This article has been published in its final form in Cognition 156, 2016, 16-29, doi: 10.1016/j.cognition.2016.07.006, © Elsevier.

# Typing pictures: Linguistic processing cascades into finger movements

Michele Scaltritti[a,b], Barbara Arfè[a], Mark, Torrance[c], & Francesca Peressotti[a]

[a] Dipartimento di Psicologia dello Sviluppo e della Socializzazione, Università degli Studi di Padova. Address: Dipartimento di Psicologia dello Sviluppo e della Socializzazione, Università degli Studi di Padova ,Via Venezia 8, 35131 - Padova, Italy. email: barbara.arfe@unipd.it; francesca.peressotti@unipd.it

[b] Aix Marseille Univ, CNRS, LPC, Marseille, France. Address: LPC-UMR 7290, Aix Marseille Univ, Case D, 3 Place Victor Hugo, 13100 – Marseille, France. email: michele.SCALTRITTI@univ-amu.fr

[c] Division of Psychology, Nottingham Trent University. Address: Division of Psychology, Nottingham Trent University, Burton Street, Nottingham NG1 4BU, UK. email: mark.torrance@ntu.ac.uk

## Author Note

Correspondence concerning this article should be addressed to: Michele Scaltritti, LPC-UMR 7290, Aix Marseille Univ, Case D, 3 Place Victor Hugo, 13100 – Marseille, France. Phone: +33 4 13550970. E mail: michele.SCALTRITTI@univ-amu.fr

**Highlights**

1. We investigated response onset and response execution in typing

2. Lexical-semantic variables influenced response onset and interkeystroke intervals

3. Orthographic variables were found to affect only response execution

4. Results seems coherent with cascaded flow of information between linguistic and motor processes

**Abstract**

The present study investigated the effect of psycholinguistic variables on measures of response latency and mean interkeystroke interval in a typewritten picture naming task, with the aim to outline the functional organization of the stages of cognitive processing and response execution associated with typewritten word production. Onset latencies were modulated by lexical and semantic variables traditionally linked to lexical retrieval, such as word frequency, age of acquisition, and naming agreement. Orthographic variables, both at the lexical and sublexical level, appear to influence just within-word interkeystroke interval, suggesting that orthographic information may play a relevant role in controlling actual response execution. Lexical-semantic variables also influenced speed of execution. This points towards cascaded flow of activation between stages of lexical access and response execution.

*Keywords:* word production; typewriting; written picture naming; cascaded activation

## 1. Introduction

Language production involves transforming a communicative intention into a physical output, be it a spoken word, a written word, or a sign. Both cognitive and motor processes are necessary to accomplish this task. Curiously, the investigation of the cognitive and the motor sides of language production has proceeded along relatively independent paths, as pointed out in the fields of both spoken (Hickok, 2014) and written (e.g., Kandel & Perret, 2015; Weingarten, Nottbusch, & Will, 2004) word production. As a result, the issue of how information flows from central cognitive processes to motor execution in word production has, to date, received relatively little attention.

During the initial cognitive levels of word-form retrieval is processing encapsulated or does it percolate into actual response execution? This question calls into play the traditional distinction between serial and cascaded models of language processing (e.g., Damian, 2003; Kello, Plaut, & MacWhinney, 2000). In serial models, information needs to be fully processed at a given stage before it can provide an input for the next process. In this scenario, variables affecting information processing within central linguistic stages of word retrieval (semantic, lexical) should play no role during subsequent output (response execution) stages. By contrast, with cascaded activation as soon as information becomes available at given level, it is immediately forwarded for processing to the next (downstream) level. Under this scenario variables affecting central linguistic processing – processing prior to output initiation – can also affect processing at the output stage.

With this in mind, our paper explores which of these two models better describes the functional relationship between language processing and response execution in typewriting. With this aim, we assessed to what extent do semantic, lexical and sublexical variables affect both response retrieval and response execution. Specifically, we studied participants providing typewritten names to pictures of everyday objects, measuring both response latency and rate of production once the response had been initiated (i.e. mean inter-keystroke interval). These data permit isolation of effects prior to and concurrent with language output. Response latency is the time elapsing from the onset of the to-be-named stimulus until the first keystroke of the response. It is considered a measure of the processing occurring before the response stage, and thus it is mostly linked with linguistic central stages of word retrieval. Mean interkeystroke interval is the average of the time intervals between the keystrokes of the response. This is typically considered more related to peripheral stages of response execution (e.g., Logan & Crump, 2011). The need to go beyond just measuring response onset time has been frequently noted (e.g., Abrams & Balota, 1991; Balota & Abrams, 1995; see also Bangert, Abrams, & Balota, 2012; Spivey & Dale, 2006; Spivey, Grosjean, & Knoblich, 2005). Typewriting makes possible precise measures of execution after response onset in the context of a genuine language production task.

The effects of semantic and lexical variables such as familiarity, age of acquisition and word frequency, which are typically considered to affect lexical retrieval at a central level, have been repeatedly reported in response times in both picture naming and writing tasks (e.g., Almeida, Knobel, Finkbeiner, & Caramazza, 2007; Barry, Hirsh, Johnston, & Williams, 2001; Belke, Brysbaert, Meyer, & Ghyselinck, 2005; Bonin, Roux, Barry, & Canell, 2012; Catling & Johnston, 2009; Caramazza, 1997; Cycowicz, Friedman, Rothstein, & Snoodgrass, 1997; Levelt, Roelofs, & Meyer, 1999; Navarrete, Scaltritti, Mulatti, & Peressotti, 2013; Peressotti, Nicoletti, Rumiati, Job, 1995; Roux & Bonin, 2012). Once word forms have been retrieved, sublexical representations then also become available and can affect the speaker or writer's performance (e.g., Gentner, Larochélle, & Grudin, 1988). However, the extent to which lexical and sublexical effects are present in post-onset measures of response execution is still unclear. In spoken production different approaches have been used in order to investigate this issue. Speech errors contain the articulatory features of the unproduced but intended target sound (e.g., Goldrick & Blumstein, 2006) and studies consistently report that the output duration of words within a sentence depends on the extent to which words can be predicted (e.g., Griffin & Bock, 1988; Gahl & Garnsey, 2004; Tily, Gahl, Arnin, Snider, Kohtari, & Bresnan, 2006). These findings support the cascaded information flow hypothesis.

However studies exploring the effect of lexical variables in the production of sounds report mixed results. In line with cascaded flow of information, studies of spontaneous speech corpora have found frequency effects on articulation durations (Gahl, 2008; Pluymaekers, Ernestus, & Baayen, 2005). In single word production, a longer duration of the initial phoneme has been reported for words entailing irregular vowel pronunciation, compared to words with a regular vowel pronunciation (Kawamoto, Kello, Jones, & Bame, 1998). Also, lexical frequency seems to affect initial phoneme durations, but not rhyme durations (Kawamoto, Kello, Higareda, & Vu, 1999; see also Mousikou & Rastle, 2015). Lexical frequency effects have actually been detected in a reading aloud task with delayed responses (Balota & Chumbley, 1985), but this result has been debated (Monsell, Doyle, Haggard, 1989; Balota & Chumbley, 1990). Additionally, the effect of neighborhood size seems to be different in single word production and in spontaneous speech. Studies with single words report more expanded vowel spaces for words from denser neighborhoods (Munson & Solomon, 2004; Yiu & Watson, 2015). Studies with spontaneous speech show that words from dense neighborhoods are phonetically reduced (Gahl, Yao & Johnson, 2012). Finally, paradigms based on semantic congruency effects produced very mixed evidence. In the seminal work by Balota, Boland and Shields (1989), semantic priming effects were shown to influence both onset and duration times. Damian (2003), however, using both a picture-word interference and a blocked naming paradigm, found semantic interference and phonological facilitation effects in response onset times, but not in response durations. Further, he showed that spoken response durations were unaffected by Stroop interference (but see Kello, et al., 2000), suggesting that interference effects do not cascade into articulatory processes.

