

**Iconicity ratings really do measure iconicity,
and they open a new window onto the nature of language**

Bodo Winter & Marcus Perlman
Dept. of English Language & Linguistics
University of Birmingham

Abstract

This paper reviews recent research using participant ratings to measure the iconicity (form-meaning resemblance) of words and signs. This method, by enabling wide coverage of lexical items and cross-linguistic comparison, has revealed systematic patterns in how iconicity is distributed across the vocabularies of different languages. These findings are consistent with established linguistic and psychological theory on iconicity, and they connect iconicity to factors like learning and acquisition, semantics, pragmatic aspects of language like playfulness, and to the semantic neighborhood density of words and signs. After taking stock of this research, we look critically at the construct validity of iconicity ratings, considering an alternative account of iconicity ratings recently put forward by Thompson, Akita, and Do (2020). They proposed that, for most vocabulary, participants might rate the iconicity of different words based on their meaning alone – specifically the degree to which it relates to the senses – independently of actual form-meaning resemblance. We argue that their hypothesis cannot account for many of the various, theory-driven results from this line of research, which strongly support the conclusion that the ratings really do measure iconicity.

Keywords: ideophones; mimetics; expressives; onomatopoeia; sound symbolism

1. Introduction

What is the nature of the linguistic sign? The fundamental question of how form relates to meaning has occupied scholars for centuries (Aristotle 350AD; Plato 1999). In the tradition of Saussure (1983), modern linguistics has largely adopted the principle of arbitrariness, assuming that, by design, the meaning of a word lacks any motivated relation to its form (Hockett 1960). Even in signed languages, where it is obvious that many signs are depictive of their meaning (e.g., Pietrandrea 2002), psychologists have argued that signers, when engaging in everyday conversation, nevertheless *treat* signs as though they are arbitrary (Bellugi & Klima 1976). Against this backdrop, the past few decades have seen an upsurge of research interest in the extent to which spoken and signed language vocabularies are characterized by *iconicity* – a resemblance between aspects of the form of a signal and aspects of its meaning (Dingemanse et al. 2015; Dingemanse, Perlman & Perniss 2020; Perniss, Thompson & Vigliocco 2010; Sidhu & Pexman 2018b). For example, in British Sign Language, the sign DRINK resembles drinking from a cup. And in English, the word *vroom* sounds like the revving of a car engine.

There is now a preponderance of evidence establishing that iconicity in its various forms, along with arbitrariness, is an important feature of human languages. Studies implicate iconicity in many different aspects of people's use of language, including cognitive processing (Sidhu, Vigliocco & Pexman 2019; Thompson 2011) and learning (Imai & Kita 2014; Laing 2019; Ortega 2017), as well as evolution (Armstrong & Wilcox 2007; Perlman, Dale & Lupyan 2015) and historical change (Flaksman 2017). This research on iconicity is highly interdisciplinary, employing a range of converging methods to study the phenomenon (reviewed in Motamedi et al., 2019; see also Dingemanse et al., 2020). Recent advances have been spurred by cross-talk between experimental approaches to form-meaning correspondences, and linguistic studies that examine iconicity in the vocabularies and grammars of natural languages (Dingemanse et al. 2015; Dingemanse, Perlman & Perniss 2020; Imai & Kita 2014). An important element of much of this work is the development of a distinctly

multi-modal and multi-semiotic theoretical framework for understanding iconicity in speech, sign, and gesture (Emmorey 2014; Ferrara & Hodge 2018; Perlman et al. 2018; Perniss, Thompson & Vigliocco 2010; Perniss, Thompson & Vigliocco 2010; Taub 2001; Thompson et al. 2020). The sum of this work makes it clear that, if we are to understand the complete nature of language, our theory will need to explain the fundamental role of iconicity.

One recent advance in the study of iconicity that bridges experimental and linguistic approaches is the use of participant ratings to measure the iconicity of words and signs (Caselli et al. 2017; Perry, Perlman & Lupyan 2015; Vinson et al. 2008b; Winter et al. 2017). Rather than relying on the intuitions of individual linguists, iconicity ratings draw on the “wisdom of the crowd”, aggregating over multiple different intuitions (Winter, 2019, Ch. 10-11). This method allows researchers to examine iconicity across large swaths of vocabulary, and it can enable more direct comparison between different kinds of languages, including between spoken and signed languages (Perlman et al. 2018; Thompson et al. 2020). In Sections 2 and 3 of this paper, we review recent research using iconicity ratings, showing how this method sheds new light onto the organization and structure of both spoken and signed vocabularies.

Yet, as we examine in Section 4, the use of iconicity ratings involves a considerable degree of abstraction away from the actual iconic correspondences involved in any particular word or sign. This raises critical questions about their construct validity, especially for spoken words where iconicity appears to be less transparent than in signs. Do iconicity ratings actually measure iconicity? Towards answering this question, we evaluate an alternative account of iconicity ratings put forward in a study of Japanese vocabulary recently published in this journal (Thompson, Akita & Do 2020). Thompson et al. proposed that, for most vocabulary, participants might rate the iconicity of different words based on their meaning alone – specifically the degree to which it relates to the senses – independently of actual form-meaning resemblance. Ultimately, we argue that their hypothesis cannot

account for many of the various, theory-driven results from this line of research, which strongly support the conclusion that the ratings really do measure iconicity.

