Journal of Memory and Language 150 (2026) 104767

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Iconicity and morphophonological processing in sign recognition: insights from behavioral and pupillary response measures

Philip Thierfelder *, Emily Koenders

Centre for Sign Linguistics and Deaf Studies, Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Hong Kong, China

Introduction

Background

Sign languages display extensive and systematic patterns of iconicity throughout their lexicons. While not every sign is iconic, iconic forms are far more widespread in sign languages than in spoken languages (see Thompson et al., 2020). The visual forms of signs can mirror the visual or functional characteristics of their referents in ways that are often readily detectable—even by non-signers. This pervasive iconicity has attracted substantial psycholinguistic interest, particularly regarding its effects on lexical recognition (recently: Caselli & Pyers, 2020; Emmorey et al., 2020; McGarry, Massa et al., 2021; McGarry, Mott et al., 2021; Zhang et al., 2023). Because current theories of lexical access were developed primarily for spoken languages with largely arbitrary mappings, investigating iconicity effects in sign recognition can reveal how language users integrate non-arbitrary form–meaning correspondences.

However, research on iconicity-related processing effects has yielded inconsistent results: some studies find facilitation (e.g., Thompson et al., 2009; Vinson et al., 2015; Zhang et al., 2023), while others report null, weak, or late-emerging effects (e.g., Bosworth & Emmorey, 2010; Emmorey et al., 2020; Mott et al., 2020). Ortega (2017) argues that iconicity must be operationalized more carefully, because its effects depend not only on degree of iconicity but also on the kind of form–meaning mapping involved and how this interacts with task demands and participant background. Similarly, recent work argues that iconicity is best investigated with complementary measures rather than a single index, since different methods capture different aspects of form–meaning resemblance (e.g., Motamedi et al., 2019; Winter & Perlman, 2021).

Psycholinguistic studies of sign languages typically operationalize iconicity using Likert-scale ratings of perceived iconicity (e.g., Baus et al., 2013; Bosworth & Emmorey, 2010; Caselli et al., 2017; Vinson et al., 2008). Although such ratings capture important aspects of perceived form–meaning resemblance, they may collapse across iconic signs that realize form–meaning mappings in qualitatively distinct ways. For instance, some iconic signs depict how a referent is handled or manipulated, whereas others represent the referent's form or shape (McDonald, 1982; Padden et al., 2010; Supalla, 1986; see also Table A1), with corresponding differences in their underlying morphophonological structure. It therefore remains unclear whether these types of iconic signs exhibit distinct processing profiles, and whether iconicity effects during lexical access are modulated by the morphophonological properties of the sign.

Classifier handshapes provide a useful test case for examining this relationship. These highly productive iconic elements instantiate distinct form–meaning mappings: entity classifiers represent whole referents by their shape or form; body-part classifiers depict specific body parts; and handling classifiers depict how a referent is manipulated or held (McDonald, 1982; Padden et al., 2010; Supalla, 1986; see also Table A1). Importantly, handshape is one of the five[1] phonological parameters that constitute signs (Brentari, 1998; Sandler & Lillo-Martin, 2006), but handshapes can also participate in iconic form–meaning mappings, functioning as "phonomorphemes" (Johnston & Schembri, 2007).[2] Classifier type thus introduces systematic variation in both morphophonological structure and the nature of iconic mapping (Koenders, 2024; Supalla, 1986, 2003). In this paper, we use "morphophonological" to refer to the dual phonological–morphemic status of classifier handshapes, and "morphological" when referring to internal structural complexity and the decomposition processes involved in recognition.

To investigate how morphophonological structure and perceived iconicity jointly affect processing, we examine lexical recognition in Hong Kong Sign Language (HKSL), where classifier handshapes are integrated into a large proportion of lexical signs, including approximately 76% of nominal signs (Koenders, 2024). Using an unprimed lexical decision task, we assess signs varying in classifier type (i.e., those with no

* Corresponding author.
E-mail address: dpthierfelder@cuhk.edu.hk (P. Thierfelder).

[1] The other four being movement, location, orientation, and non-manual markers.

[2] For example, signs articulated at the forehead location often have meanings related to cognition (e.g., THINK, SMART, REMEMBER), illustrating how a phonological parameter (location) can simultaneously carry morpho-semantic information.

https://doi.org/10.1016/j.jml.2026.104767
Received 8 August 2025; Received in revised form 11 May 2026; Accepted 11 May 2026
Available online 19 May 2026
0749-596X/© 2026 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

classifier, and those with entity, body-part, or handling classifier handshapes) and in perceived iconicity (measured via deaf signers' ratings) across three dependent measures: response times, accuracy, and pupillometry—the last of which tracks moment-by-moment processing demand during lexical access (cf. Rojas et al., 2024). We treat classifier type and iconicity ratings as complementary predictors and assess their unique contributions: if classifier type remains predictive after accounting for iconicity ratings, this would suggest that morphophonological structure contributes to processing in ways not fully captured by perceived iconicity, and if iconicity ratings explain additional variance beyond classifier type, this would suggest that perceived form–meaning resemblance contributes independently of morphophonological structure.

Iconicity in sign languages

Iconicity refers to a relationship in which the form of a word resembles or reflects some aspect of its meaning (Meir & Tkachman, 2018). While present in spoken language (e.g., onomatopoeia), iconicity is especially prominent in sign languages, where visible articulators (the hands, face, and torso) make them optimal for representing the visual features of referents (Perniss & Vigliocco, 2014; Taub, 2001).

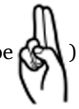
However, in the sign linguistics literature, the term iconicity has been used to refer to a range of interrelated but distinct constructs. Bross (2024) reviews three views: a sign is iconic if its form resembles (1) what it refers to in the world, (2) its linguistic meaning, or (3) a related mental image or concept. He argues against (1), since many pictorial signs lack a specific real-world referent—for example, signs for fictional beings. Option (2) is also argued to be limited because what a sign depicts often draws on background knowledge rather than core meaning—for instance, a sign may depict milking while simply meaning *milk*. Bross argues that definition (3), which states that iconicity is a perceived resemblance between aspects of a conventional sign and the signer's associated mental representation (see also Emmorey, 2014; Perniss et al., 2010), is the only account that consistently explains the full range of iconic signs in the literature, including those whose iconicity depends on historical, cultural, or language-internal knowledge.

Within this mental-representation framework, iconicity can take different forms. Anible (2020) distinguishes two types especially relevant to sign recognition: imagistic and diagrammatic iconicity. Imagistic iconicity involves concrete, pictorial form–meaning mappings that may be apparent even to non-signers. Diagrammatic iconicity involves schematic or relational mappings that are often abstract and salient only to experienced signers familiar with language-internal patterns. While not an exhaustive typology (see Bross, 2024), this contrast between pictorial and schematic mappings is directly relevant for processing research. We illustrate these distinctions with our own examples from HKSL below.

Fig. 1 shows three signs—RABBIT (a), GLASSES (b), and BASKETBALL (c)—that exhibit imagistic iconicity. The relationship between form and meaning in these signs would likely be evident even to most non-signers.

Diagrammatic iconicity can be seen across the signs in Fig. 2. The signs WEIGHT, ELEVATOR, and PERFORMANCE all incorporate the Y handshape , which in HKSL often represents entities that stand on two legs, typically humans (Tang, 2007). In the asymmetrical signs WEIGHT (2a) and ELEVATOR (2b), the non-dominant hand adopts the flat B handshape  to represent a flat surface (a scale or elevator floor, respectively). This illustrates how iconicity is systematically patterned throughout the HKSL lexicon, reflecting structured visual-semantic mappings (cf. Schembri, 2003; Supalla, 1986; Zwitserlood, 2012). While likely not highly apparent to individuals with little or no knowledge of HKSL, these iconic motivations are salient to experienced signers.

Imagistic and diagrammatic iconicity are not mutually exclusive. Signs can embody varying degrees of both. This becomes evident when contrasting the signs in Fig. 1 with diagrammatically related signs that also represent their referents imagistically (Fig. 3). In each pair (RABBIT–SNAKE, GLASSES–BADGE, BASKETBALL–BOWLING), a shared classifier handshape encodes a general semantic category while differences in location and movement distinguish specific referents.

In SNAKE and RABBIT, a body-part classifier (U handshape ) depicts anatomical features, mapping their form to the handshape and their anatomical position to an analogous location on the signer's body. For GLASSES and BADGE, an entity classifier (small C handshape ) captures the referent's shape, with the location on the signer's body reflecting where these items are typically worn. In BASKETBALL and BOWLING, a handling classifier (open 5 handshape ) mirrors the action of grasping a ball, with movement reflecting how the ball is manipulated in each sport.

Classifiers and morphological structure

Classifier handshapes operate as iconic morphemes that systematically encode visual-semantic relationships in sign languages (Schlenker et al., 2024; Supalla, 1986; Zwitserlood, 2012). Although movement, location, orientation, and non-manual markers may also be meaningful (Johnston & Schembri, 2007), classifier handshapes are central because they categorize the referent and constrain how other parameters iconically represent actions, locations, and event relations (Aikhenvald, 2000; Anible, 2020; Koenders, 2024). We follow Koenders' (2024)

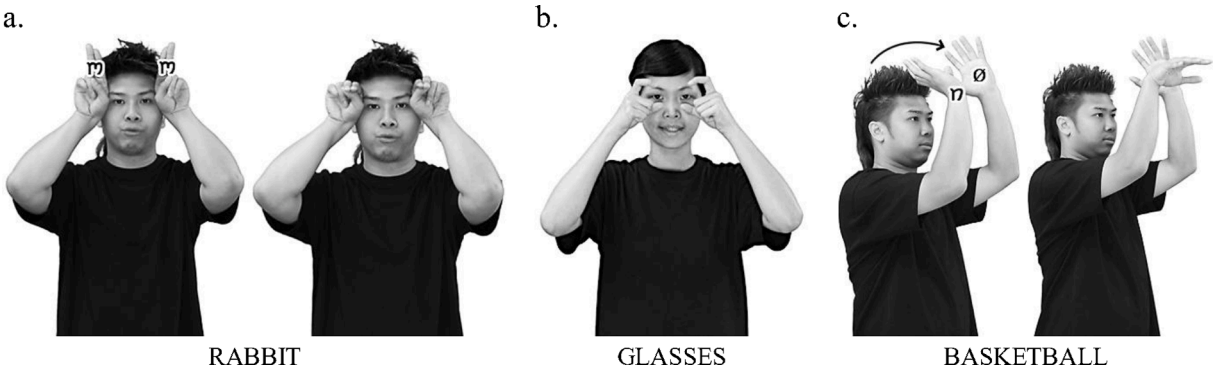


Fig. 1. Examples of signs in HKSL with imagistic iconicity. This figure shows the HKSL signs for RABBIT (a), GLASSES (b), and BASKETBALL (c), which were rated as highly iconic by deaf signers. Each sign shows a direct, pictorial resemblance to its referent (e.g., mimicking shape or action).

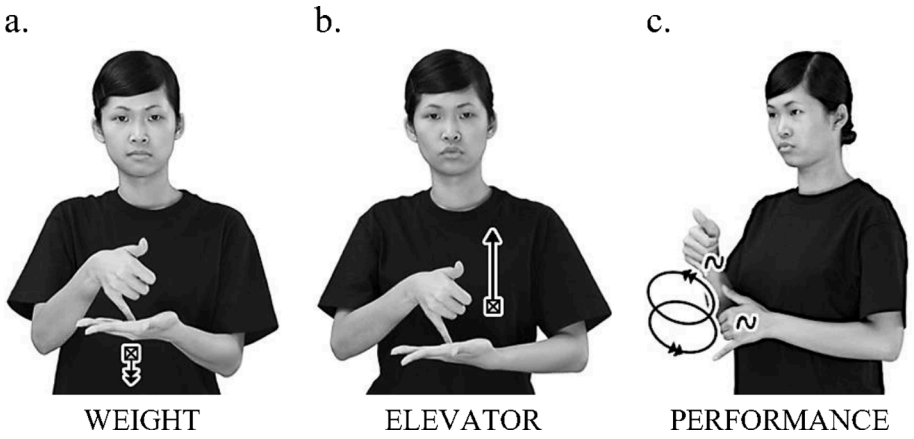


Fig. 2. Examples of iconic signs in HKSL with a diagrammatic relationship. This figure shows three HKSL signs—WEIGHT (a), ELEVATOR (b), and PERFORMANCE (c)—that exhibit diagrammatic iconicity through schematic mappings. In each sign, the Y handshape represents a person, while the other handshape (open B), seen on the left hand in (a) and (b), represents a flat surface.



Fig. 3. Examples of HKSL signs integrating imagistic and diagrammatic iconicity. This figure shows the HKSL signs for SNAKE (a), BADGE (b), and BOWLING (c), each of which pairs with a sign from Fig. 1 through a shared classifier handshape.

typology because it captures three classifier types commonly distinguished across prior accounts: entity classifiers, which depict referent shape or form; body-part classifiers, which depict bodily features; and handling classifiers, which depict how referents are manipulated. Table A1 situates this typology within prior classifier handshape accounts.

Following Zwitserlood (2008), Koenders (2024) analyzes HKSL within a Distributed Morphology framework (Halle, 1990; Halle & Marantz, 1993), treating classifier handshapes as meaningful units within the internal structure of lexical signs rather than as independent lexical items or merely phonological features. Under this analysis, the grammatical category of a sign is determined by structural context, while classifier handshapes contribute meaning by specifying referent type and how that referent is iconically represented.

This structural view motivates the present study because it links classifier type to differences in internal morphological composition. If classifier handshapes are meaningful units within lexical signs, then signs that contain them may differ from signs that do not in the number and type of meaning-bearing components they include. Figs. 4–7 schematize this contrast using representative HKSL signs from the categories considered here. Although presented as DM-style trees, the figures can be read simply as visual summaries of the signs' proposed internal structure. Bracketed labels correspond to proposed meaning-bearing components.

Classifier type therefore provides a theoretically motivated way to index one dimension of morphological structure: whether, and how, classifier handshape contributes morphemic content to the sign. Where classifier handshapes are absent, signs such as OIL (Fig. 4) are analyzed as single indivisible roots whose handshape functions as a purely phonemic unit rather than as a meaningful unit (cf. Embick, 2021).

Among signs that do contain classifier handshapes, SUN (Fig. 5) illustrates an analysis of a lexical sign with an embedded entity classifier handshape, in which the handshape encodes the referent as an orb-like entity and combines with location, movement, and orientation components associated with upper token space and outward shining or spreading. In classifier predicates, entity classifiers are typically associated with the internal argument, usually the theme (Benedicto & Brentari, 2004).

An analysis of a lexical sign with an embedded body-part classifier handshape is illustrated by BIRD (Fig. 6): the classifier handshape represents a body part of the referent and, in classifier predicates, has been analyzed as yielding an intransitive unergative structure whose single argument patterns as an external argument (Benedicto & Brentari, 2004).

