Dennett, D.C. (1995). *Intuition Pumps*. In Brockman, J., (ed.) The Third Culture: Beyond the Scientific Revolution (p. 180–197). New York, NY: Simon & Schuster.

Frank, S. L., & Bod, R. (2011). Insensitivity of the Human Sentence-Processing System to Hierarchical Structure. *Psychological Science*, 22 (6), 829–834.

Freud, S. (1891/1953). *On aphasia: A critical study*. New York: International University Press.

Jaeger, T.F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446.

Libben, G. (forthcoming). Requiem for an independent variable. *The Mental Lexicon*.

Libben, G. (2010). Compounds, semantic transparency, and morphological transcendence. In S. Olson (ed.), *New impulses in word-formation (Linguistische Berichte Sonderheft 17*), Buske, Hamburg, 212–232.

Libben, G. & Jarema, G. (2002). Mental lexicon research in the new millennium. *Brain and Language,* 81, 1–10.

Lichtheim, L. *(*1884). Uber aphasie. *Deutsches Archiv fur Klinishe Medizin, 36*, 204–268.

Loftus, G.R. (1993). A picture is worth a thousand p values: On the irrelevance of hypothesis testing in the microcomputer age. *Behavior Research Methods, Instruments, & Computers, 24*, 250–256.

Marshall, J. C. (1984). Toward a rational taxonomy of the developmental dyslexias. In R.N. Malatesha & H.A. Whitaker (Eds.), *Dyslexia: A global issue.* (pp. 211–232). The Hague: Martinus Nijhoff.

Meehl, P. & MacCorquodale, K. (1991). On a distinction between hypothetical constructs and intervening variables. In C. Anderson & K. Gunderson (Eds.) *Paul E. Meehl: Selected Philosophical & Methodological Papers* (pp. 249–263). Minneapolis, MN: University of Minnesota Press.

Popper, K. (1965). *The logic of scientific discovery*. New York, NY: Harper.

Raaijmakers, J. G., Schrijnemakers, J. M., & Gremmen, F. (2006). Corrigendum to "How to deal with "the language-as-fixed-effect fallacy": Common misconceptions and alternative solutions". *Journal of Memory and Language, 54*(4), 634.

Raaijmakers, J.G.W. (2003). A further look at the "language-as-fixed-effect fallacy'. *Canadian Journal of Experimental Psychology, 57*, 141–151.

Raaijmakers, J.G.W., Schrijnemakers, J.M.C., & Gremmen, F. (1999). How to deal with "the language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language, 41*, 416–426.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379–423.

Tremblay, A., & Tucker, B. V. (2011). The effects of N-gram probabilistic measures on the recognition and production of four-word sequences. *The Mental Lexicon, 6* (2), 302{324.

Vasishth, S. and Lewis, R.L. (2006). Human Language Processing: Symbolic Models. In: Brown, K (Ed.) *Encyclopedia of Language and Linguistics, 2nd edition*, volume 5, pages 410–419. Amsterdam, Netherlands: Elsevier.

Monaghan, P., Christiansen, M.H., Farmer, T.A. & Fitneva, S.A. (2012). Measures of phonological typicality: Robust coherence and psychological validity. In G. Libben, G. Jarema & C. Westbury (Eds.), Methodological and analytic frontiers in lexical research (pp. 13-31). Amsterdam, The Netherlands: John Benjamins.

Measures of phonological typicality

Robust coherence and psychological validity

Padraic Monaghan, Morten H. Christiansen, Thomas A. Farmer and Stanka A. Fitneva

Lancaster University, UK / Cornell University, USA / Department of Brain and Cognitive Science, University of Rochester, USA / Queens University, Canada

Phonological Typicality (PT) is a measure of the extent to which a word's phonology is typical of other words in the lexical category to which it belongs. There is a general coherence among words from the same category in terms of speech sounds, and we have found that words that are phonologically typical of their category tend to be processed more quickly and accurately than words that are less typical. In this paper we describe in greater detail the operationalisation of measures of a word's PT, and report validations of different parameterisations of the measure. For each variant of PT, we report the extent to which it reflects the coherence of the lexical categories of words in terms of their sound, as well as the extent to which the measure predicts naming and lexical decision response times from a database of monosyllabic word processing. We show that PT is robust to parameter variation, but that measures based on PT of uninflected words (lemmas) best predict response time data for naming and lexical decision of single words.

Phonological Typicality (PT) is a psycholinguistic construct that reflects the extent to which a word is typical or atypical of its lexical category, with respect to its phonology. A series of studies have indicated that measures of PT can predict variance in lexical access (Farmer, Christiansen, & Monaghan, 2006; Fitneva, Christiansen, & Monaghan, 2009; Monaghan, Chater, & Christiansen, 2003). Effects of PT thus show that access to the phonological characteristics of a word's lexical category is implicated early in lexical processing (Tanenhaus & Hare, 2007).

There has been a spate of research examining the coherence of different lexical categories with respect to their phonological and prosodic characteristics. Kelly (1992) investigated a range of sound cues that could distinguish lexical categories, including length (nouns tend to be longer than verbs in English in terms of number of syllables), consonant and vowel distribution (e.g., nouns tend to contain more coronal consonants than verbs), and stress (nouns tend to have trochaic stress, whereas verbs tend to have iambic stress; Cutler & Carter, 1987). Moreover, analyses of child-directed speech and connectionist simulations have quantified the usefulness of potential phonological cues such as syllabic complexity (Morgan, Shi, & Allopenna, 1996), stress position (Kelly & Bock, 1998), and number of syllables (Cassidy & Kelly, 2001), for distinguishing between different lexical categories. In large-scale corpus analyses, Durieux and Gillis (2001) and Monaghan, Chater, and Christiansen (2005) have tested the extent to which a combined set of phonological and prosodic cues can reflect distinctions between different lexical categories. These studies found that the cues were sufficient to distinguish lexical categories to a high degree of precision, and this was the case cross-linguistically (Monaghan, Christiansen, & Chater, 2007).

