COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 47 (2023) e13268

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13268

# Iconicity in Ideophones: Guessing, Memorizing, and Reassessing

Thomas Van Hoey,[a] Arthur L. Thompson,[a] Youngah Do,[a] Mark Dingemanse[b]

[a]*Department of Linguistics, The University of Hong Kong*
[b]*Centre for Language Studies, Radboud University*

Received 7 July 2022; received in revised form 15 February 2023; accepted 21 February 2023

**Abstract**

Iconicity, or the resemblance between form and meaning, is often ascribed to a special status and contrasted with default assumptions of arbitrariness in spoken language. But does iconicity in spoken language have a special status when it comes to learnability? A simple way to gauge learnability is to see how well something is retrieved from memory. We can further contrast this with guessability, to see (1) whether the ease of guessing the meanings of ideophones outperforms the rate at which they are remembered; and (2) how willing participants' are to reassess what they were taught in a prior task—a novel contribution of this study. We replicate prior guessing and memory tasks using ideophones and adjectives from Japanese, Korean, and Igbo. Our results show that although native Cantonese speakers guessed ideophone meanings above chance level, they memorized both ideophones and adjectives with comparable accuracy. However, response time data show that participants took significantly longer to respond correctly to adjective–meaning pairs—indicating a discrepancy in a cognitive effort that favored the recognition of ideophones. In a follow-up reassessment task, participants who were taught foil translations were more likely to choose the true translations for ideophones rather than adjectives.

Thomas Van Hoey and Arthur Lewis Thompson are co-first authors.
Correspondence should be sent to Youngah Do, Department of Linguistics, The University of Hong Kong, 930 Run Run Shaw Tower, The University of Hong Kong, Pokfulam, Hong Kong, China. Email: youngah@hku.hk or Mark Dingemanse, Centre for Language Studies, Radboud University, Nijmegen, The Netherlands, Email: mark.dingemanse@ru.nl

By comparing the findings from our guessing and memory tasks, we conclude that iconicity is more accessible if a task requires participants to actively seek out sound-meaning associations.

*Keywords:* Ideophone; 2AFC; Memory task; Iconicity; Sound symbolism

## 1. Introduction

Iconicity, or the perceived resemblance between form and meaning, is often ascribed a special status and contrasted with default assumptions of arbitrary form-meaning associations in spoken language. Several studies suggest that iconicity facilitates the learning of iconic words themselves (Fort et al., 2018; Imai & Kita, 2014). But does iconicity in spoken language have a special status when it comes to learnability? Learning is a complex and multidimensional process involving the acquisition, representation, retention, and reconstruction of knowledge and experience (Bertolo, 2001; Bruner, 1979; Matthews & Demopoulos, 2012). Here, we operationalize learnability as the ability to store and retrieve form–meaning associations. Specifically, how memorable are words with iconic mappings between form and meaning relative to words where such iconic mappings are less salient?

How well a linguistic phenomenon, such as a rule, a paradigm, or a set of words, is stored in memory has implications for its interplay with language learning in general. For example, experimental work on novel word learning has informed our understanding of how cognitive biases interact with the acquisition of phonology (see overview in Moreton & Pater, 2012a, 2012b). Words rated highly for iconicity, by native speakers, have also been correlated to earlier stages of L1 acquisition (Perry, Perlman, & Lupyan, 2015; Laing 2019)—implying at least that some iconic words are learned early, though the jury is out on whether this has downstream effects on other aspects of language learning (Nielsen & Dingemanse, 2021).

If words featuring iconicity are remembered better than other words, this could eventually help us understand the role iconicity has in language learning. For example, we might want to ask whether the perception of iconicity in a given word facilitates the learning of that word and its linguistic properties, be it from a speaker's first or second language. Work on early language learning has suggested that iconic words may help the learning of verbal meanings (Yoshida, 2004, 2012; Imai et al., 2008), and experiments with adult learners suggest that iconic words from natural languages may be easier to remember than less iconic ones (Iwasaki & Yoshioka, 2019; Lockwood et al., 2016a,b). This work has often featured words variously known as ideophones, mimetics, or expressives. Ideophones are marked words that depict sensory imagery (Akita & Dingemanse, 2019). Ideophones stand out by being marked in terms of phonology (Nuckolls, Stanley, Nielsen, & Hopper 2016; Thompson & Do, 2019; Thompson, Chan, Yeung, & Do, 2022) as well as prosody (Mok, 2001; Thompson, 2018; Van Hoey & Thompson, 2020) and morphosyntax (Beck, 2008). They often feature iconic associations between form and meaning grounded in cross-modal structural analogies (Akita & Dingemanse, 2019; Emmorey, 2014)—such as sibilant sounds like /ʃ/ being associated with scraping or roughness because fricative movements or contact with a rough texture often results in a sibilant noise (see Thompson, Van Hoey, & Do, 2021)—something that is enabled

by their highly concrete and sensory meanings (McLean, 2020; Nuckolls, 2019; Van Hoey, in press).

A few typical features of ideophones can be illustrated with the Japanese SOUND ideophone (onomatopoeia) tonton トン トン "knocking (on a door)," the Igbo MOTION ideophone *sùkwárachị sùkwárachị* "moving in a jerking motion," and the Korean TEXTURE ideophone *mulleong mulleong* 물렁 물렁 "soft but flexible (like pudding or flan)." In terms of morphology, all three examples display full reduplication. As to their prosodic patterns, *tonton* is typically produced with a higher pitch, *mulleong mulleong* with a lower pitch, while *sụ̀kwaŕaćhị sụ̀kwaŕaćhị* is spoken slower relative to surrounding words, presumably in accordance with its meaning. All three examples show how ideophones conjure up vivid sensory scenes, with a focus on sound (*tonton*), movement (*sụ̀kwaŕaćhị sụ̀kwaŕaćhị*), or texture (*mulleong mulleong*). The special properties of ideophones warrant a closer look at their potential facilitatory effects: Do the structural analogies featured in ideophones make these words easier to remember? Easier to guess? Or both?

Memory tasks and guessing task involve different cognitive processes. Memory tasks require either veridical recall or recognition which requires more cognitive resources than one-off guessing tasks where retention is not required (Archibald, 2017; Marsh, Hughes, Sörqvist, Beaman, & Jones, 2015). Retention and veridical recall aside, there are additional linguistic challenges in the memory tasks which have been used to assess the cross-linguistic learnability of ideophones. For example, participants in a memory task may have trouble remembering an ideophone from a foreign language because it is composed of strings of unfamiliar sounds which do not occur in the participant's native language. Various studies have shown that sequences of sounds not found in participants' native language impede learning (Best, 1994, 1995; Best, McRoberts, & Sithole, 1988; Flege, 1987, 1995a, 1995b). Other factors like elapsed time and the number of stimuli could also result in memory trouble or at least variability across participants. It follows, then, that a participant's failure to recall an ideophone does not necessarily entail a lack of iconic transparency; it may be due to other impeding factors.

To disentangle memory from iconicity, we use memory (veridical recall) tasks in tandem with guessing tasks, as guessing can show us how semantically transparent a word is, that is, how obvious its meaning is based on how the word sounds. The combination of guessing and memory tasks allows us to investigate the relation between semantic transparency and veridical recall. In other words, we can see whether guessing accuracy is correlated to better performance in memory tasks, something originally investigated in Lockwood, Dingemanse, and Hagoort (2016). We can also look at iconicity from a new direction by measuring the degree to which participants reassess what they have learnt from a prior memory task. We do this with a follow-up guessing task where participants are presented with the taught meaning (which is either true or coerced) and a new meaning. Participants in this reassessment task who choose the alternative, new meaning, will be said to be "flip-flopping" (see Section 3.4.3). When participants flip-flop from a taught, wrong meaning to a new, correct meaning, they potentially tap into the iconicity structure of an ideophone. That is to say, in the reassessment task, participants need to reconsider whether, for example, the Japanese ideophone *fuwafuwa*, which they might have learned as "dog barking" in the coerced condition really means "dog barking" or the newly presented (and true) alternative of "fluffy." If

flip-flopping from coerced to true meaning happens more often than the reverse, this points to the accessibility of iconic structure mappings. Furthermore, if overall performance is good for memory as well as guessing, this implies that, on some level, participants rely on principles of structure mapping to give meaning to ideophones. That is to say, the iconicity in the ideophones helps them remember and helps them guess because the iconicity lends itself to attaching meaning to otherwise unfamiliar units of sound.

We should keep in mind that the nature of how structure mapping is applied may differ between the guessing task and memory task. When guessing, participants are examining the unfamiliar word presented and then actively trying to make connections between that word's form and the two translations they are allowed to choose from. If performance is better for the guessing task, this would suggest that iconicity is only or mainly helpful when participants are *actively* searching for ways to connect sounds with meaning. Crucially, we can get insight into the "actively searching" aspect by an experimental manipulation in the memory task: Some ideophones are paired with a translation that represents their actual meanings, while others are paired with a different ("coerced") translation. The rate at which participants reassessed (or "flip-flopped"), from first memorizing a coerced translation to later guessing the correct translation, implies an active reanalysis of ideophones via structure mapping. With a memory task in isolation, in contrast, it is difficult to say whether participants actively searched for a connection between sound and meaning or instead relied on other associative means which may also be used to, say, remember a street name or phone number. In the case of unsuccessful memorization yet successful guessing, we may conclude that the iconicity of some ideophones is perceivable to nonnative speakers but only if structure mapping is actively sought out.

