LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Which Aspects of Visual Motivation Aid the Implicit Learning of Signs at First Exposure?

Julia Hofweber,[a] Lizzy Aumônier,[b] Vikki Janke,[b]
Marianne Gullberg,[c] and Chloë Marshall[a]

[a]University College London [b]University of Kent [c]Lund University

Abstract: We investigated whether sign-naïve learners can infer and learn the meaning of signs after minimal exposure to continuous, naturalistic input in the form of a weather

---

CRediT author statement – **Julia Hofweber**: conceptualization; data curation; formal analysis; methodology; investigation; visualization; writing – original draft preparation; writing – review & editing. **Lizzy Aumônier**: conceptualization; formal analysis; investigation; methodology; writing – review & editing. **Vikki Janke**: conceptualization; formal analysis; funding acquisition; methodology; project administration and supervision; writing – original draft preparation; writing – review & editing. **Marianne Gullberg**: conceptualization; formal analysis; funding acquisition; methodology; project administration and supervision; resources; writing – original draft preparation; writing – review & editing. **Chloë Marshall**: conceptualization; formal analysis; funding acquisition; investigation; methodology; project administration and supervision; resources; writing – original draft preparation; writing – review & editing.

A one–page Accessible Summary and a Video Summary in International Sign Language of this article in non–technical language is freely available in the Supporting Information online and at https://oasis-database.org

This work was funded by an International Academic Fellowship (IAF2016023 awarded to Chloë Marshall) and a Research Grant (RPG-2018-333 awarded to Chloë Marshall, Marianne Gullberg, and Vikki Janke) from the Leverhulme Trust. We thank Linnéa Lindén, Peter Roslund, Henrik Garde, and Josien Greidanus at Lund University Humanities Lab for their assistance with creating the experimental materials. We are also grateful to Claudia Civinini at the Institute of Education, University College London, for her assistance with data collection, and to Maria Jomy at the Institute of Education, University College London, for her assistance with preparing the figures.

Correspondence concerning this article should be addressed to Chloë Marshall, University College London, 20 Bedford Way, London WC1H 0AL, UK. Email: chloe.marshall@ucl.ac.uk

The handling editor for this manuscript was David Quinto-Pozos. The special issue was coordinated by Theres Grüter.

DOI: 10.1111/lang.12587

forecast in Swedish Sign Language. Participants were L1-English adults. Two experimental groups watched the forecast once ($n = 40$) or twice ($n = 42$); a control group did not ($n = 42$). Participants were then asked to assign meaning to 22 target signs. We explored predictors of meaning assignment with respect to item occurrence frequency and three facets of visual motivation: iconicity, transparency, and gesture similarity. Meaning assignment was enhanced by exposure and item frequency, thereby providing evidence for implicit language learning in a new modality, even under challenging naturalistic conditions. Accuracy was also contingent upon iconicity and transparency, but not upon gesture similarity. Meaning assignment at first exposure is thus visually motivated, although the overall low accuracy rates and further qualitative analyses suggest that visually motivated meaning assignment is not always successful.

Keywords implicit learning; sign language; iconicity; transparency; gesture; meaning

Introduction

The largely arbitrary nature of form–meaning mappings is well established (e.g., Hockett, 1959; Saussure, 1916). Nevertheless, a growing number of more recent empirical studies and review papers (see Figure 1 in Nielsen & Dingemanse, 2020) propose that *motivated* form–meaning mappings are more prevalent in language and more relevant to its processing and acquisition (including second language acquisition) than previously acknowledged. This resurgent interest has arguably been driven by a widening of the range of languages studied by psycholinguists to include Asian and African languages, where ideophones (i.e., words expressing sensory imagery) are relatively common (Akita & Dingemanse, 2019), and sign languages, where the visual modality offers greater affordance for motivated mappings between form and meaning, and which are therefore rich in iconicity (Perniss et al., 2010).

At the lexical level, iconicity can be defined in different ways (Motamedi et al., 2019), but at its heart is the notion that there is some resemblance between the referent and the visual (for sign languages) or auditory (for spoken languages) form of the word (e.g., Dingemanse et al., 2020; Emmorey, 2014; Perniss et al., 2010). For example, there is a visual resemblance between the configuration, location, and movement of the hand in the American Sign Language (ASL) sign for "bird" and the opening and closing of a bird's beak and a resemblance between the configuration, location, and movement of the arms and hands in the Turkish Sign Language sign for "bird" and the flapping of a bird's wings (see Figure 1, Emmorey, 2014). Similarly, the English phrase "pitter patter" conveys in its sound form something of the sound of lightly falling rain or of quick and light footsteps. An important issue in psycholinguistic research is whether the presence of lexical iconicity in a language has any impact on how that language is acquired, either by young first language learners or older second language learners.

There is growing evidence that iconicity *does* affect language learning. This is the case for both signed and spoken word learning in first language acquisition (e.g., Imai et al., 2008, for Japanese; Motamedi et al., 2021, for English; Thompson et al., 2012, for BSL), although the extent to which the facilitating effects of iconicity can be distinguished from the facilitating effects of neighbourhood density, frequency, prosody, and reduplication is debated (Caselli & Pyers, 2017; Laing et al., 2017; Ota et al., 2018). When Dutch-speaking adults were introduced to spoken words of Japanese, Korean, Semai, Siwu, and Ewe in laboratory studies, iconicity helped them choose between two possible meanings at above-chance levels of accuracy (Dingemanse et al., 2016). Furthermore, co-speech gestures have similar semiotic affordances for iconicity as signs, and gestures expressing content iconically have been found to promote lexical and phonological learning in children and adults (e.g., Baills et al., 2019; Kelly et al., 2009; Morett, 2014; So et al., 2012; Tellier, 2008). Therefore, iconicity does appear to play a role language learning, particularly where it is prevalent.

The aforementioned studies raise interesting issues with respect to the role of iconicity in lexical learning, including whether (a) learning is only supported in contexts where the learner is given both the form and the meaning of a lexical item and is either told explicitly about the item's iconicity or has sufficient conceptual knowledge to deduce the iconic link between form and meaning; or alternatively (b) learning is also supported in implicit contexts where the learner is *not* given the meaning and has, instead, to deduce the meaning from just the form and their conceptual knowledge of what that form might plausibly refer to. Put another way, if the learner is in a situation where they have to work out lexical meaning for themselves, is iconicity usable as a cue to meaning?

In this regard, a currently uninvestigated research question concerns whether iconicity can support the correct assignment of lexical meaning in one particular learning context: namely when a new language is presented relatively naturalistically—in other words, implicitly and in continuous (unsegmented) form—and when the modality is unfamiliar. This is an important test of the power of iconicity because it represents a challenging learning situation, but also one that is not uncommon "in the wild": not all second-language learners learn their new language in the classroom via explicit instruction and with the language considerately presented in translated, word-sized chunks. This naturalistic situation, therefore, represents a strong and ecologically valid test of the limits of iconic support for initial second-language learning. Given the greater potential for iconicity in the visuo-gestural modality than in the oral-aural modality (Goldin-Meadow & Brentari, 2017), if iconicity does help learners in naturalistic situations, it might be particularly well illustrated during the first stages of learning a sign language. Evidence from cross-modal priming shows that iconicity modulates implicit processing in hearing early learners of ASL (Mott et al., 2020), which suggests that learners draw upon iconicity when making sense of new sign input.

In the present study we invited adult hearing native speakers of English, with no previous experience of sign language, to view a short weather forecast presented solely in Swedish Sign Language. We asked (a) Are participants able to learn anything about the meaning of individual signs in this situation? (b) Are they able to learn the meaning of highly iconic signs more easily than the meaning of less iconic signs? In the remainder of this introduction, we distinguish iconicity from two related notions that we also consider in our study, namely transparency and gestural similarity. We then review essential literature on the second-language learning of form–meaning mappings at first exposure in naturalistic situations before narrowing down to our own research questions and predictions.