Thus, there is a degree of phonological coherence within lexical categories. This has been proposed to be important for acquisition of lexical categories (Braine et al., 1990; Brooks, Braine, Catalano, Brody, & Sudhalter, 1993; Cassidy & Kelly, 2001; Monaghan et al., 2005; St Clair & Monaghan, 2005). If phonological coherence is important for acquisition, then we can hypothesise that some residual effect of the acquisition process is observable in adult lexical processing. In other words, if a word is typical of its lexical category with respect to its phonology, then it ought to be accessed and processed more easily than a word that is atypical of its category in terms of phonology. In a series of studies, we have operationalised the measure of PT, and found support for these hypotheses.

Fitneva et al. (2009) demonstrated that in learning new words PT is used by seven-year-olds for lexical category assignment. They found that upon hearing a nonword containing phonological properties highly typical of verbs, children were significantly more likely to pair it with a picture of an action than they were with a picture of an object. Interestingly, English-speaking seven-year-olds in French immersion programs appeared to assign lexical category to the nonwords according to their PT in French (when the test was given in French).

In addition to directing children's learning of novel words' lexical categories, PT has an early, online effect on adults' lexical processing, influencing response times to both nouns and verbs in a lexical decision task as well as naming latencies for verbs (Monaghan et al., 2003). Moreover, across a series of studies, Farmer et al. (2006) demonstrated that PT influences both lexical and syntactic processing in adulthood. In their first study, they demonstrated that PT accounts for a significant amount of the variance in a database of lexical naming times (Spieler & Balota, 1997), even after controlling for a standard array of psycholinguistic and acoustic variables that have also been demonstrated to influence naming times. The effects of PT were not limited to words appearing in isolation, but also influenced reading times in

sentences containing various types of syntactic manipulations. Using a self-paced reading methodology, Farmer et al. (2006) conducted two additional experiments focusing on the processing of typical and atypical words occurring in unambiguous sentences. One experiment involved sentence frames selected so as to strongly predict that a noun will come next, whereas the frames in the other experiment were created to generate strong expectations for a verb. When the preceding context generated a strong expectation for an upcoming noun, noun-like nouns were read faster than verb-like nouns, and when the context was highly predictive of a verb, verb-like verbs were read faster than noun-like verbs.[1] Additionally, Farmer et al. demonstrated that PT can even bias the reading of a syntactic ambiguity created by the presence of a noun/verb homonym. When the homonym was noun-like, participants preferred the interpretation of the ambiguity that was consistent with the noun interpretation of it, and vice versa when the homonym was verb-like.

PT has also been shown to modulate the magnitude of early-occurring neural responses to violations of syntactically-driven expectations. Using magnetoencephalography (MEG), Dikker, Rabagliati, Farmer, and Pylkkanen (2010) demonstrated that the visual M100 response, a component in visual cortex that arises approximately 100–130 milliseconds after stimulus onset in response to violations of word category expectations while reading, is sensitive to PT. They found that an effect of expectedness of a noun (should a noun be next or not) was modulated by the PT of the incoming noun. In a condition where all nouns had phonological properties highly typical of nouns, the effect of expectedness was larger than in a condition where all of the nouns were neutral in terms of their phonology. That is, the magnitude of the M100 was significantly larger when a highly typical noun occurred unexpectedly, compared to when its occurrence was expected. When the nouns were not typical or atypical of other nouns (neutral), there was no difference in M100 magnitude in the expected versus the unexpected condition.

Taken together, these studies demonstrate the powerful and broad influence that lexical category-based phonological regularities, as captured by PT, have during acquisition in children and for on-line processing in adulthood. However, these previous studies of PT have been limited to a single operationalisation of the measure. In this paper, we examine alternative parameterisations of PT, reporting in greater detail than previously how the measure was calculated, and validating each parameterisation in terms of reflecting the *coherence* of the lexical category distinction, as well as its *psychological validity* as reflected by the relationship of PT to lexical decision responses times and naming latencies for monosyllabic words in English.

Method

The original operationalisation of phonological typicality

In the original measure of PT developed in Monaghan et al. (2003), and utilised in the behavioural studies (e.g., Farmer et al., 2006; Fitneva et al., 2009), we made a number of decisions in terms of parameterising the measure. We used only monosyllabic words that were unambiguous with respect to lexical category against which to measure PT, and each word made an equal contribution to the PT measure without regard to its frequency. Furthermore, we used a frequency cut-off of 1/million in the Celex corpus (Baayen, Pipenbrock & Gulikers, 1995) for a word to be included in the PT measure.

