Exploring word recognition with selected stimuli: The case for decorrelated parameters

Athanassios Protopapas (aprotopapas@phs.uoa.gr)
Department of Philosophy & History of Science (MITHE), Ano Ilissia University Campus
GR-15771 Zografos, Greece

Efthymia C. Kapnoula (ekapnoula@iowa.uiowa.edu)
Department of Psychology, E11 Seashore Hall
Iowa City, IA 52242-1409 USA

Abstract

We report a study of naming and lexical decision with 132 adult Greek speakers responding to 150 words and matched pseudowords with decorrelated frequency, length, neighborhood, syllable and bigram frequency, and transparency. This approach allowed us to individuate and accurately estimate the effects of each variable, and to assess their linearity and additivity. Significant effects of frequency, length, and syllable frequency were revealed, as well as several interactions. The results are informative for cognitive modeling of visual word recognition in more transparent orthographies.

Keywords: Visual word recognition; pseudowords; naming; lexical decision; mixed-effects models; Greek.

Models of visual word recognition posit distinct mechanisms and representations involved in the processing of orthographic stimuli. The implications of these hypotheses for response times (RT) to individual words and pseudowords are typically studied using naming and lexical decision tasks. A productive line of research concerns the effects of lexical and sublexical variables on RT distributions. In evaluating models and their properties, several variables have been examined in this light and have been found to be related to processing times, such as frequency, length, neighborhood, bigram and syllable frequency, and more (Balota, Yap, & Cortese, 2006).

A common approach to such studies has been to form groups of stimuli differing in a parameter of interest, such as a low-frequency and a high-frequency group. Interactions are then examined by crossing levels of variables in subgroups. This has led to the identification of important effects but has recently been criticized for problems stemming from selection of atypical items and restricted parameter ranges (Balota, Yap, Hutchison, & Cortese, 2012). More recent approaches have abandoned stimulus groupings in favor of regression approaches, in which wide value ranges of several variables are simultaneously entered in multivariate analyses (Yap & Balota, 2009). Advances in statistical modeling have allowed more parameters to be examined (Baayen, 2008).

This approach has culminated in the "mega study" efforts, in which huge numbers of stimuli are presented to volunteer samples of unprecedented sizes. For example, the English Lexicon Project (Balota et al., 2007) includes about 40k words and pseudowords responded to by 1,260 participants. This allows examination of the full ranges of all parameters and their combinations. Special subsets may be selected for targeted analysis, so this method allows replication of previous studies and examination of potential biases or dependencies. Moreover, these data are available for future studies on as-yet unidentified parameters without further data collection. Thus mega-studies constitute a large step forward in our quest for understanding word processing.

The multiple regression approach is also vulnerable to criticism. Regression models typically include a linear effect for each variable, assuming that this effectively "partials out" all influence of the corresponding parameter, allowing the effects of other parameters to be accurately estimated. This is only the case when all effects are linear and independent. If not, the linear modeling removes only part of an effect, conflating the remainder with other correlated variables. The problems of linearity and additivity are compounded and cannot be addressed in stimulus sets with correlated effects because it cannot be known whether a variable has a true curvilinear effect or an interaction with a correlated variable. In both cases departures from linearity will be identified but the decision to "remove" one of the two will affect the other and may do so in a nonrepresentative or nonoptimal manner. Moreover, in the absence of curvilinearity and nonadditivity, correlated parameters necessarily lead to underestimated effects because shared variance is removed when one parameter is controlled.

Linearity is potentially informative as cognitive processing models can make specific predictions about the shape of the relationship between aspects of the stimuli and the time required to respond to them. Thus it is useful to establish the fitting functions on the actual RT distributions and incorporate them in modeling (cf. Balota & Yap, 2011). Nonadditivity is also of great interest because it has implications for theoretical approaches insofar as the cause of each interaction must be understood within the context of relevant assumptions.

