Behavior Research Methods
2010, 42 (1), 96-108
doi:10.3758/BRM.42.1.96

Feedforward, -backward, and neutral transparency measures for Grade 1 and Grade 2 Greek readers

Kenneth A. Spencer
University of Hull, Hull, England

Nataly Loizidou
Frederick University, Cyprus, Greece

and

Jackie Masterson
University of London, London, England

In the present study, orthographic metrics for Greek children's Grade 1 and Grade 2 reading materials were presented. Data for five transparency metrics—three of which being neither feedforward nor feedbackward—were presented and offered for use in the research of children's reading and spelling acquisition. The analysis demonstrated the complex relationships between metrics and compared the results with those obtained for the English language. The structure of these metrics from a variety of corpus sizes was investigated, and we concluded that large corpus sizes do not necessarily make a substantial contribution to the value of such metrics when compared with smaller samples.

The regularity of orthographies varies, with those that are regular classed as "transparent" and those with irregular correspondences classed as "opaque." Finnish and Turkish are symmetrically transparent for both reading and spelling, whereas Greek and German are asymmetric, being less transparent for spelling than for reading, with letters that have clearly defined pronunciations but with phonemes that have alternative spellings. It has been proposed (Katz & Frost, 1992) that the variation between orthographies leads to differences in processing and, consequently, that transparent orthographies make literacy skills easier to learn (Oney & Goldman, 1984; L. H. Spencer & Hanley, 2003). Accuracy levels are lower and reading speed is slower for opaque languages for both normal and dyslexic children (Cossu, Gugliotta, & Marshall, 1995; Ellis et al., 2004; Frith, Wimmer, & Landerl, 1998; Landerl, Wimmer, & Frith, 1997).

Detailed linguistic analyses have measured the orthographic body/phonological rime (e.g., as in *seen*, the orthographic body <een> has the phonological rime /i:n/) transparency of English (Ziegler, Stone, & Jacobs, 1997) and of French (Ziegler, Jacobs, & Stone, 1996) for both reading and spelling, and have allowed comparisons of relative transparency. For example, when compared with English, French is 20% more consistent for reading, but is 10% less so for spelling. This large-grain, body/rime level of analysis is central to the connectionist or parallel distributed processing network model of reading (Plaut, McClelland, Seidenberg, & Patterson, 1996), in which the ease of pronunciation depends on the relative orthographic body transparency of a word. Words having body letter patterns with the same phonological rime that are always pronounced in the same way are classed as "consistent." For words classed as "inconsistent," the less typical the pronunciation, the greater the word reading difficulty. Treiman, Mullennix, Bijeljac-Babic, and Richmond-Welty (1995) saw dichotomous classification (consistent/inconsistent) as inadequate and claimed that only continuous representations of consistency allow detailed examination of transparency effects. Plaut (1999) suggested that the language mechanism gradually picks up on this continuous statistical structure among written and spoken words.

Languages may also be classified at the fine-grained grapheme–phoneme level for reading and spelling. English has been studied extensively at this level (Berndt, Reggia, & Mitchum, 1987; Carney, 1994; Gontijo, Gontijo, & Shillcock, 2003; Hanna, Hanna, & Hodges, 1966; Venezky, 1967). The present study applied this approach to the Greek language.

Variations of fine-grained word transparency for English formed an integral part of the early serial dual-route

K. A. Spencer, k.a.spencer@hull.ac.uk

Table 1
Metric Calculations for Sonograph /i/–<ει>

	Frequency	Total	Probability
Sonograph	437	27,585	.0158
Grapheme	447	27,585	.0162
Phoneme	2,412	27,585	.0874
PGC	437	2,412	.1812
GPC	437	447	.9776

Note—PGC, phoneme–grapheme correspondence; GPC, grapheme–phoneme correspondence.

models of skilled reading (Coltheart, Curtis, Atkins, & Haller, 1993). Words were divided into graphemes and their associated phonemes, which were termed *grapheme–phoneme correspondences* (GPCs) when associated with reading, and *phoneme–grapheme correspondences* (PGCs) when associated with spelling. Separate GPC and PGC conditional probabilities could be calculated from the same sound–letter components. More recently, it has been conceded that algorithm-derived GPC rules do not work for the DRC model, which now features hardwired GPC rules (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). There is also an explicit admission that the DRC model does not offer any account of how reading is learned (Coltheart, 2006). This has led to the development of the connectionist dual process (CDP) model, which is claimed to be superior because it is highly sensitive to the graded statistical consistency of spelling–sound relationships at multiple grain sizes (from letters to word bodies), and because it has a stronger developmental strand than do alternative computer models (Perry, Ziegler, & Zorzi, 2007).

Different sources and corpus sizes have been used when calculating word metrics. Large polysyllabic adult corpora of 20,000 and 17,310 words were used by Venezky (1967) and Hanna et al. (1966), whereas Seidenberg and McClelland (1989) and Coltheart et al. (1993) used a corpus of only 2,897 monosyllabic words. McGuinness (1998) argued that spelling patterns for smaller corpora of more common words would differ from more extensive analyses, but small corpora have been shown to have substantially the same fine-grained structure as do larger corpora for English (K. A. Spencer, 2009). This offers researchers the opportunity to prepare metrics that are closely associated with experimental conditions, such as metrics derived from specific children's texts rather than from general adult corpora, for experiments with children as subjects. K. A. Spencer (2007) demonstrated that in multiple regression analyses, frequency metrics derived from such a children's corpus (Children's Printed Word Database; Masterson, Stuart, Dixon, & Lovejoy, 2003) predicted substantially more individual variance in a spelling task, for school years 2–6, than did values derived from larger adult corpora. However, smaller corpora do underestimate the number of grapheme–phoneme alignments (sonographs, see below), although the underestimated sonographs fall in the long tail of very low frequencies (probabilities $< .01$) for English.

Before reading or spelling conditional probabilities can be calculated, the data on which they are based exist as a directionless association between graphemes and pho-

Table 2
Correlations Among Five Word Metrics for Corpus Sizes 0.21K–4.10K

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. Corpus 4.10K: G	–	1.00**	.98**	.96**	−.22*	−.24*	−.26*	−.25*	.38**	.39**	.41**	.41**	.68**	.68**	.64**	.65**	.30**	.29**	.18	.19
2. Corpus 1.66K: G		–	.99**	.97**	−.23*	−.24*	−.26*	−.25*	.38**	.39**	.43**	.42**	.69**	.69**	.66**	.66**	.30**	.30**	.19	.19
3. Corpus 0.70K: G			–	.97**	−.19	−.21	−.24*	−.23	.39**	.41**	.44**	.43**	.68**	.68**	.69**	.69**	.28*	.27*	.20	.20
4. Corpus 0.21K: G				–	−.15	−.16	−.17	−.18	.43**	.45**	.50**	.50**	.68**	.69**	.70**	.74**	.18	.18	.16	.17
5. Corpus 4.10K: GP					–	.99**	.96**	.94**	.27**	.23**	.24*	.11	.36**	.35**	.41**	.39**	.40**	.41**	.45**	.43**
6. Corpus 1.66K: GP						–	.98**	.95**	.24*	.21*	.20	.10	.35**	.35**	.40**	.39**	.41**	.42**	.47**	.43**
7. Corpus 0.70K: GP							–	.96**	.21	.18	.16	.09	.37**	.37**	.38**	.39**	.47**	.48**	.51**	.43**
8. Corpus 0.21K: GP								–	.10	.07	.05	.05	.37**	.38**	.38**	.39**	.47**	.46**	.48**	.42**
9. Corpus 4.10K: P									–	1.00**	.97**	.91**	.54**	.53**	.54**	.54**	−.06	−.09	−.15	−.28*
10. Corpus 1.66K: P										–	.99**	.94**	.52**	.52**	.53**	.53**	−.11	−.11	−.18	−.32*
11. Corpus 0.70K: P											–	.97**	.53**	.53**	.54**	.55**	−.12	−.14	−.20	−.35**
12. Corpus 0.21K: P												–	.47**	.48**	.49**	.55**	−.30*	−.31*	−.33**	−.39**
13. Corpus 4.10K: SG													–	1.00**	.98**	.95**	.59**	.59**	.53**	.47**
14. Corpus 1.66K: SG														–	.99**	.96**	.60**	.60**	.54**	.47**
15. Corpus 0.70K: SG															–	.97**	.59**	.59**	.53**	.46**
16. Corpus 0.21K: SG																–	.47**	.47**	.45**	.41**
17. Corpus 4.10K: PG																	–	.99**	.93**	.94**
18. Corpus 1.66K: PG																		–	.96**	.95**
19. Corpus 0.70K: PG																			–	.96**
20. Corpus 0.21K: PG																				–

Note—G, grapheme probability of occurrence; P, phoneme probability of occurrence; GP, grapheme–phoneme correspondence (reading) probability; PG, phoneme–grapheme correspondence (spelling) probability; SG, sonograph probability of occurrence. $^{*}p < .05$. $^{**}p < .01$.

