



Dynamic lexical decisions in French: Evidence for a feedback inconsistency effect

Laura Barca, Giovanni Pezzulo, Marc Ouellet, Ludovic Ferrand

► To cite this version:

Laura Barca, Giovanni Pezzulo, Marc Ouellet, Ludovic Ferrand. Dynamic lexical decisions in French: Evidence for a feedback inconsistency effect. *Acta Psychologica*, 2017, 180, pp.23-32. 10.1016/j.actpsy.2017.08.005 . hal-02350570

HAL Id: hal-02350570

<https://hal.science/hal-02350570v1>

Submitted on 12 Nov 2019

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.

Dynamic lexical decisions in French: Evidence for a feedback inconsistency effect

Laura Barca¹, Giovanni Pezzulo¹, Marc Ouellet^{2, 3, 4}, and Ludovic Ferrand²

¹ Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy

² Université Clermont Auvergne, CNRS, LAPSCO, F-63000 Clermont-Ferrand, France

³ Department of Experimental Psychology, University of Granada, Granada, Spain

⁴ Center for Research on Mind, Brain, and Behavior (CIMCYC), University of Granada, Spain

Corresponding author

Laura Barca

Institute of Cognitive Sciences and Technologies,

National Research Council (ISTC-CNR), Rome

e-mail: laura.barca@istc.cnr.it

Telephone: + 39 06 44595312

Fax: + 39 06 44595243

Abstract

Numerous studies indicate that lexical decision is a dynamic process in which various sources of information (e.g., orthographic, phonological, lexical) need to be integrated over time and that mismatches between these sources produce uncertainty during decision-making. However, the specific contribution of these different sources of information to the decision process, and the specific ways they interact over time, are still incompletely understood. Here we report two experiments conducted to investigate the role of a key characteristic of language - *orthographic consistency* - in modulating reading strategies in a lexical decision task. We aimed to test different theories of how exactly orthographic consistency influences decisions, and the ways lexical and sublexical representations are matched within it. For this, we exploited a key characteristic of the French language: the fact that retrieving spelling from phonology is hard, whereas the reverse is not the case (i.e., French is feedback-inconsistent but feedforward-consistent). Our results indicate a feedback inconsistency effect: uncertainty in our lexical decision task originates from the process of retrieving spelling from phonology, not when retrieving phonology from spelling. These results support the idea of a verification mechanism that compares orthographic sublexical information with lexical information during lexical decisions.

Keywords

Visual lexical decision, mouse tracking, orthographic consistency, pseudohomophones, feedback inconsistency.

PsycINFO classification codes

2300 Human Experimental Psychology; **2340** Cognitive Processes

1. Introduction

Lexical decision is a classic paradigm used to study the way we process linguistic stimuli. Despite numerous studies, a number of unsolved problems remain, such as, for example, the way different sources of information (e.g., lexical, orthographic) are combined to produce a lexical decision.

In a recent study, Barca and Pezzulo (2012) investigated whether the mechanisms involved in the lexical decision process are of a discrete or a dynamic nature. They recorded participants' hand's movements during a lexical decision task in which participants had to click with the mouse on one of two options presented on the screen: one lexical and one nonlexical. The task was run in Italian and the hand movements were recorded using MouseTracker software (Freeman & Ambady, 2010). The authors observed that the "lexicality effect" (better performances for lexical vs. nonlexical stimuli) did not show up at the beginning of the motor response, but during the trajectory towards the response: after the initiation of the movement, the hand was more attracted towards the competitor when the stimulus was a pseudoword than when it was a real word. The fact that this competition process occurred after the initiation of the motor response demonstrated that, first, lexical decision had not been completed before the start of the movement and, second, the lexicality effect reflected the dynamic integration of different sources of information during the decision process (for consistent results in the temporal domain, see also Barca and Pezzulo, 2015). The curved mouse trajectories observed during the execution of the motor responses could be interpreted within the *Interactive account of ventral occipitotemporal cortex (vOT) model* (Price & Devlin, 2011), if we consider that (especially) in the case of pseudowords, there is significant conflict between top-down lexical predictions and bottom-up sensory information.

Barca and Pezzulo (2012) also found that the lexicality effect was observable between low-frequency words and pseudowords, but not between high-frequency words and strings of consonants. Again, the *Interactive account of vOT model* (Price & Devlin, 2011) can explain such results. Price and Devlin (2011) explained how, when a word is read, the information from the visual areas activates the vOT which then, in turn, partially activates higher-order areas corresponding to the semantics and the phonology of the written characters. This forward or bottom-up activation is fast and followed by a process in which the activated higher-order areas send information back to the vOT. Next, recurrent interactions between the information from the visual areas and information from higher-order areas take place in the vOT until they become maximally consistent. The connections of the higher-order areas are strengthened with practice, resulting in predictive feedbacks that become more efficient with experience. Thus, extremely frequent words and strings of consonants are responded to quickly and accurately in a lexical decision task because, in the former case, a lot of the feedback information sent to the vOT matches

the information from the visual areas and, in the latter case, because no top-down predictions are involved. In both cases, no or only a few prediction errors conflict with the information from the visual areas. Although low-frequency words generate less feedback information than high-frequency words, this information still matches the information from the visual areas. By contrast, pseudowords engage prediction errors from the phonological areas, resulting in mismatching information between top-down and bottom-up processes that makes it difficult to decide that pseudowords are not words.

Similar predictions are also made by the Leaky Competing Accumulator Model (LCA) of lexical decisions (Dufau, Grainger & Ziegler, 2012). This is a computational model of word recognition in the context of a lexical decision task. In this framework, participants need to make a choice between two response nodes: *word* or *nonword*. This model is based on a previous LCA model (Usher & McClelland, 2001) for binary-decision making, but has been adapted to the particular case of a lexical decision task, in which the choice of the *nonword* response node simply depends on the lack of evidence for the *word* response node. According to this model, response towards the *word* node are the result of a noisy accumulation of evidence in favor of this node and depend on the lexical activity generated by the stimulus. This lexical activity is the reflection of the evidence for a word corresponding to this stimulus. Regarding the activation of the *nonword* response node, it is computed by deducting the lexical activity generated by the stimulus from a constant value. The activation of one node inhibits the activation of the other node and this process is thought to end when the response threshold has been reached for one of the two nodes. Importantly, the response threshold should change from trial to trial and from one experimental condition to another, depending on the response criterion value adopted by the participant.

In sum, despite their differences, both the interactive model (Price & Devlin, 2011) and the LCA model (Dufau, Grainger & Ziegler, 2012) suggest that making a lexical decision is a dynamic and competitive process, in which top-down and bottom-up information is compared and evidence is accumulated to bring about a positive or a negative response. Recording hand movements when participants move the computer's mouse towards the selected response (Barca & Pezzulo, 2012) has proved to be a useful way of studying the dynamic aspects of lexical decisions. However, it is still unclear what kind of evidence in favor of the lexical or nonlexical alternatives is accumulated over time, and the only studies addressing this question with the aid of the Mouse Tracker paradigm are in Italian (see Barca & Pezzulo, 2015, for a review).

The first aim of the present study was to replicate Barca and Pezzulo's (2012) results in a new language, French, in order to shed new light on the type of information that is accumulated over time. Specifically, the use of the French language will make it possible to gain insights into potential differences in the way evidence is accumulated between languages of different

orthographic consistencies. Barca and Pezzulo (2012) ran their experiment in Italian, a language with a highly consistent orthography (i.e., a script with a very regular correspondence between the orthographic visual form of a word and its spoken representation). The French language has a more complex spelling-sound relation than Italian, especially for vowels (with several silent letters and a large number of homophones; e.g. “saint”/“sein”/“sain”/“ceins”/“ceint” for the words pronounced [sɛ̃]). Ziegler, Jacobs and Stone (1996) performed a statistical analysis of orthographic consistency in French and found that, even though the French language is relatively consistent for the purposes of retrieving phonology from spelling (87.6% consistency for monosyllabic words), it is highly inconsistent when it comes to retrieving spelling from phonology (79.1% inconsistency for monosyllabic words). This inconsistency, namely feedback inconsistency as opposed to feedforward inconsistency, is not problematic when accessing the phonological lexical representation in a reading-aloud task (e.g., Ziegler, Perry, & Coltheart, 2003), but can be problematic when trying to activate the right orthographic lexical representation in a written lexical decision task (Ziegler, Jacobs, & Klüppel, 2001; but see Ferrand & Grainger, 2003, discussed below).

