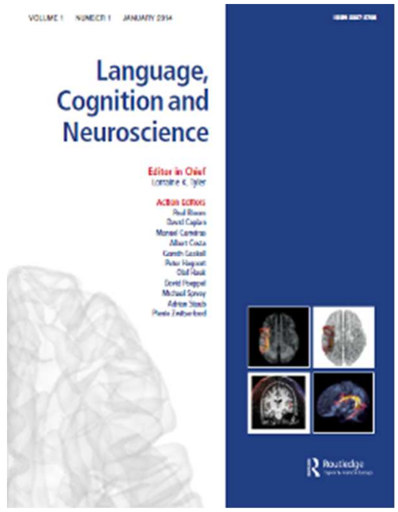

Lexical and sublexical effects on visual word recognition in Greek: Comparing human behavior to the Dual Route Cascaded (DRC) model

Journal:	*Language, Cognition and Neuroscience*
Manuscript ID	PLCP-2016-OP-10043.R3
Manuscript Type:	Original Paper
Date Submitted by the Author:	23-May-2017
Complete List of Authors:	Kapnoula, Efthymia; National and Kapodistrian University of Athens; University of Iowa, Department of Psychological & Brain Sciences; Basque Center on Cognition Brain and Language Protopapas, Athanassios; National and Kapodistrian University of Athens; University of Oslo, Department of Special Needs Education Saunders, Steven; Macquarie University, Macquarie Centre for Cognitive Science Coltheart, Max; Macquarie University, Macquarie Centre for Cognitive Science
Keywords:	visual word recognition, dual-route models, syllable frequency effect, Greek

Lexical and sublexical effects on visual word recognition in Greek: Comparing the Dual Route Cascaded (DRC) model to human behavior

Efthymia C. Kapnoula

National and Kapodistrian University of Athens,

University of Iowa,

and Basque Center on Cognition, Brain and Language

Athanassios Protopapas

National and Kapodistrian University of Athens and University of Oslo

Steven J. Saunders and Max Coltheart

Macquarie University

Author Note

Efthymia C. Kapnoula, Graduate Program in Basic and Applied Cognitive Science, National and Kapodistrian University of Athens, Greece; University of Iowa, Department of Psychological & Brain Sciences; Basque Center on Cognition, Brain and Language. Athanassios Protopapas, Department of Philosophy and History of Science, University of Athens, Greece; Department of Special Needs Education, University of Oslo, Norway. Steven J. Saunders and Max Coltheart, Macquarie Centre for Cognitive Science, Macquarie University, Sydney, Australia.

Correspondence concerning this article should be addressed to Efthymia Kapnoula, Basque Center on Cognition, Brain and Language, Mikeletegi Pasealekua, 69, 20009, Spain. E-mail: kapnoula@gmail.com

Abstract

We evaluated the Dual Route Cascaded (DRC) model of visual word recognition using Greek behavioral data on word and nonword naming and lexical decision, focusing on the effects of syllable and bigram frequency. DRC was modified to process polysyllabic Greek words and nonwords. The Greek DRC and native speakers of Greek were presented with the same sets of word and nonword stimuli, spanning a wide range on several psycholinguistic variables, and the sensitivity of the model to lexical and sublexical variables was compared to the effects of these factors on the behavioral data. DRC pronounced correctly all the stimuli and successfully simulated the effects of frequency in words, and of length and bigram frequency in nonwords. However, unlike native speakers of Greek, DRC failed to demonstrate sensitivity to word length and syllabic frequency. We discuss the significance of these findings in constraining models of visual word recognition.

Keywords: visual word recognition, dual-route models, syllable frequency effect, Greek

Lexical and sublexical effects on visual word recognition in Greek:

Comparing the Dual Route Cascaded (DRC) model to human behavior

According to the dual-route theory of word recognition and reading aloud, visually presented words are processed via two parallel routes; one lexical and one non-lexical (Forster & Chambers, 1973; Marshall & Newcombe, 1973). Within the lexical route, the orthographic information of each known word is directly linked to its phonological representation. In that way, the activation of a word's orthographic representation leads to simultaneous activation of all its phonemes. This means that the lexical route is potentially very fast. In contrast, non-lexical processing operates via the application of grapheme-to-phoneme conversion (GPC) rules, which leads to the activation of the word's phonemes in a serial (left-to-right) manner. Both words and nonwords are processed by both routes in parallel, but only known words are part of the lexical system. Therefore, words can be read via the (typically faster[1]) lexical route, whereas in the case of nonwords, the system relies primarily on the serial non-lexical route.

The Dual Route Cascaded (DRC) model (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) is one implementation of the dual-route theory and it is considered one of the most successful (Coltheart et al., 2001) and influential (Norris, 2013) models of visual world recognition. A crucial characteristic of the DRC is that GPC rules are hard-wired into the model (Coltheart, Curtis, Atkins, & Haller, 1993) and, as the name implies, they operate on representations of pre-defined size (i.e. graphemes, mapping on individual phonemes). However, hard-wired rules are not a theoretical commitment of the dual-route theory; in a different implementation of the same theory, the Connectionist Dual-Process (CDP) model (Perry, Ziegler, & Zorzi, 2007, 2010; Zorzi, Houghton, & Butterworth, 1998; Zorzi, 2010), non-lexical

[1] Since activation of the lexical route is delayed compared to the non-lexical one (see Figure 9 in Coltheart et al., 2001), this applies to words with more than two phonemes.

processing relies on mappings from orthographic to phonological elements that are learned via a connectionist network, instead of predefined GPC rules. The DRC does not explicitly aim to simulate the learning process via which conversion rules are acquired; and even when this is attempted (Pritchard, Coltheart, Marinus, & Castles, 2016), the grain size of sublexical units remains fixed at the grapheme-phoneme level. Therefore, it remains an open question whether hard-wiring the GPC rules themselves, or the size of the units they operate on, may deprive the model of the opportunity to learn certain systematicities intrinsic to the input that are not (and maybe cannot be) represented in rules at this level. That is, since the hard-wired rules operate on pre-defined representations, this may set intrinsic limitations to the model regarding the nature and size of the representations on which the non-lexical route operates (e.g. whether the basic unit of processing is the letter, the grapheme, or the syllable).

Evaluating the theoretical assumptions of DRC

One way to evaluate the theoretical assumptions of a model is by testing how successfully it can simulate human data in comparison to other models. Pritchard, Coltheart, Palethorpe, and Castles (2012) compared the two implementations of the dual-route theory in nonword reading. They reported both overall accuracy scores of how well the model's pronunciation matched the human data, as well as more detailed analyses of the types of errors performed by the humans and the models. The results showed that the overall performance of the DRC model was better than any of the CDP versions, even though the DRC did not account for the entire range of responses observed in the human data.

Such a qualitative comparison between human and model responses is one approach to evaluating computational models. An alternative approach would be to measure the effects of different factors and directly compare the pattern of results between behavioral and modeling data. Specifically, one can test whether a factor X that is found to have a significant effect on human responses also has a similar effect on the model's behavior (see discussion in Adelman & Brown, 2008). One way to do this is by comparing the proportion of variance accounted for by one variable in the human data versus the modeling data. Yap and Balota (2009) conducted an evaluation of specific factors affecting word recognition using a hierarchical regression model. The degree of agreement between the regression equations for the human and the modeling data allows the modeler not only to evaluate the sensitivity of the model to specific factors, but also to compare that sensitivity to the corresponding human data. Unfortunately, this approach is not guaranteed to produce interpretable results. For example, if predictor variables significantly co-vary then it is difficult to estimate the unique contribution of each predictor. Despite its limitations, this approach enables us to tease apart the effects of different variables and use them to evaluate models (e.g. see comparative evaluation of models in simulating stress assignment in Mousikou, Sadat, Lucas, & Rastle, 2017).

What can lexical and sublexical effects tell us about visual word recognition?

Yap and Balota (2009) confirmed and extended findings from other studies showing that word recognition/naming latencies depend on a variety of lexical and sublexical factors (see, e.g., Balota, Yap, Hutchison, & Cortese, 2012, for a review). Prominent among these stand word frequency (Forster & Chambers, 1973; Spieler & Balota, 2000), length (Barca, Burani, & Arduino, 2002; Spieler & Balota, 2000; Weekes, 1997), neighborhood density (Arduino & Burani, 2004; Barca et al., 2002; Reynolds & Besner, 2004; Spieler & Balota, 2000), and syllable frequency (Carreiras, Alvarez, & Devega, 1993; Conrad & Jacobs, 2004; Hawelka, Schuster, Gagl, & Hutzler, 2013). Therefore, it seems reasonable that a model of visual word recognition should be sensitive to these factors, similarly to human readers.