In this paper, we investigate guessability (Experiment 1) and the veridical recall (Experiment 2) of ideophones from three languages: Japanese, Korean, and Igbo. In Experiment 1, we establish a baseline by replicating Dingemanse, Schuerman, Reinisch, Tufvesson, and Mitterer (2016), which investigated how well Dutch participants guessed the meanings of Japanese ideophones. We find comparable results with a participant pool of native speakers of Cantonese. In Experiment 2, we carry out a memory (veridical recall) task following the experimental paradigm of Lockwood, Dingemanse, et al. (2016). This paradigm also contrasts ideophones with adjectives, as a point for comparison. The theoretical assumption is here that while ideophones are iconic, adjectives have the potential to be iconic, however, may not always be perceived as such. Overall, we find that while participants did not recall ideophone–meaning pairs better than adjectives (Experiment 2), they still guessed ideophone meanings with better-than-chance accuracy (Experiment 1). Furthermore, response time data from Experiment 2 show that participants took significantly longer to respond to adjectives, indicating a discrepancy in cognitive effort that favored the recognition of ideophones. Finally, in the latter half of Experiment 2, we administered a novel approach to measuring the perception of iconicity: reassessment. Participants who were taught coerced ideophone meanings in the first half of Experiment 2 were more likely to reassess what they memorized by responding with the (untaught) correct meaning in the second half of Experiment 2. All of this allows us to discuss the accessibility of iconicity when participants are cued to look for structure mappings.

## 2. Experiment 1: Guessing study

We first investigate how speakers guess the meanings of ideophones with a pool of native speakers of Cantonese. Our guessing study replicated the design of Dingemanse et al. (2016) with several modifications suitable for the current participants' pool (see 3.2 and 3.3 for the details). The original study explored whether native speakers of Dutch can guess the meanings of ideophonic stimuli from five languages (Japanese, Korean, Semai, Siwu, and Ewe) above chance level. It featured stimuli in five semantic categories (SOUND, MOTION, COLOR-VISUAL, SHAPE, and TEXTURE) and was designed as a forced choice task where in each trial, participants heard an item and chose between two possible meanings: the target meaning, or a foil meaning, that is, another item from the same sensory category. The study found that people can correctly guess aspects of the meanings of ideophones from unfamiliar languages, above chance level. Modifications made in the current study are explained in the three sections below.

### 2.1. Participants

The data pool for our study included native Hong Kong Cantonese speakers[1] ($n = 111$), the number of participants comparable to Dingemanse et al. (2016), who analyzed the data of 80 native Dutch speakers. Each participant confirmed that they were unfamiliar with Japanese, Korean, and Igbo. Upon the completion of the test, each participant received 50 HKD for their participation.

### 2.2. Stimuli

Japanese and Korean ideophones were included following Dingemanse et al. (2016). Igbo, a Niger-Congo language, was included for its typological and geographical distance from Japanese and Korean. Moreover, we assumed that Hong Kong participants would be less familiar with the phonetic nature of Igbo than with Japanese or Korean. All ideophones (Japanese: $n = 42$, Korean: $n = 45$, and Igbo: $n = 39$) were recorded by native speakers in a sound booth using an earset microphone and an Onyx Blackjack USB Recording Interface. Recordings were normalized at a root mean square (RMS) level of 70 dB using Praat (Boersma & Weenink, 2021). Note that, different from Dingemanse et al. (2016), exploring the role of prosody in guessing the meanings of ideophones was not a focus in the current study, so we only replicate the "original recordings" condition of their study. Ideophone translations were presented in Traditional Chinese. All stimuli and materials are provided in an OSF repository.[2]

---

[1] In terms of age groups: 8% between 18 and 19, 58% in their 20s, 20% in their 30s, 8% in their 40s, 5% in their 50s, and 1% 60 years old or above. In terms of gender: 69% female, 29% male, 1% non-binary, and 1% would prefer not to say.

[2] https://osf.io/463ts/

As with Dingemanse et al. (2016), the selected ideophones fall into five broad semantic categories[3]: SOUND, MOTION, COLOR-VISUAL, SHAPE, and TEXTURE as in Table 1. SOUND items included animate sounds as well as environmental sounds. MOTION terms include animate motion as well as inanimate objects' motions. COLOR-VISUAL appearance refers to colors or static visual appearances. SHAPE relates to abstract features like roundness but also included specific forms. TEXTURE refers mainly to surface feeling and consistency of referents. Specific examples of each category are provided in Table 1.

To elicit the Igbo ideophones, we adhered to the five chosen semantic categories, with a minimum of five ideophones per category, for example, see Table 1. Igbo ideophones were elicited by first drafting a list of ideophones descriptively reported throughout Emenanjo (2015, pp. 662–627), Maduka (1983), and Uchechukwu (2007, 2017). This list served to assist the consultant-led elicitation, during which our consultant, in a laboratory setting, revised the list according to which ideophones they considered most recognizable. All Igbo ideophones provided were from the Central Igbo variety. The consultant (*F*)[4] was a linguist, bilingual in English, and provided the English translations for each ideophone.

Two native Cantonese speakers, both of whom are balanced bilingual in English, provided the Traditional Chinese translations based on the English translations of the ideophones of all three languages; see the Supporting Information for the translations. The translators were instructed to control for the number of Chinese characters so that most translations are two characters in length, with the exception of 30 out of 32 SOUND ideophones. These comprise three characters because of the presence of an obligatory suffix seng[1]聲, which means "sound." Due to the resampling within the same semantic category for Experiment 1 (see below) and the even spreading of foils in Experiment 2, this discrepancy between two and three characters presented no impactful confound in the design of the experiments.

We prepared four different conditions of the experiment, each with a different random pairing of target and foil. For example, the true meaning of the Japanese ideophone *sarasara* is "smooth" (presented as 順滑 to participants). *Sarasara* belongs to the sensory category of TEXTURE and was paired, respectively, with four other TEXTURE ideophones: "slippery" 平滑, "bumpy" 凹凸, "dry, brittle" 乾燥, and "coarse, rough" 粗糙. These four different pairings were then presented depending on the condition. As in Dingemanse et al. (2016), these within-category pairings ensured that the two-alternative forced choice (2AFC) task was quite difficult.

### *2.3. Procedure*

Due to COVID-19 and the imposed social distancing regulations in Hong Kong, our study was conducted online with participants using their own computers. The experiment was presented in Cantonese with PsychoPy3 v. 2021.1.4 (Peirce et al., 2019). The training phase

[3] It should be noted that this not the only classification of sensory domains that ideophones occupy. Perhaps most striking is the absence of psychomimetic ideophones, well-known from ideophone systems in East-Asian languages (Akita, 2009); see also McLean (2020) for a more constrained system based on Japonic languages and Van Hoey (in press) for a broader typology.

[4] The consultant is from Amurri, Enugu State, Nigeria and is fluent in Central Igbo (Standard Igbo).

Table 1
Examples of ideophones per sensory category and per language (total number of types per category provided in brackets)

| Category | Igbo | Japanese | Korean |
|---|---|---|---|
| SOUND | *táwàm* “sound of a slap” (10) | *wanwan* “barking of dog” (10) | *eong’eong* “crying” (9) |
| MOTION | *yògìyògì* “wobbly” (8) | *daradara* “slow movement” (8) | *chingching* “circular motion” (10) |
| COLOR-VISUAL | *chámćhám´* “flashing” (8) | *meromero* “blurred” (5) | *chaengchaeng* “blazing sunshine” (9) |
| SHAPE | *kpụ́rụ́kpụ́rụ´* “round and chubby” (6) | *panpan* “full, bursting” (9) | *dunggeuldunggeul* “roundness” (8) |
| TEXTURE | *pị̀rị̀pị̀rị̀* “runny (diarrhea)” (7) | *nurununu* “slimy” (10) | *maekkeunmaekkeun* “smooth” (9) |

was followed by the test phase. For both phases, participants were told that they would hear words that sound like what they mean from three different languages and would have to guess the correct translation from two options, presented on the left and right sides of the screen. Participants were instructed to press either the F key or the J key on their keyboard to select the left or right stimulus. To familiarize participants with the experiment format, three sound ideophones were presented as practice items before a training phase: Japanese *wanwan* ワンワン "sound of dog barking," Korean *kungkung* "pounding from a big and heavy object/obese person walking," and Igbo *vùm* "revving of a car engine." Timings of presentation were identical to the original experiment of Dingemanse et al. (2016): per trial, participants first heard a stimulus and then, after 1,350 ms, two translations appeared. After another 1,000 ms, the stimulus was repeated and the participant chose which translation they felt sounded like the meaning of the stimulus. We moved to the next trial with a short interval of 500 ms. Every participant was tested on 126 items, which were randomly presented. For each trial, the position of the foil and the target was randomly allocated. The average duration of the actual experimental phase was around 4 min (256 s), with an average of 2 s per trial.

### 2.4. Results

Each participant was randomly allocated to a condition (condition 1: $n = 25$, condition 2: $n = 31$, condition 3: $n = 24$, and condition 4: $n = 31$). Since they were all subjected to the same ideophone stimuli and only differed in terms of foils, we did not take a post hoc sample of these groups. There were no significant differences in reaction times of the trials where participants had to choose a translation (mean reaction time= 2,038 ms). All analyses were performed with R (R Core Team, 2021). A full overview of packages and used versions is also presented in the Supporting Information.

The mean proportion of correct responses was 61%, which is above chance level (one sample *t*-test: 95% confidence intervals: 58.8% – 63.4%, $\mu = 0.5$, $p < .001$). Crucially, this is similar to the results for the non-resynthesized stimuli in Dingemanse et al. (2016). As for the language-specific results, Japanese showed the best performance ($m = 66.6\%$), followed by Igbo ($m = 60.0\%$), and then Korean ($m = 56.9\%$). For analysis by sensory category, the mean proportion of accuracy is situated above chance level ($\mu = 0.5$, $m = 61\%$, $t = 9.67$, $df = 125$, $p < .001$). As with Dingemanse et al. (2016), sound ideophones have the highest accuracy of the five semantic categories ($m = 67\%$). Aggregating by language and sensory category (Fig. 1), the same tendencies occur: Japanese has the highest accuracy, while for the other two languages there might be a language-specific interaction effect.