Studies investigating the orthographic consistency effects with a visual lexical decision task have typically shown that participants using a language with a highly inconsistent orthography rely more on semantic than on higher-order phonological information than those who use a language with a highly consistent orthography (see Lima & Castro, 2010, for a review). The language with an inconsistent orthography that has received by far the most attention in this context is English. The English and French languages differ in their orthographic inconsistency properties in that English is feedforward and feedback-inconsistent (Ziegler et al., 1997) while French is feedforward-consistent and feedback-inconsistent (Ziegler et al., 1996). If French orthography produces effects similar to those observed with other feedforward and feedback-inconsistent orthographies, this could mean that the effects of orthographic consistency do not originate when phonology is retrieved from spelling, but when participants attempt to retrieve spelling from phonology, in a kind of verification process, as proposed by Ziegler et al. (2001). This would lead us to expect a larger frequency effect (i.e., a greater difference between high and low-frequency words) in French than that reported by Barca and Pezzulo (2012) in Italian, a larger lexicality effect (i.e., a greater difference between pseudowords and low-frequency words), and also a significant difference between high-frequency words and strings of consonants - because we expect participants to favor the use of semantic at the expense of phonological information, high-frequency words should be judged faster than consonant strings. Alternatively, if the conflict originates when retrieving phonology from spelling, the observed results should be similar to those of Barca and Pezzulo (2012) because this process is relatively consistent in French.

Another goal of the present study was to test the variability of the responses. To this end, the experiment was repeated twice with a 5-minute interval. Diependale, Brysbaert and Neri (2012) showed that, contrary to what had previously been assumed, the responses of the participants to specific items in a lexical decision task were quite inconsistent from one presentation to the next (some items that were classed as “words” in one presentation were classed as “nonwords” in the other). Here, our main goal was not to analyze the repetition effects on specific items, but to look for a repetition effect in the pattern of the responses in the second presentation compared to the first. More precisely, we wanted to investigate whether the repetition of the same items would modulate the frequency effect or the strategy adopted by the participants to respond to the items. If repetitions affect the frequency effect, the participants would learn the items they have previously been exposed to. Because lexical access to the most frequent items is already fast and does not generate many prediction errors, these items would be unaffected. By contrast, low-frequency words, pseudowords and strings of consonants should be processed faster and more accurately than in the first presentation. Alternatively, if repetitions affect response strategies, we should observe different response criteria from one presentation to the other (Dufau, et al., 2012). Presenting the same items could induce the participants to lower their response threshold because they feel more confident with the items. As a consequence, participants would be expected to respond faster but less accurately (more errors and greater deviations of the hand movements) on all the items, an effect known as the speed–accuracy trade-off (see Pleskac & Busemeyer, 2010, for a review).

The second experiment explored the pseudohomophone effect. A pseudohomophone is a string of letters that is phonologically like but orthographically unlike a real word (e.g., “brane” is the pseudohomophone of “brain”). Typically, pseudohomophones are named faster than nonhomophonic nonwords but their lexical judgment is slower (e.g., Seidenberg, Petersen, MacDonald, & Plaut, 1996) and opposite baseword effect frequencies are observed in the two tasks: while pseudohomophones with a high-frequency baseword are named more slowly than those with a low-frequency baseword in a naming task (Herdman, Lefevre, & Greeham, 1996), they are judged faster in a lexical decision task (Ziegler et al., 2001). Ziegler et al. (2001) investigated the reasons for such differences between naming and lexical decision tasks with pseudohomophones and demonstrated that the pseudohomophone effect in each task depends on two distinct but related mechanisms. In a naming task, the pseudohomophone will activate an already existing phonological representation, resulting in faster naming latencies than in the case of nonhomophonic nonwords. However, the activation of the phonological lexical representations of high-frequency basewords will be strong enough to activate the corresponding orthographic lexical representations, which will need to be inhibited before naming. In a lexical decision task, the activation of an already existing phonological representation will lead to prediction errors and allow the orthographic lexical

representation of the baseword to be activated. A spelling verification process will then start. This consists in comparing the orthographic information from top-down predictions with the information from the visual input. As a consequence, the participants' responses will slow down. Because the activation of the orthographic information from higher-order areas will be earlier and more complete for high compared to low-frequency basewords, the verification process will be faster and easier for high-frequency basewords.

This verification mechanism proposed by Ziegler et al. (2001) for lexical decision tasks is based on data obtained from English participants and matches the models reviewed earlier (Dufau, et al., 2012; Price & Devlin, 2011). However, data obtained from French participants with heterographic homophones (Ferrand & Grainger, 2003) suggest that the conflict between the orthographic representations (one matching the visual input and the other corresponding to its heterographic homophone) is resolved entirely at the lexical level. As discussed earlier, feedforward and feedback orthographic consistencies are unbalanced in French: while the spelling-to-phonology correspondence is highly consistent, the phonology-to-spelling correspondence is highly inconsistent (see Ziegler et al., 1996). In English, both feedforward and feedback correspondences are highly inconsistent (Ziegler, Stone & Jacobs, 1997), making it a more balanced language. When processing homophones in a lexical decision task, French participants might refrain from the normally observed verification process because it is highly unreliable in French and encourages a strategy that would consist in inhibiting the competitor at the lexical level.

Nevertheless, Ferrand and Grainger (2003) were not able to completely discard the verification hypothesis, because their homophone stimuli were, logically, the most feedback-inconsistent stimuli compared to the stimuli of the other categories - a feature that might have biased their results. To solve this problem, we used high and low-frequency baseword pseudohomophones as in Ziegler et al. (2001). This methodology, combined with the use of online recording with the Mouse Tracker software, should make it possible to choose between these two possibilities: a conflict solved at the lexical level or via a verification between top-down and bottom-up information. If the conflict is solved at the lexical level, we should observe a slower response for high- compared to low-frequency baseword pseudohomophones, because the degree of activation of the real baseword would be stronger and more difficult to inhibit. If the conflict is solved via a verification process between the top-down and bottom-up visual information, then we should observe a faster response for high compared to low-frequency baseword pseudohomophones, because the more complete and faster activation of the real baseword would permit a faster verification. In both cases, responses should be slower for pseudohomophones than for nonhomophonic nonwords, due to the inhibition of the activated competitor at the lexical level (Ferrand & Grainger, 2003) or the slow verification process (Ziegler et al., 2001).

2. Experiment 1. Tracking visual lexical decision in French

The first aim of Experiment 1 was to replicate as closely as possible Barca and Pezzulo's (2012) experiment in French. In this experiment, the participants performed a visual lexical decision task and gave their response by moving the computer mouse toward the selected category (lexical or nonlexical, corresponding to two buttons at the top-left and top-right of a computer screen). The stimuli presented for lexical categorization were words that varied in written frequency, strings of letters and pseudowords. In Barca and Pezzulo's (2012) study, attraction toward competing alternative responses was observed only for low-frequency words and pseudowords, with the latter receiving a higher level of attraction than the former.

If French words are processed as in languages with an inconsistent orthography, then longer response latencies and trajectories should emerge for low-frequency words. Furthermore, the difference between low-frequency words and pseudowords should be greater. In addition, strings of letters should be more attracted toward the competing alternative response than high-frequency words. Alternatively, if French words are processed as in languages with a consistent orthography, the results should be very similar to those of Barca and Pezzulo (2012).

The last aim of Experiment 1 was to investigate whether or not the repetition of the same items would modulate the frequency effect or the strategies adopted by the participants to make their lexical judgments.

2.1 Method

2.1.1 Participants

The study included 30 French native speakers from Université Blaise Pascal in Clermont-Ferrand, France (mean age = 20.7, SD = 2.03). They were all right-handed with normal or corrected to normal vision.

