



HAL
open science

Effects of vertically aligned flankers during sentence reading

Jonathan Mirault, Jonathan Grainger

► **To cite this version:**

Jonathan Mirault, Jonathan Grainger. Effects of vertically aligned flankers during sentence reading. *Journal of experimental psychology. Human learning and memory*, 2024, 10.1037/xlm0001329 . halshs-04385939

HAL Id: halshs-04385939

<https://shs.hal.science/halshs-04385939v1>

Submitted on 10 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Effects of vertically aligned flankers during sentence reading

Jonathan Mirault ^{1,2} & Jonathan Grainger ^{1,3}

1 Laboratoire de Psychologie Cognitive, Aix-Marseille Université & Centre National de la Recherche Scientifique, UMR 7290, Marseille, FRANCE.

2 Pôle pilote AMPIRIC, Institut National Supérieur du Professorat et de l'Éducation, Aix-Marseille Université, Marseille, FRANCE.

3 Institute of Language, Communication and the Brain, Aix-Marseille Université, Aix-en-Provence, FRANCE

Short title: vertical flankers and sentence reading

Corresponding author:

Jonathan Mirault (jonathan.mirault@univ-amu.fr)

Laboratoire de Psychologie Cognitive

Aix-Marseille Université

3 place Victor Hugo,

13331 Marseille, FRANCE

Abstract

We examine whether information lying above and below a line of text being read can impact on reading fluency. We did so by placing length-matched flankers above and below each word in a sequence of words. We found that identical flankers facilitated sentence reading, compared with syntactically correct different text flankers, in both reading-for-meaning (Experiment 1) and grammatical decisions (Experiment 2). Experiment 3 replicated the same text facilitation in grammatical decisions and found no significant difference between different word and nonword distractors. Experiment 4 tested for an impact of case matching across targets and flankers and found a significantly greater same text facilitation when targets and flankers were in the same case. These results suggest that the same text facilitation effect might well be driven by crowding mechanisms that are more sensitive to vertically aligned information when reading horizontally aligned text.

Keywords: Sentence reading; Vertical flankers; Grammatical decisions

© 2024, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/xlm0001329

Introduction

When reading text, such as the present manuscript, each fixated word (and most content words are fixated during reading: Rayner, 1978) is typically flanked by other words to the left and right and above and below. This results in a well-established phenomenon studied by vision scientists and referred to as crowding (see Herzog & Manassi, 2015; Herzog et al., 2015; Pelli et al., 2007; Pelli & Tillman, 2008; for reviews), although, as noted by Herzog et al. (2015), crowding research started off as a branch of reading research (Legge, 2006). Crowding refers to the negative impact of visual clutter on object identification (e.g., and although crowding increases with eccentricity it is also known to affect the visibility of objects in central vision (e.g., Lev et al., 2014; see Grainger et al., 2016, for a description of the impact of crowding on letter visibility in central vision). The potential impact of crowding on text reading was nicely illustrated by Pelli et al. (2007) in their Figure 5. In a short text composed of 7 lines and roughly 8 words/nonwords per line, the informal demonstration was that when fixating the central word in the text and without making an eye movement readers are not aware that the text outside of a horizontally aligned elliptical region was nonsense (i.e., composed of nonwords). This region of uncrowded text was centered on the fixated word and spanned the line above and below this word and approximately 3-4 characters to the left and to the right of the edges of that word.

The demonstration of Pelli et al. (2007) suggests that only very limited info is available both left and right and immediately above and below the currently fixated word in a text being read. The horizontal extent of useful information during reading has been estimated using the moving window technique (McConkie & Rayner, 1975), and formally estimated to be around 4 character-spaces to the left and 15 spaces to the right (e.g., Rayner et al., 1980; but see Jordan et al., 2016, for a re-evaluation of the leftward extent of this

window). The rightward bias is thought to be a reading-specific effect, driven by reading direction (Pollatsek et al., 1981), and the attentional biases that are associated with that (Ducrot & Grainger, 2007; Engbert et al., 2005; Snell & Grainger, 2018).

The oval form of the crowding zone is typically determined by a critical spacing manipulation and an object identification task (e.g., Bouma, 1970; Pelli et al., 2007; but see Herzog et al., 2015, for a summary of the evidence against feedforward pooling accounts of crowding). The oval form of uncrowded object identification also illustrates the radial-tangential anisotropy exhibited by crowding (Toet & Levi, 1992). That is, the extent of crowding is greater for radially aligned objects than tangentially aligned objects, and the radial alignment corresponds to the horizontal meridian for text that is typically written horizontally (see Whitney & Levi, 2011, for a summary of the characteristics of crowding). Therefore, when reading horizontally aligned text the uncrowded window has a greater horizontal than vertical extent (see Chanceaux, et al., 2013; Vejnović & Zdravković, 2015; and Winsler et al., 2022, for evidence that this is indeed the case).

Nevertheless, the demonstration of Pelli et al. (2007) suggests that some information might be available from words falling above and below the fixated word (i.e., information from tangentially / vertically aligned stimuli). Snell, Mathôt, et al. (2018) tested this possibility using the flankers task, first developed by Eriksen and colleagues to investigate attentional influences on visual object perception (see Eriksen, 1995, for a review). In the reading version of the flankers task developed by Dare and Shillcock (2013), participants respond to central target words that are flanked to the left and to the right (separated by a single space) by related or unrelated flanker stimuli. Snell, Mathôt, et al. tested for simple flanker repetition effects (i.e., both flankers are the same word as the target or a different

word) and added a vertical manipulation of flanker position to the standard horizontal manipulation. For example, in the horizontal condition, the target word “rock” could be flanked to the left and to the right either by the same word “rock rock rock” or by a different word “film rock film”. In the vertical condition testing the same stimuli, the flankers appeared one line above and one line below the target. Participants made lexical decisions to the central targets. Snell et al. found a robust 39 ms flanker facilitation effect with the horizontally aligned flankers and a non-significant 2 ms effect with the vertically aligned flankers.