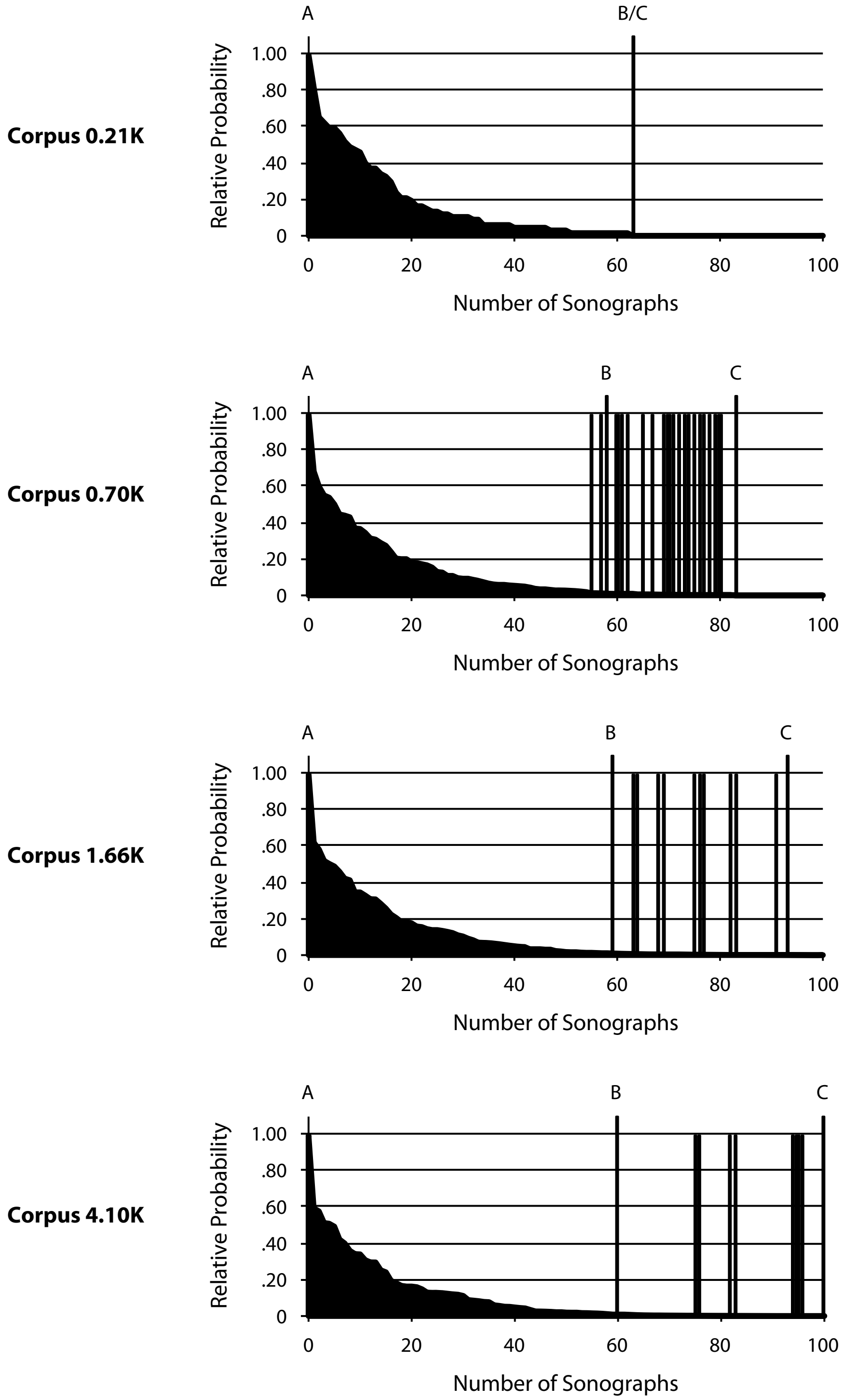

Figure 1. Sonograph probability profiles for four corpus sizes. For comparative purposes, the relative probability values are expressed as a proportion of the highest frequency sonograph. Position B indicates relative probabilities <.01. Position C indicates the total number of sonographs for the corpus. Bars in area BC indicate new sonographs for the corpus as compared with the smaller corpus above it.

nemes, each of which has a type frequency in the corpus from which it is derived. It is this type frequency that is used to calculate both GPC and PGC probability values. K. A. Spencer (2009) used the term *sonograph* to describe this directionless item and demonstrated how five basic metrics for the fine-grained level of analysis may be calculated from it. This approach has been adopted for the present analysis of the Greek language. Table 1 illustrates the cal-

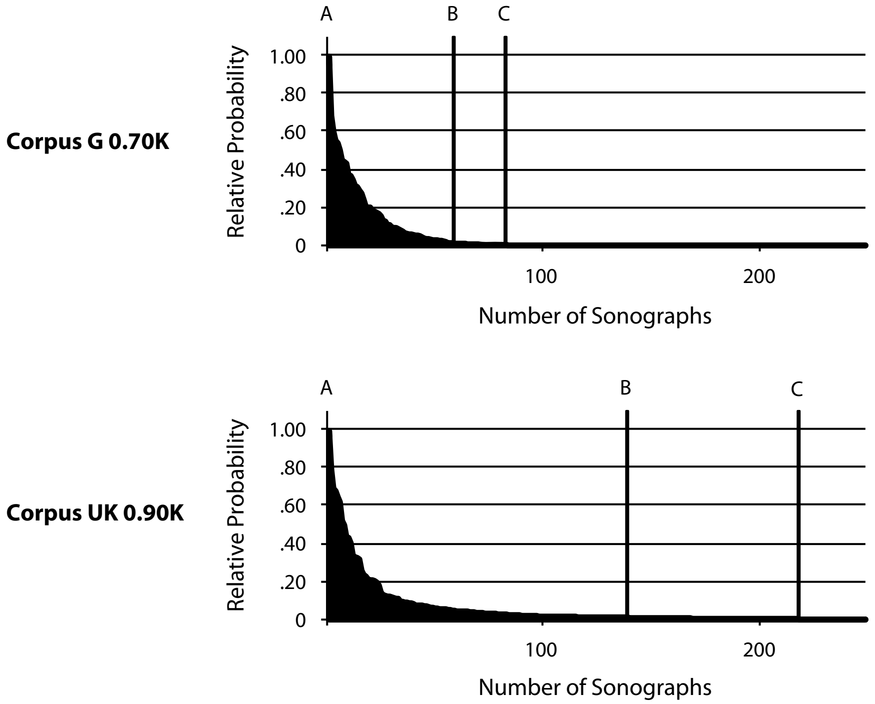

Figure 2. Sonograph probability profiles for G (Greek) and UK (British English, based on K. A. Spencer, 2009) for corpora of similar size. For comparative purposes, the relative probability values are expressed as a proportion of the highest frequency sonograph. Position B indicates relative probabilities <.01. Position C indicates the total number of sonographs for the corpus.

culation of the five metrics for one sonograph: /i/–<ει>. The 4,162 words in the present corpus are made up of a total of 27,585 constituent phonemes and their associated graphemes (27,585 sonographs). Of these, there are 2,412 occurrences of the phoneme /i/, which is represented by six graphemes (i.e., there are 6 sonographs associated with the /i/ phoneme). The <ει> grapheme occurs 437 times in association with /i/ (the remaining 1,975 occurrences are distributed among the other five graphemes associated with /i/). Thus, the unique /i/–<ει> sonograph occurs 437 times in the total number of 27,585 sonographs, with a probability of occurrence of .0158. The grapheme <ει> is usually associated with the /i/ phoneme (437 occurrences), but in a small number of cases (10), it is associated with other phonemes. So, although it is highly predictable, its probability (.98) falls short of unity.