In order to determine the distance between each pair of words, in the initial operationalisation we partitioned each word into three slots for onset, two for the vowel, and three for the coda. For example, the word *kelp* was represented as /k––ɛ–lp–/, where "–" denotes an empty slot. Each phoneme was, in turn, represented by a set of eleven phonological features derived from Harm and Seidenberg (1999) and originally based on government phonology theory (Chomsky & Halle, 1968). The features were: sonorant, consonantal, voice, nasal, degree, labial, palatal, pharyngeal, round, tongue, and radical. A key aspect of this phonological feature representation is that phonemes that are easily confused (Miller & Nicely, 1955) tend to have a similar representation, so /p/ and /b/ differ in only one of the 11 features — whether they are voiced or not — but /p/ and /f/ differ on 4 of the 11 features.

When comparing a pair of words, the phonemes were repositioned within the onset, within the vowel, and within the coda in order to determine the alignment resulting in the minimum Euclidean distance between the phonemes in the two words. In the analyses reported in this paper, we relaxed the constraint on alignments only occurring within the onset, nucleus and coda. Thus, any sequential alignment of the two words is permitted in order to minimise the distance between the words. Thus, for the words *act* and *cat*, /ækt/ and /kæt/, the closest alignment could be: /æk–t/ and /–kæt/, where "–" indicates an empty slot, such that the consonants of the coda of *act* are compared against the onset and coda of *cat*. For this alignment, the actual distance measure would be computed from comparisons between the phonemes /æ/ and the empty slot, /k/ and /k/, which would be zero, /æ/ and the empty slot again, and /t/ and /t/. For each pair of words, all possible permutations of alignment were tested, and the alignment with the lowest distance was selected.[2]

In order to make the computations involved in PT transparent, Table 1 shows a worked example of computing the phonological feature distance between *kelp* /kɛlp/ and the words *peer* /pɪəɹ/ and *street* /stɹit/. The phonological feature distance is computed by summing the squares of the differences between each phoneme slot

Table 1. Examples of Computing Phonological Feature Distance (FD) between kelp and the Words peer and street.

Position in Word	*kelp*		*peer*			*street*		
	Phoneme	Phonological features	Phoneme	Phonological Features	Sum of Squared Differences	Phoneme	Phonological Features	Sum of squared differences
1	–	{–1,–1,–1,–1,–1, –1,–1,–1,–1,–1, –1}	–	{–1,–1,–1,–1,–1, –1,–1,–1,–1,–1, –1}	0+0+0+0+0+0+ 0+0+0+0+0	s	{–0.5,1,–1,–1, 0,–1,1,–1,–1, 1,0}	0.25+4+0+0+1+ 0+4+0+0+4+1
2	k	{–1,1,–1,–1,1,–1, –1,–1,–1,–1,0}	p	{–1,1,–1,–1,1,1,0, –1,1,0,0}	0+0+0+0+0+4+ 1+0+4+1+0	t	{–1,1,–1,–1, 1,–1,1,–1,–1, 1,0}	0+0+0+0+0+0+ 4+0+0+4+0
3	ɛ	{1,–1,1, 0,–1,–1, 0,–1,–1,–1,–1}	ɪə	{1,–1,1,0,–0.5,–1, 0,–1,–0.5,–0.5, –1}	0+0+0+0+0.25+ 0+0+0+0.25+ 0.25+0	ɹ	{0.5,0,1,0,–1, –1, 1,1,1,–1, –1}	0.25+1+0+0+0+ 0+1+4+4+0+0
4	l	{0.5,0,1,0,–1,–1,1, –1,–1,1,0}	ɹ	{0.5,0,1,0,–1,–1, –1,1,1,–1,–1}	0+0+0+0+0+0+ 4+4+4+4+1	i	{1,–1,1,0,0,–1, 0,–1,–1,0,1}	0.25+1+0+0+1+ 0+1+0+0+1+1
5	p	{–1,1,–1,–1,1,1,0, –1,1,0,0}	-	{–1,–1,–1,–1,–1, –1,–1,–1,–1,–1, –1}	0+4+0+0+4+4+ 1+0+4+1+1	t	{–1,1,–1,–1, 1,–1,1,–1,–1, 1,0}	0+0+0+0+0+4+ 1+0+4+1+0
FD = Σ √(sum square differences)					12.51			15.26

in terms of its phonological features, and then taking the square root of this sum for each phoneme. For the first phoneme position, in the comparison between *kelp* and *peer*, for instance, the phonological feature representation of /k/ is {–1, 1,–1,–1,1,–1,–1,–1,–1,–1,0} and for /p/ it is {–1,1,–1,–1,1,1,0,–1,1,0,0}. Then the squared difference between the first phonological feature for this phoneme position is: $(-1 - -1)^2 = 0$. For the second position, the squared difference is $(1 - 1)^2 = 0$, for the third, fourth, and fifth positions, the squared difference is also zero, for the sixth position the squared difference is $(-1 - 1)^2 = 4$, and so on for all 11 phonological features. The sum of the squared differences for /k/ versus /p/ is then 10, so this phoneme contributes $\sqrt{10}$ to the overall distance measure. Then, the overall distance between *kelp* and *peer* is the sum of the square roots of the squared differences for each phoneme position: for *kelp* and *peer*, the phonological feature distance is $\sqrt{0} + \sqrt{10} + \sqrt{0.75} + \sqrt{17} + \sqrt{19} = 12.51$. For the distance between *kelp* and *street*, the Euclidean distance is $\sqrt{14.25} + \sqrt{8} + \sqrt{10.25} + \sqrt{5.25} + \sqrt{10} = 15.26$.

