On idle idols and ugly icons: Do homophones create interference in typing?

Merel Muylle (merel.muylle@ugent.be) and Robert J. Hartsuiker (robert.hartsuiker@ugent.be)
Department of Experimental Psychology, Ghent University, H. Dunantlaan 2
9000 Gent, Belgium

Nazbanou Nozari (bnozari@iu.edu)
Department of Psychological and Brain Sciences, Indiana University, 1101 E. 10th Street
Bloomington, IN 47405-7007, USA

Abstract

This study investigates whether homophone competitors are activated during typewriting and to which extent such activation is modulated by syntactic category. In two experiments, we compared the typewriting of homophone pairs in high vs. low conflict sentences (i.e., both homophones vs. only one homophone in the sentence, respectively) in a sentence dictation task (Experiment 1) and in a question-answering task (Experiment 2). The homophone pairs either belonged to the same or different syntactic categories. In Experiment 1, we found a homophone interference effect in accuracy, independent of conflict and syntactic category. In Experiment 2, this effect was replicated, but in addition, participants were slower to type homophones in a high vs. a low conflict context. Our results show a robust, lexically-situated homophone interference effect, regardless of conflict and syntactic category, but when deeper processing of the sentence is involved, conflict starts to play a role.

Keywords: typewriting; homophones; sentence production

Introduction

Most of us have experienced replacing words with a (heterographic) homophone (e.g., "there", "they're", and "their") when writing or typing. In fact, analyses of writing samples have shown that phonological errors and in particular homophone substitution errors are the most common type of errors (about 25% of the errors in university students) that end up in our final texts (Aitchison & Todd, 1982; Lastres López & Manalastas, 2018). A study by Bonin, Peereman, and Fayol (2001) suggests that this homophone interference effect is not lexical, but sublexical in nature. They compared the onset latencies and error rates of French homophones and matching control words in a handwritten timed picture naming task. Although there was a difference between homophones and controls in terms of accuracy, this effect disappeared when the authors controlled for orthographic knowledge (i.e., whether participants knew how to write the word correctly in an untimed naming task). Follow-up experiments revealed that sound-to-letter inconsistencies at the onset of words caused interference, suggesting that there is an early, sublexical effect of phonology in the writing process.

At first glance, the findings of Bonin et al. (2001) would imply that lexical competition plays no role in the homophone interference effect in writing. But there are two findings that challenge this conclusion: (a) Increasing lexical competition by using primes that overlap in spelling with the non-target homophone has been reported to increase the rate of homophone errors in handwriting (e.g., the prime "teacher" increased writing "beach" instead of the target "beech"), especially for the subordinate homophone form (White et al., 2008). Although this finding may be the result of sublexical competition (i.e., priming "ea" may elicit the substitution of "ee" by "ea", regardless of the word), the observation that this effect was the strongest in the less frequent, subordinate homophones is more compatible with a lexical locus of the effect. (b) In a similar paradigm using typewriting, homophone errors have been observed when the two homophones belong to the same syntactic category, but not when they belong to different syntactic categories (White et al., 2010). Since syntactic selection precedes sublexical encoding (e.g., Dell, 1984), this effect is compatible with a lexical locus of competition. Incidentally, most pairs in Bonin et al. (2001) were between-category homophones, which may have contributed to a null result.

The studies reported above paint a mixed and incomplete picture. First, a main homophone effect seems to be present or absent in different studies, which may or may not stem from methodological differences across these studies. Second, a lexical locus of the effect is implied in some studies but not others. Finally, some of the empirical reports contradict everyday observations. For example, the oft-committed mistake of misspelling “there”/“their” occurs despite the fact that the two words belong to different syntactic categories, contradicting the findings of White et al. (2010). Collectively, these contradictions call for a better controlled study of homophones.

Present study

In this study, we studied whether there is a (lexical) homophone interference effect during typewriting and whether this effect is modulated/constrained by syntactic category and conflict (i.e., whether the homophone competitor is present in the sentence or not). In Experiment 1, we used a sentence dictation task to elicit the writing of homophone (H) pairs and matched control (C) words under four conditions (see 1a-1d):

(1a) homophone-homophone:
The fashion idol turned out to be quite idle.
(1b) homophone-control:
The fashion idol turned out to be quite ugly.
(1c) control-homophone:

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

The fashion <u>icon</u> turned out to be quite <u>idle</u>.
(1d) control-control:
The fashion <u>icon</u> turned out to be quite <u>ugly</u>.