2. Iconicity ratings and age of acquisition

In one of the first studies to collect lexical norms at large scale in a signed language, Vinson et al. (2008b) examined the relationship between the iconicity of signs and their age of acquisition in British Sign Language (BSL). Deaf BSL signers watched video clips of signs and rated them on a scale of 1 to 7 for how iconic they thought each sign was, with 1 indicating that it was not iconic at all and 7 that it was highly iconic (Vinson et al. 2008b). The instructions clarified the notion of iconicity to participants with clearly iconic and arbitrary examples, explaining to participants that an iconic sign “looks like what it means” (p. 1087). A key finding of this study was a negative correlation between the rated iconicity and the estimated age of acquisition of the signs: children tended to learn signs rated high in iconicity earlier than signs rated low in iconicity, a result that was later extended to actual sign production and comprehension data from deaf children (Thompson et al. 2012). A similar relationship between rated iconicity and age of acquisition has also been found in American Sign Language (Caselli & Pyers 2017). These studies of signed languages raise the question of whether iconicity might play a similar role in the order of vocabulary acquisition in spoken languages. Indeed, experimental studies conducted in the laboratory show that young children learn iconic words more easily than arbitrary ones (Imai et al. 2008; Kantartzis, Imai & Kita 2011).

Recently, the correlation between rated iconicity and age of acquisition in signed languages has been replicated for spoken iconicity ratings. Perry et al. (2015) collected iconicity ratings for roughly 500 English and Spanish words, with the following instructions given to participants (p. 12), modelled after Vinson et al. (2008b):

“Some English words sound like what they mean. For example, SLURP sounds like the noise made when you perform this kind of drinking action. An example that does not relate to the sound of an action is TEENY, which sounds like something very small (compared to HUGE which sounds big). These words are iconic. You might be able to guess these words’ meanings even if you did not know English.”

Even among spoken languages, English – along with many other European languages – has been characterized as an iconically impoverished language, devoid of iconic word classes like ideophones and mimetics (e.g., Nuckolls, 2003; Vigliocco et al., 2014). Therefore, Perry et al. began with the null premise that words are arbitrary, and adapted their iconicity scale to range from -5 (“words that sound like the opposite of what they mean”) to +5 (“words that sound like what they mean”), putting arbitrariness (“words that do not sound like what they mean or the opposite”) at the center. The results showed that onomatopoeic words like *moo* ($M = 3.88$) and interjections like *ouch* ($M = 3.46$) received very high iconicity ratings. But other kinds of words received relatively high ratings too, for instance the adjective *sticky* ($M = 2.93$) and the verb *stop* ($M = 2.50$), in contrast to the noun *jeans* ($M = 0.0$) or the function word *here* ($M = -0.2$).

Like in BSL (Vinson et al. 2008a), in both English and Spanish, the iconicity ratings of words were correlated with their age of acquisition (Perry, Perlman & Lupyan 2015). This relationship held when onomatopoeic words were excluded from the analyses, and it was found over various rating protocols, including, for English, both written and auditory presentation of the words, and a version in which participants judged, from 0 to 100, how accurately they thought a space alien could guess the meaning of the words based only on their sound. The analyses also accounted for a number of other possible factors known to be related to age of acquisition, including word frequency, number of phonemes, number of morphemes, concreteness, systematicity (a statistical measure of form-meaning regularity;

Monaghan et al. 2014), and the degree to which people associated the word with babies.

In subsequent studies, this basic finding has been replicated and extended in several ways. In English, the relationship between iconicity and age of acquisition has been shown for an expanded set of 2,000+ words, for which iconicity ratings were also found to be correlated with children's production frequency from actual child-parent interactions (Massaro & Perlman, 2017; Perry et al., 2017). In addition, these studies found the iconicity ratings to be associated with parental input frequency, with parents using more iconic words in child-directed as opposed to adult-directed speech. Moreover, a recent study shows expanded evidence for this relationship between iconicity and word learning in English, finding the same pattern in vocabulary comprehension data from children and young adults between grades 2 and 13 and word frequency norms from school materials for students from grades 1 to 13 (Sidhu et al. 2021). In Spanish, the relationship between iconicity and age of acquisition has recently been confirmed in an analysis of almost 11,000 words, in a set that also did not include onomatopoeia (Hinojosa et al. 2020).

Thus, even though iconicity appears to be more obvious in the vocabularies of signed languages, studies with iconicity ratings show that spoken languages – at least English and Spanish – exhibit the same basic pattern of iconicity in connection with the age of acquisition and early use of vocabulary. Whatever the language, young children tend to learn iconic items earlier, and they tend to use them more frequently – a pattern that is reflected in the child-directed communication of caregivers. And, as mentioned above, these results hold when controlling for the semantic properties of words, as well as other lexical factors. In the following section, we show how research using iconicity ratings has proven fruitful well beyond these important findings.

3. Patterns of iconicity

The use of participant ratings has enabled researchers to study how iconicity is distributed across the vocabularies of various languages, signed and spoken. Indeed,

ratings have now been collected for a number of languages, including British Sign Language (Thompson et al. 2012; Vinson et al. 2008a), American Sign Language (Caselli et al. 2017; Lieberth & Gamble 1991; Occhino, Anible & Morford 2020; Sehyr & Emmorey 2019), German Sign Language (Grote 2013; Occhino et al. 2017), English (Perry, Perlman & Lupyan 2015; Winter et al. 2017), Spanish (Hinojosa et al. 2020; Perry, Perlman & Lupyan 2015), and Japanese (Thompson, Akita & Do 2020). This research – in addition to documenting the prevalence of iconicity in early learned vocabulary – has found several other notable patterns of iconicity across different languages, many of which are consistent with established linguistic and psychological theories of iconicity. These findings connect iconicity ratings to factors like semantics (Perlman et al. 2018; Winter et al. 2017), phonological markedness (Dingemanse & Thompson 2020), pragmatic aspects of language like playfulness (Dingemanse & Thompson 2020), and to psycholinguistic factors such as semantic neighborhood density (Sidhu & Pexman 2018a; Thompson et al. 2020).