Greek Orthography

Hatzigeorgiou, Mikros, and Karayiannis (2001) carried out analyses of the characteristics of written Greek based on the Hellenic National Corpus (HNC), a corpus of written modern Greek developed by the Institute for Language and Speech Processing. They reported that the average word length is 5.45 letters for the 13 million words in their corpus, and that the most familiar words are shorter than the average length of those in the whole corpus. Ktori, van Heuven, and Pitchford (2008) recently developed a lexical database of modern Greek based on the lexicon of common modern Greek and the HNC. Their analyses revealed that the majority of Greek words are between 8 and 9 letters in length, with a mean word length of 9.07 letters. However, about 50% of the summed word frequencies were accounted for by words with 4 or fewer letters, leading to the conclusion that the distribution of word length in Greek is compatible with Zipf's (1949) "principle of least effort." The average word in terms of frequency was found to be 5.7 letters long, comparable to that reported for the HNC (Hatzigeorgiou et al., 2001; Mikros, Hatzigeorgiou, & Karayiannis, 2005).

For broad phonetic transcription, modern Greek has 32 phonemes, of which 5 are vowels. These are written with 25 letters (including one end-only form), of which 7 correspond to vowels in isolation. Protopapas and Vlahou (2009) suggested that most words can be read correctly on the basis of the letter sequence alone, without the need for morphological or lexical information, and they have estimated Greek to have an overall feedforward consistency of 96% at the grapheme–phoneme level. This level of reading consistency has led Porpodas (2006) and Seymour, Aro, and Erskine (2003) to classify Greek as having a shallow orthography. However, Porpodas (2006) also recognized that for spelling, Greek is phonologically opaque, with one-to-many mappings, which means that spelling is not always predictable. Protopapas and Vlahou estimated the feedback token consistency to be 83%.

Thus, for reading correctly (but not necessarily fluently; see, e.g., Loizidou-Ieridou, 2007; Porpodas, 1999), phonological decoding is sufficient. On the other hand, very

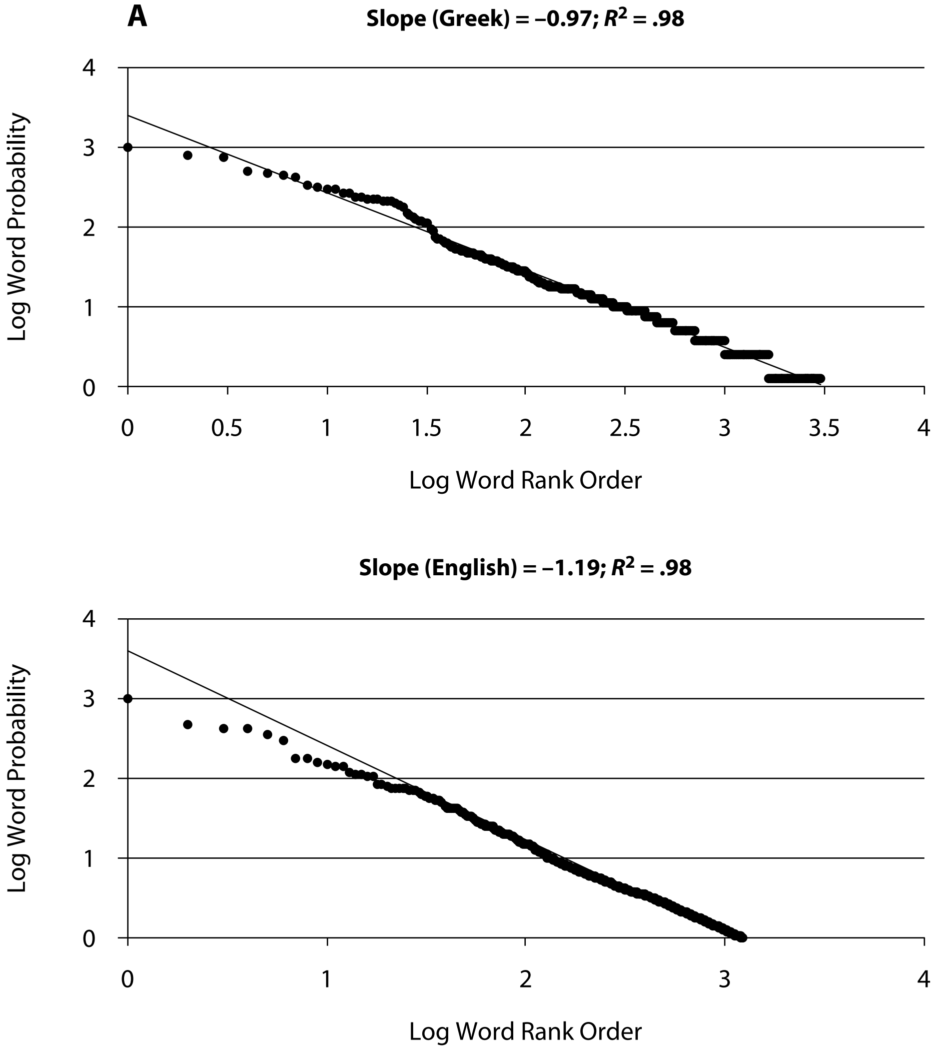

Figure 3A. Log–log plots for Greek and English words. For comparative purposes, the frequency values are expressed as a proportion of 1,000 occurrences. Word frequencies are based on ranks 1 to 3,000.

few words are regular in terms of one-to-one sound–letter correspondences, and the irregularities (mostly involving the representation of three of the five vowels) are usually affected by morphology and reflect semantic and grammatical distinctions. For example, the decision between *ο* and *ω* for the /o/ sound can be made using the rule that verbs are spelled with <*ω*> (e.g., *βάζω*, "to put") and neutral nouns with <*ο*> (e.g., *βάζο*, "flower pot"). Thus, correct spelling depends partly on decoding strategies and partly on morphological and syntactic knowledge (e.g., Chliounaki & Bryant, 2002; Loizidou-Ieridou, Masterson, & Hanley, 2009; Porpodas, 1999).

For modern Greek, the stress accent coincides with the written accent (´) of a word—for example, *μήλο* /'milo/. Only vowel letters bear a punctuation mark; additionally, the accent mark is used only when the word is of two or more syllables and when the word is written in lowercase (capital letters do not bear diacritics). There is also the diacritic dieresis (¨) to disambiguate single vowel graphemes from digraphs; for example, *αϊ* is pronounced as two separate vowels, as in *αϊβαλή* /aiva'li/.

METHOD

Corpora and Alignments

For the present analysis, a database was generated from the language textbooks used for Greek language instruction in the first two primary school grades in Cyprus (Karantzola, Kurdi, Spanelli, & Tsiagkani, 2008). This process resulted in a sample of 4,162 polysyllabic words.