In the present study we begin to address these issues by examining RTs to a set of words and pseudowords in which variables typically examined in word recognition were uncorrelated. We work in Greek, a language with higher orthographic transparency and word length than English (Protopapas & Vlahou, 2009), aiming to contribute to the cross-linguistic effort. We selected words from a corpus ensuring that there was no significant correlation among frequency, length, neighborhood, syllable and bigram frequency, and graphophonemic transparency. A set of pseudowords was then created, similarly uncorrelated, matching the word group in these variables. To the extent a reasonably wide range was sampled, this approach permits isolation of the effects of each variable and identification of individual interactions and nonlinearities. All items were presented in a naming and a lexical decision task. To maximize reliability and allow detection of small effects, we employed a relatively large sample, based on Rey, Courrieu, Schmidt-Weigand, and Jacobs (2009).

Method

Participants

The sample included 97 women and 35 men, native speakers of Greek, 18–36 years old ($M = 23.3$, $SD = 4.7$), mostly students (12–21 years of education, $M = 15.4$, $SD = 2.1$). Fourteen were left-handed.

Stimuli

A set of 150 words were selected from the IPLR word list (Protopapas, Tzakosta, Chalamandaris, & Tsiakoulis, 2012), 2–5 syllables long (4–10 letters; 4–11 phonemes). In an iterative selection process, a set of properties were retrieved along with each word, including log frequency of occurrence; number of letters, phonemes, and syllables; orthographic and phonological neighborhood (Coltheart's N); orthographic and phonological syllable frequency; letter and phoneme bigram frequency; and a nondirectional measure of graphophonemic transparency (log mean token "sonograph" probability; Spencer, 2009). A nonparametric index of association (Spearman's ρ) among all variables was calculated. The process terminated when groups of qualitatively distinct variables were not significantly correlated. The following variables were retained as most relevant for the analyses reported below: log frequency, number of letters, mean syllable and bigram frequency, and transparency. Figure 1 (top row) shows the distribution of each variable in the selected sets against the overall type and token distribution in the corpus.

A set of 150 pseudowords were constructed to resemble the words in basic phonological and orthographic structure and letter and phoneme distribution. The pseudowords were indistinguishable from the words in the target variables, as verified by the Kolmogorov-Smirnov test for equality of distributions. The results of these tests are listed in Table 1, along with the correlations among the variables for both sets.

A difficulty was encountered in pseudoword neighborhoods. Matching to the words required including items with many neighbors. Neighborhoods for long words are mainly due to grammatical inflection. Long words typically have few or no unrelated neighbors, but due to the rich inflectional system of Greek all content words have neighboring inflectional variants. Neighborhoods for pseudowords would therefore be limited to the inflectional families of word neighbors, so that some pseudowords would be strongly influenced by one lexical lemma. This was undesirable because pseudowords are known to activate lexical neighbors strongly (even assimilating their stress pattern; Protopapas, Gerakaki, & Alexandri, 2007). Therefore, using pseudowords with neighbors would result in a pseudoword set in which word activation might play a prominent role and suppress nonlexical effects. Thus we decided to minimize pseudoword neighborhoods and keep the pseudowords distinct from specific words, at the cost of matching and correlations involving neighborhoods.

Procedure

A naming and a lexical decision task were implemented in DMDX (Forster & Forster, 2003). In both tasks, each item was presented in Arial 36-pt white font at the center of a laptop 15.4″ screen for 1.9 s. A few practice and warm-up items preceded the experimental stimuli. A short break was offered halfway through each block of 150 stimuli. For lexical decision, participants responded by pressing the left and right control keys. Words and pseudowords were intermixed randomly. The "word" response was set to the participant's preferred or nonpreferred hand, approximately counterbalanced. For naming, words and pseudowords were presented in separate blocks, in counterbalanced order between participants. Responses were recorded and onset times were subsequently verified using CheckVocal (Protopapas, 2007). The order of naming and lexical decision tasks was counterbalanced. In both tasks, items were presented in a different random order for each participant. A distractor task (digit span) was administered between the two tasks to minimize carryover effects.

