Morphological decomposition in the processing of Portuguese words and pseudowords:The role of suffix salience and numerosity

Mariana Velho

UMinho | 2019

Minho University
School of Psychlogy

Mariana Dantas Oliveira Ribeiro Velho

Morphological decomposition in the processing of Portuguese words and pseudowords: The role of suffix salience and numerosity.

october 2019

Minho University
School of Psychology

Mariana Dantas Oliveira Ribeiro Velho

Morphological decomposition in the processing of Portuguese words and pseudowords: The role of suffix salience and numerosity.

Master Degree Dissertation
Integrated Master in Psychology

Work done under orientation of
PhD Helena Oliveira
&
PhD Ana Paula Soares

october 2019

Agradecimentos

À minha família por me permitirem sempre seguir o meu caminho, sem pressões, sem questionar apenas com o maior apoio e amor.

Ao Leo, por ser o meu refúgio na zona Oeste. Por me apoiar e ter os melhores abraços, sem ti nunca tinha levado isto até ao fim. Que me continues a chamar de rato de biblioteca e a gozar com o goal de ter 3 mestrados antes dos 30.

À Carolina, por berrar, rir e chorar comigo. Por ser a melhor companheira de casa e acima de tudo uma grande companheira na vida. Que venham as próximas 6 semanas de voluntariado na Índia, depois de dormir com ratos isto foi easy peasy lemon squeezy.

À Nicole, melhor companheira nesta aventura que foi o MIPSI. Que venham muitos estágios.

À Helena e à Professora Ana Paula por todo o apoio e paciência. Sem vocês o 5º ano tinha ficado por fazer ao fim de 2 meses, se hoje estou a terminá-lo é graças ao vosso apoio incansável. Que encontre sempre profissionais como vocês ao longo do meu percurso.

Ao que não está presente, mas é em grande parte responsável pelo que sou hoje, a ti Faial.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

Decomposição morfológica no processamento de palavras e pseudopalavras do português: O papel da saliência e numerosidade do sufixo

Mariana Velho

Helena Oliveira

Ana Paula Soares

Estudos realizados em diferentes línguas e recorrendo a diferentes tarefas e paradigmas, sugerem que as palavras complexas são decompostas nos seus constituintes morfológicos em etapas iniciais do reconhecimento visual de palavras. Contudo, a falta de controlo das caraterísticas das palavras usadas nesses estudos, nomeadamente no que se refere à frequência de uso dos constituintes (bases e sufixos) torna problemática a generalização desses resultados. Neste estudo procurámos testar o papel da numerosidade do sufixo (i.e., quantas palavras terminam com um dado conjunto de letras – e.g., -ice, -ite) e da saliência do sufixo (i.e., em quantas dessas palavras esse conjunto de letras é sufixo) no reconhecimento visual de palavras (e.g.., aldrabice) e pseudopalavras (e.g.., calvite, calvaca) do português, apresentadas a 48 nativos deste idioma numa tarefa de decisão lexical. As pseudopalavras foram construídas a partir da combinação de bases reais de palavras portuguesas (e.g.., calv[o]) com sufixos reais de palavras portuguesas (e.g., -ice, -ite) que apresentam elevada ou baixa numerosidade e elevada ou baixa saliência, dando origem a quatro condições experimentais (e.g., para a base alert[a]: alertaco; alertebre; alertite; alertenta). Foram usadas também palavras com sufixos de elevada numerosidade que apresentam elevada ou baixa saliência (e.g., *nojice*, *rebeldia*), em duas condições experimentais. Os resultados revelaram tempos de resposta mais elevados e maior percentagem de erro na rejeição de pseudopalavras terminadas por conjuntos de letras mais frequentes no português, como esperado. No entanto, de forma contrária às predições, a saliência do sufixo, facilitou o reconhecimento de pseudopalavras. Relativamente às palavras, os participantes foram mais rápidos e precisos a reconhecer palavras com sufixos de elevada saliência do que palavras com sufixos de baixa saliência, como hipotetizado. Estes resultados são discutidos tendo em consideração os atuais modelos do processamento morfológico.

Palavras-chave: Morfologia; Sufixo; Saliência do Sufixo; Numerosidade do Sufixo; Reconhecimento Visual de Palavras.

Morphological decomposition in the processing of Portuguese words and pseudowords: The role of suffix salience and numerosity

Mariana Velho

Helena Oliveira

Ana Paula Soares

Studies conducted in different languages using different tasks and techniques have provided evidence for morphological decomposition at early stage of visual word recognition of complex words. Several issues remain, however, open due mainly to the lack of control of the characteristics of the morphological constituents of those words (e.g., frequency of the roots and suffixes) which could have affected the results. In this study we present a lexical decision experiment performed by 48 undergraduate students from Minho University aimed to examine the role of the suffix-numerosity (i.e., how many words ends with a given letter string - e.g., ice, ite) and suffix-salience (i.e., in how many words that letter string [e.g., ice, ite] is a suffix) in the visual word recognition of Portuguese words (e.g., aldrabice) and pseudowords (e.g., calvite, calvaca). Pseudowords were made up of real roots (e.g., calv[o]) combined with high- and low-numerosity suffixes and high- and low-salience suffixes, giving rise to four experimental conditions (e.g., for the root alert[a]: alertaco; alertebre; alertite; alertenta). Ninety-six words made up of high-numerosity suffixes with high- and low-salience suffixes were also used (e.g., *nojice*, *rebeldia*, respectively), giving rise to two experimental conditions. Results from repeated measures ANOVAs revealed longer response times and more errors for pseudowords composed by letter strings that frequently are real endings in Portuguese words, as expected. However, conversely to the predictions, pseudowords made up of letter strings that most of the times constitute real suffixes in Portuguese words, facilitated (and not hindered) pseudoword processing. For words, participants were faster and more accurate when recognizing words with high- than low-salient suffixes, as expected. Results were discussed attending to the current models of morphological processing.

Keywords: Morphology; Suffix; Visual Word Recognition; Suffix-Salience; Suffix-Numerosity.

