Journal of Memory and Language xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Weighing up the evidence for sound symbolism: Distributional properties predict cue strength

Chris Westbury [a,*], Geoff Hollis [a], David M. Sidhu [b], Penny M. Pexman [b]

[a] Department of Psychology, University of Alberta, P220 Biological Sciences Building, Edmonton, AB T6G 2E9, Canada
[b] Department of Psychology, University of Calgary, Room AD255, 539 Campus Place N.W., Calgary, AB T2N 1N4, Canada

ARTICLE INFO

Article history:
Received 27 September 2016
Revision received 11 September 2017
Available online xxxx

Keywords:
Sound symbolism
Semantics
Meaning
Psycholinguistics
Experimental psychology

ABSTRACT

It is well-established that there are relationships between word meaning and certain letters or phonemes, a phenomenon known as *sound symbolism*. Most sound symbolism studies have relied on a small stimulus set chosen to maximize the probability of finding an effect for a particular semantic category. Attempts to assign weights to sound symbolic cues have been limited by a methodology that has relied largely on forced contrast judgments, which do not allow systematic assignment of weights on the sound symbolic cues. We used a novel research approach designed to allow us to assign weights to sound symbolic cues. Participants made binary yes/no judgments about thousands of randomly-generated nonwords, deciding if they were good examples for each of 18 different semantic categories. Formal cues reliably predicted membership in several of those categories. We show that there is a strong inverse relationship between the average beta weight assigned to a phonological feature, phoneme, or letter, and the frequency of that cue. Our results also extend claims about the source of sound symbolic effects, by demonstrating that different poles of the same semantic dimension differ in their predictability from form cues; that some previously unsuspected dimensions show strong symbolic effects; and that features, phonemes, and letters may all contribute to sound symbolism.

Introduction

In his *Cratylus* dialogue, Plato gives Socrates the following words:

> "Must we not begin [...] with letters; first separating the vowels, and then the consonants and mutes, into classes, according to the received distinctions of the learned; also the semivowels, which are neither vowels, nor yet mutes; and distinguishing into classes the vowels themselves? And when we have perfected the classification of things, we shall give them names, and see whether, as in the case of letters, there are any classes to which they may be all referred; and hence we shall see their natures, and see, too, whether they have in them classes as there are in the letters; and when we have well considered all this, we shall know how to apply them to what they resemble—whether one letter is used to denote one thing, or whether there is to be an admixture of several of them; just, as in painting, the painter who wants to depict anything sometimes uses purple only, or any other color, and sometimes mixes up several colors."
>
> [Plato, 360 BCE/1892]

A range of evidence (reviewed below) suggests that the sound symbolic associations Socrates was discussing, in which certain phonemes seem better suited as labels for certain meanings, are real, at least for some semantic categories. This evidence stands against one aspect of the "arbitrariness of the sign" (Saussure, 1916/1983), namely that no particular phoneme is more or less appropriate for any particular meaning. This is not a trivial matter, as arbitrariness is often taken as one of language's fundamental features (e.g., Hockett, 1963). Saussure (1916/1983) argued that there is no "reasonable basis" (p. 73) on which to discuss the appropriateness of a particular signifier for a signified, in contrast to our ability to discuss, for instance, "whether monogamy is better than polygamy" (p. 73). However, a variety of studies have shown that phonemes seem to have inherent associations with particular kinds of meanings, which suggests that such a discussion is viable and raises the possibility that these associations have affected language evolution (e.g., Berlin, 1994; Blasi, Wichmann, Hammarstrom, Stadler, & Christiansen, 2016; Johansson & Zlatev, 2013; Tanz, 1971; Ultan, 1978). Most notably, these

* Corresponding author.
E-mail address: chrisw@ualberta.ca (C. Westbury).

https://doi.org/10.1016/j.jml.2017.09.006
0749-596X/

phoneme-meaning associations include the categories *sharp/round* and *large/small*, as we review in the next section. However, there is still uncertainty as to which particular phonological features, letters, or phonemes are the best predictors of these dimensions, how they are weighted, how they are related, and *why* they are the best, in part due to the methodological limitations of much previous work.

Several associated questions remain open: Is there any systematicity as to which linguistics features can act as sound symbols? How general are the categories in which sound symbolic effects can be found? Are the predictors of one pole of the dimensions the same (with reversed sign) as the predictors of the other pole, or are the poles separately symbolized? Are all predictors of semantic dimensions equal in their predictive force? Are the effects driven by phonological features, phonemes, graphemes, or some combination? Do biphones or bigrams contribute to the effects? If so, how strongly? Are all dimensions equivalent in their reliance on formal characters (phonological features, phonemes, or letters), or are different semantic dimensions differentially sensitive to these different formal characters? When evidence from features, phonemes, and letters are in conflict, is evidence from one stronger than from the other? Is the answer the same for all semantic dimensions? In this paper, we begin to address all these questions, using a large-scale study in which hundreds of participants classified thousands of nonword strings for their suitability as exemplars of 18 semantic categories.