We ran two models with accuracy of the guess as outcome (correct vs. incorrect), and language and sensory category as fixed effect predictors, and varying intercepts for participant and item. An analysis of variance showed that there is no significant interaction between the fixed predictors ($\chi^2(8) = 4.55$, $p = 0.80$), that is, a more complex analysis with an interaction effect would not be a better fit for the data. This model showed that participants guessed Japanese items significantly better compared to the reference level ($\beta = 0.23$, logit difference: +0.27, $SE = 0.12$, $p = .02$) (see the Supporting Information for the full model). However, our main interest for this experiment concerns the different sensory categories. That is why we ran another mixed-effects logistic regression model with correctness of the guess

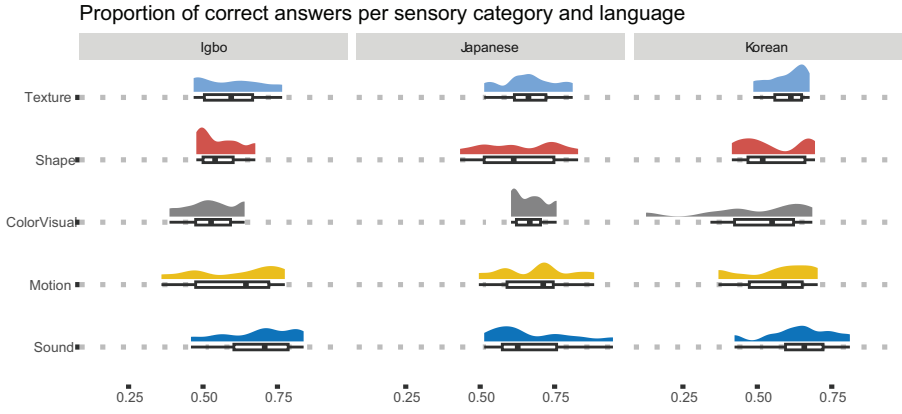


Fig. 1. Proportion of correct answers per sensory category and language.
*Note*. Shown are the boxplots and accompanying density plots, which, respectively, show the summary statistics and distribution.

Table 2
Regression weights for the analysis of fixed effect sensory category with color-visual as the intercept. Note that estimates are shown with their log-odds values. Significance codes: <0.001 '***', 0.001 '**', 0.01 '*'

| | $\beta$ | $SE$ | $p$ |
|---|---|---|---|
| intercept = $\text{sense}_{\text{Color-Visual}}$ | 0.24 | 0.11 | .034* |
| $\text{sense}_{\text{Motion}}$ | 0.21 | 0.15 | .170 |
| $\text{sense}_{\text{Shape}}$ | 0.12 | 0.16 | .440 |
| $\text{sense}_{\text{Sound}}$ | 0.52 | 0.15 | <.001*** |
| $\text{sense}_{\text{Texture}}$ | 0.33 | 0.14 | .019* |

as outcome (correct or incorrect). The predictor was the sensory category (COLOR-VISUAL, MOTION, SHAPE, SOUND, or TEXTURE). Random effects included varying intercepts of participant and item. The model's intercept, corresponding to COLOR-VISUAL, was at 0.56 and was significant. The other sensory categories were compared to COLOR-VISUAL as the baseline (Table 2). Only SOUND and TEXTURE perform better than COLOR-VISUAL. For SOUND, this was expected, both from the literature on onomatopoeia as well as the results from the replicated study. The model's area under curve was 0.65 and $d'$ 0.54, both of which indicate moderate performance.

Summarizing, we now have a baseline for the pool of Cantonese speakers, which are shown to be sensitive to iconic words, so that we can further conduct a memory task and a guessability task for ideophones. SOUND and TEXTURE ideophones were guessed best. Curiously, ideophones in the visual modality (COLOR-VISUAL, SHAPE, MOTION) were guessed worst. This may be due to the lack of supporting gestures that often co-occur with ideophones (Nuckolls, 2020).

## 3. Experiment 2: Memory task and the 2AFC reassessment task

The second experiment investigates whether participants recall the form-meaning mappings of ideophones better than adjectives. This experiment follows a study conducted by Lockwood, Dingemanse, et al. (2016). They investigated how sensitive Dutch speakers are to iconicity in Japanese ideophones and adjectives. Forty participants, all native speakers of Netherlandic Dutch, first learned combinations of true and coerced (false or sometimes opposite) translations of ideophones preselected for guessability. Participants' learning was then tested by asking whether they had learned a particular combination of an ideophone and a translation. Results show that true translations were recalled better. Next, participants were told that some of the translations they learned were incorrect and were asked to try to disregard any learned associations in performing a new guessing task that gave a choice between the true and foil translations. Overall, the true translations of Japanese ideophones were accurately selected even for ideophones learned with foil translations. Later, the same procedure was conducted using a set of Japanese adjectives, also pre-tested for guessability. Comparison with the ideophone data shows that there was no memorization effect for Japanese adjectives and, in the follow-up 2AFC reassessment task, correct translations were only selected at chance level. Our design largely follows that experiment design, though with two important differences. The Japanese adjectives we use in our study are the same as those tested in Lockwood, Dingemanse, et al. (2016), extended with semantic equivalents in Korean and Igbo, with antonyms as foils. The ideophones we use are from our own Experiment 1, with translations from different semantic categories in the same language as foils.

### *3.1. Participants*

The current study had 301 participants,[5] who reported to be native speakers of Hong Kong Cantonese and did not have knowledge of the language they were subjected to in the experiment (either Japanese, Korean, or Igbo). As an incentive for participation, we organized a lucky draw with the prize ranged from 50 HKD to 1,000 HKD. We did not exclude any participants based on outlying reaction times, because we verified that inclusion or exclusion made no significant difference for the analysis. Participants were randomly allocated to one of the 12 conditions presented in Table 3.

### *3.2. Stimuli*

The same Japanese, Korean, and Igbo ideophones used in Experiment 1 were also used in Experiment 2. Both ideophone and adjective stimuli were designed so that their translation was presented as true or coerced meanings (i.e., foils). Half of the participants learned the true meanings of stimuli, while another half learned the coerced meanings of the same stimuli.

[5] In terms of age groups: 9% in their teens, 51% in their 20s, 26% in their thirties, 9% in their 40s, 4% in their 50s, and 1% 60s or older. In terms of gender: 68% female, 30% male, 1% non-binary, and 1% preferred not to disclose this. Ninety-six percent of them were right-handed.

Table 3
Schematic overview of the four conditions in Experiment 2

| | Condition 1 | Condition 2 | Condition 3 | Condition 4 |
|---|---|---|---|---|
| Type | Ideophone | Ideophone | Adjective | Adjective |
| True | Group A | Group B | Group A | Group B |
| Coerced | Group B | Group A | Group B | Group A |

This resulted in four separate conditions (ideophones/adjectives × true/coerced) as given in Table 3.

Coerced translations for ideophones were generated first by randomly selecting a translation from an ideophone of another semantic category in the stimuli from the same language. For example, the Igbo ideophone *tawàm* has the translation "sound of a slap" 掌摑聲 and belongs to the semantic category of SOUND. The coerced translation for *tawàm* is "scattered" 散亂, which comes from the translation of an ideophone belonging to the semantic category of COLOR-VISUAL *yágáyágá*. We made sure to balance the number of pairings of semantic categories so that, for example, SOUND ideophones were not consistently assigned COLOR-VISUAL coerced translations. For Korean, three additional coerced translations "dark" 暗, "hard" 硬, and "slender" 苗條 were generated for the sake of balancing.

We used the same Japanese adjectives ($n = 38$) as those in Lockwood, Dingemanse, et al. (2016) original study and had our consultants translate them in Korean and Igbo. Both true and coerced translations were adopted from that study and translated into Cantonese by a native speaker (male in his 20s). Coerced adjective translations were simple opposites of an adjective's true translation,[6] for example, "big" (true) versus "small" (coerced). Japanese, Korean, and Igbo adjectives were recorded in a sound booth with an earset microphone and an Onyx Blackjack USB Recording Interface. All recordings were normalized. The consultants for Korean and Japanese adjectives were females in their 20s and 30s, the consultant for Igbo was a male in his 20s.

All stimuli were presented orthographically in Romanized scripts. Japanese was presented in Hepburn Romanization. Korean and Igbo were presented in altered Romanizations to cater to the phonetic perception of Cantonese-speaking participants and prevent any confusion or distraction from redundant orthographic representations. None of the alterations compromised any original minimal pairs in our stimuli. These alterations were based on the result of a perceptual pretest conducted with one native speaker of Cantonese (23 years, male). For Igbo, first, all diacritics were removed from the Ọnwụ alphabet. Next, several alterations were conducted for perceptual reasons. For example, the <g> and <k> in consonant clusters <gb> and <kp> were removed due to perceptual inaccuracy. The voiceless labial [p] was then replaced with <b>, so that <kpọ́kpọ́kpọ́> "hard" 硬 was altered to <bobobo>; <ị> was changed to <e> in keeping with our Cantonese consultant's perception of the vowel.

[6] This adoption of adjectives and their antonyms entails that both true and coerced translation fall in the same domain, while for ideophones they are tested across sensory domain. This means that we are not able to include sensory category as a relevant factor in the analysis, because they are not equal.

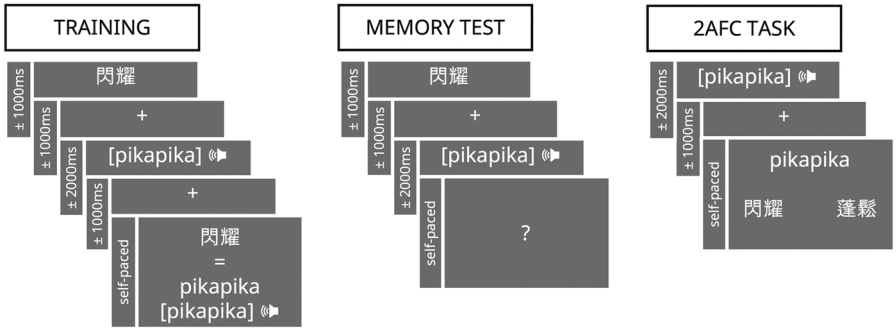


Fig. 2. Diagram that illustrates the design of Experiment 2.
*Note*. The ideophone *pikapika* means "flashing." The presented meaning *sim*[2] *jiu*[6] 閃耀 means "flashing" as well. The foil *pung*[4] *sung*[1] 蓬鬆 means "fluffy."