Background Literature

Visually Motivated Form–Meaning Mappings

Our study assesses mapping processes based on formal similarity between (a) a linguistic form that is not familiar to the participant and (b) participants' existing conceptual and visual representations. We follow Sevcikova Sehyr and Emmorey (2019) in drawing a distinction between two different types of visually motivated form–meaning mappings, namely iconicity ("the resemblance between a form and a given meaning," p. 208) and transparency ("the ability to infer the meaning of a lexical item based on the form," p. 208; i.e., where the learner is not provided with the meaning *a priori*). Although the terms "iconicity" and "transparency" are still used seemingly interchangeably by some authors (e.g., Ortega & Özyürek, 2020), we find it conceptually useful to distinguish between them because they assume different things with respect to the knowledge that learners can draw upon when learning lexical items in a new language. Crucially, neither concept is a binary phenomenon. Items are not either iconic/transparent or noniconic/nontransparent; rather, they differ in their *degree* of iconicity/transparency.

Although iconicity and transparency are highly related (highly iconic signs are also often highly transparent), they are not entirely overlapping: The meaning of highly iconic items might not be easily inferred by participants unfamiliar with the language and culture, and therefore such items might not be transparent (Klima & Bellugi, 1979; Sevcikova Sehyr & Emmorey, 2019). An example of such an iconicity–transparency discrepancy from ASL can be found in Sevcikova Sehyr and Emmorey (2019): ASL refers to the notion of “my” by placing a hand on the chest. The concrete spatial representation ‘chest’ thus represents the more abstract and metaphorical notion of personal possession. Whilst the sign MY was rated as highly iconic by participants, the majority incorrectly guessed that it meant ‘chest’, suggesting low transparency (Sevcikova Sehyr & Emmorey, 2019). Such preferences for direct over abstracted mappings in meaning assignment tasks have also been reported by Ortega et al. (2019) and might be one of the sources of iconicity–transparency discrepancies. Although many signs involve iconicity, they might not be fully transparent, which could therefore pose a challenge for learners.

Research on the impact of visually motivated form–meaning mappings on learning, both in naturalistic contexts and in experimental work, has generally focused on iconicity (e.g., Caselli & Pyers, 2017; Imai et al., 2008; Lockwood et al., 2016; Thompson et al., 2012). In order to quantify iconicity for such studies, raters are usually presented with a form and its meaning and asked to provide a form–meaning similarity judgment along a Likert scale, with the ends of the scale labelled something like *not at all similar* (i.e., arbitrary) and *very similar* (i.e., highly iconic) (Motamedi et al., 2019). Iconicity ratings obtained via this method arguably tap into the processes involved in explicit learning contexts, in which learners are given a form alongside its meaning. However, in implicit learning contexts of the sort we examine in the current study, learners have to infer meaning based on the form of a sign, the context in which it occurs, and their own language and world knowledge; the precise meaning of each sign is not provided. It is therefore not clear how relevant iconicity ratings are for predicting learning in implicit learning contexts. Transparency, on the other hand, is assessed in tasks where participants are provided solely with a form and are required to guess its meaning. This mirrors more closely what happens in implicit learning contexts. Because transparency taps into the skill of assigning meaning to a novel sign without being told explicitly about the form–meaning mapping, it might be a more accurate predictor of implicit learning.

Whereas the notion of transparency taps into the skill of inferring a sign’s meaning from its form, researchers have recently turned their attention to the converse situation, namely the sort of forms sign-naïve participants produce with their hands when required to gesture certain concepts silently (i.e., without using their voice; what Goldin-Meadow & Brentari, 2017, term ‘spontaneous signs’). In particular, the focus has been on how closely such silent gestures resemble the signs of a sign language (Ortega & Özyürek, 2020; Ortega et al, 2019). For example, Ortega et al. (2019) found that Dutch sign-naïve gesturers systematically produced gestures that were identical in form to the Sign Language of the Netherlands sign TO-CUT-WITH-SCISSORS (see their Figure 1). Signs whose elicited gestures were identical or resembled them closely—such as the aforementioned example—were also judged by sign-naïve participants to be more iconic than those signs for which the gesture–sign overlap was lower or nonexistent, e.g. LAPTOP (again, see their Figure 1). Ortega et al. (2019, p. 1) suggest that “gestures that overlap in form with signs may serve as some type of ‘manual cognates’ that help non-signing adults to break into a new language at first exposure.” The prediction is that, just as cognates boost the learning of new spoken/written words (de Groot & van Hell, 2005), a similarity between learners’ existing gestural representations (as assessed by a gesture elicitation task) and the form of the conventionalized sign may also boost learning. Although Ortega et al.’s (2019) findings point in this direction, they did not test the predicted cognate effect of gesture directly in a language-learning context. In this study, we build on Ortega et al.’s (2019) work by investigating whether such gesture similarity effects predict participants’ accuracy in inferring and learning meaning of signs.

To summarize, for the current study we conceptualize the above-described processes of form–meaning mapping based on visual similarity as a multifaceted phenomenon and investigate those different facets in order to isolate the similarity factors that support sign-naïve participants’ ability to infer and learn meanings in a sign language. Specifically, we investigate the phenomenon of visually motivated meaning assignment, and the extent to which its success is predicted by the iconicity, transparency and gesture similarity of signs.

Investigating Implicit Learning of Form–Meaning Mappings at First Exposure to Naturalistic Input

Our focus is on the extent to which form–meaning mappings can be learned during the first few minutes of exposure to a new language in a new modality (i.e., a sign language), in a context where this new language is presented unsegmented and in an untutored, noninteractive setting. We adapt a paradigm whereby novices view a specially recorded weather forecast without any specific instructions, and then, immediately afterwards, take a surprise test to determine what aspects of the new language they have (implicitly) learned from the forecast (Gullberg et al., 2010; Gullberg et al., 2012; Veroude et al., 2010). The manipulation of certain features of the input material in the forecast allows the characteristics of implicit learning to be investigated. For example, manipulating the number of times the forecast is viewed ("exposure") and the occurrence frequency of the target items in the forecast allows researchers to characterize the extent to which the number of encounters with a novel lexical item supports learning in this context.

In a series of studies using this paradigm, Gullberg and colleagues have demonstrated that "the adult learning mechanism is more powerful than is normally assumed when faced with small amounts of complex, continuous… language input" (Gullberg et al., 2010, p. 5). They showed that adult Dutch speakers were able to extract segmental, phonotactic, and lexical information about the target language (Mandarin Chinese) after watching just a few minutes of an audio-visually presented weather forecast in that language. Relevant to the present study, they found that participants performed above chance levels when deciding whether a picture of a weather symbol matched an auditorily presented word form, and that both occurrence frequency and gestural highlighting (i.e., the presenter gesturing to the weather symbol during the forecast) predicted accuracy.

The weather forecast has some advantages for our purposes. As a particular discourse type aimed at the general public, it is likely to be very familiar to participants. It should therefore generate a predictable semantic field for viewers, in terms of the words and concepts prototypically associated with weather forecasts (Moore Mauroux, 2016). This discourse genre was therefore chosen not only because it was used in previous first-exposure studies of spoken language (Gullberg et al., 2010, 2012; Veroude et al., 2010), but also because it could be adapted for presentation in Swedish Sign Language (Svenskt teckenspråk, STS) and still retain its familiarity.

The STS weather forecast used in the present study, and which we describe in detail in the Methods section, was also used in a recent study by Hofweber et al. (2022). In that study, participants undertook a sign recognition task immediately after they had viewed the forecast. Stimuli in the recognition task included a mix of target signs (which had appeared in the forecast) and distractor signs (which had not appeared). As in the present study, target items differed in their occurrence frequency in the forecast and in their degree of iconicity. Immediately after viewing each sign, participants made a yes–no decision as to whether they had seen that sign during the forecast. Frequency and iconicity cumulatively facilitated participants' recognition of target signs. Hofweber et al.'s (2022) findings extend those of Gullberg et al. (2010) from Mandarin Chinese to a sign language and suggest that the adult mechanism for implicit language learning might operate similarly for both signed and spoken languages as regards frequency, while also exploiting modality-salient properties—in this case, iconicity. However, this conclusion remains tentative because just one task was used, and that task did not test participants' access to sign meaning. It is still unclear whether adults can infer and learn the correct meaning of signs from brief naturalistic input, and whether the nature of the input (frequency, iconicity, transparency, gesture similarity) affects the accuracy of meaning assignment.

It is worth stressing that although the considerable literature on statistical and artificial language learning also examines (implicit) learning at first exposure (e.g., Christiansen, 2019; Monaghan et al., 2019), such studies typically involve either pretraining of items, much more frequent occurrence of relevant items, and/or longer exposure than a few minutes. Therefore, arguably, the paradigm used in the current study approximates a more naturalistic setting.