Results

Raw response times were logarithmically transformed and analyzed using mixed-effects models with crossed random effects for participants and items, separately for words and nonwords. Analyses were conducted using package lme4 (Bates, Maechler, & Bolker, 2012) in R (R Core Team, 2012), mainly following Baayen (2008). All variables were centered. Models reported below included fixed effects and random slopes per participant for the linear effects of trial number and preceding RT, as well as random intercepts for participants and for items. These "baseline" effects are not discussed further.

For each task and stimulus type we examined the following: (a) A "full" model, with the complete set of variables, i.e., baseline effects plus all six experimental parameters (linear effects only, not interacting). This was used to estimate the "full" effects of each parameter. (b) For each parameter, a variant of the full model excluding that parameter. The item random intercepts of this model were used to estimate the "residual" effects of each parameter. (c) A baseline model with only the baseline effects. (d) For each parameter, a variant of the baseline model including only that parameter. This was used to estimate the "single" effects of each parameter. Comparison of this model to the baseline (via likelihood ratio) determined the significance of each parameter. (e) An "augmented" model, in which quadratic effects and interactions were added to the full model in a forward-backward procedure and retained when significant (determined via likelihood ratio at $p < .05$). The linear effects of all six parameters were retained whether or not they made significant contributions, and are listed in Table 1 (4 rightmost columns).

Table 1: Correlation coefficients (Spearman's ρ) among the 6 experimental parameters, for words (above the diagonal) and pseudowords (below the diagonal), and Kolmogorov-Smirnov tests for equality of distributions between words and pseudowords. The four rightmost columns show the estimated $\hat{\beta}$ for the corresponding linear effect in the augmented model (see text).

		Correlations					K-S test		Naming		Lexical decision	
	Variable	2	3	4	5	6	D	p	Words	Pseudo	Words	Pseudo
1	frequency	$-.05$	.00	$-.02$	.10	.00			$-.0152^*$		$-.0311^*$	
2	N letters		$-.01$	$-.05$	.01	$-.11$	.06	.950	$+.0163^*$	$+.0306^*$	$+.0151^*$	$+.0361^*$
3	neighborhood	$-.50^\dagger$		$-.08$	.07	$-.02$	.70	.000	$-.0033$	$-.0668^*$	$-.0092^*$	$+.0409^*$
4	syllable freq.	$-.08$	.07		.09	$-.09$	.15	.079	$+.0032^*$	$+.0022^*$	$+.0044^*$	$+.0006$
5	bigram freq.	.08	$-.04$	.06		.06	.09	.531	$-.0039$	$-.2131^*$	$-.2578^*$	$+.1335$
6	transparency	.00	$-.03$	$-.12$	.05		.05	.983	$+.0000$	$-.0005$	$+.0018^*$	$-.0010$

Note: $^\dagger p < .0005$; for all other correlations, $p > .1$; $^*|t| \geq 2.0$

The variance of item random intercepts varied between tasks. In the baseline models, it was 4.98 for word naming, 16.03 for pseudoword naming, 9.99 for word lexical decision, and 7.90 for pseudoword lexical decision (all $\times 10^{-3}$). This sets an upper limit for the contributions of the experimental parameters, which are all item-related, indicating that there is more variance to be accounted for in pseudoword naming (the highest) than in word naming (the lowest). In comparison, the residual (error) variance of the baseline models were 13.1, 19.3, 37.9, and 36.8, respectively, suggesting that lexical decision tasks are "noisier" than naming tasks.

Linear individual effects The results for these analyses are summarized graphically in the four bottom rows of panels in Figure 1, in which the linear effect of each parameter is tested when all other parameters were in the model (residual vs. full model) and when no other parameters were in the model (single vs. baseline model). The two estimates were generally within one standard error of each other, indicating very close correspondence of the two types of analysis in estimating individual linear effects. The reduction in item variance (random intercept for items) by inclusion of each parameter was also very similar between the two approaches. There were some differences in significance but they concerned low variance proportions (1%) or were associated with pseudoword orthographic neighborhood, which was not decorrelated.

