

Arthur Lewis Thompson, Thomas Van Hoey,
Aaron Wing Cheung Chik and Youngah Do*

Iconic hand gestures from ideophones exhibit stability and emergent phonological properties: an iterated learning study

<https://doi.org/10.1515/cog-2024-0033>

Received April 2, 2024; accepted April 8, 2025; published online April 30, 2025

Abstract: Ideophones are marked words which depict sensory imagery and are usually considered iconic by native speakers (i.e., ideophones sound like what they mean). Owing to shared cross-linguistic characteristics of expressive prosody, reduplication, and unusual phonological structure, ideophones have been likened to meaning performed. Iconic hand gestures frequently occur alongside ideophones choreographed to the timing of syllables. Given the visual modality's richness in iconic affordances, these gestures have been supposed to help interlocutors infer semantic nuances and contextualize utterances, especially when an ideophone is polysemous, and may even inform speakers' mental representations of spoken language as imitative. Such gestures should therefore be learnable and replicable like any unit of language. This is what we indeed find. Using a linear iterated learning paradigm, we investigated the stability of iconic gestures from Japanese and Korean ideophones transmitted across five generations. Despite noise in the visual signal, participants' hand gestures converged, speaking to the emergence of phonological targets. Handshape configurations over time exhibited finger coordination reminiscent of unmarked handshapes observed in phonological inventories of signed languages. Well-replicated gestures were correlated with well-guessed ideophones from a spoken language study, further highlighting the complementary nature of the visual and spoken modalities in formulating mental representations.

Keywords: iconicity; gestures; ideophones; iterated learning; handshape

Arthur Lewis Thompson and Thomas Van Hoey are joint first authors.

***Corresponding author: Youngah Do,** Department of Linguistics, The University of Hong Kong, Hong Kong, China, E-mail: youngah@hku.hk. <https://orcid.org/0000-0003-1926-2230>

Arthur Lewis Thompson and Aaron Wing Cheung Chik, Department of Linguistics, The University of Hong Kong, Hong Kong, China, E-mail: arthurlewisthompson@gmail.com (A.L. Thompson), aaeronc@connect.hku.hk (A.W.C. Chik). <https://orcid.org/0000-0002-1508-1232> (A.L. Thompson). <https://orcid.org/0009-0002-9740-5380> (A.W.C. Chik)

Thomas Van Hoey, FWO & QLV, KU Leuven, Leuven, Belgium, E-mail: thomas_van_hoey@outlook.com. <https://orcid.org/0000-0002-5226-9752>

1 Introduction

Increasingly, and perhaps thanks to foundational insights of gesture and sign language researchers (Hodge and Ferrara 2022; Kendon 2004; Perniss et al. 2010), the notion that the visual modality is important for comprehension (and creation of mental representations) of spoken language can no longer be ignored (Dargue and Sweller 2020; Drijvers and Özyürek 2017; Kandana Arachchige et al. 2021; Kelly et al. 2010; Nielsen and Dingemanse 2021; Thompson and Do 2019b; Yap et al. 2011).

Ideophones are a lexical testament to this, and the role multimodality plays in their lexical comprehension. It is well-known that iconic words such as ideophones often co-occur with iconic co-speech gestures. Indeed, they are “the closest linguistic substitute[s] for [...] non-verbal, physical act[s]” (Kunene 2001: 183), often constituting small performances depicting scenes, “the next best thing to having been there” yourself (Dingemanse 2011: 299). No wonder, then, that the interplay of iconicity in spoken form and gesture, as in the case of ideophones, should be of interest to the field, an issue that we will tease apart in this study using an iterated learning paradigm.

Ideophones are widely known as “marked words that depict sensory imagery, belonging to an open lexical class” (Dingemanse 2019) – native speakers can improvise new ideophones on the spot (Assaneo et al. 2011; Nasu 2015; Nuckolls et al. 2016; Taitz et al. 2018). They are limited to a range of semantic categories such as SOUNDS, e.g., Japanese *wan-wan* “sound of a dog barking,” VISUALS, e.g., Korean *panchak-panchak* “shining; glittering,” and INNER FEELINGS, e.g., Cantonese *fi⁴li¹fe⁴le⁴* 飛哩啡呢 “in a state of disarray,” among others (Van Hoey 2023). Ideophones exhibit sound-meaning correspondences across unrelated languages (Voeltz & Kilian-Hatz 2001; Thompson et al. 2021), e.g., fricative sounds often correspond with meanings pertaining to MOTION. Non-native speakers have been shown to guess what ideophones mean above chance level for a variety of languages (Dingemanse et al. 2016; Lockwood et al. 2016; McLean, Dunn & Dingemanse 2023; Van Hoey et al. 2023), speaking to their inherent properties of iconicity. Additionally, ideophones are consistently rated as highly iconic by native speakers (Hinojosa et al. 2021; McLean, Dunn & Dingemanse 2023; Perlman et al. 2018; Thompson et al. 2020; Winter et al. 2017, 2023). Cross-linguistically, they thus belong to the iconic lexicon of languages.

As for their relationship with the visual modality, ideophones frequently occur with iconic hand gestures in natural speech (Cano 2020; Dingemanse 2013; Dingemanse and Akita 2017; Hatton 2016; Kunene 1965; Kita 1997; Mihas 2013; Nuckolls 2020; Thompson and Do 2019b), forming units known as ideophone-gesture composites. So far, research on ideophone-gesture composites has looked

at idiosyncratic and speaker-centric contexts, focusing on the communicative intent of the speaker and their success in achieving this intent. For example, studies have shown that iconic hand gestures delineate semantic nuances of polysemous ideophones (Dingemanse 2013; Nuckolls 2020), gestures are timed to prosodic phenomena associated with ideophones (Iwasaki and Yoshioka 2019; Kita 1997), and that gestures improve non-native participants' accuracy in guessing what ideophones mean (authors, under review). The same is true when iconic signs have been examined from the perspective of sign language learning (Karadöller et al. 2024).

Given that co-speech gestures can help clarify the meaning of improvised or polysemous ideophones during production or perception, it stands to reason to investigate how well such gestures are able to stand in for an ideophone's referent when divorced from the rich conversational context in which the visual information would otherwise naturally occur, and to investigate to what degree this visual information is learnable, i.e., successfully taught from one person to the next. While gestures have shown to be transmissible over time (Motamedi et al. 2019), such gestures fit into an orderly semantic classification (e.g., shapes, jobs, buildings), as opposed to the sensorial meanings of ideophones decidedly at odds with intuitive classification (de Schryver 2009: 38). This study addresses such questions within an experimental setting, focusing on an iterated learning paradigm (explained in Section 1.2). Learning involves acquiring these gestures through observation or instruction, resulting in behavioral changes in replication. Through an iterated learning paradigm, we simulate evolution of such replication by repeatedly exposing participants to these gestures and observing the transmission of replicated gestures across generations.

We first review other paradigms that have been used to address the relation between iconicity and the spoken and visual modalities (Section 1.1), before discussing iterated paradigms in general (Section 1.2), and the current study (Section 1.3).

1.1 Guessing paradigms

Guessing paradigms so far have been a preferred method for assessing the iconicity of ideophones, especially two-alternate forced-choice (2AFC) tasks. For example, Lockwood et al. (2016) show that Dutch speakers chose the correct Dutch translations for 50 Japanese ideophones (presented auditorily alongside transliterations) above chance level and do so even after they have been trained in a preceding task to remember foil ("opposite") Dutch translations. Word frequency (among other factors) pertaining to the correct and foil Dutch translations was prioritized in the stimuli design, as opposed to accounting for semantic categories

of Japanese ideophones, presumably so that Lockwood et al. (2016) could test as broad a swathe of ideophones as possible. Sensory category (COLOR-VISUAL, MOTION, SHAPE, SOUND, TEXTURE) was controlled for in a later 2AFC study conducted by Dingemanse et al. (2016) using 203 ideophones from Japanese, Korean, Semai (Mon-Khmer), Siwu (Niger-Congo), and Ewe (Niger-Congo). They show that, though all sensory categories were guessed correctly above chance level, SOUND did best. Though 2AFC design is not without its problems (Lockwood et al. 2016; McLean, Dunn & Dingemanse 2023), Dingemanse et al. (2016) made sure that foil translations were randomly selected from a set belonging to the same semantic category as true translations (e.g., true translation “smooth” TEXTURE vs. foil “slippery” TEXTURE), making the task quite difficult. Some ideophones were resynthesized so that either (a) the segment level was altered but the prosodic level preserved, or (b) the prosodic level was altered but the segment level preserved. Participants who listened to resynthesized ideophones still chose correct translations at an above chance rate, however at a rate lower than participants who listened to unaltered originals. Prosody and segments (consonants, vowels) therefore each play an equal role in the perception of iconicity and the scaffolding of perceptuomotor mappings that help make ideophones meaningful. Moreover, SOUND ideophones are presumably the most semantically transparent due to the common modality of their referent (i.e., real-world sounds) and speech.

In a follow-up guessing and memory study with L1 Cantonese participants trained on ideophones from Japanese, Korean, and Igbo (Niger-Congo), Van Hoey et al. (2023) found two additional points relevant here: (1) SOUND and TEXTURE ideophones were guessed best, and (2) participants remembered foil translations of ideophones just as well as true translations. Even though certain segments reliably co-occur with certain ideophone meanings cross-linguistically (Diffloth 1979; Hamano 1998; Thompson and Do 2019a; Thompson et al. 2021), perhaps the phonologically marked structure of ideophones simply lends itself well to depiction in general, boosted by other factors like common ground and iconic gestures.

Overall, the field tends to agree that ideophones are guessable for native and non-native speakers, and that gestures play an important role in this process; though not always, as Hatton (2016) has observed a conspicuous lack of spontaneous gestures in Pastaza Quechua SOUND ideophones. Furthermore, the range of gestures appears correlated to the semantic range of an ideophonic item Nuckolls’ (2020), and iconic gestures themselves often provide sufficient depictive qualities to be successfully guessable by non-native speakers (authors, under review). What is missing, is a detailed step-by-step observation of how ideophone-gesture composites are transmitted, an issue for which the iterated learning paradigm is well-suited.