2.1.2 Materials and stimuli

The experimental stimuli consisted of a list of 96 lexical and nonlexical items (see Appendix 1). The lexical items were 24 high- and 24 low-frequency French nouns: high-frequency words had a mean value of 179.74 occurrences per million and low-frequency words had a mean value of 3.75 occurrences per million according to the *French Lexicon Project* (Ferrand, New, Brysbaert, Keuleers, Bonin, Méot, Augustinova, & Pallier, 2010). They were four to five letters long, morphologically simple (i.e. neither derived nor compounds) and unambiguous in terms of grammatical category and meaning.

The psycholinguistic characteristics of the stimuli are presented in Table 1. The high-frequency words tended to be acquired earlier ($p < .005$) and were more familiar ($p < .005$) than the low-frequency words, but were similar in imageability, with few orthographic and phonological neighbors, and had the same length in letters. Age of acquisition and subjective frequency were taken from Ferrand, Bonin, Méot, Augustinova, New, Pallier, and Brysbaert (2008), and imageability from Ferrand, Brysbaert, Keuleers, New, Bonin, Méot, Augustinova, and Pallier (2011). The nonlexical items included 24 legal pseudowords and 24 illegal pseudowords. Pseudowords were created by changing two or more letters of real low-frequency words (not included in the list), so that they were pronounceable and orthographically similar to lexical stimuli. Pseudowords and letter strings were generated using the WordGen software (Duyck, Desmet, Verbeke & Brysbaert, 2004). To improve stimulus variation in the experimental list, half of the stimuli were strings of consonants and half were strings of vowels.

The experimental stimuli could be arranged along a “lexicality dimension line” (see also Barca & Pezzulo, 2012), ranging from highly lexical items with rich lexical representations to strings of letters that had no representations in the lexicon, could not be assembled into orthographic/phonological sequences, and did not resemble any lexical items. Between these, there were low-frequency words and pseudowords.

Table 1. Psycholinguistic characteristics of the words used in Experiment 1

	AoA	Subj. Freq.	Ima.	OLD20	PLD20	Number of letters	Subt. Freq.	RTs from Chronolex	RTs from FLP
HF	5.17	5.08	5.9	1.49	1.17	4.54	119.74	589.92	609.66
LF	7.26	3.61	5.84	1.54	1.28	4.62	3.75	661.8	655.84
	$p < .005$	$p < .005$	ns $p = .86$	ns $p = .5$	ns $p = .15$	ns $p = .56$	$p < .005$	$p < .005$	$p < .005$

Note. HF: high-frequency words; LF: low-frequency words; AoA: age of acquisition (taken from Ferrand et al., 2008); Subj. Freq.: subjective frequency (taken from Ferrand et al., 2008); Ima.: imageability (taken from Ferrand et al., 2011); OLD20: orthographic Levenshtein distance 20; PLD20: phonological Levenshtein distance 20; Subt. Freq.: subtitle frequency (taken from Ferrand et al., 2010); RTs from Chronolex: reaction times (in ms) taken from Chronolex (Ferrand et al., 2011); RTs from FLP: reaction times (in ms) taken from the French Lexicon Project (Ferrand et al., 2010).

A /START/ button was located at the bottom-center of the PC screen (DELL Precision) and a centrally presented cross served as a fixation point. Two buttons, one located at the top-left and the other at the top-right of the screen, allowed the participants to indicate their decision: lexical or nonlexical. Stimuli were presented in black, lowercase ARIAL font on a white background. The experiment was run and the data recorded using MouseTracker software (Freeman & Ambady, 2010).

2.1.3 Experimental procedure

To begin each trial, the participants clicked on the /START/ button. The fixation cross then appeared in the center of the screen and was replaced by an experimental stimulus in the same position after 300 ms. The stimulus remained on the screen for 500 ms. The participants had to respond within 2000 ms, otherwise a /TIME OUT/ message appeared. The participants were instructed to use the mouse to move the cursor to the appropriate response location (i.e., top-left button for lexical stimuli, top-right button for nonlexical stimuli) and to click it to indicate their response. The correspondence between stimulus type and button was varied across participants.

Categorization errors and movement response times (i.e. time from when the participants pressed /START/ until they reached and clicked on the selected choice) were recorded automatically. A feedback message (red cross) was displayed in the centre of the screen to indicate an incorrect response. While the participants were making their response, the x and y coordinates of the mouse trajectories were recorded (sampling rate of approximately 70 Hz). Before the experimental data were acquired, the participants performed a practice session of 12 items (6 lexical stimuli and 6 nonlexical stimuli). The experimental stimuli were presented in two blocks of 48 items each. The order of stimuli within blocks and the order of block presentation were randomized. Half of the participants categorized stimuli as lexical by clicking the top-left button and as nonlexical by clicking on the top-right button, while the other half did the opposite.

The full list of stimuli was shown a first time to permit a comparison with Barca and Pezzulo (2012). The full list was then presented to the participants a second time to try to identify a repetition effect and response variability across multiple presentations of the same items.

2.2 Results of the first presentation

2.2.1 Data processing

Mouse trajectories were rescaled into a standard coordinate space. The durations of the trajectory movements were then normalized by re-sampling the time vector into 101 time-steps using linear interpolation to permit averaging across multiple trials (for a similar procedure, see Barca, Benedetti, Pezzulo, 2015; Barca & Pezzulo, 2012; Flumini, Barca, Pezzulo, Borghi, 2014; Freeman & Ambady, 2010).

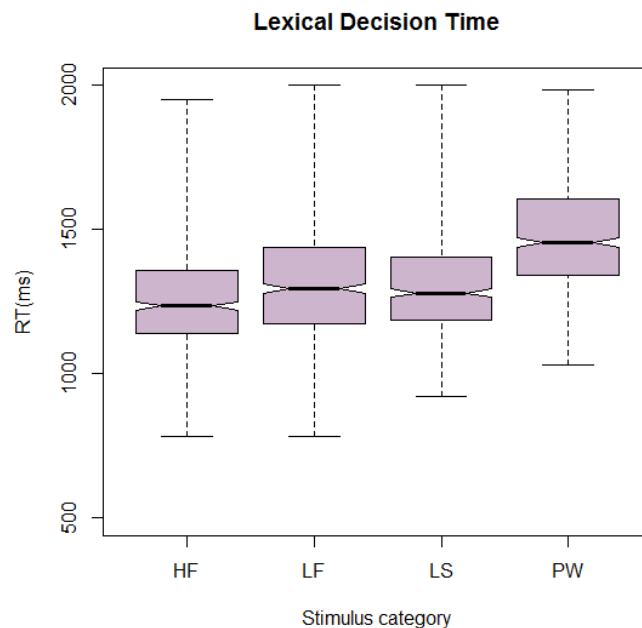
Responses exceeding the 2000 ms timeout, which accounted for .02% of the total data, were discarded from the analysis, as were incorrect responses (i.e. when the subject selected the inappropriate stimulus category), which accounted for .02% of the data. Thus, a total of .04% of the responses were discarded from subsequent analysis.

Linear Mixed-Effects Modelling (LMMs) was used to assess the impact of Lexicality on the response variables (Baayen, Davidson & Bates, 2008; Brysbaert, 2007; Bates & Maechler, 2009). Analyses were run with the lmerTest package for R (Bates & Maechler, 2009; Kuznetsova, Brockhoff & Christensen, 2016). Backward elimination of non-significant effects of the linear mixed effect model was performed with the *step* function (backward elimination of the random part is performed first, following by backward elimination of the fixed part. Finally, least squares means and their differences for the fixed part of the model are calculated. The p-values for the fixed effects are calculated from an F test and t test based on Sattethwaite's approximation). The model included Subjects and Items as Random effects factors and Stimulus category as Fixed effects factor (using R syntax, the model tested was $RT \sim \text{Lexicality} + (1|\text{subject}) + (1|\text{items})$). Separate models were run for the different dependent variables (i.e., overall response time, initiation time, etc.).