Finally, CIGARETTE (Fig. 7) illustrates an analysis of a lexical sign with an embedded handling-classifier handshape, where the handshape depicts how an agent manipulates a theme or instrument. Although CIGARETTE itself is treated here as a lexical sign, handling classifiers in verbal predicates have been analyzed as restricted to transitive structures and as encoding information about both arguments (Benedicto et al., 2007; Glück & Pfau, 1998; Kimmelman et al., 2020; Pavlič, 2016; Zwitserlood, 2003). Together, these examples illustrate the structural contrast central to the present study: signs without classifier handshapes, hereafter no-CL signs, are analyzed as morphologically simpler lexical roots, whereas signs with embedded classifier handshapes are analyzed as having more elaborate internal structure, incorporating meaning

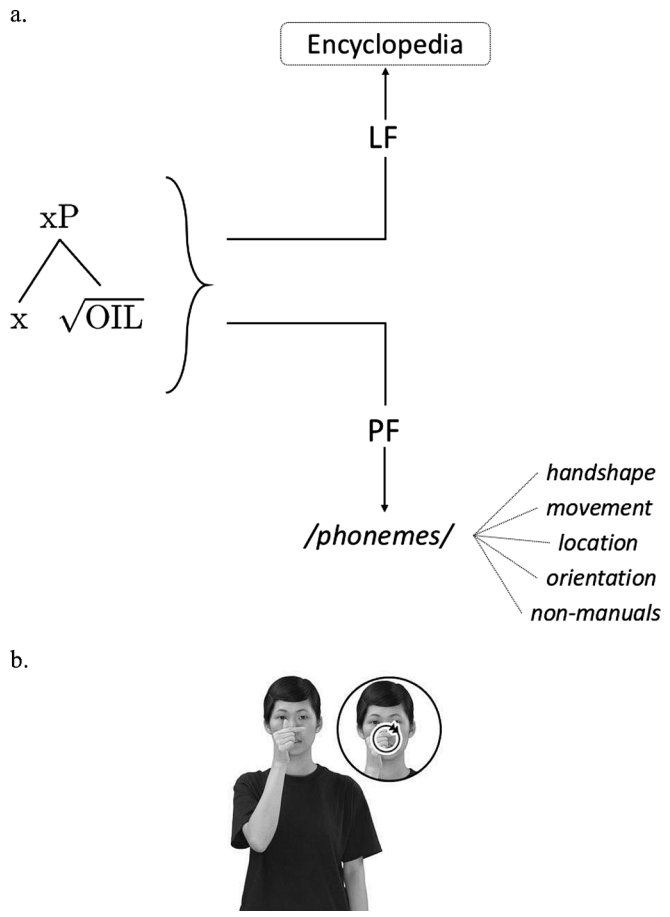


Fig. 4. The root compound $\sqrt{\text{OIL}}$. (a) shows that no featural nodes are merged under $\sqrt{\text{OIL}}$ because none of its phonological parameters carry morphemic content. (b) shows the sign as produced in HKSL.

contributed by classifier handshapes, iconic parameters, and, where relevant, argument-structural information.

Morphological processing across modalities

The decompositional view of morphological structure outlined above is consistent with empirical work on morphological processing across modalities. In spoken and written language, MEG studies show that the brain automatically decomposes complex words into morphemic components at early processing stages, with all allomorphs activating the same root (Solomyak & Marantz, 2010; Stockall & Marantz, 2006). Although decomposition appears automatic, the effort it requires varies with frequency, structure, and context.

Cross-linguistic research shows that morphological complexity can incur processing costs. In agglutinative languages[3] like Finnish, producing or comprehending morphologically complex forms (i.e., words composed of multiple morphemes) results in higher error rates, slower recognition, and increased neural activation (Hedlund et al., 2021; Hyönä et al., 2002; Hyönä & Pollatsek, 2000; Lehtonen et al., 2007). Studies on compound words in Basque and Finnish further reveal that it is not complexity per se, but the frequency of constituent morphemes that modulates processing effort, as shown by both eye-tracking and ERP measures (Bertram et al., 2000; Vergara-Martínez et al., 2009).

[3] Agglutinative languages—from the Latin *agglutinare* "to glue together"—are languages that "glue" morphemes and words together sequentially into a single word. These morphemes and words are combined without greatly altering their forms. Take the Turkish word bırak-ıl-ma-sı release-PASS-SUB-POS.3 "that he be released" which consists of a verb stem and three suffixes (Durrant, 2013).

ERP evidence from English and Korean shows that early form-based and later semantic (N400) components are differentially sensitive to regularity and frequency, with processing costs for complex forms emerging primarily for less frequent items, while high-frequency forms are recognized rapidly and likely accessed holistically (Kim et al., 2022; Rastle et al., 2015).

Sign languages offer a distinctive test case because their phonological parameters are articulated concurrently, even though perceivers extract different parameters on slightly different timescales. Emmorey and Corina (1990) report that location and orientation are accessible earliest, around 150 ms, followed by handshape at about 170 ms, allowing signers to narrow lexical candidates before movement is fully processed; recognition is typically achieved around 240 ms. This helps explain why signers recognize ASL signs after perceiving only about 51% of the sign, compared with a spoken-word recognition threshold of roughly 83% (Grosjean, 1981).

In signs containing classifier handshapes, parameters that become available early may contribute morpheme-like information. If lexical access proceeds decompositionally through sublexical units (Rastle & Davis, 2008; Taft, 2004), multiple such units may need to be accessed and integrated. Given their near-simultaneous availability, this integration may occur at least partly in parallel. Together, these considerations suggest that morphological structure may shape lexical processing in sign languages in ways that resemble, but are not identical to, spoken-language effects.

Iconicity and sign processing

Because classifier handshapes contribute to both morphophonological structure and iconic form–meaning mappings, their processing effects cannot be understood in terms of structure alone. A second relevant dimension is perceived iconicity. Numerous psycholinguistic studies have examined the role of iconicity in sign processing, but their findings are mixed, with effects varying across task demands, signer proficiency, and age of acquisition.

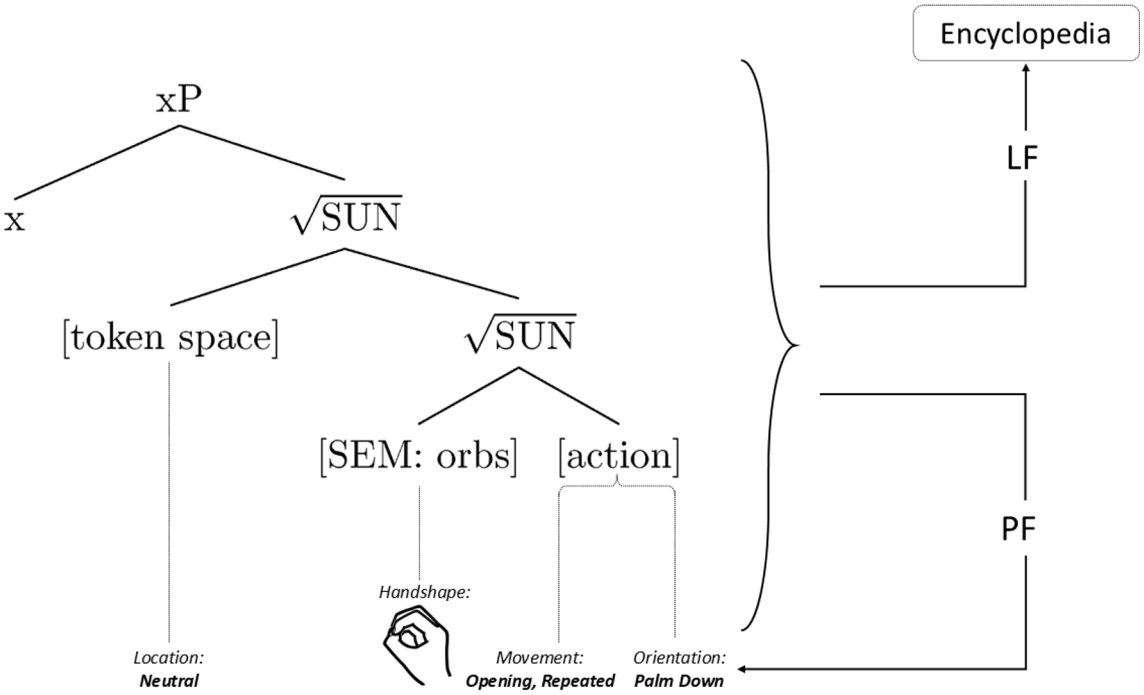
Task-based factors

Iconicity effects appear to depend strongly on task demands. In lexical decision tasks, iconicity has been found to have little facilitative effect: Bosworth and Emmorey (2010) found that semantic relatedness, but not iconicity, facilitated recognition, and Caselli et al. (2021) found that iconicity negatively affected accuracy in deaf signers with deaf parents and had no effect on response times. In contrast, tasks involving visual feature mapping may elicit stronger effects. Thompson et al. (2009) found that native ASL signers responded faster in a picture-matching task when iconic features aligned with the picture, while Vinson et al. (2015) demonstrated iconicity facilitation in BSL across sign–picture matching and picture naming. These findings are consistent with a structure-mapping account of iconicity (Emmorey, 2014; Gentner, 1983; Gentner & Markman, 1997): iconicity effects appear to be most robust when tasks make correspondences between sign form and meaning task-relevant.

Proficiency

Iconicity effects also vary with signer proficiency. Using cross-modal translation priming with ERPs, Mott et al. (2020) found that iconicity modulated translation priming earlier for hearing beginning ASL learners (600–800 ms) than for proficient deaf signers (800–1000 ms), suggesting that beginning learners recruit iconicity earlier during lexical access. Anible (2020) examined diagrammatic iconicity in an ASL–English translation recognition task and found that expert signers treated diagrammatically related distractors as phonological competitors, whereas novices treated them as semantic competitors. Anible suggests that less experienced signers rely more on iconicity during lexical access, whereas expert signers exhibit more form-based processing.

a.



b.

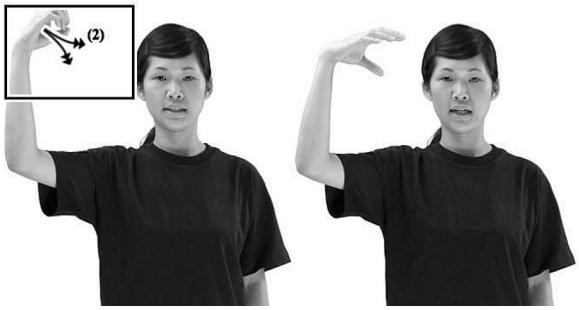


Fig. 5. The root compound $\sqrt{\text{SUN}}$. (a) shows that for the HKSL sign SUN the sign parameters are meaningful because they are associated with morphemic features merged underneath the root compound $\sqrt{\text{SUN}}$. (b) shows how the sign is produced in HKSL.

Age of acquisition

Iconicity effects also appear to vary across signer populations that differ in age of acquisition. Emmorey et al. (2020) found that iconicity exerted only a weak influence on neural activity in native ASL signers, primarily at later processing stages (1000–1200 ms). Using similar methodology with Chinese Sign Language, Zhang et al. (2023) found that highly iconic signs elicited reduced N400 negativities (300–500 ms) and reduced late positive complex amplitudes (500–800 ms), suggesting a processing advantage associated with iconicity. Zhang et al. attribute the discrepancies between their study and Emmorey et al.'s to differences in age of acquisition: their participants were deaf signers first exposed to CSL between ages 6 and 14, whereas Emmorey et al.'s participants were predominantly native signers exposed from birth.

The present study

As discussed above, findings on iconicity and lexical processing in sign languages are mixed: some studies report facilitation in particular tasks (Thompson et al., 2009; Vinson et al., 2015; Zhang et al., 2023), whereas others find only weak facilitatory effects (Emmorey et al., 2020; Mott et al., 2020) or even processing costs (Caselli et al., 2021). This variability may stem from methodological differences (such as task demands), participant background factors (such as language proficiency or age of acquisition), linguistic factors (such as the target sign language), or a combination of these influences.

In addition to these factors, variability may also depend on how iconicity itself is operationalized in a given study. A common approach is to operationalize iconicity using subjective ratings, which quantify perceived form–meaning resemblance on a single scale. These ratings are informative and have clear value for psycholinguistic work, but they do not explicitly represent morphophonological distinctions (such as classifier type) that may covary with perceived iconicity and shape how iconic mappings are realized across the lexicon. For signs with embedded classifier handshapes in particular, considering structural information alongside ratings may provide a more complete characterization of the properties that influence recognition.

As such, we adopt an integrative approach that jointly models perceived iconicity (ratings) and classifier-related structure. Focusing on HKSL, where classifier handshapes are pervasive in lexical signs (Koenders, 2024), we test whether classifier presence and type (no embedded classifier vs. embedded entity, body-part, or handling classifier) predict lexical decision measures beyond iconicity ratings and

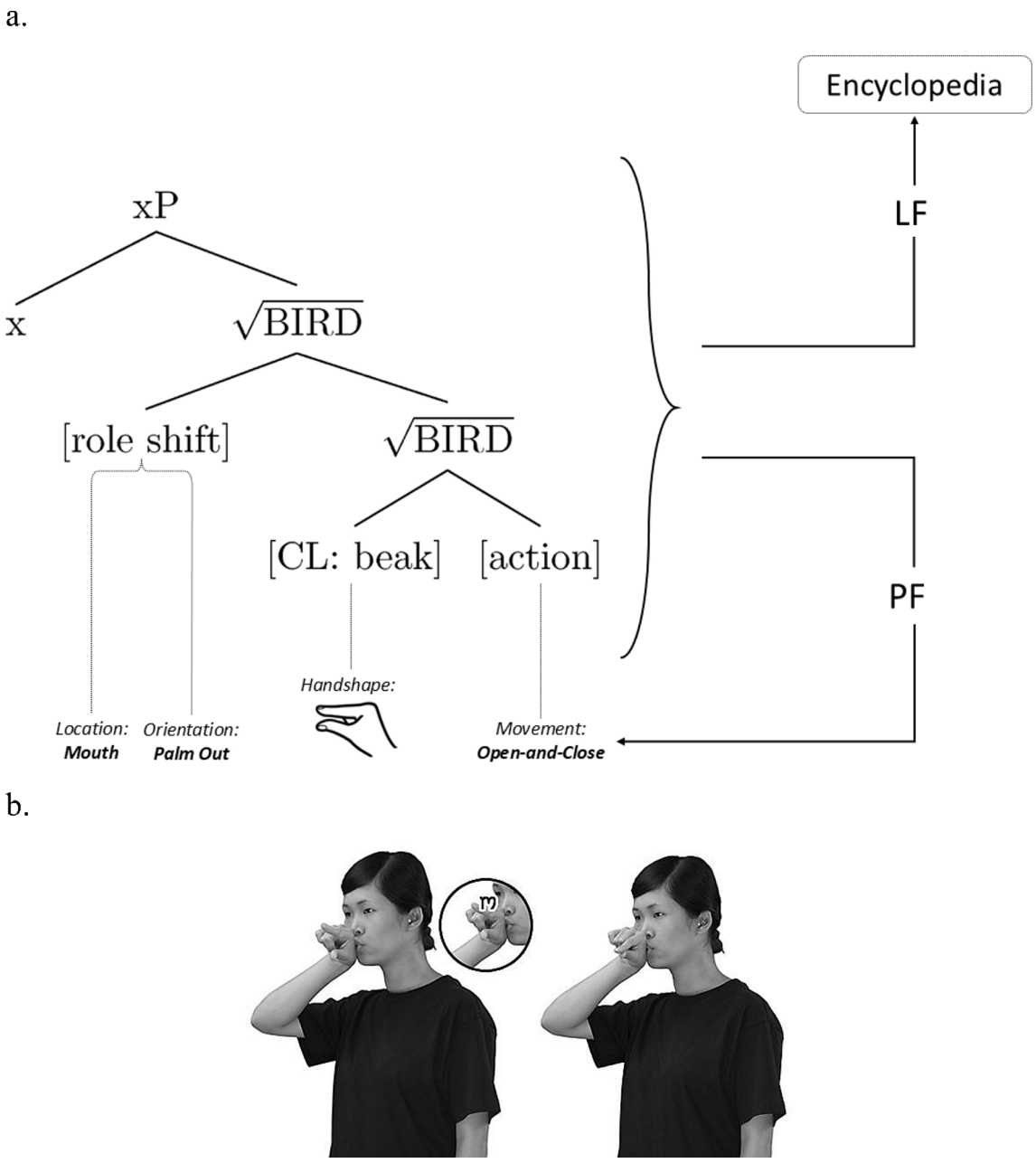


Fig. 6. The root compound $\sqrt{\text{BIRD}}$. (a) shows three morphemic nodes merged under the root compound $\sqrt{\text{BIRD}}$. (b) shows how the sign is produced in HKSL.

other lexical covariates, and whether iconicity ratings explain additional variance beyond classifier structure. Because participant language background has been implicated in variability in the processing of iconic signs (Thompson et al., 2009; Zhang et al., 2023), we also conduct exploratory analyses examining whether key effects differ as a function of age of acquisition by comparing native[4] deaf signers (who are children of deaf parents) and non-native deaf signers (acquisition by age 11).