The effect of syllabic frequency has been of particular interest, partly due to the fact that it appears to be inhibitory or facilitatory depending on the specific task demands. When the task relies heavily on lexical access (e.g. lexical decision), words with more frequent initial syllables are recognized more slowly compared to words with less frequent initial syllables (see review in Conrad, Tamm, Carreiras, & Jacobs, 2010). This inhibitory effect is thought to reflect lexical interference from competitor words that have the same first syllable as the target word; when the first syllable of a word is high-frequency, this indicates that the target word shares its first syllable with many other words that are partially activated during visual word recognition and, thus, interfere with the recognition of the target word (Carreiras et al., 1993; Mahé, Bonnefond, & Doignon-Camus, 2014).

In contrast, facilitatory effects have been observed in naming (Carreiras & Perea, 2004; Perea & Carreiras, 1998) – a shift that has been attributed to the fact that in naming tasks participants need to construct phonological output for production; under these conditions, frequent syllables are activated and produced faster (Ferrand, Segui, & Grainger, 1996). Looking closer into this complex pattern of results, Mahé and colleagues (2014) argue that the net syllable frequency effect depends on the strength of syllable activation: strongly activated syllables (e.g. due to bigrams that are not congruent with word parsing into syllable units) lead to faster activation of phonological units, which in turn activate the target word along with its lexical competitors. Thus, the net effect is inhibitory in this case. In contrast, when syllable activation is weak/slow, lexical competitors are not as strongly activated, while the target word is still directly activated by letter units, leading to an overall facilitatory effect.

Several studies have examined more closely into the nature of syllable frequency effects and how they are linked to other sublexical effects, as a way of uncovering the mechanisms that underlie visual word recognition. Syllabic frequency effects, for example, could be understood in terms of orthographic processing. According to this account, syllabic frequency effects are in essence bigram frequency effects. That is, bigrams straddling a syllable boundary (i.e. the last letter of one syllable plus the first letter of the following syllable) are less frequent than bigrams contained within syllables. This interpretation highlights the theoretical distinction between bigram and syllable frequency effects. In fact, studies examining the effect of bigram frequency have yielded contradictory results (for a review see Chetail, 2015), which may be partly due to the fact that different studies control for different factors. For example, most studies of bigram frequency effects have not controlled syllable frequency, which is correlated with bigram frequency, therefore it remains an open question whether studies reporting significant bigram frequency effects may in fact have indirectly revealed effects best attributed to syllables.

Accordingly, the orthographic interpretation of syllabic frequency effects has been questioned by the results of studies that used words without this bigram difference (Carreiras et al., 1993) or that orthogonally manipulated syllable and bigram frequency (Conrad, Carreiras, Tamm, & Jacobs, 2009). Crucially, both studies showed that the effect of syllable frequency is independent from bigram frequency. Other studies have shown that the inhibitory syllable effect is related to token, rather than type, syllable frequency (Conrad, Carreiras, & Jacobs, 2008), and to phonological, rather than orthographic, syllable units (Conrad, Grainger, & Jacobs, 2007; Mahé et al., 2013). These findings are consistent with the proposal that syllabic frequency effects arise due to the automatic phonological syllabic parsing of the visually presented word, which in turn activates other words with the same initial syllable (Mahé et al., 2014; Perea & Carreiras, 1998).

According to this account, the frequency of only the initial syllable of each word is relevant. Thus, the definition of syllabic frequency in most studies is based on word-initial syllables. This approach makes it difficult to dissociate the effect of syllable frequency from what is broadly known in the spoken word recognition literature as *cohort effects* (Allopenna, Magnuson, & Tanenhaus, 1998; Connine, Blasko, & Titone, 1993; Marslen-Wilson, Moss, & Van Halen, 1996). Cohort effects are relevant in directional access processes, which necessarily operate in spoken word recognition because speech unfolds in time starting from the beginning of the word. As the initial part of a spoken word overlaps with other existing words, a potentially large set of lexical candidates can be activated. This set becomes progressively pruned as more of the spoken word is heard and mismatching candidates are dropped. In this conceptualization, the initial syllable is of crucial importance, because it constrains the starting candidate set. Similar processes may conceivably apply in visual word recognition to the extent that left-to-right processing is operative, as in the DRC sublexical route. Thus, if the syllable frequency effect reflects a disguised cohort-type process it should indeed be dependent on the frequency of the initial syllable only. In contrast, a genuine syllable frequency effect might arise if the visual word is parsed into syllables and all of them can affect access. To evaluate these alternatives, syllable frequency could be defined as the average syllabic frequency of all the syllables in the word, to be contrasted with first-syllable frequency.

More broadly, by extracting independent estimates of cohort, syllable, and bigram frequency effects, we can identify the nature and size of the representations that are most relevant for visual word recognition. This is particularly useful when it comes to evaluating whether it is necessary for a model to be able to learn what is the unit that is most relevant for processing, or whether it is just as good to hard-code this aspect of processing into the model.

Present study

The present study aimed at evaluating the DRC model and the corresponding theoretical assumptions on which it is based, by comparing DRC simulations to Greek word and nonword reading data from human readers. We were particularly interested in assessing the degree to which different sublexical units are relevant for reading and for this reason we focused on sublexical effects. The non-lexical route of the DRC architecture operates via grapheme-to-phoneme conversion, which assumes that the *graphemic* unit has a special status in visual word recognition. However, it is unclear whether this is the case in Greek word reading – an issue linked to the broader question of whether there are cross-linguistic differences in the cognitive mechanisms and representations underlying visual word recognition (Schmalz, Robidoux, Castles, Coltheart, & Marinus, 2017).

In particular, in this study, we (a) examined the degree to which different levels of processing, and corresponding units, are relevant in Greek visual word recognition, and then proceeded to (b) testing the DRC assumption regarding the special status of graphemes, by comparing it to the behavioral data. We estimated the effects of interest using the multiple regression method proposed by Yap and Balota (2009). Both naming and lexical decision data were collected from native Greek speakers, while simulation data were produced by a Greek version of the DRC model adapted to process multisyllabic Greek words and nonwords.

Stimuli spanned a wide range of frequency and sublexical characteristics, as well as a much greater range of syllables (2–5) than most previous studies. This was mainly done because polysyllabic words are more representative in Greek (Protopapas & Vlahou, 2009), but it also allows us to draw conclusions that are applicable to a wider range of stimuli. In addition, word and nonword stimuli in the two tasks were carefully matched on several of these variables (see

Protopapas & Kapnoula, 2013, 2016; Protopapas et al., 2016, Table 1). Crucially, to facilitate accurate estimation of individual effects, our stimulus set was designed to minimize inter-correlations among several major predictor variables (Protopapas & Kapnoula, 2013, 2016, Appendix B; Protopapas et al., 2016, Table 2). This allowed us to compare, among others, the predictive power of initial syllable frequency to that of average syllable frequency, taking into account all the syllables in each word.

Greek is a language with relatively transparent orthography, with consistency and regularity in the reading direction (i.e., from graphemes to phonemes) estimated at 95%, thus being more transparent than English (69.3%; Ziegler, Stone, & Jacobs, 1997), French, and German, but less transparent than Hungarian and Italian (Protopapas & Vlahou, 2009). Given the considerable difference between the Greek and English writing systems, the primary question addressed by this study was whether the DRC (which is built predominantly based on English behavioral data) would be able to exhibit a pattern of effects similar to that observed in the Greek behavioral data.

Method

Human data collection

The behavioral data analyzed below are the same as those reported in Protopapas and Kapnoula (2013, 2016), where the reader is directed for more details about the stimuli and aspects of the method, as well general patterns of findings. All analyses reported here are novel.

Participants. In total 132 native speakers of Greek (97 women) participated in this experiment. Ages ranged from 18 to 36 years old (M = 23.3, SD = 4.7). The majority were university students (12–21 years of education, M = 15.4, SD = 2.1). Fourteen were left-handed.

Stimuli. The experimental materials consisted of 300 visual stimuli (150 words and 150 nonwords) [2] 2–5 syllables long. The words were selected from the *ILSP PsychoLinguistic Resource* (IPLR) word list (Protopapas, Tzakosta, Chalamandaris, & Tsiakoulis, 2012). The nonwords were constructed to match the criteria set for the words (i.e. phonotactically and morphologically legal, 2–5 syllables long). To facilitate assessment of reading accuracy, all stimuli were free from orthographic ambiguities – meaning they had only one possible pronunciation.