2.2.2 Accuracy and Lexical Decision time

The items were categorized correctly in 96% of cases. Participants' errors across stimulus categories were distributed as follows: .01% for High-Frequency words, .02% for Low-Frequency words .04% for Pseudowords and .01% for Letter Strings. A box and whisker diagram illustrates the spread of response times for correct categorization (Figure 1). There was no response exceeding the time deadline for High-Frequency words and Pseudowords, but 2% of Low-Frequency responses and 6% of Letter Strings exceeded the time limit.

Figure 1. Overall response time for correct lexical categorization in Experiment 1



Note. HF: High-frequency, LF: Low-frequency, LS: Letter Strings, PW: Pseudowords

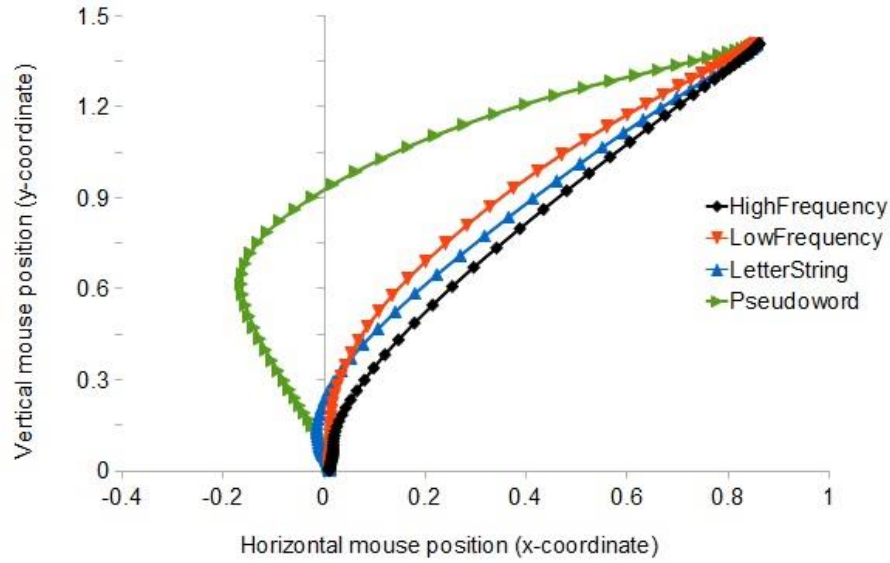
The mixed-effects models of the *Initiation Time* value, with High-Frequency words as referent category, confirmed that the participants initiated their movement responses faster for High-Frequency words than for Pseudowords (β Pseudowords = 29.03, t value = 2.33, $p < .05$), with there being no difference between High-Frequency words and the other categories (β Low-Frequency = 2.13, t value < 1 ; β Letter Strings = 9.98, t value < 1).

Overall Lexical Decision Time was modulated by stimulus lexicality, with High-Frequency words eliciting faster responses than the other categories (β Low-Frequency = 45.59, t value = 5.48, $p < .001$; β Pseudowords = 205.84, t value = 24.18, $p < .001$; β Letter Strings = 32, t value = 3.85, $p < .005$). The mean response time for Low-Frequency words was 160 milliseconds faster than for Pseudowords (computations following the application of the step function: β Low Frequency - Pseudowords = -160.31, t value = -18.75, $p < .001$), whereas it was only 13 milliseconds slower than for Letter Strings, and the difference did not reach significance (β Low Frequency - Letter Strings = 13.6, t value = 1.63, *n.s.*). Finally, Letter Strings were responded to faster than Pseudowords (β Letter Strings - Pseudowords = -174.8, t value = -20.4, $p < .001$).

2.2.3 Analyses of trajectories

Mean trajectories for the different conditions are presented in Figure 2 (remapped to the right, with the unselected response at the left corner). The mean trajectory for Low-Frequency words showed an attraction toward the Nonlexical response alternative. Similarly, the trajectories towards Pseudowords were strongly attracted to the Lexical alternative response.

Figure 2. Mouse trajectories for correct categorization in Experiment 1. Note that to facilitate visual comparison of the trajectories in the four experimental conditions, all responses (to lexical and nonlexical stimuli) were remapped to the right.



To index *the complexity of the trajectories*, the Area Under the Curve values (AUC) were analyzed using LMMs modelling with two Random-effects factors (i.e., Subjects and Trials) and one Fixed-effects factor (Stimulus category). When High-Frequency words were coded as the reference category, positive contrast coefficients for Low-Frequency words (β Low-Frequency = .22, t value = 4.48, $p < .001$) and Pseudowords (β Pseudowords = .96, t value = 19.18, $p < .001$) indicated that their mean AUC values were higher than the values for High-Frequency words, with this difference being greater for Pseudowords. No significant difference emerged between High-Frequency words and Letter Strings (β Letter Strings = .08, t value = 1.64, ns). Thus, the Low-Frequency words and Pseudowords stimuli were significantly more attracted to their own competing target category (i.e., Low-Frequency words to the nonlexical category, and Pseudowords to the lexical category) than were the High-Frequency words. This provides evidence suggesting an interference effect due to the activation of competing categories. Low-Frequency words were less attracted to the competing category than Pseudowords (β Low-Frequency - Pseudowords = -.74, t value = -14.72, $p < .001$) and more than Letter Strings (β Low-Frequency - LetterStrings = .1, t value = 2.8, $p < .005$); and Pseudowords were more attracted to the lexical category than Letter Strings (β LetterStrings - Pseudowords = -.9, t value = -17.57, $p < .001$).

To analyze *motor complexity*, we considered the x-flip values, which reflect fluctuations in trajectory between alternative responses along the x-axis. Significant parameters emerged for Low-Frequency words (β Low-Frequency = .36, t value = 2.38, $p < .05$), Pseudowords (β Pseudowords = .86, t value = 5.58, $p < .001$) and also for Letter Strings (β Letter Strings = .32, t value = 2.1, $p < .05$), indicating that these trajectories fluctuated more than that of High-Frequency words (High-

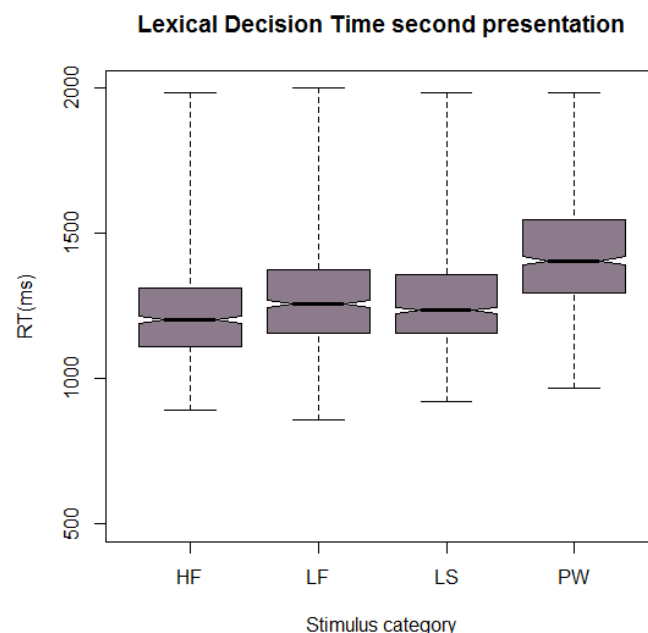
Frequency: 5.89; Low-Frequency: 6.23; Pseudowords: 6.73; Letter Strings: 6.2). The step function suggested that there was no difference between Low-Frequency words and Letter Strings (β Low Frequency - Letter Strings = -.04, t value < 1), whereas the Pseudoword trajectory fluctuated more than that of Low-Frequency words (β Low Frequency - Pseudowords = -.51, t value = -3.23, p < .005) and Letter Strings (β Letter Strings - Pseudowords = -.55, t value = -3.53, p < .005).

2.3. Results of the second presentation

2.3.1 Accuracy and Lexical Decision Time

As for the first presentation of the stimuli, accuracy was very high (96% of cases) and the pattern of the results did not change. Pseudowords elicited more errors (6.4%) and responses exceeding the time deadline (3%) than the other stimuli. Lexical Decision Times for the second presentation of the stimuli are reported in Figure 3. Overall, an increase in the speed of response of second presentation was observed (high-frequency 36 ms faster, low-frequency 35 ms, pseudowords 48 ms, and letter strings 37 ms, respectively).