The results of the Snell, Mathôt, et al. (2018) study therefore suggest that an insignificant amount of information is extracted from stimuli lying above and below the word that is fixated. Nevertheless, that particular study involved single word targets, and it might well be the case that vertically arranged flankers impact on reading behavior when this is measured across an entire sentence. The present study aimed to investigate this possibility by measuring sentence reading performance while manipulating the nature of the information falling above and below the sentence being read, with these flanking stimuli spanning the complete sentence, as in normal text reading. We asked participants to read a centrally located and horizontally aligned sequence of words, while ignoring the words placed above and below the target sequence. In Experiment 1, participants read target sentences for meaning, and in Experiments 2 - 4 they made speeded grammatical decisions to a random mix of correct sentences and ungrammatical sequences of words.

Experiment 1

Methods

Participants. One hundred participants (32 reported being women) were recruited online using the Prolific website (Palan & Schitter, 2018). They all reported being native French speakers and received monetary compensation (£9/hour) for their participation. The participants ranged in age from 18 to 71 years ($M = 29.20$ years, $SD = 9.48$) and they were naïve to the purpose of the experiment. In this, and all following experiments participants, read and accepted an online informed consent form before starting the experiment. Ethics approval was obtained from the *Comité de Protection des Personnes* SUD-EST IV (No. 17/051).

Apparatus. The experiment was created using Labvanced (Finger et al., 2017). Only recent Internet navigators (Chrome, Safari, Opera, MS Edge) on common machines (Apple, Microsoft, Linux) were allowed. Recruitment of participants was done via the Prolific website and the advertisement was active online for 1 day.

Design & Stimuli. We created 100 French 5-word sentences for the study. They ranged in length between 20 and 39 characters (including spaces) ($M = 28.75$ characters; $SD = 4.07$) with words varying in length between 2 and 15 letters ($M = 4.95$ letters; $SD = 2.35$). For each sentence, we created another sentence formed of unrelated different words matched in length in letters with each word in the target sentence (see English examples in Figure 1 below, and the OSF repository (Mirault & Grainger, 2023). for a complete list of the French stimuli). This formed the two levels of the factor Distractor Type (same vs. different). A Latin-Square

design was applied such that participants were tested in both conditions but with different target sentences, and all target sentences were tested in both conditions across two groups of participants. The presentation order was randomized for each subject.

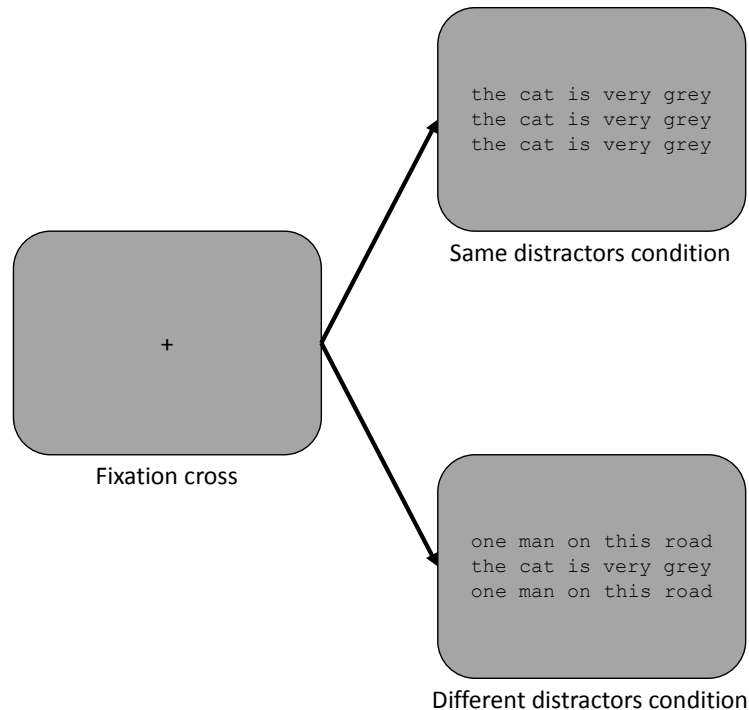


Figure 1. Examples (in English for convenience) of target and distractor sentences in the same distractor condition (upper panel) and different distractor condition (lower panel) in Experiment 1. See OSF repository (Mirault & Grainger, 2023) for the French stimuli tested in the experiment. Participants were requested to only read the sentence in the middle.

Procedure. After validating the consent form, participants read the instructions on their screen. They first started with 6 practice trials followed by the main experiment. A trial started with a fixation cross during 500 ms centered on the screen. Then, the target sentence appeared centered on the fixation cross and bounded above and below by two distractor sentences (which were the same) separated from the target sentence by an estimated 0.95° of visual angle (target center to distractor center) for an approximate viewing distance of 60 cm. Figure 1 provides a schematic of the procedure. Participants were instructed to read for comprehension and to press the space bar when then had finished reading the sentence. A gap

of 200 ms preceded presentation of the fixation cross for the next trial. Target and distractor sentences were presented in black on a light grey background in Courier New (monospaced), font size 16. In order to ensure that participants read the target sentence for meaning, a yes/no comprehension question (e.g., is the cat grey?) appeared after participants' response approximately every 10 items. Participants were instructed to ignore the sentences presented above and below the central target sentence.