For the purposes of this study, we focused on the following variables: word frequency (log number of appearances in the corpus), word length (number of letters), average orthographic bigram frequency (mean log number of appearances in word tokens), average phonological syllabic frequency (mean log number of appearances in word tokens), frequency of first orthographic bigram, frequency of first phonological syllable[3], orthographic neighborhood (defined as the number of words that have the same length as the target word and differ by only one letter, i.e., Coltheart's N; Coltheart, Davelaar, Jonasson, & Besner, 1977)[4], and grapho-phonemic transparency (computed as mean token probability of unique grapheme-phoneme mappings, termed "sonographs" following Spencer, 2009). Syllable and bigram frequencies were calculated independently of their position within the word, based on preliminary analyses indicating that position-dependent measures accounted for less variance in the behavioral data.

[2] A list of word and nonword stimuli is available in Protopapas and Kapnoula (2016; Appendix C).

[3] In our data, positional syllable frequency was not a better measure, and this is why we have not chosen to base our report on it; in our preliminary analyses, positional first syllable frequency (equivalent to a syllable-based cohort frequency) accounted for slightly less variance in naming and lexical decision RT than frequency of the first syllable measured cumulatively over all word positions.

[4] Given that this is the first investigation of its kind in Greek, we decided to use this metric of neighborhood density instead of the more recent OLD20 (Yarkoni, Balota, & Yap, 2008) in order to have data directly comparable to the majority of the literature.

These variables were collected using the IPLR (available at http://speech.ilsp.gr/iplr/, Protopapas et al., 2012) and are listed in Table 1.

---Table 1 here---

The nonwords were matched to the words as closely as possible. As shown in Table 1, they were not significantly different in most variables, according to the Kolmogorov-Smirnov non-parametric criterion. Neighborhood was an unavoidable exception, because otherwise long nonwords with many neighbors would inescapably be very similar to individual known words and their inflectional variants. That is, longer nonwords would resemble real words more strongly than shorter words, thus, confounding the effects of length and neighborhood. Therefore, to avoid systematic activation of known words we selected nonwords with as few neighbors as possible — and, for this reason, orthographic neighborhood was not included in the main nonword analyses.

---Table 2 here---

Stimuli were selected to minimize the covariance between major predictor variables, aiming to keep their unique effects as distinct as possible. As seen in Table 2, most correlations were very small and not significant for words (an exception is the first orthographic bigram and first phonological syllable frequency). For nonwords, again most correlations were not significant. An important exception was orthographic neighborhood, which was significantly correlated with length because of the length range of our stimuli and the restriction to select nonwords with few or no neighbors, which was not possible for the shorter nonwords. In addition, for the nonwords, first phonological syllable frequency was significantly correlated with first orthographic bigram and average orthographic bigram frequency, while average phonological syllabic frequency was significantly correlated with average orthographic bigram frequency.

Apparatus and procedure. Participants performed a naming and a lexical decision (LD) task. In the naming task, words and nonwords were presented in separate blocks[5]. Block order was counterbalanced between participants. Participants were asked to read each stimulus aloud. Responses were recorded via a headset microphone. In the LD task, words and nonwords were presented in the same intermixed block. Participants were asked to press one key on the keyboard for words and another for nonwords (the right and left CRTL keys). Correspondence between response type (i.e., lexicality), button placement, and handedness was counterbalanced between participants. Task order was counterbalanced between participants, to eliminate task order effects (see Protopapas & Kapnoula, 2016 for an examination of such effects). A digit-span task was inserted between the two tasks to minimize the effect of memory traces for specific stimuli. In both tasks, each stimulus was presented on a 12.1" laptop screen in white 36-pt Arial font on black background for 2000 ms. Stimulus order was randomized individually between participants (within task and block). Stimulus presentation and response recording was controlled by DMDX (Forster & Forster, 2003). The entire session lasted approximately one hour.

[5] This was done to minimize any effects on real word naming due to repeated nonword naming. As has been suggested in the literature (Coltheart & Rastle, 1994; Rastle & Coltheart, 1999), repeated exposure to nonwords can lead to a shift in the balance between the two routes, such that readers increase the contribution from the non-lexical route to phonemic activation and decrease the contribution from the lexical route. That is, as Coltheart (1978) suggested, when a reader is exposed to many nonwords in a row, they adjust their strategy either by turning down the lexical route, or turning up the non-lexical route, or both.

Model data collection

Greek DRC. The architecture of the Greek version of the DRC model is identical to that of the English version 1.2.1 (http://www.cogsci.mq.edu.au/~ssaunder/DRC/2009/10/drc-1-2-1/). However, certain changes were necessary for the DRC to handle polysyllabic Greek stimuli. In the lexical route, the two lexica (orthographic and phonological) of the Greek model contained all 217,662 words in the IPLR C corpus (Protopapas et al., 2012) and their associated frequencies. In the non-lexical route, the GPC rules were replaced by the Greek rule set of Protopapas and Vlahou (2009); originally based on Petrounias, 2002). The two routes were balanced by adjusting parameters to achieve accurate performance[6].

Additional adjustments included: (a) Increasing the number of units in each position of the visual feature and letter layers to 37 — one for each of the 36 Greek letters, and an extra unit for the 'blank', (b) increasing the number of units in each position of the phoneme layer to 33 — one for each of the 32 Greek phonemes, and one for the 'blank', and (c) increasing the number of positions in each of the visual feature, letter, and phoneme layers to 24 to accommodate long Greek words. Crucially, the Greek DRC could handle polysyllabic words correctly, including stress assignment, because stress is orthographically marked in Greek with a stress diacritic. Therefore, stress assignment was represented in the model by distinguishing between stressed and unstressed vowel letters (e.g., unstressed α differed from stressed ά).

The stimuli used in the behavioral experiment were provided as input to the model and the phonological output of the model was noted. Following Coltheart et al. (2001), the number of processing cycles was used as a proxy for response latency in the naming task. The LD task is simulated by the DRC based on the three decision criteria described by Coltheart et al. (2001): if

[6] Details about model parameters and their values are listed in the Appendix.

any one entry in the orthographic lexicon reaches a minimum activation level (A), or if the overall sum of activations among the entries in the orthographic lexicon reaches a given threshold (S), then a "yes" response is given. Otherwise, if a given number of cycles (D) has elapsed and neither of these two aforementioned thresholds were reached, then a "no" decision is made. We collected LD modeling data using the DRCLD tool (available at https://github.com/stevenjs/drcld).

Results

Descriptives and pre-processing

Spoken responses from the naming task were processed with CheckVocal (Protopapas, 2007) to check accuracy and placement of response time (RT) marks. Incorrect responses, defined as responses with any phonemic difference from the canonical pronunciation[7], were excluded from RT analysis. Average naming accuracy was 99.1% for the words and 91.1% for the nonwords. In the lexical decision task, response accuracy was 95.3% for the words and 96.4% for the nonwords. The DRC was 100% successful in pronouncing (naming task) and classifying (LD task) all stimuli. Response latency data are presented in Table 3.

---Table 3 here---

[7] Defining naming errors was straightforward because nonword stimuli had unambiguously correct pronunciations. Naming errors were mostly due to haste and inattention. For example, a nonword like δογορονήσω [ðoɣoroníso] could be misnamed as [ɣoðoroníso].

Main analyses

Stage 1: Variable selection. The aforementioned correlations (between first phonological syllable frequency and first orthographic bigram frequency for words, and among first phonological syllable frequency, average phonological syllabic frequency, and average orthographic bigram frequency for nonwords) make interpreting the data via simple linear regression problematic. Therefore we followed Yap and Balota (2009) and conducted a series of hierarchical regressions to estimate the unique variance accounted for by each variable before entering them into the final multiple regression equation. The dependent variable was RT. To allow comparisons between human and DRC data, RTs were averaged across all participants for each item (word and nonword), separately for the naming and the LD task.

The effect of the initial phoneme (dummy-coded into classes of vowel, liquid, fricative, voiced stop, and unvoiced stop consonant) was removed from the naming RT in a preliminary regression step (not shown). We then entered frequency, length, orthographic neighborhood, and grapho-phonemic transparency in the first step, for words; for nonwords, we only entered length and grapho-phonemic transparency (orthographic neighborhood was excluded, because, as explained in the **Method** section, it was impossible to de-correlate it from length). In the second step, four different models were created, by entering each of the four target variables (first phonological syllable frequency, average phonological syllabic frequency, first orthographic bigram frequency, and average orthographic bigram frequency). Each variable was then entered in a third step in every combination with second-step variables, thus creating a total of 12 models for each task and each stimulus type (see Tables 4 and 5). This allowed us to better assess the relative importance of each of our key predictors (i.e. syllable and bigram frequency measures).