1.2 Iterated learning paradigms

The present study is the first to look at gestures derived from ideophones through iterated learning. Iterated learning paradigms emulate how language is learned through cultural transmission in a compressed experimental setting. In most paradigms, at the most basic level, iterated learning involves one participant teaching another participant either words, signs, gestures, or drawings. Thus far, such studies have adopted an exploratory approach, driven by an interest in seeing how language evolves in real-time (Kirby et al. 2008). Most iterated learning experiments are thus open-ended (Scott-Phillips and Kirby 2010), and concerned with stimuli design (number of meanings learned, kinds of meanings, medium of communication) and the number of iterations before the experiment concludes. The iterative nature of the approach taken in this study allows for the examination of how gestures change and adapt over successive generations of learners, shedding light on the underlying mechanisms of cultural evolution and the transmission of symbolic systems.

Naturally, the design of the stimuli plays a crucial role in reflecting and facilitating the evolutionary objectives of the study. The number of meanings is small enough to learn in one training session but large enough so that some meanings are lost or end up converging from one generation to the next. The kinds of meanings are organized or within a shared semantic space, e.g., shapes, movements, and colors (Kirby et al. 2008), or concrete nouns and related verbs (Motamedi et al. 2019). The medium of communication is either spoken, gestured, or drawn, all of which allow for innovation on behalf of the participant. How the iteration is carried out leads to different patterns in the results (Scott-Phillips and Kirby 2010: 415). If iteration is linear, with no feedback between the teacher and the learner, then iconicity is usually preserved over time. If iteration is interactive, such as in groups or dyads, then forms lose their iconic properties in favor of communicative efficiency. These results are important to keep in mind for our study since patterns of acquisition (learning biases, cognitive biases) inform, and are perhaps inseparable from, claims made about linguistic evolution.

1.3 The present study

Our objectives are exploratory but not as open-ended as those of studies which investigate how language evolves. We are using a linear iterated learning paradigm to examine the stability and robustness of iconic forms (gestures derived from ideophones) in an implicitly acquisitional setting (see Section 2.3). Rather than designing a set of stimuli seated within a neatly organized semantic space (as in

Kirby et al. 2008; Motamedi et al. 2019), we only test items which are based on extant linguistic forms (Japanese and Korean ideophones). This also allows for comparison with prior memory and guessing studies, like Van Hoey et al. (2023) who made use of related stimuli (i.e., audio of Japanese and Korean ideophones). Only items accurately transmitted from one participant to the next five times in a row are analyzed. Accuracy is determined based on whether a participant was able to replicate a given handshape (i.e., how the hands are configured in a gesture), a concept borrowed from sign language phonology (Brentari 1998; Sandler 1989). That is, we chose handshape as our focus in determining whether or not a gesture was passed down successfully and, in doing so, set the bar relatively high (see Figure 2). According to sign language literature, handshape is the last phonological parameter L1 signing children fully acquire (Bonvillian and Siedlecki 2000; Conlin et al. 2009; Karnopp 2002; Marentette and Mayberry 2000; Siedlecki and Bonvillian 1993), and correct replication of handshape is notably difficult for hearing L2 learners of sign languages (Ortega 2017). That said, using four handshapes from American Sign Language, Occhino et al. (2020) found that less-proficient signers are better at detecting the handshape of signs which are more iconic. In terms of language evolution and the development of a sign language over time, handshape is found to be the phonological parameter most likely to undergo change (Moita et al. 2023), presumably due to the number of articulators involved (i.e., five digits) and, subsequently, the degrees of variation applicable to each. In light of the above, successful transmission of a gesture from one participant to the next is no small feat, especially given that participants were required to mentally retain multiple gestures before being asked to execute them for transmission purposes.

We hypothesize that successful transmission occurs only if visual iconicity is not obfuscated. To test this, we have created foils where iconic gestures are paired with translations which deliberately obfuscate structure mapping (e.g., gesture: *SCISSORS CUTTING* + translation: “dry, brittle”). Much like studies aimed at language evolution, we also assume that robustness coupled with systematicity implies linguistic viability (i.e., visual schemas become amenable to mental representations). The emergence of systematicity is expected to occur primarily at a phonetic and/or phonological level and may be less conspicuous than the emergence of morphological innovations, which could compromise the “perfect” replication of the original ideophonic forms. The stability of the iconic gestures across five generations, and any subtle yet systematic changes which emerge as part of a key component of their execution (i.e., handshape), speak to their semantic utility and structural conduciveness in building mental representations for the multimodal, depictive meanings behind ideophones.

2 Methods and materials

The materials and scripts containing the data analysis are available on the Open Science Framework, at <https://osf.io/dtr4p/>.

2.1 Participants

Sixty participants¹ with no sign language experience were recruited for our linear iterated learning experiment. Each participant was paid a cash reward upon the completion of the experiment which took on average around 15 min to complete.

2.2 Stimuli

We extracted the Korean (n = 46) and Japanese (n = 43) ideophones with their (mono) sensory² categories used by Dingemanse et al. (2016) and replicated by Van Hoey et al. (2023). These were presented to the two seed participants, i.e., one L1 Japanese female (33) and one L1 Korean female (31) who provided the first instances of the gestures learned by subsequent participants. Seed participants were instructed to produce an iconic gesture while saying each ideophone to yield ideophone-gesture composites. Items were discounted when no gesture was produced. For example, seed participants had difficulty producing gestures for SOUND ideophones (see Hatton 2016). Items were also discounted if the gesture produced was incongruent with the meaning described by Dingemanse et al. (2016). For example, Korean *pong dang pong dang* 풍당 풍당 is described as “(small object) falling repeatedly in water with small splashes.” Our seed participant, however, produced a THROWING gesture. While “splash” could be deduced as resultative of “throwing,” the incongruency of the visually depicted action (i.e., throwing) with the target item’s meaning (i.e., splash) led to exclusion. Any gestures exhibiting a high degree of visual similarity were eliminated so that each stimulus differed in handshape and at least one other factor (e.g., location, movement), leading to 16 items for Japanese and 25 items for Korean were retained. To balance the conditions and groups, 16 Korean items were chosen to map as closely as possible to the 16 Japanese items (a trade-off resulting in a slight imbalance in

1 38 % were 18–19 years old; 59 % were 20–29 years old; 3 % were 30–39 years old. 65 % were female, 33 % male, and 2 % non-binary.

2 Some ideophones are clearly multisensorial, e.g., *choki choki* ‘the SOUND of cutting scissors’ also of course involves the motion necessary to produce said sound. We have adopted the categories from previous literature to enable comparisons but wish to point out that this is an abstraction.

sensory categories, see OSF material). The final stimuli inventory was 16 items per language, i.e. 32 in the whole experiment.

To create foil items, a “coerced” translation was assigned to each item and balanced across five sensory categories: SOUND, TEXTURE, SHAPE, COLOR-VISUAL, and MOTION. A foil item here is thus a translation borrowed from one sensory category and assigned to a form originally belonging to another sensory category, e.g., a SOUND item is given a TEXTURE translation. All foil translations were taken from Van Hoey et al. (2023). We designed two mixed experimental conditions per language, i.e., Japanese-A, Japanese-B, Korean-A, Korean-B. Each condition contains 16 items. Condition A is comprised of eight true items and eight foil items. Condition B is the reverse of Condition A, i.e., the true items in Condition A are foils in B, while the foil items in Condition A are true in B. This mixed design ensures a balance of “easier” items and “harder” items (easier meaning more iconically motivated) per condition. Examples of stimuli and their true and foil translations are shown in Table 1.

2.3 Experiment design

The experiment was built in PsychoPy 3 (Peirce et al. 2019) and run on a local machine. Seed gestures (“generation 0”) were mapped onto true and foil meanings, resulting in two sets. These were mixed in two conditions (A and B), with half each condition having true mappings and half foil mappings. We did this for both languages, obtaining thus four groups: Jap-A, Jap-B, Kor-A, Kor-B. Each group had three transmission chains (P, Q, R). A transmission chain is a linear sequence of five participants. Each participant in the transmission chain is a generation. Participants learn only the gestures executed by the immediately preceding generation. The dimensions of the experiments required 60 participants (4 condition groups \times 3 chains \times 5 generations), see Figure 1.

2.4 Procedure

Participants were randomly assigned to a data point (language, condition, generation). They were instructed to stand in front of a monitor and camera that captured them from the waist up. The live video-feedback in the top right corner of the screen allowed them to position themselves correctly. The participant and the monitor were recorded with consent simultaneously using the web-conferencing software Zoom (Yuan 2021). Participants were verbally instructed to focus on the monitor and observe the gestures (shown twice) and their meaning, and then mimic the gestures, which they would later reproduce for recording. These instructions were repeated

Table 1: Examples of Korean and Japanese ideophone stimuli. Items were auditorily presented in the experiment but are given in transcription and native orthography for convenience here. Participants were presented with meanings in written Cantonese (Chinese characters). Translations are given between apostrophes. Sensory categories are shown in small caps.

Language	Item	True sense	Foil sense
Korean	heumeulheumeul	稀爛	苗條
	흐물흐물	“overripe (fruit), flabby”	“slim”
		TEXTURE	SHAPE
	panjakbanjak	閃耀	噴聲
	반짝반짝	“glittering”	“click one’s tongue”
		COLOR-VISUAL	SOUND
Japanese	chingching	捆綁“tie up in circular motion”	起皺
	칭칭	MOTION	“wrinkled; un-ironed”
			SHAPE
	tengteng	脹鼓	水濺聲
	탱탱	“blown up like a balloon”	“splash”
		SHAPE	SOUND
	bon	爆炸聲	平滑
	ボン	“exploding sound”	“slippery, smooth surface”
		SOUND	TEXTURE
	chokichoki	剪刀聲	乾燥
	チヨキチヨキ	“scissors snipping”	“dry, brittle”
		SOUND	TEXTURE
	kirakira	燦爛	粗糙
	キラキラ	“sparkling”	“coarse, rough”
		COLOR-VISUAL	TEXTURE
	perapera	單薄	油膩
	ペラペラ	“thin”	“oily”
		SHAPE	TEXTURE
	tonton	敲門聲	堅硬
	トントン	“knocking on door”	“very hard”
		SOUND	TEXTURE

on the monitor after the presentation of each stimulus. Participants also watched three practice items.