Analysis. We used Linear Mixed-Effects models (LME) to analyze the response times (continuous data) and Generalized Linear Mixed-Effects models (GLME) to analyze error rates (binomial data), with items and participants as crossed random effects including by-item and by-participant random intercepts and random slopes (Barr et al., 2013; Baayen et al., 2008). The model was fit with the lmer function from the lme4 package (Bates et al., 2015) in the R statistical computing environment (R Core Team, 2018). We report regression coefficients (b), standard errors (SE) and t-values for fixed effects which were deemed reliable if $|t| > 1.96$ (Baayen, 2008), and indicated as such in bold font (for this and the following analyses). Response times were log10 transformed prior to analysis.

Statistical Power. After excluding outliers, we obtained a dataset of 9,477 observations (see Results section for details) with 4,738 observations per condition, which is largely above the threshold of 1,600 data points per condition recommended by Brysbaert and Stevens (2018). Using the SIMR package (Green & McLeod, 2016) in R environment, we obtained a power of 100% (95 % CI = 0.37) with over 1000 simulations.

Results

We measured response time (RT) as the time between onset of stimulus presentation (target sentence and distractors) and the moment participants pressed the space bar on their computer keyboard. Prior to analysis we excluded 2.20 % of trials with excessively short or long RTs (i.e., < 100 ms and > 5000 ms) and then 3.08 % of trials that were +/- beyond 2.5 standard deviations from the grand mean, leaving us with a total of 9,477 observations. Condition means are shown in Table 1.

Table 1. Mean RT per experimental condition, and net effect of Distractor Type (same words vs. different words) in Experiment 1.

Response time (ms)	Mean	95% CI
Same words	1476	19.42
Different words	1607	27.79
Effect	131	

There was a significant effect of Distractor Type ($b = -0.03$; $SE = 5.26^1$; $t = -6.91$), with faster RTs in the condition where distractors were the same sentence as targets compared to the condition where they were different sentences. The complete output of the analysis, including values of intercepts and correlations between slopes and intercepts, is provided in the Open Data section for this and the following experiments.

Discussion

¹ Here and in the following experiments the beta (b) and SE values obtained in the analysis of log₁₀ RTs are multiplied by 10³ for readability.

The results of Experiment 1 revealed a robust interfering effect of different word distractors placed immediately above and below a central (horizontally aligned – see Figure 1) target sentence that had to be read for meaning. This strong impact of vertical flankers contrasts with the null effects reported by Snell, Mathôt, et al. (2018) with single words. As suggested in the Introduction, the significant effect seen in Experiment 1 is likely due to the greater amount of processing involved in sentence comprehension compared with single word reading. However, it could also be due to the different nature of the task – reading for meaning compared with lexical decision. In order to have a more comparable sentence version of the flanked lexical decision task used by Snell, Mathôt, et al. (2018), in Experiment 2 we used the grammatical decision task. As argued elsewhere (see e.g., Mirault & Grainger, 2020), the grammatical decision task can be considered to be the sentence-level equivalent of the widely used lexical decision task for single word recognition research (see also Arnon & Snider, 2010).

Experiment 2

Methods

Participants. One hundred participants (42 reported being women) were recruited online using the Prolific website (Palan & Schitter, 2018). They all reported being native French speakers and received monetary compensation (£9/hour) for their participation. The participants ranged in age from 18 to 61 years ($M = 28.78$ years, $SD = 8.69$) and they were naïve to the purpose of the experiment.

Apparatus. The same as for Experiment 1.

Design & Stimuli. We used the same set of 100 sentences as in Experiment 1 with the same set of unrelated distractor sentences. In Experiment 2 we added 100 ungrammatical sequences that were created by randomly associating 5 words that were matched in length to the correct sentences (and checking that this random selection was indeed ungrammatical), and we manipulated Distractor Type (same vs. different) in the same way as for the correct sentences (i.e., the different distractors for the ungrammatical sequences were also ungrammatical). Otherwise, the design was the same as for Experiment 1. Data concerning the grammatical and ungrammatical conditions were analyzed separately.

Procedure. This was the same as for Experiment 1 except for the task (grammatical decision). Participants were instructed to respond as rapidly and as accurately as possible if the target sequence was grammatically correct or not by pressing the right arrow (grammatical) or left arrow (ungrammatical) on their keyboard. As in Experiment 1, they were instructed to ignore the word sequences presented above and below the central target sequence.

Analysis. RTs were analyzed in the same way as for Experiment 1. We used (Generalized) Linear Mixed-Effects model (GLME) to analyze error rates (binomial data), with items and participants as crossed random effects (including by-item and by-participant random intercepts: Baayen et al., 2008) and with random slopes (Barr et al., 2013). The model was fitted with the glmer function from the lme4 package (Bates et al., 2015) in the R statistical computing environment (R Core Team, 2018). We report regression coefficients (b), standard errors (SE) and z-values for the fixed effects which were deemed reliable if $|z| > 1.96$ (Baayen, 2008).