Half of the homophone pairs belonged to the same and the other half to a different syntactic category. For each target word, we measured typing times and accuracy.

If lexical competition drives the homophone interference effect, we would predict three findings: (a) longer onset latencies, longer durations, and more errors for homophones compared to matched controls, even after controlling for potential confounds such as orthographic transparency, (b) longer latencies and lower accuracy for homophones in high-conflict sentences (i.e., with two homophones) than in low-conflict sentences (i.e., with only one homophone), and (c) longer latencies and lower accuracy for homophones belonging to the same vs. different syntactic categories. If none of these predictions are borne out, this would imply that the homophone interference effect is an artefact of orthographic non-transparency and lack of spelling knowledge, as claimed by Bonin et al. (2001).

Experiment 1

Methods[1]

Participants We collected data from 124 native speakers of American English (18-40 years old), recruited through Prolific and paid £4.50 for their participation. The sample size was determined by means of a power simulation using the mixedpower package in R (Kumle et al., 2021). If we assume a medium effect size ($d = 0.50$), 124 participants and 80 targets would yield a power of .86 to observe the interaction effect in onset latencies. Before being admitted to the study, candidates completed a typing prescreening test to ensure they were proficient typewriters (e.g., Pinet & Nozari, 2021, 2022).

Materials We selected 40 English homophone word pairs, of which 20 belonged to the same syntactic category, and 20 to a different category. To assess category membership, we used the part-of-speech information from the SUBTLEX-US database (Brysbaert et al., 2012). For the within-category pairs, we only included pairs that belonged to the same category in (about) 100% of the instances, whereas for the different-category pairs, we ensured that there was no overlap possible. For instance, the pair 'brake'-'break' was not included because both words could either be used as a noun or a verb, which makes it impossible to determine whether syntactic category puts a constraint on the competition or not.

For each homophone, we selected a control word that was closely matched in terms of consonant-vowel structure (including gemination), word length, number of syllables, number of phonemes, and syntactic category. We also ensured that, on average, there was no difference in word frequency (Zipf-score, taken from SUBTLEX-US; van Heuven et al., 2014) and phonology-to-orthography (PO) consistency (Chee et al., 2020). For this final measure, we averaged the onset and rime PO token consistency values for each syllable of the target word.[2] Furthermore, the matching controls always started with a letter that is typed with the same hand as the homophone in order to have no hand transition differences across conditions. Since we also compared within and between syntactic category homophone pairs, we ensured that, on average, stimuli in these conditions did not differ in terms of word frequency, orthographic Levenshtein distance between pairs, and PO consistency.

For each homophone pair and their matched controls, we created a sentence context in which all combinations (homophone-homophone, control-homophone, homophone-control, and control-control) would result in a plausible sentence. As such, there were 40 sets of sentences appearing in four different conditions. We ensured that the average number of words in between targets was similar across homophone pairs from the same vs. from a different category with a minimum of one word. In order to ensure that there was no difference in comprehensibility of sentences across conditions, we conducted a web-based norming study in which 40 native speakers of English judged the acceptability of spoken sentences. The sentences were recorded by means of the AI-generated speech tool Descript (using a female stock voice called 'Nancy'). The 160 sentences were counterbalanced across four lists. Each participant judged 80 target sentences and 80 filler sentences that were not acceptable in terms of their semantics. Based on the norming study, we replaced five item sets containing low frequency homophones with more frequent alternatives. Reaction times analyses from the same norming study showed no significant difference across sentences with one homophone, two homophones, and controls only, indicating that sentences were equally intelligible across conditions.

We created four lists of 40 sentences using a crossed design (i.e., from each set, one version appeared per list in a way that the conditions were counterbalanced across lists). Participants were randomly assigned to one of the four lists based on their participant number.

Procedure The experiment was programmed in JavaScript using the jsPsych library (de Leeuw, 2015) and hosted online via the JATOS platform (Lange et al., 2015) on a MindProbe server. After giving their informed consent, the participants completed the typing prescreening test. Those that passed completed a short demographic questionnaire.

Next, they completed the sentence dictation task, in which they typed down the sentence they heard as fast and accurately as possible after hearing a beep sound. They were prompted not to use any capitalization or punctuation. There were four practice trials. Each trial started with a fixation cross that was presented for 700 ms in the center of the screen. Next, the audio started playing while the response box appeared in the center of the screen. When the audio stopped,

---

[1] Both Experiment 1 and 2 were preregistered on the Open Science Framework.