In the learning phase, participants initiated a trial by pressing the space bar. A fixation cross (1s) was shown, followed by two repetitions of a video containing an ideophone-gesture composite. Next, the translation of the ideophone-gesture composite was presented in written Cantonese (see Table 1 above). After pressing space to proceed, they were given 10 s (shown with a countdown timer) to commit the ideophone-gesture composite and its translation to memory. This phase of the trial

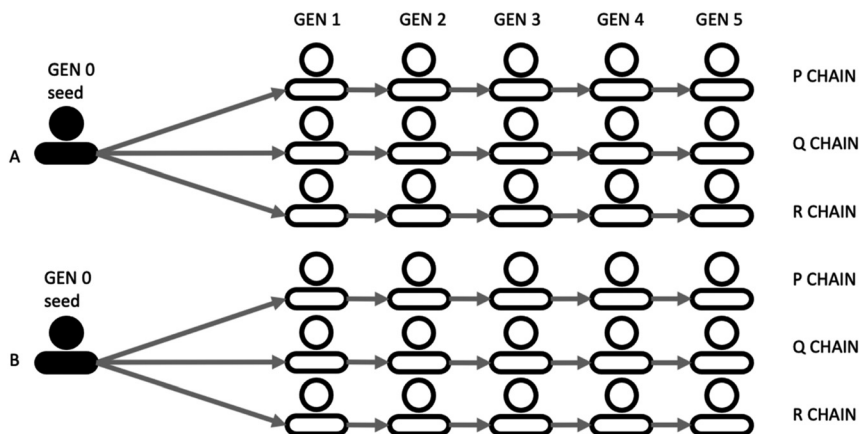


Figure 1: Experiment design for one language (either Japanese or Korean).

was also self-paced. Participants did not need more than 10 s before advancing to the next trial.

In the testing phase, participants were presented with the translations they learned, in randomized order, and asked to reproduce gestures they learned with it. They were explicitly instructed to produce gestures even if they were unable to recall the gesture learned. Before translations were presented, a fixation cross was shown (1s), followed by a “get ready” message then a three-second countdown. Research personnel intervened only when the participant intended to proceed without producing a gesture. After participants finished with their iteration, their testing phase gestures were used as input for the next participant (the following generation), akin to the game of telephone. This was done for five generations.

3 Coding and data analysis

In total, 992 data points (60 participants \times 16 items + 32 seed items) were collected and 948 entered analysis, representing those who completed the experiment. The data points can be analyzed in three ways, depending on how they are coded: binary coding, handshape coding, and the correlation between binary and handshape coding.

BINARY CODING. To see which gestures were successfully passed from one participant to the next, all videos were coded in a binary fashion as exhibiting either “resemblance” or “no resemblance” in handshape to the immediately preceding gesture in its transmission chain. Categorical differences in handshape, resulting in

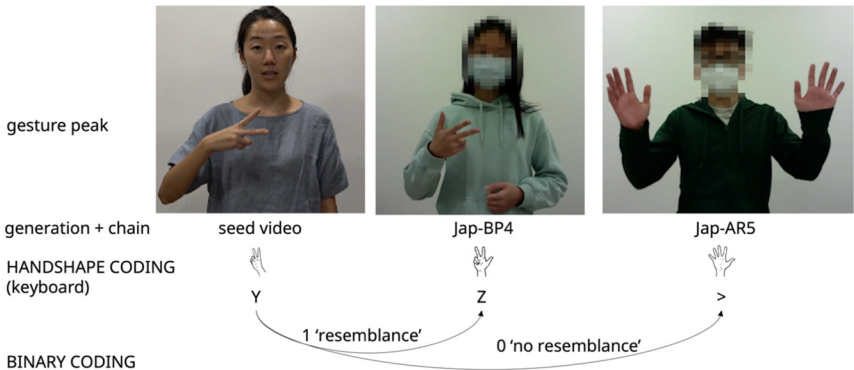


Figure 2: Didactic illustration of *chokichoki* “sound of cutting scissors.” Participants saw unmasked input video clips. Here, the transmission from [Y] to [Z] the binary coding value is 1 (successful transmission); from [Y] to [>] the binary coding value is 0 (failed transmission). In reality, BP4 and AR5 would have seen the gesture produced by the preceding generation (i.e., Jap-BP3, Jap-AR4 respectively).

seemingly no visual similarity from one participant to the next, were coded as “no resemblance” and are thus considered failed transmissions. Successfully transmitted gestures, therefore, must exhibit *at least* some visual similarity in handshape to their immediate predecessors, allowing for a degree of innovation or noise. See an example in Figure 2. This binary coding was carried out by one coder and subsequently verified by two others. No disagreement was found between coders.

HANDSHAPE CODING. Videos were later coded according to which handshape occurred at the peak of each gesture (regardless of its binary coding). This allowed us to see how handshapes evolved over time. An inventory of 47 handshape images corresponding to QWERTY keyboard symbols was compiled and agreed upon by three coders beforehand. Due to the sheer amount of coding required, the videos were divided into three batches with one coder assigned a batch. After a batch was coded that batch was then passed onto another coder for verification. No disagreement was found between the coders. Particulars about the handshape coding are available in the OSF repository.

CORRELATION BETWEEN BINARY AND HANDSHAPE CODING. The binary (“resemblance” or “no resemblance”) coding was ultimately correlated with the handshape coding. Observations about consistencies and discrepancies across the two coding systems allow us to explore how (a) successful transmission, (b) evolution across transmission, and (c) faithfulness of form to meaning (i.e., iconicity) all come together in the emergence of linguistic structure. The correlation measure is further discussed in Section 4.1.

3.1 Successful transmissions and perfect chains

A perfect chain consists of five successful binary coding values, i.e., [11111]. This means that there is no error to disrupt transmission across all five generations. We find 34 perfect chains, and 92 chains with errors in the first generation (immediately after the seed video), see Figure 3.

Transmission can be qualified in two ways: (1) IMMEDIATE SUCCESS, consisting of a simple tally of two successful transmissions (i.e., sequence of two gestures both coded as “bears resemblance”) per chain, and (2) LONGITUDINAL SUCCESS, counting the number of successful transmissions across five generations (Figure 3, e.g., there are 34 chains where none of the generations exhibited errors). Transmission chains with more longitudinal success were rewarded, while chains with less longitudinal success were penalized. Taking three successful transmissions in a row as a decent longitudinal transmission value, a logarithmic transformation with base 3 appeared the correct way of integrating such a weight. For example, the Korean ideophone *chingching* “tie up with rope, circular motion” resulted in the following number of successful transmissions per chain: three chains containing five successful transmissions (AR, BR, BP), one chain containing four successful transmissions (BQ), two chains containing zero successful transmissions (AP, AQ). Table 2 shows how the immediate success (15) and longitudinal success (3.66) were calculated for *chingching*.

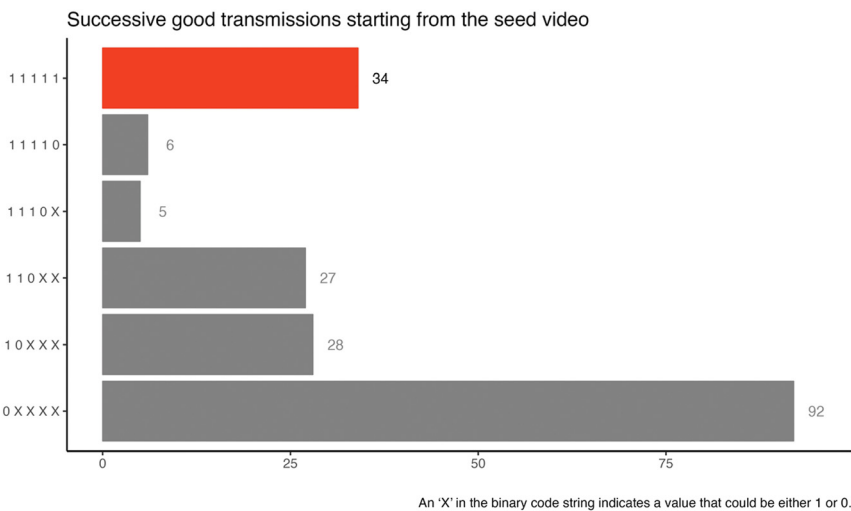


Figure 3: There are 34 instances of “perfect chains” (i.e., 5 generations with zero errors). 0 = error, 1 = accurate, X = voided generation.

Table 2: Calculation of immediate and longitudinal success for gestures generated from Korean ideophone *chingching*. The binary coding is presented with a containing generation index, e.g., [1^10^2] “generation 1 bears resemblance to the seed video, while generation 2 bears no resemblance to generation 1”. X indicates irrelevant generations after a failed transmission.

Success. trans.	Binary coding	N	Bigram	Immediate score (N × bigram)	Longitudinal weight $\log_3(\text{chains})$	Longitudinal score (N × weight)
5	$1^11^21^31^41^5$	3	4	12	1.46	4.39
4	$1^11^21^31^40^5$	1	3	3	1.26	1.26
3	$1^11^21^30^4X^5$	0	2	0	1	0
2	$1^11^20^3X^4X^5$	0	1	0	0.63	0
1	$1^10^2X^3X^4X^5$	0	0	0	0	0
0	$0^1X^2X^3X^4X^5$	2	0	0	−1	−2
Sum				15		3.66

As the relationship between immediate and longitudinal success shows (Figure 4), some gestures were transmitted more faithfully than others. Gestures derived from ideophones *chingching* (Korean), *heumeulheumeul* (Korean), *chokichoki* (Japanese), and *tonton* (Japanese) exhibited the most success (Figure 5). It is for these items that we see the first interaction between salience (see below), and the stability of iconicity and anatomical systematicity (Section 4.3).

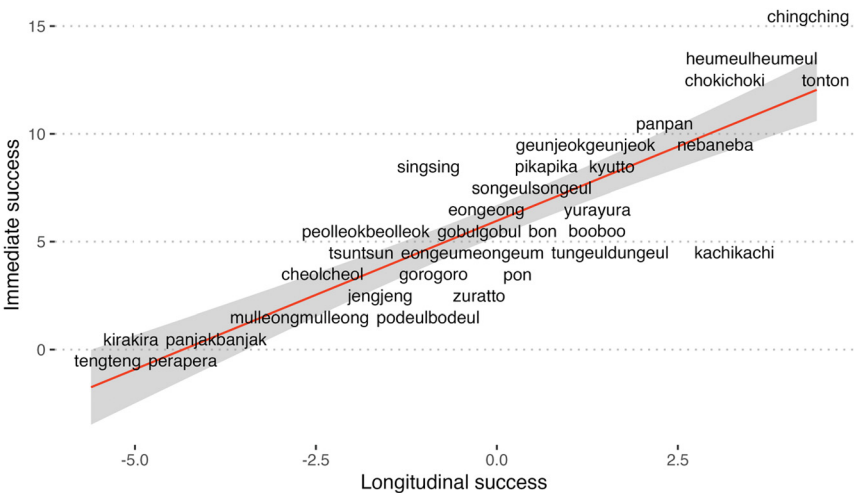


Figure 4: Item-level relation between immediate success and longitudinal success. Note that longitudinal values were centered, i.e., the x-axis shows distance from the mean ($\bar{x} = 5.97$).