---Table 4 here---

---Table 5 here---

For words, first phonological syllable frequency had a marginally significant effect on naming RTs, accounting for about 2% of the residual item variance, and a significant effect on LD RTs, accounting for about 6% of the variance. First orthographic bigram frequency was also a significant predictor, accounting for about 3% of LD item variance when entered alone in the second step. First phonological syllable frequency remained a marginally significant predictor after first orthographic bigram frequency variance was removed, but the opposite was not observed.

For nonwords, average orthographic bigram frequency had a significant effect on naming RTs, accounting for 6.6% of the residual item variance, while first orthographic bigram frequency accounted for about 4%. Average orthographic bigram frequency remained significant in the third step, after controlling for first orthographic bigram frequency, while first orthographic bigram frequency became marginally significant after controlling for average orthographic bigram frequency. None of the target variables accounted for a significant proportion of LD RT variance.

Given these results, we decided to include only first phonological syllable frequency in the analyses for the words and first phonological syllable frequency and average orthographic bigram frequency for the nonwords.

Stage 2: Task and model comparisons. Again, the effect of the initial phoneme was removed from naming RTs before entering the rest of the variables. We then conducted six multiple regression analyses, namely 2 types of stimuli × (2 tasks plus model simulation).

For the words, frequency, length, orthographic neighborhood, and grapho-phonemic transparency were entered in the first step and first phonological syllable frequency in the second step (see Table 6). Frequency and length had a significant effect in both tasks. Orthographic neighborhood and grapho-phonemic transparency were not significant in either task[8]. Overall, these four variables accounted for 39.9% of the residual item variance in naming and 33.3% in LD. When first phonological syllable frequency was entered in the second step, it accounted for an additional 6% in the LD task.

Turning to the DRC simulation data, frequency but not length had a significant effect on the number or processing cycles in the first step, for both (naming and LD) types of simulations. Grapho-phonemic transparency was a significant predictor only for the naming simulation, whereas orthographic neighborhood was marginally significant only for the LD. Overall, these four variables accounted for 45.4% of the total variance in the model naming “latency”, and 96.4% of the total variance in the model LD “latency”. When first phonological syllable frequency was entered in the second step, it did not account for significant additional variance in either of the simulations. We also analyzed the word data including both first phonological syllable frequency and average orthographic bigram frequency for a more direct comparison with the nonword data analyses. The results for the two sets of human data (naming and LD RTs) were almost identical to the reported results including only first phonological syllable frequency.

[8] Even though we used Coltheart’s N in our main analyses (because this measure of neighborhood was decorrelated from all other major predictors), we also performed the same set of analyses using OLD20 in its place. The results showed no difference between using Coltheart’s N and OLD20: Naming RT: β=.09781, p=.256; LD RT: β=.147, p=.106.

For the DRC data, we found a significant effect of average orthographic bigram frequency (accounting for 25% of the variance).

---Table 6 here---

For the nonwords, we conducted similar analyses (Table 7), excluding frequency and orthographic neighborhood from the predictor variables in the first step and including first phonological syllable frequency and average orthographic bigram frequency in the second step. Length had a significant effect in both naming and LD, while grapho-phonemic transparency was not significant in either task. Together these two variables accounted for 68.6% of the residual item variance in naming and 39.1% in LD. In the second step, both first phonological syllable frequency and average orthographic bigram frequency had a significant effect in naming. These two variables combined accounted for an additional 9% of the residual item variance. None of them was a significant predictor of RT in the LD task.

We only analyzed DRC naming data because the model's LD latency was identical for all nonwords (63 cycles)[9]. Length and grapho-phonemic transparency both had a significant effect accounting for 95.4% of the variance. In addition, when added in the second step, average orthographic bigram frequency was also a significant predictor of the model's response latency. Average orthographic bigram frequency and first phonological syllable frequency combined accounted for 4.6% of the residual item variance.

[9] Given that the maximum number of cycles is reached for all nonwords, it is reasonable that the model latencies should be identical across nonword stimuli.

---Table 7 here---

Table 8 summarizes the findings from the comparison of the individual effects between human and modeling data. As shown, DRC successfully simulated the frequency effect on RTs to known word stimuli, as well as the length effect observed for nonwords. This was expected because they reflect aspects of the model that are explicitly designed into it. However, in contrast to the participants' behavior, DRC was not sensitive to length for known words. In addition, even though our participants' responses were overall affected to a greater degree by syllable frequency compared to grapho-phonemic transparency, the opposite pattern was observed for the DRC.

---Table 8 here---

Discussion

In this study we evaluated the DRC model by conducting a qualitative and quantitative comparison of the model's responses to those of native speakers of Greek. Specifically, after selecting stimuli with relatively balanced characteristics (to avoid significant covariance between major predictor variables), we conducted multiple regressions to assess the unique variance accounted for by each of the variables. We did this separately for the model and human responses and we compared the two.

Behavioral data

Our behavioral data provide useful insights into the processes of visual word recognition and reading aloud in a relatively transparent orthography. Although word frequency was found to be a significant predictor of latency (as expected), its contribution did not eliminate other effects (i.e. length and syllabic frequency). The effect of length was also expected, given the range of our stimuli (2–5 syllables long). However, this effect was significant for both nonwords and familiar words in both tasks (naming and LD), which suggests that proficient readers are sensitive to length even in the case of familiar lexical items, consistent with findings with multisyllabic words in other languages (e.g., (Barca et al., 2002; Ferrand et al., 2010, 2011; Keuleers, Diependaele, & Brysbaert, 2010; Yap & Balota, 2009). One may argue that this finding stands in the face of a fundamental assumption of the dual-route theory, according to which familiar lexical items are primarily processed via the (faster) lexical route that allows for parallel activation of the word's sublexical parts. However, our modeling data (discussed in the next section) speak more directly to this issue of whether length effects can be observed when parallel grapho-phonemic conversion is available.

Despite the robust neighborhood effects reported in the literature for other languages, we did not observe a significant effect of neighborhood on RT in either task. This may seem counterintuitive, however, as discussed by Protopapas and Kapnoula (2016), the concept of neighborhood is problematic for Greek. This is because Greek content words are inflected by suffixation that marks different grammatical properties (such as gender and number[10]) and, as a result, different forms of the same word count as neighbors, according to common definitions of neighborhood such as Coltheart's N (Coltheart et al., 1977). Therefore, it may be that in the case of Greek (and other morphologically rich languages) this type of neighbors may not act as competitors and may instead facilitate activation of the target word. Moreover, Greek has relatively long words (token mean 5.4 letters, 2.4 syllables; Protopapas et al., 2012; type mean

[10] For example, *ταχυδρόμος* (taçiðɾˈomos), *ταχυδρόμου* (taçiðɾˈomu), *ταχυδρόμο* (taçiðɾˈomo), and *ταχυδρόμε* (taçiðɾˈomɛ) all correspond to the word *postman* in nominative, genitive, accusative, and vocative respectively.

10.1 letters, versus 7.8 in English, Balota et al., 2007, and 8.4 in French, New, Pallier, Ferrand, & Matos, 2001), resulting in much sparser neighborhoods than English, especially if inflectional variants are excluded, and, therefore, fewer opportunities for neighborhood effects.

Turning to sublexical factors, the effect of first syllable frequency has been reported in languages with relatively transparent orthographies, such as Spanish (Carreiras et al., 1993) and German (Conrad & Jacobs, 2004). Because Greek is also relatively transparent, we expected similar results. In addition, by carefully constructing stimuli sets for which our predictor variables were largely decorrelated, we could dissociate the effects of first syllable frequency, average syllable frequency, and bigram frequency (for words), unaffected by other major variables. Our results showed that first syllable frequency was the most robust predictor among these sublexical variables. This finding supports the hypothesis that syllable frequency effects stem from the interference from partially activated competitor words that share their first syllable with the target (Carreiras et al., 1993). In other words, these are phonological cohort effects.