To test these questions, we use an unprimed lexical decision task targeting single-sign recognition: participants decide whether each item is a real HKSL sign, presented without sentence or discourse context. From this task, we analyze response times and accuracy as indices of decision speed and performance, and we use pupillometry to track task-evoked processing demand during sign perception and decision-making. Because larger baseline-corrected pupillary responses are commonly associated with greater cognitive effort, and pupillometry has been shown to be sensitive to lexical recognition, semantic activation, and lexical retrieval, this measure provides a time-sensitive complement to RT and accuracy measures (Laeng et al., 2012; Schmidtke, 2018; Rojas et al., 2024). We interpret larger pupillary response (i.e., pupil dilation) as greater processing demand, while recognizing that pupillary responses may also reflect attention, arousal, uncertainty, or related cognitive factors (Kuchinke et al., 2007; Kyröläinen et al., 2015; Van Engen & McLaughlin, 2018; Thierfelder et al., 2020; Rojas et al., 2024).

In our lexical decision task, targets are presented one at a time in random order to minimize repetition, semantic relatedness, and

[4] We follow conventional usage in the sign language acquisition literature, where "native" refers to signers with deaf parents (AoA = 0). We acknowledge that this terminology is imperfect: for many signers without deaf parents, HKSL is their primary or sole language, and the native/non-native distinction as used here reflects age of first exposure rather than necessarily a claim about language dominance or proficiency.

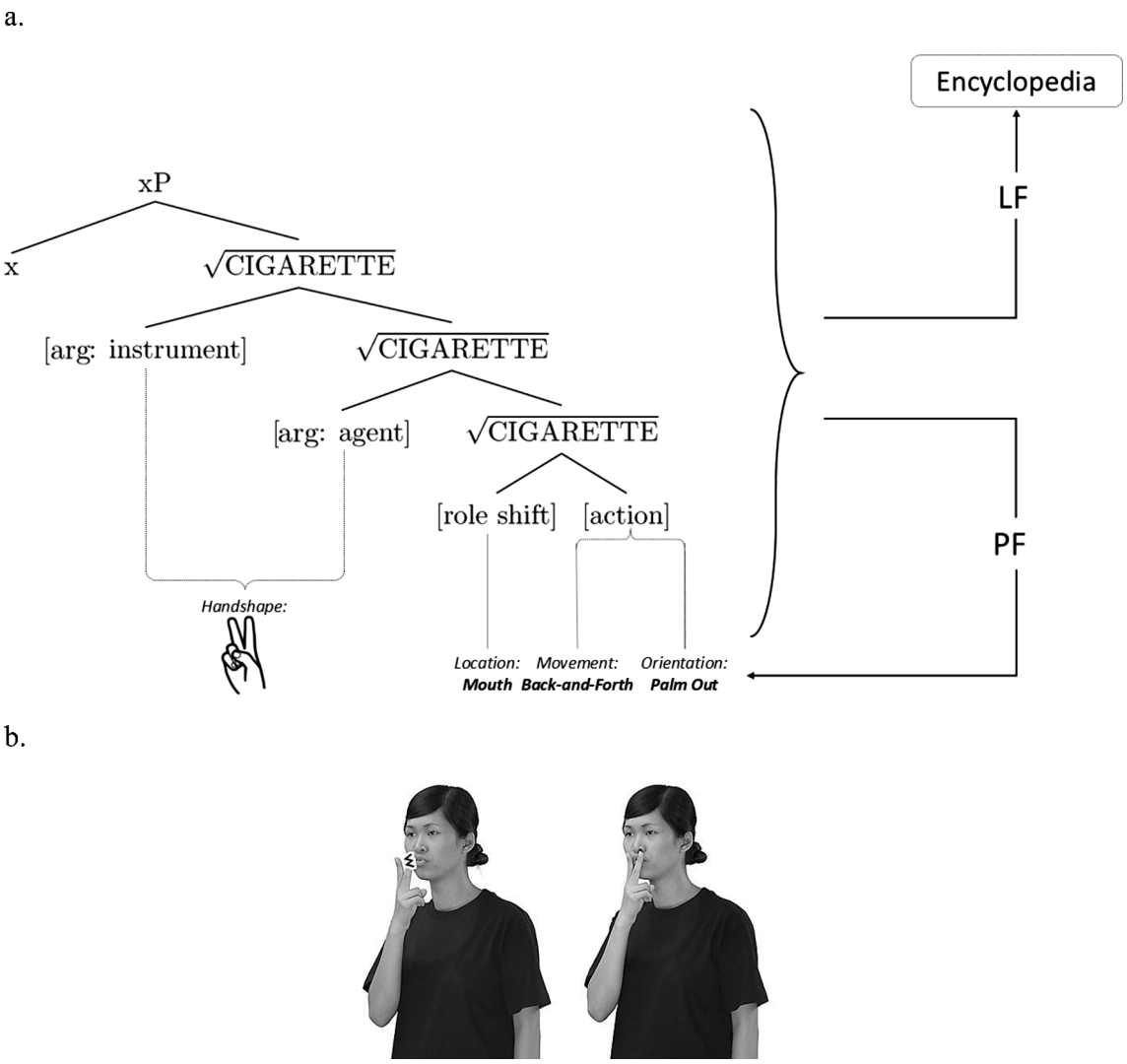


Fig. 7. The root compound √CIGARETTE. (a) shows four morphemic nodes merged under the root compound √CIGARETTE, including argument structure. (b) shows how the sign is produced in HKSL.

expectancy effects that can act as incidental priming and obscure item-level influences (e.g., classifier structure or iconicity). Trial-by-trial drift correction extends the inter-trial interval, which helps stabilize baseline pupil size and reduce carryover from the preceding trial, increasing the likelihood that RT, accuracy, and pupillary responses primarily reflect the processing demands of the current item.

The classifier categories examined here—entity, body-part, and handling—differ in morphophonological properties and in how they typically realize iconic mappings (see Koenders, 2024). Signs without embedded classifier handshapes, by contrast, lack this additional meaning-bearing component. Because classifier type is closely related to iconic mapping, apparent classifier-type effects could partly reflect perceived iconicity, and iconicity effects could be moderated by structural differences across sign types. We therefore consider classifier type and iconicity ratings jointly, with the aim of distinguishing structural effects from effects attributable to perceived form–meaning resemblance.

Method

After collecting data from the first 15 participants, we conducted a simulation-based power analysis to estimate the statistical power attainable within our recruitment constraints. Specifically, we evaluated three target sample sizes: 30, 40, and 50 participants. These values represent the realistic range of eligible, fluent deaf signers of HKSL who began learning before age 12 that we are typically able to recruit. Power was estimated using the *simr* package (version 1.0.7; Green & MacLeod, 2016), with 500 simulations conducted at each target sample size. Because the initial sample was small, the resulting estimates should be interpreted as planning-based approximations rather than definitive estimates of power. Model specifications were kept minimal to reduce overfitting and convergence problems, and we used parsimonious mixed-effects models for power estimation. Final analyses used generalized linear mixed models (GLMMs) or generalized additive mixed models (GAMMs), as appropriate for each dependent measure.

For response times, the power model used inverse-transformed RT as the outcome (Brysbaert & Stevens, 2018), as a simplified approximation to our planned final RT analyses using an inverse Gaussian or gamma GLMM. Fixed effects included sign type, with no-CL as the reference level, and sign length as a control variable, with random intercepts for participants and items. For pupillary response, the original time-series data were aggregated within each trial into two temporal epochs, early and late, defined relative to the trial midpoint. This aggregation simplified the model given the small pilot sample and avoided treating densely sampled, temporally autocorrelated observations as independent in the power simulations. Mean baseline-corrected pupillary

response was modeled with fixed effects of sign type, epoch, and their interaction, with random intercepts for participants and items. For accuracy, trial-level accuracy was coded as 0/1 and modeled using a linear mixed-effects model as a simplified approximation for the power simulations. Final accuracy analyses use a binomial GLMM, which is more appropriate for the binary outcome.

For the pupillary response model, estimated power to detect the sign type × epoch interaction was 63.6% at $N = 30$, 78.0% at $N = 40$, and 88.0% at $N = 50$. For the RT model, estimated power to detect an omnibus effect of sign type was 70.8% at $N = 30$, 72.2% at $N = 40$, and 73.2% at $N = 50$. For the accuracy model, estimated power to detect an omnibus effect of sign type was 45.0% at $N = 30$, 45.4% at $N = 40$, and 46.6% at $N = 50$. Omnibus effects were assessed by comparing models with and without the relevant fixed effect. Overall, given our final sample size of 43 deaf signers, power was adequate for the pupillary response analysis, moderate for RT, and low for accuracy. As such, accuracy results should be interpreted with particular caution. Because these simulations were based on effect-size estimates from a small initial sample and used simplified random-effects structures, the power estimates should be treated as approximate.

Participants

We recruited 43 participants. All were severely to profoundly deaf and had attended a deaf school or mainstream bimodal bilingual program. Their ages ranged from 18 to 53 ($M = 32.2$, $SD = 10.2$). Participants were first exposed to HKSL between 0 and 11 years old, with a mean age of HKSL acquisition of 3.7 years ($SD = 3.51$), and with 30.2% ($n = 13$) being native signers born to deaf parents. This study was approved by the Survey and Behavioural Research Ethics Committee of The Chinese University of Hong Kong (ref: SBRE-23-0684). All participants provided written informed consent and received monetary compensation for their participation.

Materials

The stimuli consisted of 294 video clips of HKSL signs, evenly divided into 147 real signs and 147 pseudosigns. Of the real signs, 34 did not contain a classifier handshape (no-CL signs), 34 contained a body-part classifier handshape (body-part-CL signs), 40 contained an entity classifier handshape (entity-CL signs), and 39 contained a handling classifier handshape (handling-CL signs).

Pseudosign creation

Working with the authors, a native deaf HKSL signer created all pseudosigns, following a similar process to that described in Caselli et al. (2021). They were designed to be phonotactically well-formed (e.g., conforming to symmetry and dominance conditions; Battison, 1978) while not coinciding in phonological form with any known lexical sign in HKSL. Each pseudosign was derived from a corresponding real sign by minimally modifying its manual configuration (number of hands and their spatial relationship), handshape, location, movement, or a combination of these parameters (Table 1). This approach ensured that pseudosigns remained sufficiently similar to existing signs to be plausible members of the lexicon while still being non-lexical items. To ensure that the pseudosigns appeared natural, each was produced with an arbitrary but plausible non-manual feature selected by the signer.

Table 1
Phonological parameters modified to create pseudosigns.

Changed parameter or feature	N
Location	41
Handshape	39
Movement	22
Location + Movement	15
Orientation	7
Handshape + Movement	5
Manual Configuration + Movement	5
Movement + Orientation	5
Hand Arrangement	3
Hand Arrangement + Orientation	3
Handshape + Location	2

Because our goal was to make the smallest possible change that produced an unambiguous non-sign while preserving naturalness, we most often modified location or handshape—parameters that can be altered locally without disrupting overall phonotactic well-formedness. Changes to movement, orientation, or hand arrangement were used when location or handshape changes either created another existing sign or yielded forms judged unnatural by the native signer.

The handshape inventories of the real signs and pseudosigns substantially overlapped: 42 distinct handshapes occurred in the real-sign items and 43 in the pseudosigns, with 38 shared between the two sets. All handshapes used in the pseudosigns are attested in HKSL, including in classifier constructions. However, because pseudosigns lack conventionalized meaning and were presented without discourse context, any classifier-like handshape they contain would likely not be interpreted as encoding classifier morphology. Participants could not, therefore, reject pseudosigns on the basis of handshape alone.

Real sign properties and covariate selection

To ensure that any observed effects of classifier type were not attributable to confounding stimulus characteristics, we compared the four sign types on visual, lexical, semantic, and phonological properties using one-way ANOVAs or generalized linear models as appropriate for each variable. Omnibus p-values were adjusted using the Benjamini–Hochberg procedure to control the false discovery rate across the 14 screening tests. Unless otherwise noted, p-values reported for these omnibus screening tests are BH-adjusted. For variables with reliable omnibus effects, follow-up pairwise comparisons were conducted using estimated marginal means with Tukey adjustment.

Variables showing reliable sign-type differences were considered as candidate covariates in the main analyses. Final covariate inclusion was guided by statistical imbalance, theoretical relevance, collinearity among predictors, and model-fit comparisons. Stimuli properties are summarized in Table 2.

Video-level visual properties. Pupil size is strongly influenced by low-level visual factors, particularly luminance, as well as brief visual transients, in addition to cognitive load (Beatty, 1982; Beatty & Lucero-Wagoner, 2000; Laeng et al., 2012; Mathôt, 2018). To reduce the risk that pupillary effects reflected visual rather than linguistic differences between items, we quantified two video-level properties for each sign stimulus: average luminance (overall brightness) and average optical-flow magnitude (overall visual motion).

For average luminance, each video frame was converted to grayscale and the mean pixel intensity (0–255) was computed for each frame; these values were then averaged across frames to yield a single luminance value per video. For average optical-flow magnitude, we computed dense optical flow (Farnebäck, 2003) in Python using *OpenCV* (Bradski, 2000) to obtain motion vectors at each pixel for each pair of consecutive frames. We then computed vector magnitudes (motion speeds) and averaged these magnitudes across pixels and frame pairs to derive a single optical-flow value per video. These measures were calculated to allow low-level visual properties to be controlled for in subsequent pupillometry analyses.

Average luminance differed across sign types, $F(3, 143) = 4.22$, $p = .012$. Optical-flow magnitude also differed across sign types, $F(3, 143) = 5.95$, $p = .003$. As including both luminance and optical-flow magnitude together as covariates reduced model fit, only optical-flow magnitude was included as a covariate in the final pupillometry models (see Analysis section for model comparison details).

Table 2
Properties of sign stimuli by classifier type. For each sign type, the table reports the number of items, the number of two-handed and asymmetrical signs, the number of signs with animate referents, and means (SDs) for dominant-hand (DH) and non-dominant-hand (NDH) log frequency, sign length (ms), and ratings of familiarity, iconicity, arousal, concreteness, imageability, and valence (1–7 scale), as well as average video luminance and optical-flow magnitude. Sign length was measured from 30 ms before hand-movement onset to 30 ms after the hands returned to the starting position.