Index

Introduction........8

Method........15

Participants........15

Materials........16

Procedure........18

Results........19

Discussion........24

References........27

Tables Index

Table 1........16

Table 2........18

Table 3........19

Figures Index

Figure 1........21

Figure 2........21

Figure 3........23

Introduction

Most words on a given language, including Portuguese, are complex words. However, the vast majority of studies conducted on Psycholinguistics, not only in Portuguese but across languages, use simple words. Complex words have been defined as words that have more than one morpheme and that are commonly formed by a root and, at least, one prefix (e.g., *in* + *feliz* = *infeliz* [unhappy]) or one derivational suffix (e.g., *feliz* + *mente* = *felizmente* [happily]) (Rio-Torto, Rodrigues, Pereira, Pereira, & Ribeiro 2016). The accumulated knowledge about the processing of simple words might not apply directly to the processing of complex words, especially because there is evidence that morphology plays an important role in the way individuals process and recognize words (e.g., Taft & Forster, 1975; Rastle, Davis, & New 2004). For instance, Taft and Forster (1975) in one of the first studies aimed to test whether there is a morphological decomposition at early stages of visual word recognition, used English complex pseudowords in three lexical decision task (LDT) experiments. In the first experiment, the authors defined two study conditions: one, called a real stem condition, where the words from which they generated the pseudowords met both of the following requirements: they were listed as being derived from a prefix plus stem, and this prefix contributed to the meaning of the word (e.g., re- of *rejuvenate*, being *juvenate* classified as a real stem); and the other condition, called a pseudo stem condition where derived words – which began with the same letters as a real prefix – were not listed in the dictionary as being derived from a prefix plus stem (e.g., regulate, undulate), or, if they were, the prefix no longer contributed to the meaning of the word (e.g., *devout*, *rebel*, being so -vout or -bel classified of pseudo-stems). The rationale was that a real stem would take longer to be classified as a nonword when compared to the pseudostems, since real steams should be mapped in the lexicon. As expected, they found longer response times and higher error rates for real-stem pseudowords than for pseudo-stems pseudowords, hence suggesting that real stems were perceived as being more word-like than pseudostems, which explained the interference effect observed. In a second experiment the authors tried to explore the fact that there are stems of derived complex words (i.e., bound forms – BF) that are, at the same time, real words, i.e., free forms (FF). For instance, *vent* is a free morpheme (i.e., an outlet for air) but it is also a bound morpheme as in *prevent* or *advent*, and since these two forms have quite a different function, they can be considered as two separate entries in the lexicon. So, in Experiment 2 they had two study conditions: one where the free form (FF) of the word had a higher frequency than the bound form (BF) of the same word and another where the bound form had higher frequency than the free form. They expected that morphemes with a higher free form frequency than a bound form frequency (BF > FF) would reveal less interference than the ones with higher bound form frequency than a free form frequency (BF < FF). Results showed that participants were faster when responding to the condition where the free form had higher frequency than the bound form (BF < FF) when compared to the ones with higher bound form frequency than free form frequency (BF > FF) meaning that there is interference when the frequency of the bound form is higher than the free form frequency, leading to the idea that stems of complex words are stored as lexical items. Finally, on Experiment 3, the procedure was similar to Experiment 2 (with a new set of items selected using the same criteria as in Experiment 1) but inappropriate prefixes were added, i.e., letter strings that are not prefixes (e.g., *de-*), to the real stems (e.g., *dejuvenate*) and to pseudo stems (e.g., *depertoire*). The idea was that the prefix was identified and removed, and a search begins for the entry *juvenate*, when this happens the contents of the entry are examined to see whether *de-* is a legitimate prefix. When it is found that it is not, there must be an additional search to check if *dejuvenate* is listed as a whole word, probably leading to higher interference and higher reaction times since the stem is real. However, in the pseudo stem condition (e.g., *depertoire*), the prefix *de-* is also removed, and a search begins for *pertoire*. No entry is found, and after an additional search *depertoire* is checked as not listed and the item is established as being a nonword since the stem is a pseudo stem. As expected, results showed that prefixed real stems revealed higher reaction times than prefixed pseudo stems meaning that the higher reaction times on Experiment 1 cannot be due to the uncertainty of the correct response. Based on the results of the three experiments, the authors proposed that individuals recognize complex words by parsing and analysing their morphemic subunits before accessing the mental lexicon, in a process called morphological decomposition that assumes that an early morpho-orthographic process decomposes words into bases and suffixes in a blind way.

Since this pioneering study, many authors have studied morphological processing both with words and pseudowords. For instance, Rastle, Davis, and New (2004), used a masked priming paradigm combined with a LDT presented to English native speakers with three types of word pairs: (i) transparent pairs, where there was a semantically transparent morphological relationship between primes and targets (e.g., cleaner – CLEAN); (ii) opaque pairs, where the morphological relationship between primes and targets was only apparent (e.g., corner – CORN); and (iii) unrelated pairs, where there was no morphological relationship between primes and targets (e.g., brothel – BROTH). Primes were presented for 42ms. Significant masked priming effects were observed both for the transparent and opaque pairs – cases in which primes and targets were, or appeared to be, morphologically related – and priming effects in these conditions differentiated significantly from unrelated pairs. However, as no difference was observed between transparent and opaque conditions, the authors claimed this effect resulted from a morpho-orthographic representation. Apparently, these results pointed to a morphological segmentation process that takes place at early stages of visual recognition and that operates on any printed word that contains a stem and an affix, irrespective of semantic transparency. This notion of an obligatory morphological decomposition (blind to semantic and etymological factors) supports Taft and Forster (1975) morphological decomposition statement that a morphological analysis of words is attempted prior to lexical search and that individuals recognize complex words by parsing and analysing their morphemic subunits before accessing the mental lexicon (i.e., fam[e] + ous). This model has also been corroborated more recently by Beyersmann, Coltheart, and Castles (2012) in a study, that also used the masked priming paradigm, where the results revealed significant facilitation from both truly suffixed (transparent pairs) and pseudosuffixed primes (opaque pairs) showing that priming occurred both when there was and there was not a semantically transparent relationship between the prime and the target. The significant priming in the pseudosuffixed condition added evidence to the Taft and Forster (1975) idea that there is a semantically blind morphological segmentation mechanism that decomposes any letter string with the mere appearance of being morphologically complex.

Although previous studies (e.g., Diependaele, Sandra, & Grainger, 2009; Rastle et al. 2004) have shown that the magnitude of morphological priming effects are modulated by the degree of semantic transparency between primes and targets, Feldman, O'Connor, and Martín (2009) further explored whether morpho-orthographic parsability was sufficient to account for the observed patterns of morphological facilitation in these studies matching affixes across semantically transparent, opaque related, and unrelated prime-target pairs in length, frequency, orthographic neighbourhood size, and phonological neighbourhood size. Using a masked priming paradigm combined with a LDT, the authors found that morphological facilitation was significantly greater for the semantically transparent pairs than for the opaque pairs, hence questioning the conclusions of Rastle et al. (2004). In order to explain this inconsistency, Feldman et al. (2009) pointed towards differences between both experiments regarding the stimuli used. For instance, Rastle et al.'s (2004) stimuli, had larger orthographic neighbourhoods (11.82 vs. 2.18), and had a difference in morphological family sizes (transparent targets: 4.46 vs. 3.38; opaque targets: 2.59 vs. 2.40), Feldman et al. (2009) targets were also slightly shorter (4.04 vs. 4.86). Anyway, these findings call into question the interactions between meaning (semantics) and form (orthographic/phonological) representations during early stages of complex word recognition and the autonomy of morpho-orthographic from morpho-semantic processing supporting the hybrid morphological processing model presented by Diependaele, Sandra, and Grainger (2005). This model states that there are two distinct mechanisms involved in morphological processing: a purely form-based mechanism that activates the base whenever the visual input is fully decomposable into a base and suffix (a morpho-orthographic system), and a mechanism that activates the base once the whole word has been activated whenever the visual input is fully decomposable and shares semantic features with its base (a morpho-semantic system). They also claimed that these mechanisms operate with different speeds, more specifically, the morpho-semantic system activates the root faster than the morpho-orthographic system does. Supporting this model there is the evidence obtained by Diependaele et al. (2005) which examined priming from semantically transparent and opaque suffix-derivations in French words. Three different prime exposure times (i.e., 13ms, 40ms and 67ms) were used in a cross-modal priming technique (visual primes and visual and auditory targets). Specifically, the authors presented participants with words and nonwords as targets. All word targets were free roots and were divided in three sets, each set corresponded to one of the three priming conditions. These three priming conditions were equal for the nonwords: They comprised ‘‘transparent’’, ‘‘opaque’’, and ‘‘orthographic’’ priming conditions, each consisting of 48 targets and 48 primes per condition. In the ‘‘transparent’’ condition related primes were suffix-derived forms and targets were formed by substituting one or more letters of the root (e.g., totalité – TUTAL from totalité – TOTAL). In the ‘‘opaque’’ condition targets were constructed in an identical way, but here the primes were opaque derivations (e.g., aversion – IVERSE from aversion – AVERSE). A comparison between related and unrelated primes across conditions in the three exposure times revealed for the visual condition that target processing was already facilitated – that is, lower reaction times – by transparent derivations primes at 40ms prime exposure. However, at 67ms both transparent and opaque derivations showed a robust facilitation and transparent primes caused a larger facilitative effect than opaque primes. For the auditory targets, significant facilitation effects emerged only at 67ms, both for transparent and opaque primes – and without significant differences between these two prime types. Orthographic primes at 67ms produced opposite priming effects, facilitated visual target processing, but had an inhibition effect – higher reaction times – on auditory target processing. The facilitation effect from transparent derivations emerged earlier and was systematically larger than the effect of opaque derivations when the root targets were presented visually. This effect cannot be accounted for by the sublexical account of morphological processing, since it predicts that both transparent and opaque words are treated equally at early stages of processing. These results support a view in which morphological processing is governed by a system that is both sensitive to the formal (morpho-orthographic) properties of a complex word and to its semantic (morpho-semantic) properties.