This did not result in any overlap with the original <e> of the Ọnwụ alphabet. Note that this simplification of the orthography does not entail a change in terms of potential phonosemantic mapping of the auditory stimulus (e.g., /b/ vs. /p/) but only serves to aid the Cantonese participants with the required tasks.

Similarly, for Korean, our orthography was a mix of revised Romanization of Hangul and the McCune–Reischauer Romanization system aimed at helping Cantonese participants. We chose the revised Romanization of Hangul for its lack of diacritics but maintained some aspects of the McCune–Reischauer system for phonetic reasons. For example, the McCune–Reischauer system indicates that voiced obstruents are devoiced in the word initial position while the revised Romanization of Hangul does not. Our orthographic representation reflected this, for example, <bodeulbodeul> was modified to <podeulbodeul> "fluffy" 蓬鬆. Tense consonants underwent the following alterations: <gg> became <g>; <kk> became <g>; <bb> became <b>; <pp > became <b>; and <jj> became <j>. (For a full list of alterations, see the Supporting Information.)

### *3.3. Procedure*

Lockwood, Hagoort, and Dingemanse (2016) consisted of three main parts as in Fig. 2: a training (learning) round, which was repeated once, a testing round, and a two-alternative forced-choice task. In the two training rounds, the order in which the true and coerced stimuli were presented to participants was randomized per participant but the items and conditions were fixed across participants. The meaning of each item was presented with an orthographic presentation, explained in the previous section. This was self-paced: when participants were ready to move on, a new fixation cross was presented and the next item was taught.

During the memory task, we presented to the participants either word pairs they had learned or pseudorandomized pairings of items and translation that they had not seen together before. Participants had to indicate on the keyboard whether they had seen a particular pairing or not.

Table 4
Distribution of participants across the 12 conditions

| Group | Igbo | Japanese | Korean |
|---|---|---|---|
| Ideophones 1 | 26 | 19 | 25 |
| Ideophones 2 | 22 | 21 | 25 |
| Adjectives 1 | 23 | 28 | 24 |
| Adjectives 2 | 28 | 30 | 30 |

They first saw a word for 1,000 ms with a 100-ms jitter. Next, a fixation cross announced the presentation of an audio stimulus for 2,000 ms with a 200 ms jitter. Finally, they saw a question mark and were instructed to answer as fast as possible whether they had learned the word or not.

In the final test, the reassessment task (i.e., a follow-up two alternative forced choice task), we asked participants to reassess what they had initially been instructed to commit to memory. Participants were now told that half of the word pairings they had memorized were wrong and that that actually half the stimuli had the opposite meaning to what they were originally taught. Participants were then asked to forget everything they had just memorized, and now reassess to select the translation they felt to be most natural. They were presented with an audio stimulus, for 2,000 ms with a 200-ms jitter. After a fixation cross, they were shown the orthographic item, with either the true meaning or the pseudorandomized meaning. In that last self-paced part, they were instructed to select the meaning that felt most natural to them.

### *3.4. Results*

#### *3.4.1. Memory task*

The distribution of the condition groups is shown in Table 4. As the exploratory visualization (Figs. 3 and 4) shows, there was no clear trend between accuracy depending on condition (true or coerced), type (ideophones or adjectives), or language (Igbo, Japanese, or Korean). We ran a mixed-effects logistic regression model, in which the dependent variable was the correct or incorrect identification of a learned form-meaning pairing. Our model included interactions between type (ideophone or adjective), condition (true or coerced), and language (Igbo, Japanese, Korean) and had random effects on participant, condition, and item. The model showed no significant fixed effects (see Table 5). We take this to mean that memory results are not consistent across languages and experiments. Furthermore, there was no added memory benefit depending on whether an item is ideophonic or adjectival or learned in a true condition versus a coerced condition.

Our results were not consistent with Lockwood, Dingemanse, et al.'s (2016) original study, in which they found that Japanese ideophones were learned significantly better in the true condition than in the coerced condition: 81.1% versus 71.1%. Their model estimated that ideophones of the true condition were answered 9.53% more accurately than ideophones learned in the opposite condition, with a significant fixed effect of condition ($\beta = -0.5978$,

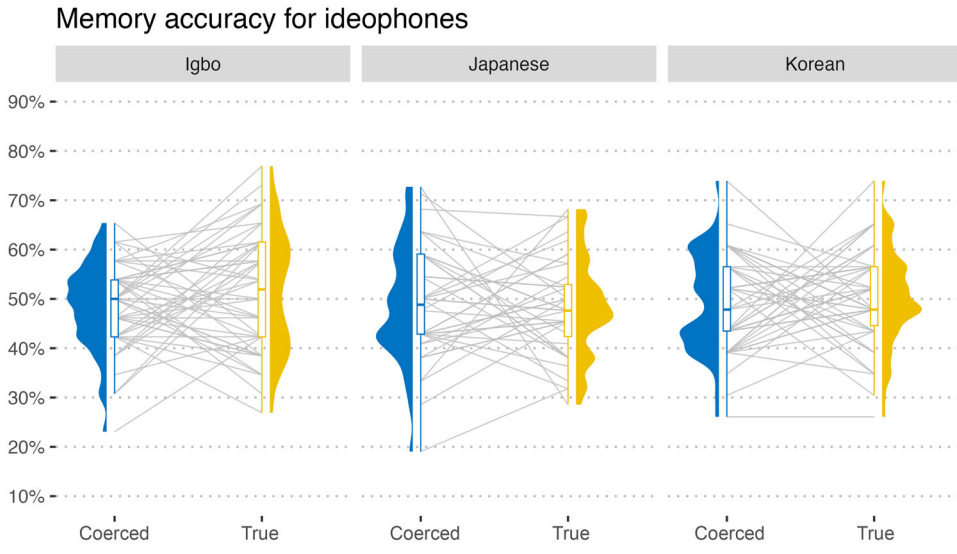


Fig. 3. Accuracy performance for ideophones in the two conditions (coerced vs. true).
*Note*. Graphs show the density distribution as well as basic summary statistics in the boxplot. Gray lines connect the mean values for ideophone types.

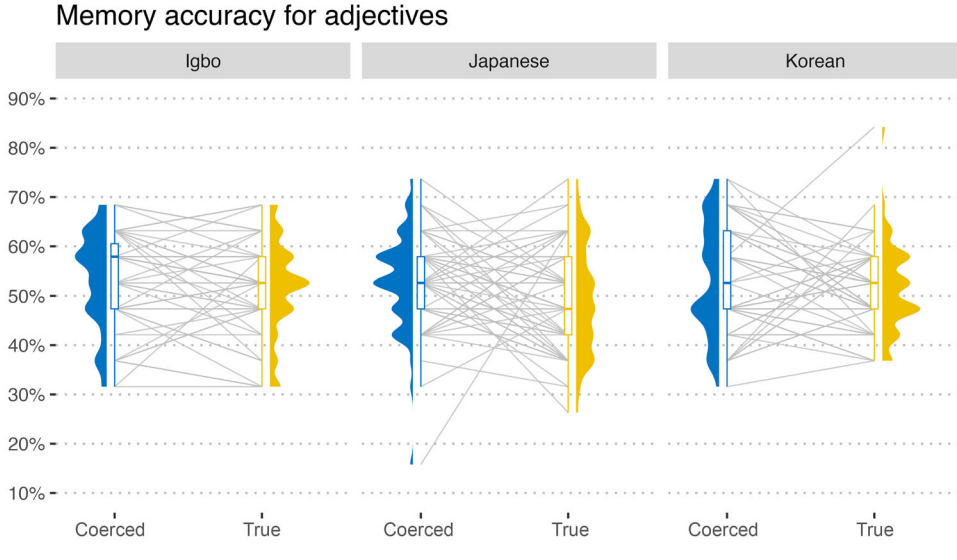


Fig. 4. Accuracy performance for adjectives in the two conditions (coerced vs. true).
*Note*. Graphs show the density distribution as well as basic summary statistics in the boxplot. Gray lines connect the mean values for adjective types.

Table 5
Regression coefficients for the memory model

| | $\beta$ | *SE* | *p* |
|---|---|---|---|
| intercept = coerced adjectives in Igbo | 0.14 | 0.1 | .186 |
| $\text{type}_{\text{ideophone}}$ | –0.17 | 0.15 | .235 |
| $\text{condition}_{\text{true}}$ | –0.08 | 0.09 | .382 |
| $\text{language}_{\text{Japanese}}$ | –0.03 | 0.15 | .855 |
| $\text{language}_{\text{Korean}}$ | 0 | 0.15 | .995 |
| $\text{type}_{\text{ideophone}}$:$\text{condition}_{\text{true}}$ | 0.2 | 0.12 | .110 |
| $\text{type}_{\text{ideophone}}$:$\text{language}_{\text{Japanese}}$ | 0.02 | 0.21 | .935 |
| $\text{type}_{\text{ideophone}}$:$\text{language}_{\text{Korean}}$ | 0 | 0.2 | .981 |
| $\text{condition}_{\text{true}}$:$\text{language}_{\text{Japanese}}$ | −0.07 | 0.13 | .580 |
| $\text{condition}_{\text{true}}$:$\text{language}_{\text{Korean}}$ | 0 | 0.13 | .974 |
| $\text{type}_{\text{ideophone}}$:$\text{condition}_{\text{true}}$:$\text{language}_{\text{Japanese}}$ | −0.07 | 0.18 | .714 |
| $\text{type}_{\text{ideophone}}$:$\text{condition}_{\text{true}}$:$\text{language}_{\text{Korean}}$ | −0.08 | 0.18 | .647 |

$p < .001$). They did not find this effect between matched adjectives in the true versus the coerced condition, respectively, 79.1% and 77% correct on average ($\beta = -0.1256$, $p = .379$). Adjectives in the true condition were remembered 1.81 percentage points more accurately than in the coerced condition. Their follow-up study (Lockwood, Hagoort et al., 2016), which looked at ideophones only, showed similar results: in the correct condition, items were remembered on average 86.7%, while in the coerced condition this was 71.3%. The model showed a fixed effect of condition ($\beta = -0.5514$, $p < .001$) and estimated that ideophones in the true condition were learned 8.1% more accurately than in the coerced condition. Contrast these numbers for Japanese with the means for true ideophones (44.8%), coerced ideophones (52.1%), true adjectives (48.1%), and coerced adjectives (53.4%). These all hover around chance level (50%).