Variable	No-CL	Body-Part-CL	Entity-CL	Handling-CL
Items Count (*n*)	34	34	40	39
Two-Handed Signs (*n*)	12	19	31	20
Asymmetrical (*n*)	3	5	17	4
Animate Referents (*n*)	6	25	5	1
DH Log Frequency	4.36 (1.57)	3.82 (1.10)	4.40 (0.72)	3.94 (0.78)
NDH Log Frequency	4.59 (1.43)	4.25 (1.27)	4.71 (0.77)	4.22 (0.88)
Sign Length (ms)	2311 (231)	2155 (172)	2257 (224)	2163 (165)
Familiarity Ratings	6.53 (0.60)	6.54 (0.82)	6.48 (0.71)	6.58 (0.71)
Iconicity Ratings	2.87 (0.81)	5.09 (1.18)	4.52 (1.31)	5.41 (1.28)
Arousal Ratings	4.42 (0.61)	4.44 (0.40)	4.37 (0.61)	4.28 (0.62)
Concreteness Ratings	4.92 (1.11)	5.60 (1.01)	5.43 (1.13)	5.86 (0.92)
Imageability Ratings	4.87 (1.26)	5.62 (0.95)	5.34 (1.31)	5.82 (0.98)
Valence Ratings	4.32 (0.82)	4.00 (0.78)	4.39 (0.68)	4.25 (0.76)
Luminance	21.36 (0.16)	21.34 (0.31)	21.51 (0.21)	21.44 (0.20)
Optical-Flow Magnitude	0.034 (0.006)	0.040 (0.009)	0.038 (0.006)	0.037 (0.006)

Lexical variables. We categorized signs as [+Animate] if they referred to humans, animals, or other sentient entities. A binomial logistic regression showed a significant effect of sign type on animacy ($\chi^2(3) = 55.44$, $p < .001$). Because animacy differed across sign types, with body-part-CL signs more likely to have animate referents, animacy was included as a covariate in all models.

Although handedness was not a primary factor of interest, it can affect the visual and motor complexity of signs and may therefore influence processing. Following standard descriptions of sign phonology (e.g., Eccarius & Brentari, 2007), we distinguish between one-handed signs, two-handed symmetrical signs (same handshape on both hands), and two-handed asymmetrical signs (distinct handshapes on each hand, with one hand typically acting as a passive base or location for the other). A binomial logistic regression showed a significant effect of sign type on the likelihood of being two-handed, $\chi^2(3) = 14.31$, $p = .006$, with entity-CL signs more likely to be two-handed than other sign types. For asymmetry, operationalized as different dominant versus non-dominant handshapes among two-handed signs, the two-handed-only analysis did not show a reliable effect of sign type after correction, $\chi^2(3) = 8.45$, $p = .058$.

Given the theoretical relevance of handedness and symmetry to articulatory and perceptual complexity, and because two-handedness differed reliably across sign types, all items were coded for this three-level handedness/symmetry factor (one-handed, two-handed symmetrical, two-handed asymmetrical), and this factor was included as a covariate in all statistical models.

Handshape frequency was determined separately for the dominant and non-dominant hands by counting the number of HKSL signs that share the same handshape in the Asian SignBank (Centre for Sign Linguistics and Deaf Studies, n.d.). Sign type did not reliably affect handshape frequency for either the dominant hand, $F(3, 143) = 2.71$, $p = .066$, or the non-dominant hand, $F(3, 78) = 1.62$, $p = .224$. Given the absence of significant effects, handshape frequency was not included as a covariate in the main analyses.

Sign length varied significantly across sign types, $F(3, 143) = 5.05$, $p = .006$, with no-CL signs being longest and handling-CL and body-part-CL signs being shorter. Given this variation, sign length was included as a covariate in all models.

Familiarity and iconicity ratings by deaf participants. Out of the 43 participants who took part in the experiment, 34 participated in an iconicity rating task, and 39 participated in a familiarity rating task. Both tasks were conducted around one month after participants completed the lexical decision experiment. HKSL instructions for both tasks were developed based on the scripts provided in Vinson et al. (2008). These instructions were recorded and shown to participants before each task. Both rating tasks used a 1–7 scale (1 = completely unfamiliar, 7 = highly familiar for the familiarity rating task; 1 = non-iconic, 7 = highly iconic for the iconicity rating task).

Deaf participants indicated similarly high levels of familiarity with signs across all sign types, $F(3, 143) = 0.15$, $p = .930$, and no pairwise comparisons were significant ($ps \geq .911$), suggesting that the items were well matched on this dimension. As predicted, iconicity ratings differed by sign type, $F(3, 143) = 32.38$, $p < .001$. Pairwise comparisons showed that body-part-CL, entity-CL, and handling-CL signs all received higher iconicity ratings than no-CL signs ($ps < .001$). Additionally, handling-CL signs were rated as more iconic than entity-CL signs, $p = .006$. Body-part-CL and handling-CL signs did not differ from each other, $p = .650$, nor did body-part-CL and entity-CL signs, $p = .172$. The distributions of iconicity ratings across the four sign types are visualized in Fig. 8.

Note that the no-CL category was not restricted to arbitrary signs: Iconicity ratings spanned the lower-to-middle range (M = 2.87, SD = 0.81), indicating that some no-CL signs were perceived as iconic (e.g., TASTE, a pointing sign directed to the mouth without a classifier handshape), although on average they were still less iconic than CL-containing signs.

Given the central importance of iconicity in our study, we assessed the inter-rater reliability for the iconicity ratings using a two-way random effects model to determine the degree of consistency among raters. The intraclass correlation coefficient [ICC(2,1)] was .97, 95% CI [.96,.97], indicating strong inter-rater consistency ($p < .001$). This high level of agreement shows that most participants rated the iconicity of the items similarly, supporting the reliability of the ratings for use in our statistical models.

Arousal, valence, concreteness, and imageability ratings. Ratings for additional psycholinguistic variables were collected from an independent sample of 41 hearing, Cantonese-speaking university students in Hong Kong (mean age = 20.5 years, *SD* = 1.26), based on the Chinese

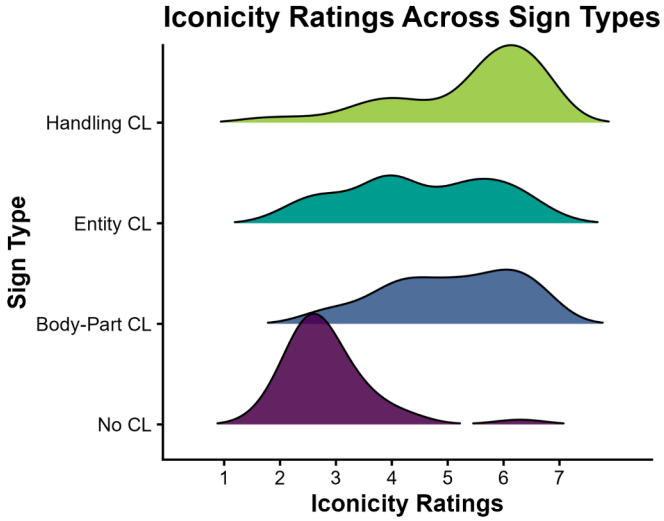


Fig. 8. Distribution of iconicity ratings (1–7 scale) across sign types.

translations of the target signs. This follows the approach of Emmorey et al. (2020), who linked English lexical norms to ASL signs via their translation equivalents. These ratings therefore index properties of the *intended meanings* of the target signs rather than properties of the signed forms themselves. We chose this approach for practical reasons, as adding four additional rating tasks (valence, arousal, concreteness, imageability) would have placed a substantial testing burden on our limited pool of deaf signers.

Participants rated valence, concreteness, imageability, and arousal using 7-point Likert scales, with instructions provided in traditional Chinese. All 147 words were rated across the four dimensions. Both the order of dimensions and the order of items within each dimension were randomized for each participant; the task took approximately one hour.

Ratings for the meanings of the words corresponding to the four sign types were similar for arousal, $F(3, 143) = 0.60$, $p = .667$, and valence, $F(3, 143) = 1.70$, $p = .215$. In contrast, there were reliable effects of sign type on both concreteness, $F(3, 143) = 5.12$, $p = .006$, and imageability, $F(3, 143) = 4.63$, $p = .008$. No-CL meanings were rated as less concrete and less imageable than the meanings of classifier-containing signs.

Concreteness and imageability were strongly correlated, $r(145) = .97$, $p < .001$, suggesting these two variables are functionally equivalent for purposes of the current study. Because of this, we included only one of these variables as a covariate in the main analyses; concreteness was selected over imageability based on AIC comparisons (see Analysis).

Because these ratings were obtained from hearing non-signers, they constitute spoken-language proxies for the corresponding properties of the signed items. Given the particular importance of concreteness in our later analyses, we conducted a small validation study in which 10 deaf HKSL signers rated the concreteness of the signed stimuli. Person-mean-centered signer ratings were strongly correlated with the hearing-based norms, $r(143) = .61$, 95% CI [.49,.70], $p < .001$, and showed no systematic mean difference (deaf − hearing = −0.02, 95% CI [−0.20, 0.16], $t(144) = -0.23$, $p = .82$). The mean absolute item-wise difference was 0.84 scale units (median = 0.67) on the 7-point scale. These results indicate that the hearing-based norms provide a reasonable proxy for deaf signers' concreteness judgments in this stimulus set.

Apparatus

Eye movements were monitored with an EyeLink 1000 Plus desktop system (SR Research Ltd., 2020) sampling at 1000 Hz. The experimental procedure was conducted on a Dell OptiPlex 7020 PC (Intel Core i7-4790 processor, 3.6 GHz; Windows 7 operating system) paired with a 24-inch ViewSonic XG2402 monitor (1920 × 1080 resolution; 144 Hz refresh rate). To reduce head movement, participants used a chin and forehead rest, which positioned them at a distance of 60 cm from the monitor.

Procedure

The experiment was programmed using Experiment Builder (Version 2.6.11; SR Research Ltd., 2024a). Prior to beginning the experiment, and at the start of each of the three subsequent blocks, each participant completed a 5-point calibration and validation of the eye tracker. For drift correction, a dot was presented at the approximate location of the signer's face, reflecting the typical gaze focus during signed conversations. Calibration and validation procedures were repeated whenever the drift correction error exceeded 1° of visual angle.

Participants were instructed to press the "s" key if the sign shown was a real HKSL sign, and the "l" key if it was not. Each video clip was presented for its entire duration, during which participants determined whether the sign was real or a pseudosign. Before the experimental trials, participants completed eight practice trials to familiarize themselves with the task. After these practice trials, participants had the opportunity to repeat the practice or ask the experimenter for clarification before proceeding to the first block.

After each s/l keyboard press throughout the experiment, participants received immediate feedback indicating whether they had correctly identified the previous sign as a real sign or a pseudosign. To minimize fatigue, participants were prompted to take a short break after every block of 98 trials (three breaks in total).

Statistical analysis

All statistical analyses were conducted in R (version 4.4.0). The primary dependent variables analyzed from the lexical decision experiment were accuracy, response time (RT), and pupil size. We employed generalized linear mixed models or generalized additive models for each variable, as appropriate. Continuous predictors were *z*-scored and categorical predictors with more than two levels (articulation type: one-handed, two-handed symmetrical, two-handed asymmetrical) were sum-coded prior to analysis.

Predictor selection and multicollinearity assessment

To avoid redundancy and ensure stable estimation of model parameters, we examined correlations among all candidate predictors and checked for multicollinearity prior to finalizing our regression models. Pairwise Pearson correlations among candidate predictors revealed that concreteness and imageability were highly correlated ($r = .97$), whereas all other pairs showed only weak to moderate correlations (all other $|r| \leq .54$). To address this, we compared model AIC (Akaike, 1998) values for versions including either concreteness or imageability. The model with concreteness provided a better fit, so concreteness was retained and imageability was excluded from the final analyses.

To further assess multicollinearity, we calculated variance inflation factors (VIFs) using a generalized linear model for accuracy and linear models for RT and pupillary response, approximating the fixed effects structure of each of the planned models. All VIFs were below 5 (accuracy: 1.08–1.37; RT: 1.08–1.55; pupil: 1.02–4.64), indicating low collinearity among the included predictors.

Pupillary responses

Preprocessing. Pupil size data were exported from SR Research EyeLink Data Viewer (Version 5.1.1; SR Research Ltd., 2024b) in 50 ms time bins. Pupil size was originally recorded in arbitrary units; to facilitate interpretability, these values were converted to approximate millimeters following the manufacturer's recommended procedure (SR-Support, 2020) by scaling against an artificial pupil (5 mm black dot) recorded under identical camera settings. Because all analyses used baseline-corrected pupil change, this scaling affects measurement units but not inferential patterns.

Preprocessing was conducted using *pupillometryR* (Forbes, 2020). Trials with more than 50% missing data were excluded (seven trials removed). The remaining data were smoothed with a 9-point median filter, and brief periods of missing data (e.g., blinks) were corrected by linear interpolation. Pupil measurements were baseline-corrected by subtracting mean pupil diameter from 450–600 ms post-stimulus onset for each trial, chosen to reduce contamination from the onset luminance reflex.

Statistical modeling. The analysis window spanned 600–2100 ms post-stimulus onset, excluding the initial pupillary constriction triggered by video onset. This window provided a 1500 ms segment encompassing the typical peak of task-evoked dilation (Schmidtke, 2018). Observations were included up to each participant's response if earlier than 2100 ms.

We fit a generalized additive mixed model (GAMM) using the bam() function in *mgcv* with a scaled t distribution. The model included parametric fixed effects for classifier type (no-CL as reference), iconicity, animacy, concreteness, articulation type, sign length, and average optical-flow magnitude.

Model comparison using AIC indicated that the model including optical-flow magnitude as the sole visual covariate provided the best fit (ΔAIC = 0), outperforming models with both optical-flow magnitude and luminance (ΔAIC = 21), luminance only (ΔAIC = 50.9), and those with no visual covariates (ΔAIC = 79.8). Accordingly, we included optical-flow magnitude (but not luminance) as the visual-control covariate in the main analyses. To control for gaze-dependent measurement artifacts, we included terms for horizontal and vertical gaze position, their quadratic terms, and their interaction.

Time was modeled with a global thin-plate spline plus classifier-type-specific deviation smooths using a shrinkage basis. The time-varying effect of iconicity was modeled with a tensor interaction. Random effects included item-level random intercepts and participant-specific smooth trajectories over time. An AR(1) structure was included to account for the non-independence of consecutive time points. The autocorrelation parameter (ρ = .94) was estimated by calculating the lag-1 autocorrelation of residuals from an initial model without AR(1) correction. Basis dimensions for smooth terms were evaluated using k.check() and increased where evidence suggested insufficient flexibility.

Exploratory nativeness/acquisition analysis. We additionally examined whether the time course of iconicity effects differed by native signer status. Here, native signers were defined as from-birth signers (AoA = 0; deaf signing parents), and non-native signers were defined as early-acquiring signers (AoA M = 5.3 years, SD = 3.01, range = 1–11). To test for group differences in the time-varying iconicity effect, we fit an additional GAMM that included (i) a shared time × iconicity tensor interaction common to both groups and (ii) an additional tensor interaction multiplied by a 0/1 nativeness indicator, providing a single omnibus test of whether the time × iconicity relationship differed between native and non-native signers. The model also included a nativeness main effect and allowed the overall time course to differ by nativeness via a nativeness-specific smooth of time, alongside the same covariates and random effects as in the primary model.

Accuracy

Accuracy (correct/incorrect) was analyzed using generalized linear mixed-effects models with a binomial distribution and logit link, implemented in *glmmTMB* (Brooks et al., 2017). A total of 6,321 trials were included (overall accuracy = 88.5%). We fit two models: the first examined classifier type effects while controlling for potential confounds; the second added iconicity to assess whether classifier-type effects were attenuated after accounting for iconicity. Both models included classifier type as a fixed effect (no-CL as reference), with covariates for concreteness, animacy, articulation type, and sign length. Random effects included intercepts for participant and item, plus by-participant random slopes for classifier type. The second model additionally included by-participant random slopes for iconicity. Where necessary to achieve convergence, correlations among by-participant random effects were dropped. Using the same native versus non-native grouping defined in the pupillometry analyses, we fit a third accuracy model adding a group × iconicity interaction to test whether the effect of iconicity on accuracy differed by nativeness/acquisition timing.