The Present Study

We investigated how accurately sign-naïve viewers can assign meaning to sign forms. While aiming to design a weather forecast input that was as faithful as possible to real-life weather forecasts, we manipulated the occurrence frequency and the iconicity of a subset of lexical items, our so-called "target" items, which were used in the meaning assignment task. Target signs were assigned to one of two frequency sets ("low" and "high" frequency, appearing three or eight times in the forecast, respectively) and were matched with respect to iconicity (as measured in a separate task with independent groups of sign-naïve participants). We also manipulated the number of times that participants viewed the forecast: zero, once, or twice.

Our research questions and predictions were:

1 Can sign-naïve adults infer and learn the correct meaning of individual signs in an implicit learning context involving briefly presented naturalistic input?

Based on Gullberg et al.'s findings (2010, 2012), which used the same paradigm in Mandarin Chinese, we expected learning under these conditions to be challenging, but possible. Specifically, we predicted that participants would more accurately assign meaning to signs if they had viewed the weather forecast input than if they had not, and that they would be more accurate if they had watched the forecast twice rather than once.

2 Does the nature of the input (frequency, iconicity, transparency, gesture similarity) affect the accuracy of meaning assignment?

We predicted that accuracy would be greater for signs with a higher occurrence frequency. The control group, who had not viewed the forecast, should not show any frequency effects. Demonstrating a lack of frequency effect for this group would confirm that there were no inherent differences in the learnability of the low and high frequency sign sets, and thus lend further support for our claim that any frequency effect found in our experimental groups indicates learning.

We also predicted that visual motivation of target sign forms, as conceptualized and measured in three ways (iconicity, transparency, and gesture similarity), would affect the accuracy of meaning assignment. Transparency was expected to yield the strongest effect given its close conceptual link with the processes involved in the meaning assignment task. In contrast to the frequency effect, we predicted that all participants, whether or not they had viewed the forecast, would be susceptible to the effects of visual motivation.

Finally, we considered that relying on visual motivation might not always lead to meaning being assigned correctly. Sign forms might bring to mind various meanings, not all of which are congruent with the sign's actual meaning.

Method

All materials are available on our Open Science Framework OSF site https://osf.io/rygh7/. In the case of materials involving videos, the materials cannot be viewed online but will need to be downloaded. The current study comprised four experimental tasks (meaning assignment, iconicity, transparency, and gesture similarity). The meaning assignment task assessed participants' ability to infer and learn novel sign meanings after viewing the signed weather report. The other tasks tapped into three different facets of visually motived mapping processes, and their aim was to provide variables that could be used to investigate item-level differences in performance on the meaning assignment task. Each experiment was granted ethical approval, including for online delivery, by the two institutions where data were collected, in southeast England. Participants gave informed consent before completing the tasks.

Table 1 Participant characteristics in the three exposure groups

Exposure group	Statistic	0x ($n = 42$)	1x ($n = 40$)	2x ($n = 42$)	p
Age (years)	*M*	25.93	26.98	27.07	.57
	SD	5.74	6.13	4.38	
Education (years)	*M*	17.95	16.91	17.86	.19
	SD	2.89	2.78	2.85	
English vocabulary knowledge	*M*	38.76	37.48	37.71	.72
(WAIS score 0–52)	*SD*	6.71	8.44	7.90	
Number of known languages	*M*	3.00	2.70	2.81	.51
	SD	1.23	1.07	1.23	
Self-rated language-learning aptitude	*M*	3.17	3.23	2.93	.44
(scale 1–7)	*SD*	1.03	0.92	1.31	
Gender split (f|m|other)	Number	34|7|1	30|10	32|10	.59

Note. WAIS = Wechsler Adult Intelligence Scale WAIS-IV (Wechsler et al., 2008).

Each participant (N = 208) took part in only one of the four experiments. The materials were based on the input material for the meaning assignment task (i.e., a weather forecast in STS).

Meaning Assignment Task

Participants

Participants were recruited using the website "Call for participants" (https://www.callforparticipants.com). Information about their demographics and language background was collected via a questionnaire administered online, based on the Language History Questionnaire 2.0 (Li et al., 2014) but with a bespoke set of questions tailored to our specific requirements. For instance, participants were asked about any prior exposure to sign languages, Makaton (a communication method that uses signs with speech and symbols), fingerspelling, or Swedish, because existing skills in these areas were exclusion criteria. As shown in Table 1, participants were aged between 18 and 40. The upper age limit was introduced to avoid confounds from age-related decline of vision, given the visual nature of the linguistic input. All were native speakers of English resident in the UK with no prior knowledge of Swedish or of any sign languages.

The participants in the meaning assignment task (N = 124) were assigned randomly to the 1x ($n = 40$), 2x ($n = 42$), or 0x ($n = 42$) exposure groups. The single exposure group (1x) watched the forecast once and the double exposure group (2x) twice. The performance of these groups was contrasted with that of the control group (0x), who undertook the meaning assignment task without having seen the forecast. Sample sizes were designed to be double those of Gullberg et al. (2010, 2012). Groups were matched for age, education, first language (i.e., English) vocabulary knowledge, number of known languages, self-rated language aptitude, and gender. Between-subject ANOVAs revealed no significant group differences for these variables (Table 1).

Materials

Weather Forecast. We created a sign language version of a weather forecast, modelled on materials used by Gullberg et al. (2010, 2012). The aim was to create as natural, engaging and professional a forecast as possible. We chose STS to control for unintended effects of participants lipreading the mouthings which frequently accompany signs; if we had used BSL, mouthings might have assisted our English participants to work out the meaning of signs. The forecast lasted four minutes and included 22 target signs on which participants were tested. Crucially, the 22 target signs differed in their occurrence frequency in the forecast: 11 "high frequency" target signs (eight occurrences) versus 11 "low frequency" target signs (three occurrences; with the exception of one item, SÖDER 'south', which occurred four times in error). The high and low frequency sets were matched for aspects of sign language phonology (e.g., locations and hand configurations) and for the number of one-handed versus two-handed signs. They were also matched for iconicity on the basis of the last author's judgment, subsequently confirmed by ratings from an independent group of 24 sign-naïve raters (see the section on the iconicity rating task for details): the ratings revealed that high ($M = 3.64$, $SD = 1.55$) and low frequency ($M = 3.68$, $SD = 1.76$) signs did not differ in their level of perceived iconicity, $F(1, 22) = 0.003$, $p = .96$, $\eta^2 = 0.000$. In addition, high and low frequency items were matched for the occurrence frequency of their English translation equivalents in the CELEX corpus (Baayen et al., 1995): low: $M = 32{,}759$, $SD = 51{,}978$; high: $M = 27{,}027$, $SD = 22{,}771$, $F(1, 22) = 0.11$, $p = .74$, $\eta^2 = 0.006$.

Meaning Assignment Task. All participants undertook the meaning assignment task. This was created and administered using the online Google Docs survey platform. They were presented with short videos of the 22 target signs they had seen in the weather forecast and asked to guess their meaning. The question format was open-ended, but participants were asked to provide only one guess by typing it into a text box below the video. The items were presented in two different pseudorandom orders: half the participants viewed the signs in order A, the other half in order B. There was no significant effect of order on accuracy rates (Mann Whitney $U = 2218$; $p = .24$). The 22 target items were preceded by two practice items. Importantly, participants were instructed to watch each video only once, in order to ascertain a high level of intuitiveness of responses.

English Vocabulary Knowledge. We considered it important that the three participant groups were matched for their first language proficiency, in case this variable affected their performance on the meaning assignment task. Given that expressive vocabulary is a reliable indicator of language proficiency, we administered the English vocabulary test from the Wechsler Adult Intelligence Scale WAIS-IV (Wechsler et al., 2008; not made available in the OSF project folder because it is proprietary). Participants were presented with 26 words both aurally and visually and asked to provide a definition for each. Responses were audio-recorded, and subsequently transcribed and scored based on the test manual. To ensure scoring was reliable, responses from a subset of 10 participants were scored by two independent judges (first and last authors). This process resulted in an interrater correlation score of Spearman's $Rho = .85, p = .002$.