One way to control for the role of semantic properties in morphological decomposition, is by using pseudowords, as, not having meaning, their processing is more based on formal aspects. Thus, pseudowords are particularly suited to study mechanisms involved on early pre-lexical morphological decomposition stages of processing since the whole-letter string cannot be successfully mapped onto an existing representation in the orthographic lexicon. Using pseudowords, Burani, Dovetto, Thornton, and Laudanna, (1997) aimed to explore the role of suffix frequency on morphological processing. The authors used Italian pseudowords made up of illegal root-suffix combinations (pseudo roots combined with derivational suffixes, e.g., *vend[ario]*). On Experiment 1, they used suffixes of two distinct frequency ranges, dividing the experiment in two different conditions: (ii) a condition where roots were combined with high-frequency suffixes (e.g., *piovario*), and the resulting pseudowords were contrasted with pseudowords in which the same roots were combined with control sequences that had analogous orthographic frequency in final position of Italian words (e.g., *piovalia*), but were not suffixes (high-frequency suffix condition); and (ii) a condition where comparison was made between pseudowords composed of root plus low-frequency suffixes (e.g., *perdigia*), and the same roots were combined with control low-frequency orthographic final sequences, e.g. *perdegio*, (low-frequency suffix condition). Roots in both sets were of medium frequency. Results revealed that the interference effect usually found on pseudowords that include real affixes was modulated by the frequency of the suffixes; that is, the authors found longer reaction times and higher error rates when pseudowords included high-frequency suffixes relative to control pseudowords. By contrast, pseudowords made up of low-frequency suffixes took no longer to be rejected than control pseudowords. These results led Burani et al. (1997) to conclude that the probability that suffixes will affect processing is modulated by their frequency, which can be accounted by the Augmented Addressed Morphology model by Caramazza, Laudanna, and Romani (1988). This model states that some complex words may be recognized through decomposition, but not others. According to this model, a letter string activates both whole-word representations (when available – that is, for known words) as well as the morphemes that comprise the word. Thus, the stimulus 'walked' will activate the access representations 'walked', 'walk-', and '-ed', as well as orthographically similar representations such as 'walks', 'walking', 'talked', 'talk', 'walked', 'winked', etc. The orthographic representation that first reaches a pre-set threshold will activate the corresponding lexical entry. They also state that lexical decision times are affected by the cumulative (root or stem) word frequency and stem-morpheme frequency of morphologically related words meaning that the more frequent an orthographic representation is the less time is needed to reach the pre-set threshold to be recognized. So, as this model explains that high-frequency words are lexicalised and maintained in our lexicon, the same might apply for suffixes, being the more frequently used suffixes recorded in our lexicon, and thus being more easily recognized, hence explaining why only the high-frequency suffixed pseudowords revealed higher response times than their respective controls.

To further explore whether the frequency of the root, of the suffix, or of the whole-word affect morphological processing of Italian derived words and pseudowords, Burani and Thornton (2003), conducted three experiments: one with pseudowords and the other two with words. On the first one, they aimed to replicate the effect of suffix frequency found by Burani et al. (1997) with Italian derivational suffixes of different frequency ranges – high, medium, and low. However, conversely to Burani et al. (1997) pseudowords, Burani and Thornton (2003) pseudowords were made up with legal orthographic sequences that did not correspond to real roots (e.g., *pruc[ezza]*). Each suffixed pseudoword was matched with a control pseudoword including the same pseudoroot in combination with an orthographic sequence that constituted the control for the suffix (e.g., *pruc[ondo]*). Results revealed longer latencies and more errors when pseudowords included a high-frequency suffix, hence showing that high-frequency suffixes create an interference effect since high-frequency suffixes are all represented on the lexicon and need to compete between them to reach a recognition threshold while the low-frequency ones are immediately reached since they are not represented on the lexicon or have fewer competitors as predicted by the Augmented Addressed Morphology model by Caramazza, Laudanna, and Romani (1988). For medium-frequency and low-frequency suffixes, the results were not significantly different from controls. In Experiments 2 and 3, the role of suffix frequency was further explored by varying both root and suffix frequency in transparent Italian derived words of low surface frequency. Derived words included suffixes belonging to two sets of different frequencies (suffixes of high-frequency were contrasted with suffixes that were of medium/low frequency – defined as low-frequency suffixes). There were four sets of materials: high-frequency roots with high-frequency suffixes (HH: e.g., *bassezza*), low-frequency roots with high-frequency suffixes (LH: e.g., *astrale*), high-frequency roots with low-frequency suffixes (HL: e.g., *testardo*), and low-frequency roots with low-frequency suffixes (LL: e.g., *serpentesco*). Suffixes were either high- or low-frequency on both token and type measures. Experiment 2 revealed that reaction times and accuracy rates were affected by the frequency of both roots and suffixes as participants were faster and more accurate when derived words included high-frequency constituents. On Experiment 3 a set of low-frequency nonderived words was added to the dataset to address the fact that for low-frequency derived words whose constituents are both low frequent. Thus, the moderate difference between the frequency of morphemes and whole-word frequency (with root and suffix only slightly higher in frequency than the whole-word) might not be large enough for morphological processing to result in benefits relative to access based on the whole-word. Results revealed that only low-frequency derived words with high-frequency roots showed faster processing relative to nonderived words of similar frequency. Thus, lexical access through activation of morphemes seems to be beneficial only for derived words with high-frequency roots, meaning that access to low-frequency derived words is not always obtained via morphological parsing but via the full-form access route instead. To conclude, Burani and Thornton (2003) stated that frequency was the major determinant of the relative probability that lexical access is based either on the whole-word route or on the morpheme route. The higher the frequency of a given lexical unit (word as a whole, root or affix) the greater the likelihood that this unit is quickly activated and processed in the different constituting morphemes. Hence, the balance between the frequency of the word as a whole (word lexical frequency) and the frequency of its constituent morphemes (both roots and affixes) is critical to determine which route will be most activated during visual word recognition of complex words. Consequently, it might be predicted from this account that a transparent derived word which has low-frequency in the language but is composed of a very frequent root and a very frequent suffix is likely to be accessed via activation of its morphemic constituents, rather than via the unit corresponding to the whole-word representation. This prediction implies that the frequencies of both the root and the suffix can affect morphological processing and ask for evidence concerning the role that these variables play at early stages of visual word recognition of complex words.