For handwriting, there is evidence that spelling processes affect motor execution. For example, motor production is slower for words with irregular spelling-to-sound mapping compared to regular words (Alfonso, Alvarez, & Kandel, 2015; Delattre, Bonin, & Barry, 2006; Kandel & Perret, 2015; Roux, McKeeff, Grosjacques, Afonso, & Kandel, 2013; but see Damian & Stadthagen-Gonzalez, 2009). Studies on handwriting have also shown that movement duration could be influenced by the phonological structure of the word. Kandel, Peerman, Grosjacques, & Fayol, (2011) investigated the role of syllabic structure in a copy task and showed that both syllable structure and bigram frequency affected mean stroke duration.

Finally, and particularly important for the present study, Logan and Zbrodoff (1998; see also Damian & Freeman, 2008) found a Stroop effect on onset latencies of typed response but not on interkeystroke intervals for typewriting, suggesting that response execution starts only once the target response, or the word form, has been selected (see also Logan & Crump, 2011). Other findings, however, seem more consistent with cascaded flow of information, showing that lexical variables exert an influence on both onset latency and interkeystroke intervals. For example, lexicality (typing a word versus typing a nonword) and word frequency have been found to affect the duration of interkeystroke intervals (e.g., Gentner, Larochélle, & Grudin, 1988; but see Baus, Strijkers, & Costa, 2013; Pinet, Ziegler, & Alario, 2016). Sublexical factors in particular, such as bigram frequency or syllabic structure, have also been found to affect interkeystroke intervals. Gentner and colleagues (1988) found that bigram frequency (i.e. the frequency with which specific letter pairs occur within written language, in this case Dutch and English) predicted interkeystroke intervals so that for higher frequency bigrams shorter intervals were found. Further, it has been demonstrated that the interval between two keystrokes is affected by syllabic boundaries (e.g., Gentner et al., 1988; Pinet et al., 2016; Weingarten et al., 2004): Interkeystroke intervals are longer when the same two letters belong to different syllables, compared to when they are part of the same syllable. These data suggest that sublexical representations are important during response execution.

## 1.1 The Present Study

Studies investigating the influence of linguistic factors on typewriting have typically adopted a factorial approach: The effect of a critical variable (or several critical variables) on behavior is assessed by comparing two groups of items which are polarized at the extremes of this variable, while controlling for other word characteristics. For typing studies, the case for a factorial and controlled approach appears particularly cogent. When dealing with interkeystroke intervals, for example, purely peripheral factors such as biomechanical and physical constraints (e.g., specific movements, hands and finger constraints, keyboard layout), are likely to be a major determinant of the results, up to the point that a computational model based mainly on these factors (Rumelhart & Norman, 1982) correlated well with the performance of human typists (.66 in Rumelhart & Norman, 1982; .57 in Gentner et al., 1988). If any peripheral factors happen to co-vary with the linguistic variables of interest, this may lead to spurious interpretations of the results. In order to convincingly demonstrate an effect of bigram frequency, for example, Gentner and colleagues (1988) had to compare Dutch and English typists. The rationale is that some digraphs have different frequency across languages, yet they obviously entail the same movements irrespective of the typist's language. The reliable bigram frequency effect found across the two languages led the authors to conclude that the effect was genuinely linguistic. For the frequency effect, Gentner and colleagues compared sequences of 4 letters occurring within high and low frequency words (e.g., *yste* in *system* and in *oyster*). In fact, a given interkeystroke interval is influenced by the two keystrokes executed before, as well as by the two keystrokes occurring afterwards (e.g., Rumelhart & Norman, 1982). By comparing the interkeystroke interval between the second and the third letter (*s* and *t*) of the same four letter sequence (*yste*) occurring within high (*system*) vs. low frequency words (*oyster*), the authors demonstrated that the frequency effect they detected was not explained just by previous and subsequent movements.

Although these examples clearly demonstrate the strength of factorial studies, the approach is not viable if we want to jointly assess the effect of multiple linguistic variables. Even if one can find the appropriate sets of items (see Cutler, 1981) the question of whether this selection affects the generalizability of the results remains, given that many criteria would affect which items were chosen. Both loss of power, as a result of transforming continuous variables into categories, and experimenter bias in item selection have been identified as drawbacks of a factorial approach (e.g., Balota, Yap, Hutchison, & Cortese, 2013; Forster, 2000; Keeulers, Lacey, Rastle, Brysbaert, 2012;

Yap & Balota, 2009). More recent psycholinguistic research has therefore developed a complementary approach involving data from relatively large numbers of subjects tested on many items. In this large item pool, linguistic variables are distributed in a natural way rather than being artificially constrained for experimental purposes (see for a review Balota, Yap, Hutchison, Cortese, 2012). Following these arguments, we decided to investigate the performance of 75 participants, each of them responding to 260 colored pictures depicting common objects.

There are many linguistic variables that potentially account for response latency and response execution during typed picture naming. We selected sets of predictor variables that best indexed activity at semantic-lexical, orthographic and sublexical representational and processing levels. The first set of predictors – *naming agreement*, *age of acquisition*, *subjective familiarity*, and *written word frequency* – have been linked to conceptual processing, lexical access, and/or the links between these two representational stages. (Naming agreement: Barry, Morrison, & Ellis, 1997; Cycowicz et al., 1997; Kan & Thompson-Schill, 2004. Age of acquisition: Belke et al., 2005; Brysbaert & Ghyselinck, 2006; Catling & Johnston, 2006; 2009; Johnston & Barry, 2005. Subjective familiarity: Alario, Ferrand, Laganaro, New, Frauenfelder, & Seguí, 2004; Bates et al., 2003; Hirsh & Funnell, 1995. Word frequency: Almeida et al., 2007; Baus et al., 2014; Bonin et al., 2012; Caramazza, Costa, Miozzo, & Bi, 2001; Delattre et al., 2006; Kandel & Perret, 2015). By studying these variables we aimed to track processing dynamics related to onset latency and mean interkeystroke intervals.

The second set of predictors - *orthographic neighborhood density*, *orthographic Levenshtein distance*, *mean frequency of occurrence of the orthographic neighbors* and *number of letters* - were selected to reflect aspects specifically related to orthographic representations. They therefore tap processes at the interface between pure (amodal) lexical representations, and modality specific lexical-orthographic representations. The effects of these variables on initial response latency and execution therefore provide an insight into whether orthographic word-forms are fully processed prior to response execution, or if their role extends beyond typing onset (for similar reasoning in handwritten production see Kandel & Perret, 2015). There is debate concerning the extent to which orthographic retrieval is phonologically mediated in writing (e.g., Bonin et al., 2012; Rapp, Benzing, & Caramazza, 1997; Miceli & Capasso, 1997; Roux & Bonin, 2012). Thus, we decided to include also a phonological variable (*phonological neighborhood density*) in this set of predictors.

Finally, we examined the effects of *mean bigram frequency*, a variable that can be assumed to reflect sublexical processes and which has already been shown to affect interkeystroke interval (e.g., Gentner et al., 1988). We expected to replicate this observation in the present analyses. We also examined whether this variable affects onset latency. Effects on both interkeystroke interval and onset latency would support the notion that fine-grained sublexical aspects are processed even before response onset.

