The Processing of the Definite Article in Brazilian Portuguese: When “the” Carries Gender and Number Marking

João Vieira[1], Elisângela Teixeira[2], Hayward J. Godwin[1] & Denis Drieghe[1]

1. School of Psychology, University of Southampton, UK

2. Departamento de Letras Vernáculas, Universidade Federal do Ceará, Brazil

Contact Information:

João Vieira
School of Psychology
University of Southampton
Southampton SO17 1 BJ
UK
Email: joao.vieira@soton.ac.uk

The raw data files and R scripts used for the analysis are publicly available at
https://osf.io/v2s3e/?view_only=03c3ccb1247f42f6ae2deea8331e11b2
This study was not preregistered.

ORCID
João Vieira - 0000-0001-5215-2020
Elisângela Teixeira - 0000-0003-3924-3985
Hayward J. Godwin - 0009-0005-1232-500X
Denis Drieghe - 0000-0001-9630-8410

Abstract

Research on eye movements during reading has shown that function words receive fewer and shorter fixations than content words. However, recent studies suggest that when matched in frequency, length, and predictability, such differences disappear. Two studies in English still indicate a special status of the article "the". Angele and Rayner (2013), using the gaze-contingent boundary paradigm, found that ungrammatical previews of "the" were skipped more often than grammatical content words, while Staub et al. (2019) found that repeated articles were noticed less often than repeated content words. We extended both studies to Brazilian Portuguese (BP), where articles carry more syntactic information (gender and number) than in English. In a gaze-contingent boundary experiment, we found that the preview of an ungrammatical definite article was skipped more often than the grammatical continuation, suggesting the mechanism of automatically skipping articles is also present in BP. Because this mechanism does not seem to be influenced by the extra information articles carry in BP compared to English, it is likely that it is the high frequency of the articles that is triggering word skipping as opposed to a special function word status. However, in the second experiment, repeated articles were noticed nearly as frequently as content words, presumably because the additional syntactic information articles carry in BP is connected to the sentence's structure in a more complex way than, for instance, English. So, in an artificial task, such as repetition detection during reading, differences between articles and content words can manifest themselves.

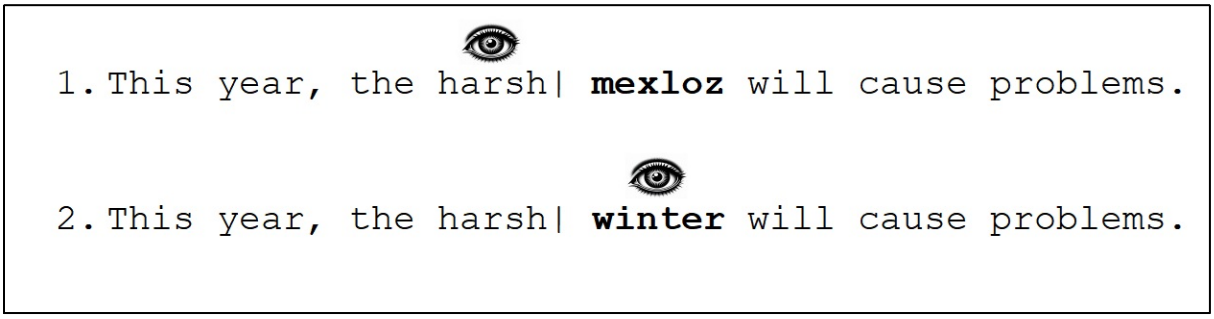
Keywords: eye movements, function words, parafoveal processing

The Processing of the Definite Article in Brazilian Portuguese: When “the” Carries Gender and Number Marking

Reading is a complex cognitive activity that involves intricate coordination between visual perception and language processing (Liversedge & Findlay, 2000). Fixations happen when our eyes remain relatively still on a focal point and, during reading, last 200-250 ms on average for proficient adult readers (Rayner, 2009). The number and duration of fixations are affected by word characteristics such as word length (Kuperman et al., 2024), frequency (Inhoff & Rayner, 1986), and predictability (Balota et al., 1985) such that shorter, more frequent, and more predictable words are skipped more often and receive shorter and fewer fixations on average when compared to longer, less frequent and less predictable words. Because our visual field only has high visual acuity in a relatively small area of about two degrees of visual angle around the fixation point, the fovea, the oculomotor system is constantly required to make quick jumps, saccades, to bring the eyes to a new fixation point. During saccades, a phenomenon called saccadic suppression occurs, rendering us virtually blind (Martin, 1974) and, for that reason, fixations are the only way through which our eyes can acquire new visual information during reading.

Besides information from the currently fixated word, useful information from the surrounding words, such as the initial letters of a word or a word’s approximate length, can be extracted from outside the fovea. These words will typically be located in the parafovea, which is an area of the visual field which extends from the foveal region to about five degrees from the fixation point in every direction. Our ability to obtain information from words in the parafovea and use such information is called parafoveal processing (see Schotter et al., 2012 for a review). One technique commonly employed to study parafoveal processing is the gaze-contingent boundary paradigm (Rayner, 1975). In this paradigm, the visual input changes dynamically depending on where the reader is looking. For example, in the sentence “This year, the harsh winter will cause problems.”, before the reader’s eyes cross the invisible boundary, illustrated by the vertical bar (see Figure 1), the parafoveal preview of the target word “winter” is replaced with random letters, “mexloz”. Once the eyes cross the boundary during the saccade, the preview “mexloz” changes to the correct target word “winter”. As this display change happens during a saccade, this change will not be noticed by the participant. Research has shown that fixation times on a target word are shorter when the preview was correct compared to when the preview was incorrect, thereby demonstrating what is called the parafoveal preview benefit (Rayner, 1975). This technique has been used extensively to examine both how much information and the nature of the information (e.g. orthographic, phonological) that is being parafoveally processed.

Figure 1



1. This year, the harsh| **mexloz** will cause problems.
2. This year, the harsh| **winter** will cause problems.

Note: The vertical bar represents the position of the invisible boundary.

Most research on eye movements during reading has focused on content words, such as nouns and verbs, while comparatively less attention has been paid to function words. Moreover, the majority of eye movement studies have examined reading in English (Siegelman et al., 2022). The goal of the current study is to investigate how definite articles are processed in Brazilian Portuguese, a language where, compared to English, articles carry more syntactic and semantic information in the form of gender and number marking (i.e. *o* and *os* are masculine singular and plural, while *a* and *as* are feminine singular and plural, respectively). The remainder of this Introduction will discuss research on function words which indicate different eye movement behaviour on function words compared to content words, before going into more recent findings challenging the idea that function words and content words are processed differentially when all relevant factors are taken into account (word length, frequency and predictability). We conclude by identifying two studies that still indicate different eye movement behaviour on function compared to content words in English and raise the issue of whether these findings would generalise to Brazilian Portuguese, a language where function words carry more content in the form of gender and number marking.

Research on reading established word class as a predictor of eye movements such that function words have consistently been reported to receive shorter and fewer fixations than content words (Drieghe et al., 2008; Gautier et al., 2000; O'Regan, 1979; Rayner, 1998). For example, in English, Drieghe et al. (2008) found that the article *the* was more likely to be skipped and receive shorter fixations than other high-frequency three-letter long words (e.g. *all*) (see O'Regan, 1979 for similar results, also in English). Likewise, in French, Gautier et al. (2000), showed that the definite plural article *les* was skipped more often than content words of the same length (e.g. *fût*). Luke and Christianson (2016) examined function and content words using data from the PROVO corpus, which is an extensive corpus of participants reading in their native English, and their analysis included the predictors of lexical predictability (i.e. predicting the exact word identity) and partial predictability (e.g. predicting the word's grammatical class). Results showed that function words with higher partial predictability received shorter fixations and were skipped more often than function words with low partial predictability. For content words, however, higher partial predictability also resulted in shorter fixations, but did not influence skipping rates, suggesting that predicting a content word's grammatical class does not provide enough information to justify skipping it, whereas it does for function words. In contrast, the authors found that lexical predictability affected word classes similarly.