Despite the fact that many studies, using different tasks, procedures, and techniques have provided evidence for morphological decomposition at early stages of visual word recognition of complex words in different languages (e.g., Crepaldi, Rastle, Coltheart, & Nickels, 2010; Rastle, Davis, & New, 2004; Taft & Forster, 1975), several questions remain under debate. For instance, it is not yet clear how does one process and recognise a morphological complex word nor the variables that underlie the speed/accuracy with which we recognize a given complex word. The work presented in this thesis aimed to directly address these issues by examining if the number of times a certain letter string ends Portuguese words (a variable called suffix numerosity) affects the speed and/or accuracy with which Portuguese complex derived words and pseudowords were recognized, both when that sequences work as real suffix in most of the times or not (a variable called suffix salience salience). Ultimately, we aimed to test if morphological effects on visual word recognition of derived complex words are mainly due to morphological or to orthographic factors. As stated before, we also aimed to further explore the role of suffix-numerosity (i.e., how many words ends with a given letter string in a language) and suffix-salience (i.e., in how many words that letter string is a suffix in the visual word recognition of Portuguese derived complex words (e.g., *aldrabice* [swindle]) and pseudowords (e.g. *calvite*, *calvaca*) made up of real roots (e.g., *calv*[o] [bald]) combined with high- and low-numerosity and high- and low-salience suffixes (e.g., -*ite*, -*aca*) in four experimental conditions. Considering that, as on previous studies, suffix-frequency and suffix-numerosity might have been confounded, according to Burani and Thornton (2003), suffix-numerosity can be a better quantitative characterization for suffixes and a stronger predictor of performance in access tasks since it is closely related to suffix productivity allowing the suffix to "emerge" as a separate processing unit. So, in the present study, we will be considering suffix-numerosity instead of suffix-frequency since it can be a better predictor of performance in access tasks. It is also important to highlight that none of the above mentioned studies, manipulated and studied the suffix-salience, being an interesting breakthrough because it will allow us to detangle the contribution of purely orthographic vs. morphological aspects might have on the Portuguese morphological processing of suffixed pseudowords and words. So, combining the two selected variables (suffix-numerosity and suffix-salience) in this study we present a lexical decision experiment aimed to examine the role of the suffix-numerosity (i.e., how many words ends with a given letter string) and suffix-salience (i.e., in how many words that letter string is a suffix) in the visual word recognition of Portuguese words and pseudowords. Pseudowords were made up of real roots combined with high- and low-numerosity suffixes and high- and low-salience suffixes, given rise to four experimental conditions. Words were also studied but only suffix-salience was manipulated being suffix-numerosity always high (due to the lack of words per suffix – on the low-numerosity suffix condition – low-numerosity suffixes were excluded from the study on the words condition) given rise to two experimental conditions.

Based on the reviewed literature, we expect to find that pseudowords with high-numerosity and high-salience suffixes would produce an interference effect as revealed by longer response times and more errors than the other three conditions. If the morphological decomposition is due to morphological effects, pseudowords with high-salience suffixes will suffer higher interference, revealing slower responses and less accurate responses, regardless of the suffix-numerosity; but, if the morphological decomposition is an effect of the frequency of occurrence of that letter string, pseudowords with high-numerosity suffixes would reveal slower and less accurate responses regardless of the suffix-salience. Regarding the processing of complex derived words, we expect that, due to being all of high-numerosity, high-salience ones would produce faster and more accurate responses than the ones with low-salience.

Method

Participants

A total of 48 undergraduates from Minho University (34 female, M_{age} = 20.2, SD = 2.72) participated in the study. All participants had Portuguese (European variant) as their native language and had normal or corrected-to-normal vision. None of the participants reported having learning or reading disabilities. Written informed consent was obtained from all the participants. The experiment was approved by the local Ethics Committee for Human Research (SECSH 003-2014) and were carried out in accordance with the code of ethics of the World Medical Association (Declaration of Helsinki). Participants received course credits in exchange for their participation accordingly to the accreditation system of the School of Psychology from the Minho University.

Materials

Sixteen suffixes that give rise to Portuguese nouns or adjectives, accordingly to the *Gramática Derivacional do Português* (Rio-Torto et al., 2016), were firstly selected. Eight of them occurred as a word-ending in a high number of Portuguese words (i.e., they were high-numerosity suffixes, *M* = 346, *SD* = 113, Range = 202 – 599), whereas the other eight occurred as a word-ending in a low number of the Portuguese words (low-numerosity suffixes, *M* = 67, *SD* = 16, Range = 13 – 94), as computed from the Procura-PALAvras (P-PAL) lexical database (Soares et al., 2018). Furthermore, half of the high-numerosity suffixes occurred as real suffix in the most (>66%) of the Portuguese words ending with that letter string (i.e., they were high-salient suffixes, e.g., -*edo*), while in the other half they worked as real suffix in less than 33% of the Portuguese words ending with that letter string (i.e., they were low-salient suffixes, e.g., -*ebre*). The same distribution was implemented for the low-numerosity suffixes (see Table 1).

Table 1

Selected suffixes distributed by condition.

	High-Numerosity (N)	Low-Numerosity (N)
High-Salience	-*ite* (319)	-*aco* (78)
	-*ola* (205)	-*edo* (52)
	-*ura* (469)	-*uda* (81)
	-*ense* (394)	-*esca* (53)
Low-Salience	-*enta* (202)	-*ebre* (13)
	-*eia* (342)	-*iça* (79)
	-*ona* (240)	-*aca* (86)
	-*ina* (599)	-*eno* (94)

Note. N = word frequency (number of words, in Portuguese, ending with that letter string).