Statistical Power. After excluding outliers, we obtained a dataset of 17,484 observations (see Result section for details) with 4,371 observations per condition which is largely above the threshold of 1,600 data points per condition recommended by Brysbaert and Stevens (2018). Using the SIMR package (Green & McLeod, 2016) in R environment (R Core Team, 2018), we obtained a power of 100% (IC 95% = 0.37) over 1000 simulations (for the correct sentences, comparing same vs. different distractors).

Results

We analyzed RTs and error rates for grammatical decisions. Prior to analysis we excluded 3 participants who performed with an accuracy lower than 75 %, as well as 2.34 % of trials with abnormally short or long RTs (i.e., < 100 ms and > 5000 ms). Condition means are shown in Table 2.

Response times

We excluded 4.51 % of trials with incorrect responses and then 3.35 % of trials with RTs lying beyond 2.5 standard deviations from the grand mean, leaving a total of 17,484 observations. For the correct sentences, we observed a significant effect of Distractor Type ($b = -0.030$; $SE = 3.80$; $t = -8.09$) with faster RTs in the same distractors condition compared with different distractors. The effect of Distractor Type was not significant for the ungrammatical sequences ($b = 0.80^3$; $SE = 2.53^4$; $t = 0.31$).

Error rates

There was a significant effect of Distractor Type for correct sentences ($b = 0.55$; $SE = 0.14$; $z = 3.91$), with fewer errors in the presence of same distractors ($M = 4.96$ %; 95 % IC = 0.60) compared to different distractors ($M = 7.04$ %; 95 % IC = 1.22).

For the ungrammatical sequences, there was also a significant effect of Distractor Type ($b = -0.65$; $SE = 0.20$; $z = -3.25$), but this time with more errors in the condition where distractors were the same ($M = 4.24\%$; $95\% \text{ IC} = 0.55$) compared to the condition where they were different ($M = 2.17\%$; $95\% \text{ IC} = 0.56$).

Table 2. Mean RT and error rate per experimental condition, and net effects of Distractor Type (same words vs. different words) in Experiment 2.

Response time (ms)	<i>Grammatical</i>		<i>Ungrammatical</i>	
	Mean	95% CI	Mean	95% CI
Same words	1399	13.40	1489	9.97
Different words	1492	16.47	1493	9.37
Effect	93		4	
Error rate (%)	Mean	95% CI	Mean	95% CI
Same words	4.96	0.60	4.24	0.55
Different words	7.04	1.22	2.17	0.56
Effect	2.08		-2.07	

Signal Detection Analysis

Although our main dependent variable was RT, and the effects of distractor type were only significant for the grammatical sequences in these analyses, the error rate analyses revealed a

different pattern of effects for the grammatical and ungrammatical sequences. Significantly fewer errors were made to grammatical sequences in the presence of same distractors, whereas significantly more errors were made to ungrammatical sequences in the presence of same distractors (see Table 2). Therefore, it might be the case that same word distractors would bias participants to respond “grammatical” and different word distractors would bias participants to respond “ungrammatical”. We examined a potential role for such response biases in driving the pattern of effects seen in Experiment 2 by using signal detection theory (Macmillan & Creelman, 2004). The sensitivity (d') measure obtained with this analysis corrects for response bias by combining probability correct grammatical responses (i.e., hits) with the probability of responding “grammatical” to ungrammatical sequences (i.e., false alarms). The obtained measure of sensitivity was practically identical for the same word distractors condition (3.39) and the different word distractor condition (3.45). This therefore allows us to reject a contamination of the key results (seen in RTs) by response bias.

Discussion

The RT results of Experiment 2 concerning correct sentences replicate the findings of Experiment 1 and moreover show the same pattern in error rates. Grammatical decisions were faster and more accurate when the target sentence was bounded above and below by the same sentence compared with a different sentence. Moreover, we note that average RTs for silent reading for meaning (Experiment 1) and grammatical decisions to correct sentences (Experiment 2) are quite similar, thus providing additional support that the grammatical decision task offers an interesting new window on sentence reading time.

RTs to ungrammatical sequences revealed a non-significant effect in the same direction as seen with grammatical sequences. Error rates, on the other hand, showed an effect in the opposite direction to that seen with grammatical sequences, hence pointing to a possible response bias – that is a tendency to respond “grammatical” with same word distractors. A signal detection analysis, correcting for response bias, nevertheless revealed very similar values of sensitivity (d') in the same word and different word distractor conditions.

Having established a significant impact of distractor words placed above and below a central, horizontally aligned, target sentence, we now sought to determine the nature of the mechanisms driving this effect. One possibility is that it is the lexical status of the different word distractors that is critical, with these different words interfering in the reading of central target words. In Experiment 3 we put this possibility to test by manipulating the lexicality of the different distractors, contrasting different words with nonwords that were matched in length but orthographically unrelated to the target words. Participants performed a grammatical decision task as in Experiment 2.

Experiment 3

Method

Participants. Sixty participants (27 reported being women) were recruited online using the Prolific website (Palan & Schitter, 2018). All reported being native French speakers and they received monetary compensation (£7.5/hour) for their participation. The participants ranged in age from 18 to 52 years ($M = 28.06$ years, $SD = 8.22$) and they were naïve to the purpose of the experiment.