Background

One of the best-known early demonstrations of sound symbolism was Köhler's (1929, 1947) report (building on closely-related observations made by Usnadze, 1924) that people had very strong intuitions about whether *maluma* (in 1947, or *baluma* in 1929)[1] or *takete* was a better name for a round or spiky shape. Köhler found that most of his participants thought that *maluma* was a better name for round things and *takete* a better name for spiky things, a result that has been much replicated (Davis, 1961; Holland & Wertheimer, 1964; Maurer, Pathman, & Mondloch, 2006; Nielsen & Rendall, 2011; Ramachandran & Hubbard, 2001; Sidhu & Pexman, 2016).[2] Sapir (1929), following up experimentally on an idea in Jespersen, 1925), undertook a related experiment in which vowels were manipulated. Sapir limited himself "to the meaning contrast 'large': 'small' as offering the most likely chance of arriving at relatively tangible results" (p. 226). Using a large number of participants, he showed that people preferred an open-front vowel *a* (/æ/) over a closed-front vowel *i* (/ɪ/) in labelling large things. This was later extended to the front-back dimension, with the finding that individuals preferred front vowels (e.g., /i/) when labelling small shapes, and back vowels (e.g., /ɔ/) when labelling large shapes (Newman, 1933).

The results of these studies have been frequently replicated. However, there are five problems with the reliance of the field on Köhler and Sapir's paradigm of forced choice experiments with a small number of contrasting strings pre-selected by the experimenter "as offering the most likely chance of arriving at relatively tangible results" (Sapir, 1929).

One problem is that such contrasts are often transparent manipulations, making it difficult to gather a lot of data per participant. Köhler could not reasonably have asked participants to make the same choice about the strings *kittatee* and *moolmer*, then *tikkikit* and *malinus*, and so on. Participants would have quickly caught on, and the task results would become uninformative. This is one reason why we have knowledge of the sound symbolic value of relatively few strings after nearly a hundred years of research into the phenomenon.

A second problem is that forced choice judgments are ambiguous in their interpretation, since they don't provide sufficient information for an outside observer to discern the basis of any choice. Concretely, when someone tells us that *mil* is a better name for a small thing than *mal*, we don't know if that person made the choice:

1. Because *mil* seemed like a good name for a small thing (a true positive) but *mal* had no interpretation either way (unclassifiable).
2. Because *mal* seemed like a good name for a large thing (a true positive) but *mil* had no interpretation (unclassifiable).
3. Because *mil* seemed liked a *poor* name for a large thing (a true negative) but *mal* seemed like a good name (a true positive).
4. Because *mal* seemed liked a *poor* name for a small thing (a true negative) but *mil* seemed like a good name (a true positive).
5. Because (as is usually assumed) *mil* and *mal* were both true positives in their respective categories.

These five possibilities are not exhaustive, because they do not take into account any considerations of *quantity*. Even in the fifth and normally assumed 'best case', we do not know from seeing a consistent response if *mil* and *mal* are *good* representatives of their respective categories, or (as we will argue from evidence later in this paper) mediocre representatives that differ just sufficiently to enable a consistent forced-choice to be made. As noted by Tukey (1969), it is very difficult to make theoretical scientific progress after merely noting *a difference* if we do not also quantify the *size* of that difference.

A third problem is that experiments relying on pre-selected contrasting strings are really experiments in 'intuition matching'. Köhler's and Sapir's experiments were not just a demonstration of sound symbolism; they were simultaneously a demonstration of the fact that other people share intuitions about sound symbolism. Of course, ultimately, sound symbolism must always rely on intuition, since the phenomenon is defined in terms of phenomenology. However, intuition matching in experimental design is scientifically dissatisfying for several reasons. One is that our conscious intuitions about a phenomenon may provide very little insight into the true nature of the phenomenon (consider, for example, the history of armchair theorizing about the nature of color prior to Newton). To the extent that a phenomenon is more complex or stranger than we are able to intuit, intuition matching experiments must leave us wondering how much of the phenomenon remains unexplained. By definition what remains to be explained lies outside of our intuitions. The second reason it is dissatisfying is that intuition matching also limits our ability to explain *why* the phenomenon of sound symbolism occurs. If we do not know why we *ourselves* feel that *k* is a good symbol of spiky things, asking other people if they feel the same way is not likely to shed any light on the question of why we all share that intuition. Much of the work on sound symbolism to date has served only to document that the phenomenon exists.

The last two problems with forced choice experiments are closely related to this intuition-matching problem. The fourth

[1] We have confirmed that the change from *baluma* to *maluma* occurred between the 2nd and 4th *printings* of the first American (1929) edition of Köhler's book. The 1st and 2nd printings occurred in the same month: April 1929. The 3rd printing (a copy of which we have not yet been able to locate) occurred in August 1929 and the 4th in October 1929. Since it is unusual to allow author edits of the same edition of a book between printings and we are aware of no comment by Köhler about this apparent change in his intuitions, we conclude that the early 1929 appearance of the nonword *baluma* was most likely *a printing error*. Thanks to Jan-Olaf Svantesson for making us aware that the change occurred between printings of the first edition rather than between editions (as implied in Westbury, 2005).

[2] We have omitted Westbury (2005) from this list as repeated failures to replicate the reported effect have cast that effect into doubt (see Westbury, in press).