Secondly, we selected 96 Portuguese simple words from the P-PAL lexical database (Soares et al., 2018) with four to six letters in length, and with a word frequency (per million words) ranging between

.02 and 186.91 (*M* = 16.23, *SD* = 31.49). These simple words were used as word roots to combine with the suffixes and to create the pseudowords. So, the 96 word roots were combined with all the 16 suffixes giving rise to the 1,536 pseudowords used in the experiment. It is worth noting that, although all the roots and suffixes exist in Portuguese, their combination does not, hence generating in each of the cases Portuguese pseudowords. These pseudowords were assigned to 16 experimental lists (96 pseudowords per list) to assure that each word root appeared only once per list but combined with one of the sixteen suffixes in different lists (e.g., the root *calv-* was combined with the suffix -*ola* on List 1 [*calvola*] and with the suffix -*ina* on List 7 [*calvina*]; the root *brev-* was combined with the suffix *-ebre* on List 5 [*brevebre*] and with the suffix -*aca* on List 7 [*brevaca*]). In each list, a quarter of the roots (*n* = 24) were assigned to one of each of the four experimental suffix conditions: high-numerosity and high-salient suffixes (N^+S^+), high-numerosity and low-salient suffixes (N^+S^-), low-numerosity and high-salient suffixes (N^-S^+) and low-numerosity and low-salient suffixes (N^-S^-). Since all the roots and suffixes underwent all the experimental conditions, the psycholinguistics characteristics that might affect processing (e.g., length, frequency, neighbourhood) were controlled for in the pseudoword stimuli.

Finally, 96 Portuguese complex words were also selected from the P-PAL database (Soares et al., 2018). They were drawn from other four high-numerosity suffixes than the ones used for pseudoword generation. Note that, for words, suffix-numerosity was not manipulated because, due to the strict control used to calculate suffix-numerosity, we were left with very few real complex words on the low-numerosity condition, making it impossible to have balanced study conditions. Half of stimuli in words had high-salient suffixes (*M* = 499, *SD*= 227, Range = 246 – 684), while the other half were low-salient (*M* = 370, *SD* = 197, Range = 240 – 597), following the same criterion as for pseudowords. Across these two conditions, words were matched on several variables known to affect the Portuguese word processing process, as word length, word frequency (per million words), number of orthographic neighbours (ON) and orthographic levenshteing distance (OLD20) (see Soares et al., 2019 for details) as obtained from P-PAL database (Soares et al., 2018). Table 2 presents the psycholinguistic characteristics (means and standard deviations) of the words used as stimuli.

Table 2

Psycholinguistic characteristics of the N+S+ and N+S- words as obtained from the P-PAL lexical database (Soares et al., 2018)

	S^+ *M* (*SD*)	S^- *M* (*SD*)	*t* – tests
Word frequency (per million words)	1.23 (2.49)	0.70 (1.56)	*p* = .22
Length (number of letters)	8.60 (1.33)	8.29 (1.18)	*p* = .11
Number of orthographic neighbours (ON)	8.60 (1.33)	8.29 (1.18)	*p* = .11
Orthographic Levenshtein Distance (OLD20)	2.55 (0.47)	2.42 (0.49)	*p* = .19

Note. S^+ = high-salience; S^- = low-salience; M = mean; SD = standard deviation; p>.05 = statistically non-significant.

All the words were presented in each of the 16 lists that contained 192 experimental stimuli (96 pseudowords + 96 words) each. A set of four words and four pseudowords with the same characteristics as the experimental stimuli were also selected for practice trials.

Procedure

Participants were tested individually in sound-proofs cabins at the Human Cognition Laboratory of the School of Psychology at the University of Minho. They were asked to decide as soon and accurately as possible if a given letter string presented at the centre of a computer screen was or not a real Portuguese word. They were instructed to press the <Z> key from the keyboard for nonwords and the <M> key for words. To familiarize participants with the task, previous to the experiment they received eight practice trials, followed by the 192 experimental trials. Breaks were given after practice, and, then after half the stimuli were presented. Each trial consisted on the presentation of two subsequent events: a fixation point (+), presented at the centre of the monitor for 500ms, followed by the target (word/pseudoword), which remained on the screen until a response was made or up to 2,500 ms had elapsed. All stimuli were presented in lowercase letter in 14–pt. Courier New font against a white background. The order in which the items were presented was randomized per participant. Each participant was randomly assigned to a list, though assuring the same number of participants per list.

The DMDX software (Forster & Forster, 2003) was used to present the stimuli and collect participants' responses (reaction times and accuracy). The procedure took about 15 minutes per participant to be completed.

Results

Analyses were conducted separately for pseudowords and words. For pseudowords, repeated analyses of variance (ANOVA) on participants' (*F*1) and items' (*F*2) data were conducted for reaction times (RTs) and percentage of errors (%E) based on a 2 (suffix numerosity: high or low) x 2 (suffix-salience: high or low) x 16 (list: 1 to 16) mixed design. In the *F*1 analyses, the two first variables entered as within subject-factors in the ANOVA, while the last entered as a between-factor to remove the error of variance due to the existence of 16 counterbalancing lists (see Pollatsek, 1995). In the *F*2 analyses, suffix-numerosity, suffix-salience and list entered as within subject-factors in the ANOVA. Prior to ANOVAs, data were cleaned based on the following trimming procedures: firstly, incorrect responses and RTs below 300 ms and above 2,000 ms were excluded from the RT analyses. Besides, a suffix per condition – the one with the higher error rate in each condition – were also excluded, which left us with three suffixes per condition. Finally, RTs more than 3 standard deviations (*SD*s) above or below each participants' mean per condition were also removed. From all these procedures, 14.21% of the raw RT data was excluded.

For words, *t*-test for paired samples comparing RTs and %E for high vs low-salience suffixes were conducted both on participants' (*t*1) and items' (*t*2) data. Table 3 presents the latency (in ms) and accuracy results (means and standard deviations) obtained by participants for pseudowords and words per experimental condition.

Table 3

Mean and SDs (in parentheses) of response times (RTs) and percentage of errors (%E) by experimental condition.

Target	Suffix conditions	RTs	%E
Pseudowords	N^+S^+	1117.3 (248.9)	11.3 (13.7)
	N^+S^-	1166.4 (246.2)	12.3 (10.8)
	N^-S^+	1109.5 (233.7)	10.8 (11.2)
	N^-S^-	1116.1 (260.6)	7.3 (10.4)
Words	N^+S^+	859.8 (161.0)	11.1 (5.2)
	N^+S^-	938.7 (176.4)	18.0 (10.9)

Note. N^+ = high-numerosity; N^- = low-numerosity; S^+ = high-salience; S^- = low-salience; RTs = reaction times (ms); %E = percentage of errors.

Pseudowords

ANOVAs for pseudowords revealed a significant main effect of suffix-numerosity, $F_1(1,32) = 6.125$, $p = .018$, $\eta_p^2 = .163$; $F_2(1,1088) = 8.293$, $p = .004$, $\eta_p^2 = .008$, showing that participants were faster when rejecting pseudowords from the low-numerosity condition ($M = 1112.8$, $SD = 246.2$) than from the high-numerosity condition ($M = 1141.9$, $SD = 247.5$), regardless of suffix salience. Moreover, the results also revealed a significant main effect of suffix-salience on participants' data, $F_1(1,32) = 8.180$, $p = .007$, $\eta_p^2 = .204$; $F_2(1,1088) = 1.706$, $p = .192$, $\eta_p^2 = .002$, showing that participants were faster when rejecting pseudowords from the high-salience condition ($M = 1113.4$, $SD = 240.2$) than from the low-salience condition ($M = 1141.2$, $SD = 253.4$). The interaction between the two factors failed to reach statistical significance in the by-participant's and by-item's analysis, $F_1(1,32) = 2.056$, $p = .161$, $\eta_p^2 = .060$; $F_2(1,1088) = 2.982$, $p = .084$, $\eta_p^2 = .003$.

