

Afonso, O, Alvarez, C and Kandel, S

Effects of grapheme-to-phoneme probability on writing durations

Afonso, O, Alvarez, C and Kandel, S (2015) Effects of grapheme-to-phoneme probability on writing durations. *Memory and Cognition*, 43 (4). pp. 579-592.

doi: 10.3758/s13421-014-0489-8

This version is available: <https://radar.brookes.ac.uk/radar/items/90ae8c57-35cd-4ff2-86de-927266e4b868/1/>

Available on RADAR: July 2016

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

This document is the postprint version of the journal article. Some differences between the published version and this version may remain and you are advised to consult the published version if you wish to cite from it.

Effects of Grapheme-to-Phoneme Probability on Writing Durations

Olivia Afonso¹, Carlos J. Álvarez¹, and Sonia Kandel^{2, 3, 4}

¹Universidad de La Laguna, Tenerife, Spain

²Université Grenoble Alpes, CNRS LPNC UMR 5105, Grenoble, France

³Université Grenoble Alpes, GIPSA-LAB (CNRS UMR 5216), Département Parole-
Cognition, Grenoble, France

⁴Institut Universitaire de France, Paris, France

Correspondence concerning this article should be addressed to Olivia Afonso,
Departamento de Psicología Cognitiva, Social y Organizacional, Campus de Guajara,
Universidad de La Laguna, 38205, Santa Cruz de Tenerife, Spain.

Phone: 0034600269934

E-mail: oafonso@ull.es

Abstract

The relative involvement of the lexical and sublexical routes across different writing tasks remains a controversial topic in the field of handwriting production research. The present article reports two experiments examining whether or not the probability of a grapheme-to-phoneme (G-P) mapping affected production during copy of polyvalent graphemes embedded in French (Experiment 1a and 1b) and Spanish (Experiment 2) known words. The relative probability of two different G-P mappings associated to the same polyvalent grapheme was manipulated (higher probability vs. lower probability). In Experiment 1a, we used the polyvalent French grapheme *E*. Writing durations revealed that the inter-letter intervals (ILIs) located before and after this letter were shorter and that the letter itself was executed faster in the condition of higher-probability of the G-P mapping (*SERVICE*, *service*) than in the lower-probability condition (*SEMAINE*, *week*). In Experiment 1b, we used the sequence *TI* (*VICTIME-MARTIEN*, *victim-Martian*), which is less frequent. In this case we failed to observe significant differences between conditions. Similar effects to those obtained in Experiment 1a were found with Spanish words in Experiment 2 using different pronunciations of the letter *C* (*DESCANSO-DESCENSO*, *rest-descent*). Altogether, these results reveal that the link between a grapheme and a phoneme is weighted according to its probability in the language. Moreover, they suggest that a two-phase route linking graphemes to phonemes and phonemes to graphemes is functional during copy.

For many years, the cognitive study of handwriting production was very limited in comparison to research conducted in the fields of language comprehension and speech production. This relative neglect has been progressively overcome, leading to a steady growth of evidence collected and theoretical models proposed about the handwriting production process. One of the most controversial topics in the literature has been the relative involvement of lexical and sublexical information during different writing tasks.

According to dual process theories, spelling may be achieved through two different processing routes. The so-called lexical route gives access to the spelling of whole-words from long-term memory so it would be used when spelling familiar words. In contrast, the sublexical route or assembled route makes use of knowledge about the links between phonology and orthography and provides a phonologically plausible spelling for nonwords or low-frequency words (Caramazza, 1988; Tainturier & Rapp, 2001). Although the existence of both routes is almost undisputed, it is less clear whether or not their use is mutually exclusive during writing. Several studies suggest that the lexical route might influence nonword spelling (Barry & Seymour, 1988; Campbell, 1983; Cuetos, 1991). However, less evidence has been provided about the influence of the assembled route during the written production of well-known words.

Additionally, little is known about the role played by each route when the orthographic form of the stimulus is present, as in the direct copy transcoding task. Recent studies have shown that sublexical information mediates several types of writing tasks (Afonso & Álvarez, 2011; Bonin, Peereman, & Fayol, 2001; Delattre, Bonin, & Barry, 2006; Qu, Damian, Zhang, & Zhu, 2011), and theoretical models have been proposed to describe the relative involvement of the lexical and the sublexical route during spelling-to-dictation (Tainturier & Rapp, 2001; see also Folk, Rapp, & Goldrick, 2002; Folk & Rapp, 2004; Rapp, Epstein, Tainturier, 2002) and/or written picture naming (Bonin, Peereman, & Fayol, 2001;

Roux & Bonin, 2011). However, the involvement of sublexical correspondences during copying has been investigated less than in other writing tasks. This is mainly due to extensive agreement about the possibility of performing this task without phonological mediation, by simply converting to graphemes the visual input (Fischer-Baum, & Rapp, 2012), through lexical access, or even resorting exclusively to visual information (Cuetos, 1991).

Copying involves two groups of processes: reading and spelling processes. When a known word has to be copied, the visual input activates a lexical entry, which in turn activates a semantic representation. Spelling processes would then lead to the retrieval of the orthographic form stored in the output lexicon corresponding to that semantic representation. Although other processing routes for copying might be available, this route is thought to be the most commonly used by an adult writer (Cuetos, 1991; Jiménez & Muñetón-Ayala, 2002). This means that the copying task is usually considered to be performed via lexical access, so an impact of the assembled route during writing is not expected. Some authors have proposed the existence of a sublexical (phonologically mediated) route for copying (Cuetos, 1991). However, such a route has been claimed to have little impact on handwriting and only in very limited circumstances, such as in the case of children, because they lack strong lexical representations.

The model introduced by Cuetos (1991) explicitly describes the mechanisms that are thought to underlie the copying task (see Figure 1). The author takes into account at least three different linguistic¹ processing routes that might be engaged during copy: a non-phonologically mediated route and two phonological routes.

¹It has also been claimed that a non-linguistic route is available for copying. This route would consist in the reproduction of letters as meaningless forms (just like replicating a drawing). Because this route would not engage any linguistic process, we will not discuss it in further detail. In any case, this route is highly unlikely to be used in normal skillful writing, since it would be extremely slow and resource-consuming.