Statistical analyses proceeded through two steps. Given the high level of correlation among some of the variables considered as predictors, the first step was to restrict the number of predictors using a principled approach. To this end, we used random forest analyses, an analytical tool that provides a measure of variable importance for each predictor, using a data-driven approach based on a collection of classification trees (e.g., Breiman, 2001; for applications in psycholinguistics, see Tagliamonte & Baayen, 2012; Sadat, Martin, Costa, & Alario, 2014). In a second step we used mixed effects regression models to estimate the effect of the selected predictors on the dependent variables (response latency and mean interkeystroke intervals). The use of mixed effect models has a number of advantages over traditional regression techniques (for discussion, see Baayen, 2008; Baayen, Davidson, & Bates, 2008). Importantly, mixed models enabled us to assess linguistic effects while taking into account by-participants variations, including inter-individual differences in terms of proficiency, and item-specific variations, including those determined by item-specific peripheral factors (e.g., specific movements entailed during typing of a given word, distance between the different letters on the keyboard, and others).

# 2. Experiment

## 2.1 Method

**2.1.1 Participants.** Eighty-six students from the University of Padova without motor or perceptual disabilities agreed to participate. Of these, 11 students were excluded from the analyses, because they were not monolingual Italian native speakers, because they had a history of specific learning disabilities, or because they had speech or language disorders. The final sample consisted of 75 students (57 females, mean age = 24.03, *SD* = 3.37).

**2.1.2 Stimuli and procedure.** Demographic data, participants' linguistic status (i.e. monolingual or bilingual) and information regarding reading/writing difficulties were collected by questionnaire before the beginning of the experiment. This questionnaire also asked participants to evaluate their own typing skill, by reporting the number of fingers used to type. This measure (number of fingers used during typewriting, henceforth named Fingers) was used in the analyses as a proxy of typewriting expertise. More precisely, the original reports were recoded into a 7 point scale in which 1 = index finger of one hand, 2 = 2 index and middle fingers of one hand, 3 = index fingers of both hands, 4 = index and middle fingers of both hands, 5 = all the fingers of one hand plus index or index and middle of the other hand, 6 = index, middle, and ring fingers of both hands, 7 = all the fingers of both hands.

The experimental stimuli consisted of the 260 pictures of the Snodgrass & Vanderwart (1980) picture set, in the colored version developed by Rossion and Pourtois (2004). The experiment was implemented within the SR Research Experiment Builder environment, with keypress times accurately captured by custom code described in Wengelin et al. (2009). The experiment took place in rooms suitable for hosting multiple participants. Each participant sat in front of a computer monitor. Pictures were displayed in random order. Participants saw a fixation point just above the screen center, displayed for a random duration between 500 and 1000 *ms* followed by 200 *ms* blank screen. A picture (fitted within 281 x 281 pixels) was then displayed and participants typed its name. Speed and accuracy were equally emphasized in the instructions. Their typed response appeared immediately below the picture in 14-point Courier New. Participants were allowed to correct themselves and to delete mistakes using the backspace key. The picture remained visible until the participant had finished typing and pressed the Enter key, at which point the experiment progressed to the next trial. The experiment was divided in 4 blocks (65 trials in each, randomly allocated). At the end of each block, participants were prompted to take a brake and then to continue with the next block.

**2.1.3 Predictor variables.** Following the rationale outlined in the introduction and previous work in the field of word production in both speech (e.g., Alario et al., 2004; Barry et al., 1997; Bates et al., 2003; Sadat et al., 2014) and writing (Bonin, Chalard, Méot, & Fayol, 2002; Bonin, Méot, Lagarrigue, & Roux, 2015), we started our analysis with 11 potential predictor variables, encompassing lexical-semantic, lexical-orthographic/phonological and sublexical effects. Each variable is described below. Values for each variable were drawn from the PhonItalia Database 1.10 (Goslin, Galluzzi, & Romani, 2014) except where noted.

***2.1.3.1 Lexical-semantic level variables.***

*Naming agreement* (range: 20.99% - 100%, M = 75.68%, SD = 25.30%). For a specific response to a picture, naming agreement was the percentage of participants who gave that particular name to that picture, representing the proportion of participants using a specific word to respond to a given item. This measure was also used to determine the responses to include in the analyses (see response scoring section). Specifically, only responses given by at least 20% of the participants were retained.

*Subjective estimates of age of acquisition* (age of acquisition; range: 1 – 3.6, M = 2.12, SD = 0.59), obtained via a questionnaire. The words considered in the analyses (268) were divided into two lists (of 134 words). Two words (*barca* – boat and *maniglia* - doorknob) were excluded from this procedure by mistake and as such were not rated for age of acquisition (and neither for familiarity). Words in the two lists were comparable for frequency of occurrence, length, and number of orthographic neighbors. Each list was presented to 20 participants in the form of a questionnaire. Participants were asked to estimate the age at which they acquired each word, using a 5-points scale where 1 = from 0 to 3 years of age, 2 = from 3 to 6 years of age, 3 = from 6 to 9 years of age, 4 = from 9 to 12 years of age, and 5 = at 12 or more years of age. Participants could also indicate that they did not know the word.

*Subjective familiarity* (familiarity estimate; range: 3.5 – 5, M = 4.57, SD = 0.32), obtained in the same way as the estimates of age of acquisition. In the 5-points scale used to assess each word's familiarity, 1 = very unfamiliar, 2 = unfamiliar, 3 = relatively familiar, 4 = familiar, 5 = very familiar.

*Frequency of occurrence of the written word in the lexicon*, expressed in logarithm (word frequency; range: 0 – 7.91, M = 3.17, SD = 1.55).

***2.1.3.2 Orthographic/phonological word-form level variables***

*Orthographic neighborhood size* (orthographic neighborhood; range: 0 – 21, M = 4.28, SD = 4.23), defined as the number of orthographic word-forms that can be created by replacing a single letter within the original word (Coltheart, Davelaar, Jonasson, & Besner, 1977).

*Mean log-frequency of the orthographic neighbors* (orthographic neighborhood frequency, range: 0 – 5.44, M = 2.01, SD = 1.23), the mean log frequency-of-occurrence of the word's neighbors.

*Orthographic Levenshtein distance* (Levenshtein distance; range: 1 – 4.45, M = 1.99, SD = 0.64), reflecting an additional measure of the similarity of a specific word with respect to the other words in the lexicon. The Levenshtein distance is the number of insertions, deletions, or substitutions required to change from the current orthographic form to all the other unique forms. The values of Levenshtein distance considered here are the mean of the 20 smallest distances found (Goslin et al., 2014; see also Yarkoni, Balota, & Yap, 2008). This measure provides a subtle measure of orthographic similarities. In visual word recognition it accounts for variance in response times that is not explained by other measures of orthographic similarity (e.g., orthographic neighborhood, see Yarkoni et al., 2008).

*Phonological neighborhood size* (phonological neighborhood; range: 0 - 21, M = 4.02, SD = 3.85), defined as the number of phonological word-forms that can be created by replacing a single phoneme within the original word with another phoneme.

*Number of letters* (letter count; range: 2 - 12, M = 6.70, SD = 1.90)

*Number of syllables* (syllable count; range: 1 - 5, M = 2.82, SD = 0.78)

***2.1.3.3 Sublexical level variables.***

*Mean bigram frequency* (range: 15,135 – 201,607, M = 113,567, SD = 35,327). This is the mean frequency of occurrence of bigrams (letter pairs) in the response word.

**2.1.4 Dependent measures.** We recorded both initial response times and interkeystroke intervals, both timed in *ms*. Response times (RTs) were measured from the onset of the picture being displayed to the time of the first keypress in the participant's response. Interkeystroke intervals were the times between each consecutive keypress during response production. These were then averaged to give a measure of word execution rate (mean IKI). Finally, we analyzed accuracy, that is, the proportion of words named and spelt correctly. Self-corrections during word production were scored as errors and did not contribute to the accuracy score. Accuracy measures are not central to the questions addressed here and thus results are reported in the Appendix A.