On the %E data, the ANOVAs revealed a main effect of suffix-numerosity on participants' data, $F_1(1,32) = 7.481$, $p = .010$, $\eta_p^2 = .189$; $F_2(1,1088) = 2.626$, $p = .105$, $\eta_p^2 = .002$, as participants were more accurate when rejecting low-numerosity suffixed pseudowords ($M = 9.0\%$, $SD = 10.9$) than high-numerosity suffixed pseudowords ($M = 11.8\%$, $SD = 12.3$). The suffix-salience main effect did not reach statistical significance. However, the interaction between suffix-numerosity and suffix-salience emerged at a marginally significant level in the by-participant´s analysis, $F_1(1,32) = 3.741$, $p = .062$, $\eta_p^2 = .105$; $F_2(1,1088) = .014$, $p = .907$, $\eta_p^2 = .000$. This effect showed that participants tended to commit less mistakes when responding to pseudowords with low-numerosity and low-salience suffixes than to pseudowords from all the other conditions. Specifically, the results revealed that on the high-numerosity condition there were no differences between the high-salience and low-salience conditions but, on the low-numerosity condition participants made less errors on the low-salience condition when compared to the high-salience condition ($p = .002$). It also reveals that on the high-salience condition there were no significant differences but on the low-salience condition participants were more accurate when responding to stimuli from the low-numerosity condition when compared to the high-numerosity one ($p = .022$). Figure 1 and Figure 2 depict the effect on reaction times and percentage of errors.

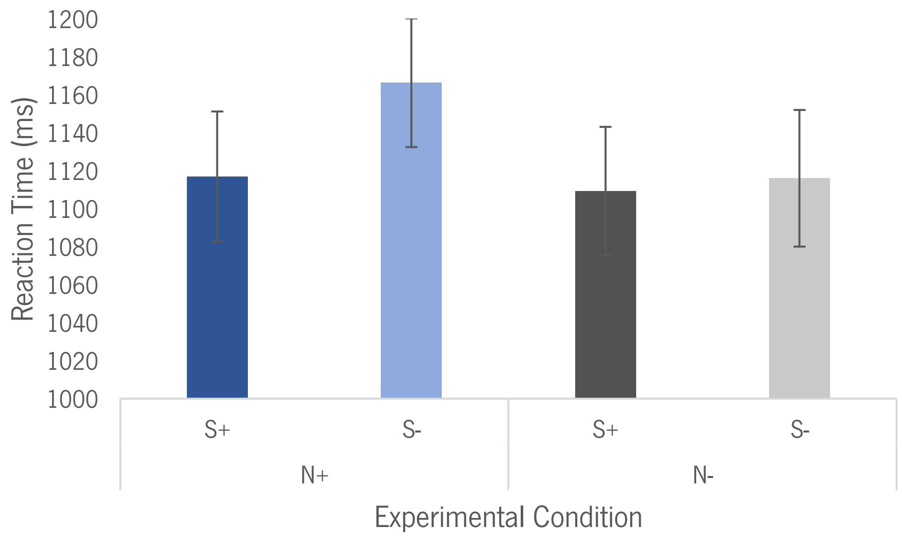

Figure 1. Graphic representation of the results per condition on reaction times (ms). Error bars represent standard errors. N+ = High-Numerosity; N- = Low-Numerosity; S+ = High-Salience; S- = Low-Salience.

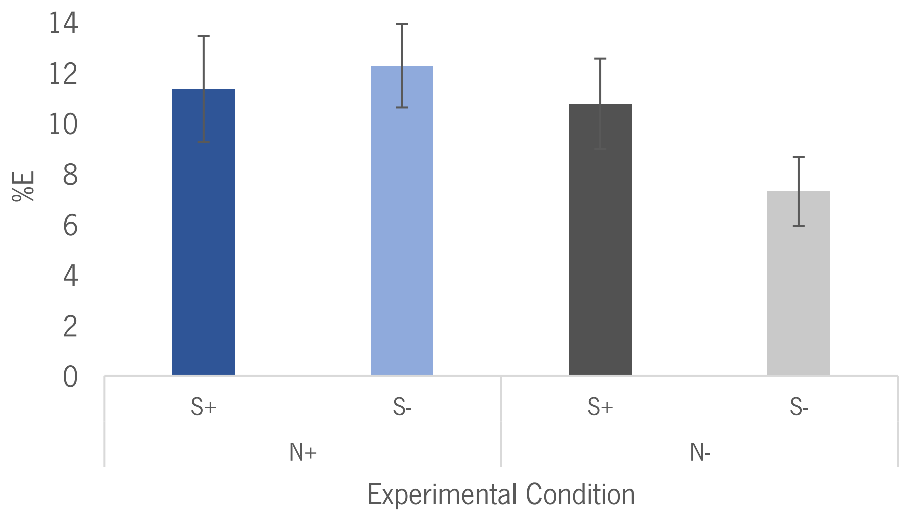

Figure 2. Graphic representation of the results per condition on percentage of errors (%E). Error bars represent standard errors. N+ = High-Numerosity; N- = Low-Numerosity; S+ = High-Salience; S- = Low-Salience.

Words

T-tests for paired samples on the time participants took to recognize Portuguese complex words, revealed a main effect of suffix-salience, $t_1(47) = -7.991$, $p < .001$; $t_2(70) = 2.423$, $p = .018$, showing that participants were faster at recognizing Portuguese words with high-salience suffixes (*M* = 859.8, *SD* = 161.0) than with low-salience suffixes (*M* = 938.8, *SD* = 176.4).

On the accuracy data, the results also showed a main effect of suffix-salience, $t_1(47) = -4.991$, $p < .001$; $t_2(70) = -.187$, $p = .512$, indicating that participants committed less errors with high-salience suffix words (*M* = 11.1%, *SD* = 5.2) when compared with low-salience suffixes (*M* = 18.0%, *SD* = 10.9).

Although the data reveal to be interesting and point towards a distinct morphological processing of pseudowords and words, they do not corroborate the advanced hypotheses, namely for pseudowords, where the suffix-salience effect did not show the expected interference effect. Actually, results demonstrate that a certain word ending being more times a suffix facilitated and not hampered pseudoword recognition. Furthermore, it is important to note that, although in our design all roots were combined with all the suffixes, hence neutralizing possible differences on the characteristics of the stimuli, the truth is that in each one of the cases, the combination of a given root with a given suffix might have created unexpected differences between the root-frequency and the suffix-numerosity that could have affected the results. This is, high-frequency roots combined with high-numerosity and high-salience suffixes might be more easily recognized by the participants than low-frequency roots combined with low-numerosity and low-salience suffixes. So, we decided to conduct a second set of analyses for the pseudowords introducing root-frequency (high vs low) as an additional within-subject factor in the analyses. The results of these analyses are presented below.