With respect to semantic factors, studies of spoken languages have found that iconicity ratings are correlated with sensory experience ratings, in both English (Sidhu & Pexman 2018a; Wertheimer 1958; Winter et al. 2017) and Spanish (Hinojosa et al. 2020). These results support the hypothesis that sensory meanings are more expressible using iconicity than abstract meanings (Dingemanse et al. 2015; Lupyan & Winter 2018; Winter et al. 2017), a claim that has also been made to account for the semantics of ideophones (Akita 2009; Dingemanse 2012; Kita 1997). More detailed analysis shows that English sound and touch words are rated as more iconic than taste and smell words, with smell words having the lowest rated iconicity in English (Winter et al. 2017). In comparison, this finding is predictably different for rated iconicity in ASL and BSL: signs stronger in tactile experience were also rated higher in iconicity, but those judged stronger in auditory experience were rated as less iconic (Perlman et al. 2018).

It is thought that iconic words and signs, by virtue of resembling the meanings they depict, communicate sensory experiences more directly. It has been argued that

this ties iconic words and signs more strongly to the concrete sensory dimensions they depict, thus making them less prone to semantic extension (Classen, 1993, Ch. 3; Meir, 2010). In support of this idea, Lupyan and Winter (2018) found that rated iconicity in English words is negatively correlated with contextual diversity: words with high rated iconicity occur in fewer different text types, suggesting that their meanings are restricted to a narrower range of contexts. Corpus evidence furthermore shows that English words with high rated iconicity are also less likely to participate in semantic extension via perceptual metaphor (*?squealing color* versus *loud color*) (Winter 2019). These results give empirical confirmation to the claim that iconic words are less prone to semantic extension (Classen, 1993, Ch. 3), which has also been proposed for ideophones in Japanese (Akita 2013) and iconic signs in signed languages (Meir 2010).

Studies, particularly of spoken languages, have also analyzed iconicity ratings according to the lexical category of words, with results that are consistent with the semantic patterns described above. As expected, in English (Perry, Perlman & Lupyan 2015; Winter et al. 2017) and in Spanish (Hinojosa et al. 2020; Perry, Perlman & Lupyan 2015), words classified as onomatopoeia were rated highest in iconicity, and similarly in Japanese (Thompson, Akita & Do 2020), ideophones were rated highest. Although much less iconic than onomatopoeia and ideophones, across all three languages, adjectives were rated relatively high in iconicity, especially compared to nouns and function words. This fits with the connection between iconicity and the sensory content of words, and with the observation that there is considerable overlap between adjectives, which prototypically refer to properties and qualities, and word classes like ideophones, which are known for conveying vivid sensory meanings (Dingemanse 2012).

Notably, ideophones are also widely used to express qualities of movement, especially manner of motion (e.g., Dingemanse, 2012; Imai & Kita, 2014). This leads to an interesting cross-linguistic prediction: languages like English that commonly express manner of motion within the main verb are likely to have more iconic verbs, whereas “verb-framed” languages like Spanish – which tend to convey path but not

manner of motion in the main verb (Talmy 1985) – will have less iconic verbs. Indeed, Perry et al. (2015) found this to be the case in their comparison of English and Spanish vocabulary, and analyses with expanded sets of vocabulary have confirmed that verbs are rated high in iconicity in English (Winter et al. 2017) compared to Spanish (Hinojosa et al. 2020). More detailed analysis of the Spanish verbs in Hinojosa’s study found further evidence that lack of manner information in the verbs drives their low iconicity.

In addition to patterns of semantics, iconicity ratings also reveal structural and pragmatic patterns of iconic words. Ideophones and similar word classes are, characteristically, structurally marked (e.g., Childs, 1994; Dingemanse, 2012). In parallel to this observation, Dingemanse and Thompson (2020) showed that English words rated high in iconicity were also more phonologically marked. Their study also found a pragmatic parallel between ideophones and spoken iconicity in English: Both are tied to playful language. It has long been suggested that in many different languages, ideophones are associated with informal discourse and language play (e.g., Klammer, 2002; Samarin, 1970), and Dingemanse and Thompson (2020) show that the same is the case for spoken iconicity in English. Intersecting with the observation that in English, funny words tend to be marked by unusual phonology (Westbury & Hollis 2019), their study showed a convergence of these pragmatic and phonological factors. Words like *zigzag*, *squeak*, and *waddle* were rated high in iconicity, and they were also righted high in funniness. These words were also structurally marked, such as featuring complex codas and onsets, or the suffix *-le*.

Finally, studies using iconicity ratings not only show which words in a vocabulary tend to be iconic; they can also reveal where within the lexicon there is less iconicity. Bühler (1934) already noted that, in theory, there ought to be an inverse relationship between iconicity and perceptual discriminability. Iconic words with similar meanings will tend to have similar forms, which can make them more difficult to perceptually discriminate (compare Gasser, 2004). This leads to the prediction that words that have a large number of close semantic neighbors will tend to be less iconic

in order to maximize their discriminability from each other. Sidhu and Pexman (2018a) used iconicity ratings to show that this is the case in English, finding a negative relationship between semantic density and rated iconicity. Interestingly, they also observed that the relationship between iconicity and sensory strength only held in sparser semantic neighborhoods. Subsequently, B. Thompson et al. (2020) looked at whether the same relationship existed in a signed language as well. They combined iconicity ratings from ASL and English together with a high-dimensional distributional model of English lexical semantics. In both languages, they found that words in semantically dense neighborhoods have less rated iconicity, particularly when considering an item's most closely related semantic neighbors.

Taken together, these studies begin to paint an increasingly detailed picture of the systematic ways that iconicity is distributed across the vocabularies of different kinds of languages. In the next section, we take a critical look at the construct validity of using participant ratings to measure iconicity. Can we take the results of these studies at face value, accept that some words or signs are more or less iconic than others, and that these differences are captured accurately by participant ratings?

4. Do iconicity ratings actually measure iconicity?

The studies discussed above show that participant ratings can be a powerful way to measure the iconicity of words and signs. They illustrate how the use of ratings offers key methodological advantages: both the ability to analyze thousands of items within a language in order to identify patterns across a lexicon, as well as to compare these patterns across different languages. The use of iconicity also affords more reproducibility than relying on singular theory-guided intuitions of individual linguists (Winter, 2019, Ch. 10-11). Yet, these advantages come at the cost of abstraction.