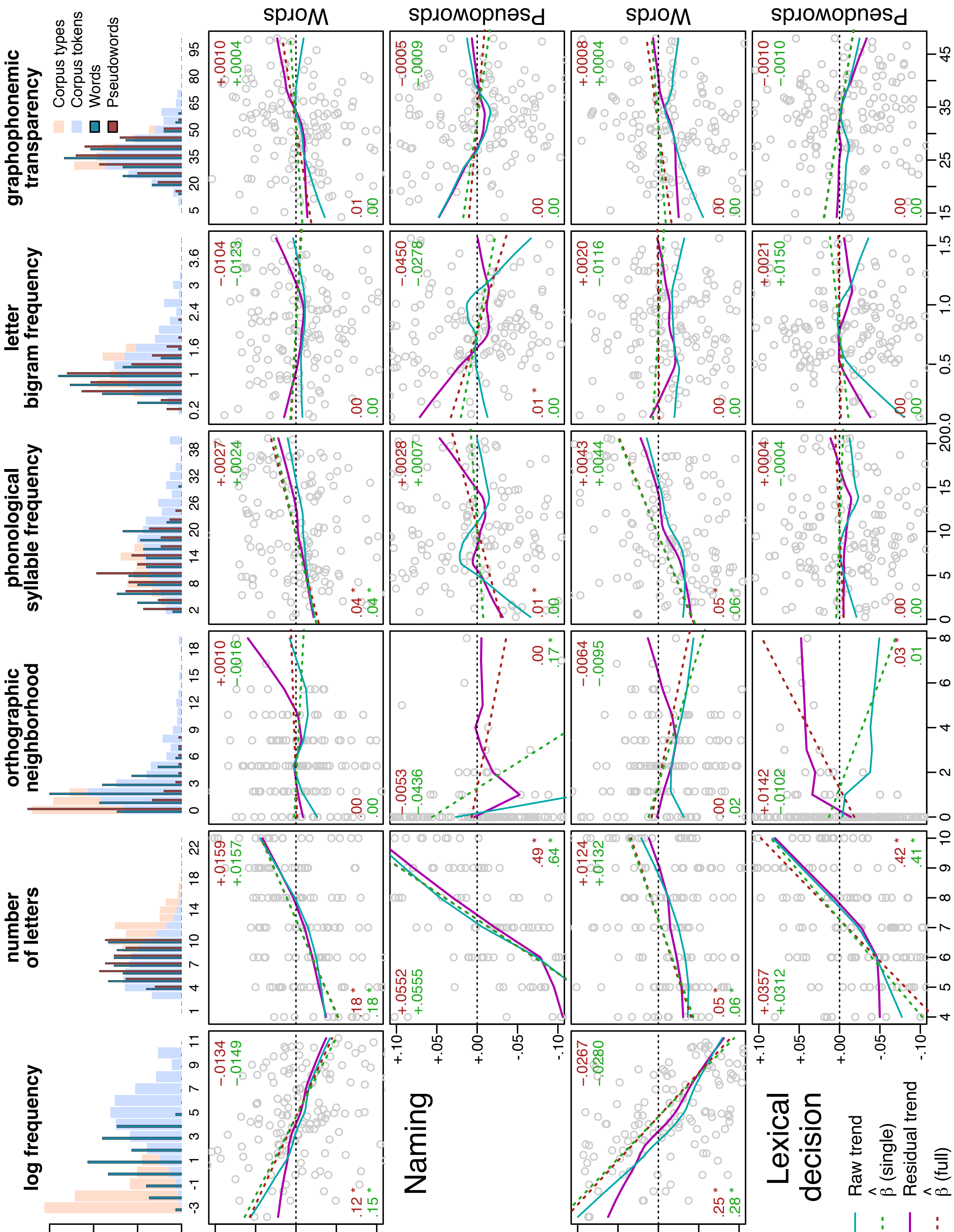
Quadratic effects Examination of the raw and residual trends in Figure 1 indicated monotonic and largely linear effects of frequency, length, and syllabic frequency, especially for words. Some curvilinearity was apparent for other parameters, particularly for pseudowords. Nonlinearities were examined for every parameter in each case by testing quadratic terms added (centered) to the full model. The quadratic effect of bigram frequency was significant in pseudoword naming ($\hat{\beta} = .1024$, $SE = .0435$), word lexical decision ($\hat{\beta} = .1720$, $SE = .0854$), and pseudoword lexical decision ($\hat{\beta} = -.0778$, $SE = .0412$). The quadratic effect of orthographic neighborhood was significant in pseudoword lexical decision ($\hat{\beta} = -.0043$, $SE = .0021$) but seems spurious in light of the severely skewed distribution of this parameter. No other quadratic terms were found to be significant. No higher-order terms or other nonlinear functions were tested.