Overall, *kelp* is a noun-like word because its average Euclidean distance to nouns is 11.83, which is less than its average Euclidean distance to verbs of 12.42. Its PT value, which we calculate by subtracting the average verb distance for a word from its average noun distance, is 11.83–12.42 = –0.61. For a more general depiction of this kind of analysis, Figure 1a shows the distance for each noun and verb for uninflected monosyllabic words. The diagonal shows the objective point at which distances to nouns and distances to verbs are equal. For PT coherence, points indicating verbs should demonstrate overall shorter distances to verbs and longer distances to nouns than for the nouns, so verbs should be to the lower right of nouns. Though there is considerable overlap, the points indicate that verbs tend to be more similar to other verbs than they are to nouns, and the majority of the verbs tend to be to the lower right of the nouns.

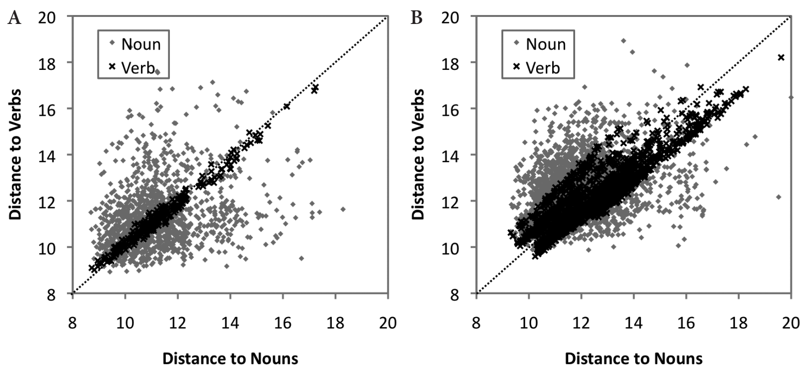

Figure 1. PT for nouns and for verbs for (A) word lemmas, and (B) wordforms using the FD measure.

Varying parameters of the operationalisation

There were thus several decisions made for operationalising PT in our original measure. We discuss alternatives for each of these decisions, before exploring the implications of making different selections at each of these decision points for the PT measure. Our aim was to determine whether PT was robust to varying the precise parameters of the measure, or whether PT effects were particularly reflected by certain choices of representation of phonological similarity between words.

Decisions about the reference vocabulary. The first decision about the vocabulary against which PT is calculated for each word is whether to include only nouns and verbs that are unambiguous with respect to their lexical category. Alternatively, all words used either as nouns or verbs or both could be used to calculate PT.

Second, in the original formulation of PT (Farmer et al., 2006; Monaghan et al., 2003), all uninflected words (lemmas) from the Celex English database (Baayen et al., 1995) were used as the reference vocabulary. However, this omitted word forms with inflectional and derivational morphology, which could have a profound influence on the calculation of PT. It is therefore important also to test both broad and limited word sets. Another alternative to using lemmas is to use words that are classified as monomorphemic in the Celex database. However, we did not test further the monomorphemic analyses, partly because there were very few monosyllabic lemmas that were classified as polymorphemic, and partly because most of these classifications appeared to be false positives as the morphology was judged automatically, so, for example, words ending in /s/ were labelled as polymorphemic, as in *axe*.

The third decision about the reference vocabulary is to determine the contribution that each word makes to the PT measure. Each word could contribute in a type analysis (as in the original operationalisation) or weighted by individual token frequency. In the following analyses, we test both type and token approaches.

The fourth decision determines whether basing PT on only monosyllabic words is sufficient to represent the vocabulary, or whether including bisyllabic words in the reference vocabulary improves the PT measure further. For now, we restrict analyses to only monosyllabic words, leaving multisyllabic words for future work. Nonetheless, we note here that the results of Farmer et al. (2006) suggest that PT scores for bisyllabic words based on analyses of monosyllabic words significantly affect word-by-word reading times in on-line sentence comprehension. Moreover, as longer words tend to be lower in frequency, monosyllabic words provide a reasonable reflection of the whole vocabulary: over 83% of the most frequent 1000 English word tokens and over 75% of the most frequent 5000 words in Celex, for instance, are monosyllabic.

Decisions about the distance measure. The phonological feature distance (FD) representation, mentioned above and illustrated in Table 1, is just one way to represent phoneme similarity, but it may not be the best computation of similarity between words in terms of their sound. Another possibility is to determine the number of phonological features that are different between the phonemes in two words. We refer to this as the phoneme feature edit distance (FE), and this is most similar to the best match to phoneme confusability in Bailey and Hahn's (2005) comparisons of syllables. Another alternative is to determine distance between two words in terms of how many phonemes are required to change in order to alter one word to another. This measure, which we refer to as the phoneme edit distance (PE), is analogous to Levenshtein's orthographic edit distance measure (Yarkoni, Balota, & Yap, 2008). Worked examples of the FE and PE distance measures for the words *kelp*, *peer*, and *street*, are shown in Table 2.

Validation of the measures

We validated the parameter variations in two ways: measuring *coherence* and *psychological validity* of each PT measurement.

Coherence. First, we tested the extent to which the measure reflected the previously observed coherence of lexical categories with respect to their phonology. For each word, we computed the mean distance for that word to all the nouns and to all the verbs. For nouns, we anticipated that the distance to other nouns would be smaller than the distance to verbs. For verbs, we anticipated that the distance to verbs would be smaller than the distance to nouns. For each parameterisation, we conducted a one-way ANOVA on the PT measure with noun/verb category as a between items factor. The coherence within a category in terms of phonology is reflected in the effect size of the main effect.