Figure 3. Overall response time for correct lexical categorization – second presentation of the stimuli in Experiment 1



Note. HF: High-frequency, LF: Low-frequency, LS: Letter Strings, PW: Pseudowords

In order to test for an effect of list repetition, different LMMs were conducted on the movement parameters, with Stimulus Category and Presentation (First vs. Second) as factors. Backward

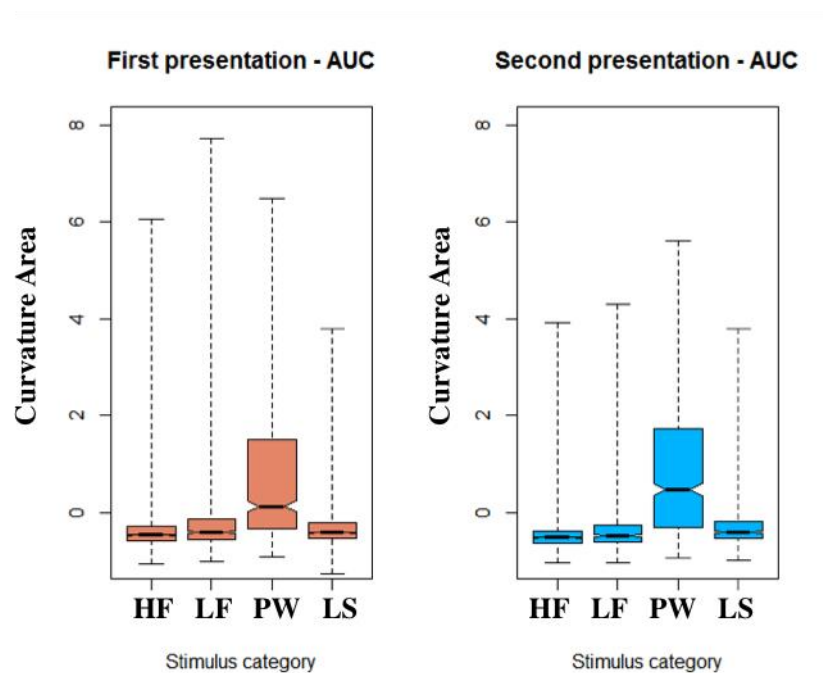
elimination of non-significant effects of the linear mixed effect model was performed with the *step* function. The p-values for the fixed effects were calculated using an F test based on Sattethwaite's approximation. Results are summarized in Table 2.

Table 2. Results of the LMMs on the movement parameters with Stimulus Category and Presentation (First vs. Second) as factors.

Variables	Stimulus Category		Presentation		Interaction	
	F value	P value	F value	P value	F value	P value
RT	444.3	p<.0001	F<1	ns	F<1	ns
init_time	2.03	ns	F<1	ns	1.05	ns
MD_time	289.8	p<.0001	F<1	ns	F<1	ns
x-flip	20.6	p<.0001	F<1	ns	1.65	ns
zAUCtogether	388.7	p<.0001	F<1	ns	4.86	p<.005

The interaction between Stimulus Category and Presentation was significant for trajectory measures of AUC, with a significant difference being observed for the Pseudoword values as a function of Presentation (β first– second presentation = -.2, t value = -3.13, $p < .005$) (see Figure 4).

Figure 4. Area Under the Curve for correct categorization in Experiment 1



The direction of the difference indicates higher AUC values for the second presentation of the Pseudowords. This might be linked to the response time difference between stimulus list presentation, with responses being faster for the second presentation.

2.4 Discussion of Experiment 1

The present results confirm the dynamic nature of the lexical decision process (Price & Devlin, 2011; Dufau, et al., 2012). Indeed, as in Barca and Pezzulo's (2012) study, most of the effects were observed after the beginning of the hand movement, with attractions towards the competitor being observed during the trajectory towards the response. They also confirm that Mouse Tracker is able to capture this process.

As in Italian, better performances were observed for high-frequency words and letter strings than for low-frequency words and pseudowords, confirming that lexicality is arranged along a "lexicality dimension line" (see Barca & Pezzulo, 2012) on which are represented, from the most to the least lexical dimensions: high-frequency words, low-frequency words, pseudowords and letter strings.

A lexicality effect was also observed between low-frequency words and pseudowords. Although a direct comparison between the datasets of Barca and Pezzulo (2012) and this study is not possible, the difference between (average) AUCs for LF vs. PW words is greater in French (.99) than in Italian (.54), suggesting that the lexicality effect is greater in French. Another interesting difference was that high-frequency words were categorized 32 ms faster than letter strings in our experiment, whereas Barca and Pezzulo (2012) did not find any difference between highly lexical (high-frequency words) and highly nonlexical (letter strings) stimuli. Finally, the frequency effect was also larger (difference of 31 ms in Italian, and 45 ms in French). These differences indicate that French, compared to Italian, participants rely more on semantic than higher-order phonological information (see Lima & Castro, 2010). Because the French language is relatively feedforward-consistent but feedback-inconsistent (Ziegler, et al., 1996), these differences suggest that the conflict between spelling and phonology in languages with feedforward and feedback-inconsistent orthography occurs when participants try to retrieve the phonology from spelling (Ziegler et al., 2001).

The results of the second presentation are extremely consistent with those of the first presentation. The participants responded faster in every condition and some of the results showed that they were less accurate (i.e., greater curvatures for pseudowords). If one assumes a dynamic model of decision-making such as the LCA model (Dufau et al., 2012) or the widely used drift-diffusion model (Ratcliff & McKoon, 2008, see also its extension to mouse tracking set-ups, Lepora

& Pezzulo, 2015), these results are consistent with the hypothesis that participants adopted a lower response threshold in the second than in the first presentation. The participants might have felt more confident when the same stimuli were presented a second time and may have traded accuracy for speed. Learning of the item orthographies would have led to faster and more accurate responses for low-frequency words, pseudowords and letter strings, but not for high-frequency words because these last items already generate fast and accurate predictions.

3 Experiment 2. Tracking visual lexical decision for pseudohomophones

In line with the predictions of dynamic models (Price & Devlin, 2011; Dufau, et al. 2012) and the results of a previous study in Italian (Barca & Pezzulo, 2012), the results of Experiment 1 indicate that French participants engage in a process that compares top-down and bottom-up information when making a lexical decision.

However, things may be different when processing heterographic homophones or pseudohomophones. While Ziegler et al. (2001) demonstrated that the conflict generated by pseudohomophones in English is solved thanks to a verification process between top-down and bottom-up information, the results reported by Ferrand and Grainger (2003) seem to indicate that the conflict between French heterographic homophones is solved entirely at the lexical level. Nevertheless, the characteristics of the French homophones (which were feedback-inconsistent) might have biased Ferrand and Grainger's (2003) results.

The use of high and low-frequency baseword pseudohomophones as in Ziegler et al. (2001) should make it possible to test whether the orthographic-spelling conflict with pseudohomophones is solved at the lexical level or thanks to a verification process. In the former case, high-frequency baseword pseudohomophones should elicit slower responses than low-frequency baseword items because the real word (the baseword) homophone would be more active and, as a consequence, more difficult to inhibit. The opposite should be true in the latter case because the greater activation of the real baseword would allow a faster verification process. Because both inhibition and verification processes need time to be achieved, it was expected that the pseudohomophones would be judged more slowly than nonpseudohomophone controls, no matter which of the process was involved.

3.1 Method

3.1.1 Participants

Experiment 2 included 42 participants (mean age = 20.4 years, SD = 1.93) with the same characteristics as in Experiment 1.