To aid comparability with recently published sources of information concerning the Greek language (HNC; Hatzigeorgiou et al., 2000; http://hnc.ilsp.gr), phonemic definitions and alignments for the children's corpus were obtained from the psycholinguistic resources available at the Institute for Language and Speech Processing (http://speech.ilsp.gr/iplr/). The analysis by Protopapas and Vlahou (2009) used 217,644 unique word forms (types) accounting for 29,557,090 occurrences (tokens), based on 36 letters:

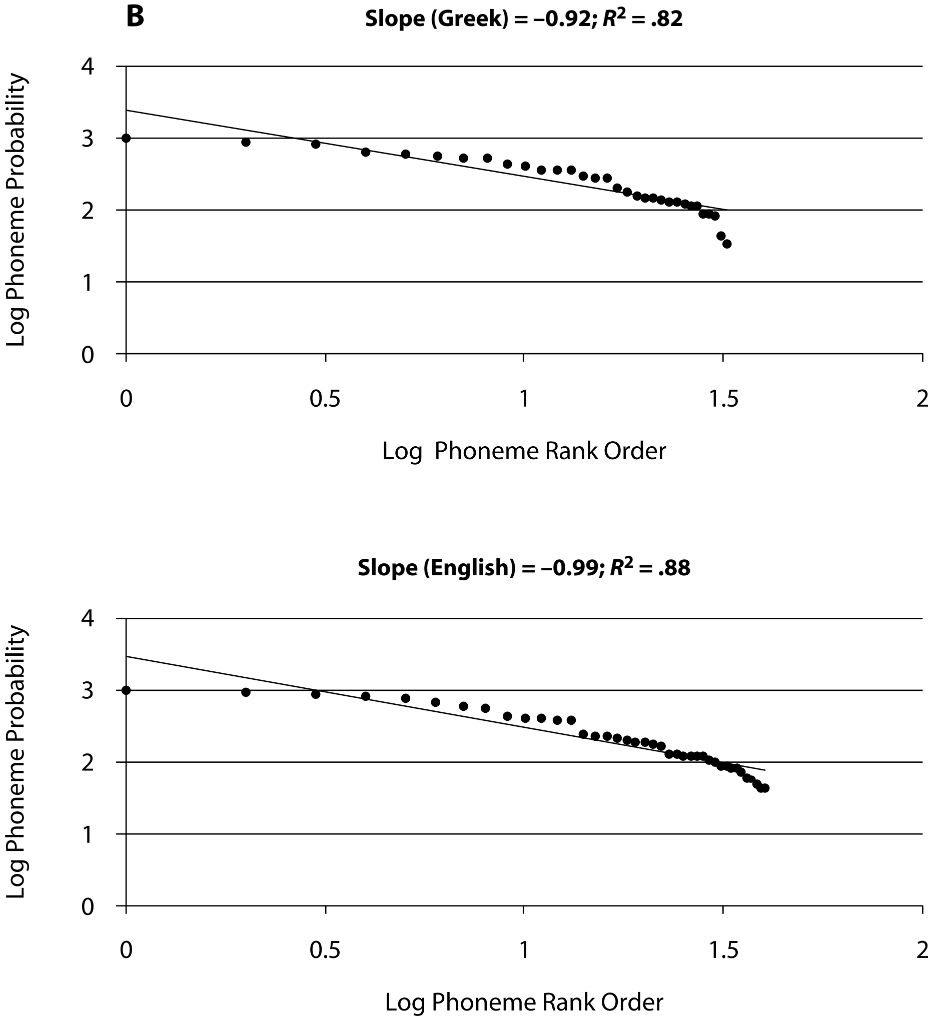

Figure 3B. Log–log plots for Greek and English phonemes. For comparative purposes, the frequency values are expressed as a proportion of 1,000 occurrences. Phoneme frequencies are based on 99% cumulative values to prevent excessive tail-end distortion.

> Of the 24 letters in the Greek alphabet, seven (the "vowel letters") have variants bearing diacritics. Specifically, all 7 may be accompanied by an acute accent, indicating stress. Two of these may carry diaeresis, indicating exception from digraph combinations. Because both types of diacritics are useful in phonological or lexical disambiguation, and because they are dictated by current spelling rules and their omission is always a spelling error, the variants of these letters with diacritics (stress mark only, diaeresis only, or both) were retained in the counts as separate letters. Including the word-final variant ς, a total of 36 letters were used in the analyses. (Protopapas & Vlahou, 2009, p. 995)

Their alignments are based on 37 phonemes:

> The resulting set of 32 phonemes, 5 of which are vowels, suffice to accurately and completely represent phonetically (broadly, at the surface realization) every Greek word in standard modern pronunciation typical of major cities such as Athens. To retain stress information, aiding in disambiguation, stressed vowels were represented as separate phonemes, bringing the total number of phonemes to 37. (Protopapas & Vlahou, 2009, p. 996)

The alignment process for the full corpus of 4,162 words resulted in a total of 100 sonographs, whereas the much larger corpus (217,664 words) of Protopapas and Vlahou yielded 118. In addition to the full corpus, a series of smaller corpora was created in order to explore the sonograph profiles for corpora of varying sizes. K. A. Spencer (2009) demonstrated that for English, increasing corpus size does not substantially alter the orthographic profile; it simply introduces additional low probability sonographs. The number of words and sonographs for the four corpora are: Corpus .21, number of words (nw) = 211, number of sonographs (nsg) = 63, frequency range (fr) ≥ 10; Corpus .71, nw = 706, nsg = 83, fr ≥ 3; Corpus 1.66, nw = 1,660, nsg = 93, fr > 1; Corpus 4.16, nw = 4,162, nsg = 100, fr = 1.

Frequency counts for each sonograph for the four corpora were produced, from which reading and spelling probabilities were calculated. The following measures were calculated: phoneme and grapheme frequency and probability of occurrence, sonograph probability of occurrence, GPC (reading) probability, and PGC (spelling) probability.

RESULTS AND DISCUSSION

The metrics for each corpus are presented in Appendices A and B. Table 2 shows that the values for the five metrics derived from the four sources are highly intercor-

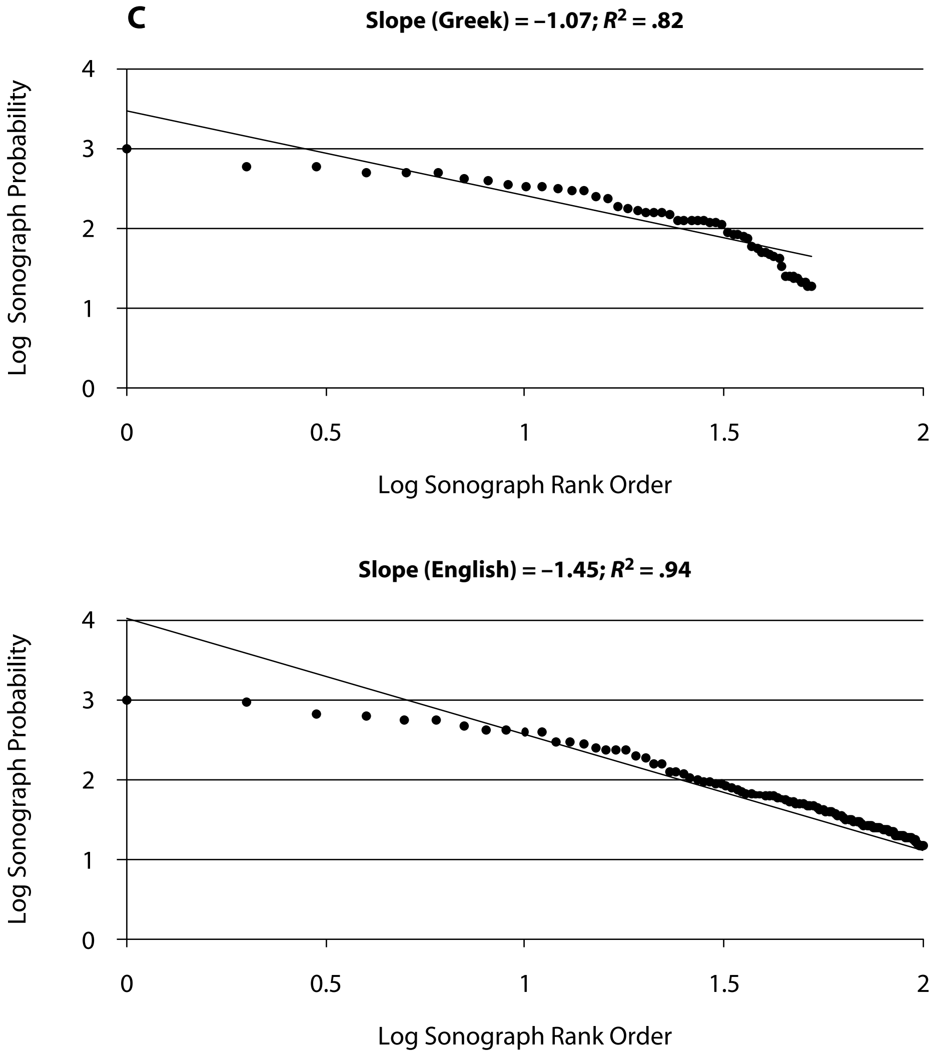

Figure 3C. Log–log plots for Greek and English sonographs. For comparative purposes, the frequency values are expressed as a proportion of 1,000 occurrences. Sonograph frequencies are based on 99% cumulative values to prevent excessive tail-end distortion.

