms	1596ms
No items	1017ms	207ms	615ms	1764ms

Source: elaborated by the authors.

The next analysis revolved around whether orthographic consistency affected latencies for the "no" items with the Auditory Lexical Decision task. A Mann-Whitney U test, using orthographic consistency as the grouping variable, demonstrated that answers for the "no" items were affected by the level of the orthographic depth of those lexical items: *Z*: -3,026; *p:* .002. This result unveils two relevant caveats for orthographic recruitment when learning a new lexicon.

First, it shows that upon encountering auditorily with an unfamiliar item, participants recruited orthography, even though this was a task that consisted only of aural stimuli. Thus, orthography was active as one of the mechanisms that aid lexical analysis, at least, for incoming unfamiliar items in a lexical decision task. This demonstrates that orthography can be necessary to auditory tasks, when subjects are compelled to conduct auditory analyses of new lexical items, which leads us to the hypothesis that linguistic systems can act in an encapsulated manner according to task demands. Previous research has posited that systems of representation (i.e., phonological, orthographic, etc.) can work in encapsulated manner to execute lexical access (Damian & Bowers, 2009). In this vein, the presence of orthographic effects can be interpreted as evidence for orthographic recruitment to be a strategic process (Cutler & Davis, 2012; Cutler, Treiman, & van Ooijen; 2010; Taft, 2011; Yoncheva et al*.,* 2013) that renders a unique processing principle that is specific to initial stages of instructed language acquisition. As written input is referred to constantly in initial stages of instructed language acquisition, orthography develops onto a system that strategically supports other linguistic processes that involve lexical knowledge for diverse tasks.

Therefore, by contrasting the results for the trained ("yes") and untrained ("no") items, we can hypothesize that orthographic effects are prevalent in earlier stages of acquisition because orthography acts as an aid for the establishment of new lexical categories. The orthographic system might assist the mapping of the phonological input to their grapho-phonic correspondences, thus leading to the creation of a "visual" lexical representation, which studies have argued to be stronger in the adult lexicon (Veivo & Järvikivi, 2013). As for the absence of such an effect for the trained words, we entertain that subjects had already formed lexical categories for them, at least in their working memory system, rendering orthography unnecessary for the lexical decision task in this category of items.

6 FINAL REMARKS

The Auditory Lexical Decision task brought to light exciting findings. Orthographic consistency did not affect subjects' responses with the trained ("yes") items in the task, even though latencies were relatively higher with opaque items. However, orthography indeed influenced latencies registered for the untrained "no" items. We argued that upon encountering

with incoming unfamiliar auditory items, subjects recruited orthography as a mechanism that aids lexical analysis in the lexical decision task. In this vein, orthographic recruitment was conceived as a strategic process that supports lexical decision in auditory tasks. This evidences a relevant caveat for second language acquisition: learners are compelled to recruit orthography in initial stages of acquisition, as this system strategically supports processes of lexical analysis, while also exerting influence onto the integration of new lexical entries in the adult lexicon (Saletta, Goffman, & Brentari, 2015). Anecdotal evidence shows that adult learners expect orthographic information to be presented along with phonological information in instructed settings, as many claim that they are able to understand what they hear once they have been presented with its written form.

As concerns the study limitations, further studies could replicate the Auditory Lexical Decision task with a correction regarding the number of "yes" and "no" trials. This task made use of an unbalanced number, as trained items ("yes" responses) encompassed 24 target words, whereas words used for the "no" responses were only 16. Moreover, a control group tested only within the L1 would be able to indicate whether such an effect is language-specific or motivated by the idiosyncrasies of the two orthographic systems in contact (Brazilian-Portuguese and English). Notwithstanding, a revised statistical model could factor in participants' proficiency level in the L2 to entertain if it bears any effect to strategic orthographic activation as a mechanism to aid lexical analysis.

REFERENCES

Antoniou M., Ettlinger M., & Wong, P. C. M. (2016) Complexity, Training Paradigm Design, and the Contribution of Memory Subsystems to Grammar Learning. *PLoS ONE*, *11*(7).

Bartolotti, J. & Marian, V. (2017), Bilinguals' Existing Languages Benefit Vocabulary Learning in a Third Language. *Language Learning, 67*: 110-140.

Bassetti, B., Escudero, P., & Hayes-Harb, R. (**2015**). Second language phonology at the interface between acoustic and orthographic input. Introduction to special issue of *Applied Psycholinguistics 36*(1), 1-6.

Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychological Science, 14*(6), 592-597.

Cutler, A., & Davis, C. (2012). An orthographic effect in phoneme processing, and its limitations. Frontiers in Psychology, 3(18), 1-7.

Cutler, A., Treiman, R., & Van Ooijen, B. (2010). Strategic deployment of orthographic knowledge in phoneme detection. Language and Speech, 53(3), 307 -320.

Damian, M. F., & Bowers, J. S. (2009). Assessing the role of orthography in speech perception and production: Evidence from picture-word interference tasks. *European Journal of Cognitive Psychology, 21*, 581-598.

Escudero, P. (2011). Speech processing in bilingual and multilingual listeners: The interrelation between sound perception and word recognition. In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *The Handbook of Laboratory Phonology* (pp. 407-417). New York: Oxford University Press.