At the level of individual words and signs, iconicity ratings are a black box. Lost in the average ratings of lexical items are qualitative details of the specific form-meaning correspondences that participants are judging as iconic (Winter 2016a). For

any given item, we know nothing about the particular form-meaning pairings that participants tap into when they perform the rating study. Thus, iconicity ratings abstract away from the direct analysis of structural analogies between elements of form and meaning that have been discussed in the context of signed languages (Emmorey 2014; Taub 2001), or in the context of specific form-meaning pairings in spoken language vocabularies (e.g., Blasi et al., 2016). Because of their black-box quality, the construct validity of iconicity ratings requires careful scrutiny and testing. What does it mean that the BSL sign *eat* received an average iconicity rating of 6.80, or that the English word *black* received an average rating of 1.19? Do these ratings actually measure iconicity?

In the case of signed vocabularies, which are widely recognized to have a large proportion of distinctly iconic items (Bellugi & Klima 1976; Pietrandrea 2002; Wescott 1971), iconicity ratings have largely been accepted as a valid measure of form-meaning resemblance. However, the construct validity of iconicity ratings in the case of spoken words – where iconicity appears less obvious than in signs – may be subject to more scepticism. Scholars have argued that fluent speakers, especially those who learned the language in early childhood, might have so deeply internalized the words of their language that they are likely to imagine iconic correspondences between sound and meaning that are not really there. Already in 1861, Müller aptly warned that “we are deceived by the constant connection of certain sounds and certain meanings in the words of our own language, and how readily we imagine that there is something in the sound to tell us the meaning of these words” (p. 351). He rhetorically asks “who does not imagine” that the word *squirrel* invokes “something of the rustling and whirling of the little animal” (p. 350).

More than 100 years later, Sutherland and Cimpian (2015) provided direct empirical evidence that language users are biased towards believing that words naturally fit their referents, regardless of whether this is actually the case. These results, together with Müller’s caution, should make us wary of using participants’ intuitions as a measure of objectively verifiable form-meaning pairings in the lexicon.

Indeed, the subjectivity of iconicity ratings is an empirically established fact, as evidenced by documented differences between the iconicity ratings of native signers of the language compared to signers of other languages (Occhino et al 2017) or non-signers (Sevcikova Sehyr & Emmorey, 2019). Many researchers concur that iconicity is, to an extent, about “perceived” resemblance that is in “the eye of the beholder” (Occhino et al. 2017). However, this does not indicate that people’s impressions of iconicity are random and without motivation. To the contrary, there is still substantial correlation across items between the ratings of signers and non-signers (Sevcikova Sehyr & Emmorey 2019). Thus, while iconicity involves a degree of subjective construal (Occhino et al. 2017; Occhino, Anible & Morford 2020; Sehyr & Emmorey 2019), there is nevertheless some consistency in this construal, and ratings provide a direct way of capturing this.

Importantly, the words of a lexicon might not all be equal with respect to the validity of iconicity ratings. For example, Thompson and Dingemanse (2020) showed that compound words tend to be rated higher in iconicity. It appears that when asked to rate words for how much they “sound like what they mean”, participants tend to confuse this with semantic transparency – knowing the meaning of the compound based on knowing its component words. Clearly, this evidences some confounding of different constructs (semantic transparency versus iconicity) for this particular class of words. Thompson and Dingemanse (2020) dealt with this concern by excluding these words from certain analyses. A more general concern is that while raters’ iconicity ratings may be viable for clearly imitative words (such as many ideophones), it is possible that the ratings are systematically biased for many standard words like typical nouns, verbs, and adjectives. On this view, most words in a vocabulary are essentially arbitrary and thus, whatever a rater happens to perceive as iconicity would be spurious.

But this raises a critical question. If iconicity ratings are not valid for standard words, then what explains the systematic patterns laid out above in Sections 2 and 3? What accounts for the various results related to age of acquisition, semantics, and

semantic neighborhood density? If raters are not tapping into iconicity, then what is driving these patterns in their responses? In their recent analysis of Japanese and English vocabulary, A. Thompson et al. (2020) propose an answer to this question, offering an alternative hypothesis to explain the results of iconicity ratings studies. In particular, they focus on the consistent finding that iconicity ratings tend to be high for adjectives and are correlated with sensory strength. Their proposal, and a critical evaluation of it, may shed some light on the construct validity of iconicity ratings.

A. Thompson et al. (2020) argue that for most spoken words – that is, those words that are not onomatopoeia or ideophones – it is a word's relationship to sensory information, rather than actual form-meaning resemblance, that motivates its iconicity rating. On this account, raters are familiar with the salient iconicity evidenced in highly sensory word classes like onomatopoeia, and then, when asked to rate other kinds of words, they extrapolate from these distinctly imitative sensory words. Thus, for example, when seeing a standard arbitrary sensory adjective, raters assume, based on the model of onomatopoeia, that this word should also receive a higher iconicity rating. As A. Thompson et al. (2020) explain:

“Speakers have an awareness of iconicity as a resemblance between form and meaning coupled with an additional relationship to sensory information. When faced with words whose resemblance between form and meaning is not immediately obvious, speakers extrapolate from their awareness of iconicity and instead rate such words according to sensory information rather than resemblance between form and meaning.”

Thompson et al. (2020) also offer an explanation for why taste and smell words such as *fragrant* and *aromatic* have relatively low rated iconicity. They suggest that these words are not actually about sensory information per se, but about the emotional effects associated with these sensory experiences. They note that this specific

hypothesis is consistent with a lexical analysis by Winter (2016b) that shows that taste and smell words have strong emotional valence.