Interactions Two-way interactions were examined by testing each pair of parameters. In word naming, there was an interaction between frequency and neighborhood ($\hat{\beta} = .0054$, $SE = .0016$). In addition, there were interactions of transparency with frequency ($\hat{\beta} = .0007$, $SE = .0003$), neighborhood ($\hat{\beta} = -.0007$, $SE = .0003$), and syllabic frequency ($\hat{\beta} = .0002$, $SE = .0001$). In pseudoword naming there was an interaction of length and neighborhood ($\hat{\beta} = -.0201$, $SE = .0068$). In word lexical decision there were interactions of neighborhood with frequency ($\hat{\beta} = .0070$, $SE = .0023$) and with transparency ($\hat{\beta} = -.0009$, $SE = .0005$). Finally, in pseudoword lexical decision there was a marginal interaction of syllabic frequency with bigram frequency ($\hat{\beta} = .0050$, $SE = .0029$). There were no interactions involving the quadratic terms. No higher-order interactions were tested.

Following the addition of the aforementioned quadratic and interactive effects, the augmented models reduced the item variance (random intercepts) by half or more. Specifically, the proportion of item variance in the baseline model that was accounted for by the six experimental parameters was .48 for word naming, .71 for pseudoword naming, .48 for word lexical decision, and .50 for pseudoword lexical decision. Additional proportions in naming could be accounted for by modeling initial phoneme classes but there was no need for that in the present approach as the respective onset effects were adequately captured in the by-item random intercepts.

Discussion

We employed a stimulus selection procedure to create a set of words and matched pseudowords with decorrelated parameters, aiming to examine curvilinear and interactive effects more accurately than with blind multiple regression. In this study we first confirmed that the linear effects are similar when estimated in full models vs. single-parameter models. Differences emerged for unmatched variables, as should be expected. Therefore this method achieves isolation of the effects of basic parameters, allowing further use in sit-



log frequency
number of letters
orthographic neighborhood
phonological syllable frequency
letter bigram frequency
graphophonemic transparency
Corpus types
Corpus tokens
Words
Pseudowords
Naming
Lexical decision
Raw trend
β̂ (single)
Residual trend
β̂ (full)

uations where complex modeling may be impractical or impossible, such as in fMRI. Use of decorrelated parameters to examine brain modulation in response to written stimuli has been previously reported for 465 monosyllabic English words (Graves, Desai, Humphries, Seidenberg, & Binder, 2010). Our analysis supports this approach and extends it to multisyllabic words and a more transparent orthography.

Our results do not strongly challenge the common assumption of linearity, as most of the effects seem well approximated by a linear function. However, our analyses were based on log RT and not raw times. If linear fits on log RT can pass more stringent tests in comparison with a richer set of curvilinear alternatives, the implications for modeling are that models should predict logarithmic RT curves. The existing analysis techniques allow the field to progress from prediction of differences between conditions, or the mere existence of associations among parameters, toward more specific predictions of the relations between participants, items, and measures, as for example in the rate-amount (Faust, Balota, Spieler, & Ferraro, 1999) and difference engine (Myerson, Hale, Zheng, Jenkins, & Widaman, 2003) approaches. Simple correlation between predicted and measured times may become inadequate as more specific derivations and variance comparisons become increasingly feasible.