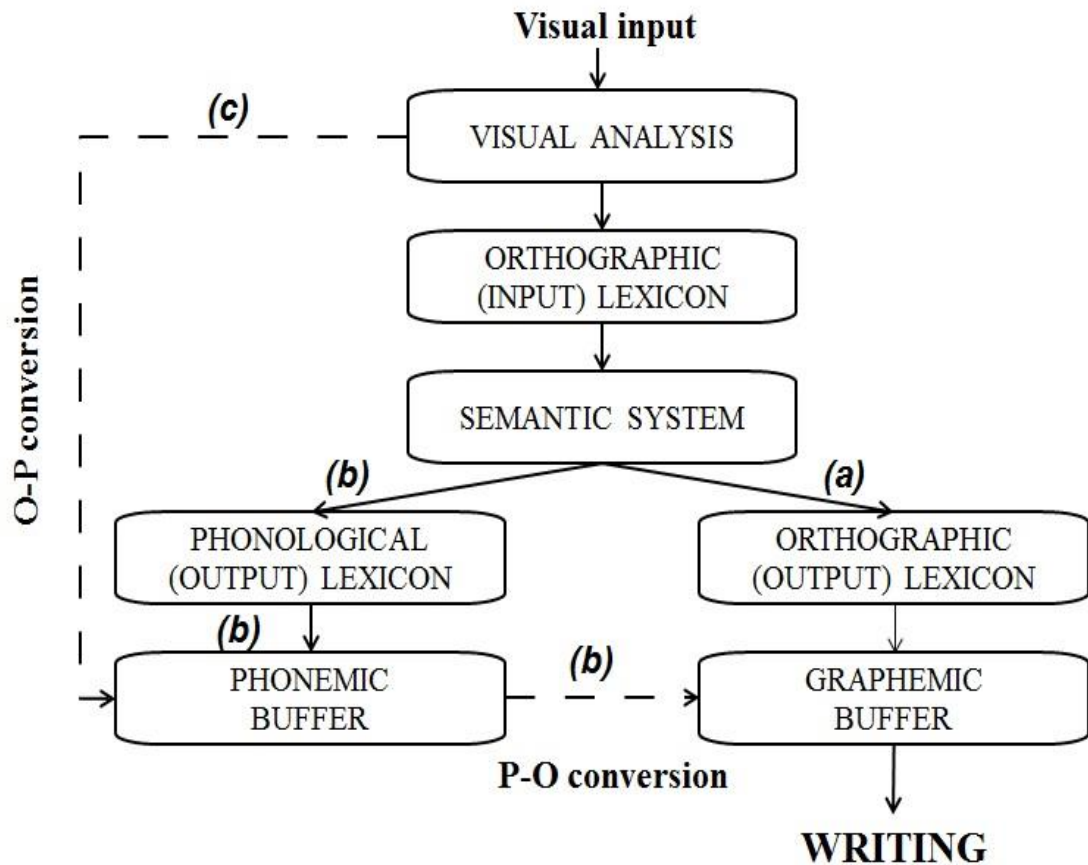


Figure 1. Model of copying, adapted from Cuetos, 2001. The dotted lines represent the routes that are thought to be used for spelling nonwords.

- (a) The lexical route is the most common path for copying. The appropriate orthographic lexical form is directly accessed via the semantic system, just after the visual input has been decoded.
- (b) In this route a semantic representation is also activated as a result of the reading process. This semantic representation, besides leading to the retrieval of the corresponding word-form from the orthographic output lexicon, would also activate a word-form in the phonological output lexicon. The individual phonemes constituting this phonological word-form would then be identified and kept in a sort of phonological buffer (called the pronunciation buffer). The phonemes maintained in this buffer would activate the corresponding graphemes according to the phoneme-to-grapheme conversion patterns permitted in the language.

(c) Finally, individual letters identified by means of the visual analysis could activate their corresponding phonemes² through orthography-to-phonology (O-P) conversion patterns. Subsequent to the retrieval of the associated phonemes, P-O correspondences would activate a phonologically plausible graphemic candidate for each phoneme. This route is claimed to operate in the absence of semantic/lexical activation, so it might produce misspellings when orthographically ambiguous phonemes (e.g., phonemes with more than one possible graphemic representation) are present in the phonological form of the input. It is proposed that the use of this route by young children explains why they produce misspellings even during copying.

Although the existence of these routes has been virtually undisputed, the use of (b) and (c) is thought to be rather limited in normal writing conditions. As mentioned above, when faced with well-known words, adult writers are likely to retrieve the corresponding word-form directly from the orthographic lexicon, so the sublexical/phonological routes for copying would be used almost exclusively by adults when they copy pseudowords (Cuetos, 1991), or by dysgraphic patients with serious damage to the lexical system. Therefore, when normal adults copy words, phonological information is believed to have little or no impact on writing latencies or durations.

In this study we aim to provide evidence about the use of a sublexical route such as (c) when adults perform a direct copy transcoding task. Specifically, we investigate whether or not the probability of G-P correspondences affect writing durations, as predicted if this route is functional.

²We use the term phoneme for the sake of clarity. Although it is widely accepted that the sublexical system links phonemes to graphemes, the precise nature (and size) of the units involved during the P-O conversion process remains unclear (Folk & Rapp, 2004).

Organization of the Sublexical System

Phonology-to-orthography (P-O) conversion patterns are thought to consist of phoneme-to-grapheme mappings weighted according to their frequency in the language. Specifically, it is claimed that the more frequent a grapheme is, the more accessible it becomes. This idea has been supported by psycholinguistic and neuropsychological evidence (Miceli, 1989; Cuetos, 2011; Barry & Seymour, 1988; Sanders & Caramazza, 1990; Goodman & Caramazza, 1986). For example, Barry and Seymour (1988) observed that P-O probability affected the spelling of nonwords. In their study, English speakers were more likely to spell the phoneme /i/ as *ee* (consistent with the most frequent mapping in the language) than as *ie* (a lower-frequency mapping). Moreover, Sanders and Caramazza (1990) reported the case of a dysgraphic patient who relied on sublexical processes to spell many words. An analysis of his phonologically plausible errors (PPEs) revealed that the relative frequency of a grapheme correlated with the frequency with which that grapheme was produced by the patient when spelling an inconsistent phoneme. This effect has been observed to interact with lexical processes during the spelling of nonwords (Barry & Seymour, 1988; Cuetos, 2011; Folk & Rapp, 2004; Tainturier, Bosse, Valdois, & Rapp, 2000; Bosse, Valdois, & Tainturier, 2003).

Moreover, phoneme-to-grapheme probability (P-G) seems to have an effect not only on nonwords, but also on word spelling. The so-called regularity effect refers to the fact that there is a processing advantage for words containing high-probability P-G mappings over low-probability P-G mappings, as manifested by shorter latencies and writing durations and/or less errors. This effect has been observed in the spelling-to-dictation task (Delattre et al., 2006), in the evaluation of some dysgraphic patients (Rapp et al., 2002), and, more importantly for the present study, in the copying task (Kandel & Valdois, 2005; Roux, McKeef, Grosjacques, Afonso, & Kandel, 2013). These convergent results have been

interpreted as evidence for the integration of the output coming from both the lexical and the sublexical route and they have led to a general consensus about the frequency-weighted organization of P-O mappings.

Likewise, the probability of the G-P mappings might well have an impact on the accessibility of the intended graphemes. In the context of the model of copying proposed by Cuetos (1991) and depicted in Figure 1, an effect of P-G probability is expected whether either route (b) or (c) is used to perform the task, since in both cases phonemes are activated before the corresponding graphemes are accessed and kept in the graphemic buffer. However, a G-P probability effect is predicted only in the case of route (c). In the case of this route being used, a more frequent G-P mapping would lead to stronger links between the grapheme and the phoneme. Thus, this phoneme would be retrieved more easily and would be mapped faster onto its corresponding grapheme. The impact of G-P probability has, however, not been experimentally addressed until now, which is hardly surprising, considering the paucity of experimental evidence regarding the organization of the sublexical route.

The finding of an effect of grapheme-to-phoneme probability during word copying would, therefore, confirm that sublexical information is functional during this task. More importantly, it would manifest that the sublexical route is sensitive not only to P-G probability, but also to G-P probability, indicating that writers convert the letters in the input to phonemes, and that these phonemes give access to the graphemes that have to be written. In the present study, we address this issue by testing graphemes that are phonologically ambiguous; so-called polyvalent³ graphemes (Alarcos-Llorach, 2011; Jiménez & Muñetón-Ayala, 2002) or feedback inconsistent graphemes (Davies & Weekes, 2005). If only P-G

³ In the present study we chose the term ‘polyvalent grapheme’ instead of the more frequent term ‘feedback/feedforward consistency’. The main reason is that the terms ‘feedback’ and ‘feedforward’ have to be inconsistently used depending on the task. Most of the studies dealing with this variable have been reading or naming studies, so feedback consistency is used to refer to P->O mappings. But in handwriting P->O mapping are feedforward connections. In contrast, the graphemes tested in our study are polyvalent graphemes independently of the input or output modality. Moreover, the term feedback consistency is strongly linked to P->O consistency in the literature, so we think that its use could lead to confusion.

probability has an impact on the organization of the sublexical system, a polyvalent grapheme should be retrieved equally fast/slowly regardless of its specific pronunciation in a given word. However, if polyvalent graphemes are accessed faster at the grapheme level when representing the more frequent pronunciation, this would suggest that G-P correspondences are functional in the course of copy.