3.1.2 Stimuli

A list of 24 high-frequency words, 24 low-frequency words and 48 pseudowords was used (see Appendix 2). The 24 high-frequency words and the 24 low-frequency words were those used in Experiment 1. Among the pseudowords, 24 were pseudohomophones taken from the PHOM database (Farioli, Grainger, & Ferrand, 2011) and the other 24 pseudowords were orthographic controls. The orthographic controls were based on the same words as the pseudohomophones and differed from their baseword by one letter, but did not sound like a real word. The pseudohomophone status was evaluated using the procedure of Stone and Van Orden (1993). Ten “judges” quickly read aloud a list of pronounceable nonwords that included the candidate pseudohomophones. For a pseudohomophone to be accepted as homophonic to a word, 9 out of the 10 judges had to have pronounced it in a way that matched its baseword. Half of the pseudowords were based on high-frequency words and the other half on low-frequency words.

3.1.3 Procedure

The procedure was exactly the same as in Experiment 1, except that in Experiment 2, there was only one presentation of the experimental list.

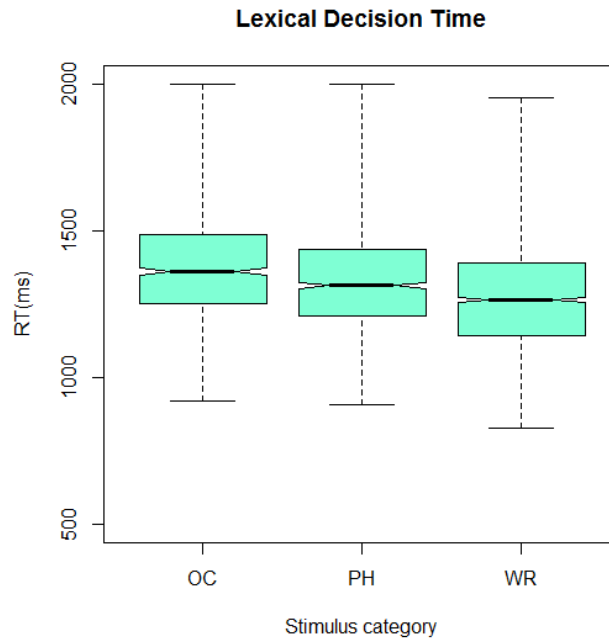
3.2 Results

3.2.1 Accuracy and Lexical Decision Time

The participants were highly accurate with 96.3% correct responses, similarly distributed across the different stimulus categories (96.5% for words, 96.7% for pseudohomophones, 95.6% for orthographic controls).

Linear Mixed-Effects Modelling (LMMs) was used to assess the impact of Lexicality on the response variables (Subjects and Items were Random-effects factors, Stimulus category was the Fixed-effects factor). Backward elimination of non-significant effects was performed with the *step* function (p-values for the fixed effects were calculated using a t test based on Sattethwaite’s approximation). In this analysis, Orthographic controls were considered as the “default” level for comparison. LMMs on *Initiation Time* did not yield significant results. *Overall Lexical Decision Times* (see Figure 5) were modulated by stimulus lexicality: words were correctly processed 48 ms faster than pseudohomophones and 87 ms faster than orthographic controls (β Pseudohomophones - Words = 47.6, t value = 7.69, $p < .001$; β OrthographicControls - Words = 93.7, t value = 15.03, $p < .001$). Moreover, pseudohomophones were processed significantly faster than orthographic controls (β OrthographicControls - Pseudohomophones = 46, t value = 6.42, $p < .001$). Similar results were observed for *Maximum Deviation Time* (β Orthographic controls - Pseudohomophones = 37.5, t value = 5.8, $p < .001$; β Orthographic controls - Words = 69.7, t value = 12.3, $p < .001$; β Pseudohomophones - Words = 32, t value = 5.73, $p < .001$).

Figure 5. Overall Lexical Decision Time in Experiment 2



Note. OC: orthographic controls, PH: pseudohomophones, WR: words

LMMs was used to test for an effect of baseword frequency, with Frequency and Type of stimulus as fixed factors. Linear Mixed Modelling of the *Initiation Time* did not yield significant results. By contrast, *Overall Lexical Decision Times* showed effects of stimulus type ($F(2, 3760) = 117.4, p < .001$), frequency ($F(1, 3818) = 29.57, p < .001$) and an interaction between the two factors ($F(2, 3817) = 19.78, p < .001$). High-frequency words were processed faster than all the other categories (p values $< .001$), and low-frequency words were processed earlier than orthographic controls irrespective of baseword frequency (p values $< .001$), but did not differ from pseudohomophones (irrespective of baseword frequency, t values < 1). Finally, pseudohomophones gave rise to shorter response times than orthographic controls, irrespective of baseword frequency. No difference was observed between orthographic controls derived from high and low-frequency words (t value < 1), or between pseudohomophones derived from high and low-frequency words (t value = -1.16 , ns). Main effects and the interaction were significant for *Maximum Deviation Time* (F Stimulus type (2, 3732) = 77.4, $p < .001$; F Frequency (1, 3805) = 19.7, $p < .001$; F Interaction (2, 3803) = 15, $p < .001$). The maximum deviation in the trajectories of the high-frequency words reached its maximum value earlier than for the other categories (p values $< .001$), and the maximum deviation for low-frequency words was earlier than for orthographic controls irrespective of baseword frequency (p

values < .001) but did not differ from pseudohomophones (irrespective of baseword frequency, (t value <1.1). Finally, Pseudohomophones reached maximum deviation earlier than orthographic controls, irrespective of baseword frequency (p values < .001).

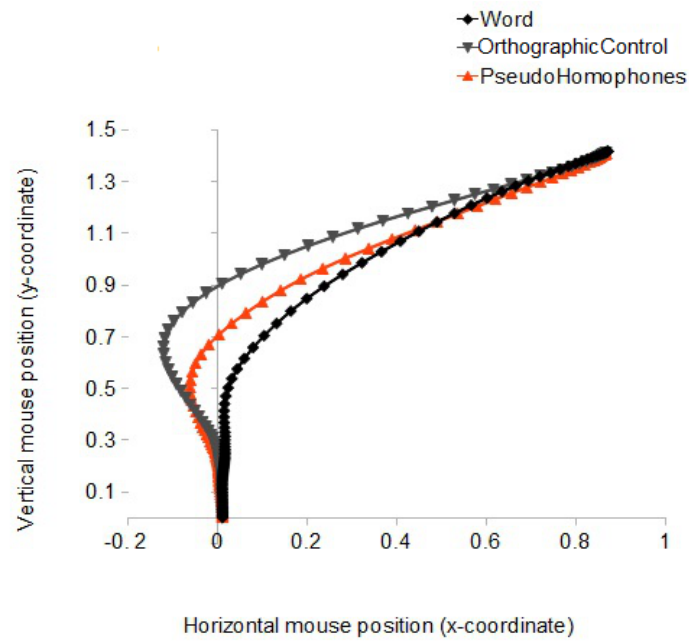
No difference was observed between orthographic controls derived from high and low-frequency words (t value <1), or between pseudohomophones derived from high and low-frequency words (t value <1).

3.2.1 Analysis of trajectories

Spatial attraction

Mean trajectories for the different conditions are presented in Figure 6. Using LMMs, differences emerged on the measure of *Area Under the Curve* between words and pseudohomophones (β Pseudohomophones - Words = .1, t value = 3.46, $p < .001$) and words and orthographic controls (β OrtographicControls - Words = .2, t value = 4.28, $p < .001$). No difference emerged between pseudohomophones and orthographic controls (t value < 1). Similar results were observed for *Curve Maximum Deviation* between words and pseudohomophones (β Pseudohomophones - Words = .2, t value = 4.20, $p < .001$) and words and orthographic controls (β OrtographicControls - Words = .2, t value = 4.58, $p < .001$). No difference emerged between pseudohomophones and orthographic controls (t value < 1). Thus, correct trajectories of both pseudohomophones and orthographic controls showed an attraction toward competing alternatives. This was not the case for lexical stimuli.