Figure 5: Stills of seed videos which yielded the most successful transmissions.

The successful items of Figure 5 are all “cognitively salient”, but not in the same manner: *Chingching* and *heumeulheumeul* are extra-systemic (Rácz 2013) – circular bi-manual movements and pinching one’s triceps, respectively, are relatively unusual gestures and therefore stand out (see Cates et al. 2013 for the importance of location) affording them a higher chance of being transmitted successfully. Compare these with the ideophone-gesture composites of *chokichoki* and *tonton*, both of which are well-entrenched and highly familiar (Schmid 2017). At the other end of the spectrum, we find *panjakbanjak* and *kirakira* (Figure 6). One reason for their unsuccessful transmission may be their meanings, depictions of multiplex light composition (Van Hoey 2020), which do not lend themselves well to salient gestures in a decontextualized set-up. This lack of contextual grounding may also be responsible for the unfaithful transmission of *tengteng* and *perapera*. In sum, the transmission at the general item-level was moderately successful for most ideophone-gesture composites, with some items underperforming and some items being transmitted extremely well.



Figure 6: Stills produces for the least successful seed video gestures.

3.2 Effects of language, condition, and semantic category

We have seen that not all gestures were transmitted with equal success. Potentially mitigating factors are language (Japanese or Korean), condition (true or foil) and sensory category of the true mapping (SOUND, MOVEMENT, COLOR-VISUAL, SHAPE, TEXTURE). An exploratory visualization of the first two factors is presented in Figure 7. Japanese gestures behave as expected: the scores for foil items tend to be much lower. For Korean gestures, however, the true items were often much less likely to be transmitted successfully than foils. We ran a linear regression model with longitudinal score as the dependent variable and the three factors of language, condition, and sensory category as the predictors. An ANOVA indicated that an interaction effect between language and condition had more explanatory power. We find a significant boosting effect for Japanese true items ($\beta = 3.55$) but an inhibitory correction for Korean true items ($\beta = -2.65$, see Table 3). In other words, gestures presented with their true (iconic) meanings are indeed more stable, as predicted. However, overall, Japanese gestures were transmitted better than Korean gestures.

There were no significant effects for sensory category, a finding that is somewhat at odds with the semantic transparency of SOUND ideophones reported in previous guessing studies (Dingemanse et al. 2016; Van Hoey et al. 2023).

However, our results are similar to previous 2AFC tasks: Figure 8 shows how our longitudinal scores compare to the reassessment accuracies in Van Hoey et al. (2023),

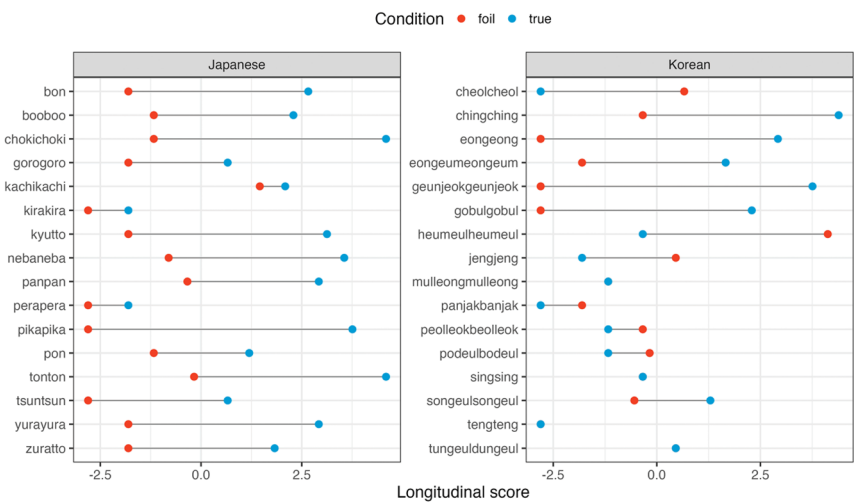


Figure 7: Longitudinal scores for each item per condition. Note that longitudinal scores were centered. Having only one dot indicates the results for both conditions are the same.

Table 3: Linear regression model with longitudinal score as the dependent variable and language, condition, and true sense as the predictors. Note that longitudinal scores were centered. The reference level is the foil condition for Japanese color-visual items.

Term	β	95 % Confidence interval	$t(56)$	p value
Intercept	−2.29	[−3.72, −0.86]	−3.21	0.002**
Condition _{true}	3.55	[2.27, 4.84]	5.55	<0.001***
Language _{korean}	0.71	[−0.6, 2.01]	1.08	0.285
Truesense _{motion}	1.08	[−0.42, 2.58]	1.45	0.154
Truesense _{shape}	0.11	[−1.4, 1.62]	0.15	0.883
Truesense _{sound}	1.25	[−0.28, 2.78]	1.63	0.108
Truesense _{texture}	1.54	[−0.01, 3.09]	1.98	0.052
Condition _{true} : language _{korean}	−2.65	[−4.47, −0.84]	−2.93	0.005**

***Indicates $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

divided by condition (true, foil) and language – the two factors that were significant in the current study. Foil items tend to have lower longitudinal scores and lower guessing accuracy, while true mappings have higher longitudinal scores and are guessed better. There may also be a language effect: Japanese items score higher than Korean items. A linear regression model (Table 4) with longitudinal score, condition, and language as the predictors and guessing accuracy as the dependent variable

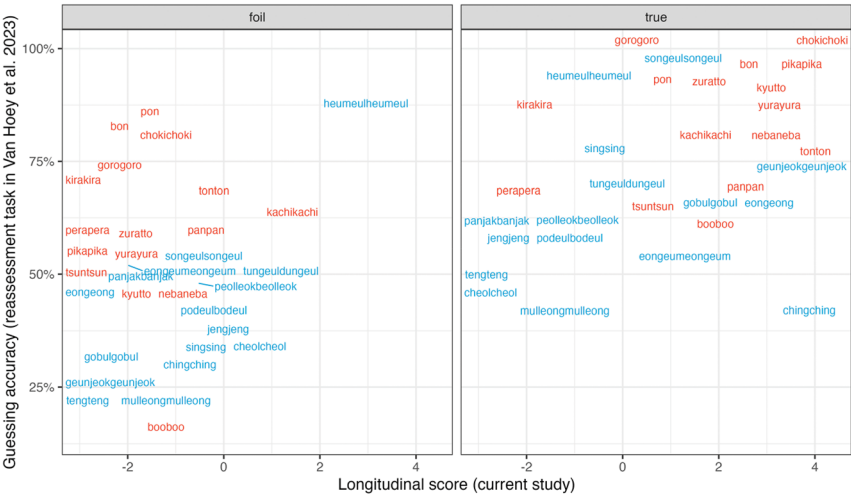


Figure 8: Comparison of longitudinal scores of gesture transmission in this study with the guessing accuracy of ideophones (audio only) in Van Hoey et al. (2023). Japanese items are in red and Korean items are in blue.

Table 4: Linear regression model with the mean reassessment scores of Van Hoey et al. (2023) as the dependent variable, and longitudinal score (current study), condition, and language as the predictors.

Term	β	95 % Confidence interval	$t(60)$	p value
Intercept	0.63	[0.56, 0.7]	18.5	<0.001***
Longitudinal Score	0.02	[0, 0.04]	2.19	0.030*
Condition _{true}	0.17	[0.09, 0.26]	3.95	<0.001***
Language _{Korean}	−0.18	[−0.25, −0.1]	−4.63	<0.001***

***Indicates $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

shows that all predictors are significant. A higher longitudinal score predicts guessability ($\beta = 0.02$, $p = 0.03$). A true condition does so even more ($\beta = 0.17$, $p < 0.001$). Korean, however, had an inhibitory effect ($\beta = -0.18$, $p < 0.001$). This means that the current study provides converging evidence with the results of Van Hoey et al. (2023). Different paradigms of iconicity (guessing vs. iterated learning) can lead to similar findings.

4 Handshape resemblance

As a starting point, we look at the correlation between binary and handshape codings (Section 4.1). Next, we look at the general entropy within handshape coding chains (Section 4.2). Finally, we break down handshapes into distinctive physical features to identify clusters of related handshapes (Section 4.3).

4.1 Correlation between the binary coding and handshape coding

The binary and handshape coding systems are correlated for 65 % (613 of 948) of the gestures entered into the analysis. This means that where two sequential gestures are both coded as “bears resemblance” in the binary code, 65 % of the time the handshape coding for those two gestures is uniform too, i.e., the handshape coding bears two identical symbols [YY].³ Such correlation means that the gesture on the whole was successfully transmitted, and the handshape of the gesture remained consistent from one participant to the next.

3 Sequences of handshapes across generations are given in square brackets, e.g., [YY] = “Y handshape followed by Y handshape”, or with a generation number index, e.g., [Y¹Z²] = “Y handshape in generation 1, Z handshape in generation 2”.

There are cases where two sequential gestures were both coded as “bears resemblance” in the binary coding yet, in the handshape coding, received two distinct symbols, e.g., [Y¹Z²] where generation 1 exhibits a different handshape from generation 2. This means that although at first pass the handshape of the gesture was perceived as successfully transmitted (binary coding), when it was later coded with more fine-grained detail (handshape coding) subtle changes were detected from one participant to the next. We presume that, in such cases, the handshapes of these gestures bear enough physical resemblance to one another to “pass” as successful transmissions. Such visual divergence, however subtle, is important because it may lead to gradual evolutionary divergence, perhaps eventually resulting in categorically distinct handshapes. Recall that, in sign languages, handshape has been found to be the phonological factor most likely to undergo change over time, even within a single generation of signers (Moita et al. 2023). If physical resemblance is to help explain why differing handshape symbols occur for gestures otherwise coded as “bears resemblance,” then we would expect such symbols to pattern together sequentially more so than others. One could say that [Z] is likely to follow [Y] (or vice versa) in a transmission chain because the difference between the two is but one finger. We would also expect semblant symbols to be coded as “bears resemblance” more so than “no resemblance.” To put it another way, two gestures in sequence both coded as “no resemblance” should exhibit drastically different handshapes, such that their handshape coding is borne out of error-related chance and is unlikely to occur in other transmission chains.