Polyvalent Graphemes

Whereas monovalent graphemes represent the same phoneme in any context (e.g., the grapheme *d* in Spanish, which is always pronounced /d/), polyvalent graphemes correspond to different phonemes depending on the surrounding letters or their relative position within a word. For instance, the grapheme *c* is polyvalent in Spanish because it may represent either the sound /k/ or /θ/ depending on the following letter (e.g., in *casa* –house– and *cesta* –basket–, /kasa/ and /θesta/ respectively). This particularity of polyvalent graphemes provides an incomparable opportunity to test two different phonemes (with different G-P probabilities) by measuring the writing duration of the same letter. For this reason polyvalent graphemes have been used in several studies to test related phenomena, such as grapheme complexity (Kandel & Spinelli, 2010) or graphemic cohesion (Spinelli, Kandel, Guerassimovitch, & Ferrand, 2012). They have also been used to study the effects of feedback/feedforward consistency in visual word recognition, spoken word production and writing (Stone, Vanhoy, & Van Orden, 1997; Ziegler, Montant, & Jacobs, 1997; Peereman, Content, & Bonin, 1998).

In the present research, we used polyvalent graphemes to test the hypothesis that G-P probability has an impact on the retrieval of the corresponding grapheme. If this is indeed the case, it has major implications for current models of handwriting, since only the use of route (c), in which P-G and G-P correspondences are applied, would predict such an effect.

We decided to address this issue in French, a language in which polyvalent graphemes are relatively common. French has a highly opaque orthography, especially concerning P-O

correspondences (Ziegler, Jacobs, & Stone, 1996). Whereas spelling in a transparent orthography may be successfully accomplished by resorting exclusively to sublexical processes, writers of a language with an opaque orthography must possess accurate orthographic (lexical) knowledge to avoid misspellings. In other words, the involvement of the sublexical route during the spelling process in French could be highly counterproductive, since it would frequently lead to errors. This means that in the following experiments we are actually stacking the deck against the possibility of finding a reliable G-P probability effect.

Experiment 1

We aimed to test whether a given grapheme is written faster when it represents a more typical pronunciation (higher G-P probability) than a less typical pronunciation (lower G-P probability). Movement duration data have indicated that stroke durations are shorter for strongly cohesive graphemes than for weakly cohesive graphemes (Spinelli et al., 2012). A strongly cohesive grapheme is a sequence of letters representing a complex grapheme that is systematically associated to the same phoneme. For example, the sequence *AU* in French is a strongly cohesive grapheme because it is always pronounced /o/. Conversely, weakly cohesive graphemes are those sequences, such as *AN*, that can be either a complex grapheme associated to one phoneme (/ã/ in *CRAN*) or two simple graphemes (/a/+n/ in *CANE*). The results reported by Spinelli and colleagues (2012) suggested that writing durations are sensitive to the orthography-to-phonology consistency of graphemes. Accordingly, frequent and infrequent G-P correspondences may yield different writing durations. The rationale is that the connection between a grapheme and a phoneme is stronger in the case of high-probability G-P correspondences than in the case of low-probability correspondences. Thus, if G-P correspondences are involved during the copying process as described in route (c), this fact would represent an advantage for more probable phonemes in comparison to less probable ones. More precisely, we predict that, given the same orthographic sequence, this

will be produced faster when representing a typical than an atypical sound for that sequence. For example, in French the grapheme *E* is most frequently pronounced /ɛ/, such as the first *E* in *SERVICE* (*service*), than /ø/, like in the first *E* in *SEMAINE* (*week*). If phonemes are activated by the graphemes in the visual input according to probabilistic principles and these phonemes activate the to-be-produced graphemes, then /ɛ/ (a higher-probability G-P mapping) would give faster access to the letter *E* than /ø/ (a lower-probability G-P mapping).

We hypothesize that a potential processing advantage of higher-probability G-P mappings over lower-probability mappings would affect central processes involved in the retrieval of phonemes and subsequent access to the corresponding graphemes. In handwriting theory, central processes have been considered to embrace those processes related to the retrieval and activation of an orthographic representation, whereas peripheral processes are concerned with the regulation of parameters of the motor response such as amplitude, orientation or force. The anticipatory vision of handwriting proposed by Van Galen and defended by other authors (Van Galen, 1991; Kandel, Peereman, Grosjacques, & Fayol, 2011) states that manipulations at the central levels of processing produce differences in writing durations corresponding to previous segments of the to-be-written sequence. This is because Van Galen's model (1991) proposes that central processes and peripheral processes are engaged in parallel but with central processes dealing with forthcoming parts of the response. Thus, and in consonance with the anticipatory vision of handwriting, we predict that an effect of the probability of G-P mapping will be observable in the writing durations obtained for the segments previous to the real-time execution of the target grapheme.

However, the influence of the probability of the G-P mapping during copying may be relatively weak, so it is possible that only those mappings which are very frequent in the language are strong enough to result in the activation of the related phoneme in time to produce a significant effect. In Experiment 1a we tested a very frequent grapheme in French:

letter *E* (*BERCEAU* versus *BESOGNE*, *cradle* and *labor* respectively). In Experiment 1b, we used a less frequent orthographic form (*TI* in *VICTIME* versus *MARTIEN*, *victim* and *Martian* respectively). We predict that the same sequence would be retrieved faster in the context of a higher-probability G-P mapping than in a lower-probability mapping.

Experiment 1a

Method.

Participants. Twenty-five students from Psychology introductory courses at the Université Pierre Mendès France took part in this experiment to fulfill a course credit requirement. All of them were native French speakers, right-handed and with no known motor or perceptive disorders. In all cases, French was their first language and they have little or no knowledge of other languages.

Materials. We selected fifty-eight stimuli containing the letter *E*. Two conditions were created according to the probability of the G-P correspondence (higher-probability versus lower-probability). We calculated the probability of the G-P mapping as follows: First, we counted the total number of times that letter *E* appeared in the words included in Lexique 3 (New, Pallier, & Ferrand, 2005). Second, we calculated the percentage of times where this letter was pronounced in a particular way. Then, we selected the two phonemic forms with the most extreme values. /ɛ/ was selected for the higher-probability condition (71% of times *E* is pronounced this way) and /ø/ was selected for the lower-probability condition (*E* is pronounced this way only 9% of times). Letter *E* was pronounced /ɛ/ in half of the experimental stimuli and /ø/ in the other half. We matched both conditions according to the position of the first *E* in the word, as well as the number of times that letter *E* was included in the word. Lexical frequency, frequency of the bigrams before and after the target sequence (e.g. in the word *BERCEAU* –[bɛrso]–, the bigrams “*BE*” and “*ER*”), word length (number of

letters and number of phonemes), orthographic neighborhood, orthographic uniqueness point and number of syllables were controlled according to the values provided by Lexique 3 (New et al., 2005). T-tests were conducted to make sure that these variables did not significantly differ across conditions (all t s < 1). The full set of stimuli with their values in these controlled variables are given in Appendix A. Ninety additional words were selected to serve as fillers, plus three more for the practice phase.