Focusing on the role of syllable frequency in word recognition, we observed a marginally significant effect of first syllable frequency on RT in the word naming task, which was significant for nonwords, and a significant effect of first syllable frequency on RT in the LD task for words. These findings are in line with the *orthographic depth hypothesis* (Katz & Frost, 1992) as they indicate that in conditions of high transparency, and, therefore, highly reliable GPC rules, readers may utilize sublexical representations to a high degree. In contrast, in less transparent languages, like English, syllables might not be utilized as much (Ziegler et al., 1997; related arguments regarding processing unit size have been made also on the basis of length effects, e.g., for Italian, Marinelli, et al, 2016, and eye movement patterns, e.g., for German, Rau et al., 2016, among others). Furthermore, our results show that phonologically defined syllables may hold a unique status among sublexical representations. That is, our findings suggest that the effect of syllable frequency is dissociable from that of bigram frequency, which was only observed for nonwords and only in the naming task. These results are in line with the idea that syllabic representations are rapidly and automatically activated during visual word recognition (Mahé et al., 2014; Perea & Carreiras, 1998) and that the course of activation of these representations is directional, consistent with left-to-right syllabic processing of the visual stimulus. Consistent with this interpretation, recent neuroimaging data using fMRI for Greek listeners responding to the same stimulus set has localized the syllabic frequency effect at left inferior frontal cortex (Protopapas et al., 2016), which is typically associated with phonological rather than orthographic representations.

Comparison between behavioral and modeling data

The Greek DRC model simulated the human pronunciation of all the stimuli (both words and nonwords) 100% accurately, although it was overly sensitive to lexicality; nonwords averaged more than twice the cycles needed for words, whereas the corresponding increase in human naming RTs was only about 30%. The DRC model's sensitivity to word frequency, as well as nonword length and bigram frequency, was in line with the behavioral data. However, in contrast to human readers, DRC did not show sensitivity to length when the stimuli were known words. Moreover, DRC showed more sensitivity to GPC transparency than to syllabic frequency, which was the opposite of the pattern observed in the behavioral responses. These inconsistencies may reflect a divergence between the Greek readers and the DRC architecture, possibly revealing a fundamental difference in terms of which unit is most relevant for processing (i.e., syllable versus grapheme).

Frequency information is represented in the model, as “a constant value [...] associated with each unit in the lexicon” (Coltheart et al., 2001) that determines the activation level of that unit given all other factors are the same. Therefore, it is not surprising that DRC successfully simulated the word frequency effect. The model also naturally accounts for the observed length effect in nonwords; unknown words can only be “read” via serial grapheme-to-phoneme conversion (through the non-lexical route) – as a result, longer nonwords take more time to be processed. In contrast, in the case of known words, graphemes are converted to phonemes in parallel (through the lexical route); this explains the model’s insensitivity to word length. If the length effect for known words was only significant in transparent orthographies, this could be viewed as a minor limitation of the model, but length effects with words have also been found in English (Perry et al., 2010; Yap & Balota, 2009). A possible solution to this problem may be to adjust the parameters that determine the relative strength of the two routes (lexical and non-lexical). We tested this by running the same analysis on the model responses generated by a version of DRC with adjusted parameter settings. As expected, we found that when the non-lexical route was strengthened length became a significant predictor of response latency ($\beta = .47$, $p < .001$); but this occurred at the expense of accuracy (3 naming errors, i.e. 2%). It is an open question whether this trade-off between the strength of the non-lexical route and naming accuracy is telling us something about how humans read words, or whether it reflects an underlying constraint of the specific model. Even though further simulations and model comparisons are needed to address this question, this may suggest that the DRC relies too heavily on lexical-level processing, compared to Greek readers, who also rely on the mapping of sublexical units, such as graphemes and syllables (possibly due to the high transparency of the language; see Katz & Frost, 1992). An alternative approach would be to switch to a parallel distributed processing description, which has been shown to simulate length effects in words (Chang, Furber, & Welbourne, 2012; see Protopapas & Outos, 2009, for an earlier model specifically targeting Greek).

With respect to the sublexical variables in the focus of this study, DRC failed to simulate the effect of syllable frequency in Greek words, showing instead sensitivity to grapho-phonemic transparency (which was, however, absent in the behavioral data). The performance of the DRC matched the behavioral results for nonwords more closely, as the model showed sensitivity to nonword length and bigram frequency, like humans readers in the naming task. There was also a marginally significant effect of the first syllable frequency on DRC latency for the nonwords, but it was in the opposite direction of that observed in the behavioral results. We believe this pattern of results to be informative as to the limitations of the DRC and our approach to modeling reading behavior more generally. Next, we turn to our interpretation of these findings.

This pattern of inconsistencies may reflect an underlying divergence between the Greek readers and the DRC architecture, perhaps in terms of which unit is most relevant for processing (e.g. whole word, syllable, or grapheme). That is, the absence of a syllable frequency effect for words could be attributed to the fact that in the DRC architecture, reading aloud of known words relies predominantly on the lexical route, in which words are processed holistically. As a result, DRC processing latency for words is largely determined by frequency (which is a characteristic of the word as a whole), but it is not affected much by sublexical characteristics, like syllable frequency. In addition, graphemes are explicitly defined as representational units in the model, whereas syllables are not, and this may be why the model only shows sensitivity to the former.

The relatively poor match of the DRC model to Greek reading behavior is also exemplified in the rigidity of its responses to pseudowords in the case of ambiguities. As noted by Protopapas and Vlahou (2009), although Greek is mostly predictable in the forward (reading) direction, a widespread case of frank ambiguity exists. Specifically, any combination of a consonant followed by an unstressed grapheme corresponding to /i/ followed by a vowel (termed CiV) can be legally pronounced either with an /i/ phoneme (a two-syllable reading) or with a palatal consonant and no /i/ (a one-syllable reading). Protopapas and Nomikou (2009) constructed pseudowords containing this ambiguity, manipulating their orthographic overlap to existing words and the preponderance of /i/ vs. palatal readings of each CiV combination. They found that Greek readers were influenced both by similar individual lexical items and by the probabilistic associations when reading these pseudowords. However, when the same items were submitted to the Greek DRC, it was found that they were all pronounced according to the implemented GPC rule, unaffected by lexical resemblance or probabilistic association. The balance between the lexical and sublexical DRC routes is chosen to ensure accurate performance on the corpus words and unambiguous nonwords; however, this balance may have precluded the model from flexibly exploiting lexical similarity and statistics to simulate the reading behavior of Greek skilled readers[11]. Thus, in this respect as well, it seems that the empirical situation of Greek skilled reading may not lend itself to the rigid distinction between lexical and sublexical processing and the absolute, hard-coded rules of the GPC route, as currently implemented in the DRC.

[11] Despite it being a rule-based model, DRC could in principle exhibit sensitivity to lexical similarity when processing nonwords if the lexical route is strengthened. Keep in mind that for a word item to be named all that is required is for activation of its lexical entry to reach a given threshold. Critically, this may happen even if the input does not match the lexical entry 100%. For example, the nonword "signeficant" may be incorrectly named as "significant" if activation for the phonological entry of the word "significant" surpasses a given threshold.

In either case, the DRC might conceivably be adjusted to simulate the effects we observe in Greek, either by adjusting the balance between lexical and non-lexical processing[12], and/or by explicitly defining syllable representations (cf. Conrad et al., 2010, for the multiple readout model). One problem with such modifications would be that the dynamics of the model might be affected in a way that interferes with the successful simulation of other effects (Conrad & Jacobs, 2004). However, we also see a theoretical problem with this kind of solution. If we explicitly define each aspect of our behavioral findings into a model, then we cannot learn much about the behavior; the model may in fact exhibit the desired effect, or pattern of effects, in a way that closely matches the behavioral data, but that would likely be the result of our explicit modification. This goes against the motivation behind using computational modeling in the first place. An alternative approach would be to design a model that is sensitive to the characteristics of the input in a way that it can *learn* what kind of representational units are relevant to processing (e.g. syllables or bigrams) and how the effects of different kinds of representations should be balanced against one another (e.g. see Seidenberg & McClelland, 1989). This might allow the model to exhibit flexibility in simulating adult readers with different learning experiences in a variety of orthographic systems.

Turning back to our initial question, raised in the introduction, even though we did not directly test the necessity of learning, we believe that our results attest to its potential usefulness. Our findings suggest that computational models built to explain a specific set of behavioral phenomena (e.g., in a specific language) may have substantial limitations in terms of their generalizability. In contrast, incorporating learning as a fundamental aspect in our modeling endeavors (i.e. even when learning itself is not the object of the simulation) may help us reach a

[12] Note that the DRC version with adjusted parameter settings mentioned above (i.e. strengthened non-lexical route) was still not able to simulate the syllable frequency effect: $\beta = -0.029$, $p = 0.728$.

unified theory of visual word recognition that is better able to account for the diverse patterns of results observed across languages and populations.