[2] We chose to use the token over the type consistency measure since it corrects for word frequency in the calculation.

a beep sound indicated that participants could initiate typing. Errors could be corrected during typing, but all keypresses were registered. Participants pressed ENTER to continue with the next trial. Finally, a blank screen appeared for 600 ms before continuing with the next trial. There were 40 experimental trials in total, with a short break after 20 trials.

Finally, there was a spell-check task at the end. This was implemented as a multiple-choice task in which participants were visually presented with the sentences from the dictation task, but with the target homophone left out. Participants were asked to select the correct sentence completion out of three options, the target, its actual homophone and a misspelled homophone (e.g., "The fashion ____ turned out to be quite ugly." Options: 'idol', 'idle', and 'idel'). For sentences that appeared in the homophone-homophone condition in the sentence dictation task, the same sentence appeared on two different trials with one of both targets left out and the other one replaced by the control word, in order to avoid influences from the competitor spelling being presented in the sentence. As such, this task also consisted of 40 trials. Once they finished, participants received the completion code and were paid.

Analyses We collected accuracy data, onset reaction times (RTs), and inter-key-intervals (IKIs) for all target words in the sentence dictation task. Since each sentence contained two target words, individual responses were scored once for the first and once for the second target. Onset RTs, mean IKIs, and accuracy scores were only included in the analyses if the targets were selected correctly in the spell-check task.

Responses to a target word were only scored as correct if no errors were made on the first typing attempt. Errors were divided into five broad categories (see Table 1 for examples): a) motor errors (when an adjacent key was hit instead of/in addition to the target key), b) planning errors (when the target letter was replaced by a letter that appeared in another position of the word, including transpositions, or by a letter that appeared in the same position in another word), c) phonological errors (phonologically plausible additions and substitutions), d) other errors, subdivided into mishearings (when the target was replaced by another word or not produced at all), conjugation errors (e.g., plurals, different tenses), segmentation errors, additions and substitutions (that did not fit motor or planning error criteria), and alternative errors (when the control word was replaced by the homophone or vice versa), e) ambiguous errors (when more than one category was possible, e.g., deletions, doubling errors), and f) mixed errors (when a target contained more than one error, belonging to different categories). Ambiguous and mixed errors were subdivided into phonological and non-phonological errors. Responses that were categorized as mishearings or alternative errors were excluded from the accuracy analysis, since participants never intended to produce the target word in these situations.

Onset RTs were operationalized as the time (in ms) needed to type the first letter of the target after pressing the spacebar, whereas the mean IKIs (in ms) were calculated by adding all individual IKIs of the target word and dividing them by the number of letters. For both, we only included trials for which the responses were correct from the start of the sentence until the end of the target word on the first attempt. Outlier RTs and mean IKIs (> 2.5 SDs above the mean) were discarded from the analyses.

The data were analyzed using (generalized) linear mixed effects models by means of the lme4 (Bates, Mächler, et al., 2015) and afex packages (Singmann et al., 2016) in R. Factors were always sum coded. In a first model, that tested the homophone interference effect and the effect of syntactic category, the fixed effects consisted of the Word Type-by-Category interaction and a main effect of Zipf score (added as control variable). Since we wanted to study the main effect of Conflict independently, this was tested in a separate model. We always started from the full model, but when it did not converge, we reduced the random effects model following the recommendations by Matuschek et al. (2017). For main and interaction effects, we report the χ^2 omnibus test results. In case of significant effects involving more than two conditions, we performed pairwise contrasts using the emmeans package (Lenth et al., 2020) to find out which conditions drove the effect. We also conducted an exploratory analysis of the errors by performing a global χ^2 test of homogeneity on the distribution of errors, followed by pairwise proportion tests in the H vs. C condition (applying Bonferroni correction).

Results

We excluded two homophone target items (124 observations or 2.5%) for which accuracy was < 50%. Also, 307 (6.2%) items that were selected incorrectly in the spell-check task were excluded. Furthermore, we excluded 136 (2.7%) errors in the homophone condition and 235 (4.7%) errors in the control condition that were mishearings or alternative errors. As such, the final dataset consisted of 4391 observations for the homophones (88.5%) and 4725 observations for the matched controls (95.3%). Of these, 2122 (42.8%) homophones and 2545 (51.3%) controls could be included in the onset RT analysis and 2122 (42.8%) homophones and 2594 (52.3%) controls in the IKI analysis.