Psychological validity. Second, we tested the extent to which each parameterisation of PT had psychological validity in terms of predicting response times for lexical decision and single word naming tasks for a large number of monosyllabic words taken from the database reported in Spieler and Balota (1997) and Balota, Cortese, Sergent-Marshall, Spieler, and Yapp (2004). This database provides naming times for 2820 words by 31 young adult participants at Washington University, and lexical decisions for 2906 words by 30 participants (Balota, Cortese, & Pilotti, 1999). In order to test the contribution of PT, we first entered several psycholinguistic variables into a regression equation, as used by Balota et al. (2004). These were: characteristics of the word's onset (which were particularly important in predicting voice onset times for the naming data), familiarity (from Balota et al., 2004), neighbourhood size (Coltheart's N, calculated from the entire vocabulary in the Celex English database), orthographic word length, and log-frequency

Table 2. Examples of Computing Feature Edit (FE) and Phoneme Edit (PE) Distance Between kelp and the Words peer and street.

Position in Word	*kelp*		*peer*				*street*			
	Phoneme	Phonological Features	Phoneme	Phonological Features	FE	PE	Phoneme	Phonological Features	FE	PE
1	-	{−1,−1,−1,−1,−1,−1,−1,−1,−1,−1,−1,−1}	–	{−1,−1,−1,−1,−1,−1,−1,−1,−1,−1,−1}	0	0	s	{−0.5,1,−1,−1, 0,−1,1,−1,−1,1,0}	6	1
2	k	{−1,1,−1,−1,1,−1,−1,−1,−1,−1,0}	p	{−1,1,−1,−1,1,1,0,−1,1,0,0}	4	1	t	{−1,1,−1,−1,1,−1,1,−1,−1,1,0}	2	1
3	ε	{1,−1,1,0,−1,−1,0,−1,−1,−1,−1}	ɪə	{1,−1,1,0,−0.5,−1,0,−1,−0.5,−0.5,−1}	3	1	ɹ	{0.5,0,1,0,−1,−1,1,1,1,−1,−1}	5	1
4	l	{0.5,0,1,0,−1,−1,1,−1,−1,1,0}	ɹ	{0.5,0,1,0,−1,−1,−1,1,1,−1,−1}	5	1	i	{1,−1,1,0,0,−1,0,−1,−1,0,1}	6	1
5	p	{−1,1,−1,−1,1,1,0,−1,1,0,0}	-	{−1,−1,−1,−1,−1,−1,−1,−1,−1,−1,−1}	7	1	t	{−1,1,−1,−1,1,−1,1,−1,−1,1,0}	4	1
Σ differences					19	4			23	5

(from the Celex English database). After these variables had been entered into the regression equation, we determined the additional contribution of PT by including distance to nouns minus distance to verbs for the verbs and distance to verbs minus distance to nouns for the nouns (so positive values indicate a word typical of its category both for the nouns and for the verbs). The standardized beta-value reflects the size of the effect of PT in predicting the behavioural data once all the other psycholinguistic factors had been taken into account. We predicted that the beta values would be negative, indicating that typicality related to reduced response times for stimuli.

Results

Coherence

The results of each parameterisation for the coherence analyses are shown in Table 3 and Figure 2. Each point in the Figure shows the Z-score of PT for nouns on the x-axis and for verbs on the y-axis. PT is calculated by subtracting a word's distance to verbs from its distance to nouns. Thus, positive values on the x-axis indicate that nouns are closer to verbs than they are to nouns overall, and negative values indicate nouns are closer to nouns than they are to verbs. Positive values on the y-axis indicate that verbs are closer to other verbs than they are to nouns,

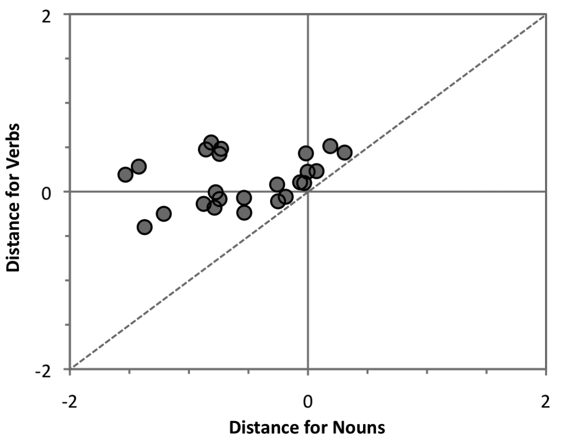

Figure 2. Coherence of nouns and verbs with respect to PT measure with different parameterisations, with Z-scores of distance for verbs and distance for nouns. Each point above the diagonal indicates that the particular parameterisation of PT reflects the phonological coherence of the vocabulary with respect to lexical category.

Table 3. Z-score of Mean Distance to Nouns Minus Distance to Verbs for Nouns (PT-N) and for Verbs (PT-V), with Different Parameterisations of the PT Measure.