related. The mean correlations among the sources are: .98 for grapheme, and .96 for phoneme probabilities of occurrence; .96 for GPC (reading), and .96 for PGC (spelling) probabilities; and .97 for sonograph probabilities of occurrence. This suggests that the metrics describing Greek remain substantially the same for corpora of varying size, from sources of writing for children. The present analysis finds that the core probability values for Greek may be obtained from relatively small samples of words, and this is reflected in Figure 1. For the most frequent words in the children's .21K corpus of 63 sonographs, all are in the relative probability[1] of occurrence range of .01 to 1.0 (shown between A and B, Figure 1). The larger .70K corpus has 20 additional sonographs, most of which are of very low probability (<.01), indicated by the bars between B and C (Figure 1). This pattern is repeated for the increasingly larger corpora. The difference between the corpus sizes of .70K and 1.66K is an additional 10 very low frequency sonographs (the bars between B and C, Figure 1), and between 1.66K and 4.10K, there are an additional 7 very low frequency sonographs.

The purpose of calculating descriptive metrics for a transparent language such as Greek is to provide continuous measures of the variables that may support current linguistic data (HNC; Hatzigeorgiou et al., 2000; Ktori et al., 2008; Protopapas & Vlahou, 2009), which may be applied in linguistic and psychological studies. This has usually involved the decomposition of very large corpora, but it appears that the present study confirmed K. A. Spencer's (2009) results for English, which demonstrated that smaller samples of words produced metrics that were very similar to those from larger samples.

Table 2 also demonstrates a relationship that reflects the difference between reading and spelling in Greek. The negative correlations between phoneme probability of occurrence and PGC probability, although small, reflect the nature of the spelling metric: The more frequent phonemes tend to be vowels, which are associated with more than one grapheme and consequently have PG values less than 1.0. Less frequent phonemes—mainly consonants—tend to have PG probabilities close to 1.0, because they are often associated with only one grapheme.

Figure 2 compares the sonograph probabilities for Greek with those for U.K. English (based on K. A. Spencer, 2009). The long tail observed in the opaque English orthography is not present in the relatively transparent Greek distribution. This clearly suggests that literacy skills for Greek will be easier to learn because there are fewer sonographs, and a larger proportion of these are more frequent and provide more opportunities for learning. This is reflected in the additional adult processing resources observed by Paulesu et al. (2000) when comparing PET brain scan activity for a transparent (Italian) and two opaque (English and French) languages.

However, Figures 3A–3C show Zipfian plots for words and phonemes in both languages, demonstrating their similarity at this structural level. Zipf's (1936, 1949) distribution of the logarithm of the rank of a signal (word, phoneme) against the logarithm of the frequency of occurrence for a human language provides a function of its potential capacity for communication. Zipf's (1949) law is based on the "principle of least effort," in which human speech and language are structured optimally by two opposing forces: unification and diversification. This process results in a balance that can be statistically represented by a regression coefficient (or slope) of −1.00, regressing the log of the rank on the log of actual frequency of occurrence. This applies to a multitude of diverse human languages, including modern Greek. Hatzigeorgiou et al. (2001) found the slope for modern Greek words to be −.98, which compares favorably with the present study's value of −.97. As would be expected for a relatively transparent language, the regression slope for Greek sonographs is also close to −1.0. For English, as a reflection of lack of orthographic transparency, the slope moves away from the ideal. Furthermore, as is shown in Figure 3, the slope has a value of −1.4, covering cumulative frequencies up to 99% (198 of the 316 sonographs) in order to avoid distortion by very low frequency sonographs (see Martindale, Gusein-Zade, McKenzie, & Borodovsky, 1996). The regression coefficient for the full range of English sonographs is −1.8. Clearly, the balancing processes that were identified by Zipf (1949) do not appear to be influencing English orthography, but they generally do apply to Greek. However, in a totally transparent orthography, the number of phonemes would dictate the number of sonographs, and this is not the case for the present Greek corpus, which has 37 phonemes and 100 sonographs. Appendix A demonstrates that, although there are more sonographs than phonemes, reading probabilities are generally high, with the mean dominant reading probability for vowels being .95. This is not the case for spelling probabilities. Appendix B shows that most vowels have spelling probabilities that are less than unity, and that the mean dominant spelling probability for vowels is .79, confirming claims that Greek is less transparent for spelling than for reading. These continuous spelling correspondence values should influence Greek children's spelling performance in a manner similar to the influence of English PGCs on spelling (K. A. Spencer, 2007).

Conclusion

Word metrics have tended to focus on feedforward and feedbackward metrics to account for variations in reading and spelling performance. The present article included the primary fine-grained metrics (sonograph probabilities), from which these secondary metrics were calculated. They were included to provide researchers, especially those in the field of children's psycholinguistics, with a range of measures that may be incorporated into models of reading and spelling to gain a more comprehensive understanding of the development of foundation literacy skills in the transparently asymmetric Greek language.

AUTHOR NOTE

Address correspondence to K. A. Spencer, Centre for Educational Studies, Hull University, Hull HU6 7RX, England (e-mail: k.a.spencer@hull.ac.uk).

REFERENCES

Berndt, R. S., Reggia, J. A., & Mitchum, C. C. (1987). Empirically derived probabilities for grapheme-to-phoneme correspondences in English. *Behavior Research Methods, Instruments, & Computers*, **19**, 1-9.

Carney, E. (1994). *A survey of English spelling.* London: Routledge.

Chliounaki, K., & Bryant, P. (2002). Construction and learning to spell. *Cognitive Development*, **17**, 1489-1499.

Coltheart, M. (2006). Dual route and connectionist models of reading: An overview. *London Review of Education*, **4**, 5-17.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, **100**, 589-608.

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**, 204-256.

Cossu, G., Gugliotta, M., & Marshall, J. (1995). Acquisition of reading and written spelling in a transparent orthography: Two nonparallel processes? *Reading & Writing: An Interdisciplinary Journal*, **7**, 9-22.

Ellis, N. C., Natsume, M., Stavropoulou, K., Hoxhallari, L., van Daal, V. H. P., Polyzoe, N., et al. (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading Research Quarterly*, **39**, 438-460.

Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in German- and English-speaking children. *Scientific Studies of Reading*, **2**, 31-54.

Gontijo, P. F. D., Gontijo, I., & Shillcock, R. (2003). Grapheme–phoneme probabilities in British English. *Behavior Research Methods, Instruments, & Computers*, **35**, 136-157.

Hanna, P. R., Hanna, J. S., & Hodges, R. E. (1966). *Phoneme–grapheme correspondences as cues to spelling improvement.* Washington, DC: U.S. Government Printing Office.

Hatzigeorgiou, N., Gavrilidou, M., Piperidis, S., Karayannis, G., Papakostopoulou, A., Spiliotopoulou, A., et al. (2000). Design and implementation of the online ILSP Greek corpus. In *Proceedings of the Second International Conference on Language Resources and Evaluation* (Vol. 3, pp. 1737-1740). Athens, Greece: LREC.

Hatzigeorgiou, N., Mikros, G., & Karayiannis, G. (2001). Word length, word frequencies and Zipf's law in the Greek language. *Journal of Quantitative Linguistics*, **8**, 175-185.

Karantzola, E., Kurdi, K., Spanelli, T., & Tsiagkani, T. (2008). *Glossa a'dimotikou* (Vols. 1 and 2). Athens, Greece: OEDB.

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 67-84). Amsterdam: Elsevier.