Apparatus and procedure. Stimuli presentation and digital recording of the responses were controlled by Ductus (Guinet & Kandel, 2010). The experiment was run on an Asus F9Eseries laptop. The experiment consisted of a copying task and it was conducted individually in a sound-proof room. Each trial started with a 200-ms fixation point (+) in the center of the screen immediately followed by the presentation of a centered 16 point lower-case word. The participants had to write the word in upper case (print handwriting was not enforced) on a lined sheet of paper placed over a digitizer (Wacom Intuos LD-1218-u) with a Intuos Inking Pen as soon and as accurately as possible. When they finished a given response, participants were instructed to place the pen over the line corresponding to the next word (just below their previous response) without making any contact with the paper. Then, the experimenter clicked the left button of the mouse to lead to a new stimulus. The to-be-copied word remained onscreen until this moment. A whole experimental session lasted around 15 minutes.

Results.

Writing durations for critical letter *E* (henceforth, LD), and the previous and following inter-letter intervals (ILI0 and ILI1, respectively) were submitted to separate analyses of variance (ANOVAs), with the probability of the G-P correspondence (higher vs. lower) as a within-subjects variable in the analysis by participants (F_1) and a between-items variable in the analysis by items (F_2). LD was measured as the time between the first contact of the pen

with the tablet for a letter and the last lift in that letter. An ILI was defined as the time between the last pen lift in a letter and the first pen down in the following letter. Responses containing misspellings and hesitations or those in which a recording error occurred were considered as errors and removed from the analysis (2.33% of the data).

A significant effect of G-P probability was observed in ILI0 durations ($F_1(1, 24) = 9.52, p < .005, MSE = 191.7, \eta_p^2 = .28$; $F_2(1, 56) = 5.85, p < .05, MSE = 286.91, \eta_p^2 = .09$), ILI1 ($F_1(1, 24) = 31.03, p < .001, MSE = 2,118.17, \eta_p^2 = .56$; $F_2(1, 56) = 18.95, p < .001, MSE = 1,981.4, \eta_p^2 = .25$), and LD ($F_1(1, 24) = 8.43, p < .01, MSE = 327.64, \eta_p^2 = .26$; $F_2(1, 56) = 5.12, p < .05, MSE = 635.59, \eta_p^2 = .08$). All three measures were faster in the condition of higher-probability of the G-P mapping. As shown in Table 1, participants produced a shorter inter-letter interval before and after the target grapheme *E*, which was also executed faster when it corresponded to the most frequent pronunciation.

G-P probability	ILI0	LD	ILI1
Higher (/ɛ/)	140	541	123
Lower (/ø/)	144	546	136

Table 1. Writing durations for the critical letter *E* (LD) and durations of the previous and following inter-letter intervals (ILI0, ILI1) in milliseconds, in Experiment 1a.

Experiment 1b

Experiment 1a showed an effect of G-P probability when the letter *E* had to be written. However, we suspected that such a pattern of results could be due to the fact that /ø/ is a highly unusual pronunciation (only 9%). The impact of a sublexical route for copying might be rather weak, especially in a language with an opaque orthography, so maybe a strong bias towards one specific pronunciation is necessary to obtain a reliable effect in French. Moreover, it is also possible that only mappings with a high absolute frequency (those mappings that appear frequently in the language) can produce an effect in writing durations. A G-P mapping has to be applied very fast to produce an effect in the production of

the grapheme during copy, since lexical processes and visual processes are otherwise likely to activate the intended grapheme before a relatively slower sublexical route such as the one depicted in (c) in Figure 1 may have an influence. Therefore, in Experiment 1b we tested words that included the orthographic sequence *TI*. This sequence is associated to two different pronunciations in French: /ti/ (the higher-probability condition) and /sj/ (the lower-probability condition). However, and unlike the case of *E*, both pronunciations of *TI* are relatively frequent, so neither of them is a particularly rare pronunciation of *TI*. Additionally, *TI* is an orthographic sequence less frequent than letter *E* (*TI* only appears in 5.9% of the French words included in Lexique 3), so French speakers are faced with the mappings tested in Experiment 1b less often than in the case of the mappings used in Experiment 1a. Thus, in Experiment 1b we can test whether differences in the probability of each pronunciation or the absolute frequency of the grapheme modulate the G-P probability effect.

Method.

Participants. Twenty-five participants from the same population as in Experiment 1a took part in this experiment. None of them participated in any of the other experiments included in this manuscript.

Materials. Fifty-two experimental stimuli including the orthographic sequence *TI* were selected from Lexique 3 (New et al., 2005). For half of the stimuli this sequence represented the sound /ti/ (66% of times the sequence *TI* is pronounced /ti/; for example, *VICTIME* –[viktim]–, *victim*), and the other half represented the sound /sj/ (28.22% of times it is pronounced as in *MARTIEN* –[marsjɛ̃]–, *Martian*). This means that /sj/, which is the lower-probability G-P correspondence is not an infrequent mapping. We controlled the same variables taken into account in Experiment 1a (all *ts* < 1). The full set of stimuli with their values in these controlled variables are given in Appendix B. Ninety additional words were selected to serve as fillers, plus 3 more for the practice phase.

Procedure, apparatus and design. The experimental procedure, apparatus, design and statistical analyses were identical to those described in Experiment 1a.

Results.

The same exclusion criteria as in Experiment 1a were applied (2.45%). No differences in this experiment reached significance, either in ILIs or in LD (all F s < 1 , except ILI0: $F_1 = 1.61$, $MSE = 101.13$, $p = .22$, $1 - \beta = .23$). Mean values for all the collected measures are shown in Table 2.

G-P probability	ILI0	LD	ILI1
Higher (/ti/)	145	363	125
Lower (/sj/)	148	364	124

Table 2. Writing durations for the critical letter *T* (LD) and durations of the previous and following inter-letters intervals (ILI0, ILI1) in milliseconds, in Experiment 1b.

Discussion.

In Experiment 1 we addressed the potential effect of the grapheme-to-phoneme probability of different phonological correspondences of polyvalent graphemes. In Experiment 1a, the letter *E* was embedded in words in which it was pronounced /ɛ/ in the higher-probability condition (71%, approximately). In the lower frequency condition, letter *E* was pronounced /ø/, which is an infrequent sound for this letter in this position (only 9%). Results from Experiment 1a showed significant effects of G-P probability in the durations of the target letter *E* and in the durations of the ILIs preceding and following this letter. In contrast to the results obtained in Experiment 1a, those from Experiment 1b showed no effects of the probability of the G-P mappings. Neither the duration of letter *T* nor ILI duration was affected by our manipulation, suggesting that differences in the probability of G-P correspondence did not impact the retrieval or execution of the grapheme. However, we consider that the absence of significant effects could be attributed, at least, to two alternative

aspects. First, it could be due to the fact that none of the phonemic forms included in Experiment 1b is particularly improbable in French. As mentioned, /sj/ is the corresponding phonemic form for the orthographic sequence *TI* almost 30% of times, so this is not a low-probability mapping. Second, the absolute frequency of the mappings in the language is considerably higher in Experiment 1a, so maybe only very frequent mappings have strong links between graphemes and phonemes, capable of producing probability effects. We consider that the results from Experiment 1a may reflect the fact that G-P correspondences are weighted by their frequency in the language. However, the results from Experiment 1b would also indicate that this G-P probability effect is rather weak, so an effect would only arise in specific conditions.