Additional analyses

Considering that 16.229 was the median value of the per million frequency of the roots, roots with frequency values below 16.229 were considered low-frequency roots (R^-), while roots above that value were considered high-frequency roots (R^+). Besides the main effects of suffix-numerosity and suffix-salience, that mimicked the effects observed in the previous analyses, the new analyses revealed a significant interaction effect of root-frequency X suffix-salience on participants' data, $F_1(1,32) = 5.888$, $p = .021$, $\eta_p^2 = .155$; $F_2(1,32) = .291$, $p = .593$, $\eta_p^2 = .009$, showing that participants were faster when rejecting pseudowords with high-salience suffixes ($p = .001$), independently of the suffix-numerosity, only for pseudo words with low-frequency roots. For high-frequency roots, no differences were found between low- and high-salience conditions. A significant suffix-numerosity X suffix-salience effect was also observed, $F_1(1,32) = 4.172$, $p = .049$, $\eta_p^2 = .115$; $F_2(1,32) = 3.900$, $p = .057$, $\eta_p^2 = .109$, indicating that, irrespectively of the root frequency, participants were slower when responding to low- than high-salience suffixes ($p = .011$), only on the high-numerosity conditions. For the low-numerosity conditions, no difference was observed between low- and high- salience conditions. Figure 3 depict the effects obtained on these additional analysis on reaction times (ms).

Regarding percentage of errors, the results revealed besides the suffix-numerosity main effect observed in the previous analysis, a main effect of root-frequency, $F_1(1,32) = 6.960$, $p = .013$, $\eta_p^2 = .179$; $F_2(1,1024) = .147$, $p = .702$, $\eta_p^2 = .000$, showing that participants were more accurate when responding to low-frequency roots (*M* = 4.7%, *SD* = 6.3) than to high-frequency roots (*M* = 5.7%, *SD* = 7.4), independently of suffix-numerosity and suffix-salience. Figure 3 also depicts the effect on percentage of errors (%E).

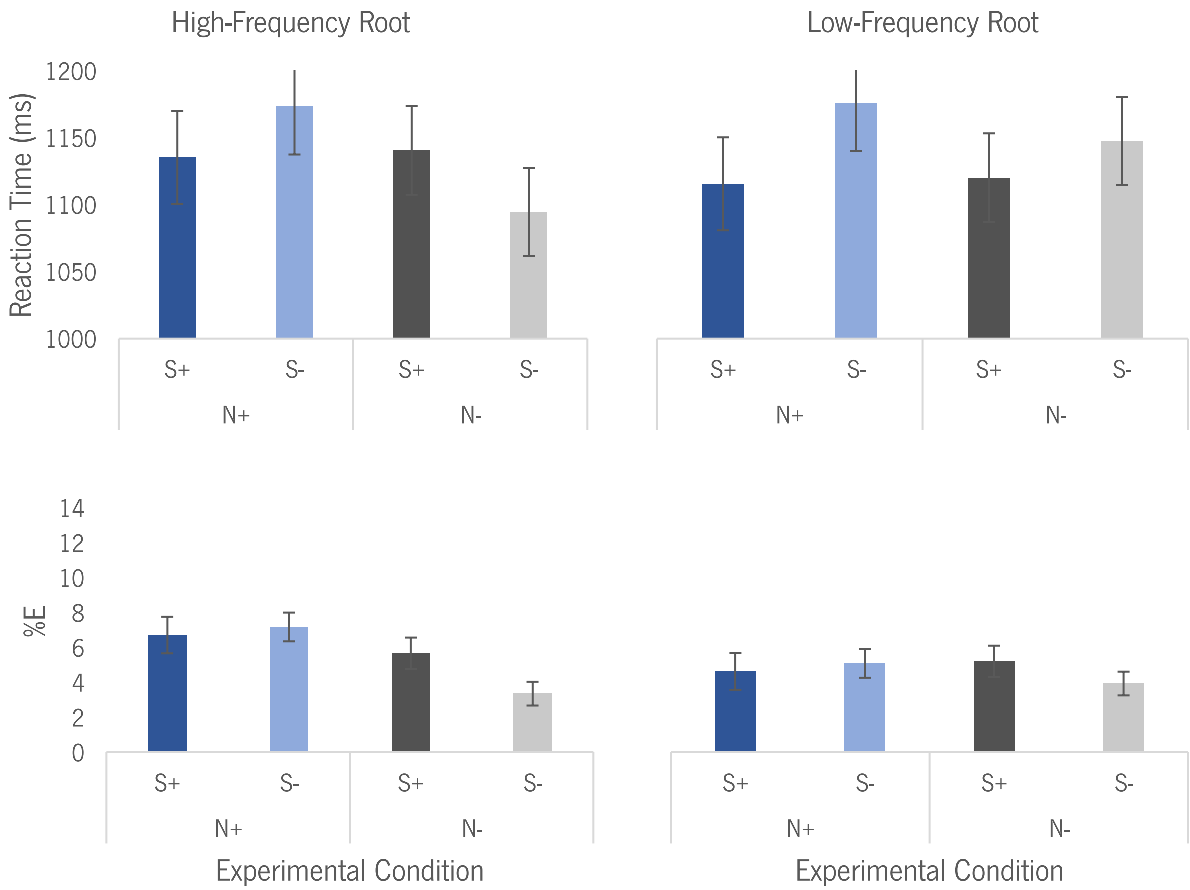

Figure 3. Graphic representation of the results on the additional analyses per condition, considering the root-frequency for both reaction times (ms) and percentage of errors (%E). Error bars represent standard error. N+ = High-Numerosity; N- = Low-Numerosity; S+ = High-Salience; S- = Low-Salience.

Discussion

The main goal of the present study was to explore one variable that seems to be critical on the visual word recognition of words and pseudowords – suffix frequency. Specifically, we intended to explore if morphological effects on visual word recognition are mainly due to morphological or to orthographic factors. To that purpose, we studied the effect of two variables directly related to suffix frequency (suffix-numerosity and suffix-salience) on the recognition of Portuguese pseudowords, and also the effect of suffix-salience on the recognition of Portuguese words. To that aim, native speakers of Portuguese performed a lexical decision task with suffixed words and pseudowords. The findings revealed that participants were faster when responding to the low-numerosity condition when compared to the high-numerosity one and more accurate when responding to pseudowords with low-numerosity suffixes when suffix-salience was low. Furthermore, when the root-frequency variable was added into the analyses the data was clear cut: when suffix-salience was low participants were faster and more accurate when responding to low-numerosity suffixes. In the case of complex words, when suffix-numerosity was high, they were faster and more accurate when responding to high-salience suffixes.