Response times

Response times (RTs) for correct responses (n = 5,593) were analyzed using generalized linear mixed-effects models with an inverse Gaussian distribution and identity link in order to accommodate the positive skew typical of RT data (Lo & Andrews, 2015), implemented in *lme4* (Bates et al., 2015). Model comparisons based on AIC indicated that the inverse Gaussian distribution provided a better fit than a gamma distribution.

To reduce the influence of extreme values and maximize overlap with trials included in the pupillometry analysis, RTs were winsorized at the 2.5th and 97.5th percentiles within each participant. This quantile-based approach is robust to positive skew (Nicklin & Plonsky, 2020). A total of 336 trials were winsorized (≈ 6%); 5,593 trials were included for analysis.

The fixed and random effects structures paralleled the accuracy models. Both models included classifier type as a fixed effect (no-CL as reference), with covariates for concreteness, animacy, articulation type, and sign length. Random effects included intercepts for participant and item, plus by-participant random slopes for classifier type; the second model additionally included by-participant random slopes for iconicity. For all RT models, correlations among by-participant random effects were dropped to achieve convergence. Using the same native versus non-native grouping defined in the pupillometry analyses, we added a group × iconicity interaction to test whether the effect of iconicity on RT differed by nativeness.

Results

Descriptive statistics for pupil measures, accuracy, and response time by sign type are presented in Table 3 (note: the three broad pupil time bins shown here are used for descriptive summaries only; statistical models used 50 ms time bins), and observed baseline-corrected pupil trajectories are shown in Fig. 9. Full model results are reported in the tables below.

Pupillary response model

Parametric coefficients for the primary pupillary response model (Model P1) are presented in Table 4. Relative to the reference level, no-CL signs, body-part-CL, entity-CL, and handling-CL signs were

Table 3
Descriptive statistics for dependent measures by sign type. Values are M (SD) based on participant-level averages. Accuracy was calculated as each participant's proportion correct within sign type. Response times were restricted to correct trials, winsorized within participant at the 2.5th and 97.5th percentiles, and then averaged within participant and sign type. For pupil measures, samples were averaged within participant, sign type, and broad descriptive time window; the table reports the mean and SD of these participant-level averages across participants. Overall pupil values were computed by averaging the three window-level means within each participant and sign type before summarizing across participants. Pupil diameter reflects the converted (non-baseline-corrected) measure in millimeters. Pupil change reflects baseline-corrected stimulus-evoked dilation relative to the 450–600 ms baseline window (referred to as the pupillary response in the text). The three broad time windows shown here are used for descriptive summaries only; statistical models used 50-ms time points across the same 600–2100 ms analysis window. CL = classifier.

Measure	No CL	Body-Part CL	Entity CL	Handling CL
Pupil diameter (mm)				
600–1100 ms	2.814 (0.356)	2.819 (0.360)	2.811 (0.363)	2.822 (0.361)
1100–1600 ms	2.939 (0.381)	2.946 (0.379)	2.945 (0.386)	2.960 (0.383)
1600–2100 ms	3.004 (0.409)	3.032 (0.405)	3.021 (0.421)	3.030 (0.422)
Full window	2.919 (0.379)	2.932 (0.379)	2.926 (0.387)	2.937 (0.386)
Pupil change (mm)				
600–1100 ms	−.006 (.047)	−.001 (.047)	−.001 (.046)	−.002 (.042)
1100–1600 ms	.118 (.083)	.126 (.084)	.133 (.086)	.135 (.081)
1600–2100 ms	.186 (.108)	.206 (.112)	.207 (.110)	.211 (.107)
Full window	.099 (.070)	.110 (.071)	.113 (.072)	.115 (.065)
Accuracy	.82 (.10)	.93 (.06)	.90 (.08)	.89 (.06)
Response time (ms)	1411 (137)	1291 (117)	1360 (123)	1372 (128)

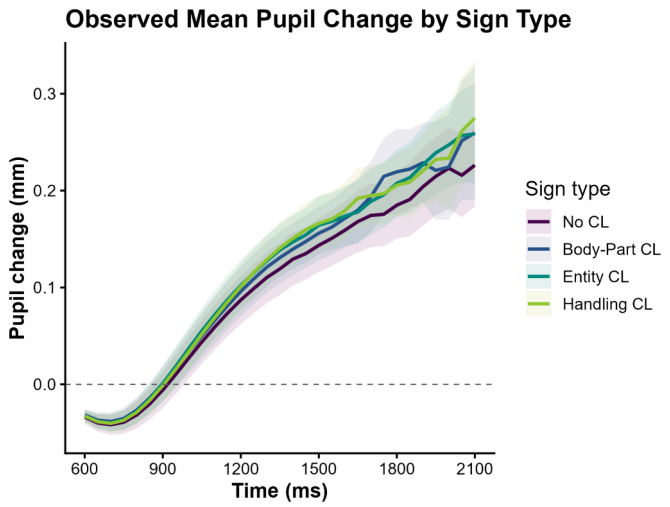


Fig. 9. Observed mean pupil change (baseline-corrected) over time by sign type. Lines represent means across participants. Shaded regions indicate 95% confidence intervals. The dashed horizontal line marks zero (no change from baseline).

Table 4
Parametric coefficients for the generalized additive mixed model predicting pupillary response (Model P1). Reference level for classifier type is no-CL (signs without an embedded classifier). Articulation type is sum-coded: Type 1 compares one-handed signs to the grand mean; Type 2 compares symmetrical two-handed signs to the grand mean. Continuous predictors were z-scored. Gaze position terms control for pupil size variation associated with eye position.

Parametric term	Estimate	SE	t	p
(Intercept)	0.106	0.011	9.76	<.001
Body-part-CL (vs. reference)	0.019	0.004	4.84	<.001
Entity-CL (vs. reference)	0.015	0.003	4.7	<.001
Handling-CL (vs. reference)	0.024	0.004	6.57	<.001
Iconicity	−0.005	0.001	−3.99	<.001
Animacy	−0.002	0.003	−0.63	.528
Concreteness	0.001	0.001	0.81	.416
Articulation type 1	−0.006	0.002	−3.78	<.001
Articulation type 2	−0.001	0.001	−0.45	.650
Sign length	−0.002	0.001	−1.63	.103
Optical-flow magnitude	−0.001	0.001	−1.18	.238
Gaze x	−0.002	2.00×10^{-4}	−7.87	<.001
Gaze y	−0.002	2.80×10^{-4}	−6.61	<.001
Gaze x^2	-2.00×10^{-5}	1.02×10^{-5}	−1.95	.051
Gaze y^2	3.17×10^{-5}	8.26×10^{-5}	0.38	.701
Gaze x × gaze y	-1.87×10^{-4}	7.63×10^{-5}	−2.45	.014

associated with significantly greater pupil dilation. Among these sign types, Tukey-corrected comparisons showed that handling-CL signs elicited greater pupil dilation than entity-CL signs (Entity-CL – Handling-CL: $\Delta b = -0.009$, SE = 0.003, $p = .007$), whereas body-part-CL signs did not differ significantly from entity-CL signs ($p = .632$) or handling-CL signs ($p = .440$). Higher iconicity predicted reduced pupil dilation. Animacy, concreteness, sign length, and optical-flow magnitude were not significant predictors.

Articulation type was sum-coded, with Articulation Type 1 indexing one-handed signs relative to the grand mean and Articulation Type 2 indexing symmetrical two-handed signs relative to the grand mean. Articulation Type 1 was associated with reduced pupil dilation, whereas Articulation Type 2 was not significant. Tukey-corrected comparisons showed that one-handed signs elicited less pupil dilation than asymmetrical two-handed signs (One-handed – Two-handed [asymmetrical]: $\Delta b = -0.013$, $SE = 0.003$, $p < .001$), and symmetrical two-handed signs elicited less pupil dilation than asymmetrical two-handed signs (Two-handed [symmetrical] – Two-handed [asymmetrical]: $\Delta b = -0.008$, $SE = 0.003$, $p = .008$). One-handed and symmetrical two-handed signs did not differ significantly ($p = .092$).

Smooth term estimates for Model P1 are presented in Table 5 and should be interpreted alongside the fitted trajectories and pairwise difference curves in Figs. 10 and 11. There was a significant nonlinear effect of time, indicating that pupil dilation varied nonlinearly across the analysis window. Time-by-classifier-type deviation smooths indicated some differences in trajectory shape across sign types (although deviation smooths with near-zero estimated degrees of freedom should be interpreted cautiously). The pairwise difference curves showed that, relative to no-CL signs, body-part-CL, entity-CL, and handling-CL signs elicited greater predicted pupil dilation over later portions of the analysis window. Handling-CL signs also elicited greater predicted pupil dilation than body-part-CL and entity-CL signs during later time windows, whereas body-part-CL and entity-CL signs did not show a reliable difference window. The significant time-by-iconicity smooth indicated that the effect of iconicity varied over time; this relationship is visualized in Fig. 12.

Accuracy models

In Model A1, body-part-CL signs were identified more accurately than no-CL signs, whereas entity-CL and handling-CL signs did not differ reliably from no-CL signs (Table 6). In Model A2, after iconicity was added, classifier-type effects were no longer significant, and iconicity positively predicted accuracy (Table 7). Concreteness also emerged as a negative predictor of accuracy in Model A2. Thus, the initial accuracy advantage for body-part-CL signs appears to be largely accounted for by iconicity. Fig. 13 shows estimated marginal mean accuracy by classifier type for Models A1 and A2, and Fig. 14 shows predicted accuracy as a function of iconicity from Model A2.

RT models

In Model R1, responses were faster for body-part-CL signs than for no-CL signs, whereas entity-CL and handling-CL signs did not differ reliably from no-CL signs (Table 8). Animacy, sign length, and the sum-coded one-handed articulation contrast were also significant predictors. In Model R2, after iconicity was added, the body-part-CL effect was no longer significant, iconicity predicted faster responses, and handling-CL signs were slower than no-CL signs (Table 9). Animacy, sign length, and the one-handed articulation contrast remained significant. However, Tukey-corrected pairwise comparisons for articulation type were not significant (all $ps \geq .142$), indicating that the articulation effect should not be interpreted as a reliable difference between specific articulation categories. Together, these results suggest that the initial RT advantage for body-part-CL signs was largely accounted for by iconicity, whereas handling-CL signs showed a processing cost once iconicity was controlled. Fig. 15 shows estimated marginal mean response times by

Table 5
Smooth term estimates for the generalized additive mixed model predicting pupillary response. edf = estimated degrees of freedom; Ref.df = reference degrees of freedom. Classifier-type deviation smooths represent departures from the overall time trajectory for each sign type. Random effects include participant-level time trajectories and item-level intercepts.

Smooth term	edf	Ref.df	F	p
Time (smooth)	19.70	21.38	38.47	<.001
Time (smooth), No-CL deviation	1.57	19.00	3.57	<.001
Time (smooth), Body-part-CL deviation	1.78×10^{-3}	19.00	0.00	<.001
Time (smooth), Entity-CL deviation	1.52×10^{-3}	19.00	0.00	.13
Time (smooth), Handling-CL deviation	1.15	19.00	0.95	<.001
Time × Iconicity (smooth)	4.40	6.94	6.98	<.001
Participant time (random smooth)	405.00	429.00	60.69	<.001
Item (random intercept)	47.70	137.00	0.54	<.001

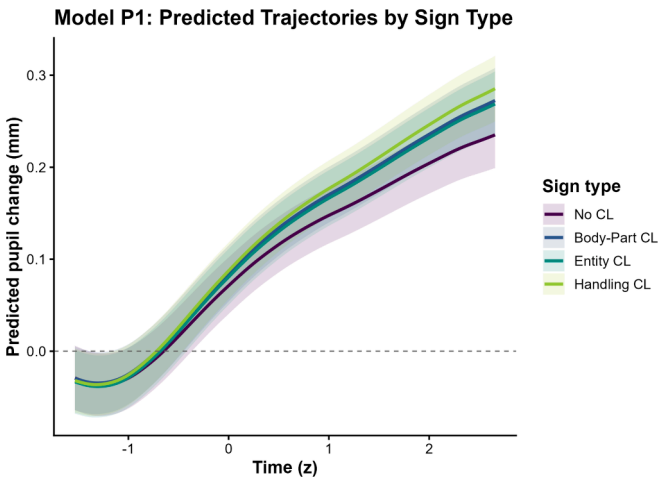


Fig. 10. Model-predicted pupil change (baseline-corrected) over time by sign type from Model P1. Lines represent estimated marginal means, marginalized over articulation type using empirical proportions. Shaded regions indicate simultaneous 95% confidence bands. Predictions are shown at reference values for all other covariates. The dashed horizontal line marks zero (no change from baseline).

classifier type for Models R1 and R2, and Fig. 16 shows predicted response time as a function of iconicity from Model R2.

Exploratory models: native vs. non-native signers

We ran an exploratory GAMM testing whether native and non-native signers differed in how iconicity related to baseline-corrected pupil change over time. The model estimated, for each combination of time and iconicity, the predicted group difference in pupillary response, expressed as Native − Non-Native. This nativeness difference in the time-by-iconicity pattern was reliable overall (edf = 5.51, $p = .003$). Fig. 17 visualizes the Native − Non-Native difference across the full time × iconicity space: warmer/lighter colors indicate regions where native signers were predicted to show greater pupil dilation than non-native signers, whereas cooler/darker colors indicate regions where native signers were predicted to show reduced pupil dilation. White contours mark regions where this group difference was statistically significant, defined as simultaneous 95% confidence intervals excluding zero. Fig. 18 visualizes the same model predictions as time-course trajectories for each nativeness group at five representative iconicity values.

Reliable group differences were concentrated in two time windows: approximately 1060–1542 ms and 1949–2100 ms post-sign onset. In the earlier window, native signers showed stronger iconicity-related modulation than non-native signers: for low-iconicity signs, native signers showed larger pupillary responses than non-native signers from 1090–1542 ms, whereas for high-iconicity signs, native signers showed reduced pupil dilation compared to non-native signers from 1060–1535 ms. In the later window, approximately 1949–2100 ms, this pattern reversed: native signers showed less pupil dilation than non-native signers for low-iconicity signs from 1957–2077 ms, but great pupil dilation for high-iconicity signs from 1972–2092 ms.

We also tested whether the effect of iconicity on behavioral outcomes differed by nativeness. For accuracy, the nativeness × iconicity interaction was significant ($b = -0.32$, $SE = 0.14$, $z = -2.34$, $p = .019$). Overall accuracy was similar between groups descriptively: 89.3% for native signers and 88.1% for non-native signers. However, the groups differed in how accuracy varied with iconicity. Non-native signers showed a larger model-predicted accuracy increase from low- to high-iconicity signs, from 78.5% to 94.3%, whereas native signers showed a smaller increase, from 84.3% to 92.8%. This difference was driven primarily by low-iconicity signs, for which native signers outperformed non-native signers by approximately 6 percentage points. This pattern is visualized in Fig. 19. In contrast, the nativeness × iconicity interaction did not reach significance for response time ($b = -0.98$, $SE = 7.03$, $t = -0.14$, $p = .889$), suggesting that the facilitative effect of iconicity on processing speed did not differ reliably between groups. Full exploratory model outputs are provided in the online supplementary materials.

Discussion

This study examined whether and to what extent embedded classifier handshapes and iconicity account for variation in lexical decision performance for HKSL signs among deaf signers, with the broader goal of identifying how iconic form–meaning mappings and classifier-related structure contribute to sign recognition. Using an unprimed lexical decision task, we measured accuracy, response time, and task-evoked pupillary response to index behavioral performance and processing effort over time during sign identification and decision making.