The analysis of both ILI durations in Experiment 1a showed shorter durations for higher-probability G-P mappings. In ILI0, this effect might reveal the greater accessibility of letter *E* when activated by its typical phonological correspondence. However, the interpretation of the effect obtained in ILI1 is not straightforward. Letter *E* is pronounced /ø/ in open syllables. Thus, in the lower-frequency condition ILI1 always coincides with the syllable boundary (for example, *SE.MAIN*, [sømɛn]). In the words included in the higher-frequency condition (*SER.VICE*, [sɛʁvis]) this ILI is always intra-syllabic, so this effect could be attributable just to differences in the position of the syllable boundary. Longer ILIs have been repeatedly found in inter-syllable than in intra-syllable intervals (Álvarez, Cottrell, & Afonso, 2009; Kandel, Álvarez, & Vallée, 2006). This pattern would fit the effect observed in Experiment 1a for ILI1. It is worth noting that we do not think this explanation can account for the rest of effects obtained in Experiment 1a. The syllable boundary is placed in both conditions after the target letter, so stimuli are comparable until ILI1. Moreover, syllable boundary effects have been thoroughly detailed in previous studies, and in those studies using upper-case letters they have not been reported to affect letter durations or previous ILIs

(Kandel et al, 2006; Álvarez et al. 2009). In Experiment 2 we tested a polyvalent Spanish grapheme to establish the source of the results reported in Experiment 1.

Experiment 2

In Experiment 2, we aimed to establish whether the effects observed in Experiment 1a were actually due to our experimental manipulation (the relative probability of the G-P correspondences) or whether they were due to the position of the syllabic boundary. To this end, we conducted Experiment 2 in Spanish. Although different results could be obtained in different languages, we expected analogous results to those obtained in French, since similar phonological effects in handwriting have been reported in both languages (Álvarez et al., 2009; Kandel et al., 2006). In Experiment 2 we tried to replicate the G-P probability effects observed in Experiment 1a with the Spanish grapheme *C*, which is pronounced /k/ or /θ/ depending on the context. For example, in the Spanish word *DESCANSO* ([des'kanso], *rest*), letter *C* is pronounced /k/. However, in the word *DESCENSO* ([des'θenso], *descent*) it is pronounced /θ/. The sound /k/ is the most frequent phonological form of this grapheme according to BuscaPalabras (Davis & Perea, 2005): approximately 64.62% of times *C* is pronounced /k/; 28.49% of times it is pronounced /θ/ (in the remaining 6.89%, *C* is part of the complex grapheme *CH*, which is pronounced /tʃ/). Although these sequences are not so frequent as *E* from Experiment 1a, they are considerably more frequent than *TI* from Experiment 1b (approximately 38.34% of the words included in BuscaPalabras contain at least one *C*). Crucially, using these sequences we can generate two frequency conditions that do not differ in the position of the syllabic boundary (e.g., *DES.CAN.SO* vs. *DES.CEN.SO*). If the effect observed for *ILI1* in Experiment 1a was a syllabic boundary effect, then it should be absent in Experiment 2.

Method

Participants. Twenty-four students from Psychology introductory courses of the

University of La Laguna took part in this experiment to fulfill a course credit requirement. All of them were native Spanish speakers, right-handed and with no known motor or perceptive disorders.

Materials. Forty tri-syllable experimental stimuli including grapheme *C* were selected. For half of the stimuli this letter was pronounced /k/ (higher-probability G-P mapping), and for the other half it was pronounced /θ/ (lower-probability G-P mapping). All the words were matched across conditions by the position of the grapheme *C*, the identity of the letter preceding this grapheme, lexical frequency, frequency type and token of the critical syllable, frequency of the critical bigram, word length (number of letters and number of phonemes), and orthographic neighborhood. T-tests were conducted to make sure that these variables did not significantly differ across conditions (all $ts < 1$). We were not able to use words with the same letter after *C* since this would lead to *C* having the same pronunciation in both conditions. For this reason, we only controlled for the frequency of the resulting bigram and syllable. The full set of stimuli with their values in the controlled variables according to BuscaPalabras (Davis & Perea, 2005) are given in Appendix C. Forty additional words were selected to serve as fillers and 2 more were added for the practice phase.

Procedure and apparatus. These were identical to those described in Experiment 1a.

Results and discussion.

The same exclusion criteria as in Experiment 1a were applied (4.12%). Writing durations (LD) for the critical grapheme *C* and the previous and following inter-letter intervals (ILI0, ILI1) were submitted to separate analyses of variance (ANOVAs), with the G-P probability (higher-probability vs. lower-probability) as a within-subject variable in the analysis by participants (F_1) and a between-items variable in the analysis by items (F_2). Mean writing durations and standard deviations obtained in Experiment 2 are given in Table 3.

G-P probability	ILI0	LD	ILI1
High (/k/)	77	137	53
Low (/θ/)	80	141	62

Table 3. Writing durations for the critical letter *C* (LD) and durations of the previous and following inter-letter intervals (ILI0, ILI1) in milliseconds, in Experiment 2.

An effect of the G-P probability was observed in LD, $F_1(1, 23) = 5.94, p < .05, MSE = 172.52, \eta_p^2 = .20$; $F_2(1, 38) = 4.15472, p < .05, MSE = 140.62, \eta_p^2 = .10$. Letter *C* was produced faster when it was pronounced in the most frequent way (i.e., /k/). In ILI0, this effect was also significant in the analysis by participants, $F_1(1, 23) = 4.86, p < .05, MSE = 105.02, \eta_p^2 = .17$, but not in the analysis by items, ($F_2 < 1$). More importantly, the effect in the G-P probability was also significant in ILI1, $F_1(1, 23) = 18.22, p < .001, MSE = 990.08, \eta_p^2 = .44$; $F_2(1, 38) = 24.21, p < .001, MSE = 864.9, \eta_p^2 = .39$. In both ILI0 and ILI1 longer inter-letter intervals were observed in the lower-frequency condition.

Experiment 2 aimed to determine whether the effects obtained in ILI0, LD and ILI1 in Experiment 1a were truly due to the probability of the G-P mappings, or whether they were better accounted for by the position of the syllabic boundary. The results revealed significant effects of G-P probability in the three measures. The duration of the interval previous to the production of the letter *C* was shorter when this letter represented its most frequently associated phoneme and the execution of this target letter was faster as well. The following interval was also shorter. We consider that this pattern of results confirms that the effects obtained in Experiment 1a cannot be attributed to the position of the syllabic boundary. Moreover, the effects observed indicate that G-P probability affects the retrieval and execution of graphemes.

General discussion

In this study we explored whether adult writers follow a two-phase route for copying, in which graphemes are firstly converted to phonemes and then P-G conversion patterns are

applied, as proposed by Cueto's model of copying. To clarify this question, we tested the potential involvement of the correspondences between graphemes and phonemes during a copying task. Specifically, we aimed to establish whether the retrieval of a to-be-written grapheme is affected by the probability of the G-P mapping. In two direct copy transcoding experiments (Experiment 1 conducted in French and Experiment 2 in Spanish) we tested polyvalent graphemes in the context of two different pronunciations, one of them representing a more probable G-P mapping than the other. Taken together, the results suggested that the selection and execution of a grapheme during a copying task is affected by the probability of the G-P mapping. Although in Experiment 1b we failed to observe reliable effects when using the orthographic sequence *TI*, Experiment 1a yielded significant differences between a more frequent pronunciation of letter *E* (/ɛ/) and a less frequent pronunciation (/ø/). The inter-letter intervals produced before (ILI0) and after (ILI1) the target letter were shorter in the higher-probability condition and the critical letter (LD) itself was produced faster. However, since ILI1 was always inter-syllabic in the words in the lower frequency condition (but not in the case of words in the higher-probability condition), it could be argued that these results reflect a syllabic boundary effect. Inter-syllabic intervals have been repeatedly reported to be significantly longer than intra-syllabic intervals (Álvarez et al., 2009; Kandel et al., 2006), so the effect observed in ILI1 could indeed have been explained by this. In order to test this possibility, Experiment 2 was conducted with the Spanish polyvalent grapheme *C*. In this case, target ILIs did not differ in their syllabic status: they were always intra-syllabic in both experimental conditions. In this experiment, the effects of grapheme-to-phoneme probability observed in Experiment 1a were replicated, including that obtained in ILI1 durations. It seems that G-P probability affects handwriting duration when adults copy words. This finding is relevant for handwriting production theory for several reasons.