It remains to be established whether the effects uncovered in this analysis are properly accounted for in linear models of log RT. The effect of frequency, in particular, seems to level off somewhat towards lower frequencies. Although this may be an artifact of nonhomogeneous sampling affecting the lower end, it is consistent with a frequency effect less steep than logarithmic. Given that frequency has been log transformed, it may be fruitful to examine alternatives (e.g., power functions, ranks, or subjective estimates of familiarity) in accounting for the frequency effect (cf. Balota et al., 2012). It is reassuring that effect estimates and variance proportions in lexical decision were substantially greater (double) than in naming, consistent with the notion of frequency as a lexical-semantic rather than surface variable.

The large effect of word length may seem surprising but it should be taken into account that multisyllabic words up to 10 letters long were involved. The Greek orthography is also relatively transparent for reading (Protopapas & Vlahou, 2009), conceivably supporting more serial approaches than English. This may explain why about half of the item variance in pseudowords was accounted for by length alone, almost as much in lexical decision as in naming. An interesting aspect of the data concerns the low-end shape of this relation, evident in pseudowords, although there may also be some flattening of the word curves. This may be related to the U-shape reported in other languages (see recent discussion in Yap & Balota, 2009) and warrants further investigation.

No significant main effects of orthographic neighborhood were revealed in our analyses. This is surprising in light of consistent reports in the literature regarding neighborhood effects. However, there are issues with Greek word neighborhoods that warrant further scrutiny. Due to extensive inflection of nouns, verbs, and adjectives, many items counted as neighbors are inflectional variants, arguably linked to a single lexical lemma (contingent on one's theory of morphological representation in the lexicon). Moreover, the number of neighbors diminishes rapidly with word length, as there are fewer instances of words in longer letter-string space. This suggests that the emphasis on neighborhood effects may have resulted from an artifact of English being the most-studied language and allowing investigation restricted to short, single-syllable words. Alternatively, more flexible indices of orthographic distance may be required to express neighborhood density (e.g., Yarkoni, Balota, & Yap, 2008).

In agreement with recent reports for other languages (Conrad, Tamm, Carreiras, & Jacobs, 2010), inhibitory phonological syllabic frequency effects were found for words in both naming and lexical decision. A smaller but significant effect was found in pseudoword naming. No comparable effect was observed with orthographic syllable frequency (not reported above), consistent with the source of such effects lying within a phonological sublexical space. In contrast to syllables, orthographic bigram effects were minor, mainly restricted to pseudowords, and partly facilitatory, with a quadratic component resisting interpretation. There were no effects of phonological bigram frequency (i.e., phoneme pairs), consistent with an explanation for bigram effects related to orthographic familiarity with letter clusters.

Figure 1: *(on previous page)* The top row shows the distribution of parameter values for the stimulus set, separately for words (blue) and pseudowords (red), in comparison to all corpus types (light peach) and tokens (light blue). Bars display proportions of items, adding up to 1.0. The other rows display the effects of each of the six experimental parameters on naming (Rows 2 and 3) and on lexical decision (Rows 4 and 5). Each box displays residual item effects (grey circles) in a model including baseline effects and all parameters except one. The red solid line plots a smoothed average (via function lowess) of these points. The dotted red line shows the effect estimate for this parameter when added to the predictor set, resulting in a full model. The teal solid line plots a smoothed average of the centered raw item means. The dotted green line shows the effect estimate of the parameter when included in a model with baseline and random effects only, absent all other parameters. The red and green numbers at the top of each panel are the corresponding effect estimates ($\hat{\beta}$) for the same-color dotted lines, whereas the numbers at the bottom of each panel are the proportions of item variance accounted for by this fixed effect; an asterisk denotes significant contribution (by likelihood ratio test). The vertical axis is scaled in log milliseconds (with respect to the grand mean intercept). Note different scaling of horizontal axes between parameters and also between distributions (top row) and effects panels.