Frequency of handshape coding symbol occurrence in relation to the binary coding is addressed through Distinctive Collexeme Analysis, a well-known corpus linguistic technique that allows one to uncover preferences between two members of a pairing. Originally, Gries and Stefanowitsch (2004) used this technique to study the strength between different lexemes and variants of a grammatical alternation, e.g. whether *to bring* is more distinctive for the prepositional dative variant or the ditransitive dative variant. Here, we apply this idea to see how likely observed handshape bigrams belong to the handshape coding or the binary coding.

First, we tallied the co-occurrences of handshape coding bigrams and binary coding values to construct contingency tables, e.g., Table 5 for [YZ] and Table 6 for

Table 5: Contingency table for the bigram [YZ].

Coding system	Bigram: [YZ]	¬ Bigram: not [YZ]	Subtotal
Handshape coding	2	474	476
Binary coding	0	196	196
Subtotal	2	670	672

Table 6: Contingency table for the bigram [Y>].

Coding system	Bigram: [Y>]	¬ Bigram: not [Y>]	Subtotal
Handshape coding	1	475	476
Binary coding	1	195	196
Subtotal	2	670	672

[Y>]. Then, we performed log-transformed Fisher-Yates Exact tests (Levshina 2015: 225) and arranged the bigrams based on those values. For illustrative purposes, we use [Y>] and [YZ] (shown in Figure 2). The value of [Y>] is -0.30; that of [YZ] is 0. This means that [Y>] is more distinctive, i.e., more likely to contain a “no resemblance” binary coding than [YZ], which is true, given that Y shows two splayed fingers, Z three splayed fingers and > a fully open hand (again, see Figure 2). All other attested bigrams and their ranking from most likely (very negative values) to least likely (very positive values) to contain a “no resemblance” coding are provided in the OSF repository.

4.2 Entropy

The first step in analyzing patterns of distribution within the handshape coding itself consists of measuring the average information content or uncertainty present in the string of handshape symbols for every transmission chain. For this, we turn to Shannon Entropy (Shannon 1948), which is calculated here based on the handshape symbols within a given transmission chain, including the seed video (superscript 0). For example, *pikapika* in the AR chain has the following handshape coding values: [⁰ 5¹ 6² 6³ >⁴ A⁵] (see OSF for descriptions of handshape symbols). As is clear, there are many different symbols corresponding to high uncertainty, hence the high entropy⁴ score of E = 1.56. Compare this to *chokichoki* in the BR chain, [Y⁰ Y¹ Y² Y³ Y⁴ Y⁵]. There is no uncertainty or surprisal in this sequence, hence the entropy value of E = 0.

We calculated the entropy for every item in a given chain. We then compared the entropy values to the amount of longest successful transmissions in those chains. For example, binary coding values [11110] indicates that there were 4 out of 5 successful (“bears resemblance”) transmissions. A chain of [01111] also contains 4 out of 5 successful transmissions – except there is a critical error following the seed video, but that error was transmitted faithfully until the fifth generation. Figure 9 shows that, in both true and foil conditions, entropy values for a given transmission chain

4 The formula for Shannon Entropy is $-\sum p(x)\log(p(x)) = \sum \log\left(\frac{1}{p(x)}\right)p(x)$.

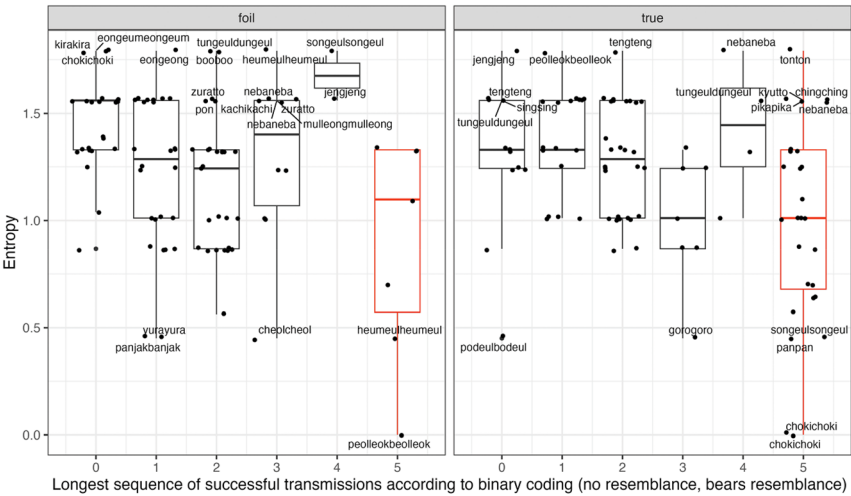


Figure 9: Entropy values per bin of the longest sequence of successful transmissions according to condition. Extreme entropy values are labelled.

tend to be lower the more successful transmissions there are in a row. A linear regression model also confirmed this ($\beta = -0.07$, $p < 0.001$) but found no significant effect of condition ($\beta = 0.01$, $p = 0.79$) to the dependent variable of entropy (models in the OSF data). This is another confirmation that the handshape coding and binary coding are indeed correlated, as we already showed in Section 4.1 through Distinctive Collexeme Analysis.

4.3 Go tile charts to describe articulatory parameters of handshapes

Figure 9 shows that, surprisingly, high entropy tends to occur in shorter sequences of successful transmissions. Furthermore, some of the longest successful sequences exhibit high entropy values, e.g., *tontonton*, *pikapika*, *nebaneba* etc. in the true condition of Figure 9. This appears paradoxical: we should expect a more consistent set of handshape symbols, the longer the sequences of successful transmissions. To account for this, we introduce a feature system for describing articulatory properties of handshapes in our data set.

Our feature system contains four binary articulatory parameters: FIRST KNUCKLE, SECOND KNUCKLE, TIPS TOUCHING, and SPLAY, see Figure 10. They are based on handshape

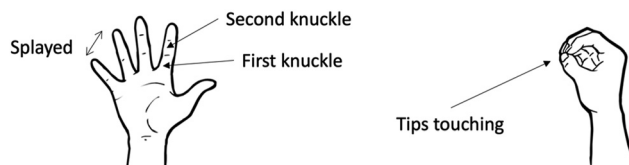


Figure 10: Binary features for describing handshapes: first knuckle, second knuckle, splay, tip touch.

schematizations known from sign language phonology (Brentari 1998; Brentari et al. 2012; Sandler 1989, 2017).

The “first knuckle” refers to the metacarpophalangeal joint where the base of the finger meets the hand. The “second knuckle” refers to the proximal interphalangeal joint in the middle of the finger. Both features pertain to whether the knuckle joint is bent or not. Knuckle bending is coded as absolute and categorical (i.e., exhibits bending versus does not exhibit bending) rather than according to degrees. Since our data does not contain any cases where the third knuckle is bent but the second knuckle is not, we did not include a third knuckle (i.e., distal interphalangeal) parameter. Continuing, “tips touching” refers to whether the end of the finger is in contact with another finger. “Splay” refers to whether the finger is actively made separate, or “spread,” from other fingers so that one or both of its sides is free from contact. If splay is assigned a value of zero, then both its sides are in contact with other fingers.

Our binary system is not about the presence or absence of a particular trait but rather has to do with if one trait is chosen over another. For example, the fact that a knuckle is not bent means that it is straight, or “extended,” and that straightness therefore adds some visual nuance to the overall appearance of a given handshape. Moreover, the degree to which one trait is active, in terms of granularity, e.g., the extent to which a joint is bent or splayed, in one finger versus another is, however, not accounted for.⁵

The feature values for a given handshape can be presented in a diagram, which we call a *GO TILE CHART*, named after the ancient game Go which uses black and white tiles. We present feature values as active (black) or not active (white), see Figure 11. We use “gray” to indicate ambiguity as whether a feature is specified or relevant, e.g., gray for the knuckles = “joint is either bent or straight.” Go tile charts provide a

⁵ Our features [+/-bent] (which apply to the first knuckle and second knuckle) as well as [+/-tip] and [+/- splay] (which apply to all fingers) would fall under “joint configuration”. Terms such as “curved” and “flat” (Brentari et al. 2012) are subsumed by variations of [+/- tips] and [+/-bent]. Explicit degree of finger curvature (i.e., bent vs curved) is not accounted for by our gesture profiles.



Figure 11: Go tile chart for our handshape coding symbol “>” denoting a Splayed-5 hand.

visually intuitive manner of assessing what makes handshapes different from each other. Go tile charts for our handshape data can be found in the OSF repository.

With the feature-based Go tile chart classification system in place, we are now able to address the paradox in Figure 9, namely, that some perfect transmission chains exhibit high entropy in terms of their handshape coding. We hypothesize that this occurs when different handshape symbols are perceived as identical (phonologically) despite subtle physical (phonetic) differences. As *chokichoki* in Figure 2 shows, the scissor-like handshape can occur as with the thumb extended [Y] or not [Z]. When tabulated into Go tile charts, [Y] and [Z] are identical except the splay feature for the thumb, i.e., there is one point of contrast between [Y] and [Z]. This speaks to a strong visual resemblance between the handshapes and thus participants who, not sensitive to such a subtle contrast, used them interchangeably.

We calculated pairwise distances between all attested handshapes using the Go tile charts. We did this by using an implementation of Levenshtein distance, where an edit from white (0) to black (1) was weighed as 1, and an edit from gray (0.5) to white or black as 0.5. After obtaining the distance matrix, we used multiple dimension scaling to reduce the matrix back to two dimensions. What emerges (Figure 12) is a map of the “natural classes” of handshapes. For example, fist-like handshapes are situated on the right, about as far away as possible from the splayed handshapes on the left. This explains in a visual manner why high entropy is still possible within longer sequences of good transmissions: they contain handshapes that belong to what we would consider to be something like the same natural class.