Firstly, this pattern of results confirms the general hypothesis that the sublexical route

mediates the word copying process in adults. Specifically, the effects of G-P probability suggest that the letters included in the visual input are first converted to phonemes, and that later these phonemes activate the graphemes that are maintained in the buffer for their subsequent production. This point is a novel idea introduced in the present work. Even though Cuetos (1991) proposed that such a route for copying should exist, until now it had not been claimed that this route might be used by experienced writers when they copy familiar words. It is generally accepted that such a route for copying would be exploited only by children and by some dysgraphic patients (because they lack strong lexical representations) or when non-lexical material is used. However, here we report evidence supporting the idea that this two-phase route is functional in the course of normal adult copying.

Although some handwriting models could easily be adapted to accommodate the effects reported here (Bonin et al., 2001; Tainturier & Rapp, 2001), it is important to notice that neither of these models has been proposed to account for the copying task, and that the route we have just described has only been assumed explicitly by Cuetos (1991). This author has pointed out the possibility of a phonologically mediated route for copying in which the individual graphemes of the visual input activate their corresponding phonemes by means of the application of the orthography-to-phonology correspondences. This is the only route proposed that predicts that differences in the level of activation of phonemes would affect the writing process. However, Cuetos has claimed that the use of phonologically mediated routes is less common than the use of the non-phonologically mediated route, in which the semantic system directly activates the appropriate orthographic word-form from the lexicon. Thus, evidence of the involvement of phonological/sublexical information during writing should be obtained especially in those cases in which the orthographic lexical representation is not available. However, we have obtained evidence of the application of a sublexical route in a

copying task involving known words and performed by skilled writers. The effects observed here reveal that, in normal writing conditions, sublexical information contributes to correctly retrieve and/or maintain the constitutive graphemes of an orthographic word-form.

Secondly, this pattern of results introduces for the first time the idea that grapheme-to-phoneme probability affects writing durations. That is, the G-P probability seems to partially determine the time needed to activate the corresponding graphemic representation and the stability of this representation. Since no effects were observed when the sequence *TI* was used (Experiment 1b), we think that G-P probability effects might be detectable only in the case of high-frequency graphemes (as *E* in Experiment 1a or *C* in Experiment 2). More evidence is necessary to establish what variables might influence this effect.

An alternative account of our data could be that these effects reflect the presence of a conflict which carries on to affect the movement durations for a few seconds, instead of differences in the accessibility of the graphemes. Delattre et al., (2006) proposed that effects obtained in writing durations might be due to the presence of a conflict which is not yet resolved when writing starts. From this point of view, a conflict might be generated by the sublexical route when faced with a lower-probability G-P mapping, producing longer writing durations in this condition than in the higher-probability mappings. However, the mechanism that would cause such a conflict is not immediately clear to us, since the sequences included in the lower-probability condition are not irregular.

It could also be argued that differences in writing durations between higher- and lower-probability G-P mappings might originate not at the grapheme level, but at a higher (previous) level of processing, in which interference would take place among the different G-P correspondences. From this point of view, more interference would be predicted between the alternative pronunciations of *TI* (Experiment 1b) than in the case of *E* (Experiment 1a), because the probabilities of the alternative G-P mappings are more similar in the former than

in the latter. However, we found the opposite result, with significant differences in the case of the higher- and lower-probability mappings of *E* but not in the case of *TI*.

In summary, although the effects reported here could be due to cascading of central processes to peripheral processes or to interference among competing G-P mappings, we propose that the effects observed in LD, ILI0 and ILI1 actually reveal that the probability of the G-P mappings affect the level of activation of the intervening phoneme, which in turn has an effect on the accessibility of the grapheme to-be-produced when P-G correspondences are applied.

It goes without saying that more evidence about the role played by G-P conversion patterns during handwriting needs to be collected. But, if confirmed in further studies, the effect of G-P probability must be taken into account by theoretical proposals. We suggest that phonological and orthographic sublexical units are interconnected and that the strength of these connections depends on the frequency of the forward and backward connections existing in the language, due to the application of P-G correspondences and G-P correspondences.

Finally, and from a methodological point of view, these results strongly support the claim made by Van Galen (1991) about the locus of central sublexical effects in handwriting. The probability of G-P mappings was observed to have an effect on the duration of the critical letter and the intervals preceding and following this letter. This pattern supports the co-occurrence of central sublexical and peripheral processing during handwriting.

To conclude, G-P probability effects were observed in French and Spanish. In spite of the fact that Spanish has a fairly transparent orthography and French has a considerably opaque orthography, phonemic representations seem to be involved in the writing production process in both languages. It has been suggested that the impact of the sublexical route is reduced in those languages with highly inconsistent P-O correspondences (Jiménez &

Muñetón-Ayala, 2002). Although it is beyond the scope of the present work to establish a detailed comparison across languages, the results of Experiment 1 and Experiment 2 suggest that the influence of sublexical information during spelling is more extended than generally thought. Further research must be carried out to elucidate the precise impact of sublexical and phonological information on the handwriting process, depending on the characteristics of each particular language and task.

References

- Afonso, O. & Álvarez, C. J. (2011). Phonological effects in handwriting production: evidence from the implicit priming paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 1474-83. doi:10.1037/a0024515
- Afonso, O. (2013). *Phonological processing during handwriting of isolated words* (Doctoral Dissertation). Universidad de La Laguna, Tenerife.
- Alarcos-Llorach, E. (2011). *Representaciones gráficas del lenguaje*. Madrid: Biblioteca Nueva.
- Álvarez, C. J., Cottrell, D., & Afonso, O. (2009). Syllabic effects in inter-letter intervals when handwriting single words in Spanish. *Applied Psycholinguistics*, 30, 205-223.
- Barry, C., & Seymour, P. H. K. (1988). Lexical priming and sound-to-spelling contingency effects in nonword spelling. *Quarterly Journal of Experimental Psychology*, 40, 5-40.
- Bonin, P., Fayol, M., & Gombert, J. E. (1997). Role of phonological and orthographic codes in picture naming and writing: An interference paradigm study. *Current Psychology of Cognition*, 16, 299-324.
- Bonin, P., Peereman, R., & Fayol, M. (2001). Do phonological codes constrain the selection of orthographic codes in written picture naming? *Journal of Memory and Language*, 45, 688-720.
- Bosse, M. L., Valdois, S., & Tainturier, M. J. (2003). Analogy without priming in early spelling development. *Reading and Writing*, 16, 543-572.
- Cuetos, F. (1991). *Psicología de la escritura: Diagnóstico y tratamiento de los trastornos de escritura* (8th ed.). Madrid: Escuela Española, S.A.
- Davies, R.A., & Weekes, B.S. (2005). Effects of feedforward and feedback consistency on reading and spelling in dyslexia. *Dyslexia*, 11, 233-252.
- Davis, C. J., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish.