**2.1.5 Response scoring.** We first removed trials where the response was incorrectly spelt (2.35% of trials), was missing (0.47%), or the participant edited the word (used cursor and delete keys) during production (9.96%). We then removed all trials for which the response was given by less than 20% of participants for that picture (i.e. naming agreement < 20%). This identified 293 admissible responses, with 32 pictures having more than one admissible response, and 1 picture having no admissible responses. We then removed responses for words that did not appear in the PhonItalia database (Goslin et al., 2014). These included multiple words (e.g., *pupazzo di neve* - snowman) and words borrowed from other languages (e.g., *clown*), and lead to the exclusion of an additional 25 (8.53%) responses. Finally, we removed trials with very long latencies >10000 ms (0.09% of the RTs). Overall 14,280 out of the original 17,945 trials were analyzed (79.58%).

For mixed-effects model analyses (see below) a reciprocal transform was applied to response latencies and mean interkeystroke intervals, as it has been shown to reduce skewness and to better approximate normality for the distribution of the models residuals (e.g., Kliegl, Masson, & Richter, 2010). We use -1000/RT and -1000/IKI to facilitate reading of the results.

**2.1.6 Statistical analysis.** First, we examined the correlations between the 11 predictors selected (Table 1). In cases in which two predictors were very strongly correlated (r > .8), we retained only the variables emerging as more important in accounting for the results using data-driven methods for assessing variable importance based on random forest analysis (Hothorn, Buehlmann, Dudoit, Molinaro, & Van Der Laan, 2006; Strobl, Boulesteix, Kneib, Augustin, & Zeileis, 2008; Strobl, Boulesteix, Zeileis, & Hothorn, 2007). This method has been indeed suggested as a black-box method to identify, amongst a larger set of variables, a smaller sample of potentially relevant predictors (Strobl, Malley, & Tutz, 2009), to be later tested with classic regression approach (Sadat et al., 2014). Random forest analysis provides a trial and error method for establishing whether a given variable is a useful predictor (Tagliamonte & Baayen, 2008). Random forests rely on the construction of multiple classification or regression trees, where predictors are recursively partitioned into subsets with increasingly homogeneous response values (Tagliamonte & Baayen, 2008; Strobl, Malley, & Tutz, 2009). It is important to note that each classification tree is based on a subset of randomly sampled data (in-bag observation). The tree's predictions can then tested against data that were not sampled (out-of-bag observation). To assess variable importance, the values of the predictors are randomly permuted, and the results of the permutation are again used to predict non-sampled (out-of-bag) observations. Clearly, if the original predictor was useful, the permuted version would be much less accurate in predicting response values. The difference in prediction accuracy before and after permutation, averaged across all trees, can thus be used as an index of variable importance (Strobl et al., 2009; Tagliamonte & Baayen, 2008; see also Breiman, 2001).

The effects of predictors selected via the random-forest procedure were then evaluated using mixed effect models. To build these models, we added predictors one at a time, assessing in each case whether the addition of the new variable determined an improvement in terms of explained variance. We first introduced variables with well-known and expected effects: semantic-lexical variables as predictors of response latencies and mean bigram frequency for mean interkeystroke intervals; we then determined the effects for the remaining predictors. For the random effect structure, we modeled random slopes for all the significant effects, including the interactions amongst them, as well as the interactions with by-participants random intercepts. In case the model failed to converge, we first removed interactions amongst random slopes, and then the interactions between random slopes and intercepts. If the model still failed to converge, we removed random slopes associated with the lesser amount of variance explained. All analyses were conducted in R (R Core Team, 2015) using packages lme4 (Bates, Maechler, Bolker, & Walker, 2015), party (Hothorn et al., 2006; Strobl et al., 2007; 2008), and rms (Harrel, 2016).

**2.2 Results**

**2.2.1 Typing Performance.** Participants mostly reported to use index and middle fingers of both hands in order to type (36 participants), followed by all fingers of both hands (29 participants). Few participants reported to use just the index fingers of both hands (6), the index and middle fingers of one hand (1), all the fingers of one hand plus index and middle fingers of the other hand (1), and the index, middle, and ring fingers of both hands (2). The mean response latency was 1354 ms (SD = 684 ms) and the mean interkeystroke interval was 204 ms (SD = 71). There was a significant correlation between Fingers and mean interkeystroke interval (Spearman's $r = -.42$, $p < .001$) suggesting that self-reported measures of proficiency were reasonably related to the actual performance.

**2.2.2 Correlations among predictors.** Correlations among pairs of predictors are reported in Table 1. We found strong correlations between Orthographic Neighborhood and Phonological

Neighborhood (.89), between Letter Count and Number of Syllables (.85), and between Levenshtein Distance and Letter Count exhibited a similarly strong correlation (.86).

----------------------

Table 1 about here

----------------------

#### 2.2.3 Onset RTs.

***2.2.3.1 Selecting predictor variables.*** As the name suggests, random forests involve different sources of randomness, such as random sampling of observations (in-bag observations used to grow the classification trees vs. out-of-bag observations used to assess predictions of the trees), random selection of subset of predictors, and random permutation of the predictors' values (for details and discussion, see Strobl et al., 2009; Tagliamonte & Baayen, 2012). To ensure that the results from our random forest analyses were stable, we decided to run each random forest analysis 4 times, and to assess the consistency of the results in terms of ranked variable importance (see below). In case of unstable results, following Strobl et al. (2009; see also Strobl et al., 2008) we modified 2 parameters of the analysis. First we increased the number of classification trees within the random forest (*ntree* parameter in the "party" R package, the default is 500), and, eventually, we increased also the number of random predictor variables sampled for each tree (*mtry* parameter in the "party" package, the default is 5), until stable solutions in terms of variable importance were met. Figure 1 reports the results in terms of variable importance obtained from 4 runs of the random forest analysis on RTs (with ntree = 1000 and mtry = 5). To interpret the results of the analysis we examined the rank order of the variables (Strobl et al., 2009; for a similar approach, Sadat et al., 2014). The results of the 4 separate runs (Figure 1) appear consistent with the only exception of Levenshtein distance and phonological neighborhood, which tended to switch ranks.

----------------------

Figure 1 about here

----------------------

On the basis of these analyses we make the following predictor choices: We retained orthographic neighborhood and discarded phonological neighborhood, as orthographic neighborhood was consistently ranked higher. We retained letter count and excluded syllable count for similar reasons. We excluded Levenshtein distance, as it was consistently ranked below letter count and orthographic neighborhood. Interestingly, we found that lexical-sematic variables – naming agreement, age of acquisition, familiarity estimate, and word frequency – which index lexical access (and upstream processes), were consistently ranked higher than other variables. This fits with the idea that conceptual processing and lexical access are indeed the lead-in processes for response onset.

***2.2.3.2 Linear mixed-effects models analysis.*** We began by fitting an intercept-only model (M0) with random intercepts for participants and items. We then proceeded by adding single predictors as fixed effects, incrementally (i.e. with a new model for each predictor). Each new model was tested against the last model that displayed a significant increase in terms of explained variance. If any predictor failed to produce a significant increase in explained variance, it was dropped from further analyses (i.e., from subsequent models). Model comparison was by likelihood ratio test. We explored the possibility of a non-linear relationship between word frequency and RTs, and between letter count and RTs, modeling these relationships with restricted cubic splines (Baayen, 2008; Baayen et al., 2006). Non-linear frequency and length effects have been found in previous research (Baayen, Feldman, & Schreuder, 2006; New, Ferrand, Pallier & Brysbaert, 2006; Yap & Balota, 2009). We tested restricted cubic splines with 3 to 7 knots, in order to evaluate different non-linear shapes.