Research in other experimental paradigms also suggests differences between the processing of content and function words. For example, letter-detection tasks, where participants have to count the number of occurrences of a given letter (e.g. e) in target words, have shown that participants are more likely to fail to notice the target letter when reading function words, such as the article *the,* compared to content words (Corcoran, 1996; also see Roy-Charland et al., 2022 for similar results in French). In a read-aloud task, Healy and Zangara (2017) presented participants with sentences containing repetitions of either the definite article *the* or the numeral *one* (e.g. in *The cougars stalked the/one prairie dog*, with either *the* or *dog* repeated or presented once). Results showed that participants failed to notice the repetition of the article *the* (i.e. they read the sentence aloud as if the repetition was not present) one-third of the time and only failed to notice the repetition of *one* in 5% of the trials, despite both words being high-frequency and having the same word length. Research on natural speech also shows differences between word classes. Bell et al. (2009) found that function words are spoken more quickly on average than content words, even when matched on frequency and predictability. Further, Juste et al. (2012) reported that children who stutter have more difficulties with function words, while older individuals are more likely to stutter on content words.

A different picture emerges from reading studies that have taken into account influences of word length, predictability and frequency, whilst not restricting themselves to examining only extremely frequent and predictable words, such as the article *the* in English, or *les* in French. Those studies where predictors of eye movements, such as frequency, length, and predictability, are controlled for, show little to no differences in processing costs between content and function words. For example, in a recent large-scale reading study in

English, Staub (2024) compared the processing costs of content and function words and found no differences in skipping rates between word classes but found unexpectedly somewhat longer first fixation and gaze durations on function words, possibly caused by a small subset of the target function words analysed. Similar results come from Schmauder et al. (2000), who, also in English, examined eye movements during reading of short sentences in which target content and function words were matched in length and frequency and found no difference in eye movement behaviour between classes.

Another large-scale study was carried out in Brazilian Portuguese (BP) using data from the RASTROS corpus of eye movements during natural reading (Vieira, 2020). This corpus consists of a dataset that includes word class information (content vs. function word), frequency, predictability, and word length for ~2,500 words in 50 short paragraphs. As mentioned, in BP, function words can carry gender and number marks, unlike, for instance, in English. Definite articles in BP can be masculine or feminine, and singular or plural. And this is also true for some longer function words (e.g. *aquele* and *aqueles* are masculine singular and plural, while *aquela* and *aquelas* are feminine singular and plural, all translate into *that/those* in English). Vieira et al. (2024) reported very limited differences in reading times between word classes. With the exception of a small difference in first fixation times that might be due to a paucity of data on low-frequency function words in the RASTROS corpus, all differences between function and content words in both reading times and skipping rates were limited to skipping rates on short words (one to four letters), where function words were skipped more often than content words. We will return to this finding in the general discussion.

Even though these latter studies show no (or very limited) differences between function and content words, two studies in English still suggest there might be differences between content and function words. One study focuses on parafoveal processing, the other examines error detection, and both examine the definite article *the* in English.

Angele and Rayner (2013) used the gaze-contingent boundary paradigm to examine word skipping of the article *the* by manipulating the parafoveal preview to be either the incorrect definite article *the* or the correct three-letter-long content word (e.g. She thought she would| *the/ace* all the tests.). The *the* preview created a syntactic irregularity such that the definite article would always appear in a grammatically illegal position (when used as a parafoveal preview). This manipulation allowed for the investigation of whether word identification would, in itself, be sufficient to trigger word skipping without taking into account syntactic information. If this were the case, the definite article should be skipped more often than the correct continuation of the sentence (e.g. *ace* in the example above) because *the* is a function word that is very frequent and carries very little information, so it should lead to successful word identification more rapidly. If syntactic information was used when deciding to skip the next word, the preview of the content word should induce more skipping compared to the syntactically illegal article. Angele and Rayner found that it was the definite article that was skipped more often than the correct continuation of the sentence, suggesting that the decision to skip a word does not take into account syntactic information, at least for short words. The question remains, however, whether it was the high frequency of *the* that triggered the high skipping rate or because *the* is a function word that carries very little meaning, which triggers ‘automatic’ skipping, a special status restricted to function words in general or even restricted to just the definite article.

The second study we will re-examine is a reading-alike study reported by Staub et al. (2019). Participants were presented with sentences that could have the definite article *the* either repeated or omitted, or have a repeated noun, and were required to read the sentences silently (e.g. *The ice cream melted in the cart before they left*, in which *the* or *cart* could be repeated, or *the* omitted). After reading each sentence, participants had to answer whether something was wrong with the sentence. Participants reported the repeated article 46% of the time (i.e. missed the repetition more than half the time), the omitted article 68% of the time, and the repeated noun 90% of the time. In a recent follow-up study, Staub et al. (2024) used the same task with sentences that had repetitions or omissions of two-letter function words (e.g., *of* and *on*) and found that repetitions and omissions of two-letter function words were both detected about two-thirds of the time. Further analyses in both the 2018 and 2024 studies revealed that eye movement behaviour was not affected by repetitions or omissions in sentences when participants did not notice the errors, possibly indicating that such errors are not initially perceived and corrected unconsciously. The suggestion is that grammar knowledge and linguistic expectations guide the system to fill in what is expected before errors are ever encoded by the brain. If the errors had to be corrected by the parsing system, even unconsciously, sentences with repetitions and omissions of words should show signs of slower reading than sentences with no errors.

Questions can be raised about whether findings on function words in English would generalise to some other languages. Unlike in English, most nouns and many function words (e.g., definite articles) in Brazilian Portuguese (BP) carry gender and number information. The morphology of the definite article is influenced by the sentence's syntax, as the article must agree in gender and number with the following noun as well as with possible adjectives and even past participles. In the sentence "The dirty table was cleaned" (*A mesa suja foi limpa,* in BP), the definite article "a" (*the*), the adjective "suja" (*dirty*), the noun "mesa" (*table*), and the past participle "limpa" (*cleaned*) must all be feminine and singular. This agreement creates syntactic requirements, which involve the articles, that are not present in English and raises the question of whether the results found by Angele and Rayner (2013) and Staub et al. (2019) would be replicated in BP.

In the current paper, we report two experiments examining the processing of definite articles in Brazilian Portuguese (BP), which always carry gender and number marking. In the first experiment, we used the gaze-contingent boundary paradigm to examine whether definite articles in BP are skipped automatically, either because of their function word status or due to their high frequency, even when placed in a grammatically illegal position, as observed in English with the article *the* (Angele & Rayner, 2013). The added content in terms of number and gender marking might result in function words in BP to behave more as content words in languages such as English, and as such the automatic skipping due to the function word status might not occur. However, if high frequency primarily drives skipping as in resulting in fast word recognition, grammatically illegal articles might still be skipped more often than correct previews even when carrying extra information.

The second experiment examines whether participants notice repeated definite articles in BP. In English, repetitions of *the* often go unnoticed (Staub et al., 2019), while repeated shorter function words like *of* or *on* are more likely to be detected (but still go unnoticed nearly one-third of the time, Staub et al., 2024). This raises the question of whether the repetition of *the* not being noticed is due to *the* being uniquely devoid of informational load in English, or is this a more general characteristic of definite articles across languages, including BP? If more complex words, including definite articles with gender and number markings, are more likely to be consciously processed, we expected the detection of repeated articles in BP to be higher than both the article *the* and two-letter function words in English.