Escudero, P. (2015). Orthography plays a limited role when learning the phonological forms of new words: The case of Spanish and English learners of novel Dutch words. *Applied Psycholinguistics, 36*(1), 7-22.

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008): Novel L2 words and assymetric lexical access. *Journal of Phonetics, 36*(2), 345-360.

Escudero, P., & Wanrooij, K. (2010). The effect of L1 orthography on L2 vowel perception. *Language and Speech, 53*(3), 343-365.

Escudero, P.; Simon, E., & Mulak, K. (2014). Learning words in a new language: orthography doesn't always help. *Bilingualism, Language and Cognition, 17*(2), 384-395.

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35,* 116-124.

Goldinger*, S. D. (*1996*). Auditory *lexical decision*. Language and Cognitive Processes, 11*(6), 559–567.

Gonçalves, A. R. (2017). *The orthographic signature in second language speech acquisition and processing.* Unpublished doctoral dissertation. Federal University of Santa Catarina, Florianópolis, Brazil.

Katz, L., & Frost, S. (2001). Phonology constrains the internal orthographic representation. *Reading and Writing: An Interdisciplinary Journal, 14*(3/4), 297-332.

Kolinsky, R. (2015). How learning to read influences language and cognition. In A. Pollatsek & R. Treiman (Eds.), *The Oxford Handbook of Reading* (pp. 377-393). Oxford: Oxford University Press.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Science*, *97*(22), 11850-11857.

Lachaud C. M., Renaud O. (2011). A tutorial for analysing human reaction times: how to filter data, manage missing values, and choose a statistical model. *Applied Psycholinguistics,* 32, 389–416

Marian, V., Bartolotti, J., Chabal, S., Shook, A. (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE 7*(8): e43230.

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 576–585.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82* (3), B101–B111.

Pytlyk, C. Shared orthography: Do shared written symbols influence the perception of L2 sounds? *The Modern Language Journal, 95*(4), 541-557.

Qu, Q.; Damian, M. (2019). Orthographic effects in Mandarin spoken language production. *Memory and Cognition*, *47*, 326–334.

Rastle, K., McCormick, S., Bayliss, L. & Davis, C. J. (2011). Orthography influences the perception and production of speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(6), 1588-1594.

Reis, A., Castro-Caldas, A. (1997). Illiteracy: a cause for biased cognitive development. *Journal of the International Neuropsychological Society*, *3*, 444-450.

Saletta, M., Goffman, L., Brentari, D. (2015). Reading skill and exposure to orthography influence speech production. *Applied Psycholinguistics,* 1-24.

Sanz, C., & Grey, S. (2015). Effects of conditions on L2 development: moving beyond accuracy. In P. Rebuschat (Ed.), *Explicit and implicit learning of languages* (pp. 301-324). John Benjamins,

Simon, E., Chambless, D., & Alves, U. K. (2010). Understanding the role of orthography in the acquisition of a non-native vowel contrast. *Language Sciences, 32*(3), 380–394.

Taft, M. (2011). Orthographic influences when processing spoken pseudowords: Theoretical implications. *Frontiers in Psychology, 2*, 1-7.

Tamminen, J., *Davis*, M. H. & *Rastle*, K. (*2015*). From specific examples to general knowledge in language learning. *Cognitive Psychology, 79*, 1-39.

Taylor, J. S. H., Davis, M. H., & Rastle, K. (2017). Comparing and validating methods of reading instruction using behavioural and neural findings in an artificial orthography. *Journal of Experimental Psychology: General, 146*(6), 826-858.

Van Assche, E., Duyck, W., & Gollan, T.H. (2016). Linking recognition and production: Cross-modal transfer effects between picture naming and lexical decision during first and second language processing in bilinguals. *Journal of Memory and Language, 89,* 37-54.

Veivo, O., & Järvikivi, J. (2013). Proficiency modulates early orthographic and phonological processing in L2 spoken word recognition. *Bilingualism: Language and Cognition, 16*(4), 864-883.

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science, 10* (2), 159–164.

Yoncheva, Y. N., Maurer, U., Zevin, J. D., McCandliss, B. D. (2013). Effects of rhyme and spelling patterns on auditory word ERPs depend on selective attention to phonology. *Brain and language, 124*(3), 238-243.

Werker, J. F. (1995). Exploring developmental changes in cross-language speech perception. In L. Gleitman & M. Liberman (eds) *Language: An Invitation to Cognitive Science,* Vol 1 (2nd ed.), p. 87-106.

Werker, J. F. (2018). Perceptual beginnings to language acquisition. *Applied Psycholinguistics, 39*(4), 703-728.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental model of speech processing. *Language Learning and Development, 1* (2), 197–234.

Werker, J. F. & Gervain, J. (2013). Speech Perception: A Foundation for Language Acquisition. In P. Zelazo (Ed.), *The Oxford Handbook of Developmental Psychology* (pp. 909-925). Oxford UK: Oxford University Press

APPENDIX A

Table 1 - Words that encompass the trained lexicon

Targets			
	/i:/	/ʌ/	Distractors
Control	geesh / keet / seeg	bup / nup/ sud	galm palb malp balsh
Experimental	deit / geib / meip	doup / soug / toud	
	geop / teog/ teob	dood/ pood / loob	

Source: elaborated by the authors.