We addressed two central questions: (i) whether signs containing embedded classifier handshapes differ in processing from signs without classifier handshapes, and (ii) whether any such differences are explained by perceived iconicity or instead persist after iconicity is controlled, consistent with additional classifier-related structural or integration demands. To address these questions, our stimuli included four sign types: signs without classifier handshapes (no-CL signs), and signs with embedded body-part, entity, or handling classifier handshapes. Analyses modeled both sign type and continuous iconicity ratings, while controlling for a range of potential confounds including animacy, concreteness, articulation type, sign length, optical-flow magnitude, and gaze position.

Pupillometry

Pupillary response offers a continuous index of processing effort. By examining pupillary response in a stimulus-locked window (600–2100 ms post-onset), we can track how cognitive demands evolve during sign identification and decision making. Our data revealed that signs with embedded classifier handshapes (i.e., body-part-CL, entity-CL, and handling-CL signs) elicited greater pupil dilation than signs without embedded classifier handshapes (i.e., no-CL signs), suggesting that these signs are associated with greater processing effort during lexical recognition.

Because these pupillary response effects persisted when variance in iconicity ratings and a range of linguistic and low-level visual confounders were controlled, we interpret this pattern as consistent with additional processing demands associated with classifier-containing signs. We propose that one mechanism underlying this cost is the need to integrate classifier-related handshape information with the movement, location, orientation, and whole-sign lexical meaning during recognition. Relative to less semantically specified handshapes, classifier handshapes carry additional meaning, and different subtypes vary in how much morphosyntactic and argument-structural information they contribute, making classifier category a theoretically motivated proxy for structural complexity. Of course, morphological complexity does not always entail greater processing effort in spoken languages, where costs may be modulated by frequency, transparency, and contextual predictability (e.g., Bertram et al., 2000; Kim et al., 2022; Rastle et al., 2015; Vergara-Martínez et al., 2009). However, the simultaneous articulation of classifier-containing signs may place particular demands on integration, as signers must identify and compose co-occurring meaning-bearing components in parallel (see Emmorey & Corina, 1990).

Two additional patterns are consistent with the interpretation that integrating simultaneous meaning-bearing components increases processing effort. The first comes from articulation type. Two-handed asymmetrical signs produced larger pupillary responses than one-

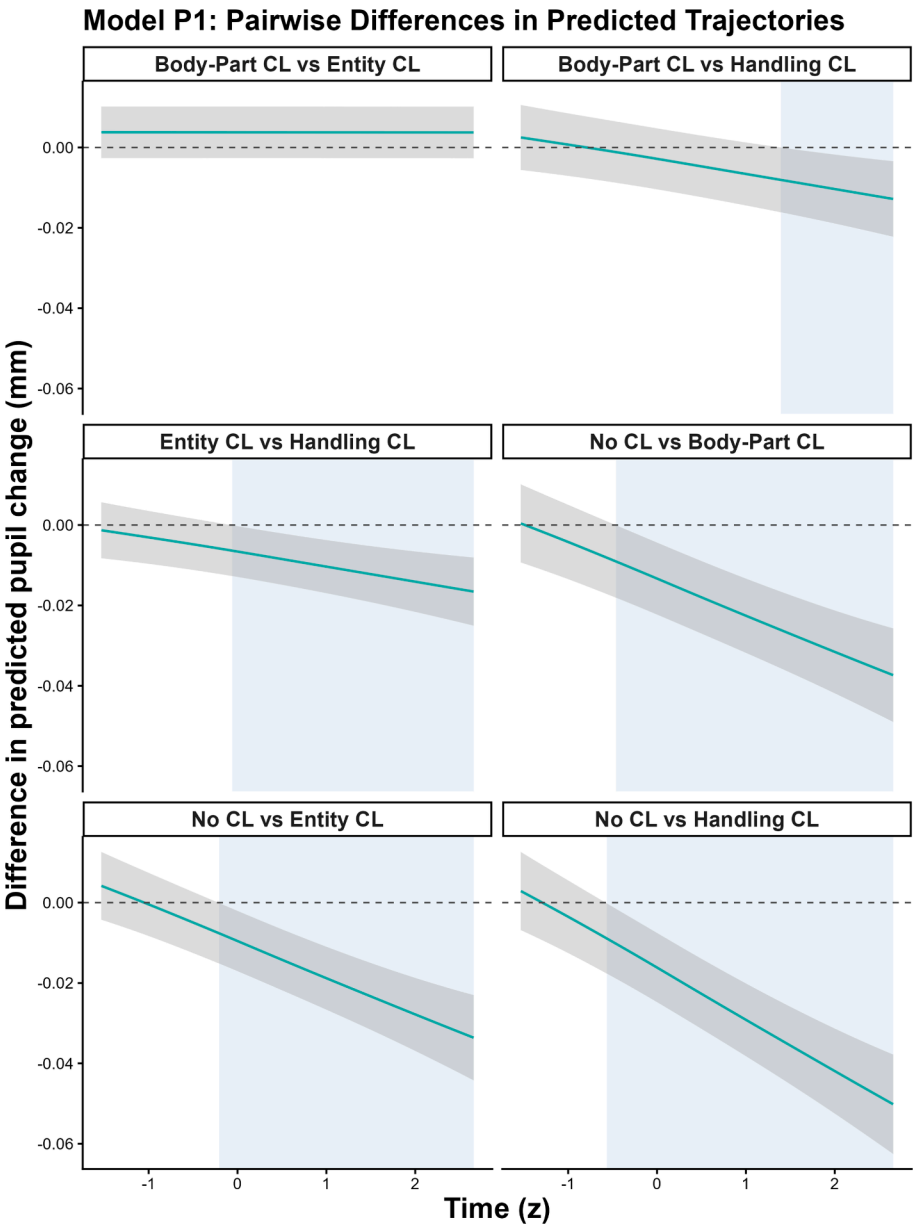


Fig. 11. Pairwise difference curves comparing predicted pupil trajectories between sign types from Model P1. Solid lines show estimated differences with simultaneous 95% confidence bands (gray ribbons). Blue shading indicates time windows where differences are statistically significant (confidence band excludes zero). Predictions are marginalized over articulation type and evaluated at reference values for all control variables. Relative to no-CL signs, body-part-CL showed greater predicted pupil change from 979–2090 ms, entity-CL from 1069–2090 ms, and handling-CL from 942–2090 ms. Handling-CL also showed greater predicted pupil change than body-part-CL from 1640–2090 ms and than entity-CL from 1122–2090 ms. The body-part-CL versus entity-CL contrast did not yield a reliable difference window. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

handed or two-handed symmetrical signs, an effect that held independent of classifier category. In asymmetrical signs, the two hands contribute distinct phonological specifications and contribute distinct semantic or morphophonological information that must be integrated into a single lexical representation. In contrast, symmetrical signs use the same handshape across both hands, and one-handed signs present only a single manual specification. The increased pupil dilation for asymmetrical signs is therefore consistent with the broader proposal that processing effort increases when distinct sublexical specifications must be unified in parallel, rather than simply with the presence of two articulators.

The second relevant pattern is that handling-CL signs elicited the greatest increases in pupil dilation among the three classifier categories, consistent with the possibility that they impose greater classifier-related integration demands. In classifier-predicate constructions, handling classifiers have been analyzed as encoding a transitive, two-participant event structure in which a hand acts on a manipulated entity (Benedicto et al., 2007; Meir et al., 2007; Zwitserlood, 2003, 2008). If lexical signs containing embedded handling-classifier handshapes retain aspects of this relational structure, they may impose greater integration demands during recognition than entity-CL or body-part-CL signs, whose internal structure may be comparatively less complex.

A related, non-mutually exclusive possibility concerns the indirectness of the form–meaning mapping in handling-CL signs, which can depict a hand–entity interaction rather than the referent itself (see Padden et al., 2015; Hwang et al., 2017). On this account, recognition may involve linking an action-based representation to referent identity, such that the depicted handling event stands for the manipulated object

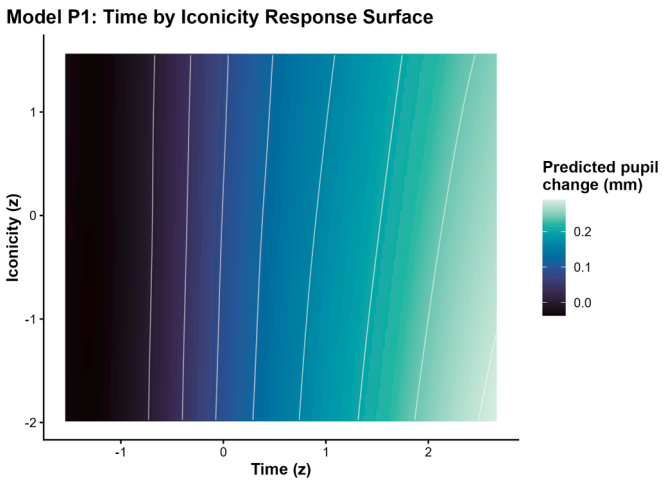


Fig. 12. Population-level response surface showing predicted pupil change as a function of time and iconicity (both *z*-scored) from Model P1. Predictions are marginalized over classifier type and articulation type using empirical weights. Contour lines connect points of equal predicted pupil change. Lighter colors indicate greater dilation.

Table 6
Fixed effects for the accuracy model without iconicity (Model A1). Reference level for classifier type is no-CL (signs without an embedded classifier). Articulation type is sum-coded: Type 1 compares one-handed signs to the grand mean; Type 2 compares symmetrical two-handed signs to the grand mean. Continuous predictors were *z*-scored.

Predictor	Estimate	SE	z	p
(Intercept)	2.62	0.44	5.91	<.001
Body-part-CL (vs. reference)	1.47	0.6	2.43	.015
Entity-CL (vs. reference)	0.77	0.53	1.46	.144
Handling-CL (vs. reference)	0.96	0.54	1.78	.075
Concreteness	−0.13	0.19	−0.69	.491
Animacy	0.08	0.52	0.15	.879
Articulation Type 1	−0.36	0.26	−1.4	.161
Articulation Type 2	−0.06	0.25	−0.25	.800
Sign Length	0.01	0.18	0.03	.975

Table 7
Fixed effects for the accuracy model with iconicity (Model A2). Reference level for classifier type is no-CL (signs without an embedded classifier). Articulation type is sum-coded: Type 1 compares one-handed signs to the grand mean; Type 2 compares symmetrical two-handed signs to the grand mean. Continuous predictors were *z*-scored.

Predictor	Estimate	SE	z	p
(Intercept)	3.66	0.46	7.93	<.001
Body-part-CL (vs. reference)	0.04	0.62	0.07	.944
Entity-CL (vs. reference)	−0.27	0.53	−0.5	.617
Handling-CL (vs. reference)	−0.64	0.58	−1.1	.272
Iconicity	1.13	0.23	5.04	<.001
Concreteness	−0.48	0.19	−2.56	.010
Animacy	−0.01	0.47	−0.02	.986
Articulation Type 1	−0.37	0.24	−1.55	.121
Articulation Type 2	−0.06	0.23	−0.26	.798
Sign Length	−0.05	0.16	−0.29	.775

or associated activity. These action-for-instrument and action-for-activity mappings have been described as metonymic mappings in sign languages (Wilcox et al., 2003). In spoken-language comprehension, metonymic mappings can carry measurable processing costs (Traxler et al., 2002; Weiland et al., 2014). By analogy, such action-to-referent mappings may have contributed to the larger pupillary response effects observed for handling-CL signs. On this interpretation, metonymic mapping may be one component of the broader processing complexity associated with handling classifiers.

Finally, higher iconicity was associated with reduced pupil dilation, consistent with facilitation from more transparent form–meaning mappings (e.g., Vinson et al., 2008; Thompson et al., 2009; Ortega, 2017). However, classifier-containing signs still elicited larger pupillary responses than no-CL signs when iconicity was controlled, suggesting that iconicity can reduce, but does not fully account for, processing demands associated with classifier-containing signs. Of course, pupillary response is not specific to morphological or classifier-related processing and may also reflect perceptual, attentional, lexical, or decision-related demands that our models did not fully isolate.

Accuracy and response time (RT)

In the baseline accuracy model, body-part-CL signs were recognized more accurately than no-CL signs; entity-CL and handling-CL signs did not differ significantly from the reference level. When iconicity was included, body-part-CL effects were attenuated to non-significance, while iconicity emerged as a robust positive predictor and concreteness as a negative predictor (regarding negative concreteness effects, see Kousta et al., 2011; Thierfelder et al., 2024).

In the baseline RT model, body-part-CL signs elicited faster responses than no-CL signs; entity-CL and handling-CL signs did not differ from the reference level. When iconicity was included, the body-part-CL advantage was no longer significant, and iconicity emerged as a significant facilitator of faster responses. Critically, handling-CL signs showed significantly slower RTs than no-CL signs once iconicity was controlled, suggesting a residual processing cost that was not apparent in the baseline model.

Notably, handling-CL and body-part-CL signs had comparably high iconicity ratings (Ms = 5.41 and 5.09, respectively; p = .650), yet the two categories showed divergent patterns once iconicity was controlled: the body-part-CL advantage was eliminated, whereas handling-CL signs revealed a processing cost. This suggests that the apparent body-part-CL RT advantage is largely accounted for by iconicity, or by correlated properties captured by the iconicity ratings, rather than by classifier category per se. By contrast, handling-CL signs appear to impose additional demands that their high iconicity does not fully compensate for, consistent with the integration demands discussed above for the pupillary data. This pattern parallels spoken-language findings where morphologically complex forms incur costs when other facilitatory factors are held constant (Kim et al., 2022; Rastle et al., 2015).

Timing of sign language acquisition

The timing of sign language acquisition was found to modulate iconicity effects for accuracy and the pupillometry time course. The iconicity × group (native versus non-native early signers) interaction was significant for accuracy: iconicity facilitated correct responses in both groups, but more strongly for non-native signers (AoA: M = 5.3 years, SD = 3.01, range = 1–11) compared to native signers (AoA = 0). This difference was driven by lower-iconicity signs, where native signers maintained relatively high accuracy, while non-native signers had slightly lower accuracy in sign identification.

The pupillometry results revealed group differences that emerged in two late windows (~1060–1542 ms and ~1949–2100 ms post-onset). In the earlier window, native signers showed reduced pupil dilation compared to non-native signers for high-iconicity signs and larger pupillary responses for low-iconicity signs. In the later window, the direction of the native iconicity pattern reversed, yielding the opposite native–non-native contrast (i.e., native signers showed greater pupil dilation for signs with higher iconicity and reduced pupil dilation for signs with lower iconicity). In contrast, for non-native signers, the iconicity effect on pupillary response was relatively constant over time, with more iconic signs associated with slightly reduced pupil dilation throughout the trial.