Table 2 gives models comparisons for models that showed a significant increase in explained variance. The first predictors added were fingers (M1) and trial order. Trial order failed to determine an increase in explained variance ($\chi^2$ (1) = 0.48, p = .49), and was thus discarded from further analysis. We then considered lexical-semantic predictors, adding predictors in the following order: naming agreement (M2), word frequency (M3), then word frequency in non-linear terms (using restricted cubic splines with 3 to 7 knots), age of acquisition (M4), and familiarity estimates. All predictors gave an increase of explained variance, with the exception of word frequency in non-linear terms (all $\chi s^2 < 2.20$, all ps > .38), and familiarity estimates ($\chi^2$ (1) = 2.57, p = .11), which were therefore discarded. Next, we considered lexical-orthographic variables (orthographic neighborhood, orthographic neighborhood frequency, letter count, and non-linear letter count), and the sublexical predictor (bigram frequency). None of these explained additional variance ($\chi^2 < 3.43$, p > .17 in all cases).

The final effects model included random slopes for word frequency and age of acquisition, with no correlations (neither amongst slopes, nor between slopes and intercepts). Parameters of the final model are listed in Table 3. Correlations of fixed effects were weak (< .2), except for a moderate correlation between Age of Acquisition and Word Frequency (.44).[1]

----------------------

Table 2 about here

----------------------

As expected, major determinants of RTs are variables associated with semantic and lexical levels of word-retrieval processes. Reaction times were predicted by naming agreement, with pictures that elicited fewer names being named more quickly, and were faster for words with earlier age of acquisition and with higher frequency.

----------------------

Table 3 about here

----------------------

**2.2.4 Mean interkeystroke intervals.**

***2.2.4.1 Selecting predictor variables.*** Figure 2 reports variable importance measures obtained in 4 runs of random forest analysis on mean interkeystroke intervals (using default parameters).

----------------------

Figure 2 about here

----------------------

The variable with highest values of variable importance was mean bigram frequency. As for the highly correlated variables, the analysis showed higher values for orthographic neighborhood and letter count than for phonological neighborhood, syllable count and Levenshtein Distance. We thus decided to drop these latter variables from subsequent analyses.

***2.2.4.2 Linear mixed-effects models analysis.*** In building our mixed-effects models, we followed the same strategy as for RTs. Model comparison is summarized in Table 4. Order of entering predictors into the model was such that it assessed the effects of lexical-semantic variables after first partialling out sublexical and orthographic effects.

Self-reported typing skill (fingers; M1) and trial order (M2) both showed significant effects. We then considered the sublexical variable mean bigram frequency, which also improved model fit (M3). We then tested predictors relating to orthographic word-forms. We first added orthographic neighborhood, which improved fit (M4), and then orthographic neighborhood frequency and letter count (linear) neither of which reached significance ($\chi^2 < 1.35$, $p > .24$ in both cases). However the non-linear effect of letter count modeled using restricted cubic splines with 3 knots explained additional variance. Finally we considered, in order, naming agreement, word frequency, age of acquisition, and familiarity estimate. Naming agreement and word frequency significantly improved fit, while age of acquisition and familiarity did not (all $\chi^2 < 1.85$, ps > .17). Non-linear effects of frequency failed to reach significance (restricted cubic spline with 3 knots, $\chi^2 (1) = 3.10$, p = .08; 4 to 7 knots, $\chi^2 < 4.02$, p > .20 in all cases).

-----------------------

Table 4 about here

-----------------------

As for the random-effects structure, we retained the random slopes for orthographic neighborhood and word frequency. Parameters of the final model are listed in Table 5. Correlations among fixed effects were moderately low (all correlations were between 0 and .38), except between letter count and orthographic neighborhood (.67).[1]

-----------------------

Table 5 about here

-----------------------

Our results suggest that variables related to orthographic word-form and to finer grained sublexical aspects of orthography play an important role in determining mean interkeystroke latency. Mean interkeystroke intervals decrease with increasing bigram frequency (replicating Gentner et al., 1988). Mean interkeystroke intervals decrease with increasing number of letters, at least until a certain length (around 8 letters). This decrease seems to reach a plateau and to stop for longer words, as it appears from the representation in Figure 3.

-----------------------

Figure 3 about here

-----------------------

Words with an infrequent orthographic structure were produced more slowly, with longer interkeystroke intervals for words with few orthographic neighbors. Interkeystroke intervals were shorter for words with higher frequency and for words with higher naming agreement. Both of these lexical semantic effects were also found in the onset RT analysis.

## 3. Discussion

In the present article we assessed the influence of different psycholinguistic variables on two measures of picture typing performance, namely response latency and interkeystroke interval. Response latencies provide an indication of the time necessary to retrieve the word form. Effects on interkeystroke intervals suggest processing that was not complete at typing onset. Our goal was to investigate the extent to which semantic-lexical, orthographic and sublexical variables act differentially on response latency and interkeystroke intervals. Our analysis reveals two main findings: (i) As predicted, variables associated with processing up to and including lexical access (word frequency, age of acquisition, and name agreement) significantly affected onset RTs. However, notably, two of these three variables - word frequency and name agreement - also significantly affected mean interkeystroke intervals, indicating effects that persisted beyond typing onset. (ii) Variables related to the orthographic structure of the name given to the picture affected interkeystroke intervals. This was true not just for bigram frequency – a sublexical effect – but also for variables related to the orthographic word-forms, such as orthographic neighborhood. Importantly, these latter variables did not predict onset RTs. Below we outline what we consider to be the theoretical relevance of these findings.

### 3.1 The Effect of Lexical-Semantic Variables on Onset RTs and Interkeystroke Intervals.

The influence of lexical-semantic variables on onset RTs was consistent with prior research on spoken and written naming. Onset RTs, in typing as in other modalities, must necessarily capture (among other things) the time needed to retrieve the lexical entry that corresponds to the name of the stimulus picture. What is relevant here is that the effects of some of these variables appear to persist after participants had started to output their response. This result is in line with the idea that lexical access/retrieval and motor processes of the typing response may not represent two serially organized stages, rather two continuous stages through which activation flows in a cascaded fashion.

Support for this latter view comes from the influential model of movement organization proposed by Glover (2004), which distinguishes between planning and online control. The *planning system* selects a motor program that can achieve the actor's current goal, taking into account the specific environmental and the bio-mechanical constraints. Once a motor program has been selected, the planning system will determine when to initiate a movement. This stage is informed by cognitive processes and is therefore influenced by variables that determine the speed with which information is retrieved from memory. The *online control system* is devoted to minimizing the spatial error of the movement, quickly monitoring and, if necessary, adjusting motor programs on the fly. Relating this to our findings, response latency, which is largely determined by semantic-lexical factors, would mainly reflect the actions of the planning system. Interkeystroke intervals would reflect activity in the control system, which is chiefly concerned with how keypresses are activated in sequence, as well as with the finer aspects of the orthographic structure of the word to be typed. In explaining the timecourse of motor output, Glover (2004) argued that the two movement stages, planning and control, partially overlap. In fact the very early stages of the movement – the initial kinematic parameterization of the movement – might still be under the influence of the planning system. This potentially accounts for the effects of semantic / lexical variables on interkeystroke intervals. Our findings might indicate that during a typing task the two stages of action overlap. Planning is entirely responsible for movement onset, and continues to influence action as movement unfolds. [2]

Alternative interpretations of effects that we interpreted as markers of cascaded activation are possible. Word frequency could affect interkeystroke intervals not because lexical retrieval processes percolate into motor response execution, but because higher frequency words are also typed more often thus yielding a practice-driven facilitation at the level of response execution. This practice effect may involve the fine-grained motor transitions across keystrokes, an aspect which should also be captured by bigram frequency. In this respect, it is important to note that the frequency effect we detected was significant after the effect of bigram frequency was partialled out. However, evidence suggests that words represent chunking units for motor-response programming (Crump & Logan, 2010a), and that practice with specific words (e.g., "*dog*") may produce a facilitation effect over and above the practice of specific bigrams (*do* and *og*) and single letters (*d*, *o*, and *g*).

