

Cognitive Science (2016) 1–19

Copyright © 2016 Cognitive Science Society, Inc. All rights reserved.

ISSN: 0364-0213 print / 1551-6709 online

DOI: 10.1111/cogs.12361

Non-Arbitrariness in Mapping Word Form to Meaning: Cross-Linguistic Formal Markers of Word Concreteness

Jamie Reilly,^{a,b} Jinyi Hung,^{a,b} Chris Westbury^c

^a*Eleanor M. Saffran Center for Cognitive Neuroscience, Temple University*

^b*Department of Communication Sciences and Disorders, Temple University*

^c*Department of Psychology, University of Alberta*

Received 25 September 2014; received in revised form 24 September 2015; accepted 17 December 2015

Abstract

Arbitrary symbolism is a linguistic doctrine that predicts an orthogonal relationship between word forms and their corresponding meanings. Recent corpora analyses have demonstrated violations of arbitrary symbolism with respect to concreteness, a variable characterizing the sensorimotor salience of a word. In addition to qualitative semantic differences, abstract and concrete words are also marked by distinct morphophonological structures such as length and morphological complexity. Native English speakers show sensitivity to these markers in tasks such as auditory word recognition and naming. One unanswered question is whether this violation of arbitrariness reflects an idiosyncratic property of the English lexicon or whether word concreteness is a marked phenomenon across other natural languages. We isolated concrete and abstract English nouns ($N = 400$), and translated each into Russian, Arabic, Dutch, Mandarin, Hindi, Korean, Hebrew, and American Sign Language. We conducted offline acoustic analyses of abstract and concrete word length discrepancies across languages. In a separate experiment, native English speakers ($N = 56$) with no prior knowledge of these foreign languages judged concreteness of these nouns (e.g., Can you see, hear, feel, or touch this? Yes/No). Each naïve participant heard pre-recorded words presented in randomized blocks of three foreign languages following a brief listening exposure to a narrative sample from each respective language. Concrete and abstract words differed by length across five of eight languages, and prediction accuracy exceeded chance for four of eight languages. These results suggest that word concreteness is a marked phenomenon across several of the world's most widely spoken languages. We interpret these findings as supportive of an adaptive cognitive heuristic that allows listeners to exploit non-arbitrary mappings of word form to word meaning.

Keywords: Concreteness; Imageability; Semantic Memory; Phonology; Cross-Linguistic; Psycholinguistics; Arbitrary Symbolism

Correspondence should be sent to Jamie Reilly, Temple University, Weiss Hall, Philadelphia, PA 19122.
E-mail: reillyj@temple.edu

1. Introduction

Our empirical knowledge of language structure has largely been informed by the ways that people acquire, comprehend, and produce concrete words such as *dog*, *desk*, and *drum*. Yet a unique property of the human mind is its capacity for representing abstract concepts such as *irreducibility*, *irrationality*, and *irrelevance*. Virtually all documented languages are rife with abstract words that denote feelings, ideas, social concepts, and introspective states. For example, *waldeinsamkeit* (German) denotes the feeling of being alone in a forest, and *karoshi* (Japanese) denotes the phenomenon of working oneself to death. The question of how we process concrete relative to abstract words remains a central topic for the studies of cognition, language, and consciousness. The *word concreteness effect* describes the collective advantage that concrete words manifest over abstract in a multitude of cognitive domains, including age-of-acquisition, reading and spelling accuracy, word recognition, serial recall, and naming. Researchers have historically attributed concreteness effects to differences in the semantic structures of abstract and concrete words. However, concrete word advantages are also at least in part attributable to differences in the sound structures of concrete and abstract words.

Corpus analyses have demonstrated that abstract words are on average longer and more derivationally complex than concrete words (Reilly & Kean, 2007; Westbury & Moroschan, 2009). Numerous other formal cues (many co-varying with word length) mark the abstract–concrete dichotomy, including syllable stress patterns, phonotactic probability, compounding, phonological neighborhood density, and derivational complexity. English etymology is a potential latent factor that may account for many of the observed differences; abstract nouns are more commonly derived from Latinate, whereas concrete nouns are more often Germanic (Love, 2014).

Corpus analyses demonstrate patterns upon which language users might bootstrap from low-level sound structure to concreteness. Yet one cannot infer, *prima facie*, that any observed pattern in the data impacts language processing. The direct test of such a hypothesis involves analyzing whether sound structure and word concreteness interact in natural language processing. Mounting evidence supports the presence of such interactivity within domains such as naming and spoken word recognition where it is now reasonably well accepted that listeners exploit phonological cues to discriminate between word types, including open and closed class words (Shi, Morgan, & Allopenna, 1998; Shi, Werker, & Morgan, 1999), nouns and verbs (Durieux & Gillis, 2001; Langenmayr, Gozutok, & Gust, 2001; Monaghan, Christiansen, & Fitneva, 2011), and antonym pairs in foreign languages (Koriat, 1975).

We have only an incipient understanding regarding the extent to which word form moderates acquisition and/or processing as a function of word concreteness. Our initial studies of this phenomenon involved probing metalinguistic awareness of abstract–concrete word differences through pseudoword strings (Reilly, 2005; Reilly, Westbury, Kean, & Peelle, 2012). We specifically manipulated the length and phonological complexity of nonwords and asked healthy adults to make judgments of concreteness for each randomly

presented string (i.e., Can you see, hear, or touch this?). Participants reliably rated shorter nonwords with many orthographic neighbors as concrete, whereas longer nonwords with fewer neighbors were rated as abstract. We reasoned that knowledge of this statistical regularity could prove adaptive in facilitating lexical access for abstract and concrete words, and subsequently tested this prediction by examining the performance of patients with semantic dementia, a neurodegenerative condition that manifests as relatively circumscribed deficit in word and object knowledge (Reilly, Cross, Troiani, & Grossman, 2007). We hypothesized that patients who experience a relatively focal semantic impairment would demonstrate a pathological overreliance upon their preserved implicit phonological knowledge. We presented patients with spoken words varied factorially by length (short/long) and concreteness (abstract/concrete) and asked them to judge (yes/no) whether each word was concrete or abstract. Patients with semantic dementia often misclassified long concrete words (e.g., *apartment*) as abstract and short abstract words (e.g., *fate*) as concrete, evidence supporting the application of a word length heuristic in making semantic decisions.