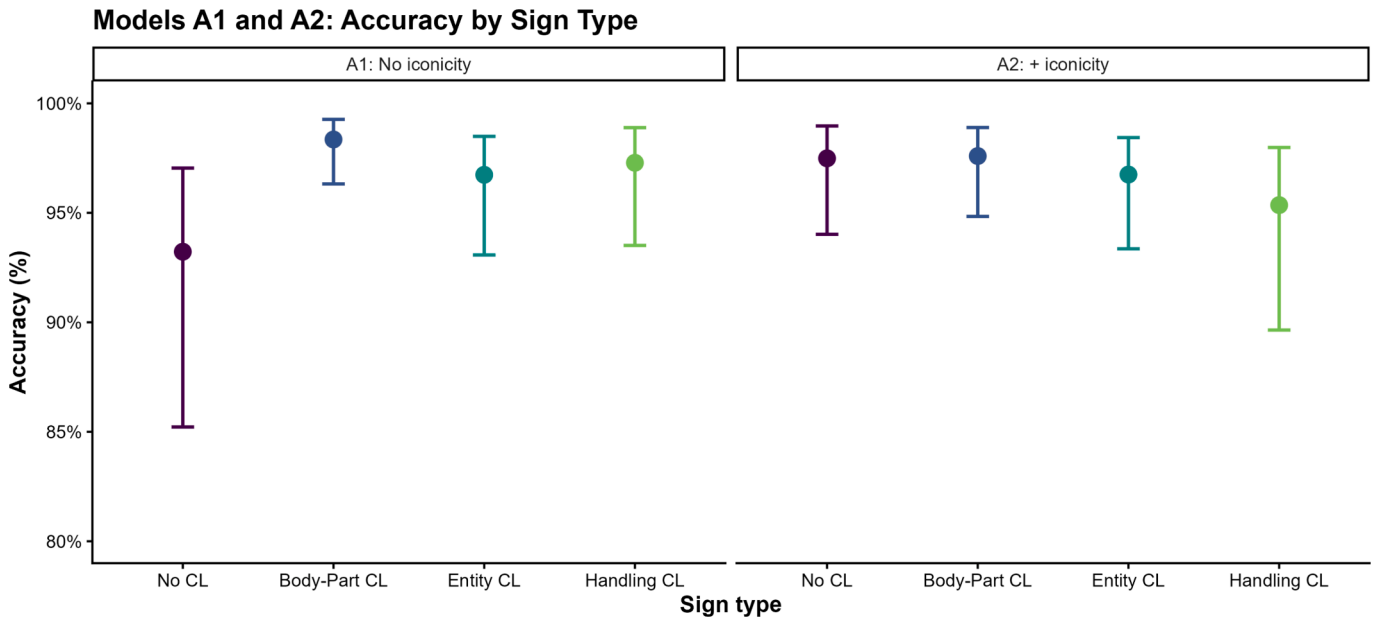


Fig. 13. Estimated marginal mean accuracy (response scale) by classifier type, comparing Model A1 (without iconicity covariate) and Model A2 (with iconicity covariate). Error bars indicate 95% confidence intervals. Estimates are marginalized over all other covariates.

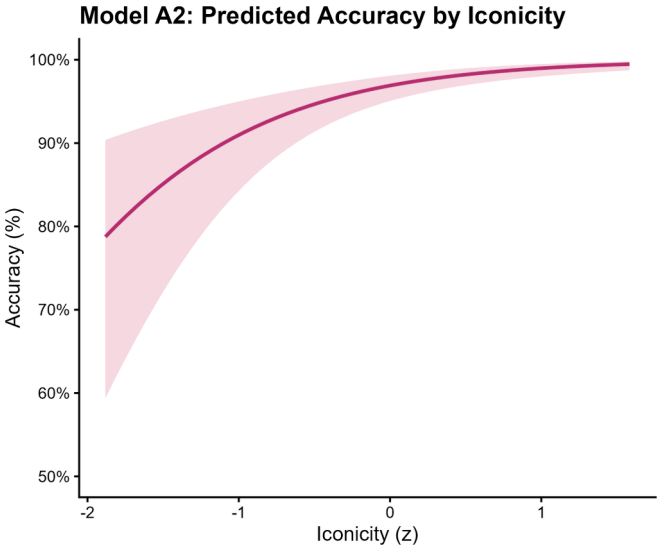


Fig. 14. Predicted accuracy as a function of iconicity (*z*-scored) from Model A2. Line represents estimated marginal mean on the response scale; shaded region indicates 95% confidence interval. Predictions are marginalized over all other covariates at reference values.

Overall, the data indicate that iconicity does not map onto a single, uniform pupillary response signature: non-native signers showed a relatively stable iconicity effect across time, whereas native signers showed a temporally dynamic pattern. One tentative interpretation is that, for native signers, high iconicity may reduce mid-latency processing effort but later engage additional decision-related evaluation in the lexical decision task, yielding the late reversal. For example, highly depictive, gesture-like forms may elicit increased late-stage evaluation of lexical status or heightened uncertainty before response in a lexical decision task (cf. Caselli et al., 2021). More generally, experience-dependent differences in how signers use iconic mappings during recognition have been reported in other paradigms (e.g., Anible, 2020), consistent with the view that acquisition history can shape the dynamics of iconicity effects.

Table 8
Fixed effects for the response time model without iconicity (Model R1). Reference level for classifier type is no-CL (signs without an embedded classifier). Articulation type is sum-coded: Type 1 compares one-handed signs to the grand mean; Type 2 compares symmetrical two-handed signs to the grand mean. Continuous predictors were *z*-scored.

Predictor	Estimate	SE	t	p
Intercept	1508.36	14.35	105.12	<.001
Body-part-CL (vs. reference)	−43.53	16.82	−2.59	.010
Entity-CL (vs. reference)	3.32	17.12	0.19	.846
Handling-CL (vs. reference)	−0.99	14.74	−0.07	.946
Concreteness	9.10	11.92	0.76	.446
Animacy	−78.40	13.36	−5.87	<.001
Articulation Type 1	30.81	13.07	2.36	.018
Articulation Type 2	−3.30	13.57	−0.24	.808
Sign Length	55.20	10.50	5.26	<.001

Table 9
Fixed effects for the response time model with iconicity (Model R2). Reference level for classifier type is no-CL (signs without an embedded classifier). Articulation type is sum-coded: Type 1 compares one-handed signs to the grand mean; Type 2 compares symmetrical two-handed signs to the grand mean. Continuous predictors were *z*-scored.

Predictor	Estimate	SE	t	p
Intercept	1474.23	15.53	94.91	<.001
Body-part-CL (vs. reference)	−21.32	17.98	−1.19	.236
Entity-CL (vs. reference)	20.72	14.69	1.41	.158
Handling-CL (vs. reference)	34.19	14.86	2.30	.021
Iconicity	−37.39	11.55	−3.24	.001
Concreteness	19.60	12.21	1.61	.109
Animacy	−74.92	16.14	−4.64	<.001
Articulation Type 1	30.63	14.04	2.18	.029
Articulation Type 2	−3.35	13.36	−0.25	.802
Sign Length	56.35	11.40	4.94	<.001

Prior work has yielded mixed results on the timing and even the direction of iconicity effects in sign recognition (Caselli et al., 2021; Emmorey et al., 2020; Zhang et al., 2023). These differences likely reflect a combination of factors (task demands, language, stimulus sets, and participant characteristics). Our exploratory analyses align with

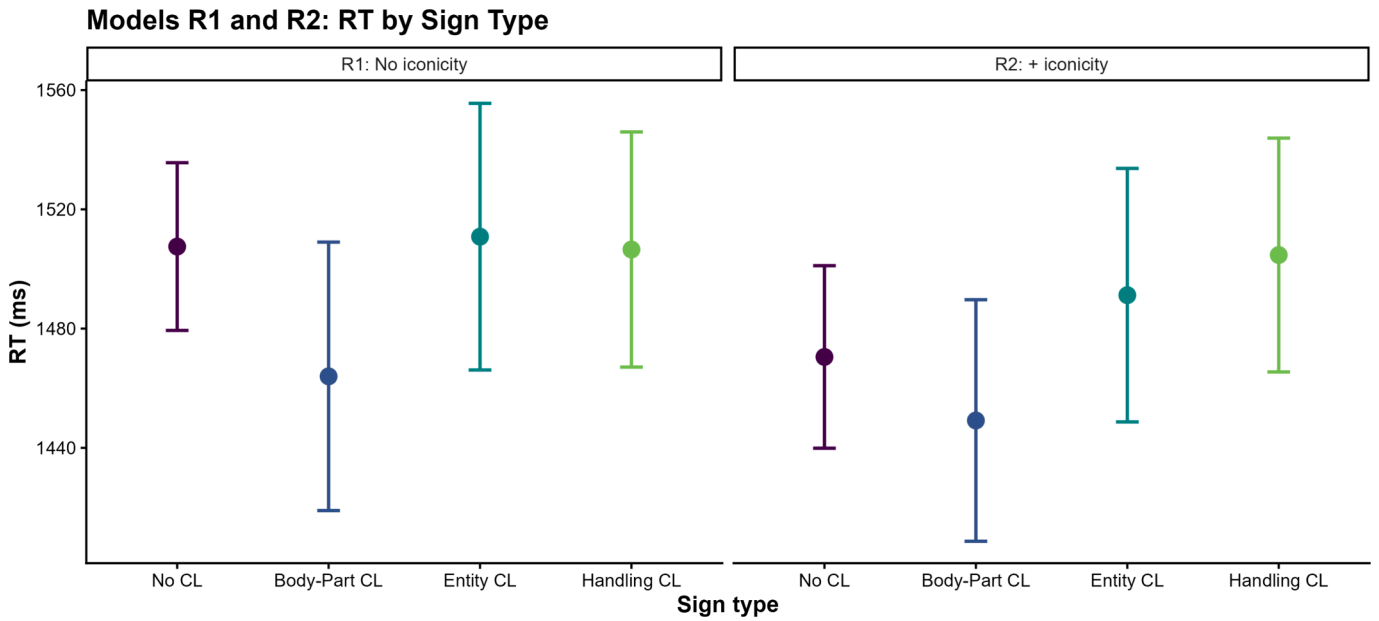


Fig. 15. Estimated marginal mean response times by classifier type, comparing Model R1 (without iconicity covariate) and Model R2 (with iconicity covariate). Error bars indicate 95% confidence intervals. Estimates are marginalized over all other covariates.

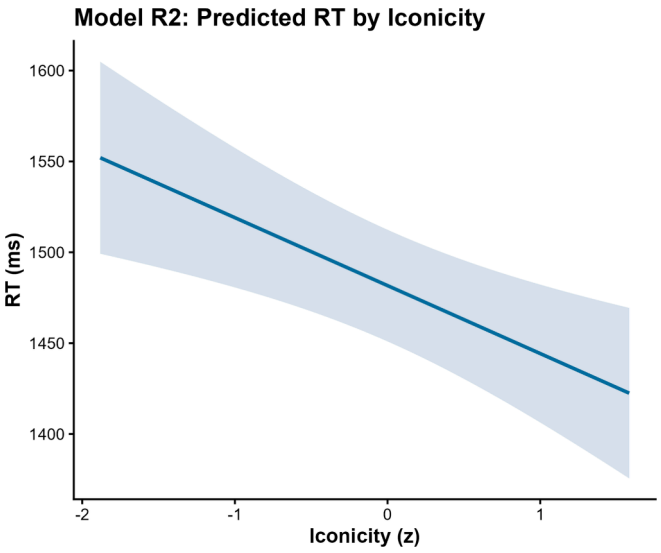


Fig. 16. Predicted response time as a function of iconicity (z-scored) from Model R2. Line represents estimated marginal mean; shaded region indicates 95% confidence interval. Predictions are marginalized over all other covariates at reference values.

prior work in indicating that acquisition history is one such factor: native and non-native signers showed different time courses of the iconicity–pupil relationship and a different magnitude of the iconicity benefit for accuracy, though there was no such effect in RT. This pattern supports the view that iconicity effects are not uniform across signers and may depend on language experience.

Synthesis and theoretical implications

Across measures, the results suggest two partially independent influences on lexical sign recognition: facilitation from iconic form–meaning mappings and increased processing demands associated with classifier-related internal structure. Mechanistically, we propose that iconicity facilitates recognition by providing form-based cues to meaning during lexical access, whereas classifier-containing signs may require additional integration of meaningful handshape information with other morphophonological features of the sign, as well as with whole-sign meaning.

Iconicity predicted faster response times, higher accuracy, and reduced pupil dilation. At the same time, signs containing embedded classifier handshapes elicited greater pupil dilation than signs without classifier handshapes, and handling-CL signs showed a residual RT cost after controlling for iconicity. Notably, entity-CL signs did not differ from no-CL signs in accuracy or RT, yet still elicited greater pupil dilation, suggesting that pupillometry captures processing demands that do not always surface in behavioral outcomes. Together, these findings indicate that iconicity and classifier-related structure make separable contributions to lexical sign recognition: iconicity facilitates recognition, whereas classifier-related structure in lexical signs can impose additional processing demands even when behavioral performance is unaffected.

This pattern is consistent with an interactive, partially decompositional account of sign recognition, in which lexical access draws on both whole-sign representations and meaningful sublexical structure. Visually available parameters such as handshape, movement, location, orientation, and non-manual features may contribute to lexical activation while also serving, in some signs, as meaning-bearing components of the sign. For signs with embedded classifier handshapes, this internal structure may include meaning contributed by the classifier handshape, consistent with analyses of classifier handshapes as meaningful units in sign languages (Koenders, 2024; Zwitserlood, 2003, 2008). The decompositional component of this interpretation is also consistent with broader accounts of lexical processing, in which complex forms are processed, at least in part, through their internal morphological structure (Taft, 2004; Rastle & Davis, 2008; Stockall & Marantz, 2006; Solomyak & Marantz, 2010). The pattern is not well explained by a strictly holistic or iconicity-only account, which would not predict additional processing effects associated with classifier-containing signs beyond what iconicity captures. Nor is it explained by a structure-only account, which would not predict the independent facilitatory effect of iconicity across measures.

On this account, recognition involves the concurrent use of whole-sign and sublexical information. As a sign's articulation unfolds, perceivers use early-available phonological cues to narrow down lexical

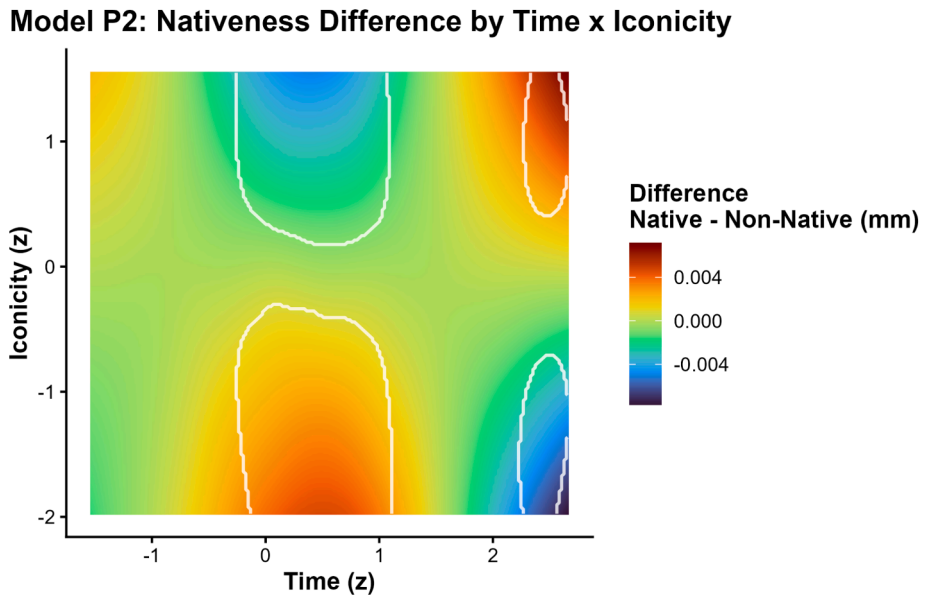


Fig. 17. Difference surface showing the estimated Native − Non-Native contrast in the time × iconicity interaction from Model P2. Warmer colors indicate greater predicted baseline-corrected pupil change for native signers than non-native signers; cooler colors indicate greater predicted pupil change for non-native signers than native signers. White contours indicate regions where simultaneous 95% confidence intervals exclude zero.