Homophones vs. Controls Participants were less accurate when typing homophones ($M = .79, SD = .56$) vs. controls ($M = .88$, $SD = .46$; $\chi^2(1) = 22.02$, $p < .001$), but there was no effect of Word Type in onset RTs (H: $M = 198$, $SD = 160$; C: $M = 195$, $SD = 156$; $\chi^2(1) = 0.45$, $p = .50$) and mean IKIs (H: $M = 140$, $SD = 61$; C: $M = 140$, $SD = 55$; $\chi^2(1) = 0.39$, $p = .53$).

High vs. Low Conflict For accuracy, there was a main effect of Conflict ($\chi^2(2) = 125.09$, $p < .001$), but the pairwise comparisons showed no difference between high and low conflict ($Z = 0.64$, $p = .52$; see Figure 1, Panel A). There was no effect of Conflict in onset RTs ($\chi^2(2) = 0.63$, $p = .73$) and mean IKIs ($\chi^2(2) = 0.77$, $p = .68$).

Same vs. Different Category The interaction between Word Type and Category was not significant in any measure (accuracy: $\chi^2(1) = 0.06$, $p = .80$; onset RT: $\chi^2(1) = 0.04$, $p = .84$; mean IKI: $\chi^2(1) = 0.86$, $p = .35$), nor was there any main effect of Category (all *p*s > .29; see Figure 1, Panel B).

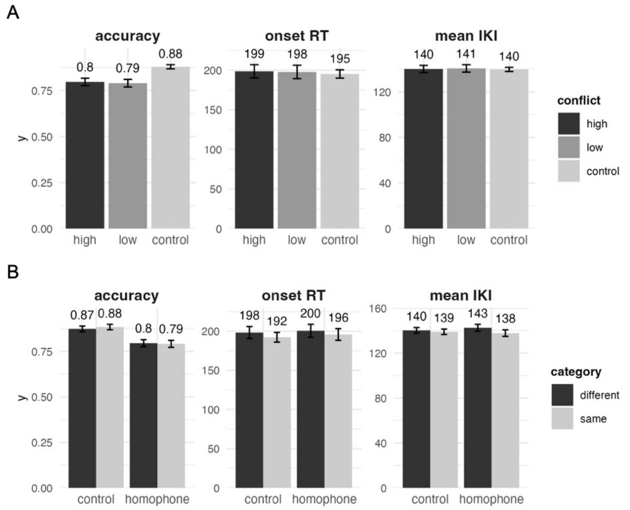

Figure 1. Comparison across conflict conditions (Panel A) and across syntax conditions (Panel B) in Experiment 1.

Exploratory error analysis The distribution of errors was significantly different across homophones and control words ($\chi^2(6) = 336.99$, $p < .001$). There were more phonological, ambiguous, and mixed errors, but fewer other errors in homophones than in controls (see Table 1). The higher proportion of ambiguous and mixed errors in homophones was due to a larger proportion of phonologically plausible errors, whereas the proportion of non-phonological errors was similar across both conditions.

Table 1. Comparison of errors in Experiment 1, illustrated with examples (Ex.) for the target "idol".

Error type	Ex.	H (%)	C (%)	$\chi^2(1)$	*p*
Ambiguous		6.4	3.4	44.77	<.001
-phonological	*idl*	3.7	0.8	-	-
-non-phonological	*odol*	2.7	2.7	-	-
Mixed		1.5	0.5	21.65	<.001
-phonological	*iidle*	1.2	0.2	-	-
-non-phonological	*dool*	0.3	0.3	-	-
Motor	*idok*	1.7	2.5	7.16	.052
Other		4.3	6.9	30.10	<.001
-addition	*itdol*	<0.1	<0.1	-	-
-alternative	*icon*	<0.1	0.1	-	-
-conjugation	*idols*	0.9	0.6	-	-
-mishearing	*model*	3.0	4.6	-	-
-segmentation	*i dol*	0.0	1.0	-	-
-substitution	*izol*	0.3	0.5	-	-
Phonological	*idle*	7.6	1.3	230.06	<.001
Planning	*diol*	1.6	1.7	0.03	1

Discussion

Experiment 1 showed that even after controlling for other factors in the design, participants still committed more errors during the typing of homophones compared to control words. This finding establishes a true homophone effect that cannot be attributed to confounds such as orthographic transparency or poor knowledge of spelling. The error analysis revealed that this difference was mostly driven by an increase in phonological errors in homophones, in line with our hypothesis. Interestingly, the effect was not sensitive to conflict or syntactic category. On the one hand, the finding of the homophone effect across syntactic categories matches the prevalence of cross-category errors, such as “there” and “their”. On the other hand, the finding is at odds with the report of White et al. (2010). Similarly, the absence of an effect of conflict, which is predicted by the lexical competition account, seems to contradict the findings of White et al. (2008).