1.1. Linguistic arbitrariness and word length

Swiss linguist Ferdinand de Saussure is credited with integrating *l'arbitraire du signe* (arbitrariness of the sign) into the formal study of language (Saussure, 1916). Saussure's oft-cited example was that of the word *tree*. In this particular example, the signifier ("tree") is arbitrarily related to the signified (a leafy green object in the world). Thus, there is nothing inherently *treelike* about the phoneme triplet /tri/. Analogously, the sound structure of an abstract word such as *waldeinsamkeit* does not map onto the feeling of being alone in the woods. Distributional evidence strongly favors arbitrary symbolism as the driving principle of lexical organization across mature languages. The presence of a predictive (non-orthogonal) relationship between word form (a signifier variable) and word concreteness (a conceptual variable) violates the assumption of linguistic arbitrariness. Nevertheless, this is an exception backed by compelling empirical data. Corpora, coupled with behavioral patterns in English, suggest that such interactive effects do manifest in language processing (for an extensive treatment of violations of arbitrariness in English phonology, see also Monaghan, Shillcock, Christiansen, & Kirby, 2014). Such a systematic sound–meaning mapping further suggests a potential middle ground where certain global attributes of word meaning (e.g., concreteness) might reasonably be inferred from lower level structural components (e.g., length).

It has long been recognized that there exist systematic relations between word length and a variety of other lexical and grammatical factors. Zipf's (1949), for example, describes an inverse relation between word length and lexical frequency. As languages evolve, pressures for optimizing communicative efficiency cause highly frequent words such as *automobile* to spontaneously truncate to *auto*. One might accordingly speculate that the abstract–concrete word length discrepancies observed in English corpora analyses reflect a Zipf-like process whereby concrete words are more frequently encountered than abstract words. However, lemma frequency data do not support this contention. The

correlation between word frequency and concreteness across thousands of English nouns is negligible (Reilly & Kean, 2007). Piantadosi, Tily, and Gibson (2011) recently advanced a nuanced perspective on Zipf's Law, arguing that word length is optimized for information content beyond simple frequency of occurrence. In a work of remarkable computational breadth, the authors examined relationships between orthographic word length and information content across 11 natural languages for target words situated within a text-based narrative context via a corpus-based analysis using the Google N-Gram database. For 10 of these languages, the correlation between word length and information content held in the predicted direction (i.e., longer words convey more information content). In response, Reilly and Kean (2011) raised the question of whether word length discrepancies between abstract and concrete nouns correspondingly mark differences in information content. In a reply, Piantadosi et al. (2011) argued that this is indeed the case. Abstract nouns are both statistically longer, *and* they convey more information content than concrete nouns. On this account, information content is one potential driver for formal differences that mark abstract and concrete words.

We hypothesize that concreteness is a key semantic distinction that is formally marked across many languages and that this distinction is predominantly marked by word length. Such cues may prove adaptive toward facilitating rapid online "routing" of concrete and abstract words for qualitatively different post-lexical semantic processing strategies (Reilly, Peelle, Garcia, & Crutch, 2016). This hypothesis finds parallel support in an extensive literature regarding syntactic bootstrapping, where language learners use sound to parse the grammatical distinction between nouns and verbs in running discourse comprehension (Chomsky & Halle, 1968; Kelly, 1992; Monaghan, Chater, & Christiansen, 2005).

Our aim here was to evaluate whether form-concreteness correspondence is an idiosyncratic property of the English lexicon, or whether the relationship is apparent across other natural languages. We reasoned that if similar acoustic phonetic markers of word concreteness exist across unrelated languages, then naive listeners might detect such cues to aid in "guessing" the concreteness of unfamiliar words. To follow, we report a combination of behavioral experimentation and corpus analyses of abstract and concrete nouns across eight widely spoken (or signed) languages, including Russian, Arabic, Dutch, Mandarin, Hindi, Korean, Hebrew, and American Sign Language (ASL).

2. Method

2.1. Participants

Participants included 56 young adult, monolingual English speakers (47 females; mean age = 19.77; range 18–23 years). Participants were by self-report free of language learning disabilities, dyslexia, or brain injury. We queried previous foreign language exposure via written questionnaire to ensure naiveté with the languages they would be tested on.

We conducted this behavioral study over the course of a semester and terminated data collection when we acquired at least 20 responses per item across all languages tested.

2.2. Materials

We first obtained a large pool of English nouns ($N > 600$) with concreteness ratings from the Medical Research Council (MRC) Psycholinguistic database (for aggregation and scaling procedures, see Coltheart, 1981)¹. Concreteness ratings are typically derived through Likert scales whereby participants rate the extent to which a word can be experienced through the senses. Abstract words are typically marked by lower ratings on this scale. Of note, the abstract–concrete dichotomy is neither absolute, nor are there firm numerical cutoffs for concreteness values that constitute abstract words. We isolated the tails of the concreteness distribution (highly abstract/highly concrete) by first filtering for part-of-speech (i.e., nouns) using the MRC database search delimiting function. We subsequently eliminated low frequency and archaic words, homophones, compound words, and any remaining words with ambiguous grammatical roles (e.g., content). These selection criteria yielded a sample of highly abstract and concrete nouns ($N = 200$ each). On a standard 100–700 point scale, the mean abstract word rating for this sample was 303.9, and the mean concrete word rating was 590.8 [$p_{\text{diff}} < .001$].