With this account, Thompson et al. offer an interesting alternative to explain certain findings from studies with iconicity ratings. However, their proposal generally fails to account for a number of other findings, including most of those discussed in Sections 3 and 4 above, and even including some of their own results. First, consider Thompson et al.'s hypothesis that participants rate words according to sensory information rather than form-meaning resemblance. This is simply not borne out by the data. For example, in English, the correlation between iconicity and the sensory strength of words, although statistically robust, is modest in strength. Using sensory ratings from Juhasz and Yap (2013), Winter (2016a) reports a correlation coefficient of $r = 0.18$ and an R^2 value of 0.03, corresponding to 3% described variance. Although this correlation is reliably different from zero, it is far too low for the two constructs to be co-extensive. After all, an R^2 value of 0.03 suggests that for this dataset, 97% of the variation in iconicity ratings is not due to sensory ratings. In Spanish vocabulary, the correlation coefficient between iconicity ratings and sensory ratings was slightly higher with $r = 0.24$ and $R^2 = 0.06$ (Hinojosa et al. 2020), but still far too low to reduce the iconicity ratings to sensory strength alone.

Second, the attempt at explaining away the iconicity rating result with respect to smell words – that they mainly express emotional rather than sensory information – is inconsistent with existing data. Using Juhasz and Yap's (2013) data, smell adjectives actually receive, on average, high sensory experience ratings ($M = 4.45$), higher than even auditory adjectives ($M = 3.21$), many of which are onomatopoeic. The low rated iconicity of English smell words, as found by Winter et al. (2017), is in line with literature finding that smell is one of the least common semantic domains to be depicted by ideophones (Dingemanse 2012). These results suggest that it is not sensory experience alone that matters, but the modality-specific affordances to create iconic form-meaning links. It is also noteworthy that Thompson et al. (2020) do not mention within-category variation, not all of which can easily be reduced to sensory

experience differences. For example, some smell words receive high iconicity ratings (*whiffy, musty, musky*) and others low iconicity ratings (*fetid, acrid, odorous*). The same applies to taste words, for which there are words rated relatively high in iconicity (*chewy, nutty, creamy*), as well as words rated relatively low (*lemony, malty, herby*). Thompson et al.'s critique only addresses words related to taste and smell, and is only targeted at the average between-group differences, whereas other studies on the iconicity ratings have looked across all sensory modalities at both within and between-category variation (Perlman et al. 2018; Winter et al. 2017). Thompson et al., for example, offer no account for why English tactile words receive high iconicity ratings, as discussed in detail in Winter et al. (2017).

A third point to consider in Thompson et al.'s (2020) proposal is the hard line that is drawn between imitative words – ideophones in Japanese and onomatopoeia in English – and standard vocabulary, such as typical nouns, verbs, adjectives and function words. They report an etymological analysis of iconicity ratings from Japanese and English, which appears to support this categorical distinction. Their analysis focused on words in each language that are not recently borrowed – Yamato words in Japanese that are derived from Old Japanese, and Germanic words in English. Using dictionaries to identify words with iconic origins, they compared the rated iconicity of these words to those without recognized iconic origins. Importantly, the set included only standard vocabulary, and did not include words identified as ideophones or onomatopoeia. Thompson et al. report that, in Japanese, words with iconic etymologies were not significantly more iconic than their non-iconic counterparts. For English, Thompson et al.'s data shows that words which have iconic origins according to the Oxford English Dictionary are rated as *more* iconic ($M = 2.37$) than other words ($M = 0.74$), a reliable difference with a large effect size ($t = 6.25$, $df = 401$, $p < 0.0001$, Cohen's $d = 1.76$) (see the following Open Science Framework Repository: <https://osf.io/xj3hb/>). These results suggest that iconicity ratings correspond to independent observations made by lexicographers. The question remains why Thompson et al. did not find a similar pattern in Japanese vocabulary. It

may be that the very prominent and large inventory of ideophones in Japanese leads speakers to make a sharper categorical distinction between iconic and arbitrary vocabulary, whereas this distinction is more graded in English, which has a less marked and less expansive ideophone system.

These critical points notwithstanding, alternative proposals like Thompson et al.'s are important for a rigorous evaluation of the construct validity of iconicity ratings, which is especially needed for spoken languages where lexical iconicity does not appear as clear-cut and well-structured as in signed languages. Construct validity should not be assumed, but it needs to be empirically demonstrated. As Thompson et al. observe, one potential source of validation is the use of foreign word guessing studies in which participants attempt to guess the meanings of foreign words from a set of alternatives. Although a number of these experiments have been conducted (Brown, Black & Horowitz 1955; Dingemanse et al. 2016), only a few have linked guessing accuracy to iconicity ratings. Yet, while limited in scope, this work does support the idea that iconicity ratings provide a valid measure of iconicity. For example, there is some evidence from signed languages: Lieberth and Gamble (1991) found that ASL signs rated higher in iconicity were guessed more correctly than lower-rated signs, a finding that has later been replicated and extended by Sehry and Emmorey (2019). Data is less direct for spoken languages, but it is also suggestive of an iconicity-based interpretation. In a study of communication with non-linguistic vocalizations, Perlman and Lupyan (2018) showed that vocalizations that were rated higher in iconicity were guessed more accurately and learned more quickly by an independent set of participants.

5. Conclusion

We have argued the case, based on a substantial amount of cross-linguistic data, that iconicity ratings really do measure iconicity. As a methodology, ratings offer a specific way of operationalizing iconicity (Dingemanse, Perlman & Perniss 2020; Motamedi et al. 2019) that has both advantages and disadvantages: Their power comes from the

ability to survey iconicity across large swaths of a vocabulary, which allows broad comparisons between languages. In doing so, they open a new window onto vocabulary structure and the balance between iconicity and arbitrariness in different kinds of languages, spoken and signed. Yet, iconicity ratings have the disadvantage that they obscure the details of specific mappings between form and meaning. Thus, they work best as one tool among many, part of a larger “iconicity toolbox” of methods (Motamedi et al. 2019), and they are most informative in conjunction with psychological experiments and detailed linguistic analyses. No single method has privileged access to iconicity, and we should be careful to appoint any one method to be the “litmus test” for this concept (Thompson, Akita & Do 2020). Dąbrowska (2016) observed that “when it comes to understanding something as complex as human language, it will be most productive to use every method that is available” (p. 57). The study of iconicity is no different in this respect. Only by integrating ideas from different disciplines will we get a fuller picture of the fundamental role of iconicity in the nature of language.