Methods

Experiment 1

In Experiment 1, participants read sentences with a two-letter-long target content word with four possible parafoveal previews (see Table 1). In the *normal* condition, the parafoveal preview of the target word was not manipulated, so it was the correct target word (e.g. *pé*, foot in English); in the *illegal article* condition, the preview was changed to a definite article (i.e. *os* or *as*, the in English); in the *random letter* condition, random letters were used as a preview (e.g. *qx*); and in the *wrong accent* condition, the correct target word received an accent when it did not have one, or had the correct accent changed when it originally had one (e.g. *pè* instead of *pé*). An invisible boundary was placed at the beginning of the empty space before the target word and when the participant's gaze crossed the boundary, the correct target word was presented.

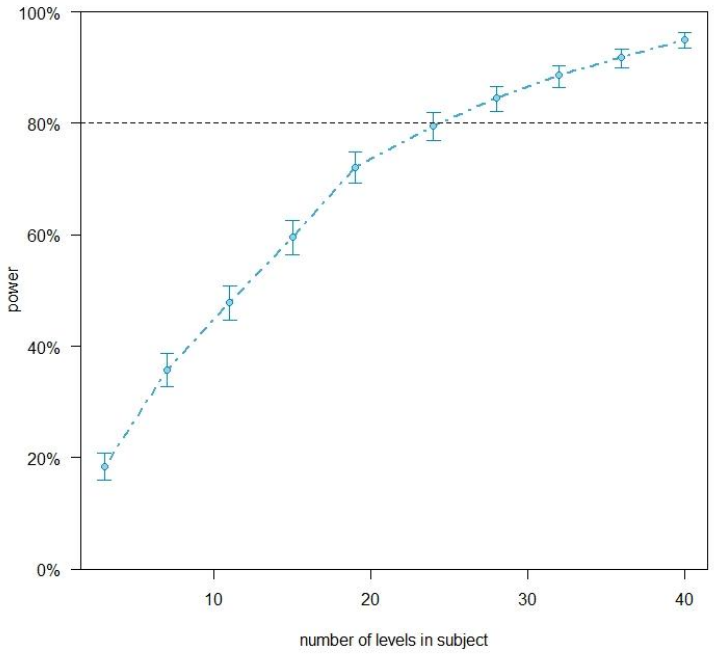
Sentences were created so that the target word position would always be syntactically illegal for an article, so the preview in the *illegal article* condition always created a grammatical error until the eyes crossed the boundary. We included the *wrong accent* condition to examine an additional research question whether accent information can be parafoveally processed in BP. Previous research in Spanish found effects of wrong accent on late processing, such as longer total reading and re-reading times on target words with the wrong accent preview, but showed no influence on early processing such as skipping rates (Marcet & Perea, 2022). In our study, accent information was always three characters from the invisible boundary. All words with added or changed accents in our stimuli do not exist in BP (e.g. *pè* is not a word), so, if accent information is picked up parafoveally, we expect less skipping of and longer fixation times on target words with the wrong accent information in comparison to the normal condition (e.g. *pè* vs *pé,* respectively).

Participants

Forty-six undergraduate students (30 female; mean age: 21.6; range: 18 - 31) from the Federal University of Ceará, in Fortaleza, Brazil took part in the experiment. All participants were native speakers of Brazilian Portuguese, had normal or corrected to normal vision, and reported no reading difficulties. To estimate the minimum sample size needed, we conducted power simulations using the *simr* package (version 1.0.7, Green & MacLeod, 2016) in R Studio (version 4.1.1) using data from the first ten participants in our study. We used the effect size reported by Angele and Rayner (2013) when comparing the probability of fixating the target word when it was the normal condition (i.e. correct continuation of the sentence) versus the illegal article preview (β = – -.47). Our simulations indicated that a sample of approximately 25 participants was necessary to detect an equivalent effect in our data with a minimum of 80% power (Figure 2). We collected data from 46 participants.

Figure 2

Power Analysis for Experiment 1



Note: the horizontal line represents how many participants were needed to reach a minimum of 80% power.

Material

Sixty experimental items were created with four conditions each. All items had a two-letter-long target content word that could be a verb, adjective, adverb, or noun. By utilising the gaze-contingent paradigm (Rayner, 1975), the preview of the target word was manipulated according to one of four conditions: In the *normal* (N) condition, the preview of the target word was the correct word; in the *illegal article* (IA) condition, the preview was replaced by a grammatically illegal plural definite article, either *os* or *as*[1] (equivalent to *the* in English); in the *random letters* (RL) condition, the preview was made of random letters; and in the *accent* (AC) condition, the preview either included an added accent (e.g. the preview of *ir* was *iř*) or displayed a wrong accent when the word already had an accent (e.g. the preview of *pé* would be *pè*). Table 1 illustrates all conditions. The bar indicates where the invisible boundary was placed.

Table 1

An Example Stimulus for All Conditions

Condition	Sentence
Normal	O bebê machucou o seu delicado| **pé** pisando na areia quente[2].
Illegal Article	O bebê machucou o seu delicado| **os** pisando na areia quente.
Random Letters	O bebê machucou o seu delicado| **qx** pisando na areia quente.
Accent	O bebê machucou o seu delicado| **pè** pisando na areia quente.

The mean Zipf frequency of the content target words was 4.7 (range: 2.9-6.3), and 6.9 for the definite article *os*, and 6.8 for *as* (Boos et al., 2014). All target words were content words, out of which two were adjectives, three adverbs, 18 verbs, and 37 nouns. In the construction of experimental sentences, particular attention was given to ensuring that the definite article would always be grammatically incorrect in the position of the target word.

[1] When used correctly, the gender and number marks of the definite article, in BP, will match those of the following noun. In our experiment, since the definite articles were grammatically illegal and did not precede a noun, we made sure half the trials had a masculine article and half had a feminine article.

[2] In English: *The baby hurt his delicate **foot** by stepping on hot sand*.

The pre-target word always had at least four letters, to ensure that when the pre-target word was typically fixated, the preview was in an area with relatively good visual acuity. In addition to the 60 experimental items, 20 fillers with no preview manipulations were included in the study. Four lists were created from the 60 experimental trials following a Latin square design. Each list had 15 items from each experimental condition, along with all 20 fillers. Four practice trials were presented at the beginning of the experiment.

To ensure target words were not predictable, a cloze task (Taylor, 1953) was conducted with 15 participants who did not participate in the reading experiment. All participants read the sentence up to the target word and answered what they thought the next word was. None of the target words used in this experiment were answers given in the cloze task.

Apparatus and Procedure

Eye-movement data were collected using an SR Research Eyelink 1000 eye-tracker at a sample rate of 1000 Hz. Sentences were displayed in Courier New font, size 18, on a light-grey background. Participants read sentences on a Dr. Office, 21.5’, 75 Hz monitor. The distance between the participant’s eyes and the monitor was 65 cm, which amounted to approximately three letters per degree of visual angle.

Participants were asked to read silently and for comprehension. A three-point calibration was done at the beginning of the experiment until maximum error was below 0.5 degrees for the three points and below 0.3 degrees for the central calibration point where the target word was always located close to. In addition, a re-calibration was carried out after participants asked for a rest, or when the drift correction at the beginning of the trial amounted to more than 0.5 degrees. Following the initial four practice trials, participants read all 60 trial items and 20 fillers in a pseudorandom order. Participants were instructed to press the UP arrow on the keyboard in front of them when they had finished a sentence and wanted to proceed. After one-quarter of the sentences, participants were presented with Yes/No comprehension questions, which they answered by pressing either the LEFT (Yes) or RIGHT (No) keys on the keyboard. All participants had an accuracy of at least 85%, indicating they were reading for comprehension.