As our results indicate, there are apparently fewer benefits for memorizing the pairings between true and coerced items. A number of possible reasons may account for this. There are some differences in stimuli selection: Our stimuli were composed of the ideophones adopted from Dingemanse et al. (2016), a set not pre-selected for guessability, and the adjectives from Lockwood, Dingemanse, et al. (2016), pre-selected for guessability like the ideophones in that study. This means the adjectives had a leg up relative to the ideophones, different from the original studies in which all ideophones and adjectives were selected in similar ways. Another possibly relevant difference is that we test words from multiple languages with potentially different properties and baseline iconicity levels. Some consider Igbo adjectives to comprise a closed class with as few as eight members (Emenanjo, 2015). Stative verbs and abstract nouns are often used to create adjectival meanings in Igbo and, for similar languages, Ameka (2001) argues that ideophones are also heavily relied on as well. This somewhat periphrastic nature of how adjectival meaning is expressed in Igbo could lead to a greater variability in baseline iconicity of our Igbo adjective translations than researchers might expect for adjectives from some other languages. Finally, there are differences in

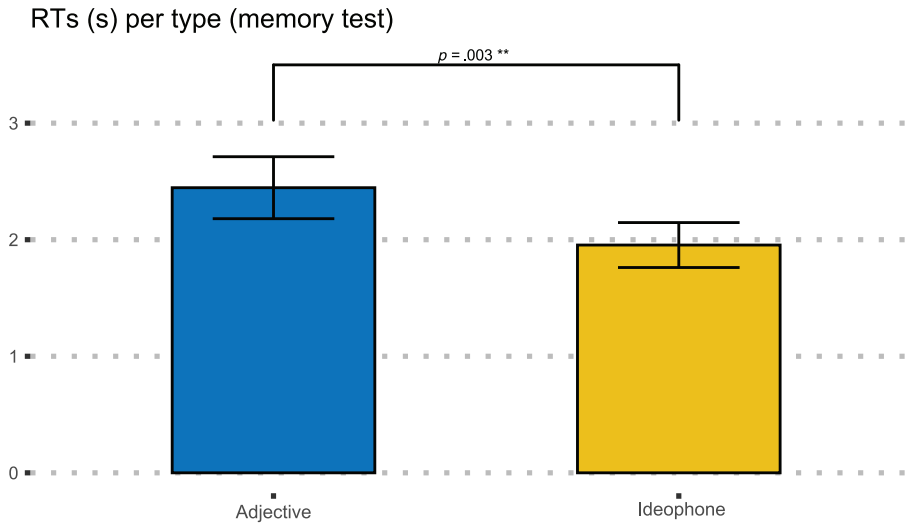


Fig. 5. Response times (s) per type in the memory test.

the linguistic backgrounds of the participant pools and in the size of the participant pools (301 participants in our study vs. $32 + 30 + 40 = 102$ participants in the two studies by Lockwood, Dingemanse et al., 2016, and Lockwood, Hagoort et al., 2016).

To further explore the two conditions, we decided to inspect response times in the memory task. Based on prior work, one might expect that items may be responded to faster if participants can rely on iconic or structural cues to meaning (Kovi ć, Plunkett, & Westermann, 2010), conceptual domain (Nygaard, Cook, & Namy, 2009), or word class (Farmer, Christiansen, & Monaghan, 2006). To the extent that ideophones provide such cues, they should be responded to faster. As Fig. 5 shows, responses to ideophones were significantly faster than adjectives. A linear regression with type (adjective or ideophone) as predictor and reaction time as dependent variable showed a significant intercept for adjectives ($\beta = 2.44$, $p < .001$) and a significant difference for ideophones ($\beta = -0.49$, $p = .003$). When we look at the reaction times for different types (adjective or ideophone) and conditions (coerced or true), it turns out that ideophones were responded to fast regardless of condition, whereas our participants were significantly slower when guessing true adjective form-meaning pairings, as presented in Fig. 6. The results of a mixed-effects linear regression in which type and condition interact to predict reaction time show a significant effect of true condition and an interaction between ideophone and true condition. Coefficients are presented in Table 6.

In sum, adjectives presented with their actual translations trigger the slowest responses while ideophones are fast across conditions. While response times do not clearly correspond to differences in accuracy, it is possible that participants are tapping into the structural markedness of ideophones in general. The similarly low response time for adjectives with antonymic translations in the coerced condition is in line with what Nygaard et al. (2009)

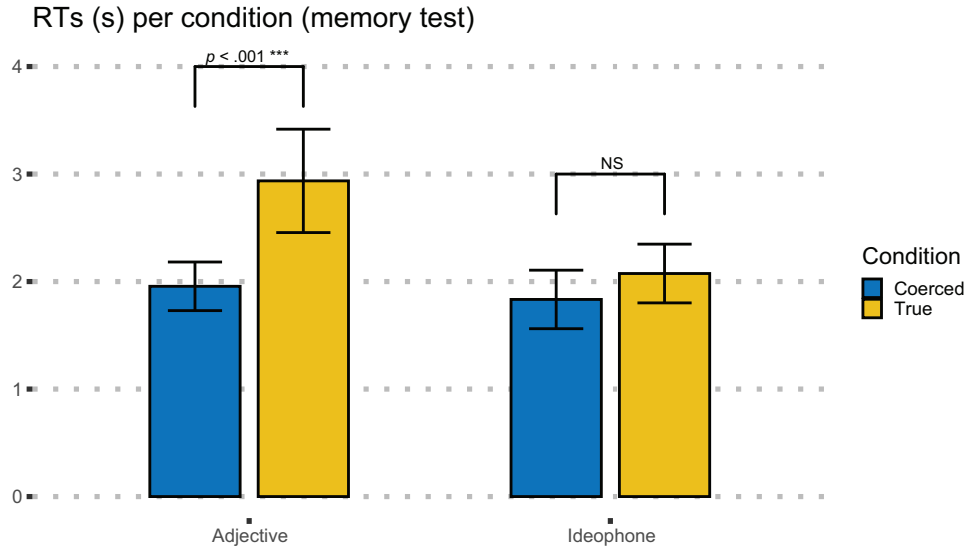


Fig. 6. Response times (s) per type and condition in the memory task.

Table 6
Regression analysis for response times in the memory task

| | $\beta$ | $SE$ | $p$ |
|---|---|---|---|
| intercept = $\text{type}_{\text{adjective}} \times \text{condition}_{\text{coerced}}$ | 1.96 | 0.25 | <.001 |
| $\text{type}_{\text{ideophone}}$ | −0.11 | 0.36 | .757 |
| $\text{condition}_{\text{true}}$ | 0.98 | 0.23 | <.001 |
| $\text{type}_{\text{ideophone}} \times \text{condition}_{\text{true}}$ | −0.74 | 0.32 | .021 |

found for antonyms of Japanese words, but the markedly higher response times to adjectives in the true condition are not. Follow-up work is needed here.

### *3.4.2. Two alternative forced choice reassessment task*

In the follow-up 2AFC reassessment task, participants selected the true meanings of both ideophones and adjectives above chance level ($\mu = 0.5$). For ideophones, the mean was at 60.9% ($\mu = 0.5$, $t = 7.48$, $df = 126$, $p < .001$, 95% CIs = [0.58–0.64]) and for adjectives at 57.8% ($\mu = 0.5$, $t = 5.59$, $df = 113$, $p < .001$, 95% CIs = [0.55–0.61]). An unpaired two-sample $t$-test with the Welch correction showed that there was no significant difference between these two means ( $t_{(238.94)} = -1.509$, $p = .13$). These results were as expected: Lockwood, Dingemanse, et al. (2016) report that participants selected the true meanings of Japanese ideophones with 72.3% accuracy on average and adjectives with 63% accuracy.

Upon closer inspection (Fig. 7) it can be seen that there is a difference between conditions in our study: Items taught in the true condition perform extremely well, while coerced items are less likely to get to chance level. Such trends for different conditions were expected:

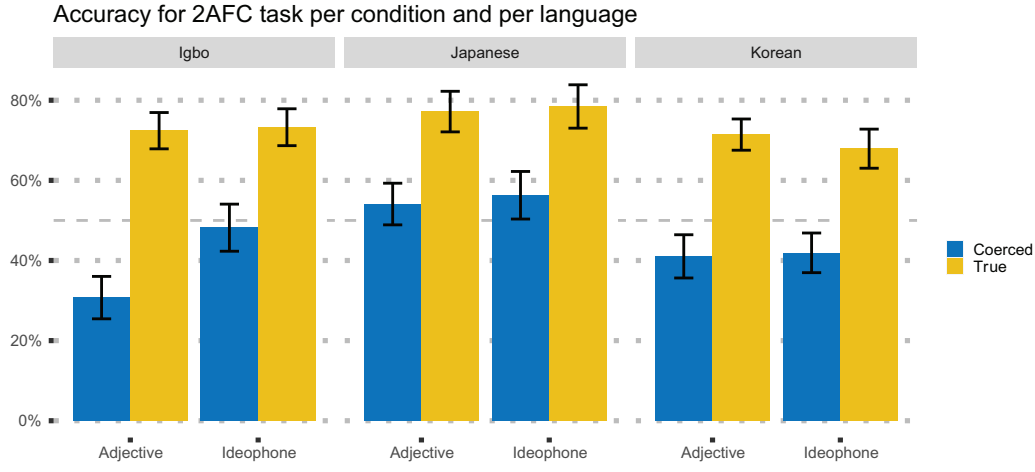


Fig. 7. Accuracy percentages for the 2AFC reassessment task presented per condition (coerced vs. true) and per language (Igbo, Japanese, Korean) for both adjectives and ideophones.