References

- Akita, Kimi. 2009. *A grammar of sound-symbolic words in Japanese: Theoretical approaches to iconic and lexical properties of mimetics*. Kobe: Kobe University PhD Thesis.
- Akita, Kimi. 2013. Constraints on the semantic extension of onomatopoeia. *Public Journal of Semiotics* 5(1). 21–37. <https://doi.org/10.37693/pjos.2013.5.9646>.
- Aristotle. 350AD. *On Interpretation*. <http://classics.mit.edu/Aristotle/interpretation.html> (9 November, 2020).
- Armstrong, David F. & Sherman E. Wilcox. 2007. *The Gestural Origin of Language*. Oxford ; New York: Oxford University Press.
- Bellugi, U. & E. S. Klima. 1976. Two faces of sign: iconic and abstract. *Annals of the New York Academy of Sciences* 280. 514–538.
- Blasi, Damián E., Søren Wichmann, Harald Hammarström, Peter F. Stadler & Morten H. Christiansen. 2016. Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences* 113(39). 10818–10823. <https://doi.org/10.1073/pnas.1605782113>.
- Brown, R. W., A. H. Black & A. E. Horowitz. 1955. Phonetic symbolism in natural languages. *Journal of Abnormal Psychology* 50(3). 388–393.
- Bühler, Karl. 1934. *Theory of Language: The representational function of language*. Philadelphia, NETHERLANDS: John Benjamins Publishing Company.

- <http://ebookcentral.proquest.com/lib/bham/detail.action?docID=688912> (20 November, 2018).
- Caselli, Naomi K. & Jennie E. Pyers. 2017. The road to language learning is not entirely iconic: Iconicity, neighborhood density, and frequency facilitate acquisition of sign language. *Psychological science* 28(7). 979–987.
- Caselli, Naomi K., Zed Sevcikova Sehyr, Ariel M. Cohen-Goldberg & Karen Emmorey. 2017. ASL-LEX: A lexical database of American Sign Language. *Behavior Research Methods* 49(2). 784–801.
- Childs, G. Tucker. 1994. African ideophones. (Ed.) Leanne Hinton, Johanna Nichols & John J. Ohala. *Sound symbolism*. Cambridge University Press Cambridge 178–206.
- Classen, Constance. 1993. *Worlds of Sense: Exploring the Senses in History and Across Cultures*.
https://books.google.co.uk/books/about/Worlds_of_Sense.html?id=k9QOAAAAQAAJ&redir_esc=y (2 December, 2020).
- Dąbrowska, Ewa. 2016. Looking into introspection. In Grzegorz Drożdż (ed.), *Studies in Lexicogrammar: Theory and applications*, 55–74. Amsterdam: John Benjamins.
- Dingemanse, Mark. 2012. Advances in the cross-linguistic study of ideophones. *Language and Linguistics compass* 6(10). 654–672.
- Dingemanse, Mark, Damián E. Blasi, Gary Lupyan, Morten H. Christiansen & Padraic Monaghan. 2015. Arbitrariness, iconicity, and systematicity in language. *Trends in cognitive sciences* 19(10). 603–615.
- Dingemanse, Mark, Marcus Perlman & Pamela Perniss. 2020. Construals of iconicity: experimental approaches to form–meaning resemblances in language. *Language and Cognition*. Cambridge University Press 12(1). 1–14.
<https://doi.org/10.1017/langcog.2019.48>.
- Dingemanse, Mark, Will Schuermer, Eva Reinisch, Sylvia Tufvesson & Holger Mitterer. 2016. What sound symbolism can and cannot do: Testing the iconicity of ideophones from five languages. *Language* 92(2). e117–e133.
- Dingemanse, Mark & Bill Thompson. 2020. Playful iconicity: structural markedness underlies the relation between funniness and iconicity. *Language and Cognition* 1–22.
<https://doi.org/10.1017/langcog.2019.49>.
- Emmorey, Karen. 2014. Iconicity as structure mapping. *Philosophical Transactions of the Royal Society B: Biological Sciences*. The Royal Society 369(1651). 20130301.
<https://doi.org/10.1098/rstb.2013.0301>.
- Ferrara, Lindsay & Gabrielle Hodge. 2018. Language as Description, Indication, and Depiction. *Frontiers in Psychology*. Frontiers 9.
<https://doi.org/10.3389/fpsyg.2018.00716>.
<https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00716/full> (31 July, 2020).
- Flaksman, Maria. 2017. Iconic treadmill hypothesis: The reasons behind continuous onomatopoeic coinage. In Angelika Zirker, Matthias Bauer, Olga Fischer & Christina Ljungberg (eds.), *Iconicity in Language and Literature*, vol. 15. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/ill.15.02fla>.
<https://benjamins.com/catalog/ill.15.02fla> (21 December, 2019).
- Gasser, Michael. 2004. The origins of arbitrariness in language. In Kenneth Forbus, Dedre Gentner & Terry Regier (eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, 434–439. Mahwah, NJ: Erlbaum.