Data from four participants were removed due to their detection of the display change more than ten times[3]. Participants received R$20,00 upon completing this experiment. The experimental procedure was approved by the ethics committees at both the Federal University of Ceará and the University of Southampton. Before commencing the experiments, participants read and signed an informed consent form.

Data Availability

All materials, including R scripts, stimuli, and eye movement data are available at https://osf.io/v2s3e/?view_only=03c3ccb1247f42f6ae2deea8331e11b2

Results

The initial processing of the data was done using Data Viewer (SR Research). First, all fixations that were no more than 0.5 degrees away and shorter than 80 ms were merged. Then, we followed the same procedure but for fixations shorter than 40 ms and no more than 1.25 degrees away. Lastly, all remaining fixations shorter than 80 ms, longer than 800 ms, or outside of interest areas were removed. Additionally, trials were removed if display changes occurred due to: a fixation before the boundary; by the eyes crossing the boundary prior to

[3] Hohenstein and Kliegl (2014) and Veldre and Andrews (2018) reported that display change awareness does not seem to impact patterns of results in display change experiments. Contrastingly, White et al. (2005) reported differences in eye movements when participants noticed the changes. In some studies (e.g. Drieghe & Chan, 2023; Zang et al., 2020), authors report removing participants who noticed a certain number of changes (i.e. 10), while in others (e.g. Schwalm & Radach, 2023), authors do not mention removing participants. We chose to take a conservative approach and removed participants who noticed more than ten display changes.

landing before the boundary, a phenomenon known as “hooking”; or a display change being triggered too late after the eyes crossed the boundary (10 ms or more) as Slattery, Angele, & Rayner (2011) found that changes delayed by more than 10 ms impact eye movement behaviour regardless of whether participants notice the change. Approximately 9% of trials were removed based on these criteria. Further, each dependent measure was checked for outliers and observations of 2.5 standard deviations or more from the participants’ mean were removed, resulting in another ~3% of the data being excluded. Our analysis is centred on the critical region. However, we note that analyses on the pre-target region did not show any effect of the preview manipulation (i.e. parafoveal-on-foveal effects).

The following dependent variables were analysed: skipping rates, which is how often a word was skipped during first pass; first fixation duration (FFD), the duration of the first fixation on a word; single fixation duration (SFD), the duration of a fixation on a word when it received exactly one fixation; gaze duration (Gaze), which is the sum of all fixations on a word before the eyes leave the word; go past time (Go Past) is how long it takes for the eyes to go past a word to the right the first time it is read, including possible regressions; and regression out (Reg Out), how often regressions originated from the target region. The means and standard deviations of all dependent measures on the target word are in Table 2.

Table 2

Condition Means Across Subjects on Target Words

		Fixation times (ms)			Probabilities	
Condition	SFD	FFD	Gaze	Go Past	Skipping	Reg Out
Normal	234 (71)	233 (73)	235 (74)	309 (195)	65% (0.48)	21% (0.41)
Illegal Article	251 (100)	264 (123)	263 (110)	409 (400)	76% (0.43)	30% (0.46)
Random Letters	270 (103)	287 (128)	281 (112	427 (338)	60% (0.49)	32% (0.47)
Wrong Accent	239 (82)	239 (83)	248 (93)	304 (178)	62% (0.49)	17% (0.38)

Note: Standard deviations across subjects are in parentheses.

Our data were analysed using linear mixed models with the lme4 package, version 1.126 (Bates et al., 2015) in R Studio (version 4.1.1). For all the models, we used sliding differences contrasts for the parafoveal preview manipulation. Contrast 1 was such that the *accent* condition was compared to the *normal* condition, Contrast 2 compared the *normal* condition to the *illegal article* condition, and Contrast 3 compared the *illegal article* condition to the *random letters* condition. The initial model included preview contrasts as a fixed factor and a maximal random-effects structure: random intercepts for participants and items, plus by participant and by item slopes for the contrasts. If the model failed to converge, we simplified it by removing elements of the random structure with either a perfect correlation, or else associated with the smallest variance. We repeated this process until the model converged. For binomial measures, we used logistic GLMMs and followed the same process. The results for all the models are in Table 3. Before going into the individual eye movement measures, we note that the contrast comparing the wrong accent and normal conditions did not yield any statistically significant effects (all *t* values below 1.98), so we will not focus on this contrast for the remainder of the results section.

Regarding skipping rates, the preview with *random letters* was skipped less often than the condition with the *illegal article*, while the *illegal article* condition was skipped considerably more often than the *normal* condition.

Focussing on the remaining eye movement measures, for the comparison between the normal preview and the illegal article, the normal preview condition had shorter first fixation times, gaze duration and go-past times. The numerical difference in single fixation duration was not significant and there was no difference in Regression Out. For the comparison between the article and random letters, first fixation durations, single fixation duration and gaze duration were longer in the random letter condition compared to the illegal article condition with no differences observed in the later measures (go past times and regressions out).

Table 3

Linear Mixed Models for Analyses of Target Words

		Estimate	SE	t/z	p
Skipping Rates	Intercept	0.75	0.13	5.77	< **0.001**
	Normal - Accent	0.18	0.13	1.32	0.19
	Article - Normal	0.58	0.14	4.07	< **0.001**
	Random - Article	-0.87	0.14	-6.17	**0.01**
FFD	Intercept	5.45	0.03	210.27	< **0.001**
	Normal - Accent	-0.03	0.04	-0.84	0.4
	Article - Normal	0.09	0.04	2.23	**0.03**
	Random - Article	0.09	0.04	2.18	**0.03**
SFD	Intercept	5.44	0.03	215.53	< **0.001**
	Normal - Accent	-0.02	0.03	-0.73	0.47
	Article - Normal	0.06	0.04	1.56	0.12
	Random - Article	0.07	0.04	1.72	**0.09**
Gaze	Intercept	5.46	0.03	212.29	< **0.001**
	Normal - Accent	-0.05	0.04	-1.37	0.17
	Article - Normal	0.09	0.04	2.17	**0.03**
	Random - Article	0.07	0.04	1.74	**0.08**
Go Past	Intercept	5.69	0.04	126.66	< **0.001**
	Normal - Accent	-0.01	0.05	-0.18	0.86
	Article - Normal	0.18	0.06	3.28	**0.001**
	Random - Article	0.08	0.05	1.43	0.15
Reg Out	Intercept	-1.32	0.18	-7.43	< **0.001**
	Normal - Accent	0.27	0.27	0.99	0.32
	Article - Normal	0.52	0.28	1.83	**0.07**
	Random - Article	0.09	0.26	0.35	0.73

Note: p values in bold are significant.

Discussion

The main objective of Experiment 1 was to examine, in Brazilian Portuguese, whether grammatically illegal parafoveal previews of definite articles are skipped more often than target content words with the same length, as shown to be the case in English, where articles carry comparatively less information than in Brazilian Portuguese (Angele & Rayner, 2013). The second aim of Experiment 1 was to examine, also in Brazilian Portuguese, whether accent information is processed parafoveally. Participants read sentences with a parafoveal preview that could be the target content word (no manipulation), an article in a grammatically illegal position, random letters, or the correct content word with either a wrong accent or an added accent when the word did not have one.

Results showed that accent information did not influence skipping rates or reading times on target words, likely because the change was too subtle to be captured parafoveally, even if only three characters away from the invisible boundary. In Spanish, Marcet and Perea (2022) found inflated times on words missing accents, but in their study, participants directly fixated on the words missing the accents, while our study focused on whether accent information is picked up during parafoveal processing, which possibly explains the different results between the two studies.