Procedure

Data collection took place during the COVID-19 pandemic in 2020–2021 and was conducted in a quasi-face-to-face style using the Microsoft Teams software. The experimenter (the first author) met with each participant individually on Teams and observed the participant completing the tasks. The exposure group participants first watched a short video of a weather forecast in STS. They were provided with a link to the forecast, and watched the video with their cameras on, so the experimenter could see they were concentrating on the task. Our aim was to replicate an implicit learning context, so instructions relating to the input were kept to a minimum. Participants in the single exposure group were told to watch the signer as she signed the forecast. Participants in the double exposure group were told that they would be watching the film twice back-to-back. The viewings were followed by the surprise meaning assignment task, in which they had to provide the meaning of 22 signs from the forecast. The control group completed only the meaning assignment task, having been told that all signs in the task had originally occurred in a weather forecast presented in Sweden in STS. They were thus given a semantic clue to put them on a par with the exposure groups in terms of narrowing down the scope of their semantic search field when assigning meaning.

After the meaning assignment task, participants completed the demographic and language background questionnaire on SurveyMonkey and the vocabulary test.

The accuracy of all responses in the meaning assignment task was scored by the first author and by a second, trained, scorer, and the interrater reliability suggested high levels of scoring consistency (Spearman's $Rho = .94$, $p < .001$). Responses were scored as accurate when they were semantically identical to the target item. The spreadsheet labelled "correct responses" at https://osf.io/rygh7/ lists the responses coded as accurate and inaccurate for each of the target items.

Iconicity Task

Participants

We recruited 24 new participants (19 females) for the iconicity task. All had English as their first language and were resident in the UK. All were students (from bachelor level to PhD level) from a university in southeast England. Their mean age was 20 years ($SD = 3$, range 18–32). None reported knowing any sign language, which was an important consideration because iconicity ratings are known to differ between sign-naïve and fluent sign language users (see Motamedi et al., 2019).

Materials and Procedure

The design of the iconicity rating task was based on recommendations by Motamedi et al. (2019). Participants were shown individual target signs and, simultaneously, the written English translation. On a response sheet, they circled a number on a scale from 1 to 7 to indicate how iconic they considered each sign to be.

Stimuli were presented using Microsoft PowerPoint. Two pseudorandomized orders were created, and 12 participants did Order 1 and the other 12 did Order 2. Before the target signs were presented, they saw four practice signs (unrelated to weather), in order to familiarize them with the task. Two filler signs occurred at the end of the list to avoid the last target signs being rushed due to their positioning.

All participants were tested face-to-face, individually, in the lab in 2019–2020, before the COVID-19 pandemic.

Transparency Task

Participants

We recruited 30 new participants (21 females) for this task. Again, all had English as their first language and were resident in the UK. The majority ($n = 24$) were students (from bachelor level to PhD level) from a university in southeast England. Others were recruited by opportunity sampling through the researchers' contacts. Their mean age was 23 years ($SD = 4$, range 18–39). None reported any knowledge of a sign language.

Materials and Procedure

Sevcikova Sehyr and Emmorey (2019) found that nonsigners attributed correct meaning to signs only approximately 10% of the time. However, in their task they provided no context for the signs. We wanted to ensure that participants had the same context as those in the meaning assignment task, so we asked them to read an English-language transcript of the STS weather forecast before undertaking the transparency task. The frequency of the target signs in the STS forecast was maintained in the English version.

Stimuli were presented via a PowerPoint presentation. Two pseudorandomized orders were created, and participants were assigned randomly to Order 1 or Order 2 (15 for each order). Participants had a sheet with written instructions and space to record their responses.

Twenty-two participants were tested face-to-face, but following the temporary closure of the university's buildings for part of 2020, the remaining eight were tested online using Microsoft Teams.

Response accuracy was coded by the second author, with all coding checked by the last author. Any disagreements were discussed until agreement was reached. For each item, a transparency score was created, which was a proportion score computed as the number of correct responses divided by the total number of responses for that item.

Gesture Similarity Task

Participants

We recruited 30 new participants (21 females). All had English as their first language and were resident in the UK. The majority ($n = 29$) were students (from Bachelor level to PhD level) from two universities in southeast England, while the other was a staff member. Their mean age was 20 years ($SD = 3$, range 18–30). None reported any knowledge of a sign language.

Materials and Procedure

Participants were presented with the auditory and written English forms of the target words. Two pseudorandomized orders were created, and 14 participants were presented with Order 1 and 16 with Order 2. Participants were asked to silently produce a manual gesture for each word. Note that silent gestures must be distinguished from the co-speech gestures that hearing speakers spontaneously produce with speech. Instead, these are "hand movements that communicate information to another person while consciously avoiding the use of speech" (Ortega & Özyürek, 2020, p. 52) or "spontaneous signs" (Goldin-Meadow & Brentari, 2017, p. 9).

The testing session was video-recorded for later coding. Silent gestures were coded with respect to three aspects of their similarity to their corresponding STS signs:

Phonological Similarity. Gestures were coded for whether they had full, partial or no phonological overlap with the STS sign (as per Ortega et al., 2019) at the level of handshape, orientation, location and movement. Full overlap scored 2 points, partial overlap 1 point, no overlap 0 points.

Representational Similarity. Gestures were coded for whether they represented the referent using the acting, representing, drawing/moulding, or personifying modes of representation (as per Ortega & Özyürek's 2020 adaptation of Müller's 2014 classification, whereby "acting" means the body represents itself and depicts intransitive actions as well as how objects are manipulated, "representing" means the hand(s) adopt(s) the form of the referent, "drawing/moulding" means the hands outline or shape the referent, and "personifying" means that the body serves as a map for a comparable nonhuman body). If the gesture and STS sign drew on the same mode of representation, this scored 1; if they drew on different modes of representation, this scored 0.

Source of Iconicity. Gestures were coded for whether they drew on the same source of iconicity as the STS sign. For example, if the gesture for "snow" drew on the light downwards movement of snow, then the source of iconicity was coded as the same (a score of 1), but if, for example, it drew on the shape of snowflakes it was coded as different (a score of 0).

An overall gesture similarity score was then calculated for each item, by summing the three types of scores (maximum score = 4). The scores for each item were averaged over the number of participants—which was usually 30, but occasionally a participant missed a gesture, in which case the denominator was adjusted accordingly.

Fourteen participants were tested face-to-face, but the remaining 16 were tested online. Coding of the entire dataset was carried out by the last author. Gestures from four of the 30 participants were independently coded by the second author in order to calculate interrater reliability. An intraclass correlation to measure consistency of coding for the overall score using a two-way random effects model (where both people effects and measures effects are random)

Table 2 Accuracy rates by exposure group

Exposure group		0x ($n = 42$)	1x ($n = 40$)	2x ($n = 42$)
Accuracy rate (%)	*M*	9.20	13.86	14.07
	SD	6.39	9.15	9.67
	95% CI	[7.21, 11.19]	[10.39, 16.79]	[11.06, 17.08]

revealed excellent interrater consistency (Cicchetti, 1994), intraclass correlation = .824 (95% $CI = .743–.881$), $p < .001$.

Analyses

In all analyses, we used the lmer.test package in *R* (Kuznetsova et al., 2017), which allows for the use of mixed models and provides the results of significance testing automatically in the form of a *p*-value. When taking random effects into consideration, we assumed a maximally conservative approach, allowing both items and subjects to vary by both intercept and slope (Winter, 2020). The alpha level was set at .05. The summary tables of our results are presented using the style adopted by Ortega et al. (2019). The full data set, analysis scripts, as well as the raw model outputs are available at https://osf.io/rygh7/.