Word Set	Unambiguous/ All N/V	Type/ Token	Distance Measure	N	PT-N	PT-V	F	η^2
Forms	U	Type	FD	4104	−.743	.425	1328***	.244
	U	Token	FD		−.730	.483	1441***	.260
	U	Type	FE		−.857	.475	1672***	.290
	U	Token	FE		−.813	.554	1806***	.306
	U	Type	PE		−1.533	.191	2294***	.359
	U	Token	PE		−1.421	.282	2337***	.363
	A	Type	FD	8174	−.785	−.181	538***	.062
	A	Token	FD		−.743	−.084	647***	.073
	A	Type	FE		−.876	−.138	718***	.081
	A	Token	FE		−.775	−.007	840***	.093
	A	Type	PE		−1.372	−.400	941***	.103
	A	Token	PE		−1.211	−.250	1016***	.111
Lemma	U	Type	FD	1580	−.259	.080	25***	.016
	U	Token	FD		.188	.513	22***	.014
	U	Type	FE		−.535	−.237	18***	.011
	U	Token	FE		−.003	.224	11***	.007
	U	Type	PE		−.537	−.070	46***	.029
	U	Token	PE		−.017	.431	43***	.027
	A	Type	FD	4716	.072	.231	23***	.005
	A	Token	FD		−.187	−.058	18***	.004
	A	Type	FE		−.066	.105	30***	.006
	A	Token	FE		−.251	−.109	24***	.005
	A	Type	PE		.308	.440	20***	.004
	A	Token	PE		−.032	.097	15***	.003

Note. U = unambiguous nouns and verbs, A = all nouns and verbs, FD = feature distance, FE = feature edit, PE = phoneme edit. N is the size of the corpus. For significance of F-value, *** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .1$.

whereas negative values indicate that verbs are closer to nouns than to other verbs. Over the whole vocabulary of nouns and verbs, then, points above the diagonal indicate phonological coherence of nouns and verbs as reflected in the particular parameterisation of PT.

In the one-way ANOVAs of monosyllabic nouns and verbs, shown in Table 3, PT was significantly different for all parameterisations. The largest effect sizes were found for the word form analyses, which is because the PT measures are partially reflecting the morphology of the word, and morphology is most richly expressed in this word set (shown in Figure 1b for the FD measure — note that the coherence appears greater than for the word lemma analysis in Figure 1a). However, the coherence was still observed in the word lemmas set, without inflectional morphology, consistent with previous work exploring distinct phonological cues to lexical category in word sets with and without morphology (Monaghan et al., 2005, 2007; Onnis & Christiansen, 2008). Together, these results indicate that the PT measure is robust across different ways of assessing the phonological information available to distinguish nouns and verbs.

Psychological validity

In our analyses of the word naming dataset from Balota et al. (2004), there were 2377 words classified as either nouns or verbs in CELEX (according to their most frequent usage). 1764 of these were nouns, and 613 were verbs. For the lexical decision data, there were 2446 words classified as either nouns or verbs, 1815 nouns and 631 verbs. As we were interested in the effects of PT on nouns and verbs, we only used words classified as belonging to these categories. Table 4 shows the results of the regression analyses partially replicating steps 1 and 2 of Balota et al. (2004), with onset variables entered at step 1, and psycholinguistic variables entered at step 2. At step 3, we tested for the effect of PT by subtracting the mean distance to all the verbs from the mean distance to all the nouns for each word. If the word was a noun, we then took the negative of this value (so a phonologically typical noun would have a positive score), and if the word was a verb, then we kept the original value (so a phonologically typical verb would also have a positive score).

The Table reports that the essential results of Balota et al. (2004) were replicated on this subset of the words (just the nouns and verbs) in the Balota et al. database. The onset variables had greatest effect for the naming task, and related only weakly to lexical decision. The psycholinguistic variables were all strong predictors of variance in naming responses, and frequency and familiarity were strong predictors for lexical decision times.

For Step 3, the measures of PT based on monosyllables were able to predict variance in response times to lexical decisions for nearly all parameterisations, and the PT measures based on word lemmas as a reference vocabulary were also able to predict variance in naming responses, as shown in Table 4. This is similar to the effects reported in Monaghan et al. (2003) for a single parameterisation of the PT measure. Though the PT measures contribute only a small amount to explaining

Table 4. Regression Analyses of Word Naming and Lexical Decision Response Times, with Onset Variables Entered at Step 1, Psycholinguistic Variables Entered at Step 2, and Different Parameterisations of PT at Step 3.

	Reference Vocabulary			Distance	Naming		Lexical Decision	
	Word Set	Unambiguous/ All NV	Type/Token	Measure	β	R^2	β	R^2
Step 1								
Onset variables					[−.252,.419]	.25	[−.126,.135]	.02
Step 2								
Psycholing. Vbles						.42		.41
Length					.122***		.035	
Log-Freq					−.117***		−.288***	
Neighbors					−.179***		.023	
Familiarity					−.185***		−.379***	
Step 3								
PT	Forms	U	Type	FD	−.003	.42	−.065***	.41
		U	Token	FD	−.001	.42	−.065***	.41
		U	Type	FE	.001	.42	−.062***	.41
		U	Token	FE	.004	.42	−.060***	.41
		U	Type	PE	−.019	.42	−.061***	.41
		U	Token	PE	−.019	.42	−.060***	.41
		A	Type	FD	−.003	.42	−.067***	.41
		A	Token	FD	.001	.42	−.066***	.41
		A	Type	FE	.000	.42	−.064***	.41
		A	Token	FE	.004	.42	−.060***	.41