With respect to effects of naming agreement, which also persisted beyond typing onset, this variable may be related to different levels of processing (e.g., Barry et al., 1997; Vitkovitch & Tyrrel, 1995). For example, there might be difficulties related to the structural processing of the picture stimuli, which can be hard to identify (e.g., orange or lemon). Even when pictures are not conceptually unambiguous, they can activate alternative synonyms or near synonyms (e.g., couch and sofa) that would then compete for selection (Vitkovitch & Tyrrel, 1995; see also Barry et al., 1997). These processes occur before response selection. In a serial model, a participant would retrieve the name for a given picture, and then start the motor response. Any conflict about the correct response should be resolved before the beginning of the production processes. However, if co-activated representations interfere with response execution, by slowing down typing rate, this can be consistent with a cascaded account, in which activated but unselected representation can influence stages occurring after response selection due to the continuous nature of information flow within the system. Another possibility is that co-activation of several representations contributes to raise the level of uncertainty of the response to be given and this might affect not only response onset but also the way in which the response is actually executed (e.g. Tzagarakis, Ince, Leuthold, & Pellizzer, 2010).

Finally, in contrast with other lexical-semantic variables age of acquisition significantly affected onset RTs, but not mean interkeystroke intervals (but see additional analyses in Appendix B). It is possible that the limited range of values for age of acquisition (1-3.6) contributed to reduce the likelihood to obtain significant effects, especially in adult typists, who may be less affected by age of acquisition than younger typists.

### 3.2 The Selective Influence of Orthographic Variables on Interkeystroke Intervals

In our study is that the effects of variables related to measures of orthographic processing are detected solely on interkeystroke intervals. For bigram frequency, this result is not surprising and is in line with previous findings (Gentner et al., 1988; Pinet et al., 2016). The effect of orthographic neighborhood selectively detected on interkeystroke intervals and not in response latency is more interesting. As already pointed out, this result may suggest that some aspects of the word-level orthographic representation become relevant during response execution without affecting retrieval stages. Again, in terms of Glover (2004), this might suggest that the control system modulates the motor program required to execute the word triggering the retrieval of fine-grained orthographic representations, which discriminate between orthographically similar words. This finding is consistent with some recent evidence which showed orthographic regularity effects on letters duration in handwriting (Roux, et al., 2013), and with data showing that orthographic representations are continuously rehearsed during spelling processes (Colombo, Arfe, & Bronte, 2012).

This result, however, appears to be at odds with respect to other empirical evidence suggesting that all the keystrokes needed to type the word are retrieved in parallel and are available at the time of the first keystroke. For example, Crump and Logan (2010b) presented word or pseudoword primes before a probe letter. Touch typists were instructed to type the probed letter.

Reaction times were faster when the probe letter was part of the word prime. This priming effect was reliable even when the probe letter represented the medial or the final letter of the word-prime, suggesting that all keystrokes were activated in parallel upon prime presentation (for converging electrophysiological evidence see Logan, Miller, & Strayer, 2011).

Differently, our results suggest that orthographic representations are still influential while participants are actually typing the responses. We first explored the possibility that the differences between these two scenarios depended on the level of typing skills. Particularly, our participants' typing skill was not assessed with a pre-test, and we had no a priori measure of their typing performance. It is thus possible that our participants had a lower level of proficiency and that such difference is responsible for the apparent incongruence between results. Therefore, we conducted separate analyses for the fast and the slow typists of our sample, as identified by a median-split on their mean interkeystroke interval. For both groups, results replicated those displayed in final models for the whole sample. Importantly, for both fast and slow typists lexical-orthographic predictors failed to show any significant effects in terms of response latency (all $\chi^2$s $< 1.73$, all ps $>$ .18), and modulated just interkeystroke intervals.[3] Our data thus do not provide any evidence supporting the notion that orthographic predictors (or any other predictor) display differential effects as a function of typing proficiency. In addition, the idea that skill-level could enhance advanced processing of orthographic information has not been fully confirmed in the case of handwriting. Specifically, while skilled adults (Roux et al., 2013) and children older than 10 (Kandel & Perret, 2015) display spelling-to-sound regularity effects in motor response execution, less skilled writers (8 years old children) deal with irregularity mostly before response onset (Kandel & Perret, 2015). This pattern clearly contradicts the claim that the more proficient the participant, the more information is processed before response initiation. In sum, although we recognize that the issue of proficiency needs to be further addressed with specific research, our data offer no evidence of proficiency-modulated pattern.

Alternatively it may be that, consistent with suggestions from previous literature (e.g., Crump & Logan, 2010b; Logan, et al., 2011), orthographic word-forms are retrieved in advance of response initiation, but the lexical-orthographic predictors we considered may play little or no role in these retrieval processes. The effects of these variables may instead act during response execution on control system processes that come in to play after a motor program has already been selected. Glover (2004) suggested that during response execution the control system continuously monitors movement and occasionally adjusts it in order to prevent errors. It may be that in order to ensure that the movement is accurate the control system makes use of word-level orthographic representations, thus becoming sensitive to the familiarity of the orthographic structure of the word. Less frequent orthographic structures (i.e. words with few neighbors) that are more likely to be erroneously typed, need stronger monitoring and control, resulting in slower typing.

Finally, in line with the hypothesis of advanced response planning, we observed an effect of number of letters on mean interkeystroke intervals: The average interkeystroke interval decreases from shorter words to words with more letters, until reaching a plateau for longer words. This result is consistent with the idea that response execution in typewriting is subjected to advance planning, and more precisely, that response typing has been planned to occur within a fixed time. In this respect, the isochrony principle states that the velocity of a movement is proportionally linked to its linear extension so as to permit the execution time to be maintained approximately constant (Viviani & McCollum, 1983). It has been suggested that this principle links velocity to the amplitude of a movement plan. Reference to this type of temporal regularity in human motor behavior was first made in the literature more than a hundred years ago (Binet & Courtier, 1893) and it has been noted in a variety of well-rehearsed actions. Studies on writing movements, for instance, have shown that it takes the same time to write a letter or a word at different sizes, implying that there are proportional changes in velocity (Lacquaniti, Terzuolo & Viviani, 1983). This type of relationship between the linear extension of a movement and velocity appears to be a rather common feature pertaining not only to writing but also to typing (Viviani & Terzuolo, 1982), as found here. It is interesting to note that the adoption of the isochrony principle implies that during movement planning orthographic information regarding word length is fully specified. Remember that in order to execute a response within a fixed time, longer words need to be typed faster than shorter words. We interpret this result as consistent with the idea that orthographic word forms have to be already retrieved at the moment of response initiation.

**3.3 Conclusions**

To summarize, our results demonstrate for the first time that, during typewriting, activation at linguistic processing levels cascade into response execution. Further we found that both lexical and sublexical orthographic variables affect output rate after response initiation, but do not affect the time taken to initiate a response. This suggests that the influence of orthographic information persists once motor programs are initiated.

**Acknowledgements:** The data presented in this study were collected within a larger multilingual normative study: Torrance, M., et al (submitted). *Timed Written Naming in 14 European Languages*. Michele Scaltritti's work has been partially carried out within the Labex BLRI (ANR-11-LABX-0036), and has benefited from support from the French government, managed by the French National Agency for Research (ANR), under the project title Investments of the Future A*MIDEX (ANR-11-IDEX-0001-02). The National Science Foundation under Grant No. 1349042 has supported Francesca Peressotti's work.