Lockwood, Dingemanse, et al. (2016) found for ideophones taught in the true condition a 2AFC accuracy of 75.1% versus 69.5% for ideophones in the coerced condition. For adjectives, they report, respectively, 65.1% and 60.8%. Lockwood, Hagoort, et al. (2016), respectively, report for ideophones 77.3% and 68.6%. Summarizing, our data confirm the trends identified before but show that the differences between the conditions are much more dramatic across languages than any differences between word classes.

A mixed effects logistic regression model was built with accuracy (correct or incorrect) as the dependent variable and type (adjective or ideophone), condition (true or coerced), and language (Igbo, Japanese, or Korean) as predictive variables. We verified that there are interaction effects between the fixed variables and incorporated condition per participant and item as random effects. The model's explanatory power is substantial ($R^2 = 0.45$). With Igbo adjectives in the coerced condition as the intercept, we find the following significant fixed effects (see Table 7). Note that if items were ideophones, they were reassessed more correctly than if they were adjectives, especially if they were taught in the true condition.

One follow-up question was whether there is any difference between participants moving from one part of the experiment (memory task) to another (reassessment task). It can be hypothesized that participants who were good at the memory task will be very good at selecting the correct answer in the reassessment task when the condition is true but may suffer more from the confusion that the coerced conditions bring. An "average" participant, on the other hand, may have less variation: not doing terribly well in the memory task but also not shining in the reassessment task. To analyze this, we present the plots in Fig. 8. It can be seen there is no general pattern in whether high scores for one task also correlate to high scores in the other. Neither Pearson's rank correlation nor Spearman's correlation showed any clear patterns in any direction (values hovered around 0).

Table 7
Regression analysis results for the 2AFC reassessment task

| | $\beta$ | $SE$ | $p$ |
|---|---|---|---|
| intercept = coerced adjectives in Igbo | −1.33 | 0.24 | < .001 |
| $\text{type}_{\text{ideophone}}$ | 1.13 | 0.34 | .001 |
| $\text{condition}_{\text{true}}$ | 2.60 | 0.31 | < .001 |
| $\text{language}_{\text{Japanese}}$ | 1.51 | 0.33 | < .001 |
| $\text{language}_{\text{Korean}}$ | 0.76 | 0.34 | .024 |
| $\text{type}_{\text{ideophone}}$: $\text{condition}_{\text{true}}$ | −1.09 | 0.43 | .012 |
| $\text{type}_{\text{ideophone}}$: $\text{language}_{\text{Japanese}}$ | −1.04 | 0.48 | .032 |
| $\text{type}_{\text{ideophone}}$: $\text{language}_{\text{Korean}}$ | −0.98 | 0.47 | .039 |
| $\text{condition}_{\text{true}}$: $\text{language}_{\text{Japanese}}$ | −1.23 | 0.42 | .003 |
| $\text{condition}_{\text{true}}$: $\text{language}_{\text{Korean}}$ | −0.89 | 0.43 | .038 |
| $\text{type}_{\text{ideophone}}$: $\text{condition}_{\text{true}}$: $\text{language}_{\text{Japanese}}$ | 1.05 | 0.62 | .089 |
| $\text{type}_{\text{ideophone}}$: $\text{condition}_{\text{true}}$: $\text{language}_{\text{Korean}}$ | 0.71 | 0.60 | .237 |

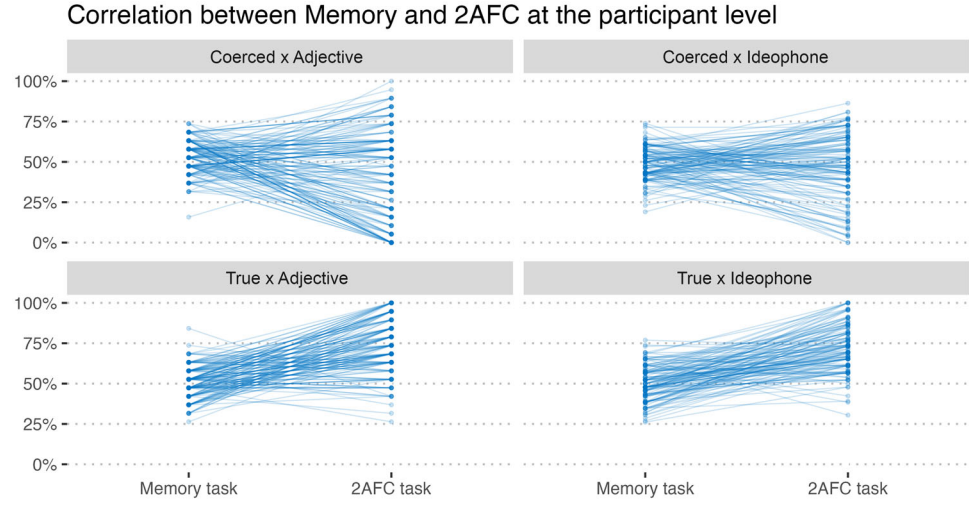


Fig. 8. Correlation between the memory task and the 2AFC reassessment task at the participant level.

We also ran an analysis at the item level, focused on cognates between Cantonese on the one hand and Japanese and Korean stimuli on the other. The analysis comparing the means of cognates versus non-cognates within a task and a condition did not show any significant results. (See the Supporting Information for the full analysis.)

### *3.4.3. Likelihood of reassessment*

The memory task (Section 3.4.1) showed no significant effect between adjectives and ideophones, nor between coerced or true mappings. This did not corroborate what Lockwood, Dingemanse, et al. (2016) had found. The 2AFC reassessment task, on the other hand, did show significant differences, in line with the findings of Lockwood, Dingemanse, et al.

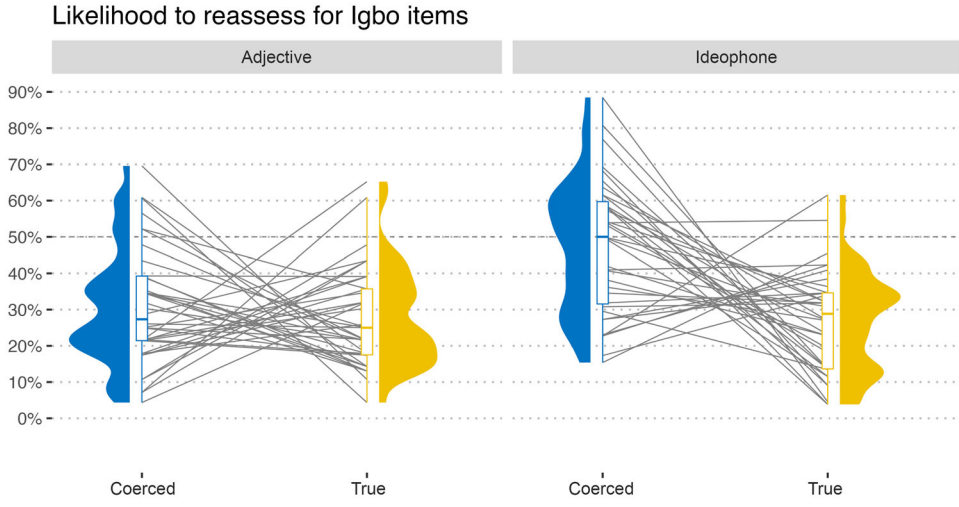


Fig. 9. Likelihood of reassessment for Igbo items.

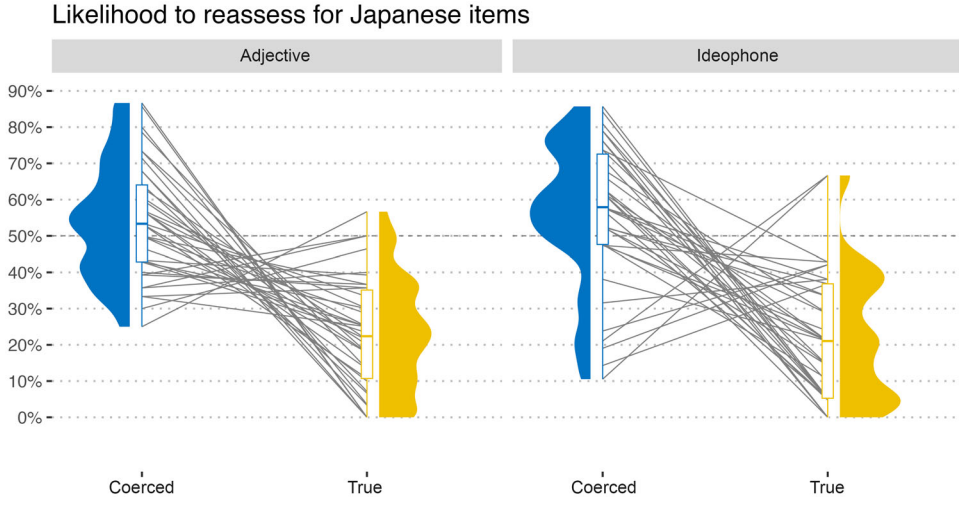


Fig. 10. Likelihood of reassessment for Japanese items.

(2016). Let us further inspect the results from the reassessment task by investigating the likelihood of abandoning what had been taught in the exposure round in favor of an alternative choice, that is, "flip-flopping" to the alternative.

As can be seen from Fig. 9 to Fig. 11, items learned in the true condition were overall less likely to be reassessed. In other words, participants were more confident for items that they learned with a true definition (i.e., an actual form-meaning pairing) than a coerced meaning. There are some differences between ideophones and adjectives as well. Ideophones, overall, were less likely to be reassessed if learned in the true condition yet more likely to be


Fig. 11. Likelihood of reassessment for Korean items.