Returning to the main aim of Experiment 1, two-letter words are often skipped during reading, as evidenced by the fact that our target words were skipped about two-thirds of the time in the *normal* condition. Still, skipping rates were even higher in the *illegal article* condition, replicating results found in English (Angele & Rayner, 2014). There are two possible explanations for the results, and one does not necessarily exclude the other. First, as suggested by Angele et al. (2014), the extremely high frequency of definite articles may be the cause of their higher skipping rates. The very fast processing time for these highly frequent words will lead to high skipping rates, as the syntactic information does not seem to influence the decision to skip the word. Secondly, these results could also be caused by the function word status, which could lead to automatic skipping.

Regarding fixation duration on target words, the overall pattern showed that *random letters* resulted in longer fixations than in the *illegal articles*, while the *normal* condition received the shortest fixations overall. This is probably because preview benefits (Rayner,

1975) were only present in the *normal* condition, while random letters likely feed noise into the system, causing the longest fixations (Hutzler et al., 2013). Further, our results showed that readers were more likely to revisit earlier parts of the sentence after fixating on the target word in the *illegal article* condition than in the *normal* condition.

In summary, our findings showed that, in BP, parafoveal previews with grammatically illegal definite articles were skipped more often than the correct continuation of the sentence, likely due to their high frequency, their function word status or both. In addition, wrong accent information was not picked up parafoveally and did not influence reading behaviours.

Experiment 2

In Experiment 2, participants read a series of sentences that could either have a repeated definite article (e.g. *os* or *as*) or a repeated content word (e.g. *luas*, moons in English). In the *normal* condition, sentences were presented without any repetition, while in the *repeated article* condition, the article was repeated (e.g. *os os*) and in the *repeated content word* condition, the content word was repeated (e.g. *luas luas*). Otherwise, sentences had no other errors.

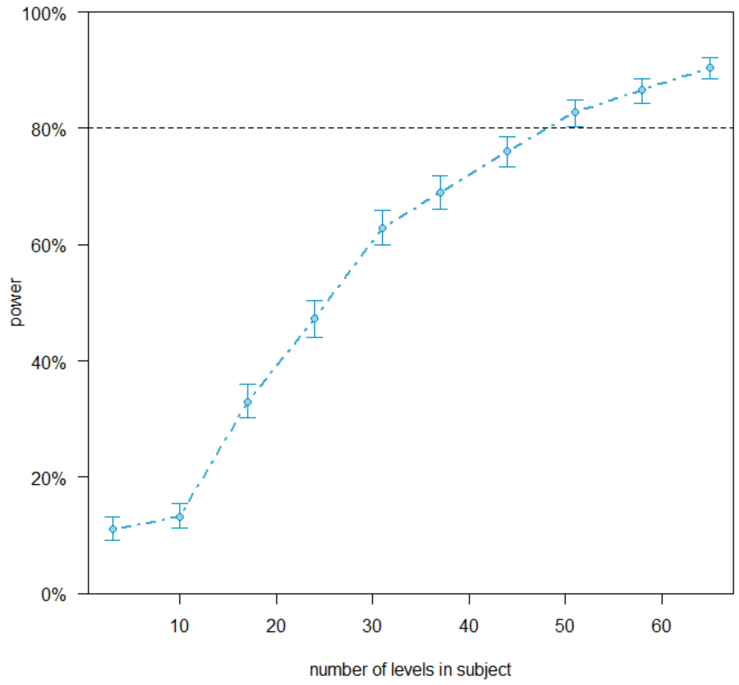
Methods

Participants

Sixty-two undergraduate students (39 female; mean age: 21.6; range: 18 - 32) from the Federal University of Ceará, in Fortaleza, Brazil took part in the experiment. All participants were native speakers of Brazilian Portuguese and had normal or corrected to normal vision. To estimate the minimum sample size needed, we performed power simulations using the *simr* package (Green & MacLeod, 2016) in R Studio (version 4.1.1) using data from the first ten participants. We used the effect size reported by Staub et al. (2018) for the interaction between repetition manipulation (“repeated the” vs “repeated noun”) and fixation patterns (whether one or both instances of the repeated words were fixated) ($\beta = -1.18$) on the accuracy of repetition detection. Our simulations showed that we would need a sample of approximately 48 participants to detect an equivalent effect in our data with at least 80% power (Figure 3). We collected data from 62 participants.

Figure 3

Power Analysis for Experiment 2



Note: the line represents how many participants were needed to reach a minimum of 80% power.

Materials

Sixty experimental items were created with three conditions each. The *grammatical* (G) condition had no errors, the *repeated article* (RA) condition had the critical article in the sentence repeated, and the *repeated content word* (RC) had the critical content word repeated. Table 4 illustrates all three conditions:

Table 4

Example Stimulus for All Conditions

Condition	Sentence
Grammatical	O estudante adora olhar **as luas** de Saturno com seu telescópio[4].

[4] In English: The student loves to look at the|the moons|moons of Saturn with his telescope.

Repeated Article	O estudante adora olhar **as as** luas de Saturno com seu telescópio.
Repeated CW	O estudante adora olhar as **luas luas** de Saturno com seu telescópio.

The critical content words in our study had either three or four characters (mean 3.9) and the definite article always had two characters (*os* and *as*, *the* in English). The mean zipf frequency of our content words was 4.1 (range: 3 to 6.7) (Boos et al., 2014), accordingly, the majority of the words were classified as high-frequency, exceeding a Zipf score of four. Twenty-seven of our trials had the feminine plural version (*as*) definite article and thirty-three had the masculine plural version[5] (*os*). We also included twenty filler trials which had no manipulation, resulting in a total of 80 trials.

Three lists were created from the sixty experimental trials. Each list had 20 items from each condition, plus the 20 fillers and four practice trials, presented at the beginning of the experiment. Each participant read 84 sentences.

Apparatus and Procedure

Apparatus and calibration procedures were the same as in Experiment 1. After completing four practice trials, participants proceeded to read all 60 experimental items and 20 filler sentences in a pseudo-randomised order. After every trial, participants were presented with the question “Was there anything wrong with the previous sentence?” (note that the question was presented in Portuguese), to which they pressed LEFT (Yes) or RIGHT (No) on the keyboard in front of them to answer. Additionally, a YES/NO comprehension question was asked after 50% of filler sentences, which they also answered using the keyboard. Participants were paid R$20,00 after completing the experiment. Two participants had an accuracy rate below 80% on the comprehension questions and were removed.

[5] Participants’ accuracy in detecting the definite article repetition was the same regardless of its gender.

Results

Initial eye movement data treatment on DataViewer (SR Research) followed the same procedure explained in Experiment 1. In this Experiment, we only analysed fixation probability on target words as the measure of eye movement behaviour. Therefore, no data were excluded based on standard deviation thresholds. We ran a linear mixed model with contrasts with weights of -0.5 and 0.5. One contrast compared condition (RCW vs. RA), and the other contrast compared fixations on one instance versus both instances. Both contrasts were included as fixed factors, along with their interaction. The few cases where both instances were skipped were removed from this analysis (~6%). We employed the same strategy described in Experiment 1 to determine the optimal random effects structure. The model output is in Table 5.