Figure 6. Mouse trajectories for correct categorization in Experiment 2



If we take account of the frequency of the word stimuli and the frequency of the basewords for the pseudohomophones and orthographic controls, then stimulus type and the interaction were significant on both measures of *Area Under the Curve* (F Stimulus type (2, 3799) = 11, $p < .001$; F Frequency (1, 3858) = 2.27, ns; F Interaction (2, 3857) = 18.49, $p < .001$) and *Curve Maximum Deviation* (F Stimulus type (2, 3828) = 13.9, $p < .001$; F Frequency (1, 3868) = 1.71, ns; F Interaction (2, 3868) = 16.94, $p < .001$), with significant difference between the MD values for high-frequency words and low-frequency words, pseudohomophones (irrespective of baseword frequency, p values $< .001$) and orthographic controls (irrespective of baseword frequency, p values $< .001$). The same pattern emerged for the AUC values.

The analysis of x-flip values showed a main effect of Frequency (F (1, 3839) = 12.82, $p < .001$), with neither the effect of Stimulus type (F (2, 3839) = 3.89, ns) nor the interaction (F (2, 3837) = 1.89, ns) reaching the threshold of significance.

3.3 Discussion of Experiment 2

The results obtained in Experiment 2 were somewhat puzzling and could not be accounted for by any of the a priori hypotheses. Performance was better for pseudohomophones than for nonhomophonic pseudoword controls and there was no difference between high and low-frequency baseword pseudohomophones.

Our guess is that the composition of the list caused the participants to adopt a different reading strategy. Many studies have demonstrated that list composition could modulate participants'

strategy when recognizing written words (among others Grainger, Spinelli, & Ferrand, 2000; Kinoshita, Lupker, & Rastle, 2004; Lupker, Brown, & Colombo, 1997; Pagliuca, Arduino, Barca, & Burani, 2008). However, the most relevant work for this study is undoubtedly that of McCann and Besner (1987; see also Peressotti & Colombo, 2011). In their study, when pseudohomophones and nonhomophonic pseudoword controls were mixed in the same list, the results were similar to ours, i.e., faster reaction times for pseudohomophones than for nonhomophonic pseudowords controls and no baseword frequency effect.

Based on the results of the present experiment and those of previous studies (McCann & Besner, 1987; Peressotti & Colombo, 2011), we suggest that the presence of pseudohomophones without the presence of real homophones might have encouraged the participants to process the phonology of the written stimuli more deeply. If this were the case, it would explain why the pseudohomophones were processed faster than the orthographic control pseudowords, as they would be in a reading-aloud task (Seidenberg, et al., 1996). In addition, because the effects of the frequency of the baseword are opposite when reading aloud compared to when making a lexical decision (Ziegler et al., 2001), the inhibition of the orthographic lexical representation of very frequent basewords (when processing the phonology) would be masked by the facilitation of the verification process.

The effects of this reading strategy would have had little effect on lexical decisions on the high and low-frequency real words since the information from both the phonological and semantic higher-order areas matched the information from the visual input.

4. General discussion

The first important observation resulting from the present study is that the dynamic nature of the mechanisms involved in the lexical decision process was confirmed (see Barca & Pezzulo, 2012) and is consistent in both Italian and French. Most of the effects could not be observed at the initiation of the response movement, but occurred during this movement. This means that the competition between information sources - or between top-down predictions and bottom-up information in the IAC model (Price & Devlin, 2011) or the LCA model (Dufau et al., 2012) - was active during the execution of the response movement, which is in keeping with dynamic models that see decision and action as intertwined and not sequential processes (Lepora & Pezzulo, 2015; Pezzulo & Cisek, 2016; Pezzulo et al., 2015; Spivey, 2007).

Another observation was that lexicality could be arranged along a “lexicality dimension line”. The fact that the best performances were observed for high-frequency words and strings of letters, and the worst performance for low-frequency words and pseudowords, suggests that the different components of this “lexicality dimension line” are arranged as in Barca and Pezzulo (2012): the

most lexical stimuli were the high-frequency words, followed by the low-frequency words and then the pseudowords and finally the letter strings, which were the least lexical of the stimuli.

As in Barca and Pezzulo (2012), a lexicality effect was observed, but this effect seemed to be greater in the present study. This observation, combined with the observation of a frequency effect that also seemed to be greater and a lexicality effect between high-frequency words and strings of letters, suggests that French participants might rely more on the semantic information from higher-order areas than Italian participants do when processing written words. This way of processing words is typical in feedforward and feedback-inconsistent orthographies (see Lima & Castro, 2010, for a review). Since the French language has a relatively feedforward-consistent and feedback-inconsistent orthography, our results suggest that the conflict between phonology and spelling in other languages with a highly inconsistent orthography occurs in a kind of a verification process at the point when spelling is being retrieved from phonology (see Ziegler, et al., 2001, for a similar proposal).

The results of Experiment 2 using pseudohomophones were surprising as they violated the predictions of all the previously formulated hypotheses - that is, better performance for pseudohomophones than for nonhomophonic pseudowords and no baseword frequency effect for pseudohomophones. Like McCann and Besner (1987) and Peressotti and Colombo (2011), we attribute this effect to list composition. The presence of pseudohomophones with nonhomophonic pseudoword controls but no homophones and/or proper basewords may have induced the participants to adopt a different reading strategy involving a deeper activation of the phonology of the written stimuli. This would explain why, as in a reading-aloud task (Seidenberg, et al., 1996; Grainger et al., 2000), pseudohomophones were processed faster than nonhomophonic controls. The absence of a baseword frequency effect could be explained as follows. On one hand, the orthographic representation is more highly activated for high than for low-frequency basewords and this needs to be inhibited when the stimulus is about to be read. On the other, high-frequency basewords facilitate the verification process more than low-frequency basewords because their orthographic representation is accessed earlier and is more accurate. These inhibition and facilitation effects would cancel each other out. This mechanism would explain our results and those of other studies in which the use of pseudohomophones was similar to ours (McCann & Besner, 1987; Peressotti & Colombo, 2011).

A final point of interest in this study concerned the repetition effect. Most previous studies using a lexical decision task have assumed that the participants' responses to the same items across various presentations are invariable. However, Diependale et al. (2012) demonstrated that variability in response to specific items is high across multiple presentations. Here, we tested whether the pattern of results would change when the same stimuli are presented a second time.

More precisely, we checked for a learning effect or a change in the response criteria. Our results indicate that the participants did not learn the stimuli from the first presentation but instead exhibited a lower response threshold. This result has implications for studies in which the same item has to be repeated several times, as in electrophysiological and neuroimaging studies where multiple presentations of the same stimulus are necessary in order to improve the signal-to-noise ratio or are required by the experimental paradigm (e.g., Barca, Cornelissen, Simpson, Urooj, Woods, Ellis, 2011). The pattern of results observed in such studies would reflect a process that differs from a first-time presentation. An avenue for future research will be to directly compare languages having different levels of feedforward and feedback orthographic consistency. A direct comparison would make it possible to test in more detail our hypothesis of a feedback inconsistency effect, or show that it operates jointly with other effects. As an example, greater lexicality and frequency effects could imply that two conflicts between spelling-orthographic correspondences have to be resolved, one in the feedforward direction and the other in the feedback direction. Until a direct comparison with the data obtained from feedforward and feedback-inconsistent languages is performed, the generalization of the present results to such languages will require caution.

5. Acknowledgement

The author Marc Ouellet would like to acknowledge the support of a Juan de la Cierva Postdoctoral fellowship (JCI-2012-13046) from the Spanish MEC.