**References**

Abrams, R.A., & Balota, D.A. (1991). Mental chronometry: Beyond reaction time. *Psychological Science, 2*, 153-157.

Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers, 36*, 140–155.

Afonso, O., Álvarez, C. J., & Kandel, S. (2014). Effects of grapheme-to-phoneme probability on writing durations. *Memory & cognition*, *43*(4), 579-592.

Almeida, J., Knobel, M., Finkbeiner, M., & Caramazza, A. (2007). The locus of the frequency effect in picture naming: When recognizing is not enough. *Psychonomic Bulletin & Review, 14*, 1177–1182.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Baayen, R. H., Feldman, L. F., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory & Language, 53*, 496–512.

Balota, D. A., & Abrams, R. A. (1995). Mental chronometry: Beyond onset latencies in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1289-1302.

Balota, D. A., Boland, J. E., & Shields, L. W. (1989). Priming in pronunciation: Beyond pattern recognition and onset latency. *Journal of Memory and Language, 28*, 14-36.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Access and/or production? *Journal of Memory and Language, 24*, 89-106.

Balota, D.A., & Chumbley, J.I. (1990). Where are the effects of frequency in visual word recognition tasks? Right where we said they were! Comment on S. Monsell, M.C. Doyle, and P.N. Haggard (1989). *Journal of Experimental Psychology: General, 119*, 231-237.

Balota, D. A., Yap, M. J., Hutchison, K. A., & Cortese, M. J. (2013). Megastudies: What do millions (or so) of trials tell us about lexical processing? In J. S. Adelman (Ed.), *Visual Word Recognition Volume 1: Models and methods, orthography and phonology* (pp. 90-115). New York, NY: Psychology Press.

Bangert, A.S., Abrams, R.A., & Balota, D.A. (2012). Reaching for words and nonwords: Interactive effects of word frequency and stimulus quality on the characteristics of reaching movements. *Psychonomic Bulletin & Review, 19*, 513-520.

Barry, C., Hirsch, K. W., Johnston, R. A., & Williams, C. L. (2001). Age of acquisition, word frequency, and the locus of repetition priming of picture naming. *Journal of Memory & Language, 44*, 350-375.

Barry, C., Morrison, C. M., & Ellis, A.W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *Quarterly Journal of Experimental Psychology, 50A*, 560-585.

Bates, E., D'Amico, S., Jacobsen, T., Szekely, A., Andonova, E., Devescovi, A., et al. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review, 10*, 344-380.

Baus, C., Strijkers, K., & Costa, A. (2013). When does word frequency influence written production? Frontiers in Psychology, 4. doi: 10.3389/ fpsyg.2013.00963

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*, 1-48.

Belke, E., Brysbaert, M., Meyer, A. S., & Ghyselinck, M. (2005). Age of acquisition effects in picture naming: Evidence for a lexical–semantic competition hypothesis. *Cognition, 96*, B45–B54.

Binet, A., & Courtier, J. (1893). Sur la vitesse des mouvements graphiques. *Revue Philosophique de la France et de l'Étranger*, *35*, 664-671.

Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2002). The determinants of spoken and written picture naming latencies. *British Journal of Psychology, 93*, 89-114.

Bonin, P., Méot, A., Lagarrigue, A, & Roux, S.. (2015).Written object naming, spelling to dictation, and immediate copying: Different tasks, different pathways? *Quarterly Journal of Experimental Psychology, 68*, 1268-1294.

Bonin, P., Roux, J.-S., Barry, C., & Canell, L. (2012). Evidence for a limited cascading account of written word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 1741–1758.

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.

Brysbaert, M., & Ghyselinck, M. (2006). The effect of age of acquisition: Partly frequency related, partly frequency independent. *Visual Cognition, 13*, 992–1011

Caramazza, A. (1997). How many levels of processing are there in lexical access? Cognitive *Neuropsychology, 14*, 177-208.

Caramazza, A., Costa, A., Miozzo, M., & Bi,Y.(2001).The specific word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1430–1450.

Catling, J. C., & Johnston, R. A. (2006). The effects of age of acquisition on an object classification task. *Visual Cognition, 13*, 968– 980.

Catling, J. C., & Johnston, R. A. (2009). The varying effects of age of acquisition. *Quarterly Journal of Experimental Psychology, 62*, 50–62.

Coles, M. G. H (1989). Modern mind-brain reading: Psychophysiology, physiology, and cognition. Psychophysiology, 26, 251-269.

Colombo, L., Arfé, B., & Bronte, T. (2012). The influence of pho- nological mechanisms in written spelling of profoundly deaf children. *Reading and Writing, 25*, 2021–2038.

Coltheart, M., Davelaar, E., Jonasson, J., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), Attention and performance VI (pp. 535–555). Hillsdale, NJ: Erlbaum.

Crump, M. J. C., & Logan, G. D. (2010a). Episodic contributions to sequential control: Learning from a typist's touch. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 662-672.

Crump, M. J. C., & Logan, G. D. (2010b). Hierarchical control and skilled typing: Evidence for word-level control over the execution of individual keystrokes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 1369-1380.

Cutler, A. (1981). Making up materials is a confounded nuisance: Or will we be able to run any psycholinguistic experiments in 1990? *Cognition, 10*, 65–70.

Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology, 65*, 171-237.

Damian, M. F. (2003). Articulatory duration in single-word speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 416-431.

Damian, M. F., & Freeman, N. H. (2008). Flexibility and inflexibility response components: A Stroop study with typewritten output. *Acta Psychologica, 128*, 91–101.

Damian, M. F., & Stadthagen-Gonzalez, H. (2009). Advance planning of form properties in the written production of single and multiple words. *Language and Cognitive Processes, 24*, 555–579.

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.

Forster, K. I. (2000). The potential for experimenter bias effects in word recognition experiments. *Memory & Cognition, 28*, 1109–1115.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language, 84*, 474–496.

Gahl, S., & Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: syntactic probabilities affect pronunciation variation. *Language* 80, 748–775.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language, 66*, 789-806.

Gentner, D.R., Larochelle,S., & Grudin,J.(1988). Lexical, sublexical, and peripheral effects in skilled typewriting. *Cognitive Psychology, 20*, 524–548.

Glover, S. (2004). Planning and control in action. *Behavioral and brain sciences*, *27*(01), 57-69.

Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: evidence from tongue twisters. *Language and Cognitive Processes, 21*, 649–683.

Goslin, J., Galluzzi, C., & Romani, C. (2014).  Phonitalia: a phonological lexicon for Italian. *Behavior Research Methods,46*, 872-886.

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production, *Journal of Memory and Language, 38*, 313–338.

Harrell, F. E., Jr. (2016). *rms: Regression Modeling Strategies (R package version 4.4-2)*. (retrieved from https://CRAN.R-project.org/package=rms)

Hickok G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience, 29*, 2-20.

Hirsh, K. W., & Funnell, E. (1995). Those old, familiar things: Age of acquisition, familiarity and lexical access in progressive aphasia. *Journal of Neurolinguistics, 9*, 23-32.

Hothorn, T., Buehlmann, P., Dudoit, S., Molinaro, A., & Van Der Laan, M. (2006). Survival ensembles. *Biostatistics, 7*, 355–373.

Johnston, R. A., & Barry, C. (2005). Age of acquisition effects in the semantic processing of pictures. *Memory & Cognition, 33*, 905– 912.

Kan, I. P., & Thompson-Schill, S. L. (2004a). Effect of name agreement on prefrontal activity during overt ad covert picture naming. *Cognitive, Affective and Behavioral Neuroscience, 4*, 43–57.