One explanation for these differences could be the nature of our task. In a spelling-to-dictation task, there is a strong activation of phonological information (since the input is phonological in nature), while the processing of meaning is less important for this task. The shallow lexical-semantic processing of this task could weaken lexical competition, and consequently the influence of conflict and syntactic category. Moreover, when we type, we usually transform ideas and meanings into letter strings, so a spelling-to-dictation task is not the best representative of our normal typing activities. To address this concern, in Experiment 2, we designed a question-answering task, in which participants were instructed to type replies to questions based on a picture while repeating all elements from that question. To that end, we slightly modified the stimuli from Experiment 1 to create questions (see 2a-2d):

(2a) homophone-homophone:
Who shouted at the fashion idol that he was quite idle?
(2b) homophone-control:
Who shouted at the fashion idol that he was quite ugly?
(2c) control-homophone:
Who shouted at the fashion icon that he was quite idle?
(2d) control-control:
Who shouted at the fashion icon that he was quite ugly?

Figure 2. Example picture from the question-answering task (for the idol-idle pair).

(In this case, a picture of a shouting photographer was shown, see Figure 2, so the expected answer was "The photographer shouted at the fashion idol/icon that he was quite idle/ugly"). Although these questions were still orally presented, the participants had to process the meaning, retrieve new lexical items, and build a syntactic structure in order to answer them.

Experiment 2

Methods

Participants Another 124 participants took part in this experiment. The selection criteria (including the pre-screening typing test) were the same as in the previous experiment. A new power simulation in which we updated the values of the parameters based on the model output of Experiment 1 yielded a power of .96 to detect an interaction effect at d= 0.50.

Materials We used the same target words and similar target sentences as in Experiment 1. Some of the targets where replaced if the proportion correct responses in the spell-check task was < .50 (*N* =1) or if the proportion correct final responses in the sentence dictation task was < .50 (*N* = 6). For the replacement, we chose homophone-pairs and controls that had a higher frequency than the original targets.

For each target sentence, we created a Wh-question that did not query the target words, but another aspect of the sentence. As such, some target sentences were slightly altered, for instance by adding extra information (see 2a-2d). The Wh-questions were recorded in Descript, using the same voice as in Experiment 1.

Next, we created a picture for each set of questions (*N* = 40) by means of an open-source AI text-to-image converter (dream.ai/create). We ensured that the same picture was appropriate for all versions of the same question.

In order to check whether all questions and pictures were clear, we conducted a web-based norming study in which 40 native speakers of English were asked to answer the questions by using the picture. Based on the results, we adapted 18 pictures, mainly because there was low naming agreement (<.70 same response), and reformulated some of the questions to make the response more homogenous (e.g., 'Where' was changed to 'In which room'). Crucially, the participants gave a sensible response in at least 89% of the cases for each item, showing that the questions were clear and relatively easy to answer based on the picture.

Procedure The procedure was identical to Experiment 1, except that instead of hearing a sentence, participants heard a question, and typed a response based on a picture. They were instructed to respond as fast and as accurately as possible and repeat all elements from the question in their response without capitalization or punctuation. Practice trials started with a fixation cross that was presented for 700 ms in the center of the screen. Next, the audio started playing and once finished, the picture appeared in the center of the screen, with the response box below. Once participants finished typing they pressed ENTER. They then saw the expected response appearing above their original responses, so they could compare. Finally, a blank screen appeared for 600 ms before continuing to the next trial. After five practice trials, participants continued onto the main task if at least 3/5 responses were correct. Otherwise, they got more blocks of five trials until they reached 3/5 accuracy in one block. There were 40 experimental trials divided into two blocks with a short break. Each block started with a catch trial, in which participants received feedback if their response was not correct in order to remind them to mention all elements from the question. These catch trials were not analyzed.

Analyses We ran the same analyses as in Experiment 1.

