PRESSURE FOR AND AGAINST NON-ARBITRARINESS: EVIDENCE FROM PHONESTHEMES

DAVID A. HASLETT*[1], ZHENGUANG G. CAI[1,2]

*Corresponding Author: david.a.haslett@gmail.com
[1]Dept. of Linguistics and Modern Languages, The Chinese University of Hong Kong
[2]Brain and Mind Institute, The Chinese University of Hong Kong

Words with related meanings tend to sound dissimilar, and this arbitrariness allows for large vocabularies that make nuanced distinctions (like *glint* versus *shine*). Languages violate arbitrariness in constrained ways, such as patterns in form that provide cues indicating word class (e.g., English nouns and verbs tend to be stressed on different syllables; Kelly & Bock, 1988), but phonesthemes (i.e., sound clusters associated with semantic features, like /gl/ with light, as in *glint* and *gleam*) more directly defy arbitrariness by pertaining to word meaning proper. We provide evidence from two experiments that phonesthemes influence how people process word meanings (e.g., they're slower to decide that *glove* is unrelated to light, compared to a control like *mitten*) and that phonesthemes aid in the recognition of visually degraded words. These findings support that, in English, non-arbitrariness interferes with word meaning access and that the lexicon evolves to balance competing needs: Arbitrariness maintains subtle distinctions, but non-arbitrariness can facilitate recognition in noisy environments.

1. Introduction

Word forms tend to be arbitrarily assigned to meanings (de Saussure, 1983; Hockett, 1960). For example, *bench* and *sofa* refer to similar concepts but sound nothing alike. Evidence from simulations and novel word learning suggests that this arbitrariness occurs not by chance but in response to communicative pressure. When people need to distinguish referents, cultural evolution increases arbitrariness (Kirby, Tamariz, Cornish, & Smith, 2015), which is consistent with the finding that arbitrariness makes it easier to acquire a large vocabulary (Gasser, 2004) and to learn specific word meanings (Monaghan, Christiansen, & Fitneva, 2011). There is selective pressure for an arbitrary lexicon. But non-arbitrariness can be advantageous, too, by facilitating categorization. For example, people read nouns in sentences faster if those nouns sound more like other nouns than like verbs (Farmer, Christiansen, & Monaghan, 2006). Importantly, such patterning in form (i.e., systematicity) pertains to low-level semantics, like word class (Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015), and so is functionally compatible with arbitrariness at the level of individual word meanings, like *bench* and *sofa* (Monaghan et al., 2011).

However, words that overlap substantially in meaning (not just that are members of the same broad category) sometimes sound similar, as in phonesthemic words. Phonesthemes are sound clusters associated with semantic features, like /gl/ with light (e.g., *glimmer* and *glaze*; Firth, 1930), and there is speculation that words which contain phonesthemes but lack the corresponding features, such as *gland* and *glove* (unrelated to light), are apt to acquire that feature (Bolinger, 1965; Blust, 2003). For example, English words beginning with the phonestheme /fl/ (as in *flap* or *flow*) have grown more likely to relate to "moving through the air" (Smith, 2016). Such semantic changes defy the tendency for the lexicon to evolve to be more arbitrary or to limit non-arbitrariness to low-level semantics. Novel word learning experiments are consistent with this corpus analysis. People tend to pair novel words containing phonesthemes (e.g., *glep*, like *glow* and *glimmer*) with definitions containing semantic features associated with that phonestheme (e.g., something related to light), and when asked to invent labels for such definitions, people are more likely to produce novel words containing the corresponding phonesthemes (Hutchins, 1998; Abelin, 1999; Magnus, 2001).

Phonesthemes might motivate semantic change in an efficient lexicon even if the resulting non-arbitrariness hinders our ability to distinguish similar concepts (like *glint* versus *gleam*). Consider how ambiguity is sometimes advantageous: Languages reuse sound sequences that are easier to produce, and while this results in homophones, it eases the burden of articulation on speakers (Piantadosi, Tily, & Gibson, 2012). Similar-sounding words with related meanings might facilitate comprehension, too. For example, Bergen (2004) found that phonesthemic primes sped up recognition of targets containing the same phonestheme (*glow* → *glitter*) more than formal (*glove*) or semantic (*shine*) primes did.