Once a suitable item pool was established, we then enlisted native speakers to translate the target words into Arabic, Mandarin, Dutch, Hindi, Hebrew, Korean, Russian, and ASL. We recorded native speakers' spoken production of the word list and later spliced each word into an individual audio mp3 file. Similarly, we video recorded a fluent signer as she produced the same item list in ASL and later spliced the sign language videos into individual segments. Each video began with the signer at rest, followed by all relevant motions providing handshape, location, and movement cues. The sign videos also included a full complement of non-manual markers (e.g., facial expressions, torso movement). Each video clip terminated after the speaker placed her hands down, and in this manner, each sign encapsulated the periods both before and after the signer produced a single word. We reasoned that this presentation method ameliorated the effects of co-articulation incurred when signing a list of words, as well as potential reliability concerns arising from splicing the videos at the immediate onset or offset of a sign.

Two blinded raters first scored the original recordings for clarity (i.e., distortions induced by microphone errors, hesitations, and restarts). We then discarded inaudible recordings and asked the original native speakers to record new versions of initially distorted items. Once auditory quality of the item pool was ascertained, we subsequently eliminated English cognates (i.e., recognizable English root words). We then conducted a post hoc cross-validation procedure to verify the accuracy of all translations. We did so by submitting all of the foreign language translations to Google Translate (Google, Inc.) for back-translation to English. We subsequently eliminated all items that did not include the original English target word within the list of primary translation terms. We also eliminated translations that differed with respect to word sense. For example, a Korean native speaker translated the English abstract noun, *aspect*, as 측면. Google

For the signed stimuli, we manually coded length based on the visual duration of the gestured sign from the onset to offset of hand motion (i.e., rest-to-rest).

2.4. Behavioral testing procedures

We pseudorandomly assigned a subset of three foreign languages to each participant. Within the experimental block, the order of exposure to each of these languages was again randomized.

Participants were first seated in a quiet testing room and fitted with noise-canceling headphones at a computer running E-Prime 2.0 Professional stimulus delivery software (Psychology Software Tools Inc, 2014). Participants first completed 2 min of passive exposure to each foreign language. These sound clips consisted of translations of a standardized narrative sample (Van Riper, 1963) recorded/videotaped by the same native speakers who produced the word stimuli. The purpose of this exposure was two-fold. First, it provided a brief introduction to the unique sound system of the language each participant would soon hear. Second, this exposure was critical for diminishing the influence of English and/or the previously tested foreign language.

After completing a brief familiarization sequence, participants heard or viewed all items from each of their assigned foreign languages in a completely randomized order. We then asked participants to make binary categorical judgments of concreteness. We did so by adapting verbiage for continuous concreteness scales (e.g., rate the extent to which *dog* can be experienced through the senses) to a more explicit, categorical format (for specific wording examples see also Brysbaert et al., 2014; Clark & Paivio, 2004; Coltheart, 1981). Namely, we required participants to signal a Yes/No response via keypress to the question, “Can you see, hear, smell, taste or touch this?” Trials advanced after a 1000 ms interstimulus interval with one short (30 s) break at the midpoint.

2.5. Data analyses

We analyzed word length differences across each language using syllable length and acoustic duration as the dependent measures, controlling for multiple comparisons via Bonferroni correction. The behavioral experiment involved a two-alternative, forced-choice guess (i.e., Can you see, hear, or touch this? Yes/No). We examined response sensitivity, accuracy, and bias using a standard signal detection measure (d-prime) for each language (Wickens, 2002). That is, each participant produced three unique d-prime scores corresponding to their guessing accuracy for each assigned language. We evaluated departures from chance guessing through a parametric one-sample t-test for each language, evaluating whether the d-prime scores across participants were significantly different from zero.

We further analyzed the behavioral guessing results using a mixed-effects logistic regression model. The outcome variable was response accuracy (hit/miss) for each target word. We nested a series of predictors including language, word concreteness, acoustic duration, and the interaction term concreteness * acoustic duration within individual

subjects. Prior to running the full model, we z-transformed and mean-centered acoustic duration; these normalization procedures were necessary to contrast length differences between and within languages with differing baseline word lengths. Participants and items were modeled as random effects, while all other predictors were treated as fixed effects. We analyzed the data via a generalized linear mixed model using maximum likelihood estimation with Laplace Approximation using the R-statistical program (packages lme4 and glmer R Core Team, 2013). Model comparison was undertaken by comparing estimated values for each model of the Akaike Information Criterion, which is a measure of relative information loss (Akaike, 1974).

3. Results

3.1. Abstract–concrete word length differences

Table 1 illustrates differences in acoustic duration and syllable length. Collapsed across all spoken languages, the average acoustic duration of abstract nouns was longer than concrete by a margin of 136 ms [$t(1829) = 13.2, p < .001; \eta^2 = 0.09$]. Abstract and concrete words significantly differed by acoustic duration across Russian, Hebrew, Hindi, Dutch, and ASL. This length discrepancy reversed for ASL, in which concrete signs took longer to communicate by a margin of 893.38 ms [$t(256) = 11.08, p < .001, \eta^2 = 0.32$]. The rank order of the magnitude of these acoustic length discrepancies across the individual languages was ASL > Russian > Dutch > Hindi > Hebrew > Mandarin > Arabic > Korean. Abstract nouns were also longer as gauged by an average syllable length discrepancy of 0.5 syllables [$t(1829) = 11.62, p < .001; \eta^2 = 0.07$]. These length differences were driven by statistically significant syllable discrepancies across Russian, Hindi, Dutch, and Hebrew. The rank order of the magnitude of these syllable length discrepancies across the individual languages was Russian > Dutch > Hindi > Hebrew.

3.2. Behavioral results

Figs. 1 and 2 and Table 2 reflect guessing accuracy across languages. Response accuracies modestly exceeded chance probability across four of eight languages (i.e., ASL, Russian, Dutch, Hindi). Table 2 summarizes the magnitude of the differences between abstract and concrete word length for each language.

Logistic mixed-effects (LME) modeling is problematic with the full dataset because ASL is a distant outlier on both length and accuracy. ASL has both longer average word duration (an average [SD] of 3,115 [782] ms compared to 759 [231] ms in all other languages) and a higher probability of a word being correctly judged as abstract or concrete (62.9% vs. an average of 52.0% in all other languages). We therefore modeled the effects of all spoken languages together. We consider ASL separately below.