The contrast comparing fixations on one instance vs. both instances of repeated words was not significant, suggesting accuracy was not influenced by first pass fixations. There was a significant difference in accuracy between conditions such that participants were more likely to detect the repetition in the RCW condition than in the RA condition. The interaction was not significant. Participants reported the error more often in the RCW condition (97%) than in the RA condition (92%), while seldom falsely reporting an error in the G condition (Table 6). These numbers form a striking difference with what Staub et al. (2019) found in English (~46% detection rate of repeated articles overall and ~89% for content words), likely caused by the additional syntactic information articles carry in Brazilian Portuguese. Across all conditions, fixation patterns during first pass had no significant impact on repetition detection (Table 7), showing participants noticed the repetition most of the time on both conditions, regardless of where their eyes landed. In contrast, Staub et al. found that participants were more likely to notice the repetition of articles when fixating both instances (66%) in comparison to fixating only the first instance

(34%) or only the second instance (46%), an indication that foveal processing increased the chances of detecting the error in English.

Table 5

Linear Mixed Models for Analyses of Accuracy

Fixed Effects	β	SE	z	P
Intercept	3.01	0.26	11.26	**< 0.001**
Both vs. One	0.5	0.35	1.43	0.15
RCW vs. RA	1.4	0.51	2.73	**0.01**
BothVsOne*RCWvRA	-0.17	0.53	-0.32	0.74

Note: All significant p values are in bold.

Table 6

Accuracy in Detecting Word Repetition

	Repetition Report	
Condition	Percentage	SD
G	2%	15
RA	92%	18
RCW	97%	27

Table 7

Accuracy in Detecting Repeated Words Per Fixation Pattern During First Pass

Condition			
RA	Fixation Pattern	Fixation Probability	Error Report
	Both Fixated	23% (42)	95% (23)
	Both Skipped	11% (31)	91% (29)
	First Only Fix	22% (41)	90% (30)
	Second Only Fix	44% (50)	92% (26)
RCW			
	Both Fixated	71% (46)	97% (16)
	Both Skipped	2% (17)	94% (24)
	First Only Fix	13% (34)	95% (22)
	Second Only Fix	14% (50)	96% (19)
G			
	Both Fixated	28% (45)	97% (16)
	Both Skipped	37% (19)	98% (15)

Article Only	14% (35)	98% (11)
CW Only	54% (50)	98% (16)

Note: Standard deviations are inside the parentheses. In the RA and RCW conditions, Both Fixated refers to when both instances of the target word were fixated, and Both Skipped refers to when both instances were skipped. In the G condition, both measures refer to when the article and the content word in the sentence were either fixated or skipped.

Discussion

The objective of Experiment 2 was to investigate how often participants notice the repetition of definite articles compared to the repetition of short content words in Brazilian Portuguese. In English, Staub et al. (2019) found that participants only noticed the repetition of the definite article ~46% of the time, while content words were noticed nine out of ten times. In our study, participants silently read sentences that featured no repetitions or had either a definite article or a short content word repeated and had to report whether they had noticed something wrong in the sentence. Because of the additional gender and number information in articles in Brazilian Portuguese, we predicted that participants would notice the repetition of definite articles more often than participants detect the repetition of the definite article in English, but less frequently than the repetition of our content words (Staub et al., 2019).

Participants showed near-ceiling accuracy in detecting repetitions in both conditions, but repeated definite articles were still noticed slightly less often than content words (92% vs 97%, respectively). Interestingly, while results follow the same direction as in English (45% for articles and 89% for content words; Staub et al., 2018), the difference between conditions is considerably smaller in BP. Using the same task as Staub et al. (2019), a study by Staub et al. (2024) had participants read sentences which could have repeated or omitted two-letter long function words in English (e.g. *of* and *in*) and found that participants noticed the repetitions and omissions roughly two-thirds of the time. In our study, repeated articles, which were also always two-letter long, were detected more often than two-letter function words in English and much more often than the article *the*. Together, these results seem to suggest that the repetition of the definite article *the* in English may be particularly easy to overlook, even when compared to shorter English function words, but the same is not true for definite articles in BP.

General Discussion

Research on eye movements during reading has traditionally considered word class as a predictor of eye movements, such that function words typically receive shorter and fewer fixations than content words (Drieghe et al., 2008; Gautier et al., 2000; O’Regan, 1970). More recently, two large-scale studies have provided converging evidence for the claim that, once word length, frequency, and predictability are taken into account, there is little to no difference in processing costs, as indicated in eye movement measures, between content and function words both in English (Staub, 2024) and Brazilian Portuguese (Vieira et al., 2024). Here, we focus on two studies in English that indicate that there might still be something special about the definite article *the*. First, Angele and Rayner (2013) found that the article is often skipped even when used as a grammatically illegal parafoveal preview, and second, Staub et al. (2019) showed that repetitions of the definite article *the* are perceived by readers less often than repetitions of content words (also see Healy & Zangara, 2017, who found similar results in a reading-aloud experiment, also in English). The main objective of this study was to investigate whether different patterns would be observed when the indefinite articles carry more information than they do in English, such as in Brazilian Portuguese, where the indefinite article has number and gender marking.

In our study, results from Experiment 1 showed that definite articles used as grammatically illegal parafoveal previews were skipped more often than the correct target content words, even with the additional information the articles carry in BP. The findings from Angele and Rayner (2013) indicated that word recognition in itself is sufficient to trigger the decision to skip the next word and that syntactic information is not taken into account when making this decision. The question remained whether this was due to the very high frequency of the article or due to a special status taken up by function words resulting in automatic skipping of words that are known to carry very little meaning.

There are two studies that point towards the high skipping rate of indefinite articles being, for the most part, if not entirely, due to their high frequency. The first study is reported by Angele et al. (2014) as a direct follow-up to the 2013 study, in which they used the gaze-contingent paradigm again, and presented different high-frequency words (beyond the definite article *the*) as the syntactically illegal parafoveal previews of three-letter-long words. They found that high-frequency short target words were also skipped more often than the correct continuation of the sentences. This suggests that at least some of the high skipping of syntactically illegal article previews could be due to their high frequency and not due to their special function word status. The second study is the one we report here. If there is a special status assigned to function words because the system has learned that these words contain little to no meaning, it is reasonable to assume that this mechanism would be less pronounced in a language where function words contain more information compared to English. However, in the current study in Brazilian Portuguese, we closely replicate the findings of increased skipping of a syntactically illegal indefinite article. Combining these studies, we can state that there is good evidence for the high skipping of syntactically illegal previews of words being due to their high frequency.

We also mention two related studies examining reading in Chinese. Zang et al. (2018) found that syntactically illegal previews of the very high-frequency structural particle (的), were skipped more often than the correct continuation of the sentence (e.g. a verb). In a follow-up, Zang et al. (2020) also found that other high-frequency syntactically wrong previews of content words in Chinese were skipped more often than lower-frequency, correct continuations. Even though these studies did not directly examine the (non-existent) indefinite article in Chinese, the reported results are clearly compatible with the findings from Angele et al. (2013; 2014).

Some indications for a special status of function words for word skipping still exist even in a language such as Brazilian Portuguese. Vieira et al. (2024) found that, in Brazilian Portuguese, even when frequency and predictability were entered as predictors in the model, very short function words were skipped more often than very short content words. However, this might be due to a specific property of the language where, once readers have determined the next word is a function word (which will happen more often when the parafoveal word is a short word), they may have learned to automatically skip the function word as the information contained in it (gender and number) is reliably repeated in the content word (as in “The girl found the dolls she wanted”, *A menina encontrou **as** bonec**as** que ela queria*, in BP, where the gender and number information in the article *as* is repeated at the end of the following noun “bonecas”). A replication of this finding in a language where the function word contains information that is not transparently repeated in the noun (e.g. German) would, therefore, be extremely interesting to elucidate whether this phenomenon is language-specific.