Table 8
Regression analysis results for the likelihood of reassessment

| | $\beta$ | $SE$ | $p$ |
|---|---|---|---|
| intercept = coerced adjectives in Igbo | −1.07 | 0.15 | < .001 |
| $\text{condition}_{\text{true}}$ | −0.17 | 0.11 | .110 |
| $\text{language}_{\text{Japanese}}$ | 1.21 | 0.20 | < .001 |
| $\text{language}_{\text{Korean}}$ | 0.62 | 0.21 | .003 |
| $\text{type}_{\text{ideophone}}$ | 0.81 | 0.21 | < .001 |
| $\text{condition}_{\text{true}}$: $\text{language}_{\text{Japanese}}$ | −1.39 | 0.15 | < .001 |
| $\text{condition}_{\text{true}}$: $\text{language}_{\text{Korean}}$ | -0.48 | 0.15 | .001 |
| $\text{type}_{\text{ideophone}}$: $\text{condition}_{\text{true}}$ | −0.65 | 0.14 | < .001 |
| $\text{type}_{\text{ideophone}}$: $\text{language}_{\text{Japanese}}$ | −0.71 | 0.30 | .018 |
| $\text{type}_{\text{ideophone}}$: $\text{language}_{\text{Korean}}$ | −0.74 | 0.29 | .0012 |
| $\text{type}_{\text{ideophone}}$: $\text{condition}_{\text{true}}$: $\text{language}_{\text{Japanese}}$ | 0.53 | 0.21 | .011 |
| $\text{type}_{\text{ideophone}}$: $\text{condition}_{\text{true}}$: $\text{language}_{\text{Korean}}$ | 0.85 | 0.20 | < .001 |

reassessed if learned in the coerced condition than adjectives. A mixed effects logistic regression model was conducted with likelihood to be reassessed as the dependent variable (yes or no), and as fixed effects: condition (true or coerced), type (ideophone or adjective), and language (Igbo, Japanese, or Korean). A comparison between models showed that these fixed effects interact with each other. The random effects consisted of participant and item. With Igbo adjectives in the coerced condition as the intercept, we find (Table 8) that ideophones are significantly more likely to be reassessed than adjectives in the coerced condition, but not in the true condition, where they resist this. Compared to the baseline intercept, there also appear some difference across the languages: participants exhibited more reassessment with Korean and even more so for Japanese.

From Fig. 10, it can be surmised that Japanese adjectives behave similarly to Japanese ideophones. This is likely due to there being comparatively more (partly) iconic adjectives in the Japanese set, resulting in a larger average confidence for true adjectives. We wondered whether phonological overlap might have influenced the reassessment task. However, when incorporating the Levenshtein distance between stimuli and the Cantonese pronunciation of the character stimuli in the 2AFC model, we found no significant effects (see the Supporting Information), also ruling out the related point of reassessment.

We ran the analysis on two subsets of the data, depending on the value of the condition variable. In the coerced condition, we still saw the effect of ideophones ($\beta = 1.16$, $p < .001$). This means that ideophones when compared to adjectives, are more likely to have their alternative chosen in the coerced condition. In the true condition, on the other hand, we expect that ideophones, when compared to adjectives, are more likely to resist the selection of their alternative. This is corroborated by running the analysis on the subset of true condition items, where for the ideophone predictor we do not find a significant effect ($\beta = -0.02$, $p = .95$).

Finally, we also conducted a short analysis (see the Supporting Information) on the likelihood to reassess for different sensory categories within the ideophone data. Note that sensory category assignment was not available for the adjective data; hence, a comparison between item types was impossible. Within the ideophone data, a logistic mixed-effects model with random effects for participant and item and real sensory category (SOUND, COLOR-VISUAL, TEXTURE, SHAPE, MOTION) and condition (coerced or true) showed that sound items are significantly more likely to be reassessed (logit $\beta = 0.39$, $p = .001$) when compared to the reference level of color-visual ideophones in the coerced condition (logit $\beta = -0.20$, $p = .059$). The effect of the condition is significant and negative (true condition: logit $\beta = -.77$, $p < .001$). This effect is boosted for sound when interacting in the true condition (logit $\beta = -0.803$, $p < .001$). This means that, at least for ideophones, we can say that participants were better at tapping into auditory iconicity, which was not as available in ideophones belonging to other modalities, like the visual modality.

## 4. Discussion

We have studied how people guess, memorize, and reassess form-meaning associations they are exposed to in a learning task. Collectively, our findings allow us to better understand the accessibility of iconicity across languages. The accessibility of iconicity refers to an ability to create structure mappings between sound and meaning, based on what is given in a stimulus that facilitates interpretation of that stimulus. Further, the "interpretation of that stimulus" refers either to an ability to convert a stimulus into a representation more easily stored in the mind, or an ability to correctly infer its communicative intent, its meaning. By comparing the findings of our memory task (Experiment 2) with those of our guessing task (Experiment 1), it may be inferred that iconicity is more accessible if the task at hand requires participants to actively seek out structure mappings. We can come to this conclusion if we start by considering what participants did during our guessing task. They were told that they will see and hear words that sound like what they mean and must guess the correct translation.

Later, in the reassessment task of Experiment 2, participants were told that, even though they memorized these words in the previous task, half of what they were trained to remember was wrong. In this way, participants were asked to be active in two ways: (1) to deduce the meanings of stimuli in a way that relies on their supposed sound-meaning nature and (2) to reevaluate prior knowledge with this sound-meaning nature now in mind (see "Likelihood of Reassessment"). Both (1) and (2) are arguably more active than the rote memorization of foreign words coupled with translations. Interestingly, based on the regression analysis, the Japanese and Korean adjectives reassessed best overall were those with Chinese origins, for example, the Cantonese word *fuk*$^{1}$*zaap*$^{6}$ 複雜 "complicated" is *fukuzatsu* 複雑 and *bokjap* (複雜) in Japanese and Korean, respectively. This further alludes to an active approach on the participants' part in that they may have been potentially associating task stimuli with sound-meaning relations in their mother tongue. As shown above, however, there was no special statistically significant benefit of cognates versus non-cognates within the adjective groups. This means that even if there is an influence of cognate status for the adjectives it is highly likely confined to a few items only.

All of this is not to say that structure mapping may not be a possible strategy for better short-term memorization; it is just that if the option to use structure mapping is under the table, as opposed to on top of it, some participants may need to kick to know it is there and of use. Not all participants kick. Thus, if the nature of the task does not invite participants to seek out structure mappings, then those structure mappings are less accessible.

Our memory task results (Experiment 2) were not fully consistent with previous work (Lockwood, Dingemanse et al., 2016; Lockwood, Hagoort et al., 2016), a difference that motivates further research looking into possible effects of differences in stimulus selection methods, linguistic properties of stimuli, and participant pools. Importantly, however, we found that participants who were first trained to memorize coerced translations of ideophones were still more likely to choose the true translation when asked to reassess what they were originally instructed to memorize. Moreover, adjectives were reassessed (flip-flopping) correctly less often than ideophones. Reassessment also differed according to language: Japanese ideophones triggered more flip-flopping between the memory and reassessment tasks, followed by Korean.

Lockwood, Dingemanse, et al. (2016), whose memory task we replicated, would seem to contradict our conclusion about the accessibility of iconicity and memory tasks because Dutch participants memorized true meanings of Japanese ideophones better than coerced meanings and better than adjectives. However, a difference in stimulus selection may have stacked the deck against ideophones in our design. Prior to conducting the learning study, Lockwood, Dingemanse, et al. (2016) drew up a list of 95 ideophones and 87 adjectives in Japanese and then administered a pre-test, a guessing task. The 38 best guessed items in each word class were kept for the stimuli design of the main task, a memory task. In contrast, in our replication, we pitted their 38 pre-tested Japanese adjectives (and semantic equivalents in Korean and Igbo) against a set of ideophones from all three languages (from our own Experiment 1) that did not undergo the same preselection for guessability. So, the Japanese adjectives, but not the ideophones, represent a subset that is groomed for transparency and relative ease of

access. We leave it to future work to explore the consequences of stimulus selection for the accessibility and memorability of iconic cues across word classes in more detail.

Reaction time data from memory tasks have been used to argue for the presence of iconic affordances in words. Nygaard et al. (2009) make this argument by showing that native English participants responded faster to Japanese items (none of which were ideophones) with true word–meaning pairings as opposed to opposite word–meaning pairing or random word–meaning pairings. Interestingly, our results both align and contradict Nygaard et al. (2009). Ideophones are generally considered to be iconic and yielded the shortest RTs in the testing phase of our memory task, regardless of true or coerced conditions. However, contrary to Nygaard et al. (2009), adjectives in the true condition yielded RTs which far exceeded those of adjectives in the coerced condition. It is possible that the phonological structure of ideophones generally lends itself better to the perception of form-meaning mappings (no matter whether these mappings are accurate or not), while the phonological structure of adjectives is more arbitrary and thus leads to RTs which iconicity cannot easily answer for.

It is also important to note that iconic associations are context-dependent and subjective to some degree (Occhino, Anible, Wilkinson, & Morford, 2017) and that participants' own linguistic backgrounds may influence the way they perceive potential structure mappings. The influence of cross-linguistic differences in iconicity and of participants' language backgrounds are both factors which are usually neglected in studies of iconicity, but our results show that there is a lot of variability there, pointing to possibly language-relativized forms of iconicity; and we cannot exclude that iconic sensitivity itself may be shaped and constrained by participants' own language experiences. Indeed, a recent study found shared sound-meaning relations in ideophones across 13 languages, speaking to the potential for cross-linguistic structure mappings but only for meanings related to sound and movement (Thompson et al., 2021), speaking to the potential diversity of language-specific mappings in other semantic categories covered by ideophones.

Human analogical thinking is highly flexible. The price of this flexibility is that structure mappings can be prolific yet at the same time ephemeral. Propped up by expressive prosody, contextual cues, and a system of representational conventions that is to some degree inflected by language, ideophones can feel to native speakers as the ultimate union of sound and sense, with form and meaning seemingly inevitably connected. Divorced from this rich context and studied in isolation, with forms regularized and meanings reduced to single words, only the strongest and most salient form–meaning associations survive: those in the domain of sound and motion. Studied in simplified recall and forced choice guessing tasks, iconic mappings in ideophones seem to offer diminishing returns. This paradigm can be seen as a way of putting iconicity to the test in fairly severe conditions; in such conditions, the effects are relatively moderate, in line with prior work (Dingemanse et al., 2016), and some word class differences even seem to dissipate.