Ktori, M., van Heuven, W. J. B., & Pitchford, N. J. (2008). GreekLex: A lexical database of modern Greek. *Behavior Research Methods*, **40**, 773-783.

Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German–English comparison. *Cognition*, **63**, 315-334.

Loizidou-Ieridou, N. (2007). *Literacy development and reading difficulties in Greek-speaking Cypriot children aged between 6 and 11 years.* Unpublished doctoral dissertation, University of Essex.

Loizidou-Ieridou, N., Masterson, J., & Hanley, R. (2009). *Spelling development in 6–11 year old Greek-speaking Cypriot children.* Manuscript submitted for publication.

Martindale, C., Gusein-Zade, S. M., McKenzie, D., & Borodovsky, M. Y. (1996). Comparison of equations describing the ranked frequency distributions of graphemes and phonemes. *Journal of Quantitative Linguistics*, **3**, 106-112.

Masterson, J., Stuart, M., Dixon, M., & Lovejoy, S. (2003). Children's printed word database (Economic and social research council project R00023406). Retrieved September 28, 2007, from www.essex.ac.uk/psychology/cpwd/.

McGuinness, D. (1998). *Why children can't read: And what we can do about it.* London: Penguin.

Mikros, G., Hatzigeorgiou, N., & Karayiannis, G. (2005). Basic quantitative characteristics of the modern Greek language using the Hellenic National Corpus. *Journal of Quantitative Linguistics*, **12**, 167-184.

Oney, B., & Goldman, S. (1984). Decoding and comprehension skills in Turkish and English: Effects of regularity of grapheme–phoneme correspondences. *Journal of Educational Psychology*, **76**, 557-568.

Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S. F., et al. (2000). A cultural effect on brain function. *Nature Neuroscience*, **3**, 91-96.

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.

Plaut, D. C. (1999). A connectionist approach to word reading and acquired dyslexia: Extension to sequential processing. *Cognitive Science*, **23**, 543-568.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, **103**, 56-115.

Porpodas, C. D. (1999). Patterns of phonological and memory processing in beginning readers and spellers of Greek. *Journal of Learning Disabilities*, **32**, 406-416.

Porpodas, C. D. (2006). Literacy acquisition in Greek: Research review of the role of phonological and cognitive factors. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 189-199). Mahwah, NJ: Erlbaum.

Protopapas, A., & Vlahou, E. (2009). A comparative quantitative analysis of Greek orthographic transparency. *Behavior Research Methods*, **41**, 991-1008.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, **96**, 523-568.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, **94**, 143-174.

Spencer, K. A. (2007). Predicting children's word-spelling difficulty for common English words from measures of orthographic transparency, phonemic and graphemic length and word frequency. *British Journal of Psychology*, **98**, 305-338.

Spencer, K. A. (2009). Feedforward, -backward, and neutral transparency measures for British English. *Behavior Research Methods*, **41**, 220-227. doi:10.3758/BRM.41.1.220

Spencer, L. H., & Hanley, J. R. (2003). Effects of orthographic transparency on reading and phoneme awareness in children learning to read in Wales. *British Journal of Psychology*, **94**, 1-28.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, **124**, 107-136.

Venezky, R. L. (1967). English orthography: Its graphical structure and its relation to sound. *Reading Research Quarterly*, **2**, 75-106.

Ziegler, J. C., Jacobs, A. M., & Stone, G. O. (1996). Statistical analysis of the bidirectional inconsistency of spelling and sound in French. *Behavior Research Methods, Instruments, & Computers*, **28**, 504-515.

Ziegler, J. C., 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. *Behavior Research Methods, Instruments, & Computers*, **29**, 600-618.

Zipf, G. K. (1936). *The psycho-biology of language: An introduction to dynamic psychology.* London: Routledge.

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Cambridge, MA: Addison-Wesley.

NOTE

1. For comparative purposes, probabilities are expressed relative to the most frequent sonograph.

APPENDIX A
Grapheme and GPC Probability

	Grapheme Probability					GPC Probability			
	Corpus Size					Corpus Size			
Grapheme	4.10K	1.66K	0.70K	0.21K	Phoneme	4.10K	1.66K	0.70K	0.21K
α	.0972	.0941	.0917	.0895	a	.9948	.9891	.9758	.9054
					ˈa	.0052	.0109	.0242	.0946
ά	.0389	.0457	.0540	.0484	ˈa	1.0000	1.0000	1.0000	1.0000
αι	.0055	.0058	.0058	.0080	ε	.9934	.9821	.9524	.8570
					ˈe	.0066	.0179	.0476	.1429
αί	.0023	.0029	.0025	.0012	ˈe	1.0000	1.0000	1.0000	1.0000
β	.0116	.0125	.0114	.0036	v	1.0000	1.0000	1.0000	1.0000
γ	.0199	.0215	.0199	.0157	ɣ	.6095	.6029	.5833	.6923
					ʝ	.3832	.3971	.4170	.3080
					ŋ	.0073	–	–	–
γγ	.0005	.0003	.0003	–	ɟ	.7333	.3333	–	–
					g	.2667	.6667	1.0000	–
γει	.0000	.0001	.0003	–	ʝ	1.0000	1.0000	1.0000	–
γι	.0015	.0022	.0030	.0048	ʝ	1.0000	1.0000	1.0000	1.0000
γκ	.0014	.0009	.0008	–	g	.7436	.5556	–	–
					ɟ	.2564	.4444	1.0000	–
γκι	.0001	.0001	–	–	ɟ	1.0000	1.0000	–	–
γυ	.0001	.0001	–	–	ʝ	1.0000	1.0000	–	–
δ	.0177	.0192	.0216	.0190	ð	1.0000	1.0000	1.0000	1.0000
ε	.0580	.0561	.0512	.0496	ε	.9675	.9707	.9676	.9268
					ˈe	.0325	.0293	.0324	.0732
έ	.0235	.0265	.0277	.0302	ˈe	1.0000	1.0000	1.0000	1.0000
ει	.0162	.0179	.0199	.0133	i	.9776	.9598	.9167	.9091
					ˈi	.0134	.0230	.0556	.0909
					ʝ	.0045	.0115	.0139	–
					ç	.0045	.0057	.0139	–
εί	.0058	.0063	.0066	.0133	ˈi	1.0000	1.0000	1.0000	1.0000
ζ	.0083	.0064	.0047	.0024	z	1.0000	1.0000	1.0000	1.0000
η	.0249	.0246	.0190	.0270	i	.9869	.9791	.9560	.9090
					ˈi	.0102	.0209	.0441	.0909
					ʝ	.0029	–	–	–
ή	.0126	.0148	.0163	.0169	ˈi	1.0000	1.0000	1.0000	1.0000
θ	.0108	.0097	.0072	.0097	θ	1.0000	1.0000	1.0000	1.0000
ι	.0371	.0369	.0352	.0363	i	.8074	.7827	.7402	.7667
					ʝ	.1173	.1142	.1417	.1000
					ç	.0674	.0836	.0866	.0667
					ɲ	.0059	.0139	.0157	.0333
					ˈi	.0020	.0056	.0157	.0333
ί	.0158	.0171	.0180	.0169	ˈi	1.0000	1.0000	1.0000	1.0000
ϊ	.0006	.0004	–	–	i	1.0000	1.0000	–	–
κ	.0412	.0408	.0380	.0351	k	.6989	.7053	.7445	.7586
					c	.3011	.2947	.2555	.2414
κι	.0018	.0013	.0028	.0012	c	1.0000	1.0000	1.0000	1.0000
κκ	.0004	.0004	.0006	.0012	c	.9000	1.0000	1.0000	1.0000
					k	.1000	–	–	–
λ	.0331	.0324	.0310	.0302	l	1.0000	1.0000	1.0000	1.0000
λει	.0003	.0004	.0003	–	ʎ	1.0000	1.0000	1.0000	–
λι	.0021	.0020	.0022	.0012	ʎ	1.0000	1.0000	1.0000	1.0000
λλ	.0024	.0029	.0039	.0073	l	1.0000	1.0000	1.0000	1.0000
λλι	.0001	.0002	–	–	ʎ	1.0000	1.0000	–	–
μ	.0332	.0324	.0332	.0387	m	1.0000	1.0000	1.0000	1.0000
μμ	.0008	.0007	.0008	.0012	m	1.0000	1.0000	1.0000	1.0000
μπ	.0041	.0046	.0047	.0048	b	1.0000	1.0000	1.0000	1.0000
ν	.0495	.0471	.0396	.0423	n	1.0000	1.0000	1.0000	1.0000
νι	.0016	.0012	.0008	–	ɲ	1.0000	1.0000	1.0000	–
νν	.0004	.0007	.0006	–	n	1.0000	1.0000	1.0000	–
ννι	.0001	.0001	.0003	–	ɲ	1.0000	1.0000	1.0000	–
νοι	.0001	–	–	–	ɲ	1.0000	–	–	–
ντ	.0080	.0060	.0044	.0024	d	1.0000	1.0000	1.0000	1.0000
ντζ	.0002	.0002	.0003	–	ʣ	1.0000	1.0000	1.0000	–
ξ	.0086	.0077	.0086	.0048	ks	1.0000	1.0000	1.0000	1.0000
ο	.0410	.0396	.0404	.0484	o	.9991	.9974	.9932	1.0000
					ˈo	.0009	.0026	.0068	–
ό	.0189	.0210	.0247	.0326	ˈo	1.0000	1.0000	1.0000	1.0000
οί	.0060	.0059	.0086	.0070	i	.8424	.7368	.6129	.5000
					ç	.1333	.2632	.3871	.5000