Results

Research Question 1: Meaning Assignment as a Function of Exposure

As can be seen from Table 2, accuracy rates for the meaning assignment task were low in all three groups, suggesting that participants found the task challenging. Our first analysis assessed the impact of exposure group (0x, 1x, 2x) on accuracy. To take into consideration random effects of subjects and items, we conducted linear mixed effects models as per above. Binary variables were centred using sum-coding by assigning the values -1 and $+1$, as suggested by Winter (2020). Due to the limitations associated with modelling independent variables with three levels in R, the comparison was conducted in two steps, again as per Winter (2020). First, we compared accuracy in the control group to accuracy in the two exposure groups, after which we compared the accuracy levels between the two exposure groups (1x, 2x). To investigate the effects of exposure group on the binary variable "accuracy" (correct, incorrect), we conducted two separate glmer models whose outputs are presented in Tables 3 and 4. As illustrated in Figure 1, the two experimental groups performed significantly more accurately on meaning assignment than the control group. This suggested that meaning assignment based on learning generated higher

Table 3 Model output glmer comparison Exposure group 0x versus Exposure groups 1x and 2x

Accuracy predictors	*β*	*SE*	95% CI	*Z*	*p*
Intercept	−3.49	0.45	[−4.37, −2.61]	−7.76	< .001
Exposure 0x vs. 1x/2x	0.57	0.19	[0.20, 0.94]	2.93	< .001

Note. Glmer (accuracy ∼ Exposure0x1x2x + (1+1|subject) + (1+1|item), data, family="binomial")

Table 4 Model output glmer comparison Exposure group 1x versus Exposure group 2x

Accuracy predictors	*β*	*SE*	95% CI	*Z*	*p*
Intercept	−2.97	0.47	[−3.89, 2.05]	−6.36	< .001
Exposure 1x vs. 2x	0.0037	0.23	[−0.45, 0.45]	0.02	.99

Note. Glmer (accuracy ∼ Exposure1x2x + (1+1|subject) + (1+1|item), data, family="binomial")

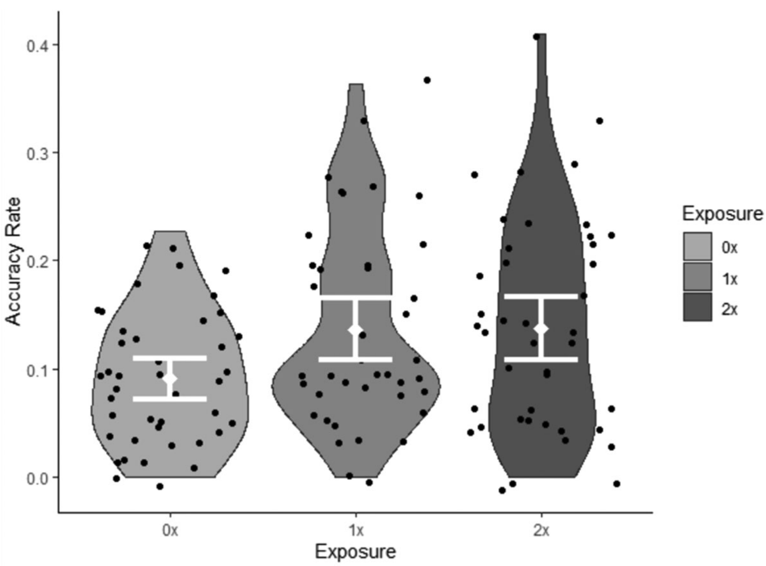

Figure 1 Accuracy of meaning assignment across exposure groups. Error bars indicate 95%CI.

Table 5 Predictors of Accuracy in the 0x Exposure group

Accuracy predictors	β	*SE*	95% CI	*Z*	*p*
Intercept	−4.87	0.73	[−6.30, 3.44]	−6.67	<.001
Frequency	0.39	0.37	[−0.34, 1.12]	1.04	.30
Transparency	5.87	1.11	[3.69, 8.05]	5.26	<.001
Iconicity	0.22	0.23	[−0.23, 0.67]	0.97	.33
Gesture similarity	−0.20	0.30	[−0.79, 0.39]	−0.66	.51

Note. Glmer(accuracy0x∼frequency+transparency+iconicity+gesturesim+(1+1|subject)+(1+1|item),data,family="binomial")

accuracy rates than meaning assignment based on inferencing. However, the double exposure group did not outperform the single exposure group, indicating that doubling the amount of exposure to the input materials did not increase accuracy further.

Research Question 2: The Effects of Input-Related Factors on Meaning Assignment Accuracy

We started by conducting correlational analyses for the three variables assessing visually motivated facets of form–meaning mappings (i.e., iconicity, transparency, and gesture similarity) because we expected these to be correlated. There were positive correlations between iconicity and transparency ($r = .58$, $p = .005$) and between iconicity and gesture similarity ($r = .72, p < .001$) but no relationship between transparency and gesture similarity ($r = .18, p = .41$). Given our theoretical reasons for exploring these three facets individually, as well as the lack of correlation between transparency and gesture similarity, we kept the three sets of scores separate in subsequent analyses.

To assess the effects of input-related factors on the accuracy of meaning assignment, the following variables were entered into the models as predictors: occurrence frequency in the forecast (3x versus 8x), iconicity scores, transparency scores, and gesture similarity scores. Two separate glmer models were conducted. The first model assessed the predictors of meaning assignment accuracy in the control group (see Table 5); the second assessed the predictors of meaning assignment in the two experimental groups (see Table 6). In the control group, accuracy was only predicted by transparency: meaning was assigned more accurately to more transparent items. Crucially, frequency did not predict accuracy in the control group, confirming that performance for the high and low frequency items was identical in a group that had not seen the forecast. In contrast, in the single and double exposure groups, frequency did have

Table 6 Predictors of Accuracy in the 1x and 2x Exposure groups

Accuracy predictors	*β*	*SE*	95% CI	*Z*	*p*
Intercept	−5.57	0.71	[−6.96, 4.18]	−7.91	< .001
Frequency	0.88	0.40	[0.10, 1.66]	2.21	.03
Transparency	5.01	1.20	[2.66, 7.36]	4.16	< .001
Iconicity	0.62	0.26	[0.11, 1.13]	2.35	.02
Gesture similarity	−0.46	0.34	[−1.13, 0.21]	−1.37	.17

Note. Glmer(accuracy1x2x∼frequency+transparency+iconicity+gesturesim+(1+1|subject)+(1+1|item),data,family="binomial")

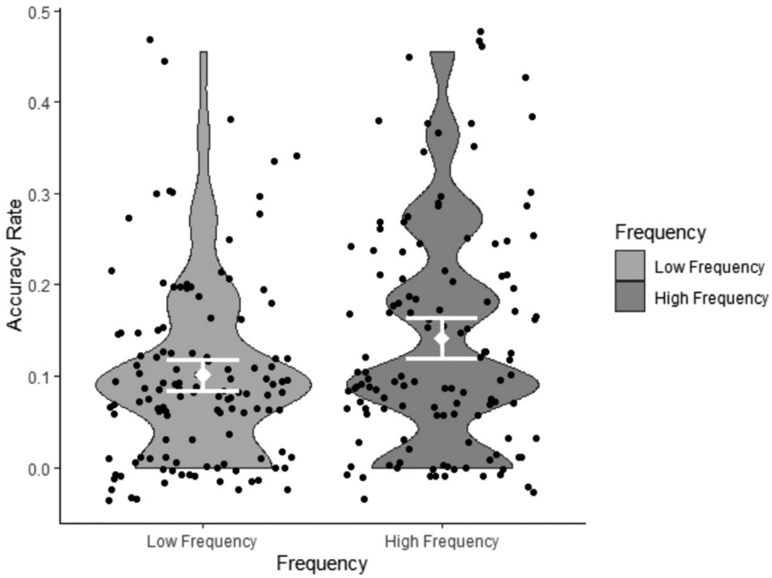

Figure 2 Accuracy rates by item frequency in the combined 1x and 2x exposure groups, summarized by subject. Error bars indicate 95%CI.

a positive effect on accuracy (see Figure 2). The strongest predictor of accuracy was transparency, as illustrated in Figure 3, but, in contrast to the control group, iconicity also predicted accuracy.