At the same time, the potential to perceive iconic form-meaning association is never far off. Our novel measure of reassessment (flip-flopping) probes the degree to which people are willing to reconsider form-meaning mappings to which they were previously exposed. We find that people are more likely to choose the new true meaning of ideophones (but not adjectives) despite having been trained on a false meaning in a previous task, especially within the

auditory modality. This is in-line with an approach to iconicity known as structure mapping (see Emmorey, 2014), which argues true meanings of ideophones are more attractive than previously trained false meanings given how well the true meanings are perceived to match to an ideophone's form.

Our results, in short, nuance simplistic notions about iconicity and at the same time bring home the human potential for perceiving similarity in words and their meanings across languages and across semantic domains.

## Acknowledgment

This study was funded by the General Research Fund (17603120) from the Research Grant Council of Hong Kong. This study was not preregistered.

## Open Research Badges

This article has earned Open Data and Open Materials badges. Data are available at osf.io/463ts/ and materials are available at https://osf.io/463ts/.

## References

Akita, K. (2009). *A grammar of sound-symbolic words in Japanese: Theoretical approaches to iconic and lexical properties of mimetics (日本語音象徴語文法:擬音　擬態語の類像的　語彙的特性への理論的アプローチ)* [Unpublished doctoral dissertation]. Kobe University.

Akita, K., & Dingemanse, M. (2019). Ideophones (mimetics, expressives). In *Oxford research encyclopedia of linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.477

Ameka, F. K. (2001). Ideophones and the nature of the adjective word class in Ewe. In F. K. E. Voeltz & C. Kilian-Hatz (Eds.), *Ideophones* (pp. 25–48). Amsterdam, The Netherlands: John Benjamins.

Archibald, L. M. (2017). Working memory and language learning: A review. *Child Language Teaching and Therapy*, *33*(1), 5–17.

Beck, D. (2008). Ideophones, adverbs, and predicate qualification in Upper Necaxa Totonac. *International Journal of American Linguistics*, *74*(1), 1–46.

Bertolo, S. (2001). *Language acquisition and learnability*. Cambridge, England: Cambridge University Press.

Best, C. T. (1994). The emergence of language-specific phonemic influences in infant speech perception. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from recognizing speech sounds to spoken word* (pp. 167–224). Cambridge, MA: MIT Press.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). York, England: York Press.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345–360. https://doi.org/10.1037/0096-1523.14.3.345

Boersma, P., & Weenink, D. (2021). Praat: Doing phonetics by computer (6.1.56) [Computer software]. www.praat.org

Bruner, J. S. (1979). Learning how to do things with words. In D. Aaronson, & R. Rieber (Eds.), *Psycholinguistic research: Implications and applications.*Hillsdale NJ: Erlbaum.

Dingemanse, M., Schuerman, W., Reinisch, E., Tufvesson, S., & Mitterer, H. (2016). What sound symbolism can and cannot do: Testing the iconicity of ideophones from five languages. *Language*, *92*(2), e117–e133. https://doi.org/10.1353/lan.2016.0034

Emenanjo, E. N. (2015). *A grammar of contemporary Igbo: Constituents, features and processes*. M and J Grand Orbit Communications.

Emmorey, K. (2014). Iconicity as structure mapping. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130301. https://doi.org/10.1098/rstb.2013.0301

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*(1), 47–65. https://doi.org/10.1016/S0095-4470(19)30537-6

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences*, *103*(32), 12203–12208. https://doi.org/10.1073/pnas.0602173103

Flege, J. E. (1995a). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–276). York, England: York Press.

Flege, J. E. (1995b). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics*, *16*, 425–442.

Fort, M., Lammertink, I., Peperkamp, S., Guevara-Rukoz, A., Fikkert, P., & Tsuji, S. (2018). Symbouki: A meta-analysis on the emergence of sound symbolism in early language acquisition. *Developmental Science*, e12659. https://doi.org/10.1111/desc.12659

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 1–13. https://doi.org/10.1098/rstb.2013.0298

Iwasaki, N., & Yoshioka, K. (2019). Iconicity in L2 Japanese speakers' multi-modal language use mimetics and co-speech gesture in relation to L1 and Japanese proficiency. In K. Akita & P. Pardeshi (Eds.), *Ideophones, mimetics and expressives* (pp. 265–302). Amsterdam, The Netherlands: John Benjamins.

Ković, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, *114*(1), 19–28. https://doi.org/10.1016/j.cognition.2009.08.016

Laing, C. (2019). A role for onomatopoeia in early language: Evidence from phonological development. *Language and Cognition*, *11*(2), 173–187. https://doi.org/10.1017/langcog.2018.23

Lockwood, G., Dingemanse, M., & Hagoort, P. (2016). Sound-symbolism boosts novel word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(8), 1274–1281. https://doi.org/10.1037/xlm0000235

Lockwood, G., Hagoort, P., & Dingemanse, M. (2016). How iconicity helps people learn new words: Neural correlates and individual differences in sound-symbolic bootstrapping. *Collabra*, *2*(1), 1–15. https://doi.org/10.1525/collabra.42

Maduka, O. N. (1983). Igbo ideophones and the lexicon. *Journal of the Linguistic Association of Nigeria*, *2*, 23–29.

Marsh, J. E., Hughes, R. W., Sörqvist, P., Beaman, C. P., & Jones, D. M. (2015). Erroneous and veridical recall are not two sides of the same coin: Evidence from semantic distraction in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1728.

Matthews, R. J., & Demopoulos, W. (Eds.). (2012). *Learnability and linguistic theory* (Vol. 9). Berlin, Germany: Springer Science & Business Media.

McLean, B. (2020). Revising an implicational hierarchy for the meanings of ideophones, with special reference to Japonic. *Linguistic Typology*, *25*(3), 507–549. https://doi.org/10.1515/lingty-2020-2063

Mok, W. E. (2001). *Chinese sound symbolism: A phonological perspective* [Unpublished doctoral dissertation]. University of Hawai'i.

Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in language learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1152–1164. https://doi.org/10.1037/a0027747

Moreton, E., & Pater, J. (2012a). Structure and substance in artificial-phonology learning, part I: Structure. *Language and Linguistics Compass*, *6*(11), 686–701. https://doi.org/10.1002/lnc3.363

Moreton, E., & Pater, J. (2012b). Structure and substance in artificial-phonology learning, part II: Substance. *Language and Linguistics Compass*, *6*(11), 702–718. https://doi.org/10.1002/lnc3.366

Nielsen, A. K., & Dingemanse, M. (2021). Iconicity in word learning and beyond: A critical review. *Language and Speech*, *64*(1), 1–21. https://doi.org/10.1177/0023830920914339

Nuckolls, J. B. (2019). The sensori-semantic clustering of ideophonic meaning in Pastaza Quichua. In K. Akita & P. Pardeshi (Eds.), *Ideophones, mimetics and expressives* (pp. 167–198). Amsterdam, The Netherlands: John Benjamins.

Nuckolls, J. B. (2020). "How do you even know what ideophones mean?": Gestures' contributions to ideophone semantics in Quichua. *Gesture*, *19*(2–3), 161–195. https://doi.org/10.1075/gest.20005.nuc

Nuckolls, J. B., Stanley, J. A., Nielsen, E., & Hopper, R. (2016). The systematic stretching and contracting of ideophonic phonology in Pastaza Quichua. *International Journal of American Linguistics*, *82*(1), 95–116.

Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, *112*(1), 181–186. https://doi.org/10.1016/j.cognition.2009.04.001

Occhino, C., Anible, B., Wilkinson, E., & Morford, J. P. (2017). Iconicity is in the eye of the beholder: How language experience affects perceived iconicity. *Gesture*, *16*(1), 100–126. https://doi.org/10.1075/gest.16.1.04occ

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Perry, L. K., Perlman, M., & Lupyan, G. (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

R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Thompson, A. L. (2018). Are tones in the expressive lexicon iconic? Evidence from three Chinese languages. *PLoS ONE*, *13*(12), e0204270. https://doi.org/10.1371/journal.pone.0204270

Thompson, A. L., & Do, Y. (2019). Defining iconicity: An articulation-based methodology for explaining the phonological structure of ideophones. *Glossa: A Journal of General Linguistics*, *4*(1), 72. https://doi.org/10.5334/gjgl.872

Thompson, A. L., Van Hoey, T., & Do, Y. (2021). Articulatory features of phonemes pattern to iconic meanings: Evidence from cross-linguistic ideophones. *Cognitive Linguistics*, *32*(4), 563–608. https://doi.org/10.1515/cog-2020-0055

Thompson, A. L., Chan, M.P.Y, Yeung, P.H, & Do, Y. (2022). Structural markedness and depiction: The case of lower sequential predictability in Cantonese ideophones. *The Mental Lexicon*, *17*(2), 300–324.

Uchechukwu, C. (2007). The Phenomenon of ideophonic pairing in Igbo. *Journal of West African Languages*, *34*(1), 26–38.

Uchechukwu, C. (2017). Ideophonic pairing as a pan-Igbo phenomenon. *Igbo Language Studies*, *2*, 77–84.

Van Hoey, T. & Thompson, A. L. (2020). The Chinese ideophone database (CHIDEOD). *Cahiers de linguistique Asie orientale*, *49*(2), 136–167. https://doi.org/10.1163/19606028-bja10006.

Van Hoey, T. (in press). A semantic map for ideophones. In T. F. Li (Ed.), *Handbook of cognitive semantics*. Brill. Preprint doi: 10.31219/osf.io/muhpd

Yoshida, H. (2004). *Iconicity in language learning: The role of mimetics in word learning tasks* [Unpublished doctoral thesis]. Indiana University, Bloomington, IN.

Yoshida, H. (2012). A cross-linguistic study of sound symbolism in children's verb learning. *Journal of Cognition and Development*, *13*(2), 232–265. https://doi.org/10.1080/15248372.2011.573515