Conclusion and further directions

This is the first study to evaluate a computational model of visual word recognition and reading aloud with Greek stimuli by directly comparing modeling against behavioral data. Our behavioral results showed that visual word recognition in Greek is affected by both lexical and sublexical factors; less frequent, longer words with higher frequency first syllable were recognized more slowly. DRC was successful in simulating many aspects of the behavioral data, but our evaluation also revealed significant limitations of the model. For example, DRC showed insensitivity to length and syllabic frequency effects (which were present in the behavioral data). Such inconsistencies may reflect underlying differences, for example, in terms of the representational units used in processing (e.g. syllable versus graphemes). Certain adjustments might lead to improved simulation outcomes (e.g. parameter adjustment and/or explicitly defining syllabic representations). However, an alternative approach would be for these effects to arise within a system in which they are not explicitly hard-wired. For example, given the learning procedure via which GPC rules are acquired in the CDP++ model, it would be interesting to see whether a syllabic level of representation would naturally emerge.

References

Adelman, J., & Brown, G. (2008). Methods Of Testing And Diagnosing Model Error: Dual And Single Route Cascaded Models Of Reading Aloud. *Journal of Memory and Language*, *59*(4), 524–544.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking The Time Course Of Spoken Word Recognition Using Eye Movements: Evidence For Continuous Mapping Models. *Journal of Memory and Language*, *38*(4), 419–439.

Arduino, L. S., & Burani, C. (2004). Neighborhood Effects On Nonword Visual Processing In A Language With Shallow Orthography. *Journal of Psycholinguistic Research*, *33*(1), 75–95.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459.

Barca, L., Burani, C., & Arduino, L. S. (2002). Word Naming Times And Psycholinguistic Norms For Italian Nouns. *Behavior Research Methods, Instruments, & Computers*, *34*(3), 424–434.

Carreiras, M., Alvarez, C. J. C., & Devega, M. (1993). Syllable Frequency And Visual Word Recognition In Spanish. *Journal of Memory and Language*, *32*(6), 766–780.

Carreiras, M., & Perea, M. (2004). Naming Pseudowords In Spanish: Effects Of Syllable Frequency. *Brain and Language*, *90*(1), 393–400.

Chang, Y., Furber, S., & Welbourne, S. (2012). “Serial” Effects In Parallel Models Of Reading. *Cognitive Psychology*, *64*(4), 267–291.

Chetail, F. (2015). Reconsidering The Role Of Orthographic Redundancy In Visual Word Recognition. *Frontiers in Psychology*, *6*, 645.

Coltheart, M. (1978). Lexical Access In Simple Reading Tasks. In G. Underwood (Ed.), *Strategies of Information Processing* (pp. 151–216). London: Academic Press.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models Of Reading Aloud: Dual-route And Parallel-distributed-processing Approaches. *Psychological Review*, *100*(4), 589.

Coltheart, M., Davelaar, E., Jonasson, T., & Besner, D. (1977). Access To The Internal Lexicon.

Coltheart, M., & Rastle, K. (1994). Serial Processing In Reading Aloud: Evidence For Dual-route Models Of Reading. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(6), 1197.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A Dual Route Cascaded Model Of Visual Word Recognition And Reading Aloud. *Psychological Review*, *108*, 204–56.

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do The Beginnings Of Spoken Words Have A Special Status In Auditory Word Recognition? *Journal of Memory and Language*, *32*(2), 193–210.

Conrad, M., Carreiras, M., & Jacobs, A. (2008). Contrasting Effects Of Token And Type Syllable Frequency In Lexical Decision. *Language and Cognitive Processes*, *23*(2), 296–326.

Conrad, M., Carreiras, M., Tamm, S., & Jacobs, A. M. (2009). Syllables And Bigrams: Orthographic Redundancy And Syllabic Units Affect Visual Word Recognition At Different Processing Levels. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 461–479.

Conrad, M., Grainger, J., & Jacobs, A. (2007). Phonology As The Source Of Syllable Frequency Effects In Visual Word Recognition: Evidence From French. *Memory & Cognition*, *35*(5), 974–983.

Conrad, M., & Jacobs, A. (2004). Replicating Syllable Frequency Effects In Spanish In German: One More Challenge To Computational Models Of Visual Word Recognition. *Language and Cognitive Processes*, *19*(3), 369–390.

Conrad, M., Tamm, S., Carreiras, M., & Jacobs, A. M. (2010). Simulating Syllable Frequency Effects Within An Interactive Activation Framework. *European Journal of Cognitive Psychology*, *22*(5), 861–893.

Ferrand, L., Brysbaert, M., Keuleers, E., New, B., Bonin, P., Méot, A., Augustinova, M., & Pallier, C. (2011). Comparing Word Processing Times In Naming, Lexical Decision, And Progressive Demasking: Evidence From Chronolex. *Frontiers in Psychology*, *2*, 306.

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., Augustinova, M., & Pallier, C. (2010). The French Lexicon Project: Lexical Decision Data For 38,840 French Words And 38,840 Pseudowords. *Behavior Research Methods*, *42*(2), 488–496.

Ferrand, L., Segui, J., & Grainger, J. (1996). Masked Priming Of Word And Picture Naming: The Role Of Syllabic Units. *Journal of Memory and Language*, *35*(5), 708–723.

Forster, K. I., & Chambers, S. M. (1973). Lexical Access And Naming Time. *Journal Of Verbal Learning And Verbal Behavior*, *12*, 627–635.

Hawelka, S., Schuster, S., Gagl, B., & Hutzler, F. (2013). Beyond Single Syllables: The Effect Of First Syllable Frequency And Orthographic Similarity On Eye Movements During Silent Reading. *Language and Cognitive Processes*, *28*(8), 1134–1153.

Katz, L., & Frost, R. (1992). The Reading Process Is Different For Different Orthographies: The Orthographic Depth Hypothesis. *Advances in Psychology*.

Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice Effects In Large-Scale Visual Word Recognition Studies: A Lexical Decision Study On 14,000 Dutch Mono- And Disyllabic Words And Nonwords. *Frontiers in Psychology*, *1*, 174.

Mahé, G., Bonnefond, A., & Doignon-Camus, N. (2014). The Time Course Of The Syllable Frequency Effect In Visual Word Recognition: Evidence For Both Facilitatory And Inhibitory Effects In French. *Reading and Writing*, *27*(1), 171–187.

Marinelli, C. V., Romani, C., Burani, C., McGowan, V. A., & Zoccolotti, P. (2016). Costs And Benefits Of Orthographic Inconsistency In Reading: Evidence From A Cross-Linguistic Comparison. (M. Perea, Ed.)*PLOS ONE*, *11*(6), e0157457.

Marshall, J. C., & Newcombe, F. (1973). Patterns Of Paralexia: A Psycholinguistic Approach. *Journal of Psycholinguistic Research, 2*, 175–199.

Marslen-Wilson, W. D., Moss, H. E., & Van Halen, S. (1996). Perceptual Distance And Competition In Lexical Access. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 1376–1392.

Mousikou, P., Sadat, J., Lucas, R., & Rastle, K. (2017). Moving Beyond The Monosyllable In Models Of Skilled Reading: Mega-study Of Disyllabic Nonword Reading. *Journal of Memory and Language*, *93*, 169–192.

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une Base De Données Lexicales Du Français Contemporain Sur Internet : LEXIQUE™//A Lexical Database For Contemporary French : LEXIQUE™. *L'année Psychologique*, *101*(3), 447–462.

Norris, D. (2013). Models Of Visual Word Recognition. *Trends in Cognitive Sciences*, *17*(10), 517–524.

Perea, M., & Carreiras, M. (1998). Effects Of Syllable Frequency And Syllable Neighborhood Frequency In Visual Word Recognition. *Journal of Experimental Psychology Human Perception and Performance*, *24*, 134–144.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested Incremental Modeling In The Development Of Computational Theories: The CDP+ Model Of Reading Aloud. *Psychological Review*, *114*, 273–315.

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond Single Syllables: Large-scale Modeling Of Reading Aloud With The Connectionist Dual Process (CDP++) Model. *Cognitive Psychology*, *61*(2), 106–151.