- Grote, Klaudia. 2013. *“Modality relativity”: The influence of sign language and spoken language on conceptual categorization*. Hochschulbibliothek der Rheinisch-Westfälischen Technischen Hochschule Aachen PhD Thesis.
- Hinojosa, J. A., J. Haro, S. Magallares, J. A. Duñabeitia & P. Ferré. 2020. Iconicity ratings for 10,995 Spanish words and their relationship with psycholinguistic variables. *Behavior Research Methods*. Springer 1–14. <https://doi.org/10.3758/s13428-020-01496-z>.
- Hockett, Charles F. 1960. The Origin of Speech. *Scientific American* 203(3). 88–96. <https://doi.org/10.1038/scientificamerican0960-88>.
- Imai, Mutsumi & Sotaro Kita. 2014. The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Phil. Trans. R. Soc. B* 369(1651). 20130298. <https://doi.org/10.1098/rstb.2013.0298>.
- Imai, Mutsumi, Sotaro Kita, Miho Nagumo & Hiroyuki Okada. 2008. Sound symbolism facilitates early verb learning. *Cognition* 109(1). 54–65. <https://doi.org/10.1016/j.cognition.2008.07.015>.
- Juhasz, Barbara J. & Melvin J. Yap. 2013. Sensory experience ratings for over 5,000 mono- and disyllabic words. *Behavior Research Methods* 45(1). 160–168.
- Kantartzis, Katerina, Mutsumi Imai & Sotaro Kita. 2011. Japanese Sound-Symbolism Facilitates Word Learning in English-Speaking Children. *Cognitive Science* 35(3). 575–586. <https://doi.org/10.1111/j.1551-6709.2010.01169.x>.
- Kita, Sotaro. 1997. Two-dimensional semantic analysis of Japanese mimetics. *Linguistics* 35(2). 379–416. <https://doi.org/10.1515/ling.1997.35.2.379>.
- Klamer, Marian. 2002. Semantically motivated lexical patterns: A study of Dutch and Kambera expressives. *Language*. JSTOR 258–286.
- Laing, Catherine. 2019. A role for onomatopoeia in early language: evidence from phonological development. *Language and Cognition* 11. 173–187. <https://doi.org/10.1017/langcog.2018.23>.
- Lieberth, Ann K. & Mary Ellen Bellile Gamble. 1991. The role of iconicity in sign language learning by hearing adults. *Journal of Communication Disorders*. Elsevier 24(2). 89–99. [https://doi.org/10.1016/0021-9924\(91\)90013-9](https://doi.org/10.1016/0021-9924(91)90013-9).
- Lupyan, Gary & Bodo Winter. 2018. Language is more abstract than you think, or, why aren’t languages more iconic? *Philosophical Transactions of the Royal Society B: Biological Sciences* 373(1752). 20170137.
- Massaro, Dominic W. & Marcus Perlman. 2017. Quantifying Iconicity’s Contribution during Language Acquisition: Implications for Vocabulary Learning. *Frontiers in Communication* 2. <https://doi.org/10.3389/fcomm.2017.00004>. <https://www.frontiersin.org/articles/10.3389/fcomm.2017.00004/full> (2 October, 2019).
- Meir, Irit. 2010. Iconicity and metaphor: Constraints on metaphorical extension of iconic forms. *Language* 86(4). 865–896.
- Monaghan, Padraic, Richard C. Shillcock, Morten H. Christiansen & Simon Kirby. 2014. How arbitrary is language? *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1651). <https://doi.org/10.1098/rstb.2013.0299>. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4123678/> (24 October, 2018).
- Motamedi, Yasamin, Hannah Little, Alan Nielsen & Justin Sulik. 2019. The iconicity toolbox: empirical approaches to measuring iconicity. *Language and Cognition*. Cambridge University Press 11(2). 188–207.

- Müller, Friedrich Max. 1861. *Lectures on the science of language*. Cambridge Library Collection [2013]. London: Longman, Green & Co.
- Nuckolls, Janis B. 2003. To be or not to be ideophonically impoverished. In *Proceedings of the Eleventh Annual Symposium about Language and Society*. Austin, TX: Citeseer.
- Occhino, Corrine, Benjamin Anible & Jill P. Morford. 2020. The role of iconicity, construal, and proficiency in the online processing of handshape. *Language and Cognition*. Cambridge University Press 12(1). 114–137. <https://doi.org/10.1017/langcog.2020.1>.
- Occhino, Corrine, Benjamin Anible, Erin Wilkinson & Jill P. Morford. 2017. Iconicity is in the eye of the beholder: How language experience affects perceived iconicity. *Gesture*. John Benjamins 16(1). 100–126. <https://doi.org/10.1075/gest.16.1.04occ>.
- Ortega, Gerardo. 2017. Iconicity and Sign Lexical Acquisition: A Review. *Frontiers in Psychology* 8. <https://doi.org/10.3389/fpsyg.2017.01280>.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5539242/> (29 December, 2018).
- Perlman, Marcus, Rick Dale & Gary Lupyan. 2015. Iconicity can ground the creation of vocal symbols. *Royal Society Open Science* 2(8). 150152.
<https://doi.org/10.1098/rsos.150152>.
- Perlman, Marcus, Hannah Little, Bill Thompson & Robin L. Thompson. 2018. Iconicity in signed and spoken vocabulary: A comparison between American Sign Language, British Sign Language, English, and Spanish. *Frontiers in Psychology* 9. 1433.
<https://doi.org/10.3389/fpsyg.2018.01433>.
- Perlman, Marcus & Gary Lupyan. 2018. People can create iconic vocalizations to communicate various meanings to naïve listeners. *Scientific Reports*. Nature Publishing Group 8(1). 1–14. <https://doi.org/10.1038/s41598-018-20961-6>.
- Perniss, Pamela, Robin L. Thompson & Gabriella Vigliocco. 2010. Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology* 1.
- Perry, Lynn K., Marcus Perlman & Gary Lupyan. 2015. Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PloS one* 10(9). e0137147.
<https://doi.org/10.1371/journal.pone.0137147>.
- Perry, Lynn K., Marcus Perlman, Bodo Winter, Dominic W. Massaro & Gary Lupyan. 2017. Iconicity in the speech of children and adults. *Developmental Science* e12572.
<https://doi.org/10.1111/desc.12572>.
- Pietrandrea, Paola. 2002. Iconicity and Arbitrariness in Italian Sign Language. *Sign Language Studies* 2(3). 296–321. <https://doi.org/10.1353/sls.2002.0012>.
- Plato. 1999. *Cratylus*. Reprinted with corr. Indianapolis: Hackett.
- Samarin, William J. 1970. Inventory and choice in expressive language. *Word*. Taylor & Francis 26(2). 153–169. <https://doi.org/10.1080/00437956.1970.11435590>.
- Saussure, Ferdinand de. 1983. *Course in General Linguistics*. La Salle, IL: Open Court.
- Sehyr, Zed Sevcikova & Karen Emmorey. 2019. The perceived mapping between form and meaning in American Sign Language depends on linguistic knowledge and task: evidence from iconicity and transparency judgments. *Language and Cognition*. Cambridge University Press 11(2). 208–234.
<https://doi.org/10.1017/langcog.2019.18>.
- Sevcikova Sehyr, Zed & Karen Emmorey. 2019. The perceived mapping between form and meaning in American Sign Language depends on linguistic knowledge and task: evidence from iconicity and transparency judgments. *Language and Cognition* 11(2). 208–234. <https://doi.org/10.1017/langcog.2019.18>.