Design & Stimuli. These were the same as for Experiment 2, except that an additional 100 sentences and 100 ungrammatical sequences were created due to the addition of a third level of the Distractor Type variable (in order to have the same number of items per condition as in the previous experiments). The third type of distractor sequence consisted of unpronounceable nonwords matched in length with the words of the target sequence. The nonword distractors were orthographically unrelated to the target words. Thus, for both the grammatical and ungrammatical sequences, the Distractor Type factor had 3 levels (same words vs. different words vs. nonwords), and the Latin-square involved three groups of participants. The statistical analyses involved pairwise (same word vs. different word; different word vs. nonword).

Procedure. This was the same as for Experiment 2.

Statistical Power. After excluding outliers, we obtained a dataset of 10,617 observations (see Result section for details) with 1,769 observations per condition which is above the threshold of 1,600 data points per condition recommended by Brysbaert and Stevens (2018). Using the SIMR package (Green & McLeod, 2016) in R environment (R Core Team, 2018), we obtained a power of 100% (95 % CI = 0.37) with over 1000 simulations.

Results

We analyzed RTs and error rates in the grammatical decision task. We first excluded 3 participants with an average accuracy lower than 75 %, as well as 1.29 % of trials with abnormally short or long RTs (i.e., < 100 ms and > 5000 ms). The remaining dataset was composed of 11,815 observations. Condition means are shown in Table 3.

Response times

We excluded trials with incorrect responses (6.98 %) and then 3.39 % of trials that were +/- beyond 2.5 standard deviations from the grand mean, leaving us with a total of 10,617 observations.

For the correct sentences, we found a significant difference between same word and different word distractors ($b = -0.02$; $SE = 0.003$; $t = -8.46$) with faster RTs with identical distractors relative to different distractors. However, the difference between the different word distractors and the nonwords distractors was not significant ($b = -0.006$; $SE = 0.003$; $t = -1.68$). There were no significant differences in the pairwise contrasts for the ungrammatical sequences (Same words vs. Different words: $b = -0.007$; $SE = 0.004$; $t = -1.72$; Different words vs. Nonwords: $b = -0.008$; $SE = 0.004$; $t = -1.95$).

Error rates

For the correct sentences, the difference between the same word distractors and the different word distractors was not significant ($b = 0.20$; $SE = 0.16$; $z = 1.24$), and neither was the difference between the different word and the nonword distractor conditions ($b = -0.21$; $SE = 0.15$; $z = -1.40$). For the ungrammatical targets, we found a significant difference between the same word distractors and the different word distractors ($b = -0.44$; $SE = 0.13$; $z = -3.27$), with fewer errors in the condition where distractors were different compared to the condition where they were identical. There was no significant difference between the different word and the nonword distractors ($b = -0.13$; $SE = 0.13$; $z = -0.95$).

Table 3. Mean RT and error rates for the three types of distractors in Experiment 3.

Response time (ms)	<i>Grammatical</i>		<i>Ungrammatical</i>	
	Mean	95% CI	Mean	95% CI
Same words	1358	17.96	1520	19.33
Different words	1458	22.80	1555	18.05
Nonwords	1420	18.16	1521	18.82
Error rate (%)	Mean	95% CI	Mean	95% CI
Same words	4.82	0.89	9.93	1.40
Different words	5.63	0.96	7.09	0.86
Nonwords	6.42	0.80	7.97	1.07

Discussion

The results of Experiment 3 replicated the difference between same word and different word distractors found in Experiments 1 and 2 in RTs to correct sentence targets. Crucially, at the same time there was no significant difference between the different word and the nonword distractors. We can therefore reject the distractor lexicality account of the observed difference between same word and different word distractors, according to which the effect is driven by different word distractors interfering in target sentence reading due to their lexical status. An alternative explanation is that the distractor interference effect is driven by mismatching visual features between target and distractor words. Experiment 4 was designed to test this alternative explanation, by manipulating the letter case of distractor sequences. Given that visual feature mismatch is greater with different case – same word distractors (e.g., a-A vs. a-a) compared with same case – same word distractors, we predicted that the former should

hinder target processing more than the latter (see Fournet et al., 2022, for a study showing effects of letter case in the grammatical decision task).

Experiment 4

Method

Participants. One hundred participants (51 reported being women) were recruited online using the Prolific website (Palan & Schitter, 2018). They all reported being native French speakers and received monetary compensation (£7.5/hour) for their participation. The participants ranged in age from 18 to 51 years ($M = 26.71$ years, $SD = 7.24$) and they were naïve to the purpose of the experiment.

Design & Stimuli. We used the same sentences and ungrammatical sequences as Experiment 2 and we manipulated Distractor Type (same words vs. different words) in the same way as Experiment 2. Experiment 4 included a manipulation of Distractor Case (lowercase vs. uppercase). Target sequences were always presented in lowercase. This led to a 2 (Distractor Type) X 2 (Distractor Case) factorial design. The data concerning the grammatical and ungrammatical conditions were analyzed separately.

Procedure. This was the same as for Experiments 2 and 3.

Analysis. The results were analyzed in the same way as for the previous Experiments. We used sum coding in order to evaluate main effects and their interaction, and the *emmeans*

function (Lenth et al., 2019) to evaluate the significance of the main effects of Distractor Type and Distractor Case if there was a significant interaction between these factors.

Statistical Power. After excluding outliers, we obtained a dataset of 18,708 observations (see Result section for details) with 2,338 observations per condition which is largely above the threshold of 1,600 data points per condition recommended by Brysbaert and Stevens (2018). Using the SIMR package (Green & McLeod, 2016) in the R environment (R Core Team, 2018), we obtained a power of 100% (IC 95% = 0.37) with over 1000 simulations (for the correct sentences and the contrast between same vs. different distractors).