While we have these natural classes, not every handshape exemplar is equally frequent or salient in our data. In Figure 13, we augmented the conceptual space of Figure 12 by including frequency information and pathway frequency between handshapes (successful transmissions only). It is immediately apparent that there are three handshapes extremely high in frequency: [>] is a fully splayed hand

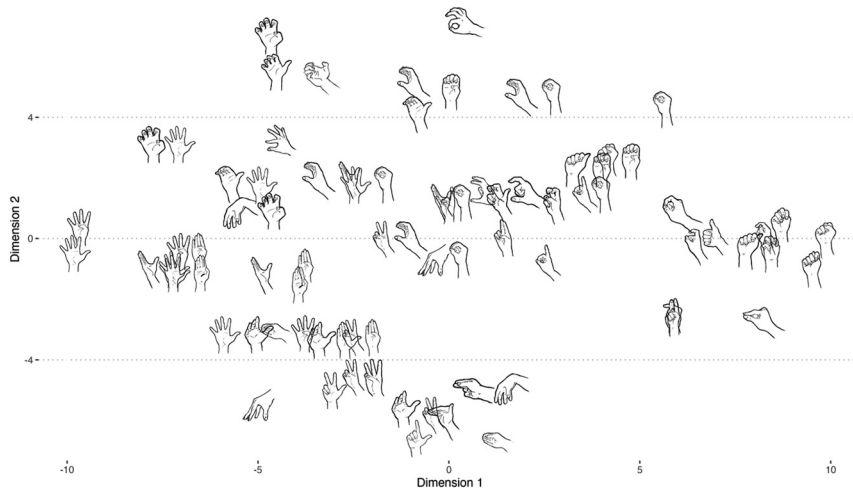


Figure 12: Natural classes among handshapes based on their Go tile chart values.

(i.e., Splayed-5 in Figure 10), [6] is a closed fist, and [A] a loose fist.⁶ In other words, handshapes gravitate towards two extremes: a completely open hand or a completely closed hand. Looking at the thickness of the pathway lines, we observe that highly frequent handshapes are connected to most handshapes, indicating that they are a natural evolutionary destination for handshapes as they progress from one participant to the next (directional arrows we omitted to increase the legibility of the plot). There are noticeable a few thicker lines spanning the plot from [6] to [A]. In cases like these, the handshape presumably did not play a salient role in the successful transmission; rather, it was based on the gesture as a whole, including factors like motion, location, palm orientation etc.

There is of course an anatomical reason participants ended up with such extreme handshapes, which has to do with the opposability of the thumb. In Figure 14, we present a Sankey diagram of finger coordination in perfect transmission chains (see red bar in Figure 3) based on the Go tile chart values. If all Go tile values were the same for a handshape (Figure 14), they are counted as “fully coordinated” (blue). If only the thumb is different, it is “coordinated except thumb” (red). When multiple fingers have a different feature configuration, it is “uncoordinated” (green). The blue and red group remain the largest, while the green group becomes smaller over time. This shift from uncoordinated to coordinated

⁶ Interestingly, these handshapes are unmarked in terms of physiological configuration, age of acquisition, and cross-linguistic attestedness in the sign language literature (Ann, 1993; Eccarius 2002; Marentette & Mayberry 2000; Wong 2008).

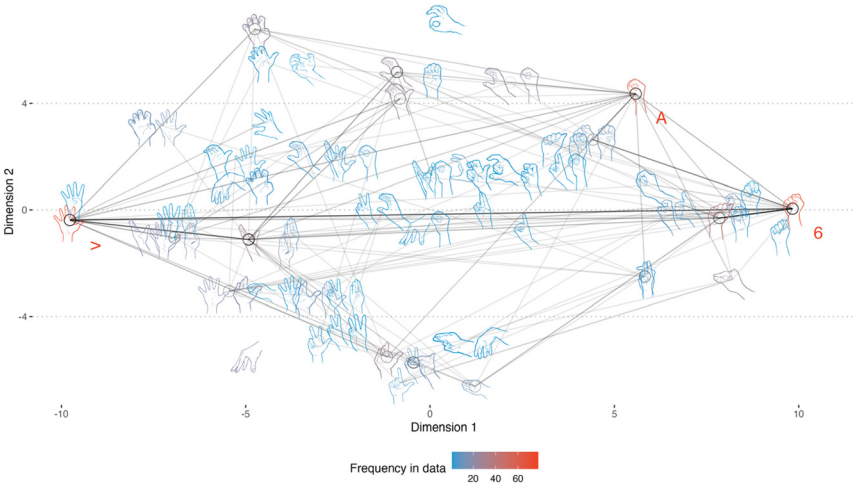


Figure 13: Frequent paths between handshapes. Pathways reflect successful transmissions.

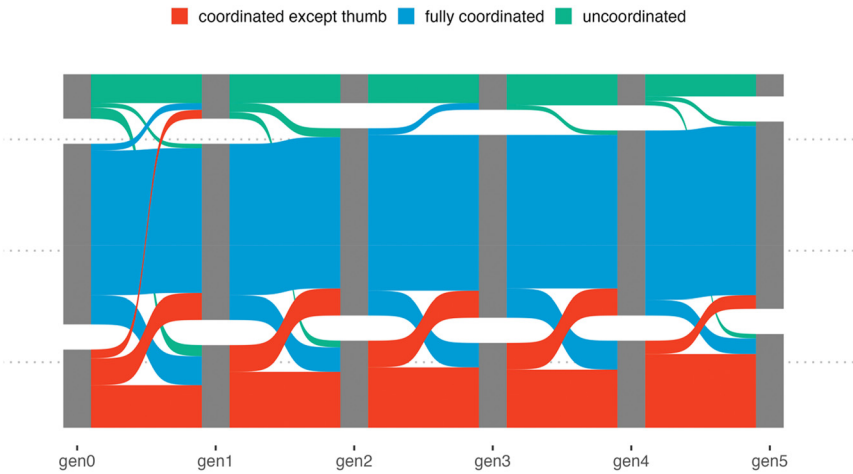


Figure 14: Sankey diagram of the perfect transmission chains, indicating how uncoordinated feature settings decrease in number as the experiment progresses.

would suggest, according to the notion of handshape complexity (Brentari et al. 2012: 7–8), that handshapes become more systematic over time because the fingers gradually assume the same joint configurations, like *perapera* (Figure 5) or *tonton* (Figure 6) – their salience and motivation do not necessary lie in the handshape configuration (presumably in motion in these two cases). Some uncoordinated

gestures, however, resist this drift. Most notable is the example of *chokichoki* “sound of cutting scissors” in the true condition: people are familiar with this handshape (Figure 5). This is a nice illustration of the power of iconicity: it can resist a general drift if items are sufficiently motivated. At the same time, it serves as a reminder that iconicity is in the eye of the beholder (Occhino et al. 2017): if you don’t see the mapping, you are more likely to go along with drifts within the system.

5 Discussion and conclusion

This paper uses a linear iterated learning paradigm to examine the robustness and stability of iconic gestures taken from ideophone-gesture composites as they are transmitted across chains of five participants (L1 Cantonese) unfamiliar with the languages used to derive the target gesture stimuli (Japanese, Korean). Our study is grounded in two key observations from the literature: 1) iconic gestures frequently occur alongside ideophones, to such an extent that we can speak of ideophone-gesture composites, and 2) ideophones belong to an open lexical class. Indeed, ideophones are often improvised in natural discourse. Any co-occurring iconic gestures are assumed to act as clarifying or disambiguating an improvised ideophone’s referent, speaking to their importance. Our primary research question – Do iconic gestures derived from ideophone-gesture composites exhibit more successful transmissions? – is answered affirmatively by way of our binary assessment of whether a given iteration resembles its predecessor (Section 3, binary coding). Our gestures, despite being divorced from natural discourse, presumably rich with other meaningful contextual clues, on their own were able to stand in for ideophone referents. Moreover, these referents, span a range of sensory categories, beyond noun and verbs (objects and actions), speaking to the expressive potential of gestural inventories at the disposal of spoken language users.

In our linear iterated learning paradigm, we put the visual signal in focus, to see whether gestures derived from ideophone-gesture composites develop language-like traits, such as conventionalization, when transmitted from one generation to the next without interference from other factors, like participants having to simultaneously process spoken language input from an unfamiliar language. Transmission was successful both immediately (from one generation to the next) and longitudinally (across all five generations) (Section 3.1). This means that, despite differences in linguistic origin of the stimuli, sensory category of those stimuli, emergence of anatomically driven systematicity, or subtle structural changes in the gestures per participant across generations – iconic gestures were learned better overall (Section 3.2). The implication here is the linguistic viability of these gestures reflected by their gradual conventionalization as a trend toward communicative efficiency. Similar

findings have been described by Moita et al. (2023) in an emerging sign language. However, in that case, a linguistic system of sorts (i.e., a young sign language) is already in place. Our study shows that, even without such a system in place, phonological traits begin to emerge as physical features are filtered out in favor of conventionalization, even without a neat/shared semantic space (e.g., all stimuli are shape-related) as is often the case in iterated learning paradigm studies (see Motamedi et al. 2019; Kirby et al. 2008).

Our binary coding (“no resemblance,” “bears resemblance”) categorically assesses whether or not a gesture as a whole resembles its predecessor and is therefore a successful transmission or not. To ascertain whether any idiosyncrasies or physical changes did in fact emerge from one participant to the next, we devised a set of discrete symbols to identify what handshape occurs at the gesture peak. We call this our handshape coding. We noticed that some transmissions were still considered successful (“bears resemblance”) despite differing in handshape coding from one participant to the next. If treated as units of language, handshapes make up a core component of sign language phonology (Sandler 1989; 2017; Brentari 1998) and yet are notoriously difficult for both native and non-native learners alike to acquire (see Siedlecki and Bonvillian 1993; Bonvillian and Siedlecki 2000; Conlin et al. 2009; Ortega 2017 among others). Handshapes are furthermore likely to undergo significant structural changes as a sign language is developing over time (Moita et al. 2023). Inconsistency in handshape from one participant to the next is thus to be expected and perhaps meaningful. For this reason, we then broke down our handshapes into four binary features to describe their physical characteristics. These features helped us explain why some transmissions were still considered successful (“bears resemblance”) despite exhibiting physical differences in handshape from one participant to the next. Such instances result in a high degree of entropy when the handshape symbols for a given transmission chain are concatenated into a string, e.g., [⁰ 5¹ 6² 6³ >⁴ A⁵] where superscript 0 represents the seed video. Such “noise” among the handshapes is not random but phonetic (Section 4.3). Our four binary features were then converted in a distance matrix showing clusters of different handshapes that were nevertheless very similar in terms of physicality (e.g., imagine all the different ways one can make a fist). Their similarities render an impression of visual relatedness, which in turn speaks to the potential formation of a phonological target. Indeed, this visual relatedness is echoed by sign language notation conventions which group handshapes according to physical appearance (Eccarius and Brentari 2008). The formation of a phonological target is exemplified by the handshape symbols in the string [Y¹ Z² Y³], where each letter represents the handshape of a given generation (superscript) in a transmission sequence. Despite seeing [Z] executed by the previous generation, Generation 3 was still able to produce the [Y] executed by Generation 1.