The results of LME model for all languages with the exception of ASL are summarized in Tables 3 and 4. The best-fitting model for correctly categorizing a word included

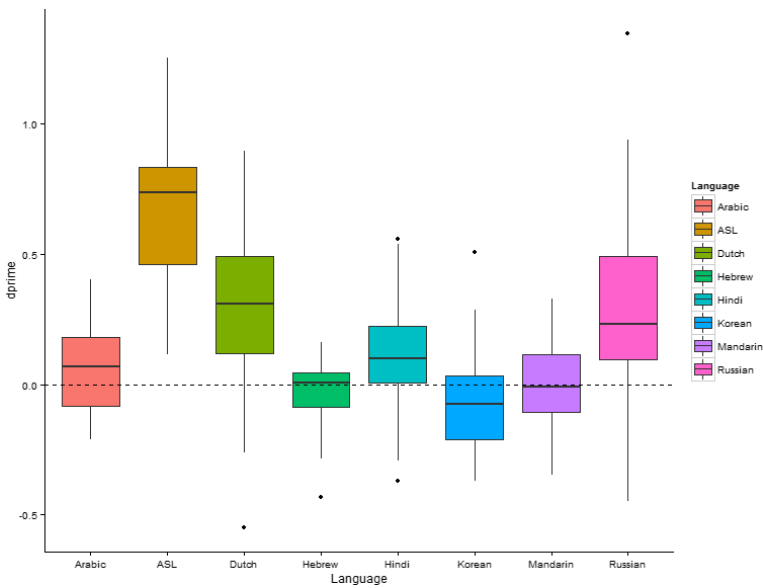


Fig. 1. Abstract-concrete prediction accuracy across languages.

random effects of item and subject, along with a three-way interaction between length, concreteness category, and language.

The key finding is the interaction between length and concreteness category, which is shown graphically in Fig. 3. Words were generally more likely to be judged as concrete when they are shorter in duration. The interaction reflects the fact that abstract words are therefore less likely to be judged correctly (that is, more likely to erroneously categorized as concrete) when they are short than when they are long, whereas concrete words are less likely to be correctly classified as concrete when they are long than when they are short. Words that are the closest to being classified at chance levels are intermediate in length, roughly 700–900 ms long.

In order to be better able to understand the interaction of this concreteness \times duration effect with language, we undertook individual LME analyses of each language, using an analogous model structure to the model used with the entire dataset: random effects of item and subject with interacting fixed effects of concreteness and duration. The results of these eight analyses are presented in Table 5. Seven of the eight languages (all but Mandarin) showed a reliable interaction between concreteness and stimulus duration.

The regression weights in Table 5 show that ASL is very different from the spoken languages with respect to our interests. It is the only language in which concrete words were less likely to be judged concrete and one of only two (with Hebrew) for which longer words were more likely to be judged concrete, as measured by the sign on the weight of the main effect for length. Only the model for ASL produced estimates that longer words were concrete (see Fig. 4). The reversal in ASL was potentially driven by a

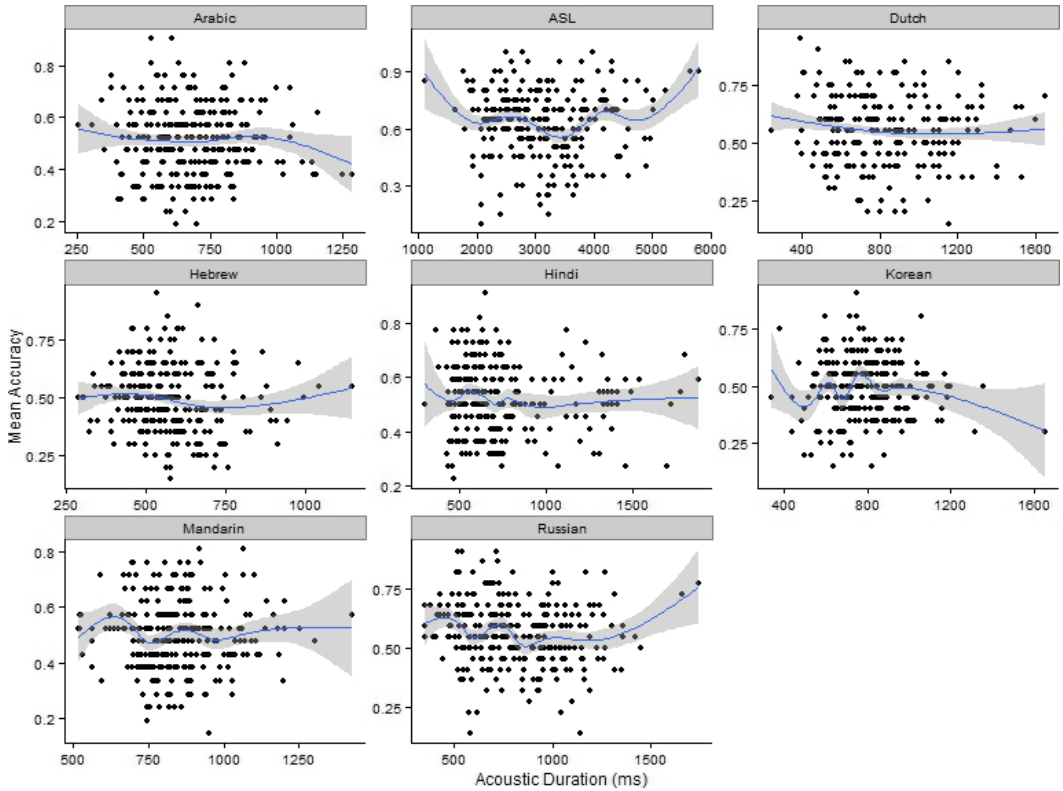


Fig. 2. Word duration by judgment accuracy of concreteness across languages.