APPENDIX A (Continued)

	Grapheme Probability					GPC Probability			
	Corpus Size					Corpus Size			
Grapheme	4.10K	1.66K	0.70K	0.21K	Phoneme	4.10K	1.66K	0.70K	0.21K
					ɲ	.0242	-	–	–
οί	.0014	.0009	.0011	–	ˈi	1.0000	1.0000	1.0000	–
ου	.0156	.0146	.0152	.0109	u	.9930	.9930	.9818	1.0000
					ˈu	.0070	.0070	.0182	–
ού	.0123	.0114	.0078	.0048	ˈu	1.0000	1.0000	1.0000	1.0000
π	.0350	.0387	.0452	.0508	p	1.0000	1.0000	1.0000	1.0000
ππ	.0003	.0005	.0006	–	p	1.0000	1.0000	1.0000	–
ρ	.0577	.0580	.0615	.0532	ɾ	1.0000	1.0000	1.0000	1.0000
ρρ	.0003	.0002	–	–	ɾ	1.0000	1.0000	–	–
σ	.0504	.0442	.0391	.0399	s	.9511	.9721	.9929	1.0000
					z	.0489	.0279	.0071	–
ς	.0287	.0307	.0330	.0375	s	1.0000	1.0000	1.0000	1.0000
σσ	.0008	.0011	.0011	–	s	1.0000	1.0000	1.0000	–
τ	.0499	.0486	.0485	.0665	t	1.0000	1.0000	1.0000	1.0000
τζ	.0003	.0003	–	–	ʤ	1.0000	1.0000	–	–
τσ	.0018	.0015	.0019	.0012	ʦ	1.0000	1.0000	1.0000	1.0000
υ	.0153	.0145	.0136	.0097	i	.7470	.7092	.7143	.3750
					f	.1371	.2128	.1837	.5000
					v	.1064	.0638	.0612	–
					ʝ	.0047	.0071	.0204	.1250
					ɲ	.0024	.0071	.0204	–
					ç	.0024	–	–	–
ύ	.0070	.0065	.0061	.0085	ˈi	.7098	.7619	.8636	1.0000
					v	.1813	.1587	.1364	–
					f	.1088	.0794	–	–
υι	.0000	–	–	–	ç	1.0000	–	–	–
φ	.0162	.0164	.0155	.0097	f	1.0000	1.0000	1.0000	1.0000
χ	.0171	.0184	.0197	.0121	x	.7288	.7039	.7042	.3000
					ç	.2712	.2961	.2958	.7000
χι	.0004	.0004	–	–	ç	1.0000	1.0000	–	–
ψ	.0033	.0031	.0025	.0012	ps	1.0000	1.0000	1.0000	1.0000
ω	.0145	.0139	.0166	.0157	o	.9875	.9704	.9667	1.0000
					ˈo	.0125	.0296	.0333	–
ώ	.0075	.0065	.0083	.0133	ˈo	1.0000	1.0000	1.0000	1.0000

Note—Grapheme Probability is the grapheme probability of occurrence. GPC Probability, grapheme–morpheme correspondence (reading) probability.