- Behavior Research Methods*, 37, 665-671.
- Delattre, M., Bonin, P., & Barry, C. (2006). Written spelling to dictation: Do irregularity effects persist on writing durations? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 1330-1340.
- Fischer-Baum, S., & Rapp, B. (2012). Underlying cause(s) of letter perseveration errors. *Neuropsychologia*, 50, 305-318.
- Folk, J. R., & Rapp, B. (2004). Interaction of lexical and sublexical information in spelling: Evidence from nonword priming. *Applied Psycholinguistics*, 25, 565-585.
- Folk, J. R., Rapp, B., & Goldrick, M. (2002). Lexical/sublexical interaction in spelling: What's the point? *Cognitive Neuropsychology*, 19, 653-671.
- Goodman, R. A., & Caramazza, A. (1986). Aspects of the spelling process: Evidence from a case of acquired dysgraphia. *Language and Cognitive Processes*, 1, 263-296.
- Guinet, E., & Kandel, S. (2010). Ductus: A software package for the study of handwriting production. *Behavior Research Methods*, 42, 326-332.
- Jiménez, J. E., & Muñetón-Ayala, M. (2002). *Dificultades de aprendizaje en escritura. Aplicaciones de la psicolingüística y de las nuevas tecnologías*. Madrid: Trotta.
- Kandel, S., Álvarez, C. J., & Valleé, N. (2006). Syllables as processing units in a handwriting production. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 18-31.
- Kandel, S., Peereman, R., Grosjacques, G., & Fayol, M. (2011). For a psycholinguistic model of handwriting production: Testing the syllable-bigram controversy. *Journal of Experimental Psychology: Human Perception and Performance*, 37(4), 1310-1322.
- Kandel, S., & Spinelli, E. (2010). Processing complex graphemes in handwriting production. *Memory & Cognition*, 38, 762-770.
- Kandel, S., & Valdois, S. (2005). The effect of orthographic regularity on children's

- handwriting production. *Current Psychology Letters: Brain, Behavior and Cognition*, 17,3.
- Miceli, G. (1989). A cognitive model of spelling: Evidence from cognitively-impaired subjects. In P. G. Aaron and R. Malatesha Joshi (Eds.), *Reading and writing disorders in different orthographic systems*. Dordrecht: Kluwer Academic Publishers.
- New, B., Pallier, C., & Ferrand, L. (2005). *Lexique 3*. www.lequixe.org.
- Peereman, R., Content, A., & Bonin, P. (1998). Is perception a two-way street? The case of feedback consistency in visual word recognition. *Journal of Memory and Language*, 39, 151-174.
- Qu, Q., Damian, M. F., Zhang, Q., & Zhu, X. (2011). Phonology contributes to writing: Evidence from written word production in a nonalphabetic script. *Psychological Science*, 22, 1107-1112.
- Roux, S., McKeef, T. J., Grosjacques, G., Afonso, O., & Kandel, S. (2013). The interaction between central and peripheral processes in handwriting production. *Cognition*, 127, 235-241.
- Sanders, R. J., & Caramazza, A. (1990). Operation of the phoneme-to-grapheme conversion mechanism in a brain injured patient. *Reading and Writing*, 2, 61-82.
- Spinelli, E., Kandel, S., Guerassimovitch, H., Ferrand, L. (2012). Graphemic cohesion effect in reading and writing complex graphemes. *Language and Cognitive Processes*, 27, 770-795.
- Stone, G. O., Vanhoy, M., & Van Orden, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language*, 36, 337-359.
- Tainturier, M. J., Bosse, M. L., Valdois, S., & Rapp, B. (2000). *Lexical neighborhood effects in pseudoword spelling*. Forty-first Annual meeting of the Psychonomic Society, New

Orleans (USA), 16-19 November.

- Tainturier, M. J., & Rapp, B. (2001). The spelling process. In B. Rapp (Ed.), *The Handbook of Cognitive Neuropsychology: What Deficits Reveal about the Human Mind* (pp. 263–289). Philadelphia: Psychology Press.
- Van Galen, G. P. (1991). Handwriting: Issues for a psychomotor theory. *Human Movement Science, 10*, 165-191.
- Zhang, Q., & Damian, M. F. (2010). Impact of phonology on the generation of handwritten responses: Evidence from picture-word interference tasks. *Memory and Cognition, 38*, 519-528.
- Ziegler, J. C., Jacobs, A. M., & Stone, G. O. (1996). Statistical analysis of the bidirectional inconsistency of spelling and sound in French. *Behavior Research Methods, Instruments, & Computers, 28*, 504-515.
- Ziegler, J. C., Montant, M., & Jacobs, A. M. (1997). The feedback consistency effect in lexical decision and naming. *Journal of Memory and Language, 37*, 533-554.

Acknowledgments

Olivia Afonso was the recipient of a post-graduate Grant AP-2007-00065 from the Spanish Government. This research was partially supported by Grant PSI 2010-15184 from the Spanish Government. The authors would like to thank Margaret Gillon Dowens for her helpful comments on earlier versions of this article.

Appendix A

Experimental stimuli used in Experiment 1a.

Higher G-P probability	WF	B0F	B2FR	NL	NP	N	OUP	NS
Averti	2.16	949	1010	6	6	0	6	3
Berceau	12.43	233	1010	7	5	1	6	2
Berger	3.92	233	1010	6	5	7	6	2
Berline	2.09	233	1010	7	6	0	7	2
Berlue	1.08	233	1010	6	5	0	6	2
Cerque	0.07	2620	1010	6	4	4	4	1
Fièvre	38.58	169	23	6	5	2	4	1
Mercure	1.76	969	1010	7	6	0	7	2
Permis	4.53	1047	1010	6	5	3	6	2
Persan	1.08	1047	1010	6	5	0	5	2
Persil	2.36	1047	1010	6	5	0	5	2
Pervers	3.38	1047	1010	7	6	0	7	2
Pester	0.61	1047	613	6	5	8	6	2
Presto	1.22	1100	613	6	6	1	6	2
Segment	0.88	2399	88	7	5	1	7	2
Seigle	2.09	2399	274	6	4	1	5	1
Sergent	20.88	2399	1010	7	5	4	7	2
Sermon	3.85	2399	1010	6	5	0	6	2
Serpent	13.24	2399	1010	7	5	4	7	2
Serveur	5.27	2399	1010	7	6	1	6	2
Service	106.28	2399	1010	7	6	1	6	2
Servir	74.59	2399	1010	6	6	4	6	2
Ternir	1.49	662	1010	6	6	4	6	2
Trèfle	4.19	192	1	6	5	0	4	1
Verges	5.88	446	1010	6	5	4	6	2
Vermeil	0.88	446	1010	7	6	0	5	2
Verser	9.86	446	1010	6	5	5	6	2
Vertige	24.26	446	1010	7	6	2	7	2
Veston	15.27	446	613	6	5	1	5	2
Mean	12.56	1181.03	842.93	6.38	5.31	2	5.86	1.9
Lower G-P probability	WF	B0F	B2FR	NL	NP	N	OUP	NS
Bedaine	1.42	233	26	7	5	0	4	2
Bedeau	1.69	233	26	6	4	0	4	2
Belote	3.99	233	532	6	5	1	6	2
Besace	2.43	233	613	6	5	0	5	2
Besogne	10.74	233	613	7	5	2	7	2
Brebis	7.03	1100	13	6	5	0	4	2
Brevet	3.11	1100	351	6	5	0	6	2
Cerise	1.22	2620	1010	6	5	1	6	2
Crever	29.05	1100	351	6	5	2	6	2
Devenir	2.84	8323	351	7	6	1	5	2
Deviser	6.22	8323	351	6	5	3	6	2
Grenat	3.85	1100	2700	6	5	0	6	2
Menacer	5.34	969	2700	7	6	2	7	3
Mesurer	16.35	969	613	7	6	2	7	3
Pelade	0.41	1047	532	6	5	1	5	2
Pelage	5.74	1047	532	6	5	2	5	2
Pelote	4.19	1047	532	6	5	4	6	2
Peluche	5.34	1047	532	7	5	1	7	2
Pelure	2.03	1047	532	6	5	1	5	2
Penaud	3.04	1047	2700	6	4	1	5	2
Regain	3.18	1255	88	6	4	0	5	2
Repris	1.55	1255	190	6	5	3	6	2
Secouer	14.19	2399	272	7	5	2	7	2
Secret	70.81	2399	272	6	5	1	5	2
Semaine	111.81	2399	1143	7	5	0	7	2
Sevrage	1.01	2399	351	7	6	1	5	2
Tenable	0.68	662	2700	7	6	0	5	2
Tenant	2.3	662	2700	6	4	3	6	2
Velours	35.88	446	532	7	5	0	6	2
Mean	12.32	1618.17	822.69	6.38	5.03	1.17	5.65	2.07