Table 2
Prediction accuracy

	D'	%Acc	Min	Max	<i>t</i> -value	<i>df</i>	<i>p</i> -value
Russian	0.32	0.56 (0.07)	0.41	0.75	3.795	21	.001
Hindi	0.11	0.52 (0.04)	0.43	0.61	2.321	21	.030
Korean	-0.05	0.49 (0.04)	0.42	0.61	-0.966	19	.346
Arabic	0.05	0.51 (0.03)	0.46	0.58	1.338	21	.196
Mandarin	-0.01	0.50 (0.03)	0.44	0.56	-0.189	20	.852
Dutch	0.28	0.55 (0.06)	0.39	0.66	3.735	19	.001
Hebrew	-0.04	0.49 (0.03)	0.41	0.53	-1.223	19	.236
ASL	0.70	0.63 (0.05)	0.51	0.74	11.264	19	.000

higher degree of iconicity within concrete signs, a point we revisit in the general discussion to follow.

The distribution of guessing scores was characterized by a cluster of participants ($N = 24$) who performed at chance for all three languages, whereas the remaining participants ($N = 32$) performed above chance in at least one of the three languages to which

Table 3
Logistic mixed-effects model fitting for all spoken languages

Model	Specification	AIC	Improvement
BASE1	(1 Subject)	52,836	N/A
BASE2	(1 Item) + (1 Subject)	52,611	> 1,000,000 x
M1	BASE2 + Concreteness	52,412	> 1,000,000 x
M2	BASE2 + Concreteness + Duration	52,404	55 x
M3	BASE2 + Concreteness × Duration	52,310	> 1,000,000 x
M4	BASE2 + Concreteness × Duration + Language	52,287	> 98,700 x
M5	BASE2 + Concreteness × Duration × Language	52,251	> 1,000,000 x

Notes. Model assessment table across spoken languages comparing models by Akaike Information Criterion value. See Table 4 for specification of the best model M5.

Table 4
Fixed effects from the best-fitting model [M5 in Table 3] across spoken languages

	Estimate	SE	<i>z</i>	<i>p</i>
(Intercept)	-0.17	0.05	-3.18	0.0015
Concreteness	0.36	0.07	5.34	0.00000010
zAcousticDuration	0.28	0.07	4.06	0.000049
Dutch	0.21	0.08	2.43	0.01
Hebrew	-0.32	0.10	-3.20	0.0014
Hindi	0.07	0.07	1.01	0.31
Korean	-0.03	0.07	-0.38	0.70
Mandarin	-0.09	0.08	-1.14	0.26
Russian	0.22	0.08	2.79	0.0053
Concreteness:zAcousticDuration	-0.46	0.09	-5.35	0.000000089
Concreteness:Dutch	-0.12	0.11	-1.09	0.28
Concreteness:Hebrew	0.56	0.13	4.25	0.00
Concreteness:Hindi	-0.15	0.10	-1.46	0.14
Concreteness:Korean	-0.06	0.09	-0.59	0.56
Concreteness:Mandarin	0.16	0.10	1.64	0.10
Concreteness:Russian	-0.16	0.10	-1.58	0.11
zAcousticDuration:Dutch	-0.20	0.08	-2.46	0.01
zAcousticDuration:Hebrew	-0.54	0.13	-4.24	0.000022
zAcousticDuration:Hindi	-0.22	0.07	-2.95	0.0032
zAcousticDuration:Korean	-0.10	0.10	-1.00	0.32
zAcousticDuration:Mandarin	-0.20	0.10	-1.93	0.05
zAcousticDuration:Russian	-0.14	0.08	-1.70	0.09
Concreteness:zAcousticDuration:Dutch	0.24	0.12	2.07	0.04
Concreteness:zAcousticDuration:Hebrew	0.82	0.15	5.37	0.000000080
Concreteness:zAcousticDuration:Hindi	0.23	0.12	1.96	0.05
Concreteness:zAcousticDuration:Korean	0.12	0.12	0.99	0.32
Concreteness:zAcousticDuration:Mandarin	0.38	0.13	2.89	0.0038
Concreteness:zAcousticDuration:Russian	0.04	0.11	0.38	0.71

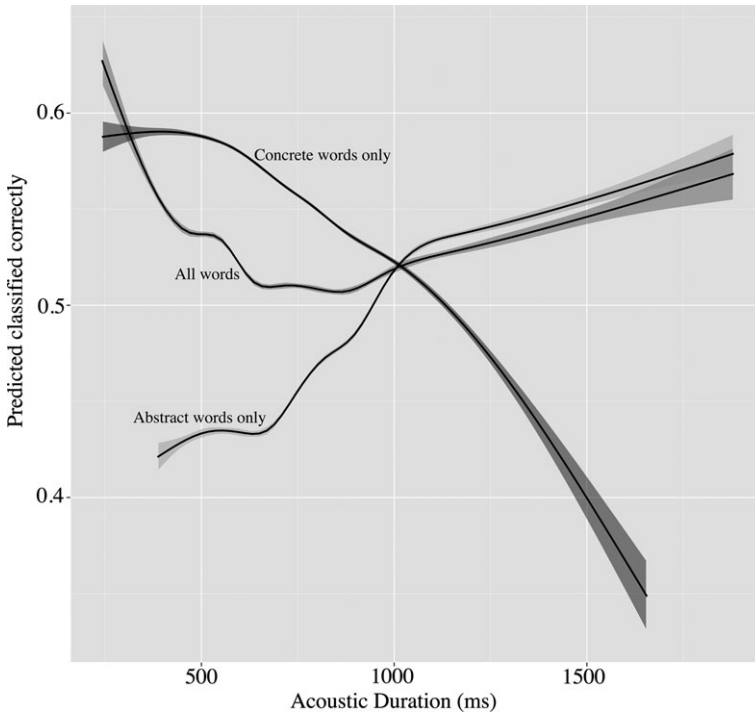


Fig. 3. GAM-smoothed estimated effects of acoustic duration on classification for spoken words only. Results reflect the best logistic mixed effects (LME) model from Table 3 [M5, Table 4], shown with 95% confidence bounds.