6. References

- Baayen, R. H., Davidson, D. J., & Bates, M. D. (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.
- Barca, L., Benedetti, F., Pezzulo, G. (2015). The effects of phonological similarity on the semantic categorization of pictorial and lexical stimuli: evidence from continuous behavioral measures. *Journal of Cognitive Psychology*, 28, 2, 159-170.
- Barca, L., Cornelissen, P., Simpson, M., Urooj, U., Woods, W., Ellis, A. W. (2011). The neural basis of the right visual field advantage in reading: An MEG analysis using virtual electrodes. *Brain & Language*, 118, 53–71.
- Barca, L., & Pezzulo, G. (2012). Unfolding visual lexical decision in time. *PLoS ONE*, 7(4), e35932.
- Barca, L., & Pezzulo, G. (2015). Second thoughts: Continuous and discrete revision processes during visual lexical decision. *PLoS ONE*, 10(2), e0116193.
- Barca, L., Pezzulo, G., Castrataro, M., Rinaldi, P. & Caselli, M. C. (2013). Visual word recognition in deaf readers: lexicality is modulated by communication mode. *PLoS ONE*, 8(3), e59080.
- Bates, D., & Maechler, M. (2009) *lme4: Linear mixed-effects models using s4 classes* [Computer software manual]. Available: <http://CRAN.R-project.org/package=lme4> (R package version 0.999375–32).
- Brysbaert, M. (2007). *'The language-as-fixed-effect fallacy': Some simple SPSS solutions to a complex problem (Version 2.0)*. Royal Holloway, University of London.
- Diependale, K., Brysbaert, M. & Neri, P. (2012). How noisy is the lexical decision? *Frontiers in Psychology*, 3: 348.
- Dufau, S., Grainger, J., & Ziegler, J. C. (2012). How to say "no" to a nonword: A leaky competing accumulator model of lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 1117-1128.
- Duyck, W., Desmet, T., Verbeke, L., & Brysbaert, M. (2004). WordGen: A Tool for Word Selection and Non-Word Generation in Dutch, German, English, and French. *Behavior Research Methods, Instruments & Computers*, 36, 488-499.
- Farioli, F., Grainger, J., & Ferrand, L. (2011). PHOM: une base de données de 14 000 pseudohomophones [PHOM: A database of 14,000 pseudohomophones]. *L'Année Psychologique*, 111, 725-751.
- Ferrand, L. & Grainger, J. (2003). Homophone interference effects in visual word recognition. *The Quarterly Journal of Experimental Psychology*, 56A, 403–419.
- Ferrand, L., Bonin, P., Méot, A., Augustinova, M., New, B., Pallier, C., & Brysbaert, M. (2008). Age of acquisition and subjective frequency estimates for all generally known monosyllabic

- French words and their relation with other psycholinguistic variables. *Behavior Research Methods*, 40, 1049-1054.
- Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., Augustinova, M., & Pallier, C. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, 42, 488-496.
- Ferrand, L., Brysbaert, M., Keuleers, E., New, B., Bonin, P., Méot, A., Augustinova, M., & Pallier, C. (2011). Comparing word processing times in naming, lexical decision, and progressive demasking: Evidence from Chronolex. *Frontiers in Psychology*, 2:306. doi:10.3389/fpsyg.2011.00306
- Flumini, A., Barca, L., Pezzulo, G., & Borghi, A. (2014). How do you hold your mouse? Tracking the compatibility effect between hand posture and stimulus size. *Psychological Research*, 79, 6, 928-938.
- Freeman, J. B., & Ambady, N. (2010) Mousetracker: software for studying real-time mental processing using a computer mouse-tracking method. *Behavioral Research Methods* 42, 226–241.
- Grainger, J., Spinelli, E., & Ferrand, L. (2000). Effects of baseword frequency and orthographic neighborhood size in pseudohomophone naming. *Journal of Memory and Language*, 42, 88-102.
- Herdman, C. M., Lefevre, J. -A., & Greeham, S. L. (1996). Base-word frequency and pseudohomophone naming. *Quarterly Journal of Experimental Psychology*, 49A, 1044-1061.
- Kinoshita, S., Lupker, S. J., & Rastle, K. (2004). Modulation of regularity and lexicality effects in reading aloud. *Memory and Cognition*, 32, 1255-1264.
- Kuznetsova, A., Brockhoff, P.B., & Christensen, R.H.B. (2016). *Package 'lmerTest'* [package manual]. Available: <https://cran.r-project.org/web/packages/lmerTest/index.html>
- Lepora, N., Pezzulo, G. (2015) Embodied choice: How action influences perceptual decision making. *PLOS Computational Biology*, 11(4): e1004110
- Lima, C. F., & Castro, S. L. (2010). Reading strategies in orthographies of intermediate depth are flexible: Modulation of length effects in Portuguese. *European Journal of Cognitive Psychology*, 22, 190-215.
- Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 23, 570-590.
- McCann, R. S., & Besner, D. (1987). Reading pseudohomophones: Implications for models of pronunciation assembly and the locus of word-frequency effects in naming. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 14–24.

- Pagliuca, G., Arduino, L. S., Barca, L., & Burani, C. (2008). Fully transparent orthography, yet lexical reading aloud: New evidence from Italian. *Language and Cognitive Processes*, 23, 422-433.
- Peressotti, F., & Colombo, L. (2011). Reading aloud pseudohomophones in Italian: Always an advantage. *Memory and Cognition*, 40, 466-82.
- Pezzulo, G., & Cisek P. (2016) Navigating the Affordance Landscape: Feedback Control as a Process Model of Behavior and Cognition. *Trends in Cognitive Sciences* 20 (6), 414-424
- Pezzulo, G., Rigoli, F., & Friston, K. (2015) Active Inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology* 134, 17-35
- Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, 117, 864–901.
- Price, C. J., & Devlin, J. T. (2011). The Interactive Activation Account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences*, 15, 246–53.
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, 20, 873–922.
- Schmalz, X., Marinus, E., Coltheart, M., & Castles, A. (2015). Getting to the bottom of orthographic depth. *Psychonomic Bulletin and Review*, 22, 1614-1629.
- Seidenberg, M. S., Petersen, A., MacDonald, M. C., & Plaut, D. C. (1996). Pseudohomophone effects and models of word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 48-62.
- Spivey, M. J. (2007). The continuity of mind. *Oxford University Press*, USA.
- Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proc Natl Acad Sci U S A*, 102, 10393–10398.
- Stone, G. S., & Van Orden, G. (1993). Strategic control of processing in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 744-774.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, 108, 550-592.
- Ziegler, J. C., Jacobs, A. M., & Klüppel, D. (2001). Pseudohomophone effects in lexical decision: Still a challenge for current word recognition models. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 547-559.
- Ziegler, J. C., Jacobs, A. M., & Stone, G. (1996). Statistical analysis of the bidirectional inconsistency of spelling and sound in French. *Behavior Research Methods, Instruments, & Computers*, 28, 504-515.
- Ziegler, J. C., Perry, C., & Coltheart, M. (2003). Speed of lexical and nonlexical processing in French: The case of the regularity effect. *Psychonomic Bulletin & Review*, 10, 947-953.

Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). What's the pronunciation for -OUGH and the spelling for /u/? A database for computing feedforward and feedback inconsistency in English. *Behavior Research Methods, Instruments, & Computers*, 29, 600-618.

7. Appendix 1

Stimuli used in Experiment 1

High Frequency words	Low Frequency words	Legal pseudowords	Illegal strings of letters
arbre	aigle	aucle	lnvlc
boîte	berge	merve	cfjdp
carte	bille	sulle	fgnsm
chef	bûche	fûgne	rcsjk
chien	bulle	rille	jmbtd
choix	cidre	pivre	gjevqm
ciel	crabe	flace	dcshp
fête	crêpe	grêde	stqm
goût	cube	puje	fcjt
livre	cycle	pycre	dtldr
lune	fable	tadre	rfsb
paix	flûte	brûre	vgdr
pièce	frein	froin	iéyao
pied	gant	galt	eoyîi
pluie	grue	grou	ioyuè
pont	myope	myage	eaêiu
rêve	niche	voche	aioya
robe	ogre	ébre	eiaué
route	perle	parne	ueyoa
sœur	pneu	freu	aeô
table	pôle	rèle	iuoe
train	rame	bage	yaeu
vent	ride	bime	ouaé
ville	veuf	vauf	auéi

8. Appendix 2

Stimuli used in Experiment 2

Pseudohomophones	Orthographic controls
feible	foible
forse	forpe
fylm	falm
juje	jube
onkle	onfle
paine	poine
prize	prine
rouje	roude
roze	rove
sible	rible
sukre	supre
vyvre	vavre
fraize	fraipe
garce	garte
greine	groine
jifle	tifle
klou	plou
meigre	moigre
naje	nale
pakte	palte
sime	pime
thèze	thène
tygre	tagre
vygne	vagne