- Sidhu, David M. & Penny M. Pexman. 2018a. Lonely sensational icons: semantic neighbourhood density, sensory experience and iconicity. *Language, Cognition and Neuroscience* 33(1). 25–31.
- Sidhu, David M. & Penny M. Pexman. 2018b. Five mechanisms of sound symbolic association. *Psychonomic Bulletin & Review* 25(5). 1619–1643.
- Sidhu, David M., Gabriella Vigliocco & Penny M. Pexman. 2019. Effects of iconicity in lexical decision. *Language and Cognition* 1–18. <https://doi.org/10.1017/langcog.2019.36>.
- Sidhu, David, Jennifer Williamson, Velina Slavova & Penny M. Pexman. 2021. An investigation of iconic language development in four datasets. *PsyArXiv*. PsyArXiv. <https://psyarxiv.com/qv9pg/>.
- Sutherland, Shelbie L. & Andrei Cimpian. 2015. An explanatory heuristic gives rise to the belief that words are well suited for their referents. *Cognition*. Elsevier 143. 228–240. <https://doi.org/10.1016/j.cognition.2015.07.002>.
- Talmy, Leonard. 1985. Lexicalization patterns: semantic structure in lexical forms. In Timothy Shopen (ed.), *Language typology and syntactic description*, vol. 3. Cambridge, UK: Cambridge University Press.
- Taub, Sarah F. 2001. *Language from the Body: Iconicity and Metaphor in American Sign Language*. Cambridge University Press.
- Thompson, Arthur Lewis, Kimi Akita & Youngah Do. 2020. Iconicity ratings across the Japanese lexicon: A comparative study with English. *Linguistics Vanguard*.
- Thompson, Bill, Marcus Perlman, Gary Lupyan, Zed Sevcikova Sehyr & Karen Emmorey. 2020. A data-driven approach to the semantics of iconicity in American Sign Language and English. *Language and Cognition*. Cambridge University Press 12(1). 182–202. <https://doi.org/10.1017/langcog.2019.52>.
- Thompson, Robin L. 2011. Iconicity in Language Processing and Acquisition: What Signed Languages Reveal. *Language and Linguistics Compass* 5(9). 603–616. <https://doi.org/10.1111/j.1749-818X.2011.00301.x>.
- Thompson, Robin L., David P. Vinson, Bencie Woll & Gabriella Vigliocco. 2012. The road to language learning is iconic: Evidence from British Sign Language. *Psychological science* 23(12). 1443–1448.
- Vigliocco, Gabriella, Pamela Perniss & David Vinson. 2014. Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1651). <https://doi.org/10.1098/rstb.2013.0292>. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4123671/> (21 May, 2019).
- Vinson, David P., Kearsy Cormier, Tanya Denmark, Adam Schembri & Gabriella Vigliocco. 2008a. The British Sign Language (BSL) norms for age of acquisition, familiarity, and iconicity. *Behavior research methods* 40(4). 1079–1087.
- Vinson, David P., Kearsy Cormier, Tanya Denmark, Adam Schembri & Gabriella Vigliocco. 2008b. The British Sign Language (BSL) norms for age of acquisition, familiarity, and iconicity. *Behavior Research Methods* 40(4). 1079–1087. <https://doi.org/10.3758/BRM.40.4.1079>.
- Wertheimer, Michael. 1958. The relation between the sound of a word and its meaning. *The American Journal of Psychology*. JSTOR 71(2). 412–415. <https://doi.org/10.2307/1420089>.
- Wescott, Roger W. 1971. Linguistic Iconism. *Language* 47(2). 416. <https://doi.org/10.2307/412089>.

- Westbury, Chris & Geoff Hollis. 2019. Wiggly, squiffy, lummoX, and boobs: What makes some words funny? *Journal of Experimental Psychology: General* 148(1). 97–123. <https://doi.org/10.1037/xge0000467>.
- Winter, Bodo. 2016a. *The sensory structure of the English lexicon*. Merced: University of California, Merced PhD.
- Winter, Bodo. 2016b. Taste and smell words form an affectively loaded and emotionally flexible part of the English lexicon. *Language, Cognition and Neuroscience* 31(8). 975–988.
- Winter, Bodo. 2019. *Sensory linguistics: Language, perception, and metaphor*. Amsterdam: John Benjamins.
- Winter, Bodo, Marcus Perlman, Lynn K. Perry & Gary Lupyan. 2017. Which words are most iconic? Iconicity in English sensory words. *Interaction Studies* 18(3). 433–454. <https://doi.org/10.1075/is.18.3.07win>.