Table 5
Summary of logistic mixed effect modeling by individual language

Language	Concreteness		Duration		CNC * Duration	
	Estimate	<i>p</i>	Estimate	<i>p</i>	Estimate	<i>p</i>
Arabic	0.50	< 2e -16	0.20	1.10E-05	-0.34	7.50E-09
ASL	-0.38	0.002	-0.03	NS	0.26	0.04
Dutch	0.17	NS	0.10	NS	-0.27	0.02
Hebrew	0.63	< 2e -16	-0.15	0.01	0.20	0.003
Hindi	0.21	0.005	0.08	0.04	-0.30	0.003
Korean	0.24	0.0002	0.14	0.009	-0.25	0.0001
Mandarin	0.48	< 2e -16	0.05	NS	-0.05	NS
Russian	0.10	NS	0.15	0.002	-0.45	4.27E-08

Note. Reliable interactions between concreteness and stimuli duration are shown in bold type.

they were assigned. There are a number of possible explanations for this trend. The most effective strategy in executing this task requires that participants spontaneously invoke their metalinguistic knowledge of formal markers of abstract/concrete words in English. That is, participants can strategically apply a word length heuristic by sampling what they

know of the probabilistic pattern of English (e.g., *independence*, *consolidation*, and *honesty* are abstract words, whereas *cat*, *dog*, and *desk* are concrete words). When participants adopt word length as a primary strategy, they can extrapolate from English to each of the languages they were randomly assigned. In contrast, adopting no strategy or being assigned a language in which there is no baseline abstract–concrete length discrepancy would produce accuracies equivalent to random guessing.

We conducted a post hoc analysis of the chance responders by examining the distribution of languages they were randomly assigned to make guessing judgments. For languages with marked differences in length via corpus analysis between abstract/concrete words (Russian, Hebrew, Hindi, Dutch, and ASL), a word length heuristic should produce guessing accuracies above chance. In contrast, a word length guessing heuristic would not be effective for languages where concreteness is not marked by length. Participants were randomly assigned three languages for prediction. Inspection of the distribution of assigned languages revealed that the chance group was more often assigned languages in which there was no length discrepancy (i.e., Arabic, Korean, and Mandarin) (46% of the chance group vs. 31% in the responder group). A second potential source of individual differences is baseline foreign language expertise. At intake, we assessed foreign language expertise via a questionnaire to ensure that participants did not have prior knowledge of the foreign languages to which they were assigned. We evaluated whether multilingualism was independent of foreign language abstract/concrete guessing perfor-

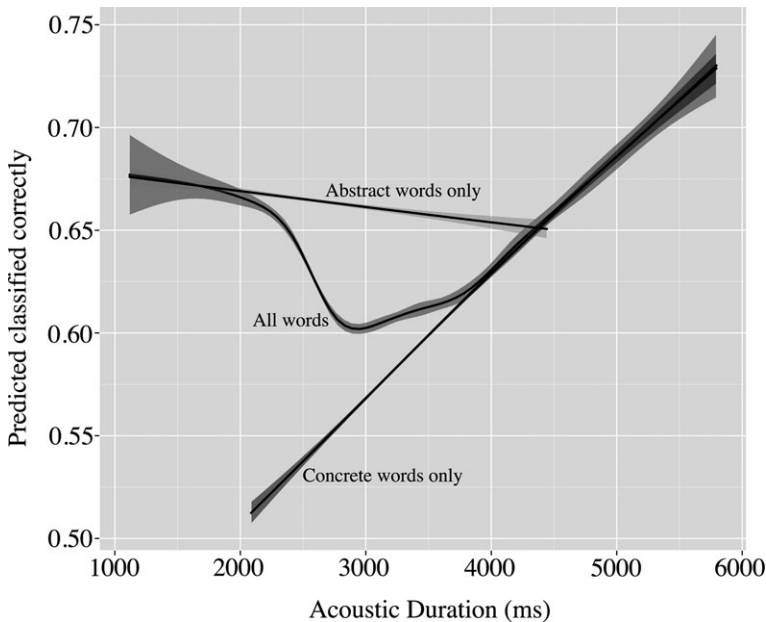


Fig. 4. GAM-smoothed estimated effects of acoustic duration on correct concreteness classification for ASL words only. Results reflect the best logistic mixed effects (LME) model from Table 5, shown with 95% confidence bounds.

mance by generating a two-way contingency table of cell counts from the original sample of 56 participants. We binarized the column variable as multilingualism (i.e., monolingual or multilingual) and the row variable as performance on the concrete–abstract guessing experiment (chance vs. responder). This non-parametric contrast demonstrated that multilingualism was independent of prediction accuracy on the concreteness judgment task [$\chi^2(1) = 0.14, p > .05$].

4. General discussion

Pattern induction is an essential component of language processing. Its effects are evident in early infancy in service of adaptively signaling word boundaries, assigning syntactic roles, and mapping sounds to concepts (Nygaard, Cook, & Namy, 2009; Saffran & Thiessen, 2003; St. Clair, Monaghan, & Christiansen, 2010). It is now reasonably well accepted that humans exploit regularities in the sound systems of our native languages to speed the efficiency of word recognition and to mark the particular role that a word plays in running speech or text. Our aim in this work was to demonstrate that similar violations of arbitrary symbolism also exist with respect to the relation between word form and word meaning (i.e., concreteness) and that these markers may transcend linguistic boundaries. For five of the eight languages we analyzed here with respect to word length (Russian, Hebrew, Hindi, Dutch, and ASL), this appears to be the case. When considered in conjunction with our earlier corpus analyses of English, this represents a sizeable number of speakers for whom abstract and concrete concepts are formally marked. For scope, these languages alone approach 1 billion speakers (Lewis, Simons, & Fennig, 2015).

We must first acknowledge several methodological limitations before interpreting the results. Acoustic length and total syllables are crude predictors of word form, and it is unclear whether native speakers in each of the respective languages we analyzed show sensitivity to length. Our rationale for analyzing word length was two-fold. First, length appears to be among the primary drivers of nonword concreteness judgments in English speakers (Reilly et al., 2012). Second, many languages have not yet been exhaustively cataloged with an inventory of psycholinguistic norms. Thus, our results reflect only a coarse proof of concept that acoustic-phonetic differences potentially mark abstract and concrete words across languages other than English. Far greater specificity is necessary to delineate such markers within their native linguistic contexts.