APPENDIX B
Phoneme, PGC, and Sonograph Probability

		Phoneme Probability				PGC Probability				Sonograph Probability			
		Corpus Size				Corpus Size				Corpus Size			
Ph	Gr	4.10K	1.66K	0.70K	0.21K	4.10K	1.66K	0.70K	0.21K	4.10K	1.66K	0.70K	0.21K
a	α	.0966	.0930	.0895	.0810	1.0000	1.0000	1.0000	1.0000	.0966	.0930	.0895	.0810
ˈa	ά	.0394	.0468	.0562	.0568	.9871	.9780	.9606	.8511	.0389	.0457	.0540	.0484
ˈa	α					.0129	.0220	.0394	.1489	.0005	.0010	.0022	.0085
b	μπ	.0041	.0046	.0047	.0048	1.0000	1.0000	1.0000	1.0000	.0041	.0046	.0047	.0048
c	κ	.0146	.0138	.0130	.0109	.8507	.8731	.7447	.7778	.0124	.0120	.0097	.0085
c	κι					.1269	.0970	.2128	.1111	.0018	.0013	.0028	.0012
c	κκ					.0224	.0299	.0426	.1111	.0003	.0004	.0006	.0012
ç	χ	.0084	.0106	.0125	.0145	.5494	.5146	.4667	.5833	.0046	.0054	.0058	.0085
ç	ι					.2961	.2913	.2444	.1667	.0025	.0031	.0030	.0024
ç	οι					.0944	.1456	.2667	.2500	.0008	.0015	.0033	.0036
ç	χι					.0429	.0388	–	–	.0004	.0004	–	–
ç	ει					.0086	.0097	.0222	–	.0001	.0001	.0003	–
ç	υ					.0043	–	–	–	.0000	–	–	–
ç	υι					.0043	–	–	–	.0000	–	–	–
d	ντ	.0080	.0060	.0044	.0024	1.0000	1.0000	1.0000	1.0000	.0080	.0060	.0044	.0024
ð	δ	.0177	.0192	.0216	.0193	1.0000	1.0000	1.0000	1.0000	.0177	.0192	.0216	.0193
ʣ	τζ	.0005	.0005	.0003	–	.6429	.6000	–	–	.0003	.0003	–	–
ʣ	ντζ					.3571	.4000	1.0000	–	.0002	.0002	.0003	–
f	φ	.0191	.0200	.0180	.0145	.8498	.8205	.8615	.6667	.0162	.0164	.0155	.0097
f	υ					.1103	.1538	.1385	.3333	.0021	.0031	.0025	.0048
f	ύ					.0399	.0256	–	–	.0008	.0005	–	–
ɡ	γκ	.0012	.0007	.0003	–	.8788	.7143	–	–	.0011	.0005	–	–
ɡ	γγ					.1212	.2857	1.0000	–	.0001	.0002	.0003	–
ɣ	γ	.0121	.0130	.0116	.0109	1.0000	1.0000	1.0000	1.0000	.0121	.0130	.0116	.0109
i	ι	.0874	.0851	.0773	.0713	.3425	.3394	.3369	.3898	.0299	.0289	.0260	.0278
i	η					.2807	.2826	.2330	.3390	.0245	.0241	.0180	.0242
i	ει					.1812	.2017	.2366	.1695	.0158	.0172	.0183	.0121
i	υ					.1310	.1208	.1254	.0508	.0115	.0103	.0097	.0036
i	οι					.0576	.0507	.0681	.0508	.0050	.0043	.0053	.0036
i	ϊ					.0070	.0048	–	–	.0006	.0004	–	–
ˈi	ί	.0411	.0451	.0499	.0605	.3850	.3781	.3611	.2800	.0158	.0171	.0180	.0169
ˈi	ή					.3057	.3280	.3278	.2800	.0126	.0148	.0163	.0169
ˈi	εί					.1410	.1390	.1333	.2200	.0058	.0063	.0066	.0133
ˈi	ύ					.1207	.1093	.1056	.1400	.0050	.0049	.0053	.0085
ˈi	οί					.0344	.0205	.0222	–	.0014	.0009	.0011	–
ˈi	η					.0062	.0114	.0167	.0400	.0003	.0005	.0008	.0024
ˈi	ει					.0053	.0091	.0222	.0200	.0002	.0004	.0011	.0012
ˈi	ι					.0018	.0046	.0111	.0200	.0001	.0002	.0006	.0012
ɟ	γγ	.0009	.0006	.0008	–	.4583	.1667	–	–	.0004	.0001	–	–
ɟ	γκ					.4167	.6667	1.0000	–	.0004	.0004	.0008	–
ɟ	γκι					.1250	.1667	–	–	.0001	.0001	–	–
ʝ	γ	.0138	.0154	.0172	.0145	.5526	.5533	.4839	.3333	.0076	.0085	.0083	.0048
ʝ	ι					.3158	.2733	.2903	.2500	.0044	.0042	.0050	.0036
ʝ	γι					.1079	.1400	.1774	.3333	.0015	.0022	.0030	.0048
ʝ	γυ					.0053	.0067	–	–	.0001	.0001	–	–
ʝ	ει					.0053	.0133	.0161	–	.0001	.0002	.0003	–
ʝ	η					.0053	–	–	–	.0001	–	–	–
ʝ	υ					.0053	.0067	.0161	.0833	.0001	.0001	.0003	.0012
ʝ	γει					.0026	.0067	.0161	–	.0000	.0001	.0003	–
k	κ	.0288	.0288	.0283	.0266	.9987	1.0000	1.0000	1.0000	.0288	.0288	.0283	.0266
k	κκ					.0013	–	–	–	.0000	–	–	–
ks	ξ	.0086	.0077	.0086	.0048	1.0000	1.0000	1.0000	1.0000	.0086	.0077	.0086	.0048
l	λ	.0355	.0353	.0349	.0375	.9325	.9184	.8889	.8065	.0331	.0324	.0310	.0302
l	λλ					.0675	.0816	.1111	.1935	.0024	.0029	.0039	.0073
m	μ	.0341	.0331	.0341	.0399	.9755	.9783	.9756	.9697	.0332	.0324	.0332	.0387
m	μμ					.0245	.0217	.0244	.0303	.0008	.0007	.0008	.0012
n	ν	.0499	.0478	.0402	.0423	.9913	.9849	.9862	1.0000	.0495	.0471	.0396	.0423
n	νν					.0087	.0151	.0138	–	.0004	.0007	.0006	–
ɲ	νι	.0022	.0020	.0019	.0012	.7333	.6316	.4286	–	.0016	.0012	.0008	–
ɲ	ι					.1000	.2632	.2857	1.0000	.0002	.0005	.0006	.0012
ɲ	οι					.0667	–	–	–	.0001	–	–	–
ɲ	νοι					.0500	–	–	–	.0001	–	–	–
ɲ	ννι					.0333	.0526	.1429	–	.0001	.0001	.0003	–
ɲ	υ					.0167	.0526	.1429	–	.0000	.0001	.0003	–
ŋ	γ	.0001	–	–	–	1.0000	–	–	–	.0001	–	–	–

APPENDIX B (Continued)

		Phoneme Probability				PGC Probability				Sonograph Probability			
		Corpus Size				Corpus Size				Corpus Size			
Ph	Gr	4.10K	1.66K	0.70K	0.21K	4.10K	1.66K	0.70K	0.21K	4.10K	1.66K	0.70K	0.21K
o	ο	.0554	.0529	.0562	.0641	.7407	.7456	.7143	.7547	.0410	.0395	.0402	.0484
o	ω					.2593	.2544	.2857	.2453	.0144	.0135	.0161	.0157
ˈo	ό	.0265	.0280	.0338	.0459	.7104	.7500	.7295	.7105	.0189	.0210	.0247	.0326
ˈo	ώ					.2814	.2316	.2459	.2895	.0075	.0065	.0083	.0133
ˈo	ω					.0068	.0147	.0164	–	.0002	.0004	.0006	–
ˈo	ο					.0014	.0037	.0082	–	.0000	.0001	.0003	–
p	π	.0353	.0392	.0457	.0508	.9908	.9869	.9879	1.0000	.0350	.0387	.0452	.0508
p	ππ					.0092	.0131	.0121	–	.0003	.0005	.0006	–
ps	ψ	.0033	.0031	.0025	.0012	1.0000	1.0000	1.0000	1.0000	.0033	.0031	.0025	.0012
ɾ	ρ	.0580	.0582	.0615	.0532	.9950	.9965	1.0000	1.0000	.0577	.0580	.0615	.0532
ɾ	ρρ					.0050	.0035	–	–	.0003	.0002	–	–
s	σ	.0775	.0748	.0729	.0774	.6183	.5742	.5323	.5156	.0479	.0430	.0388	.0399
s	ς					.3709	.4107	.4525	.4844	.0287	.0307	.0330	.0375
s	σσ					.0108	.0151	.0152	–	.0008	.0011	.0011	–
t	τ	.0499	.0486	.0485	.0665	1.0000	1.0000	1.0000	1.0000	.0499	.0486	.0485	.0665
ʦ	τσ	.0018	.0015	.0019	.0012	1.0000	1.0000	1.0000	1.0000	.0018	.0015	.0019	.0012
u	ου	.0155	.0145	.0150	.0109	1.0000	1.0000	1.0000	1.0000	.0155	.0145	.0150	.0109
ˈu	ού	.0124	.0115	.0080	.0048	.9912	.9911	.9655	1.0000	.0123	.0114	.0078	.0048
ˈu	ου					.0088	.0089	.0345	–	.0001	.0001	.0003	–
v	β	.0145	.0145	.0130	.0036	.7995	.8652	.8723	1.0000	.0116	.0125	.0114	.0036
v	υ					.1128	.0638	.0638	–	.0016	.0009	.0008	–
v	ύ					.0877	.0709	.0638	–	.0013	.0010	.0008	–
x	χ	.0125	.0130	.0139	.0036	1.0000	1.0000	1.0000	1.0000	.0125	.0130	.0139	.0036
ʎ	λι	.0025	.0026	.0025	.0012	.8529	.7600	.8889	1.0000	.0021	.0020	.0022	.0012
ʎ	λει					.1029	.1600	.1111	–	.0003	.0004	.0003	–
ʎ	λλι					.0441	.0800	–	–	.0001	.0002	–	–
z	ζ	.0108	.0076	.0050	.0024	.7710	.8378	.9444	1.0000	.0083	.0064	.0047	.0024
z	σ					.2290	.1622	.0556	–	.0025	.0012	.0003	–
ε	ε	.0616	.0601	.0551	.0532	.9111	.9060	.8995	.8636	.0561	.0545	.0496	.0459
ε	αι					.0889	.0940	.1005	.1364	.0055	.0057	.0055	.0073
ˈε	έ	.0277	.0311	.0321	.0363	.8480	.8515	.8621	.8333	.0235	.0265	.0277	.0302
ˈε	αί					.0826	.0924	.0776	.0333	.0023	.0029	.0025	.0012
ˈε	ε					.0682	.0528	.0517	.1000	.0019	.0016	.0017	.0036
ˈε	αι					.0013	.0033	.0086	.0333	.0000	.0001	.0003	.0012
θ	θ	.0108	.0097	.0072	.0097	1.0000	1.0000	1.0000	1.0000	.0108	.0097	.0072	.0097

Note—PGC Probability, phoneme–grapheme correspondence (spelling) probability; Sonograph Probability, sonograph probability of occurrence; Ph, phoneme; Gr, grapheme.

(Manuscript received March 23, 2009;
revision accepted for publication July 10, 2009.)