Results

One homophone item was excluded due to accuracy below 50% (62 observations or 1.2%). We excluded another 371 (7.5%) observations for the homophones that had erroneous responses on the spell-check task and 302 (6%) mishearings and alternative errors in the homophones and 439 (8.9%) of these errors in the controls. A total of 4225 (85.2%) observations for the homophones and 4521 (91.1%) observations for the controls were included in the final dataset. Of these, 1336 (26.9%) homophone observations and 1522 (30.7%) controls could be included in the onset RT analysis and 1313 (26.5%) homophones and 1503 (30.3%) controls in the IKI analysis.

Homophones vs. Controls Similar to Experiment 1, we observed a main effect in accuracy ($\chi^2(1) = 18.67$, $p < .001$), with participants committing more errors in the homophones ($M = .80$, $SD = .55$) than in the control words ($M = .85$, $SD = .50$). There was no main effect of Word Type in onset RTs (H: $M = 222$, $SD = 212$; C: $M = 220$, $SD = 192$; $\chi^2(1) = 0.00$, $p = .99$) and mean IKIs (H: $M = 149$, $SD = 75$; C: $M = 150$, $SD = 64$; $\chi^2(1) = 0.01$, $p = .94$).

High vs. Low Conflict For accuracy, there was a main effect of conflict ($\chi^2(2) = 43.14$, $p < .001$), but the difference between high and low conflict was not significant (Z = 1.95, p = .051; see Figure 3, Panel A).The effect of conflict was not significant in onset RTs ($\chi^2(2) = 4.99$, $p = .08$), but it was in mean IKIs ($\chi^2(2) = 15.83$, $p < .001$). Participants were slower to type the full homophone in the high vs. the low conflict condition ($t(2684.41) = 3.93$, $p < .001$).

Same vs. Different Category There was a main effect of Category in the mean IKIs ($\chi^2(1) = 6.26$, $p = .01$), which indicated that participants were slower in general when typing targets belonging to the same category sample compared to the different category sample (see Figure 3, Panel B). There was no Word Type-by-Category interaction effect in any measure (accuracy: $\chi^2(1) = 0.35$, $p = .56$; onset RT: $\chi^2(1) = 0.22$, $p = .64$; mean IKI: $\chi^2(1) = 0.60$, $p = .44$). Despite the absence of a significant interaction in accuracy on the group level, an exploratory analysis revealed a

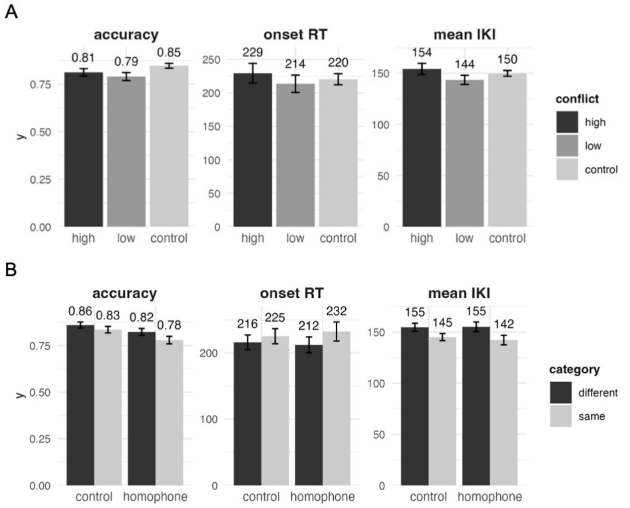

Figure 3. Comparison across conflict conditions (Panel A) and across syntax conditions (Panel B) in Experiment 2.

negative correlation between participants' overall accuracy and the difference score between categories in homophones ($r = -0.29$, $p = 0.001$), but not in controls ($r = -0.04$, $p = 0.62$). This shows that less accurate participants were more sensitive to competitors from the same vs. a different category than highly accurate participants.

Exploratory error analysis The error pattern was significantly different across both word types ($\chi 2(6) = 268.73$, $p < .001$), with more phonological and fewer other and motor errors in the homophones vs. controls (see Table 2).

Table 2: Comparison of errors in Experiment 2 illustrated with examples (Ex.) for the target "idol".