A global analysis of the handshape symbols in all the gestures that were successfully transmitted across five generations, reveals that handshapes become more systematic over time. Across generations fingers coordinate so as to copy each other (Section 4.3), aside from the thumb which is opposable and thus exceptional when it comes to coordination. Currently, finger coordination is mitigated by iconicity, so that fingers end up coordinating only if the visual relationship to their referent is iconic (as opposed to our foil items). Interestingly, the finger coordination we observe echoes a phenomenon in signed languages (Brentari 1998; Brentari and Eccarius 2010: 290, 294; 2011: 136–137; Hara 2016), whereby unmarked handshapes and handshapes with low complexity both exhibit finger coordination. In our study, a preference for fingers to be in coordination could, if more generations were observed, eventually lead to a bleaching of iconicity in the signal and an opening to arbitrariness. Presumably, and following observations from iterated learning studies in the language evolution literature (e.g., Scott-Phillips and Kirby 2010), this preference for “coordination,” or sameness, would over time lead to merging of handshapes so that the set of overall handshapes learned would shrink, resulting in a kind of homophony or redundancy in the signal. This emergent systematicity is perhaps a catalyst for an evolutionary process which could serve to make the set easier to learn (fewer items) and to execute (regularized formations due to coordination).

In terms of limitations, even though the foils in our study are theoretically not iconic since their forms and meanings were conflated by design, our results cannot tell us much about arbitrariness except that without structural systematicity – such as artificial compositionality and combinatoriality in a neat semantic space, e.g., shapes, actions, objects (Kirby et al. 2008), or nouns and verbs (Motamedi et al. 2019) – our foils are incredibly difficult, if not impossible, to learn. Does arbitrariness ever exist in such a vacuum though? It might be more appropriate to say that our foils were “pseudo-arbitrary” as they lacked the basic mechanics which would have otherwise allowed them to function as full-fledged linguistic units, such as systematicity, compositionality, and combinatoriality (Hockett 1960; Kirby et al. 2015; Tria et al. 2012). Arbitrariness itself probably does not emerge without structural systematicity and definable semantic spaces already in place anyway (Gasser 2004; Kirby et al. 2014; Monaghan et al. 2014; but see Carr et al. 2017). For example, in iterated learning studies looking at language evolution, stimuli often consist of shapes, movements, or colors (clearly defined semantic spaces), whereas the forms are often consonants and vowels in familiar sequences, such as CVCV (adhering to a systematic structure) (Carr et al. 2017; Scott-Phillips and Kirby 2010). In our study, only a handful foils were transmitted successfully across five generations, regardless of condition (see Section 3.1). However, as our focus is on the viability of gestures

from ideophone-gesture composites, and less so the more open-ended question of language evolution, a full analysis of foil transmissions and errors therein is beyond the scope of this paper.

It is likely that our results would have turned out very differently had we relied on an interactional (as opposed to a linear) iterated learning paradigm (Fay et al. 2014; Kirby et al. 2014; Scott-Phillips and Kirby 2010). Our choice harkens back to our original research question about learnability and mental representation of ideophone-gesture composites. Given that ideophones are often found in narratives and expressive contexts, and by design are meant to depict (rather than convey a concisely coded message), our overarching research question is essentially asking “How learnable are visual representations of these depictive words if those visual representations are treated like units of language?” We test that learnability by means of correct and incorrect transmission (binary coding) from one person to the next. Studies which make use of interactional iterated learning paradigms require a back-and-forth between participants and are thus aimed at a different research question: “How and when does change arise to accommodate more efficient comprehension or production?” Indeed, researchers investigating language evolution have already remarked that linear paradigms reveal whether stimuli are learnable, while interactional paradigms show reveal stimuli are expressive enough to be disambiguated from other stimuli. So, if our study were to involve an element of interaction, how would one interpret such results regarding the learnability of preexisting linguistic forms (i.e., from Japanese and Korean)? It is not clear. Finally, the interactional paradigm is not appropriate here because, unlike other studies, we use stimuli which were not intended to serve as a “sandbox” version of a complete language.

To sum up, our study has shown that gestural aspects of ideophone-gesture components are robustly learnable, replicable, and exhibit systematicity in handshape across five generations of participants. Variation in handshapes across successful transmissions is phonetically related and explainable in terms of emergent phonological targets. Beyond our study, spoken language equivalents of the well transmitted gestures were also shown to be guessed with accuracy at an above chance level by non-native speakers in Van Hoey et al. (2023), further highlighting the complementary nature of the visual and spoken modalities in formulating mental representations of ideophones. Methodologically, we have shown that guessing paradigms and iterated learning paradigms can lead to similar constataions (Section 3.2), but that the iterated learning paradigm further allows us to inspect the emergence of structural changes at a more fine-grained level, on top of those general conclusions.

Data availability statement

Supplementary data for this paper are available in the OSF repository (<https://osf.io/dtr4p/>). This repository holds the data, an R markdown detailing how the different figures and numbers were obtained, as well as the original seed videos and blurred experiment videos.

Acknowledgments: We would like to thank Chihiro Yokoyama and Jinyoung Hwang for providing the seed gestures, as well as the participants for the experiments, and the members of the Language Development Lab at the University of Hong Kong for their early feedback. Additionally, we are grateful to the audience of our presentation at DGfS 2023 in Cologne and the reviewers of our paper at Cognitive Linguistics for their thoughtful comments.

Competing interests: Not applicable.

Research funding: This work was supported by General Research Funds 17602723 and 17607522 awarded to the corresponding author by the Research Grant Council of Hong Kong.

References

- Ann, Jean. 1993. *A linguistic investigation of physiology and handshape*. Tucson, AZ: University of Arizona PhD dissertation.
- Assaneo, María Florencia, Juan Ignacio Nichols & Marcos Alberto Trevisan. 2011. The anatomy of onomatopoeia. (ed.) Mariano Sigman. *PLoS One* 6(12). e28317.
- Bonvillian, John D. & Theodore Siedlecki. 2000. Young children's acquisition of the formational aspects of American Sign Language: Parental report findings. *Sign Language Studies* 1(1). 45–64.
- Brentari, Diane. 1998. *A prosodic model of sign language phonology*. Cambridge: MIT Press.
- Brentari, Diane & Petra Eccarius. 2010. Handshape contrasts in sign language phonology. In Diane Brentari (ed.), *Sign languages* (Cambridge Language Surveys), 284–311. Cambridge: Cambridge University Press.
- Brentari, Diane & Petra Eccarius. 2011. When does a system become phonological? Potential sources of handshape contrast in sign languages. In Rachel Channon & Harry Van Der Hulst (eds.), *Formational units in sign languages*, 125–150. Nijmegen: De Gruyter Mouton & Ishara Press.
- Brentari, Diane, Marie Coppola, Laura Mazzoni & Susan Goldin-Meadow. 2012. When does a system become phonological? Handshape production in gesturers, signers, and homesigners. *Natural Language & Linguistic Theory* 30(1). 1–31.
- Cano, María Graciela. 2020. *The categorization of ideophone-gesture composites in Quichua narratives*. Provo, UT: Brigham Young University Master thesis.
- Carr, Jon W., Kenny Smith, Hannah Cornish & Simon Kirby. 2017. The cultural evolution of structured languages in an open-ended, continuous world. *Cognitive Science* 41(4). 892–923.
- Cates, Deborah, Eva Gutiérrez, Sarah Hafer, Ryan Barrett & David Corina. 2013. Location, location, location. *Sign Language Studies*. Gallaudet University Press 13(4). 433–461.

- Conlin, Kimberly E., Gene R. Mirus, Claude Mauk & Richard P. Meier. 2009. The acquisition of first signs: Place, handshape, and movement. In Charlene Chamberlain, Jill Patterson Morford & Rachel I. Mayberry (eds.), *Language acquisition by eye*, 51–69. New York: Psychology Press.
- Dargue, Nicole & Naomi Sweller. 2020. Learning stories through gesture: Gesture's effects on child and adult narrative comprehension. *Educational Psychology Review* 32(1). 249–276.
- Diffloth, Gérard. 1979. Expressive phonology and prosaic phonology in Mon-Khmer. In Theraphan L. Thongkum (ed.), *Studies in Mon-Khmer and Thai phonology and phonetics in honor of E. Henderson*, 49–59. Bangkok: Chulalongkorn University Press.
- Dingemanse, Mark. 2011. *The meaning and use of ideophones in Siwu*. Nijmegen: Radboud University Nijmegen dissertation.
- Dingemanse, Mark. 2013. Ideophones and gesture in everyday speech. *Gesture* 13(2). 143–165.
- Dingemanse, Mark. 2019. "Ideophone" as a comparative concept. In Kimi Akita & Prashant Pardeshi (eds.), *Ideophones, mimetics and expressives* (Iconicity in Language and Literature, ILL 16), 13–33. Amsterdam: John Benjamins.
- Dingemanse, Mark & Kimi Akita. 2017. An inverse relation between expressiveness and grammatical integration: On the morphosyntactic typology of ideophones, with special reference to Japanese. *Journal of Linguistics* 53(3). 1–32.
- 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.
- Drijvers, Linda & Asli Özyürek. 2017. Visual context enhanced: The joint contribution of iconic gestures and visible speech to degraded speech comprehension. *Journal of Speech, Language, and Hearing Research* 60(1). 212–222.
- Eccarius, Petra. 2002. *Finding common ground: A comparison of handshape across multiple sign languages*. West Lafayette: Purdue University Master thesis.
- Eccarius, Petra & Diane Brentari. 2008. Handshape coding made easier: A theoretically based notation for phonological transcription. *Sign Language and Linguistics* 11(1). 69–101.
- Fay, Nicolas, Mark Ellison & Simon Garrod. 2014. Iconicity: From sign to system in human communication and language. *Pragmatics and Cognition* 22(2). 244–263.
- Gasser, Michael. 2004. The origins of arbitrariness in language. In Kenneth D. Forbus, Dedre Gentner & Terry Regier (eds.), *Proceedings of the 26th annual meeting of the Cognitive Science Society*, 434–439. Merced: UC Merced.
- Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collocation analysis: A corpus-based perspective on `alternations. *International Journal of Corpus Linguistics* 9(1). 97–129.
- Hamano, Shoko. 1998. *The sound-symbolic system of Japanese* (Studies in Japanese Linguistics). Stanford, Calif.; Tokyo: CSLI Publications; Kurocio.
- Hara, Daisuke. 2016. An information-based approach to the syllable formation of Japanese Sign Language. In Masahiko Minami (ed.), *Handbook of Japanese applied linguistics*, 457–482. Berlin: De Gruyter.
- Hatton, Sarah Ann. 2016. *The onomatopoeic ideophone-gesture relationship in Pastaza Quichua*. Provo: Brigham Young University Master thesis.
- Hinojosa, José A., Juan Haro, Sara Magallares, Jon Andoni Duñabeitia & Pilar Ferré. 2021. Iconicity ratings for 10,995 Spanish words and their relationship with psycholinguistic variables. *Behavior Research Methods* 53. 1262–1275.
- Hockett, Charles F. 1960. The origin of speech. *Scientific American* 203(3). 88–96.
- Hodge, Gabrielle & Lindsay Ferrara. 2022. Iconicity as multimodal, polysemiotic, and plurifunctional. *Frontiers in Psychology* 13. 808896.