Another potential limitation applies to the primary semantic variable of interest. Many language researchers draw a clear distinction between the psycholinguistic constructs of concreteness (the extent to which a word can be experienced through the senses) and imageability (the extent to which a word can evoke a mental image) (Kousta, Vigliocco, Vinson, Andrews, & Del Campo, 2011). Our methods do not permit decorrelation of these two constructs. Nevertheless, our aims are not necessarily compromised by this limitation. Concreteness and imageability likely share many qualitative semantic processing attributes (e.g., associative organization, emotion and magnitude as salient features for abstract words), which formal cues may facilitate access to (Crutch, Troche, Reilly, &

Ridgway, 2013; Reilly et al., 2016). Caveats acknowledged, we turn to theoretical interpretation of the findings.

4.1. *Relations between word length, concreteness, and information content*

The current results demonstrate that word length and concreteness are correlated constructs across some of the world's most widely spoken languages. Yet this length effect is not universal. Several of the languages we queried showed no clear length discrepancies between abstract and concrete words. One potential explanation is that the relationship between information content and word length is not a language universal. Another account relates to morphology as a moderating variable. One of the primary drivers of word length inflation related to noun abstractness in English is derivational morphology. English derives many of its abstract words through inflecting concrete stems (e.g., *friend* → *friendliness*). Affixation often conveys abstractness while simultaneously increasing word length and a variety of other phonological factors such as syllable stress placement and neighborhood density. In our corpus analyses, the languages that most robustly demonstrated concrete–abstract word length differences (e.g., Russian, Dutch) share this property of English morphology; that is, word stems are affixed.

One might look to the morphological structures of languages that did not show an abstract–concrete length discrepancy in corpus analyses (i.e., Mandarin, Korean, and Arabic) for an explanation.

Mandarin lacks inflectional morphology and is, therefore, less likely to produce abstract words via affixation (i.e., most morphemes are monosyllabic). Although Korean's system of derivational morphology is more diverse than Mandarin, most Korean abstract words were borrowed from Chinese. Finally, Arabic morphology substantively differs from English in that derivation is not achieved through the addition of prefixes and suffixes, but instead through a system of introflexion where vowel internal constituents are altered within root forms. Thus, morphology is one potential moderating factor in accounting for cross-linguistic relations between word length and a range of other lexical or semantic variables (e.g., information content or concreteness).

The fact that not all languages marked concreteness by length poses a challenge for the account of length–concreteness (or information content) as a true language universal. Another potential challenge regards the patterns of ASL we observed. This was the single language that elicited a reversal of the typical concreteness effect (i.e., concrete signs took longer to unfold than abstract). One explanation for this pattern is that concrete signs are more likely to show iconicity and that the production of iconic signs inflates word length (see also Vinson, Thompson, Skinner, & Vigliocco, 2015). Another possibility is that signed and spoken languages optimize the relation between length and information content in different ways.

Lewis, Sugarman, and Frank (2014) proposed an alternate perspective to that advanced by Piantadosi and colleagues, arguing that word lengths are optimized for information complexity. In this work, the authors cite words such as *brick* and *engine* as exemplifying varying degrees of informational complexity. Adults' information complexity ratings of a

corpus of words ($N = 500$) bore out this prediction in that word length was strongly positively correlated with information complexity ($R = 0.66$). In a second experiment, Lewis and colleagues examined preferential mapping between nonwords of varying lengths (e.g., *tupa* vs. *tupabugorn*) and geometric shapes (geons) of variable visual complexity. Again, participants selected the complex shapes as matches for longer nonword names. Although Lewis and colleagues did not explicitly consider abstractness as a metric of information complexity, their hypothesis has special relevance for our prediction that word length potentially confers a concrete object bias during early language development. That is, infants may show bias for mapping short and uninflected word forms to concrete objects, while reserving longer and/or more acoustically complex words for abstract concepts.

4.2. Concluding remarks

A growing body of literature supports the claim that listeners use distributional cues to aid in language learning by using word length and syllable stress placement to rapidly assign syntactic roles during online language comprehension (Kelly, 1992; Monaghan, Christiansen, & Chater, 2007; Reali & Christiansen, 2005). In this respect, distributional cues can facilitate sentence comprehension by tuning the listener's attention to individual syntactic elements. In the current work, we evaluated the possibility that similar distributional cues might also inform listeners about word concreteness, a key semantic distinction in natural language processing. There exist a number of potential advantages afforded by such a concreteness processing heuristic both in terms of early word learning and in the mature language systems of adults. However, it is also clear that much remains to be learned about the scope and universality of this violation of linguistic arbitrariness.

Acknowledgments

We are grateful to Amelia Wisniewski-Barker for her assistance with translation cross-validation, Daniel Mirman for his guidance on multilevel modeling procedures, and Sameer Ashaie for assistance with acoustic analysis. This work was funded by US Public Health Service grant R01 DC013063 (JR).

Note

1. The MRC concreteness norms are widely used in psycholinguistic research; however, the ratings are now over 35 years old. Word frequency norms in English have radically shifted over this period, reflecting the natural evolution of language use (Brysbaert & New, 2009). It is unclear whether concreteness is subject to a similar shift. We examined stability of the MRC concreteness norms for our dataset relative to a more contemporary database of concreteness norms (Brysbaert, Warriner,