Error type	Ex.	H (%)	C (%)	$\chi^2(1)$	*p*
Ambiguous		4.7	3.9	3.63	.40
-phonological	*idl*	2.0	0.6	-	-
-non-phonological	*odol*	2.7	3.3	-	-
Mixed		1.0	0.7	1.72	1
-phonological	*iidle*	0.8	0.2	-	-
-non-phonological	*dool*	0.2	0.5	-	-
Motor	*idok*	1.8	3.3	19.88	<.001
Other		8.7	11.7	22.29	<.001
-addition	*itdol*	0.2	0.2	-	-
-alternative	*icon*	0.4	2.0	-	-
-conjugation	*idols*	1.2	1.5	-	-
-mishearing	*model*	6.3	6.9	-	-
-substitution	*izol*	0.7	1.1	-	-
Phonological	*idle*	7.0	1.0	227.08	<.001
Planning	*diol*	2.0	2.3	1.24	1

Discussion

Despite using a different task, Experiment 2 replicated the main result of Experiment 1, i.e., a significant homophone interference effect. The increase in errors for homophones vs. controls was again mainly situated in the phonological category. However, the change to the design revealed some differences from Experiment 1: participants were now slower to type homophones in high-conflict vs. low-conflict sentences. This finding indicates that the presence of the homophone competitor in the sentence makes it more challenging to type the correct alternative, in line with the lexical competition hypothesis. The effect of syntax was less clear and limited to generally less accurate participants.

General Discussion

We investigated homophone interference effects and their interaction with conflict and syntactic category in a sentence dictation task and in a question-answering task. Both tasks, despite controlling for a host of potentially confounding variables and only including items sanctioned by the spell-check task, elicited a robust homophone interference effect, reflected in higher error rates on homophones compared to controls. This finding is at odds with the report of Bonin et al.'s (2001) study, in which the homophone interference effect disappeared after controlling for orthographic knowledge. Instead, our finding points to an interference effect that arises due to competition from another lexical item with similar phonology.

The lexical competition hypothesis made two more predictions, namely, an increase of homophone interference in high-conflict contexts and for same-category homophones. Neither prediction was supported in Experiment 1. However, when the task was changed from spelling-to-dictation to producing sentences in response to a question, we observed some evidence of sensitivity to conflict, as predicted by the lexical competition account. The effect of syntax remained elusive, and only showed up in a post-hoc analysis; participants who were overall poorer in spelling were more affected by homophones within the same syntactic category than those with stronger spelling abilities. This finding is partially aligned with the report of White et al. (2010) for a role of syntax in homophone interference. However, it contrasts with that report in demonstrating that homophone interference, although potentially sensitive to syntactic category, is not strongly bound by syntax.

Collectively, our results suggest that homophones cause robust interference during typewriting, but the nature of this interference is modulated by task. When the task involves producing sentence from meaning, the influence of lexical competition on homophone interference becomes clear.

References

Aitchison, J., & Todd, P. (1982). Slips of the mind and slips of the pen. In R. N. St. Clair & W. von Raffler-Engel (Eds.), *Language and cognitive styles: Patterns of neurolinguistic and psycholinguistic development* (pp. 180–194). Swets & Zeitlinger.

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods*, *44*(4), 991–997. https://doi.org/10.3758/s13428-012-0190-4

Chee, Q. W., Chow, K. J., Yap, M. J., & Goh, W. D. (2020). Consistency norms for 37,677 English words. *Behavior Research Methods*, *52*(6), 2535–2555. https://doi.org/10.3758/s13428-020-01391-7

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Dell, G. S. (1984). Representation of serial order in speech: Evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(2), 222–233. https://doi.org/10.1037/0278-7393.10.2.222

Kumle, L., Võ, M. L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods*, *53*(6), 2528–2543. https://doi.org/10.3758/s13428-021-01546-0

Lange, K., Kühn, S., & Filevich, E. (2015). "Just another tool for online studies" (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0130834

Lastres López, C., & Manalastas, G. (2018). Errors in L1 and L2 university students' writing in English: Grammar, spelling and punctuation. *RAEL: Revista Electrónica de Lingüística Aplicada*, *16*(1), 118–134. http://hdl.handle.net/10347/19927

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2020). Package ' emmeans .' *CRAN Repository*.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Pinet, S., & Nozari, N. (2021). The role of visual feedback in detecting and correcting typing errors: A signal detection approach. *Journal of Memory and Language*, *117*, 104193. https://doi.org/10.1016/j.jml.2020.104193

Pinet, S., & Nozari, N. (2022). Correction without consciousness in complex tasks: Evidence from typing. *Journal of Cognition*, *5*(1), 1-14. https://doi.org/10.5334/joc.202

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521