Results

We excluded 3 participants who performed with an accuracy lower than 75 % and 3.56 % of trials with abnormally short or long RTs (i.e., < 100 ms and > 5,000 ms). Condition means are shown in Table 4 (RTs) and in Table 5 (error rates).

Response times

We excluded 4.76 % of trials with incorrect responses and then 3.48 % of trials with RTs lying beyond 2.5 standard deviations from the grand mean, leaving a total of 17,196 observations.

For the correct sentence targets, we observed a significant effect of Distractor Type ($b = -0.03$; $SE = 0.003$; $t = \mathbf{9.94}$), with faster RTs in the condition with same distractors compared with different distractors. We also found a significant effect of Distractor Case ($b = 0.008$; $SE = 0.003$; $t = \mathbf{2.38}$) with longer RTs when distractors were in uppercase compared with lowercase distractors (targets were always in lowercase). Crucially, there was also a

significant interaction ($b = 0.02$; $SE = 0.004$; $t = \mathbf{4.93}$), with the effect of Distractor Type being greater with same case distractors ($b = 0.03$; $SE = 3.62 \times 10^{-3}$; $t = \mathbf{9.94}$) compared with different case distractors ($b = 0.01$; $SE = 3.65 \times 10^{-3}$; $t = \mathbf{3.21}$).

For the ungrammatical sequences, the effect of Distractor Type was not significant ($b = 0.02$; $SE = 0.01$; $t = 1.75$) and the effect of Distractor Case was not significant ($b = 0.003$; $SE = 0.01$; $t = 0.28$). The interaction was also not significant ($b = 0.01$; $SE = 0.01$; $t = 0.83$).

Table 4. Mean RT (ms) per experimental condition and net effects of Distractor Type in Experiment 4.

	<i>Grammatical</i>		<i>Ungrammatical</i>	
	Mean	95% CI	Mean	95% CI
Same case				
Same words	1385	27.65	1476	26.49
Different words	1493	29.04	1427	21.89
Effect	108		-49	
Different case				
Same words	1421	23.01	1432	21.61
Different words	1467	26.87	1405	24.04
Effect	46		-27	

Error rates

There was a significant effect of Distractor Type for the correct sentence targets ($b = 0.36$; $SE = 0.13$; $z = 2.71$), with fewer errors in the presence of same distractors compared with different distractors. The effect of Distractor Case was not significant ($b = -0.19$; $SE = 0.13$; $z = -1.45$) and neither was the interaction between the two factors ($b = 0.05$; $SE = 0.18$; $z = 0.27$).

Table 5. Mean error rates (%) per experimental condition and net effects of Distractor Type in Experiment 4

	<i>Grammatical</i>		<i>Ungrammatical</i>	
	Mean	95% CI	Mean	95% CI
Same case				
Same words	5.56	1.07	5.82	1.18
Different words	7.33	1.30	1.12	0.64
Effect	1.77		-4.7	
Different case				
Same words	6.10	1.00	2.57	0.72
Different words	8.48	1.43	1.66	0.82
Effect	2.38		-0.91	

For the ungrammatical sequences, there was also a significant effect of Distractor Type ($b = -1.73$; $SE = 0.37$; $z = -4.69$), but this time with more errors in the condition where distractors were the same compared to the condition where they were different. The effect of Distractor Case was not significant ($b = -0.32$; $SE = 0.44$; $z = -0.72$), but the interaction between the two factors was significant ($b = 1.37$; $SE = 0.52$; $z = 2.64$). The effect of Distractor Type was greater with same case distractors ($b = 1.74$; $SE = 0.37$; $z = 4.69$) compared with different case distractors ($b = -0.36$; $SE = 0.36$; $z = 0.98$).

Discussion

The results of Experiment 4 show that the impact of different word distractors relative to same word distractors is modulated by whether the distractors are in the same case as the target sequence or not. Although there were significant effects of distractor type in the different case condition (in RTs to correct sentence targets), the effects were significantly greater in the same case condition. These results provide support for a role for mismatching visual features in driving the same word distractor effects seen in the preceding experiments. The fact that there was also an effect with different case distractors can be accounted for by the visual overlap between lowercase and uppercase letters, albeit less so than with same case letters. We interpret this finding in terms of a low-level crowding effect, the precise nature of which will be discussed below.

General Discussion

The present study examined the impact of stimuli lying one line above and one line below the line of text that was to be read by participants. Based on prior work with isolated words, we

expected a negligible influence of distractor stimuli lying above or below the target word, contrary to the robust effects of distractors that are horizontally aligned with the target (Snell, Mathôt et al., 2018). On the basis of work on crowding and visual object identification (e.g., Pelli & Tillman, 2008), however, we hypothesized that vertically aligned distractors should have some impact on reading, and that this influence might be best observable during sentence reading (as opposed to single word reading). That is, the impact of vertically oriented crowding might be negligible in single word recognition, but this impact might be observable when reading a complete sentence, given the cumulative impact over a sequence of words.