There were no significant effects of consistency except in the augmented model for word lexical decision. This stands in contrast to large effects of regularity and consistency reported for English and may be attributed to the greater transparency of Greek. Nevertheless, there were consistent interactions involving transparency in word naming and lexical decision. Higher-probability grapheme-phoneme tokens were associated with increased RTs, an effect enhanced for higher-frequency words and diminished in larger neighborhoods. Instead of a pure consistency effect whereby more frequent mappings are decoded more rapidly, here we may have a situation in which systematic mappings permit greater confusion among lexical candidates. The fact that this only occurred for words—the (nonsignificant) trends for pseudowords being negative—suggests that it may be related to orthographic N being a poor index of neighborhood effects.

It should be kept in mind that our findings for pseudowords must be interpreted with caution as the stimulus set was not fully controlled and decorrelated due to the aforementioned neighborhood issue. This is not a major limitation of the study because most cognitive models typically focus on words and do not emphasize pseudoword processing.

Overall, the variance accounted for by our parameters was substantial but far from the 80% estimate Rey et al. (2009) gave for reproducible item variance in samples of this size. Although inclusion of initial phoneme class raised this proportion considerably, 80% was achieved only for pseudoword naming. Word naming lagged behind at 63%, indicating that major sources of systematic item variance remain to be brought into the models (Adelman, Marquis, Sabatos-DeVito, & Estes, in press). Morphological and semantic variables are obvious candidates to be examined in follow-up studies.

Acknowledgments

Supported by the European Union (European Social Fund) and Greek national funds through the National Strategic Reference Framework – Research Funding Program THALIS-UOA-COGMEK: Cognitive mechanisms in the perception, representation, and organization of knowledge.

References

Adelman, J. S., Marquis, S. J., Sabatos-DeVito, M. G., & Estes, Z. (in press). The unexplained nature of reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

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

Balota, D., & Yap, M. (2011). Moving beyond the mean in studies of mental chronometry: The power of response time distributional analyses. *Current Directions in Psychological Science*, *20*(3), 160–166.

Balota, D., Yap, M., & Cortese, M. (2006). Visual word recognition: The journey from features to meaning (a travel update). In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 285–375). Amsterdam, The Netherlands: Academic Press.

Balota, D., Yap, M., Hutchison, K., & Cortese, M. (2012). Megastudies: What do millions (or so) of trials tell us about lexical processing? In J. S. Adelman (Ed.), *Visual word recognition* (Vol. 1, pp. 90–115). Psychology Press.

Balota, D., Yap, M., Hutchison, K., Cortese, M., Kessler, B., Loftis, B., et al. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445–459.

Bates, D., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes.* (R package version 0.999999-0)

Conrad, M., Tamm, S., Carreiras, M., & Jacobs, A. (2010). Simulating syllable frequency effects within an interactive activation framework. *European Journal of Cognitive Psychology*, *22*(5), 861–893.

Faust, M., Balota, D., Spieler, D., & Ferraro, F. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin*, *125*(6), 777–799.

Forster, K., & Forster, J. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods*, *35*(1), 116–124.

Graves, W., Desai, R., Humphries, C., Seidenberg, M., & Binder, J. (2010). Neural systems for reading aloud: A multiparametric approach. *Cerebral Cortex*, *20*, 1799–15.

Myerson, J., Hale, S., Zheng, Y., Jenkins, L., & Widaman, K. (2003). The difference engine: A model of diversity in cognition. *Psychonomic Bulletin & Review*, *10*, 262–288.

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., Gerakaki, S., & Alexandri, S. (2007). Sources of information for stress assignment in reading Greek. *Applied Psycholinguistics*, *28*(4), 695.

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*(4), 991–1008.

R Core Team. (2012). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rey, A., Courrieu, P., Schmidt-Weigand, F., & Jacobs, A. (2009). Item performance in visual word recognition. *Psychonomic Bulletin & Review*, *16*(3), 600–608.

Spencer, K. (2009). Feedforward, -backward, and neutral transparency measures for British English. *Behavior Research Methods*, *41*(1), 220–227.

Yap, M., & Balota, D. (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 Colthearts N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*(5), 971–979.