Additionally, in Experiment 1, we found standard preview benefits (Rayner, 1975) such that target words received longer fixations in the condition with the article preview than in the condition with no preview manipulation. Interestingly, when the preview was random letters, fixation times were longer than in the article condition, even though both conditions provided no preview benefit. These findings are compatible with previously made suggestions that using random letters in preview manipulation may not be an ideal baseline, as they introduce noise rather than provide a meaningful point of comparison (see Hutzler et al., 2013 for a discussion on this topic).

In a research question separate from the main focus of this paper, Experiment 1 also showed that accent information in BP was not picked up parafoveally, and did not impact reading in any way. The manipulation in our study was such that either the accent information was added when a word did not originally have one, or in words where there is an accent, the accent was incorrect. Previous research in Spanish found no change in early processing of target words missing accent marks, but did find inflated fixation times in later measures, including rereading, on words missing accents (Marcet & Perea, 2022). However, it is important to note that this previous study did not explore parafoveal processing by means of the display change paradigm. The lack of an effect of wrong accents on any measures in our study, therefore, suggests that such information may not be processed parafoveally, and only happens after a word is fixated. Contrastingly, *diacritics* (accent-like marks in Arabic which convey complete vowel information) have been shown to influence reading behaviours on both foveal (Hermena et al., 2015) and parafoveal (Hermena et al., 2016) processing, suggesting that in languages where accent and accent-like marks carry comparatively more information, readers may detect changes as early as during parafoveal processing.

In Experiment 2, we showed that in an error detection task, native readers of BP had near-ceiling-level accuracy in detecting repetitions of both definite articles and content words, while still detecting repeated articles slightly less often than content words. In English, however, participants have been reported to fail to notice the repetition of the definite article in English more than 50% of the time (Staub et al., 2019) and roughly one-third of the time for shorter, two-letter-long function words (Staub et al., 2024). One additional finding from Staub et al. (2024) is that fixation durations and skipping rates were not statistically different between control trials and trials with ungrammatical repetitions in cases where participants failed to notice the repetition, likely because the system "ignored" the error by inferring it was not there, based on language knowledge and expectations. In other words, the error was corrected by the system before ever reaching awareness. Because our participants noticed the repetition more than nine out of ten times, we do not have enough data to statistically analyse whether the same pattern repeats in BP.

Unlike the function words used in the experiments in English by Staub et al. (2019; i.e. the article *the*) and Staub et al. (2024; e.g. *on*, and *of*), definite articles in BP carry gender and number information, which are conveyed through letters, changing the morphology of the articles depending on the marks they carry. Because of that additional information, articles in BP influence, or are influenced by, the syntactic structure of the sentence in more ways, than, for example *the* in English, by forcing an agreement in gender and number across multiple words. Additionally, articles in BP can appear in more positions in the sentence than in English. In English, the definite article *the* usually appears before a noun or an adjective preceding the noun (e.g. the car; the new car), while, in BP, definite articles also appear in front of possessive pronouns, such as in "Eu pintei **a** minha casa" (*I painted [the] my house*, in English), and in contractions with certain prepositions, such as in "Ele entrou **na** minha casa" (*He entered in [the] my house*, in English). In this sense, definite articles in BP also interact with the sentence syntax in more ways than in English, potentially requiring more attention from readers, more similar to what happens to most content words in English and BP. The fact that, in English, errors on shorter function words (e.g *of* and *in*) are noticed more often than repetitions of *the* also suggests that the definite article is particularly empty, both in meaning and syntactic influence, and that when words carry more information, even other function words (e.g. *of* and *on*), the system is more likely to detect anomalies.

While the additional information present in definite articles in BP is likely behind participants noticing their repetition nine out of ten times in our Experiment, we believe the way this extra information affects reading is because of how it needs to align with the sentence’s syntax, shaping one’s expectations during the process of reading. This is not to say that the article *the* does not have an impact on reading, but that having an L1 language with articles with more information may impact one’s reading behaviours and expectations. For example, native speakers of Cantonese who are second language speakers of English are not as efficient as native speakers of English in using *the* as a cue to identify referents in the context, possibly because the syntactic structure of Cantonese has classifiers and not determiners such as *the* before nouns (Kang & Ge, 2022).

In Experiment 2, we only examined how often repetitions of articles and content words were noticed by participants, excluding omissions of articles. In some cases, the syntactic structure in BP allows for articles to be omitted, which would cause participants to only notice the error later in the sentence, so we preferred to avoid including a condition with an omitted article in the current. However, considering our findings, future research would be wise to explore whether and how the omission of the article is noticed in BP and other languages where function words carry more syntactic information.

In the current study, we employed very different tasks in each experiment. In Experiment 1, we examined eye movement behaviours of participants reading sentences with parafoveal previews containing a grammatical violation and, in Experiment 2, we asked participants to actively report whether they had detected errors in the sentences they read. Even though both tasks involved a type of grammatical violation, participants were influenced by errors quite differently in each experiment. Eye-tracking is an online technique that captures cognitive processes translated into eye movements as the eyes move along the text, while error detection tasks require additional input from participants, as well as explicit attention to the linguistic structure. In Experiment 1, participants did not consciously notice the illegal preview with the article and skipped them most of the time, while in Experiment 2, the repetition of the article allowed for foveal processing and participants detected the error more than 90% of the time. Therefore, the task in Experiment 2 is quite artificial in the sense that participants face the errors multiple times in a short duration and learn that they have to find them. It is very likely, as proposed by Staub et al. (2019), that in natural reading situations, repetitions of articles, even in BP, would be noticed (even) less often.

To summarise, in Experiment 1, we analysed parafoveal processing by examining whether a grammatically illegal article would be skipped, while in Experiment 2, we analysed how salient the repetition of definite articles is in BP, by asking participants whether they had noticed the error. Experiment 1 is compatible with the notion that, due to the very high frequency of the article, word processing is very fast and leads to increased word skipping that does not take syntactical information into account (see also Angele & Rayner, 2013; Angele et al., 2014). Experiment 2 shows that in error detection, a repeated article is noticed slightly less often than a repeated content word in BP. This is different from English, where failure to detect the article repetition seems far more prevalent (Staub et al., 2019), and is in all likelihood due to the article containing more syntactic and semantic information in BP compared to English, and how multiple words influence articles in the sentence. Combined, the results from both studies highlight the automaticity of article skipping, probably due to their high frequency, even when they have more information in BP compared to English, and how articles engage with the complex syntactic structure in BP.

References

Angele, B., & Rayner, K. (2013). Processing *the* in the parafovea: Are articles skipped automatically? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(2), 649–662. https://doi.org/10.1037/a0029294

Angele, B., Laishley, A. E., Rayner, K., & Liversedge, S. P. (2014). The effect of high- and low-frequency previews and sentential fit on word skipping during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(4), 1181–1203. https://doi.org/10.1037/a0036396

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, 17(3), 364–390. https://doi.org/10.1016/0010-0285(85)90013-1

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

Bates, E., Devescovi, A., Hernandez, A., & Pizzamiglio, L. (1996). Gender priming in Italian. *Perception & Psychophysics*, *58*(7), 992–1004. https://doi.org/10.3758/bf03206827

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92–111. https://doi.org/10.1016/j.jml.2008.06.003

Boos, R., Prestes, K., Villavicencio, A., & Padró, M. (2014). BRWAC: A WaCky corpus for Brazilian Portuguese. In *Lecture notes in computer science* (pp. 201–206). https://doi.org/10.1007/978-3-319-09761-9_22

Corcoran D. W. (1966). An acoustic factor in letter cancellation. *Nature*, 210(5036), 658. https://doi.org/10.1038/210658a0