Petrounias, E. (2002). Neoellinikı Grammatikı Kai Sigkritikı Análisi, Tómos A: Fonitikı Kai Eisagogı Sti Fonologıa [Modern Greek Grammar And Comparative Analysis, Vol A: Phonetics. *Thessaloniki: Ziti*.

Pritchard, S. C., Coltheart, M., Marinus, E., & Castles, A. (2016). Modelling The Implicit Learning Of Phonological Decoding From Training On Whole-word Spellings And Pronunciations. *Scientific Studies of Reading*, *20*(1), 49–63.

Protopapas, A. (2007). CheckVocal: A Program To Facilitate Checking The Accuracy And Response Time Of Vocal Responses From DMDX. *Behavior Research Methods*, *39*, 859–862.

Protopapas, A., & Kapnoula, E. C. (2013). Exploring Word Recognition With Selected Stimuli: The Case For Decorrelated Parameters. In & I. W. M. Knauff, M. Pauen, N. Sebanz (Ed.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1169–1174). Berlin, Germany.

Protopapas, A., & Kapnoula, E. C. (2016). Short-term And Long-term Effects On Visual Word Recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(4), 542–565.

Protopapas, A., & Nomikou, E. (2009). Rules Vs. Lexical Statistics In Greek Nonword Reading. In N. Taatgen & H. van Rijn (Eds.), *31st Annual Conference of the Cognitive Science*

Society. (pp. 1030–1035). Amsterdam, Netherlands: Cognitive Science Society.

Protopapas, A., Orfanidou, E., Taylor, J. S. H., Karavasilis, E., Kapnoula, E. C., Panagiotaropoulou, G., Velonakis, G., Poulou, L. S., Smyrnis, N., & Kelekis, D. (2016). Evaluating Cognitive Models Of Visual Word Recognition Using FMRI: Effects Of Lexical And Sublexical Variables. *NeuroImage*, *128*, 328–341.

Protopapas, A., & Outos, K. D. (2009). Reading Aloud Multisyllabic Words: A Single-Route Connectionist Model For Greek. In *9th International Conference on Cognitive Modeling, Manchester, UK, 24–26*. Manchester, UK.

Protopapas, A., Tzakosta, M., Chalamandaris, A., & Tsiakoulis, P. (2012). IPLR: An Online Resource For Greek Word-level And Sublexical Information. *Language Resources and Evaluation*, *46*(3), 449–459.

Protopapas, A., & Vlahou, E. (2009). A Comparative Quantitative Analysis Of Greek Orthographic Transparency. *Behavior Research Methods*, *41*, 991–1008.

Rastle, K., & Coltheart, M. (1999). Serial And Strategic Effects In Reading Aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(2), 482.

Rau, A. K., Moll, K., Moeller, K., Huber, S., Snowling, M. J., & Landerl, K. (2016). Same Same, But Different: Word And Sentence Reading In German And English. *Scientific Studies of Reading*, *20*(3), 203–219.

Reynolds, M., & Besner, D. (2004). Neighbourhood Density, Word Frequency, And Spelling-Sound Regularity Effects In Naming: Similarities And Differences Between Skilled Readers And The Dual Route Cascaded Computational Model. *Canadian Journal of Experimental Psychology*, *15*, 13–31.

Schmalz, X., Robidoux, S., Castles, A., Coltheart, M., & Marinus, E. (2017). German And English Bodies: No Evidence For Cross-Linguistic Differences In Preferred Orthographic Grain Size. *Collabra: Psychology*, *3*(1), 5.

Seidenberg, M. S., & McClelland, J. L. (1989). A Distributed, Developmental Model Of Word Recognition And Naming. *Psychological Review*, *96*(4), 523–568.

Spencer, K. A. (2009). Feedforward, -backward, And Neutral Transparency Measures For British English. *Behavior Research Methods*, *41*(1), 220–7.

Spieler, D. H., & Balota, D. A. (2000). Factors Influencing Word Naming In Younger And Older Adults. *Psychology and Aging*, *15*, 225–231.

Weekes, B. S. (1997). Differential Effects Of Number Of Letters On Word And Nonword Naming Latency. *The Quarterly Journal of Experimental Psychology Section A*, *50*(2), 439–456.

Yap, M., & Balota, D. A. (2009). Visual Word Recognition Of Multisyllabic Words. *Journal of Memory and Language*, *60*(4), 502–529.

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving Beyond Coltheart's N: A New Measure Of Orthographic Similarity. *Psychonomic Bulletin & Review*, *15*(5), 971–979.

Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). What Is The Pronunciation For -ough And

The Spelling For /u/? A Database For Computing Feedforward And Feedback Consistency In English. *Corpus*, *29*, 600–618.

Zorzi, M. (2010). The Connectionist Dual Process (CDP) Approach To Modelling Reading Aloud. *European Journal of Cognitive Psychology*, *22*(5), 836–860.

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two Routes Or One In Reading Aloud? A Connectionist Dual-process Model. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(4), 1131.

Appendix

List of parameter values for Greek DRC

The values of the following parameters were adjusted using a process of trial-and-error until the model produced no errors on any vocabulary word, or on any of a set of 1000 nonwords, with the minimum reading phonology set to 0.9 to simulate reading at leisure:

Parameter	Value
General Parameters	
ActivationRate	0.2
FrequencyScale	0.05
MinReadingPhonology	0.4
FeatureLevel Parameters	
FeatureLetterExcitation	0.005
FeatureLetterInhibition	0.15
Letter Level Parameters	
LetterOrthlexExcitation	0.02
LetterOrthlexInhibition	0.74
LetterLateralInhibition	0
LetterNoise	0
LetterDecay	0
LetterUnsupportedDecay	0
Orthographic Lexicon (Orthlex)	
OrthlexPhonlexExcitation	0.3
OrthlexPhonlexInhibition	0
OrthlexLetterExcitation	0.2
OrthlexLetterInhibition	0
OrthlexLateralInhibition	0.35
OrthlexNoise	0
OrthlexDecay	0
OrthlexUnsupportedDecay	0
OrthlexThreshold	0
Phonological Lexicon (Phonlex)	
PhonlexPhonemeExcitation	0.09
PhonlexPhonemeInhibition	0
PhonlexOrthlexExcitation	0.25

PhonlexOrthlexInhibition	0
PhonlexStressExcitation	0
PhonlexStressInhibition	0
PhonlexLateralInhibition	0.35
PhonlexNoise	0
PhonlexDecay	0
PhonlexUnsupportedDecay	0
PhonlexThreshold	0
Phoneme Level Parameters	
PhonemePhonlexExcitation	0.02
PhonemePhonlexInhibition	0.16
PhonemeLateralInhibition	0.147
PhonemeNoise	0
PhonemeDecay	0
PhonemeUnsupportedDecay	0.05
PhonemeThreshold	0
GPC Route Parameters	
GPCPhonemeExcitation	0.051
GPCRhymePhonemeExcitation	0
GPCStressExcitation	0
GPCInputThreshold	0
GPCOutputThreshold	0
GPCCriticalPhonology	0.05
GPCOnset	26
GPCOnsetThreshold	0
GPCNoise	0
Stress Level Parameters	
StressPhonlexExcitation	0
StressPhonlexInhibition	0
StressLateralInhibition	0
StressNoise	0
StressDecay	0
StressUnsupportedDecay	0
StressThreshold	0

Table 1

Descriptive statistics for the stimuli (words and nonwords) versus corpus statistics and comparison between words and nonwords via the Kolmogorov-Smirnov (K-S) test for the equality of distributions. This information is also presented in Protopapas and Kapnoula (2016) and Protopapas et al. (2016).

	Words		Nonwords		K-S test (words-nonwords)		Corpus tokens	
	M	*SD*	*M*	*SD*	*Z*	*p*	*M*	*SD*
Frequency	0.877	1.887	-	-	-	-	6.220	3.270
Length	7.240	1.931	7.307	1.839	0.520	.950	5.431	3.166
BA	0.020	0.001	0.020	0.001	0.808	.531	0.021	0.004
SA	12.162	1.054	11.865	1.464	1.097	.180	13.248	1.054
B1	0.020	0.002	0.021	0.002	0.751	.626	0.021	0.004
S1	13.388	1.600	13.136	1.667	0.981	.290	13.756	1.273
ON	2.173	1.501	0.387	1.214	6.062	<.001	5.880	4.830
GPT	3.454	0.264	3.458	0.258	0.462	.983	3.561	0.497

Note: Frequency as log number of occurrences; Length in letters; BA = mean log bigram frequency (all bigrams); SA = mean log syllable frequency (all syllables); B1 = log frequency of the initial bigram; S1 = log frequency of the initial syllable; ON = orthographic neighborhood size; GPT = grapho-phonemic transparency, as mean log token sonograph probability.