In Experiment 1, we asked participants to read sentences for meaning (pressing a response key when finished), and we checked for reading comprehension approximately every 10 sentences. Target sentences occupied a single line of text and were bounded one line above and one line below either by the same sentence or a different sentence composed of the same number of words of the same length (see Figure 1). Sentence reading was facilitated by the presence of same sentence distractors. In Experiments 2-4 we used the grammatical decision task (decide as rapidly and as accurately as possible if a sequence of words forms a grammatically correct sentence or not) as an alternative measure of sentence processing speed. The results concerning grammatical decisions to correct sentences in Experiment 2 replicated the findings of Experiment 1. Moreover, the grammatical decision RTs to correct sentences in Experiment 2 aligned well with the silent reading for meaning RTs obtained in Experiment 1 (average RT in Experiment 1 = 1518 ms; average RT in Experiment 2 = 1446 ms).

Building on the novel finding of an effect of distractor type (same words vs. different words) in Experiments 1 and 2, in the following Experiments we investigated the precise nature of this effect by manipulating the lexical status of distractors (Experiment 3), and the letter case of distractors (Experiment 4). The key results are that the lexical status of distractors had no significant influence, whereas the case of distractors did. That is, the interfering effect of different distractors (compared with the same words condition) did not depend on whether the different distractors were words or nonwords. On the other hand, RTs to grammatically correct sentences were hindered more by different case distractors than same case distractors. Overall, these findings point to a low-level locus of the interfering effect of different distractors found in the present experiments. We return to discuss the possible nature of this low-level interference below. First, we must consider the possible role played by response biases in the grammatical decision task.

We first remind readers that the main dependent variable in the speeded version of the grammatical decision task is RT. However, the results of Experiments 2-4 using this task suggested that there might be a response bias, such that same word distractors would bias participants to respond “grammatical” and different word distractors would bias participants to respond “ungrammatical”. We examined a potential role for such a response bias in Experiment 2, that implemented a simple “same word” vs. “different word” distractor manipulation. To do so we used an analysis based on signal detection theory (Macmillan & Creelman, 2004). The sensitivity (d') measure obtained with this analysis corrects for response bias by combining probability correct grammatical responses (hits) with the probability of responding “grammatical” to ungrammatical sequences (false alarms). The d' values obtained in this analysis were very similar in the same word and different word

distractor conditions. We therefore conclude that it is unlikely that response biases were affecting the results obtained with our main dependent variable which was RT.

Our general conclusion is that the interfering effects of different distractors are crowding effects. However, the precise mechanisms underlying crowding remain hotly debated (see e.g., Herzog & Manassi, 2015; Herzog et al., 2015). In the Introduction we mentioned the studies summarizing investigations of the spatial extent of crowding using a critical spacing manipulation (e.g., Pelli et al., 2007). Associated with this important empirical work, inspired by the seminal work of Bouma (1970), is the theoretical notion that crowding reflects the excessive pooling of visual features lying outside the limits of a crowding zone the shape and size of which is determined by several factors. One such factor is notably the radial-tangential asymmetry of the zone of uncrowded perception (Toet & Levi, 1992) which accounts for the greater horizontal than vertical extent of the uncrowded region for reading horizontally aligned text. Spatial pooling of visual features impairs visual object identification, hence the impact of vertically aligned flankers in the present study when these do not match the text being read. However, as noted in the reviews of Herzog and colleagues (Herzog & Manassi, 2015; Herzog et al., 2015), there is evidence that additional mechanisms play a key role in crowding. According to one such mechanism, referred to as grouping, feature-level interactions can be determined by more global characteristics of the visual scene, extending beyond the classic window determined by Bouma's law. Another such mechanism would be recurrent processing, whereby higher-level object characteristics determine on-going processing of low-level feature information. Examining exactly how such mechanisms might have impacted on the present results is beyond the scope of the present study, although we would point out that a potential impact of recurrent processing fits well with our cascaded-interactive approach to reading (e.g., Snell & Grainger, 2017).

One potential avenue for future research would be to examine how the vertical distractor effects seen in the present study are affected by healthy ageing and mild forms of Alzheimer's disease. It is well established that although more automatic processes are relatively spared in such populations, processes that require some forms of inhibitory control are impaired (Hasher & Zacks, 1988). This impairment is thought to be the cause of increased distractor interference effects during text reading in the afore-mentioned populations (Connelly et al., 1991; Duchek et al., 1998). In these studies, distractor stimuli were inserted in a passage of text to be read for meaning (60 distractors for a target text of approximately 125 words). Distractors were inserted in the target text (with no immediate repetitions of distractors) and were distinguished from the target text either by a change in font (Connelly et al., 1991) or by using lowercase italics for distractors vs. uppercase bold for the main text (Duchek et al., 1998). The presence of distractors significantly interfered with reading times (compared with a no distractor condition), and more so with semantically related distractors. Crucially, this effect of distractor interference was found to be significantly greater in the older participants and participants diagnosed with Alzheimer's disease, hence pointing to a role for inhibitory control in driving the effects. However, given the low-level nature of the distractor interference effects observed in the present study, we suspect that inhibitory control mechanisms are not involved. We hypothesize that such control mechanisms would more likely be involved when distractors occur on the same line of text and would be due to the parallel word processing that can occur in such conditions (see Snell & Grainger, 2019, for a summary of the evidence). That is, the simultaneous processing of irrelevant word identities while reading the target text would require some form of inhibitory control in order to maintain fluent reading.