Table 4. *(continued)*

Reference Vocabulary			Distance Measure	Naming		Lexical Decision	
Word Set	Unambiguous/ All NV	Type/Token		β	R^2	β	R^2
	A	Type	PE	−.018	.42	−.062***	.41
	A	Token	PE	−.018	.42	−.060***	.41
Lemmas	U	Type	FD	−.051**	.42	−.050**	.41
	U	Token	FD	−.048**	.42	−.008	.41
	U	Type	FE	−.043**	.42	−.059***	.41
	U	Token	FE	−.050**	.42	−.024	.41
	U	Type	PE	−.050**	.42	−.058***	.41
	U	Token	PE	−.049**	.42	−.021	.41
	A	Type	FD	−.041*	.42	−.025	.41
	A	Token	FD	−.041*	.42	−.039*	.41
	A	Type	FE	−.042**	.42	−.022	.41
	A	Token	FE	−.049**	.42	−.034*	.41
	A	Type	PE	−.032*	.42	−.024	.41
	A	Token	PE	−.038*	.42	−.046**	.41

Note. U = unambiguous nouns and verbs, A = all nouns and verbs, FD = feature distance, FE = feature edit, PE = phoneme edit. For standardized beta values: *** $p < .001$, ** $p < .01$, * $p < .05$.

the variance in responses, after accounting for the aforementioned variables they are highly significant for most of the parameterisations. For lexical decision the particular choice of parameters was not so critical for predicting reaction times, though weighting by frequency (token analyses) affected the predictiveness of the word lemmas analyses.

For naming, the parameterisation was more fragile — only if the reference vocabulary was word lemmas was the effect observed. It may be that the effect of PT on lexical access is masked by the contribution of inflectional morphology — the typicality of the word root may make the greatest contribution to predicting word processing.

Discussion

The PT measure aims to reflect the extent to which a word's phonology is similar to that of other words of the same lexical category. A typical noun, for instance, sounds more like other nouns than it does sound like verbs. However, to develop a measure of PT, a number of decisions have to be made — what does "similar" actually mean, are words that are ambiguous or unambiguous with respect to lexical category to be included in the measure, should the word set include morphological variants, and does a word's frequency have an influence on the typicality of other words' phonology with respect to their lexical category? We have shown in this paper that the precise decisions about the reference vocabulary used to generate the PT measure that we made in our initial formulation have an influence on the extent to which the vocabulary is shown to be coherent with respect to phonology within lexical categories, as well as the extent to which the PT measure has psychological validity in terms of being able to predict large datasets of naming and lexical decision response times.

In terms of coherence, the word forms obviously show the greatest effect, though the analyses of the lemma word sets confirm that morphology is not the only word property that results in phonological similarity among the lexical categories of nouns and verbs. The precise measure of similarity did not have a large effect on the validity of the PT measure with respect to coherence. Even the simplest measure — the number of phonemes that have to be adjusted to convert one word to another — reflected coherence of the categories as strongly as the more sophisticated measures of phoneme feature similarity.

In terms of the psychological validity of the PT measures, as reflected by predicting variance in response times to naming and lexical decision, we found that variants of the PT measure could predict lexical access for both of these tasks. For naming, this depended on comparing a word's typicality to word lemmas — when inflected word forms were also included, the predictiveness of PT for word naming was reduced. The effect of inflectional morphology was to obscure the true variance for predicting effects of PT for naming. However, for lexical decision, PT showed a larger effect, and was more robust to different parameterisations.

Yet, the most interesting aspect of this validation is that the effect of PT in these analyses was free from context. In previous studies of PT, the context has had an effect on processing. In Fitneva et al. (2009), pictures of objects and actions provided a visual context indicating the lexical category of the novel words that children were asked to learn — words that either conformed to or contradicted the phonology of its lexical category. In Farmer et al. (2006), PT was manipulated within a predictive sentential context. Indeed, Staub et al. (2009) and Farmer et al., (2011) have shown that when the context is weakened, the effect of PT is reduced. Yet, the regression analyses demonstrate that contextual information is not critical for eliciting effects of PT. These are subtle effects, and perhaps can only be revealed without the presence of a predictive context by large sets of stimuli, but they are nonetheless highly significant, and, in the case of the lexical decision data particularly, highly robust to decisions about the reference vocabulary.

It is of key theoretical and methodological importance that the effects of PT established in the literature now can be confidently interpreted as not being due to a particular parameterisation of the PT measure. The extent to which a word resembles other words of the same category with respect to its phonology has been shown to have an influence on acquisition of the vocabulary, as well as the lexical categories to which the words belong (Braine et al., 1992; Brooks et al., 1994; Monaghan et al., 2005). The influence of phonology with respect to lexical category for vocabulary *learning* appears to be observable in tasks that directly assess lexical acquisition (Fitneva et al., 2009), as well as access to the adult vocabulary in both predictive sentence contexts (Farmer et al., 2006) and when the word appears without any context in the analyses of the word naming and lexical decision databases present here. Effects of PT show that accessing a single word is interconnected with properties, both phonological and syntactic, of the entire vocabulary.

Determining how phonology and syntax become interconnected in terms of PT across development is an important topic for future studies. Such research promises to offer potential insights into the acquisition of phonology, lexical items, grammatical categories, and syntax, as well as how these developmental processes may interact. Investigations of the PT of children's first words, for example, will be highly informative about whether PT may play a role in structuring the vocabulary from the very onset of word learning. It would be also important to examine the possibility that PT operates based on the subset of words that the child knows at any given time. Another important question is how PT may interact with the acquisition of syntax. Many words have an ambiguous syntactic status in early

language development. PT may help solidify lexical category knowledge or it may emerge as a factor influencing word learning only after the syntactic and semantic properties of vocabulary items are more firmly established.