The mean accuracy rate of 12.38% across the three groups indicated that the proportion of accurate responses was low. An analysis focused purely on accuracy would therefore miss the richness and variety of the majority of responses, i.e., the inaccurate responses. To gain a more detailed picture of

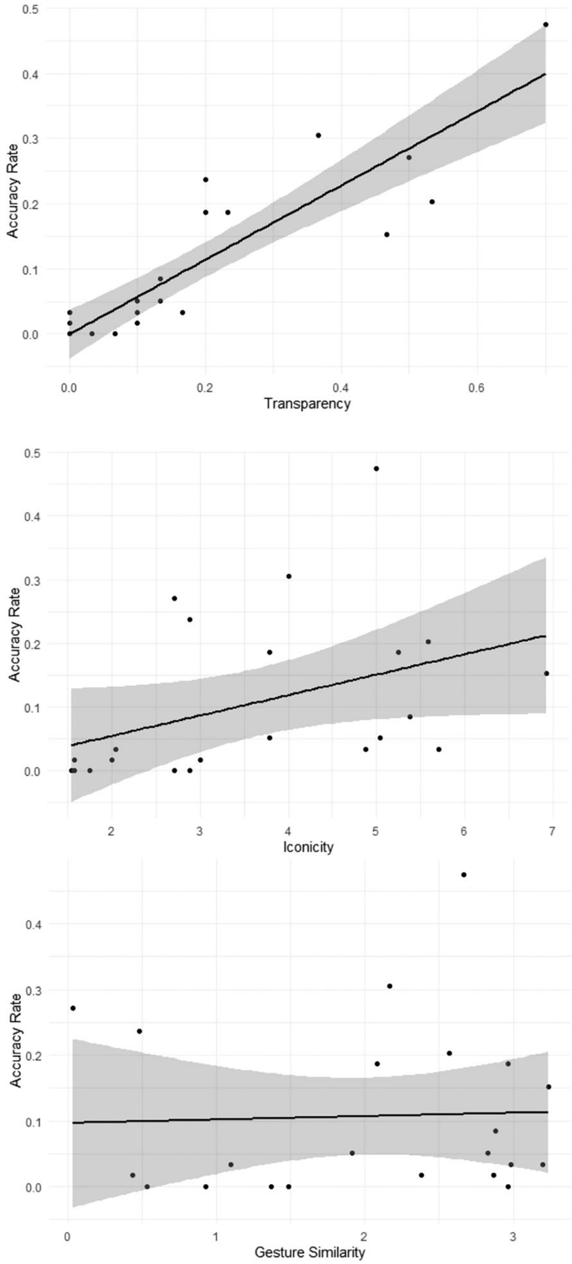

Figure 3 Predictors of accuracy relating to visually motivated meaning assignment in the combined 1x and 2x exposure groups, summarized by item.

how participants assigned meaning to signs in both accurate and inaccurate responses, we conducted qualitative analyses across the whole dataset. The aim was to better understand the nature of visually motivated meaning assignment processes—in other words, the way in which participants linked visually salient properties of signs to their existing world knowledge for the purpose of assigning meaning. Initial exploratory analyses revealed that participants frequently drew upon visual aspects of signs when constructing sign meaning, but that doing so did not always result in accurate responses: in some cases, visual motivation inspired incorrect responses that were semantically far removed from the target. In other words, participants picked up on a visually salient aspect of the sign that was unrelated to its actual STS meaning, which led them astray rather than getting them closer to the target response. For example, the STS sign for MOUNTAIN involves hand movements resembling the rubbing together of two surfaces, which does not directly infer the notion of mountain. This sign is actually a homonym of the sign for STONE, whose visual motivation is the rubbing of two stones together to make fire (K. Schönström, personal communication, August 10, 2020). The form of the sign led several participants to assign it the incorrect meaning "washing board," based on the visual analogy to the rubbing movement.

To investigate this observation further, we placed participants' responses from the meaning assignment task into three categories:

Successful visually motivated meaning assignment: responses motivated by a visually salient aspect of the sign form, which facilitated responses that were semantically related to the target meaning (e.g., the response "rain" for the sign SNOW, which depicts the shared visuo-semantic aspect "precipitation" by presenting both hands open and moving downwards with the fingers fluttering). In other words, this category includes both accurate and near-accurate responses.

Unsuccessful visually motivated meaning assignment: responses motivated by a visually salient aspect of the sign form that is semantically unrelated to its meaning, resulting in an inaccurate response (e.g., the response "washing board" for the sign MOUNTAIN, which visually evokes the rubbing together of two surfaces).

No visually motivated meaning assignment: responses, in which iconicity did not appear to influence the participant's response, such as seemingly random responses or weather-related responses that were visually and semantically unrelated to the target (e.g., the response "drizzle" for the sign TODAY, which does not involve any visual properties evoking "precipitation").

To reduce the risk of biases and subjectivity, participants' responses were coded independently by three raters (the first, second and fourth authors).

Table 7 Visual motivation in meaning assignment

Visually motivated meaning assignment		k	%
Response type	Successful	643	24
	Unsuccessful	1582	58
	None	503	18
Total		2728	100

Interrater reliability was calculated using methods proposed for capturing agreement between multiple raters about categorical data coding (Fleiss, 1971; Landis & Koch, 1977). The analyses revealed that the raters converged in their categorizations to a "moderate" degree (*Fleiss' kappa* $= .48$, $p < .001$). Although raters diverged in their categorization of individual responses, the overall pattern of results was consistent. In this section, we present the results based on the codings of just one rater, the first author. (A full list of codings by each rater is provided here: https://osf.io/rygh7/)

The categorization revealed that the majority of responses (over 80%) were visually motivated (Table 7). Only 18% of guesses displayed no apparent visual motivation. This suggests that participants drew heavily on visual motivation as a resource for meaning assignment. However, this strategy did not always yield correct responses. In fact, 58% of responses represented cases of visually motivated *un*successful meaning assignment. Many visually motivated yet unsuccessful responses involved non-concrete signs that would involve metaphorical derivations to provide a correct response, such as space-time abstractions.

Discussion

This study investigated whether sign-naïve adults are able to infer and learn the meaning of lexical items in an implicit learning context that involves naturalistic input not only in a new language but also in a new modality. A further contribution was that we investigated features of the input that might predict successful learning. In particular, we took a more systematic approach to investigating the potential impact of iconicity (i.e., visual motivation of form–meaning mappings) compared to previous studies of either signed or spoken language learning. Our results suggest that, although accuracy was low, learners were able to derive some information about signs' meaning from the input (i.e., the STS weather forecast), and that they used visual motivation as a resource.

Our first research question asked whether sign-naïve adults are indeed able to infer and learn the correct meaning of individual signs in this challenging learning situation. Our experimental between-subjects exposure manipulation indicated that they can: Meaning assignment was more accurate for participants who had viewed the weather forecast video than for those who had not. This suggests that although participants who had not viewed the forecast could achieve a certain level of success on the task solely by inferring meaning, those who had viewed the forecast were going beyond inference: They had been able to learn something about form–meaning mappings, too. With Gullberg et al. (2010, 2012) as a precedent, it was expected that learning under these conditions—while possible—would be challenging. Accuracy rates in our study were indeed low, at just 14% for each of the two exposure conditions. However, seeing the forecast twice did not lead to greater accuracy than watching it just once. We cannot directly compare the accuracy rates between our task and Gullberg et al.'s because theirs was a forced-choice task whereby participants were required to make a judgment as to whether form–meaning mappings were correct or not. As a result, scores of 50% in that task represented chance performance. However, both our findings—low accuracy rates, and the lack of a boost from double exposure—were reported by Gullberg et al. (2012).

What our findings do show is that adults are able to learn form–meaning mappings on first exposure to a language, even when it occurs in a novel modality. Our findings also contribute to ongoing discussions in the field of second language acquisition about implicit learning at first exposure, which have hitherto been limited to spoken and written language (see Hofweber et al., 2022, for an overview).

Our second question asked whether and how the nature of the input affected meaning assignment. Here, too, we have evidence that those who viewed the forecast were able to learn from it. One of our input manipulations concerned frequency. Half the target signs appeared eight times in the forecast ("high frequency signs") and half appeared just three times ("low frequency signs"). For participants in the control group, there were no significant differences in accuracy on these two sets of signs, confirming that there was nothing inherently "easier" about one set compared to the other when it came to inferring meaning. Therefore, the advantage for the high frequency set seen in the two experimental groups was driven by learning those signs that had been seen more frequently had been learned better. A frequency effect was also reported by Gullberg et al. (2012). The finding of a frequency effect in both studies does make the lack of exposure boost (i.e., the finding of no greater accuracy in the double exposure group compared to the single exposure group) puzzling. Gullberg et al. (2012) concluded that eight instances of a lexical item are sufficient to allow a learner to map meaning to a new word in an unknown language. Our results support this conclusion. Nevertheless, one might have predicted that those high frequency signs seen a total of 16 times by participants in the double exposure group would have achieved an even higher level of accuracy, contrary to fact. These findings are consistent with Hofweber et al.'s (2022) study of sign recognition which used the same STS weather forecast and also found no advantage for the double exposure group. We speculate that in both studies the lack of advantage for double exposure could be due either a ceiling effect or to a habituation effect (whereby participants paid less attention to the input materials when watching them for a second time).