Regarding complex words, results corroborate the ones obtained in previous studies on morphological processing (e.g., Burani & Thornton, 2003; Taft & Forster, 1975), as they showed that words ending in a letter sequence that works as a suffix most of the times are recognized faster and more accurately than words ending in a letter sequence that, even though presenting a similar frequency of occurrence in the language, does not work as a real suffix most of the times. In our case, and accordingly to our initial hypothesis, since words were all of high-numerosity, the ones with low-salience suffixes were the ones to reveal lower reaction times and less percentage of errors. These findings seem to be in accordance with parallel dual-route models (see, e.g., Caramazza et al., 1988; Burani and Laudanna 1992; Chialant and Caramazza 1995; Schreuder and Baayen 1995) which states that frequency is the major determinant of the relative probability that lexical access is either whole-word based or morpheme based. This is, words composed by more than one morpheme may activate in parallel two types of access units, namely units corresponding to the whole word and units corresponding to the morphemes included in the stimuli. The assumption underlying these models is that the higher the frequency of a given lexical unit (i.e., a word, a root or an affix) the greater the likelihood that this unit is quickly activated and processed in the different components. The probability of the lexical access being provided by either whole-word or morpheme processing is determined by the balance of the frequencies of the whole-word and the frequency of its constituent morphemes, a more relative frequency than an absolute frequency (i.e., a derived word low frequent in the language that is composed by a very frequent root and a very frequent suffix is likely to be accessed via activation of its morphemes rather than via the whole-word).

On the other hand, pseudowords results only partially corroborated our initial hypotheses. Results revealed higher reaction times and percentage of errors on the high-numerosity conditions showing an effect of frequency of occurrence of a certain letter string over morphological effects since for morphology to play a role it was expected that the high-salience condition would also originate higher reaction times and higher percentage of errors. While no salience effect emerged on the percentage of errors, a significant effect emerged on the reaction times, though it was on the opposite direction: high-salience suffixes revealed faster reaction times than low-salience ones.

In general, our results are similar to those of Experiment 1 by Burani and Thornton (2003), where pseudowords with a high-frequency suffix also revealed longer reaction times and lower accuracy when compared to the control condition that was a letter string with similar frequency as the suffix. On the present study pseudowords with high-numerosity and low-salience suffixes had a similar pattern of results, however, differences on the stimuli must be acknowledged. For instance, what Burani and Thornton (2003) considered a pseudosuffix (i.e., a letter string with the same frequency as an analogous suffix but that did not work as a suffix) might be similar to what we considered to be a low-salience suffix. However, in our case, it is still a suffix but one that occurs less frequently as a suffix in the language. Besides, pseudowords with low-numerosity, independently of the suffix salience, were the ones that were faster to be rejected as words. This result might be accounted for by the Augmented Addressed Morphology model by Caramazza, Laudanni, and Romani (1988), since pseudowords like the ones we constructed to the present study might be more prone to morphemic segmentation because of the very high-numerosity of the suffix. So, a letter string (suffix) activates both whole-word representations (when available – that is, for known words) as well as the morphemes that comprise the word. Thus, the stimuli will activate different access representations (as in 'walked' will activate 'walked', 'walk-', and '-ed'), as well as orthographically similar representations (such as 'walks', 'walking', 'talked', 'talk', 'walked'). The orthographic representation that first reaches a pre-set threshold will activate the corresponding lexical entry. So, since low-numerosity suffixes are less familiar (i.e., they occur less times) they are immediately retrieved as being less likely to be part of a word leading to faster reaction times to be rejected as a word. Now, looking to the pseudowords with low-salience suffixes, the ones with high-numerosity were the ones with higher reaction times, meaning that when salience was low and numerosity was high, an interference effect might have occurred due to the existence of many orthographic competitors due to the frequent use, leading to higher reaction times. These findings were also verified in Burani et al. (1997), where high-frequency suffixes in pseudowords lead to more interference in deciding about the lexicality of those pseudowords when compared to non-suffixed pseudowords. They even concluded that the higher the frequency of an orthographic string, the greater the interference on lexicality decision and consequently the longer is the rejection time.

In summary, the reported results in the present study on suffixed words and pseudowords point towards orthography playing an important role on the visual word recognition (i.e., numerosity) side by side with morphological factors (i.e., salience). However, we have to acknowledge that the fact that all the final letter strings in our study were existing suffixes in Portuguese, do not allow us to completely isolate the effect of pure orthographic frequency variables. So, on future studies, one important factor to explore is the importance of the orthographic frequency being important to add, for example, letter strings that are not suffixes but are frequent on the Portuguese language. Only this way one would be able to explore pure frequency effects without morphological interference. If similar results with high-numerosity endings (that are not real suffixes) keeps occurring, then, it would demonstrate that orthographic frequency, independently of the morphological status, is an important variable in the so called “morphological decomposition effect”. Until then we can only assume that orthography seems to play an important role, but we cannot rule out the role of the morphology, since in our study all the stimuli presented to the participants included real suffixes. This study also emphasises the fact that root-frequency should be another variable to consider in future studies trying to control and manipulate it since the results presented in this study seem to provide some evidences that root can also have an important role on the parsing process.

References

Beyersmann, E., Coltheart, M., & Castles, A. (2012). Parallel processing of whole words and morphemes in visual word recognition. *The Quarterly Journal of Experimental Psychology*, *65*(9), 1798-1819.

Burani, C., Dovetto, F. M., Thornton, A. M., & Laudanna, A. (1997). Accessing and naming suffixed pseudo-words. In *Yearbook of morphology 1996* (pp. 55-72). Springer, Dordrecht.

Burani, C., & Thornton, A. M. (2003). The interplay of root, suffix and whole-word frequency in processing derived words. *Trends in Linguistics Studies and Monographs*, *151*, 157-208.

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, *28*(3), 297-332.

Crepaldi, D., Rastle, K., Coltheart, M., & Nickels, L. (2010). 'Fell'primes 'fall', but does 'bell'prime 'ball'? Masked priming with irregularly-inflected primes. *Journal of Memory and Language*, *63*(1), 83-99.

Diependaele, K., Sandra, D., & Grainger, J. (2005). Masked cross-modal morphological priming: Unravelling morpho-orthographic and morpho-semantic influences in early word recognition. *Language and Cognitive Processes*, *20*(1-2), 75-114.

Diependaele, K., Sandra, D., & Grainger, J. (2009). Semantic transparency and masked morphological priming: The case of prefixed words. *Memory & Cognition*, *37*(6), 895-908.

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

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior research methods, instruments, & computers*, *35*(1), 116-124.

Longtin, C. M., Segui, J., & Halle, P. A. (2003). Morphological priming without morphological relationship. *Language and Cognitive Processes*, *18*(3), 313-334.

Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(3), 785.

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, *11*(6), 1090-1098.

Rio-Torto, G., Rodrigues, A., Pereira, I., Pereira, R., & Ribeiro, S. (2016). *Gramática derivacional do Português*. Imprensa da Universidade de Coimbra (2ª edição).

Soares, A. P., Iriarte, Á., de Almeida, J. J., Simões, A., Costa, A., Machado, J., & Perea, M. (2018). Procura-PALavras (P-PAL): A web-based interface for a new European Portuguese lexical database. *Behavior Research Methods*, *50*(4), 1461-1481.

Soares, A. P., Lages, A., Silva, A., Comesaña, M., Sousa, I., Pinheiro, A. P., & Perea, M. (2019). Psycholinguistic variables in visual word recognition and pronunciation of European Portuguese words: a mega-study approach. *Language, Cognition and Neuroscience*, *34*(6), 689-719.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal learning and verbal behavior*, *14*(6), 638-647.