There is suggestive evidence that phonesthemes motivate semantic change and facilitate word recognition, but those studies rely on correlations and single phonesthemes (Smith, 2016; Blust, 2003), on novel words rather than known words learned in rich, realistic contexts (Hutchins, 1998; Magnus, 2001; Abelin, 1999), and on priming rather than the ambient influence of the structure of the lexicon (Bergen, 2004). And while previous studies have found that non-arbitrariness facilitates processing of part of speech (e.g., Farmer et al., 2006), those experiments do not apply to words that overlap in high-level conceptual semantics. We therefore conducted two experiments to provide evidence that phonesthemes affect how people process the meanings of English words and that the resulting non-arbitrariness facilitates recognition in some conditions. All data and scripts are available at osf.io/sna92.

2. Experiment 1

We first investigated whether people associate a word containing a phonestheme (e.g., /gl/) with the corresponding semantic feature (light) even when the word is unrelated to that feature (*glove*). In an online experiment (implemented on Gorilla.sc; Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2018), 57 adult native English speakers (recruited from Prolific.co) decided whether a cue word associated with a phonestheme's feature (e.g., *shine*, with light) was semantically related to a target that either contains that phonestheme (*glove*) or is a control (*mitten*). We manipulated the target (pseudo-phonesthemic vs. control), for a one-factor, two-level design, manipulated within items and within subjects. In each trial, the cue (e.g., *shine*) was unrelated to the target (*glove*/*mitten*), and the pseudo-phonesthemic and the control targets were synonyms matched on semantic distance from the cue, as measured by a distributional semantic model (computed using the LSAfun package in R; Günther, Dudschig, & Kaup, 2015). The targets were also matched on formal similarity to cues (Levenshtein distance), word frequency, age of acquisition, concreteness rating, bigram frequency, orthographic neighbourhood density (OLD20), and length (all from the English Lexicon Project; Balota et al., 2007). We divided 126 items (unrelated cue–target pairs) and 126 fillers (related pairs) into six lists, each comprising 21 pseudo-phonesthemic targets, 21 control targets, and 42 fillers. We assigned each participant to one list. To incentivize fast and accurate responses, participants had only two seconds to make each decision, with three seconds of feedback following timeouts and incorrect responses. We excluded participants with accuracy below 75% (seven of 64 participants) and trials faster than 200ms (one trial). Linear mixed effects modelling conducted using the lmerTest package in R (Kuznetsova, Brockhoff, & Christensen, 2017) indicates that pseudo-phonesthemic targets were judged to be unrelated to cues significantly more slowly than controls were (see Tables 1 and 2 and Figure 1). Importantly, semantic relatedness decisions involve word meaning access, not just recognition or low-level categorization, so these effects stem from a phenomenon that is distinct from statistical regularities within a broad category, as Dingemanse et al. (2015) define systematicity. Phonesthemes pertain to word meaning proper and so are incompatible with arbitrariness.

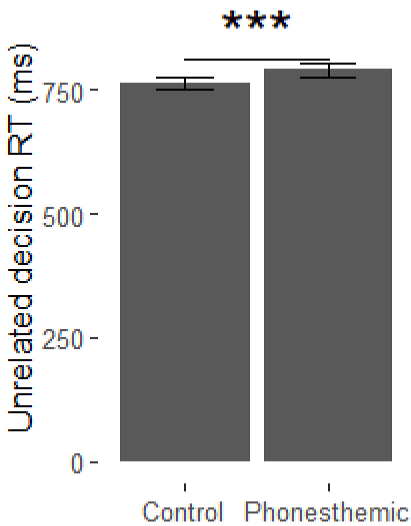

Figure 1: Semantic relatedness decision RT (correct unrelated decisions only) for targets containing phonesthemes (*glove*) versus arbitrary controls (*mitten*) following unrelated cues (*shine*) in Experiment. 1. Error bars indicate 95% confidence intervals.

Table 1: Comparison of RT by target type (sum contrast coded: pseudo-phonesthemic = 1, control = -1) for correct unrelated semantic relatedness decisions in Experiment 1, fitting the random intercepts of subject and item, with target type added as a random slope to the random intercept of subject, as justified by forward model comparison (cf. Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017).

	β	*SE*	*t*	*p*
Intercept	775.18	16.21	47.8	< .001
Pseudo-phonesthemic	13.59	3.79	3.6	< .001

Table 2: Comparison of accuracy (proportion of unrelated decisions) by target type (sum contrast coded: pseudo-phonesthemic = 1, control = -1) for semantic relatedness decisions in Experiment 1, fitting the random intercepts of subject and item, with target type added as a random slope to both random intercepts, as justified by forward model comparison.