Drieghe, D., & Seem, R. C. (2023). Parafoveal processing of repeated words during reading. *Psychonomic Bulletin & Review*, *29*(4), 1451–1460. https://doi.org/10.3758/s13423-021-02054-0

Drieghe, D., Pollatsek, A., Staub, A., & Rayner, K. (2008). The word grouping hypothesis and eye movements during reading. *Journal of Experimental Psychology Learning Memory and Cognition*, *34*(6), 1552–1560. https://doi.org/10.1037/a0013017

Friederici, A. D., & Jacobsen, T. (1999). Processing grammatical gender during language comprehension. *Journal of Psycholinguistic Research, 28*(5), 467–484. https:// doi.org/ 10. 1023/A:10232 64209 610

Gautier, V., O'Regan, J. K., & Le Gargasson, J. F. (2000). 'The-skipping' revisited in French: programming saccades to skip the article 'les'. *Vision Research*, 40(18), 2517–2531. https://doi.org/10.1016/s0042-6989(00)00089-4

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

Healy, A. F., & Zangara, T. K. (2017). Examining misses in reading aloud repeated words. *Quarterly Journal of Experimental Psychology*, *70*(3), 373–377. https://doi.org/10.1080/17470218.2016.1218521

Hermena, E. W., Drieghe, D., Hellmuth, S., & Liversedge, S. P. (2015). Processing of Arabic diacritical marks: Phonological–syntactic disambiguation of homographic verbs and visual crowding effects. *Journal of Experimental Psychology Human Perception & Performance*, *41*(2), 494–507. https://doi.org/10.1037/xhp0000032

Hermena, E. W., Liversedge, S. P., & Drieghe, D. (2016). Parafoveal processing of Arabic diacritical marks. *Journal of Experimental Psychology Human Perception & Performance*, *42*(12), 2021–2038. https://doi.org/10.1037/xhp0000294

Hohenstein, S., & Kliegl, R. (2014). Semantic preview benefit during reading. *Journal of Experimental Psychology Learning Memory and Cognition*, *40*(1), 166–190. https://doi.org/10.1037/a0033670

Hutzler, F., Fuchs, I., Gagl, B., Schuster, S., Richlan, F., Braun, M., & Hawelka, S. (2013). Parafoveal X-masks interfere with foveal word recognition: evidence from fixation-related brain potentials. *Frontiers in Systems Neuroscience*, *7*. https://doi.org/10.3389/fnsys.2013.00033

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics, 40*(6), 431–439. https://doi.org/10.3758/BF03208203

Juste, F. S., Sassi, F. C., & de Andrade, C. R. F. (2012). Exchange of disfluency with age from function to content words in Brazilian Portuguese speakers who do and do not stutter. *Clinical Linguistics & Phonetics*, 26(11–12), 946–961. https://doi.org/10.3109/02699206.2012.728278

Kang, X., & Ge, H. (2022). Tracking Object-State representations during Real-Time language comprehension by native and non-native speakers of English. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.819243

Kuperman, V., Schroeder, S., & Gnetov, D. (2023). Word length and frequency effects on text reading are highly similar in 12 alphabetic languages. *Journal of Memory and Language*, 135, 104497. https://doi.org/10.1016/j.jml.2023.104497

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, 4(1), 6–14. https://doi.org/10.1016/s1364-6613(99)01418-7

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, 88, 22–60. https://doi.org/10.1016/j.cogpsych.2016.06.002

Marcet, A., & Perea, M. (2022). Does omitting the accent mark in a word affect sentence reading? Evidence from Spanish. *Quarterly Journal of Experimental Psychology*, *75*(1), 148–155. https://doi.org/10.1177/17470218211044694

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin,* 81(12), 899–917. https://doi.org/10.1037/h0037368

O'Regan, K. (1979). Saccade size control in reading: Evidence for the linguistic control hypothesis. *Perception & Psychophysics*, *25*(6), 501–509. https://doi.org/10.3758/bf03213829

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology, 7*(1), 65–81. https://doi.org/10.1016/0010-0285(75)90005-5

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Roy-Charland, A., Collin, M.-M., & Richard, J. (2022). The development of the missing-letter effect revisited: The role of word frequency and word function. *Experimental Psychology,* 69(5), 275–283. https://doi.org/10.1027/1618-3169/a000565

Schmauder, A. R., Morris, R. K., & Poynor, D. V. (2000). Lexical processing and text integration of function and content words: Evidence from priming and eye fixations. *Memory & Cognition, 28*(7), 1098–1108. https://doi.org/10.3758/BF03211811

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention Perception & Psychophysics*, 74(1), 5–35. https://doi.org/10.3758/s13414-011-0219-2

Schwalm, L., & Radach, R. (2023). Parafoveal syntactic processing from word N + 2 during reading: the case of gender-specific German articles. *Psychological Research*, *87*(8), 2511–2532. https://doi.org/10.1007/s00426-023-01833-9

Siegelman, N., Schroeder, S., Acartürk, C., Ahn, H., Alexeeva, S., Amenta, S., Bertram, R., Bonandrini, R., Brysbaert, M., Chernova, D., Da Fonseca, S. M., Dirix, N., Duyck, W., Fella, A., Frost, R., Gattei, C. A., Kalaitzi, A., Kwon, N., Lõo, K., . . . Kuperman, V. (2022). Expanding horizons of cross-linguistic research on reading: The Multilingual Eye-movement Corpus (MECO). *Behavior Research Methods*, *54*(6), 2843–2863. https://doi.org/10.3758/s13428-021-01772-6

Staub, A. (2024). The function/content word distinction and eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 50(6), 967–984. https://doi.org/10.1037/xlm0001301

Staub, A., Dodge, S., & Cohen, A. L. (2019). Failure to detect function word repetitions and omissions in reading: Are eye movements to blame? *Psychonomic Bulletin & Review*, 26(1), 340–346. https://doi.org/10.3758/s13423-018-1492-z

Staub, A., McMurray, H., & Wickett, A. (2024). Perceptual inference corrects function word errors in reading: Errors that are not noticed do not disrupt eye movements. *Cognitive Psychology*, *154*, 101691. https://doi.org/10.1016/j.cogpsych.2024.101691

Taylor, W. L. (1953). “Cloze Procedure”: a new tool for measuring readability. *Journalism Quarterly*, 30(4), 415–433.

Veldre, A., & Andrews, S. (2018). How does foveal processing difficulty affect parafoveal processing during reading? *Journal of Memory and Language*, *103*, 74–90. https://doi.org/10.1016/j.jml.2018.08.001

Vieira, J. M. M. (2020). *The Brazilian Portuguese eye tracking corpus with a predictability study focusing on lexical and partial prediction* [Master’s thesis, Federal University of Ceará]. UFC, Biblioteca Universitária. http://www.repositorio.ufc.br/handle/riufc/55798

Vieira, J. M. M., Teixeira, E., Rodrigues, E. D. S., Godwin, H. J., & Drieghe, D. (2024). EXPRESS: When function words carry content. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/17470218241307582

White, S. J., Rayner, K., & Liversedge, S. P. (2005). Eye movements and the modulation of parafoveal processing by foveal processing difficulty: A reexamination. *Psychonomic Bulletin & Review*, *12*(5), 891–896. https://doi.org/10.3758/bf03196782

Zang, C., Du, H., Bai, X., Yan, G., & Liversedge, S.P. (2020). Word skipping in Chinese reading: The role of high-frequency preview and syntactic felicity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *46*(4), 603-620.

Zang, C., Zhang, M., Bai, X., Yan, G., Angele, B., & Liversedge, S.P. (2018). Skipping of the very high frequency structural particle de in Chinese reading, *The Quarterly Journal of Experimental Psychology, 71*(1), 152-160.