Now to our particular interest regarding the nature of the input, which concerns the visual motivation of form–meaning mappings. Specifically, we investigated whether the measures we had obtained from separate groups of participants for three facets of visual motivation—namely, iconicity (Motamedi et al., 2019), transparency (Sevcikova Sehyr & Emmorey, 2019), and gesture similarity (Ortega et al., 2019)—predicted successful inference and learning of meanings at the item level. We found that, of those three, transparency was the most influential. It predicted accuracy in the control group and the two exposure groups, whereas iconicity was a (weaker) predictor of performance only in the exposure groups. In contrast, gesture similarity played no role at all. The role for transparency in the control group is not surprising because an item's transparency is by its very nature a mark of how easily its meaning is inferred. However, the finding that iconicity was *not* a significant predictor for this group, independently of transparency, supports our claim that considering only iconicity is not sufficient to understand the role of visually motivated forms in learning, and it is relevant to consider the different facets of visually motivated form–meaning mappings. Furthermore, the finding that iconicity was a predictor of performance in the exposure groups suggests that some degree of learning from the input materials was necessary for iconicity to play a role. The effect of iconicity is consistent with the findings of Hofweber et al. (2022), where sign recognition was more accurate for highly iconic items. Those authors speculated that participants were endeavouring to construct meaning as they viewed the STS forecast, even though meaning per se was not tested by the recognition task. The results of the current study—which directly investigated meaning—lend support to that interpretation.

In contrast, the similarity between our target signs and the gestural forms that sign-naïve adults create when asked to silently gesture the relevant meanings (i.e., our gesture similarity score) was not relevant either to inferring or to learning sign meanings. Contrary to predictions derived from Ortega et al. (2019), we did not observe “manual cognate” effects from gesture. This suggests that learners might not be helped by their gestural inventory when it comes to assigning meaning to newly encountered signs. However, it is also possible that the absence of an effect of gesture cognates is due to the specific semantic context we investigated (i.e., the weather forecast), where target items appear to lend themselves less easily to the acting mode of representation that predominates in Ortega et al.’s (2019) dataset. Instead, the majority of our items elicited representing gestures. Another explanation could be sought in differences between language production and comprehension. Whilst gesture similarity was measured by an elicitation task involving production, the meaning assignment task taps into comprehension processes. Hence, the mapping directions in the two tasks contrasted: from meaning to form in the gesture elicitation task, and from form to meaning in the meaning assignment task. Nevertheless, on the basis of the combined findings from our iconicity, transparency and gestural measures, we suggest that when confronted with sign language input, sign-naïve participants consult their general visual world knowledge directly, rather than making recourse to their L1 gestural repertoire.

The visual motivation for mapping sign forms onto existing knowledge was further explored in qualitative analyses of both the accurate and the inaccurate responses (with the latter making up the majority of responses). The data supported the conclusion that participants’ mapping attempts were strongly driven by visual aspects of the sign form. However, in the majority of cases participants made incorrect mappings, possibly due to the large number of possible form–meaning mappings. This rare success in using visual motivation as a strategy for assigning meaning could partially explain why understanding and learning a sign language is harder than one might expect given the relatively high levels of iconicity in sign languages compared to spoken languages.

Limitations and Future Directions

For practical reasons, we studied the implicit learning of just one sign language, by one L1 group, with one set of input materials. Our results therefore need to be replicated in different sign languages, in adults with different L1s, and with input materials other than a weather forecast, so as to determine the extent to which our results for implicit learning generalize across languages, populations, and contexts. Cross-linguistic comparisons would also be valuable. We deliberately chose participants whose L1 was not Swedish in order to avoid them gaining any support from being able to read the signers’ mouthings, because these often reflect the articulation of words in the spoken language with which the sign language is in community contact (i.e., spoken Swedish for STS; Mesch & Schönström, 2021). If we had carried out the task with Swedish speakers, they might have benefitted from these visual cues on the mouth and assigned meaning to signs more accurately. And yet, this is an important population: For many people who need or choose to learn a sign language, this is precisely the learning situation that they are in—they will be learning, for example, BSL in Britain or ASL in the U.S. and will already be (L1 or L2) speakers of English. How well learners of sign languages are able to use mouthings to map meaning to manual sign forms is an interesting avenue for research. Relatedly, much L2 learning of sign languages takes place by adults (deaf or hearing) who are already fluent in at least one other sign language. It would be interesting to consider the challenges they face when the same sign forms occur in their L1 and L2 with different meanings. The L2 learning of sign languages is rarely studied, and the answers to these questions are still incomplete at best (Marshall et al., 2021; Schönström, 2021).

It would also be useful to compare different types of input materials, including input materials with fewer visual contextual cues. For example, the weather forecast is a particular semantic domain, and the fact that we found no effect of gesture similarity on learning could be due to a limitation of this domain. We are not yet able to conclude that learners never map signs that they encounter onto their own gestural repertoire; for other semantic domains, they might. It would also be beneficial to examine how individual differences in learners' cognitive and language-learning profiles influence their ability to assign meanings to signs, particularly given the considerable individual variation in how well participants were able to learn in this implicit situation.

In the introduction, we argued that our experimental context represents a strong and ecologically valid test of the limits of iconic support for second language learning, and that because there is greater potential for iconicity in the manual modality (Goldin-Meadow & Brentari, 2017), iconicity might be particularly relevant to learning a sign language. We have demonstrated that iconicity *does* support the learning of form–meaning mappings in signs. However, our unpacking of the notion of visually motivated forms enabled a more nuanced picture of iconicity, and we have demonstrated that transparency might be more relevant than iconicity in situations where learners are not provided with the meaning of signs. A pertinent question is whether this is also the case for spoken languages, where iconic form–meaning mappings are predominantly auditorily motivated rather than visually motivated. Is it harder for learners to map meaning to auditory forms compared to visual forms, and is the distinction between iconicity and transparency also relevant in the initial stages of spoken language learning from auditory input? Do learners of spoken languages, like the participants in our study, make many unsuccessful attempts at matching meanings to forms when they draw upon iconicity in the learning process? Furthermore, the spoken language encountered in natural contexts is fundamentally multimodal; learners of spoken languages have the option to draw upon word-external visual cues, such as manual gesture (as revealed by Gullberg et al., 2010, 2012; and more generally by Baills et al., 2019; Kelly et al., 2009; Morett, 2014; So et al., 2012; Tellier, 2008, *inter alia*). How do learners use iconic information across modalities in the service of language learning at first exposure? It remains to be seen whether *auditory* iconicity is involved in implicit, first exposure spoken language learning to the same extent as visual iconicity was in our study. Certainly, studies of intermediate and advanced learners of Japanese indicate that they find it challenging to map meanings to mimetic words (see Iwasaki & Yoshioka, 2019, for a review), which suggests that doing so would also be challenging in first exposure contexts. Studies investigating the latter would help shed further light on just how universal the particular mechanisms of language learning are.

Conclusion

In conclusion, our results show that new learners of a sign language can, after just four minutes of exposure to naturalistic input, learn some form–meaning mappings successfully, particularly if those mappings are visually motivated. However, the learning task is challenging, and whilst relying on visual motivation can be a successful strategy, it can also lead learners astray. The joint findings of the present study and those of Hofweber et al. (2022) indicate that lexcical learning is possible from naturalistic input on first exposure to sign. The discoveries complement work on spoken languages and reveal that adult language-learning mechanisms can operate across modalities.

Final revised version accepted 27 March 2023

Open Research Badges

This article has earned Open Data and Open Materials badges for making publicly available the digitally-shareable data and the components of the research methods needed to reproduce the reported procedure and results. All data and materials that the authors have used and have the right to share are available at https://osf.io/rygh7. All proprietary materials have been precisely identified in the manuscript.

References

Akita, K., & Dingemanse, M. (2019). Ideophones (mimetics, expressives). In *Oxford Research Encyclopedia for Linguistics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.477

Baayen, R., Piepenbrock, R., & Gulikers, L. (1995). CELEX2 LDC96L14. [Data set]. Linguistic Data Consortium.