- Iwasaki, Noriko & Keiko Yoshioka. 2019. Iconicity in L2 Japanese speakers' multi-modal language use: Mimetics and co-speech gesture in relation to L1 and Japanese proficiency. In Kimi Akita & Prashant Pardeshi (eds.), *Ideophones, mimetics and expressives (Iconicity in Language and Literature, ILL 16)*, 265–302. Amsterdam: John Benjamins.
- Kandana Arachchige, Kendra G., Isabelle Simoes Loureiro, Wivine Blekic, Mandy Rossignol & Laurent Lefebvre. 2021. The role of iconic gestures in speech comprehension: An overview of various methodologies. *Frontiers in Psychology* 12. 634074.
- Karadöller, Dilay Z., David Peeters, Francie Manhardt, Aslı Özyürek & Gerardo Ortega. 2024. Iconicity and gesture jointly facilitate learning of second language signs at first exposure in hearing nonsigners. *Language Learning*. lang.12636. <https://doi.org/10.1111/lang.12636>.
- Karnopp, Lodenir Becker. 2002. Phonology acquisition in Brazilian Sign Language. In Gary Morgan & Bencie Woll (eds.), *Directions in sign language acquisition*, 2, 29–53. Amsterdam: John Benjamins.
- Kelly, Spencer D., Aslı Özyürek & Eric Maris. 2010. Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science* 21(2). 260–267.
- Kendon, Adam. 2004. *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.
- Kirby, Simon, Hannah Cornish & Kenny Smith. 2008. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences* 105(31). 10681–10686.
- Kirby, Simon, Tom Griffiths & Kenny Smith. 2014. Iterated learning and the evolution of language. *Current Opinion in Neurobiology* 28. 108–114.
- Kirby, Simon, Monica Tamariz, Hannah Cornish & Kenny Smith. 2015. Compression and communication in the cultural evolution of linguistic structure. *Cognition* 141. 87–102.
- Kita, Sotaro. 1997. Two-dimensional semantic analysis of Japanese mimetics. *Linguistics* 35. 379–415.
- Kunene, Daniel P. 1965. The ideophone in Southern Sotho. *Journal of African Languages* 4. 19–39.
- Kunene, Daniel P. 2001. Speaking the act: The ideophone as a linguistic rebel. In Erhard Friedrich Karl Voeltz & Christa Kilian-Hatz (eds.), *Ideophones (Typological studies in language 44)*, 183–192. Amsterdam: John Benjamins.
- Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam; Philadelphia: John Benjamins.
- Lockwood, Gwilym, Peter Hagoort & Mark Dingemanse. 2016. How iconicity helps people learn new words: Neural correlates and individual differences in sound-symbolic bootstrapping. *Collabra* 2(1). 1–15.
- Marentette, Paula F. & Rachel I. Mayberry. 2000. Principles for an emerging phonological system: A case study of early ASL acquisition. In Charlene Chamberlain, Jill Patterson Morford & Rachel I. Mayberry (eds.), *Language acquisition by eye*, 71–90. New York: Psychology Press.
- McLean, Bonnie, Michael Dunn & Dingemanse Mark. 2023. Two measures are better than one: Combining iconicity ratings and guessing experiments for a more nuanced picture of iconicity in the lexicon. *Language and Cognition* 15(4). 1–24.
- Mihás, Elena. 2013. Composite ideophone-gesture utterances in the Ashéninka Perené 'community of practice', an Amazonian Arawak society from Central-Eastern Peru. *Gesture* 13(1). 28–62.
- Moita, Mara, Ana Maria Abreu & Ana Mineiro. 2023. Iconicity in the emergence of a phonological system? *Journal of Language Evolution* 8(1). 1–17.
- 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). 20130299.
- Motamedi, Yasamin, Marieke Schouwstra, Kenny Smith, Jennifer Culbertson & Simon Kirby. 2019. Evolving artificial sign languages in the lab: From improvised gesture to systematic sign. *Cognition* 192. 103964.

- Nasu, Akio. 2015. The phonological lexicon and mimetic phonology. In Haruo Kubozono (ed.), *Handbook of Japanese phonetics and phonology*, 253–288. Berlin: De Gruyter Mouton.
- Nielsen, Alan K. S. & Mark Dingemanse. 2021. Iconicity in word learning and beyond: A critical review. *Language and Speech* 64(1). 1–21.
- Nuckolls, Janis B. 2020. “How do you even know what ideophones mean?”: Gestures’ contributions to ideophone semantics in Quichua. *Gesture* 19(2–3). 161–195.
- Nuckolls, Janis B., Joseph A. Stanley, Elizabeth Nielsen & Roseanna Hopper. 2016. The systematic stretching and contracting of ideophonic phonology in Pastaza Quichua. *International Journal of American Linguistics* 82(1). 95–116.
- Occhino, Corrine, Benjamin Anible & Jill P. Morford. 2020. The role of iconicity, construal, and proficiency in the online processing of handshape. *Language and Cognition* 12(1). 114–117.
- 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* 16(1). 100–126.
- Ortega, Gerardo. 2017. Iconicity and sign lexical acquisition: A review. *Frontiers in Psychology* 8(1280). 1–14.
- Peirce, Jonathan, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman & Jonas Kristoffer Lindeløv. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* 51(1). 195–203.
- 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.
- 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. <https://doi.org/10.3389/fpsyg.2010.00227>.
- Rácz, Péter. 2013. *Saliency in sociolinguistics: A quantitative approach*, Vol. 84. Berlin: De Gruyter Mouton.
- Sandler, Wendy. 1989. *Phonological representation of the sign: Linearity and nonlinearity in American Sign Language*. Berlin: De Gruyter Mouton.
- Sandler, Wendy. 2017. The challenge of sign language phonology. *Annual Review of Linguistics* 3(1). 43–63.
- Schmid, Hans-Jörg (ed.). 2017. *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge* (Language and the Human Lifespan). Washington, DC: Berlin: American Psychological Association; De Gruyter Mouton.
- Schryver, Gilles-Maurice de. 2009. The lexicographic treatment of ideophones in Zulu. *Lexicos (AFRILEX)* 19. 34–54.
- Scott-Phillips, Thomas C. & Simon Kirby. 2010. Language evolution in the laboratory. *Trends in Cognitive Sciences* 14(9). 411–417.
- Shannon, Claude E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27(3). 379–423.
- Siedlecki, Theodore & John D. Bonvillian. 1993. Location, handshape & movement: Young children’s acquisition of the formational aspects of American Sign Language. *Sign Language Studies* 78(1). 31–52.
- Taitz, Alan, M. Florencia Assaneo, Natalia Elisei, Mónica Trípodí, Laurent Cohen, Jacobo D. Sitt & Marcos A. Trevisan. 2018. The audiovisual structure of onomatopoeias: An intrusion of real-world physics in lexical creation. *PLoS One* 13(3). e0193466.
- Thompson, Arthur Lewis & Youngah Do. 2019a. Defining iconicity: An articulation-based methodology for explaining the phonological structure of ideophones. *Glossa: A Journal of General Linguistics* 4(1). 72.
- Thompson, Arthur Lewis & Youngah Do. 2019b. Unconventional spoken iconicity follows a conventional structure: Evidence from demonstrations. *Speech Communication* 113. 36–46.

- Thompson, Arthur Lewis, Kimi Akita & Youngah Do. 2020. Iconicity ratings across the Japanese lexicon: A comparative study with English. *Linguistics Vanguard* 6(1). 20190088.
- Thompson, Arthur Lewis, Thomas Van Hoey & Youngah Do. 2021. Articulatory features of phonemes pattern to iconic meanings: Evidence from cross-linguistic ideophones. *Cognitive Linguistics* 32(4). 563–608.
- Tria, Francesca, Bruno Galantucci & Vittorio Loreto. 2012. Naming a structured world: A cultural route to duality of patterning. (Ed.) Emmanuel Andreas Stamatakis. *PLoS One* 7(6). e37744.
- Van Hoey, Thomas. 2020. *Prototypicality and salience in Chinese ideophones: A cognitive and corpus linguistics approach* 中文擬聲(態)詞的原型與顯著特徵:以認知與語料庫語言學方法探討. Taipei: National Taiwan University PhD dissertation.
- Van Hoey, Thomas. 2023. A semantic map for ideophones. In Thomas Fuyin Li (ed.) *Handbook of cognitive semantics*, Vol. 2, 129–175. Leiden: Brill.
- Van Hoey, Thomas, Arthur L. Thompson, Youngah Do & Mark Dingemanse. 2023. Iconicity in ideophones: Guessing, memorizing, and reassessing. *Cognitive Science* 47(4). e13268.
- Voeltz, Erhard Friedrich Karl & Christa Kilian-Hatz (eds.). 2001. *Ideophones (Typological Studies in language 44)*. Amsterdam; Philadelphia: J. Benjamins.
- Winter, Bodo, Marcus Perlman, Lynn K. Perry & Gary Lupyan. 2017. Which words are most iconic? Iconicity in English sensory words. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems* 18(3). 443–464.
- Winter, Bodo, Gary Lupyan, Lynn K. Perry, Mark Dingemanse & Perlman Marcus. 2023. Iconicity ratings for 14,000+ English words. *Behavior Research Methods* 56. <https://doi.org/10.3758/s13428-023-02112-6>.
- Wong, Yuet On. 2008. *Acquisition of handshape in Hong Kong Sign Language: A case study*. Hong Kong: Chinese University of Hong Kong Master thesis.
- Yap, De-Fu, So Wing-Chee, Ju-Min Melvin Yap, Ying-Quan Tan & Ruo-Li Serene Teoh. 2011. Iconic gestures prime words. *Cognitive Science* 35(1). 171–183.
- Yuan, Eric. 2021. *Zoom*. San Jose, CA: Zoom Video Communications Inc.