Note. WF = word frequency; B0F = frequency of the previous bigram; B2F = frequency of the posterior bigram; NL = number of letters; NP = number of phonemes; N = orthographic neighborhood; OUP = orthographic uniqueness point; NS = number of syllables.

Appendix B

Experimental stimuli used in Experiment 1b.

Higher G-P probability	WF	B0F	B2FR	NL	NP	N	OUP	NS
Centime	2.36	1154	279	7	5	1	7	2
Comptine	1.22	98	928	8	5	0	6	2
Cultivé	2.64	63	318	7	7	2	7	3
Destiné	1.22	618	928	7	7	4	7	3
Émotif	0.61	218	162	6	6	0	6	2
Entier	56.69	1154	1231	6	4	2	5	3
Fertile	2.57	526	648	7	6	0	7	2
Hématite	0.07	607	849	8	6	2	7	2
Identité	22.64	1154	849	8	7	0	8	4
Intime	23.85	1154	279	6	4	3	6	2
Légitime	13.24	849	279	8	7	1	8	3
Litige	0.74	849	294	6	5	0	6	3
Maritime	3.58	849	279	8	7	0	6	2
Mitigé	0.2	849	294	6	6	0	6	2
Notice	1.96	218	298	6	5	1	5	4
Obstiné	3.24	618	928	7	7	2	7	2
Patine	2.84	697	928	6	5	8	6	3
Platine	3.18	697	928	7	6	2	7	2
Ratatiné	0.81	697	928	8	8	1	8	3
Routine	9.53	593	928	7	5	1	7	3
Satiné	0.2	697	928	6	6	4	6	3
Solstice	1.28	618	298	8	7	0	4	2
Ultime	21.49	63	279	6	5	1	6	2
Ventilé	0.14	1154	648	7	6	2	7	3
Vestige	1.62	618	294	7	6	1	6	2
Victime	28.45	290	279	7	6	0	7	2
Mean	7.35	657.77	587.73	6.92	5.92	1.46	6.46	2.54
Lower G-P probability	WF	B0F	B2FR	NL	NP	N	OUP	NS
Action	72.91	290	839	6	5	0	6	2
Ambition	19.32	849	839	8	6	0	8	3
Caution	2.23	593	839	7	5	0	7	2
Diction	1.49	290	839	7	6	2	7	2
Dotation	0.41	697	839	8	7	5	4	3
Édition	10.61	849	839	7	6	0	5	3
Fiction	4.32	290	839	7	6	3	7	2
Fixation	1.89	697	839	8	8	0	7	3
Initial	4.19	849	146	7	7	1	7	3
Initié	1.89	849	63	6	6	2	6	3
Initier	3.45	849	1231	7	6	1	6	3
Lotion	0.54	218	839	6	5	3	6	2
Martial	1.76	526	146	7	7	1	7	2
Martien	0.61	526	1231	7	6	0	7	2
Motion	0.68	218	839	6	5	3	5	2
Mutation	3.18	697	839	8	7	0	7	3
Nation	31.96	697	839	6	5	4	6	2
Notion	10.61	218	839	6	5	4	5	2
Nuptial	1.35	98	146	7	7	0	7	2
Option	1.08	98	839	6	5	0	6	2
Ponction	0.74	290	839	8	6	2	8	2
Potion	2.3	218	839	6	5	3	5	2
Ration	5.95	697	839	6	5	2	6	2
Relation	10.2	697	839	8	7	0	8	3
Section	16.35	290	839	7	6	0	7	2
Taxation	0.07	697	839	8	8	0	5	3
Mean	8.08	511.04	759.35	6.92	6.04	1.38	6.35	2.38

Note. WF = word frequency; B0F = frequency of the previous bigram; B2F = frequency of the posterior bigram; NL = number of letters; NP = number of phonemes; N = orthographic neighborhood; OUP = orthographic uniqueness point; NS = number of syllables.

Appendix C

Experimental stimuli used in Experiment 2.

Higher G-P probability	WF	BF	NL	N	SF_TP	SF_TK
Aducir	2.5	317.14	6	0	49	227.32
Asceta	0.18	239.46	6	0	130	825.54
Boceto	1.07	239.46	6	1	130	825.54
Decente	7.68	139.29	7	3	30	181.96
Descenso	20.54	205.18	8	1	30	181.96
Docena	25	239.46	6	2	130	825.54
Docente	2.5	139.29	7	7	30	181.96
Escena	68.04	239.46	6	1	130	825.54
Faceta	3.39	239.46	6	2	130	825.54
Incienso	4.82	803.39	8	0	19	396.79
Mecedor	0.36	139.29	7	0	130	825.54
Obsceno	3.21	325.71	7	1	130	825.54
Pecera	0.71	239.46	6	0	130	825.54
Perecer	1.43	327.68	7	2	23	251.61
Piscina	10.71	792.86	7	0	170	1724.46
Precintar	0.18	260.54	9	0	10	39.82
Procesar	1.61	205.18	8	2	130	825.54
Recinto	16.25	218.93	7	0	10	39.82
Recitar	4.11	218.93	7	2	170	1724.46
Vicioso	5.71	218.93	7	1	25	246.96
Mean	9	287.46	6.9	0.95	86.8	631.35
Lower G-P probability	WF	BF	NL	N	SF_TP	SF_TK
Bocata	0.54	261.25	6	3	200	981.61
Bocazas	0.18	341.25	7	0	200	981.61
Decorar	0.71	142.5	7	2	94	283.93
Descanso	32.86	209.11	8	1	23	219.46
Disecar	0.54	160.89	7	2	92	350.18
Educар	3.57	269.46	6	0	49	227.32
Encuesta	15.89	349.64	8	0	7	38.57
Escama	0.36	261.25	6	3	200	981.61
Escudo	8.93	158.04	6	1	92	1089.82
Fecundo	1.61	163.75	7	2	8	10.54
Locura	25.18	158.04	6	0	92	1089.82
Pecador	1.79	341.25	7	3	200	981.61
Pescado	18.93	338.39	7	1	200	981.61
Picante	4.64	341.25	7	0	23	219.46
Precursor	4.64	279.46	9	0	11	145.89
Procurar	6.79	284.64	8	0	92	1089.82
Recado	6.96	261.25	6	5	200	981.61
Recorte	4.64	142.5	7	1	20	108.75
Rescate	16.25	338.39	7	0	200	981.61
Vacuna	16.96	158.04	6	0	92	1089.82
Mean	8.6	248.02	6.9	1.2	104.75	641.73

Note. WF = word frequency; BF = frequency of the previous bigram; NL = number of letters; N = orthographic neighborhood; SF_TP = syllable frequency type; SF_TK = syllable frequency token .