& Kuperman, 2014). The Pearson bivariate correlation between these two datasets was $R = .95$. Moreover, none of the original items were misclassified (e.g., abstract as concrete) using more contemporary norms. The strength of this relationship indicates relative stability of concreteness across time. Stimuli along with their respective norms from both MRC and Brysbaert et al. are freely available for download at <http://www.reilly-coglab.com/data/>.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. doi:10.1109/TAC.1974.1100705.
- Brysbaert, M., Warriner, A., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46(3), 904–911. doi:10.3758/s13428-013-0403-5
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.
- Clark, J. M., & Paivio, A. (2004). Extensions of the Paivio, Yuille, and Madigan (1968) norms. *Behavior Research Methods, Instruments & Computers*. Special Issue: Web-Based Archive of Norms, Stimuli, and Data: Part 1, 36(3), 371–383.
- Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, 33, 497–505.
- Crutch, S. J., Troche, J., Reilly, J., & Ridgway, G. R. (2013). Abstract conceptual feature ratings: The role of emotion, magnitude, and other cognitive domains in the organization of abstract conceptual knowledge. *Frontiers in Human Neuroscience*, 7, 186. doi:10.3389/fnhum.2013.00186
- Durieux, G., & Gillis, S. (2001). Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition, Volume 1. In J. Weissenborn & B. Höhle (Eds.), *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition* (1st ed.) (pp. 189–229). Amsterdam: John Benjamins.
- Eprime 2.0 Professional [Computer software]. (2015). Sharpsburg, PA: Available from <https://www.pstnet.com/eprime.cfm>
- Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, 99(2), 349–364.
- Kessler, B., & Treiman, R. (1997). Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language*, 37(3), 295–311.
- Koriat, A. (1975). Phonetic symbolism and feeling of knowing. *Memory & Cognition*, 3(5), 545–548. doi:10.3758/BF03197529
- Kousta, S.-T. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: why emotion matters. *Journal of Experimental Psychology General*, 140(1), 14–34. doi:10.1037/a0021446
- Langenmayr, A., Gozutok, M., & Gust, J. (2001). Remembering more nouns than verbs in lists of foreign-language words as an indicator of syntactic phonetic symbolism. *Perceptual and Motor Skills*, 93(3), 843–850.
- Lewis, M. P., Simons, G. F., & Fennig, C. D. (2015). Summary by language size. In *Ethnologue: Languages of the World* (19th ed.) (online version). Dallas, TX: SIL International. Retrieved February 22, 2016 from <http://www.ethnologue.com>
- Lewis, M., Sugarman, E., & Frank, M. (2014). The structure of the lexicon reflects principles of communication. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 845–850).
- Love, J. (2014). English vs. English: On the concrete and the abstract, the Germanic and the Latinate. Retrieved March 15, 2015, from <https://theamericanscholar.org/english-vs-english/#.VR6yelPF8Xk>

- Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition*, *96*(2), 143–182.
- Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*, *55*(4), 259–305.
- Monaghan, P., Christiansen, M. H., & Fitneva, S. A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology. General*, *140*(3), 325–347. doi:10.1037/a0022924
- Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society B*, *369*(20130299), 1–12.
- Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, *112*(1), 181–186.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Reply to Reilly and Kean: Clarifications on word length and information content. *Proceedings of the National Academy of Sciences*, *108*(20), E109–E109. doi:10.1073/pnas.1103550108
- R Core Team. (2013). R: A language and environment for statistical computing. Retrieved from <http://www.r-project.org/>
- Reali, F., & Christiansen, M. H. (2005). Uncovering the richness of the stimulus: Structure dependence and indirect statistical evidence. *Cognitive Science*, *29*(6), 1007–1028.
- Reilly, J. (2005). A tale of two imageabilities: An interaction of sound and meaning in natural language perception. Dissertation Abstracts International: Section B: The Sciences and Engineering, US. Retrieved from <https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2005-99024-143&site=ehost-live>
- Reilly, J., Cross, K., Troiani, V., & Grossman, M. (2007). Single-word semantic judgements in semantic dementia: Do phonology and grammatical class count? *Aphasiology*, *21*(6-8), 558–569. doi:10.1080/02687030701191986
- Reilly, J., & Kean, J. (2007). Formal distinctiveness of high- and low-imageability nouns: Analyses and theoretical implications. *Cognitive Science*, *31*(1), 157–168. doi:10.1080/03640210709336988
- Reilly, J., & Kean, J. (2011). Information content and word frequency in natural language: Word length matters. *Proceedings of the National Academy of Sciences USA*, *108*(20), doi:1103035108 [pii]10.1073/pnas.1103035108
- Reilly, J., Peelle, J. E., Garcia, A., & Crutch, S. J. (2016, in press). Linking somatic and symbolic representation in semantic memory: The Dynamic Multilevel Reactivation Framework. *Psychonomic Bulletin and Review*.
- Reilly, J., Westbury, C., Kean, J., & Peelle, J. E. (2012). Arbitrary symbolism in natural language revisited: When word forms carry meaning. *PLoS ONE*, *7*(8), e42286. doi:10.1371/journal.pone.0042286
- Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology*, *39*(3), 484–494.
- Saussure, F. de (1916). *Cours de linguistique generale*. Lausanne, Paris: Payot.
- Shi, R., Morgan, J. L., & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *Journal of Child Language*, *25*(01), 169–201.
- Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, *72*(2), B11–B21. doi:10.1016/S0010-0277(99)00047-5
- St. Clair, M. C., Monaghan, P., & Christiansen, M. H. (2010). Learning grammatical categories from distributional cues: Flexible frames for language acquisition. *Cognition*, *116*(3), 341–360.
- Swaab, T. Y., Boudewyn, M. A., Long, D. L., Luck, S. J., Kring, A. M., Ragland, J. D., Ranganath, C., Lesh, T., Niendam, T., Solomon, M., Mangum, G. R., & Carter, C. S. (2013). Spared and impaired spoken discourse processing in schizophrenia: Effects of local and global language context. *Journal of Neuroscience*, *33*(39), 15578–15587. doi:10.1523/JNEUROSCI.0965-13.2013
- Van Riper, C. (1963). *Speech correction* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

- Vinson, D., Thompson, R. L., Skinner, R., & Vigliocco, G. (2015). A faster path between meaning and form? Iconicity facilitates sign recognition and production in British Sign Language. *Journal of Memory and Language*, 82, 56–85. doi:10.1016/j.jml.2015.03.002
- Westbury, C., & Moroschan, G. (2009). Imageability x phonology interactions in lexical access. *The Mental Lexicon*, 4(1), 115–145.
- Wickens, T. D. (2002). *Elementary signal detection theory*. New York: Oxford University Press.
- Zipf, G. (1949). *Human behavior and the principle of least effort*. New York: Addison Wesley.