Table 2

Nonparametric correlation coefficients (Spearman's ρ) between independent variables for words (above the diagonal) and nonwords (below the diagonal). This information is also presented in Protopapas and Kapnoula (2016) and Protopapas et al. (2016).

	Length	BA	SA	B1	S1	ON	GPT
Frequency	−.049	.103	.042	−.010	.115	−.002	.002
Length		.008	.105	−.005	.017	−.007	−.106
BA	.096		.083	.171*	.062	.071	.056
SA	.088	.213**		−.034	.073	.029	.054
B1	−.027	.017	−.087		.332***	−.038	.115
S1	−.064	−.196*	.025	.254**		.026	−.022
ON	−.504***	−.041	−.015	.097	.064		−.018
GPT	.003	.043	−.144	.126	.002	−.032	

Note: Frequency as log number of occurrences; Length in letters; BA = mean log bigram frequency (all bigrams); SA = mean log syllable frequency (all syllables); B1 = log frequency of the initial bigram; S1 = log frequency of the initial syllable; ON = orthographic neighborhood size; GPT = grapho-phonemic transparency, as mean log token sonograph probability; All DFs = 148. *p<.05, **p<.01, ***p<.001

Table 3

Response Times (in ms for humans and processing cycles for the DRC) for words and nonwords in the two tasks

	Naming			
	Humans		DRC	
	M	*SD*	*M*	*SD*
Words	547.3	42.5	70.4	3.3
Nonwords	712.6	87.8	152.1	11
	Lexical Decision			
	Humans		DRC	
	M	*SD*	*M*	*SD*
Words	716.6	85.8	44.4	1.5
Nonwords	824.1	84.4	63	0

Table 4

Stage 1 regression analysis predicting human latencies for words

		DV: Naming RT			DV: Lexical Decision RT		
Step	Variable	β	ΔR^2	p	β	ΔR^2	p
1	Frequency	−.420		< .001	−.502		<.001
	Length	.446		< .001	.223		.001
	ON	−.024		.712	−.105		.124
	GPT	.022		.740	.044		.519
			.383	< .001		.315	<.001
2	S1	.142	.020	.084	.244	.060	.003
3	SA	.037	.001	.654	.117	.014	.155
3	B1	.031	.001	.711	.048	.002	.563
3	BA	−.002	.000	.982	−.012	.000	.881
2	SA	.039	.002	.631	.118	.014	.149
3	S1	.141	.020	.086	.244	.059	.003
3	B1	.105	.011	.200	.181	.033	.027
3	BA	.001	.000	.986	−.007	.000	.931
2	B1	.102	.010	.214	.170	.029	.037
3	S1	.090	.008	.273	.160	.026	.050
3	SA	.048	.002	.563	.134	.018	.103
3	BA	−.001	.000	.878	−.030	.001	.714
2	BA	.002	.000	.981	−.005	.000	.948
3	S1	.141	.020	.084	.245	.060	.003
3	SA	.039	.002	.632	.118	.014	.149
3	B1	.102	.010	.215	.171	.029	.036

Note: DV = dependent variable; Frequency as log number of occurrences; Length in letters; BA = mean log bigram frequency (all bigrams); SA = mean log syllable frequency (all syllables); B1 = log frequency of the initial bigram; S1 = log frequency of the initial syllable; ON = orthographic neighborhood size; GPT = grapho-phonemic transparency, as mean log token sonograph probability; DFs for step 1 = 145; DFs for steps 2-3 = 148.

Table 5

Stage 1 regression analysis predicting human latencies for nonwords

		DV: Naming RT			DV: Lexical Decision RT		
Step	Variable	β	ΔR^2	p	β	ΔR^2	p
1	Length	.828		< .001	.622		< .001
	GPT	−.030		.514	−.063		.333
			.681	< .001		.382	< .001
2	S1	.203	.041	.013	.122	.015	.137
3	SA	.138	.019	.093	−.007	.000	.931
3	B1	−.118	.014	.151	.017	.000	.835
3	BA	−.222	.049	.006	.118	.014	.151
2	SA	.138	019	.092	−005	.000	.951
3	S1	.202	.041	.013	.122	.015	.136
3	B1	−.033	.001	.692	.059	.004	.471
3	BA	−.282	.079	.000	.094	.009	.252
2	B1	−.044	.002	.590	.060	.004	.467
3	S1	.218	.048	.007	.101	.010	.217
3	SA	.134	.018	.101	.000	.000	.999
3	BA	−.256	.066	.002	.093	.009	.260
2	BA	−.257	.066	.002	.093	.009	.256
3	S1	.158	.025	.053	.141	.020	.086
3	SA	.186	.034	.023	−.020	.000	.806
3	B1	−.042	.002	.613	.059	.003	.477

Note: DV = dependent variable; Frequency as log number of occurrences; Length in letters; BA = mean log bigram frequency (all bigrams); SA = mean log syllable frequency (all syllables); B1 = log frequency of the initial bigram; S1 = log frequency of the initial syllable; ON = orthographic neighborhood size; GPT = grapho-phonemic transparency, as mean log token sonograph probability; DFs for step 1 = 147; DFs for steps 2-3 = 148.

Table 6

Multiple regression analysis predicting human and DRC latencies for words.

		Dependent variable					
		Naming RT (humans)			Naming RT (DRC)		
Step	Variable	β	ΔR^2	p	β	ΔR^2	p
1	Frequency	−.420		< .001	−.649		< .001
	Length	.446		< .001	.029		.634
	ON	−.024		.712	.100		.106
	GPT	.022		.740	−.172		.006
			.399	< .001		.454	< .001
2	S1	.142	.020	.084	.010	.000	.902
		Lexical Decision RT (humans)			Lexical Decision RT (DRC)		
Step	Variable	β	ΔR^2	p	β	ΔR^2	p
1	Frequency	−.502		< .001	−0.983		< .001
	Length	.223		.001	−0.014		.382
	ON	−.105		.124	.0265		.096
	GPT	.044		.519	−0.010		.516
			.333	< .001		.964	< .001
2	S1	.244	.060	.003	.093	.009	.257

Note: Frequency as log number of occurrences; Length in letters; S1 = log frequency of the initial syllable; ON = orthographic neighborhood size; GPT = grapho-phonemic transparency, as mean log token sonograph probability; DFs for step 1 = 145; DFs for step 2 = 148.

Table 7

Multiple regression analysis predicting human and DRC latencies for nonwords

		Dependent variable					
		Naming RT (humans)			Naming RT (DRC)		
Step	Variable	*β*	*ΔR²*	*p*	*β*	*ΔR²*	*p*
1	Length	.828		< .001	.973		< .001
	GPT	−.030		.514	.080		< .001
			.686	< .001		.954	< .001
2	S1	.159		.050	−.156		.060
	BA	−.226		.005	−.179		.031
			.090	.001		.046	.032
		Lexical Decision RT (humans)			Lexical Decision RT (DRC)		
Step	Variable	*β*	*ΔR²*	*p*	*β*	*ΔR²*	*p*
1	Length	.622		.001	*No variation*		
	GPT	−.063		.333			
			.391	< .001			
2	S1	.146		.081			
	BA	.122		.144			
			.029	.114			

Note: Frequency as log number of occurrences; Length in letters; BA = mean log bigram frequency (all bigrams); S1 = log frequency of the initial syllable; GPT = grapho-phonemic transparency, as mean log token sonograph probability; All DFs = 147.

Table 8

Direct comparison of effects between behavioral and model data

	Words			
	Naming task		Lexical decision	
	Humans	DRC	Humans	DRC
Frequency	yes	yes	yes	yes
Length	yes	no	no	no
ON	no	no	no	marginally
GPT	no	yes	yes	no
S1	marginally	no	no	no
	Nonwords			
	Naming task		Lexical decision	
	Humans	DRC	Humans	DRC
Frequency	-	-	-	-
Length	yes	yes	yes	-
ON	-	-	-	-
GPT	no	yes	yes	-
S1	yes	no	no	-
BA	yes	yes	yes	-

Note: Frequency as log number of occurrences; Length in letters; BA = mean log bigram frequency (all bigrams); S1 = log frequency of the initial syllable; ON = orthographic neighborhood size; GPT = grapho-phonemic transparency, as mean log token sonograph probability.