Kandel, S., & Perret, C. (2015). How does the interaction between spelling and motor processes build up during writing acquisition? *Cognition*, *136*, 325-336.

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.

Kawamoto, A. H., Kello, C. T., Higareda, I., & Vu, J. V. (1999). Parallel processing and initial phoneme criterion in naming words: Evidence from frequency effects on onset and rime duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 362-381.

Kawamoto, A. H., Kello, C. T., Jones, R., & Bame, K. (1998). Initial phoneme versus whole-word criterion to initiate pronunciation: Evidence based on response latency and initial phoneme duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 862-885.

Keeulers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British lexicon project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods, 44*, 287-304.

Kello, C. T., Plaut, D. C., & MacWhinney, B. (2000). The task-dependence of staged versus cascaded processing: An empirical and computational study of Stroop interference in speech production. *Journal of Experimental Psychology: General, 129*, 340–361.

Kliegl, R., Masson, M. E. J., & Richter, E. M. (2010). A linear mixed model analysis of masked repetition priming. *Visual Cognition, 18*, 655–681.

Lacquaniti, F., Terzuolo, C., & Viviani, P. (1983). The law relating the kinematic and figural aspects of drawing movements. *Acta psychologica*, *54*(1), 115-130).

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral & Brain Sciences, 22*, 1-75.

Logan, G. D., & Crump, M. J. (2011). Hierarchical control of cognitive processes: the case for skilled typewriting. *The Psychology of Learning and Motivation, 54*, 1-27.

Logan, G. D., & Zbrodoff, N. J. (1998). Stroop-type interference: Congruity effects in colour naming with typewritten responses. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 978–992.

Logan, G. D., Miller, A. E., & Strayer, D. L. (2011). Electrophysiological evidence for parallel response selection in skilled typists. *Psychological Science, 22*, 54–56.

Miceli, G., & Capasso, R. (1997). Semantic errors as neuropsychological evidence for the independence and the interaction of orthographic and phonological word forms. *Language and Cognitive Processes, 12*, 733-764.

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: where are they? *Journal of Experimental Psychology: General, 118*, 43-71.

Mousikou, P., & Rastle, K. (2015). Lexical frequency effects on articulation: a comparison of picture naming and reading aloud. *Frontiers in psychology*, *6*.

Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research, 47*, 1048-1058.

Navarrete, E., Scaltritti, M., Mulatti, C., & Peressotti, F. (2013). Age-of-acquisition effects in delayed picture-naming tasks. *Psychonomic Bulletin & Review, 20*, 148-153.

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Re-examining word length effects in visual word recognition: New evidence from the English lexicon project. *Psychonomic Bulletin & Review, 13*, 45–52.

Peressotti, F., Job, R., Rumiati, R., & Nicoletti, R. (1995). Levels of representation in word processing. Visual Cognition, 2, 421–450.

Pinet, S., Ziegler, J. C., & Alario, F.-X. (2016). Typing is writing: Linguistic properties modulate typing execution. *Psychonomic Bulletin & Review*. Advance Online Publication. doi: 10.3758/s13423-016-1044-3

Pluymaekers, M., Ernestus, M., & Baayen, R. (2006). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, *62*, 146-159.

R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rapp, B., Benzing, L., & Caramazza, A. (1997). The autonomy of lexical orthography. *Cognitive Neuropsychology, 14*, 71-104.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception, 33*, 217-236.

Roux, J.-S., & Bonin, P. (2012). Cascaded processing in written naming: Evidence from the picture–picture interference paradigm. *Language and Cognitive Processes, 27*, 734-769.

Roux, J.-S., McKeeff, T. J., Grosjacques, G., Afonso, O., & Kandel, S. (2013). The interaction between central and peripheral processes in handwriting production. *Cognition, 127*, 235–241.

Rumelhart, D. E., & Norman, D. A. (1982). Simulating a skilled typist: A study of skilled cognitive-motor performance. *Cognitive Science, 6*, 1–36.

Sadat, J., Martin, C. D., Costa, A., Alario, F. X. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology, 68*, 33-58.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory, 6*, 174-215.

Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. Current Directions in Psychological Science, 15, 207–211.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 10393-10398.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics, 9*(307).

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics, 8*(25).

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods, 14*, 323-348.

Tagliamonte, S., & Baayen, R. H. (2012). Models, forests and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change, 24*, 135–178.

Tily, H., Gahl, S., Arnon, I., Snider, N., Kothari, A., & Bresnan, J. (2009). Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition, 1*, 147–165.

Torrance, M., Nottbusch, G., Alves, R. A., Arfé, B., Chanquoy, L., Chukharev-Hudilainen, E., … Wengelin, Å. (2016). *Timed picture naming in 14 European languages*. Manuscript submitted for publication.

Tzagarakis, C., Ince, N. F., Leuthold, A. C., & Pellizzer, G. (2010). Beta-band activity during motor planning reflects response uncertainty. *The Journal of Neuroscience*, *30*, 11270-11277.

Vitkovitch, M., & Tyrrell, L. (1995). Sources of disagreement in object naming. *The Quarterly Journal of Experimental Psychology, 48*, 822–848.

Viviani, P., & McCollum, G. (1983). The relation between linear extent and velocity in drawing movements. *Neuroscience*, *10*(1), 211-218.

Viviani, P., & Terzuolo, C. (1982). Trajectory determines movement dynamics. *Neuroscience*, *7*(2), 431-437.

Weingarten, R., Nottbusch, G., & Will, U. (2004). Morphemes, syllables and graphemes in written word production. In T. Pechman & C. Habel (Eds.), *Multidisciplinary approaches to language production* (pp. 529-572). Berlin, NY: Mouton de Gruyter.

Wengelin, Å., Torrance, M., Holmqvist, K., Simpson, S., Galbraith, D., Johansson, V., & Johansson, R. (2009). Combined eyetracking and keystroke-logging methods for studying cognitive processes in text production. *Behavior Research Methods, 41*, 337-351.

Wrum, L. H., & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language, 72*, 37-48.

Yap, M.J., & Balota, D.A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language, 60*, 502-529.

Yarkoni, T., Balota, D. A., & Yap, M. J. (2008). Beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15*, 971–979.

Yiu, L.K. & Watson, D.G. (2015). When overlap leads to competition: Effects of phonological encoding on word duration. *Psychonomic Bulletin and Review,* 22, 1701-1708.

**Footnotes**

1. For both RTs and mean IKIs we performed analyses using residualized predictors in order to decrease correlations between fixed effects. In particular, for RTs we residualized age of acquisition against word frequency, and used residualized age of acquisition as a predictor in the model. For mean IKIs, we regressed letter count on orthographic neighborhood. The results, in terms of significance and direction of the effects, were the same as when using unresidualized predictors, with a reduction of the correlations between the fixed effects involved in the residualization.

2. Although we have no elements to speculate about the actual extension of the temporal overlap between planning and control stages, this reasoning raises the interesting possibility that cascaded effects may be more easily detectable in the initial phase of the response execution, i.e. in the first keystrokes. Indeed, this prediction has been already considered in typing (Damian & Freeman, 2008), and more thoroughly discussed and examined in spoken production (e.g., Kawamoto et al., 1998; 1999). As suggested by an anonymous reviewer, we performed additional analyses in which we separately considered the average interkeystroke intervals corresponding to the first and the second half of each response. The results obtained seem to support the idea that the lexical variables mainly affect the movements required to type first half of the response rather than the second half. This analysis is fully reported in Appendix B.

3. Our fast group had an average IKI of 171 ms (SD = 50), which closely resembles the values reported in studies focused on highly skilled typists (in particular, see Logan et al., 2011).