More generally, the robustness of PT effects has implications for the modularity of language processing. Traditionally, phonological and syntactic information have been considered to involve separate and independent levels of processing (e.g., Chomsky & Halle, 1968; Frazier, 1995; Hockett, 1963; Levelt, 1999). Yet, PT effects point to permeability of and interactivity between phonological and syntactic information in lexical processing. This raises the intriguing possibility that the processing of syntactic properties may be observable for all tasks involving isolated words and, conversely, that phonological properties may be important contributors to both sentence and discourse processing. Thus, future studies of PT may provide further support for the notion that lexical and syntactic processing are intrinsically interconnected, with PT providing a key window into those interactions.

Notes

1. Staub, Grant, Clifton, & Rayner (2009) reported a failure to replicate the effects of these two studies, but Farmer, Monaghan, Misyak, and Christiansen (2011) replicated the original effect and demonstrated that Staub et al. had altered critical features of the original experiment resulting in a reduction of the observed effect of PT due to weakening of contextual cues to lexical category.

2. Another adjustment from the original implementation of Monaghan et al. (2003) used in the analyses reported below is that the phonological representation of the vowel in the current analyses was a single slot, such that diphthongs were encoded as an average of the phonological features of the two vowels from which they are composed. In the original implementation, vowels occupied up to two slots. This was in order to increase the similarity between words containing single vowels, when one was a short vowel and the other was a diphthong or a long vowel (e.g., the long /i/ and the short /ɪ/ were previously distinguished by an additional vowel slot, and were therefore distant in the similarity space).

References

Baayen, R. H., Pipenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database* (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania,.

Bailey, T. M., & Hahn, U. (2005). Phoneme similarity and confusability. *Journal of Memory & Language, 52,* 347–370.

Balota, D. A., Cortese, M. J., & Pilotti, M. (1999). Item-level analyses of lexical decision performance: Results from a mega-study. In *Abstracts of the 40th Annual Meeting of the Psychonomics Society* (p. 44). Los Angeles, CA: Psychonomic Society.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yapp, M. J. (2004). Visual word recognition for single syllable words. *Journal of Experimental Psychology: General, 133*, 283–316.

Braine, M. D. S., Brody, R. E., Brooks, P. J., Sudhalter, V., Ross, J. A., Catalano, L., et al. (1990). Exploring language acquisition in children with a miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory & Language, 29*, 591–610.

Brooks, P. B., Braine, M. D. S., Catalano, L., Brody, R. E., & Sudhalter, V. (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of Memory & Language, 32*, 79–95.

Cassidy, K. W., & Kelly, M. H. (2001). Children's use of phonology to infer grammatical class in vocabulary learning. *Psychonomic Bulletin & Review, 8*(3), 519–523.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language, 2*, 133–142.

Dikker, S., Rabagliati, H., Farmer, T. A., & Pylkkanen, L. (2010). Early occipital sensitivity to syntactic category is based on form typicality. *Psychological Science, 21*, 629–634.

Durieux, G., & Gillis, S. (2001). Predicting grammatical classes from phonological cues: An empirical test. In J. Weissenborn & B. Höhle (Eds.), *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition* (Vol. 1, pp. 189–229). Amsterdam: John Benjamins.

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences, 103*, 12203–12208.

Farmer, T. A., Monaghan, P., Misyak, J. B., & Christiansen, M. H., (2011). Phonological typicality influences sentence processing in predictive contexts: A reply to Staub et al. (2009). *Journal of Experimental Psychology: Learning, Memory and Cognition 37*, 1318–1325.

Fitneva, S., Christiansen, M. H., & Monaghan, P. (2009). From sound to syntax: Phonological constraints on children's lexical categorization of new words. *Journal of Child Language, 36*, 967–997.

Frazier, L. (1995). Constraint satisfaction as a theory of sentence processing. *Journal of Psycholinguistic Research, 24*, 437–468.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading and dyslexia: Insights from connectionist models. *Psychological Review*, *106*, 491–528.

Hockett, C. F. (1963) The problem of universals in language. In Greenberg, J. H. (Ed.), *Universals of language* (pp. 1–29). Bradford: MIT Press.

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review, 99*, 349–364.

Kelly, M. H., & Bock, J. K. (1988) Stress in time. *Journal of Experimental Psychology: Human Perception and Performance*, 389–403.

Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences, 3*, 223–232.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America, 27*, 338–352.

Monaghan, P., Chater, N., & Christiansen, M. H. (2003). Inequality between the classes: Phonological and distributional typicality as predictors of lexical processing. *Proceedings of the 25th Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential contribution of phonological and distributional cues in grammatical categorisation. *Cognition, 96,* 143–182.

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The Phonological Distributional coherence Hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology, 55*, 259–305.

Morgan J. L., Shi R., & Allopenna P. (1996) Perceptual bases of rudimentary grammatical categories: Toward a broader conceptualization of bootstrapping. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax* (pp. 263–286). Mahwah, NJ: Erlbaum.

Onnis, L., & Christiansen, M. H. (2008). Lexical categories at the edge of the word. *Cognitive Science, 32*, 184–221.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science, 8*, 411– 416.

Staub, A., Grant, M., Clifton, C., & Rayner, K. (2009). Phonological typicality does not influence fixation durations in normal reading. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 35*(3), 806–814.

St Clair, M. C., & Monaghan, P. (2005). Categorizing grammar: Differential effects of preceding and succeeding contextual cues. *Proceedings of the 27th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.

Tanenhaus, M. K., & Hare, M. (2007). Phonological typicality and sentence processing. *Trends in Cognitive Sciences, 11*(3), 93–95.

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