Baills, F., Suárez-González, N., González-Fuente, S., & Prieto, P. (2019). Observing and producing pitch gestures facilitates the learning of Mandarin Chinese tones and words. *Studies in Second Language Acquisition*, *41*(1), 33–58. https://doi.org/10.1017/S0272263118000074

Caselli, N., & Pyers, J. (2017). The road to language learning is not entirely iconic: Iconicity, neighborhood density, and frequency facilitate acquisition of sign language. *Psychological Science*, *28*, 979–987. https://doi.org/10.1177/0956797617700498

Christiansen, M. (2019). Implicit statistical learning: A tale of two literatures. *Topics in Cognitive Science*, *11*(3), 468–481. https://doi.org/10.1111/tops.12332

Cicchetti, D. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*, 284–290. https://doi.org/10.1037/1040-3590.6.4.284

de Groot, A., & van Hell, J. (2005). The learning of foreign language vocabulary. In J. Kroll & A. De Groot (Eds.), *Handbook of Bilingualism. Psycholinguistic Approaches* (pp. 9–49). Oxford University Press.

Dingemanse, M., Perlman, M., & Perniss, P. (2020). Construals of iconicity: Experimental approaches to form–meaning resemblances in language. *Language and Cognition*, *12*(1), 1–14. https://doi.org/10.1017/langcog.2019.48

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

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

Fleiss, J. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*, 378–382. https://doi.org/10.1037/h0031619

Goldin-Meadow, S., & Brentari, D. (2017). Gesture, sign and language: The coming of age of sign language and gesture studies. *Brain and Behavioral Sciences*, *40*, e46. https://doi.org/10.1017/S0140525x15001247

Gullberg, M., Roberts, L., & Dimroth, C. (2012). What word-level knowledge can adult learners acquire after minimal exposure to a new language? *International*

Review of Applied Linguistics in Language Teaching, *50*, 239–276. https://doi.org/10.1515/iral-2012-0010
Gullberg, M., Roberts, L., Dimroth, C., Veroude, K., & Indefrey, P. (2010). Adult language learning after minimal exposure to an unknown natural language. *Language Learning*, *60*, 5–24. https://doi.org/10.1111/j.1467-9922.2010.00598.x
Hockett, C. (1959). Animal "languages" and human language. *Human Biology*, *31*(1), 32–39.
Hofweber, J., Aumônier, L., Janke, V., Gullberg, M., & Marshall, C. (2022). Breaking into language in a new modality: The role of input and individual differences in recognizing signs. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.895880
Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism between a word and an action facilitates early verb learning. *Cognition*, *109*(1), 54–65. https://doi.org/10.1016/j.cognition.2008.07.015
Iwasaki, N., & Yashioka, 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). John Benjamins Publishing Company.
Kelly, S., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, *24*(2), 313–334. https://doi.org/10.1080/01690960802365567
Klima, E., & Bellugi, U. (1979). *The signs of language*. Harvard University Press.
Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13
Laing, C., Vihman, M., & Keren-Portnoy, T. (2017). How salient are onomatopoeia in the early input? A prosodic analysis of infant-directed speech. *Journal of Child Language*, *44*, 1117–1139. https://doi.org/10.1017/S0305000916000428
Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310
Li, P., Zhang, F., Tsai, E., & Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition*, *17*(3), 673–680. https://doi.org/10.1017/S1366728913000606
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
Marshall, C., Bel, A., Gulamani, S., & Morgan, G. (2021). How are signed languages learned as second languages? *Language and Linguistics Compass*, *15*(1), e12403. https://doi.org/10.1111/lnc3.12403
Mesch, J., & Schönström, K. (2021). Use and acquisition of mouth actions in L2 sign language learners: A corpus-based approach. *Sign Language and Linguistics*, *24*(1), 36–62. https://doi.org/10.1075/sll.19003.mes

Monaghan, P., Schoetensack, C., & Rebuschat, P. (2019). A single paradigm for implicit and statistical learning. *Topics in Cognitive Science*, *11*(3), 536–554. https://doi.org/10.1111/tops.12439

Moore Mauroux, S. (2016). Patterns and variation in the weather forecast: Can prosodic features be predicted too? *Anglophonia*, *21*. https://doi.org/10.4000/anglophonia.755

Morett, L. M. (2014). When hands speak louder than words: The role of gesture in the communication, encoding, and recall of words in a novel second language. *The Modern Language Journal*, *98*(3), 834–853. https://doi.org/10.1111/modl.12125

Motamedi, Y., Little, H., Nielsen, A., & Sulik, J. (2019). The iconicity toolbox: Empirical approaches to measuring iconicity. *Language and Cognition*, *11*(2), 188–207. https://doi.org/10.1017/langcog.2019.14

Motamedi, Y., Murgiano, M., Perniss, P., Wonnacott, L., Marshall, C., Goldin-Meadow, S., & Vigliocco, G. (2021). Linking language to sensory experience: Onomatopoeia in early language development. *Developmental Science*, *24*(3), e13066. https://doi.org/10.1111/desc.13066

Mott, M., Midgley, K., Holcomb, P., & Emmorey, K. (2020). Cross-modal translation priming and iconicity effects in deaf signers and hearing learners of American Sign Language. *Bilingualism: Language and Cognition*, *23*(5), 1032–1044. https://doi.org/10.1017/S1366728919000889

Müller, C. (2014). Gestural modes of representation as techniques of depiction. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & J. Bressem (Eds.), *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction 38.2* (pp. 1687–1702). De Gruyter Mouton.

Nielsen, A., & Dingemanse, M. (2020). Iconicity in word learning and beyond: A critical review. *Language and Speech*, *64*(1), 52–72. https://doi.org/10.1177/0023830920914339

Ortega, G., & Özyürek, A. (2020). Systematic mappings between semantic categories and types of iconic representations in the manual modality: A normed database of silent gesture. *Behaviour Research Methods*, *52*, 51–67. https://doi.org/10.3758/s13428-019-01204-6

Ortega, G., Schiefner, A., & Özyürek, A. (2019). Hearing non-signers use their gestures to predict iconic form–meaning mappings at first exposure to signs. *Cognition*, *191*, 103996. https://doi.org/10.1016/j.cognition.2019.06.008

Ota, M., Davies-Jenkins, N., & Skarabela, B. (2018). Why *choo-choo* is better than *train*: The role of register-specific words in early vocabulary growth. *Cognitive Science*, *42*(6), 1974–1999. https://doi.org/10.1111/cogs.12628

Perniss, P., Thompson, R., & Vigliocco, G. (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

de Saussure, F. (1916). *Cours de linguistique générale* (C. Bally & A. Sechhaye, Eds.). Payot.

Schönström, K. (2021). Sign languages and second language acquisition research: An introduction. *Journal of the European Second Language Association*, *5*, 30–43, https://doi.org/10.22599/jesla.73

Sevcikova Sehyr, Z., & Emmorey, K. (2019). The perceived mapping between form and meaning in American Sign Language depends on linguistic knowledge and task: Evidence from iconicity and transparency judgments. *Language and Cognition*, *11*(2), 208–234. https://doi.org/10.1017/langcog.2019.18

So, W. C., Sim Chen-Hui, C., & Low Wei-Shan, J. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall? *Language and Cognitive Processes*, *27*(5), 665–681. https://doi.org/10.1080/01690965.2011.573220

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, *8*(2), 219–235. https://doi.org/10.1075/gest.8.2.06tel

Thompson, R. L., Vinson, D., Woll, B., & Vigliocco, G. (2012). The road to language learning is iconic: Evidence from British Sign Language. *Psychological Science*, *23*(12), 1443–1448. https://doi.org/10.1177/0956797612459763

Veroude, K., Norris, D., Shumskaya, E., Gullberg, M., & Indefrey, P. (2010). Functional connectivity between brain regions involved in learning words of a new language. *Brain and Language*, *113*(1), 21–27. https://doi.org/10.1016/j.bandl.2009.12.005

Wechsler, D., Coalson, D., & Raiford, S. (2008). *WAIS-IV technical and interpretive manual*. Pearson.

Winter, B. (2020). *Statistics for linguists: An introduction using R*. Routledge.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Accessible Summary