	β	*SE*	*z*	*p*
Intercept	2.95	0.17	17.3	< .001
Pseudo-phonesthemic	-0.16	0.15	1.1	.284

3. Experiment 2

We next investigated whether phonesthemes facilitate word recognition in noisy conditions. In another online experiment, 59 participants made lexical decisions about phonesthemic words (e.g., *glitter*, related to light and containing the /gl/ phonestheme) or controls (*sparkle*) that were presented clearly or were visually degraded (i.e., the first and second half of the word flickering separately on the screen, each half seen twice for 100ms). It was a 2 (target: phonesthemic vs.

control) x 2 (presentation: clear vs. degraded) design, with both factors manipulated within subjects and presentation manipulated within items. Target type was manipulated between items, but phonesthemic targets were paired with synonymous control targets and, as in Experiment 1, were matched on word frequency, AoA, concreteness, OLD20, bigram frequency, and length. We evenly divided 216 items (108 phonesthemic and 108 control targets) and 216 non-word fillers into eight lists and assigned each participant to one list. We again incentivized fast, accurate responses with a time limit and penalties, with the same exclusion criteria as Experiment 1, excluding five of 64 participants. Logistic and linear mixed effects modelling indicates that decisions were significantly faster and more accurate for phonesthemic than control words and were faster for degraded words that contained phonesthemes (i.e., a significant interaction of target type and presentation; see Tables 3 and 4 and Figure 2). Note that while the main effect of target type is significant in both analyses, the simple effect of target type (i.e., when using treatment coding, with control targets and clear presentation as the reference levels, thereby analyzing the effect of target type in the clear condition; cf. Brehm & Alday, 2022) is not significant in the RT analysis, which suggests that the advantage for phonesthemes in the degraded condition is the real source of the main effect on RT (see supplementary scripts on osf.io/sna92).

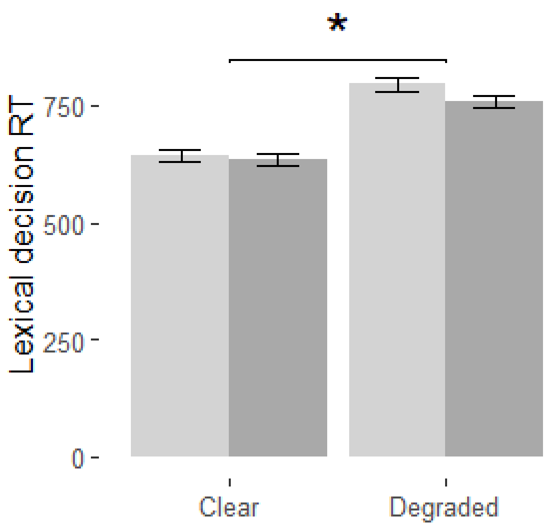

Figure 2: Lexical decision RT (correct decisions only) for phonesthemic targets (dark grey; e.g., *glitter*) versus controls (light grey; e.g., *sparkle*) presented clearly or visually degraded in Experiment 2. Error bars indicate 95% confidence intervals.

Table 3: Comparison of RT by target type (sum contrast coded: phonesthemic = 1, control = -1) and presentation type (visually degraded = 1, clear = -1) for correct lexical decisions in Experiment 2, with presentation type added as a random slope to the random intercepts of subject and item, as justified by forward model comparison.

	β	*SE*	*t*	*p*
Intercept	710.14	14.24	49.9	< .001
Degraded presentation	69.36	4.03	17.2	< .001
Phonesthemic target	-11.52	4.18	2.8	.006
Degraded : Phonesthemic	-6.71	2.90	2.3	.022

Table 4: Comparison of accuracy by target type (sum contrast coded: phonesthemic = 1, control = -1) and presentation type (visually degraded = 1, clear = -1) in Experiment 2, fitting the random intercepts of subject and item, with presentation type added as a random slope to the random intercept of item, as justified by forward model comparison.

	β	*SE*	*z*	*p*
Intercept	2.95	0.17	16.9	< .001
Degraded presentation	-0.47	0.12	4.0	< .001
Phonesthemic target	0.44	0.13	3.3	.001
Degraded : Phonesthemic	-0.04	0.07	0.5	.583

4. Discussion and Conclusion

We found that phonesthemes affect how people process word meanings. In Experiment 1, participants were slower to decide that words containing phonesthemes (e.g., *glove*, compared to *mitten*) were unrelated to meanings typically associated with that phonestheme (e.g., *shine*). We also found that phonesthemes confer an advantage in noisy conditions. In Experiment 2, people made faster lexical decisions about visually degraded words when those words were phonesthemic than when they were controls (e.g., *glitter* versus *sparkle*).