In conclusion, we suspect that the impact of stimuli lying above or below a line of text being read is limited to influences driven by mechanisms associated with the general phenomenon referred to as crowding (Bouma, 1970; Herzog et al., 2015; Pelli & Tillman, 2008; Whitney & Levi, 2011). Of course, future in-lab experimentation with an eye-tracker will be necessary to know the extent to which readers move their eyes between lines, and therefore extent to which the present results might be due to a number of such vertical eye-movements operating during sentence reading. However, such eye movements should, in principle, facilitate orthographic influences rather than the purely visual influences that we observed. Moreover, given the hypothesized low-level nature of the effects we have reported, we are led to predict no role for inhibitory control in driving the effects of vertically aligned distractors.

Acknowledgments

This research was funded by ERC advanced grant 742141. We thank Manuel Perea for his advice concerning the statistical analyses.

Open Data

All data and scripts used for data analysis are available at: <https://osf.io/r67gm/>

References

- Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*, 67–82.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge: Cambridge University Press.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255-278.
- Bates, D. J., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*, 177-178.
- Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1).
- Chanceaux, M., Mathôt, S., & Grainger, J. (2013). Flank to the left, flank to the right: Testing the modified receptive field hypothesis of letter-specific crowding. *Journal of Cognitive Psychology*, *25*, 774-780.
- Connelly, S. L., Hasher, L., & Zacks, R. T. (1991). Age and reading: the impact of distraction. *Psychology and Aging*, *6*(4), 533.
- Dare, N., & Shillcock, R. (2013). Serial and parallel processing in reading: Investigating the effects of parafoveal orthographic information on non-isolated word recognition. *Quarterly Journal of Experimental Psychology*, *66*, 487-504.
- Duchek, J. M., Hunt, L., Ball, K., Buckles, V., & Morris, J. C. (1998). Attention and driving performance in Alzheimer's disease. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *53*(2), P130-P141.
- Ducrot, S. & Grainger, J. (2007). Deployment of spatial attention to words in central and peripheral vision. *Perception & Psychophysics*, *69*, 578-590.
- Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: a dynamical model of saccade generation during reading. *Psychological Review*, *112*, 777-813.
- Eriksen, C. W. (1995). The flankers task and response competition: A useful tool for investigating a variety of cognitive problems. *Visual Cognition*, *2*, 101-118.
- Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). LabVanced: a unified JavaScript framework for online studies. In *2017 International Conference on Computational Social Science IC2S2 (Cologne)*.

- Fournet, C., Mirault, J., Perea, M., & Grainger, J. (2022). Effects of letter case on processing sequences of written words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, in press.
- Grainger, J., Dufau, S., & Ziegler, J.C. (2016). A vision of reading. *Trends in Cognitive Sciences*, 20, 171-179.
- 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, 493–498.
- Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. *Psychology of learning and motivation*, 22, 193-225.
- Herzog, M.H., & Manassi, M. (2015). Uncorking the bottleneck of crowding: A fresh look at object recognition. *Current Opinion in Behavioral Sciences*, 1, 86–93.
- Herzog, M.H., Sayim, B., Chicherov, V., & Manassi, M. (2015). Crowding, grouping, and object recognition: A matter of appearance. *Journal of Vision*, 15, 5.
- Jordan, T. R., McGowan, V. A., Kurtev, S., & Paterson, K. B. (2016). A further look at post view effects in reading: An eye-movements study of influences from the left of fixation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 296-307.
- Legge, G. E. (2006). *Psychophysics of reading in normal and low vision*. CRC Press.
- Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). Package ‘emmeans’. R package version, 1(3.2).
- Lev, M., Yehezkel, O., Polat, U. (2014). Uncovering foveal crowding. *Scientific Reports*, 4, 4067.
- Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.
- McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, 17, 578-586.
- Mirault, J., & Grainger, J. (2023). Effects of vertically aligned flankers during sentence reading. Retrieved from osf.io/r67gm
- Mirault, J. & Grainger, J. (2020). On the time it takes to judge grammaticality. *The Quarterly Journal of Experimental Psychology*, 73, 1460-1465.
- Palan, S., & Schitter, C. (2018). Prolific. ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22-27.
- Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience*, 11, 1129–1135.

- Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision*, 7(2), 1-36.
- Pollatsek, A., Bolozky, S., Well, A. D., & Rayner, K. (1981). Asymmetries in the perceptual span for Israeli readers. *Brain and language*, 14, 174-180.
- R Core Team, R. (2018). R: A language and environment for statistical computing. 2018.
- Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin*, 85, 618-660.
- Rayner, K., Well, A. D., & Pollatsek, A. (1980). Asymmetry of the effective visual field in reading. *Perception & Psychophysics*, 27, 537-544.
- Snell, J., & Grainger, J. (2017). The sentence superiority effect revisited. *Cognition*, 168, 217-221.
- Snell, J. & Grainger, J. (2018). Parallel word processing in the flanker paradigm has a rightward bias. *Attention, Perception, & Psychophysics*, 80, 1512-1519.
- Snell, J., & Grainger, J. (2019). Readers are parallel processors. *Trends in Cognitive Sciences*, 23(7), 537-546.
- Snell, J., Mathôt, S., Mirault, J., & Grainger, J. (2018). Parallel graded attention in reading: A pupillometric study. *Scientific Reports*, 8 (1), 3743.
- Toet, A., & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research*, 32, 1349-1357.
- Vejnović D, Zdravković S. (2015). Side flankers produce less crowding, but only for letters. *Cognition*, 143, 217-227.
- Whitney, D., & Levi, D. M. (2011). Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, 15, 160–168.
- Winsler, K., Grainger, J., & Holcomb, P.J. (2022). On letter-specific crowding and reading: Evidence from ERPs. *Neuropsychologia*, 176, 108396.