Experiment 1 supports that non-arbitrariness can interfere with word meaning access (e.g., Gasser, 2004; Monaghan et al., 2011) and is consistent with studies suggesting that phonesthemes, and similar-sounding words in general, motivate semantic change (Smith, 2016; Blust, 2003; Haslett & Cai, 2021). As form influences interpretation, subtle shifts in word meanings can accumulate. At first glance, these two implications seem to be at odds. If non-arbitrariness impedes processing, then it should decrease rather than increase over generations, assuming the lexicon evolves for the sake of efficiency (e.g., Gibson et al., 2019). However, Experiment 2 can help resolve this apparent inconsistency. While non-arbitrariness slows access to the precise meaning of a word, it facilitates recognition in challenging conditions. Compare this to the polysemy advantage in lexical decisions. People are faster to recognize words with multiple related senses (e.g., *plant* means both “situate” and “organism with roots”) than homonyms with multiple unrelated meanings (e.g., *bark* in the context of a dog versus a tree), perhaps because related senses share a broad attractor basin in semantic space (Rodd, Gaskell, & Marslen-Wilson, 2002, 2004). We found that words which sound similar (rather than identical) to words with related meanings confer the same sort of advantage when words are harder to read. It is possible, though entirely speculative, that words' overlapping in both form and meaning is reflected in the organization of semantic space, providing a conglomerate target when encountering those words in noisy conditions.

These are small effects (a 28ms disadvantage for pseudo-phonesthemic words in Experiment 1, a 35ms advantage for phonesthemic words in the degraded condition in Experiment 2), but they are on par with comparable studies. For example, Rodd et al. (2002) reported a 29ms disadvantage for homonyms, compared to unambiguous words, and a 33ms advantage for polysemous words with many senses, compared to those with few senses. Are differences on the order of tens of milliseconds sufficient to drive language evolution, leading people to favour some words over others or to alter words' sounds or meanings? Such minute advantages, even if imperceptible to the mind's naked eye, are consistent with what Christiansen and Chater (2016) call the now-or-never bottleneck: We need to process words fast in order to shuttle information along. Still, how fast is fast enough is an open empirical question, one that must weigh the value of word recognition against distinguishing word meanings.

Previous research has explored the competing pressures of word meaning individuation and word learning (e.g., Monaghan et al., 2011; Kirby et al., 2015). Here, we have provided evidence that patterns in sound and meaning influence interpretation and can thereby impede word meaning access but that those patterns also facilitate recognition in noisy conditions. There are communicative pressures both for and against phonesthemes, depending on depth of processing.

References

Abelin, Å. (1999). Phonesthemes in Swedish. In *Proceedings of XIV International Conference of Phonetic Sciences*, *99*, 1333-1336.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior research methods, 52*(1), 388-407.

Balota, D., et al. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445-459.

Bergen, B. (2004) The Psychological Reality of Phonesthemes. *Language, 80*(2), 291-311.

Blust, R. (2003). The Phonestheme ŋ- in Austronesian Languages. *Oceanic Linguistics, 42*(1), 187-212.
Bolinger, D. (1965). *Forms of English: Accent, Morpheme, Order*. Harvard UP.
Brehm, L., & Alday, P. M. (2022). Contrast coding choices in a decade of mixed models. *Journal of Memory and Language*, 125.
Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and brain sciences*, *39*.
Farmer, T., Christiansen, M., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences, 103*(32), 12203-122208.
Firth, J. R. (1930). *Speech*. Oxford UP.
Gasser, M. (2004). The origins of arbitrariness in language. *Proceedings of the Annual Conference of the Cognitive Science Society, 26*.
Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun – An R package for computations based on Latent Semantic Analysis. *Behavior Research Methods, 47*, 930-944.
Haslett, D. & Cai, Z. (2021). New neighbours make bad fences: Form-based semantic shifts in word learning. *Psychonomic Bulletin & Review*.
Kelly, M. H., & Bock, J. K. (1988). Stress in time. *Journal of experimental psychology: human perception and performance*, *14*(3), 389.
Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition, 141*, 87-102.
Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1-26.
Magnus, M. (2001.) *What's in a word? Evidence for phonosemantics*. Dissertation, University of Trondheim.
Monaghan, P., Christiansen, M. H., & Fitneva, S. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General, 140*(3), 325-347.
Piantadosi, S., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280-291.
Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of memory and language, 46*(2), 245-266.
Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive science, 28*(1).
de Saussure, F. (1983). *Course in General Linguistics*. Open Court.
Smith, C. (2016). Tracking semantic change in fl- monomorphemes in the *Oxford English Dictionary*. *Journal of Historical Linguistics*, *6*(2), 165-200.