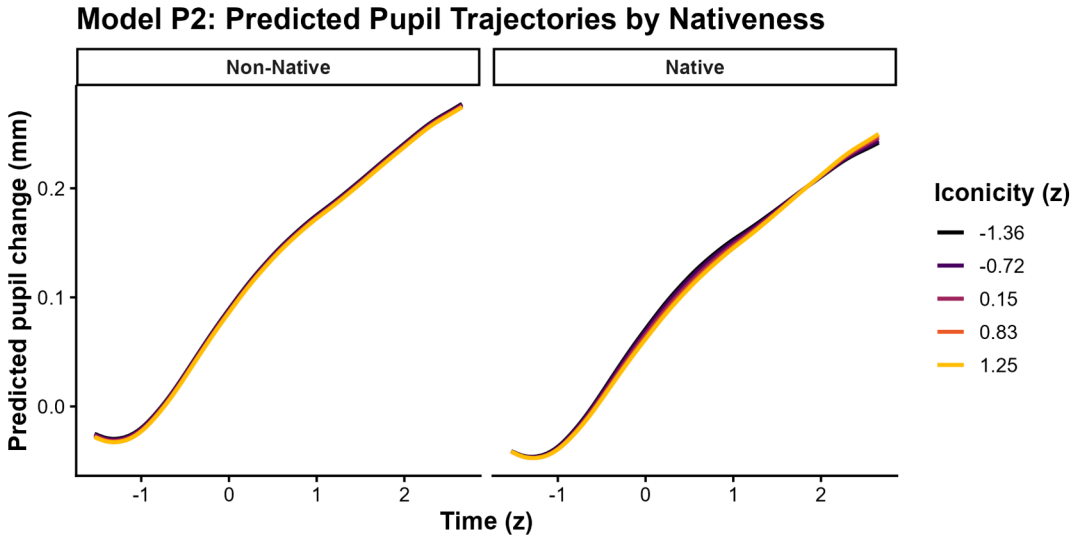


Fig. 18. Population-level model predictions of baseline-corrected pupil change over time for non-native signers (left) and native signers (right) at five iconicity quantiles: 10th, 30th, 50th, 70th, and 90th percentiles of iconicity *z*-scores. Predictions are marginalized over classifier type and articulation type using empirical weights, with other covariates held at their centered or reference values and participant/item random effects excluded.

candidates (Emmorey & Corina, 1990; Grosjean, 1981), while also mapping aspects of the perceived form onto conceptual representations. Consistent with structure-mapping theory, iconicity may facilitate recognition when the perceptual structure of a sign aligns with salient aspects of the referent or event representation (Emmorey, 2014; Gentner, 1983; Gentner & Markman, 1997; see also Perniss et al., 2010). For classifier-containing signs, however, classifier handshapes introduce additional meaningful elements that must be integrated with other sign parameters during online recognition—a process that may draw on additional cognitive resources even when iconicity supports the broader form–meaning mapping. The observed processing profile thus reflects the balance between iconic facilitation and morphophonological integration cost.

Finally, group effects in the exploratory analysis suggest that the contribution of iconicity to sign recognition may be modulated by linguistic experience. Non-native early signers showed a stronger accuracy benefit from iconicity, consistent with greater reliance on whole-sign form–meaning mappings. Native signers showed a more temporally dynamic pattern of iconicity effects in the pupil data, which may reflect differences in how iconic and sublexical information are recruited during recognition. This interpretation remains tentative, but it is consistent with prior work showing that signer experience and age of acquisition can modulate iconicity effects during sign processing (Anible, 2020; Mott et al., 2020; Zhang et al., 2023). Overall, the findings support a hybrid account in which lexical sign recognition draws on both iconic form–meaning mappings and classifier-related morphophonological structure, with the role of iconicity shaped in part by linguistic experience.

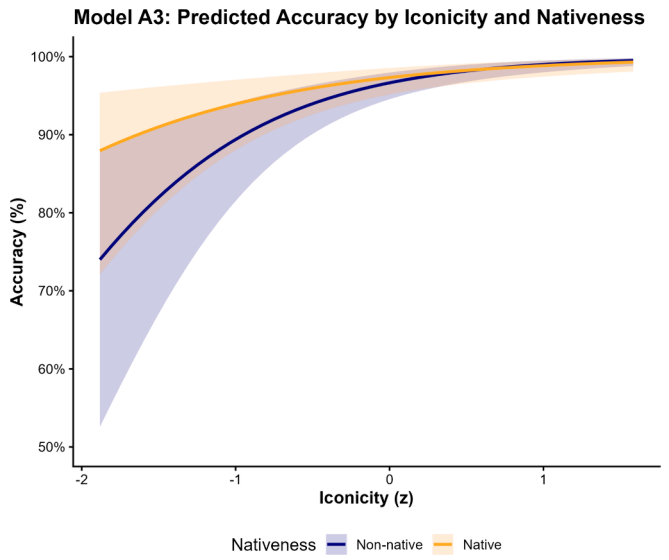


Fig. 19. Predicted accuracy as a function of iconicity for native and non-native signers from Model A3. Lines represent estimated marginal means; shaded regions indicate 95% confidence intervals. Predictions are averaged over classifier type and articulation type.

Limitations and future directions

Power and design constraints limit the strength and scope of some inferences. Although the sample ($N = 43$) provided sufficient sensitivity for the pupillometry analyses, power was low for accuracy and the native subgroup was relatively small ($n = 13$), so nativeness interactions and null effects, especially in accuracy, should be interpreted cautiously and tested in larger, more balanced replications.

A second limitation is task-specificity. An unprimed lexical decision task with trial-by-trial feedback likely blends lexical access with decision and verification processes, so some RT and pupil effects may reflect late-stage checking rather than recognition per se. Future work should evaluate whether the same patterns hold in tasks with reduced metalinguistic decision demands, no trial-by-trial feedback, or more implicit measures of lexical access.

Third, sign type was not a fully controlled manipulation of structural complexity. Rather, we used body-part-, entity-, and handling-CL status as an index of classifier-related morphophonological structure that may involve different integration demands. Although we controlled for several lexical and visual covariates, classifier categories may covary with dimensions not isolated here, such as gesture-likeness, action- or event-to-referent mapping, visual salience, or the analyzability of individual morphophonological components. Future studies that manipulate these dimensions more directly would provide stronger tests of the classifier-integration interpretation.

Finally, while pupillometry provides a sensitive online index of processing effort, it is not process-specific: larger dilations for classifier-containing signs are consistent with added integration demands but could also reflect other factors, such as attention, perceptual difficulty, or decision uncertainty. Follow-up studies pairing pupillometry with complementary methods, such as EEG/MEG, targeted experimental manipulations, or fine-grained analyses of pseudosign-rejection errors, will be important for isolating the mechanisms underlying these effects.

Conclusion

This study examined how embedded classifier handshapes and iconicity contribute to lexical recognition of HKSL signs, using accuracy, response time (RT), and task-evoked pupillary responses in a lexical decision task to capture both behavioral outcomes and processing effort. Across measures, iconicity facilitated lexical decision performance, whereas classifier-containing signs—especially handling-CL signs—were associated with increased processing effort, most clearly reflected in sustained pupillary responses and, when iconicity was controlled, in RT. Language experience further appeared to shape these dynamics, with native and non-native signers showing different time courses of iconicity effects in pupillometry.

We propose that iconicity and classifier-related structure reflect partially independent components of an interactive recognition process: iconicity facilitates form–meaning mapping, whereas classifier-related morphophonological structure may increase the demands of integrating multiple simultaneous sublexical elements into a coherent lexical representation. Together, these findings suggest that models of lexical processing in sign languages should incorporate both iconic form–meaning mappings and the morphophonological structure of lexical signs.

CRediT authorship contribution statement

Philip Thierfelder: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Emily Koenders:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the Department of Linguistics and Modern Languages, The Chinese University of Hong Kong. We are grateful to all of the participants for generously contributing their time and energy to this research. We especially thank Sandy Xu for allowing us to use her image in the experimental stimuli and for helping us create the pseudosigns through her native signing intuitions. We also thank Dr. Jessica Zhao for her tireless support, encouragement, and guidance throughout this project. Finally, we are deeply indebted to the Hong Kong Deaf community for teaching us Hong Kong Sign Language and for continually supporting our research.

Appendix

Table A1
Overview of the overlap and differences between existing classifier handshape typologies used in sign linguistics discourse.

Study	Sign Language	Proposed Classifier Types			
Koenders (2024)	*HKSL*	Entity Classifiers	Handling Classifiers	Body Part Classifiers	Other
McDonald (1982)	*ASL*	X-Type Object	Handle X-Type Object		
Supalla (1986)	*ASL*	SASS; Semantic; Tool	Instrumental	Body; Body Part	
Engberg-Pedersen (1993)	*DTS*	Whole Entity; Extension	Handle	Limb	
Valli and Lucas (2000)	*ASL*	Whole Entity; Depth-and-Width; Extent; Perimeter; On-Surface	Instrumental		Surface
Schembri (1996)	*Auslan*	Whole Entity; Collectives; Extent; Depth-and-Width; Perimeter-Shape	Handling (Holding; Touch; Instrumental)	Body Part	Surface
Wallin (1996)	*STS*	Non-Agentive	Agentive		
Zwitserlood (2003)	*NGT*	Entity (Incl. SASS)	Handling		
Benedicto and Brentari (2004)	*ASL*	Whole Entity; Extension	Handle	Limb	
Eccarius and Brentari (2007)	*ASL, HKSL, DSGS*	Whole Entity	Handle	Limb	
Tang (2007)	*HKSL*	SASS; Semantic; Tracing	Handling; Touch	Body	
He (2011)	*TJSL*	Whole Entity	Handling		

Data availability

The data, materials, analysis code, and power analysis are available on OSF at https://osf.io/d8c2r/.

References

Aikhenvald, A. Y. (2000). *Classifiers: A Typology of Noun Categorization Devices*. Oxford University Press.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected papers of Hirotugu Akaike* (pp. 199–213). Springer.

Anible, B. (2020). Iconicity in American Sign Language–English translation recognition. *Language and Cognition, 12*(1), 138–163. https://doi.org/10.1017/langcog.2019.51

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Battison, R. (1978). *Lexical borrowing in American Sign Language*. Linstok Press.

Baus, C., Carreiras, M., & Emmorey, K. (2013). When does iconicity in sign language matter? *Language and Cognitive Processes, 28*(3), 261–271.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91*(2), 276–292.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 142–162). Cambridge University Press.

Benedicto, E., & Brentari, D. (2004). Where did all the arguments go?: Argument-changing properties of classifiers in ASL. *Natural Language & Linguistic Theory, 22*(4), 743–810. https://doi.org/10.1007/s11049-003-4698-2

Benedicto, E., Cvejanov, S., & Quer, J. (2007). Valency in classifier predicates: A syntactic analysis. *Lingua, 117*(7), 1202–1215.

Bertram, R., Hyönä, J., & Laine, M. (2000). The role of context in morphological processing: Evidence from Finnish. *Language and Cognitive Processes, 15*(4–5), 367–388. https://doi.org/10.1080/01690960050119634

Bosworth, R. G., & Emmorey, K. (2010). Effects of iconicity and semantic relatedness on lexical access in American Sign Language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(6), 1573–1581.

Bradski, G. (2000). The OpenCV library. *Dr Dobb's Journal: Software Tools for the Professional Programmer, 25*(11), 120–123.

Brentari, D. (1998). *A prosodic model of sign language phonology*. MIT Press.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal, 9*(2), 378–400.

Bross, F. (2024). What is iconicity?: The view from sign languages. *Sign Language & Linguistics, 27*(1), 73–102. https://doi.org/10.1075/sll.22003.bro

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition, 1*(1).

Caselli, N. K., Emmorey, K., & Cohen-Goldberg, A. M. (2021). The signed mental lexicon: Effects of phonological neighborhood density, iconicity, and childhood language experience. *Journal of Memory and Language, 121*, Article 104282. https://doi.org/10.1016/j.jml.2021.104282

Caselli, N. K., & Pyers, J. E. (2020). Degree and not type of iconicity affects sign language vocabulary acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(1), 127–139. https://doi.org/10.1037/xlm0000713

Caselli, N. K., Sehyr, Z. S., Cohen-Goldberg, A. M., & Emmorey, K. (2017). ASL-LEX: A lexical database of American Sign Language. *Behavior Research Methods, 49*(2), 784–801. https://doi.org/10.3758/s13428-016-0742-0

Centre for Sign Linguistics and Deaf Studies. (n.d.). Asian SignBank. The Chinese University of Hong Kong. https://www.cslds.org/asb/.

Durrant, P. (2013). Formulaicity in an agglutinating language: The case of Turkish. *Corpus Linguistics and Linguistic Theory, 9*(1), 1–38. https://doi.org/10.1515/cllt-2013-0009

Eccarius, P., & Brentari, D. (2007). Symmetry and dominance: A cross-linguistic study of signs and classifier constructions. *Lingua, 117*(7), 1169–1201.

Embick, D. (2021). The motivation for roots in distributed Morphology. *Annual Review of Linguistics, 7*, 69–88. https://doi.org/10.1146/annurev-linguistics-040620-061341

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

Emmorey, K., & Corina, D. (1990). Lexical recognition in sign language: Effects of phonetic structure and morphology. *Perceptual and Motor Skills, 71*, 1227–1252.

Emmorey, K., Winsler, K., Midgley, K. J., Grainger, J., & Holcomb, P. J. (2020). Neurophysiological correlates of frequency, concreteness, and iconicity in American Sign Language. *Neurobiology of Language, 1*(2), 249–267. https://doi.org/10.1162/nol_a_00012

Engberg-Pedersen, E. (1993). *Space in Danish Sign Language: The Semantics and Morphosyntax of the use of Space in a Visual Language*. Signum Press.

Farnebäck, G. (2003). Two-frame motion estimation based on polynomial expansion. In J. Bigun, & T. Gustavsson (Eds.), *Image analysis* (Vol. 2749, pp. 363–370). Springer. https://doi.org/10.1007/3-540-45103-X_50.

Forbes, S. (2020). PupillometryR: An R package for preparing and analysing pupillometry data. *Journal of Open Source Software, 5*(50), Article 2285. https://doi.org/10.21105/joss.02285.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*(2), 155–170.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*(1), 45–56.

Glück, S., & Pfau, R. (1998). On classifying classification as a class of inflection in German Sign Language. *Proceedings of ConSole, 6*(1), 59–74.

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Grosjean, F. (1981). Sign and word recognition: A first comparison. *Sign Language Studies, 32*(1), 195–220.

Halle, M. (1990). *An approach to morphology. NELS, 20*, 150–184.

Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale, & S. J. Keyser (Eds.), *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger* (pp. 111–176). MIT Press.

He, J. (2011). Instrument classifier predicates in Tianjin Sign Language [Master's thesis, the Chinese University of Hong Kong]. *CUHK Electronic Theses & Dissertations Collection.*

Hedlund, L. A., Wikman, P., Hut, S. C. A., & Leminen, A. (2021). Neural responses to Finnish inflected forms during overt and covert production: The role of stem frequency and stem allomorphy. *Journal of Neurolinguistics, 57*, Article 100953. https://doi.org/10.1016/j.jneuroling.2020.100953

Hwang, S.-O., Tomita, N., Morgan, H., Ergin, R., İlkbaşaran, D., Seegers, S., Lepic, R., & Padden, C. (2017). Of the body and the hands: Patterned iconicity for semantic categories. *Language and Cognition, 9*(4), 573–602. https://doi.org/10.1017/langcog.2016.28

Hyönä, J., & Pollatsek, A. (2000). Processing of Finnish compound words in reading. In: A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), Reading as a perceptual process (pp. 65–87). Elsevier. https://doi.org/10.1016/B978-008043642-5/50006-1.

Hyönä, J., Vainio, S., & Laine, M. (2002). A morphological effect obtains for isolated words but not for words in sentence context. *European Journal of Cognitive Psychology, 14*(4), 417–433. https://doi.org/10.1080/09541440143000131

Johnston, T., & Schembri, A. (2007). *Australian Sign Language (Auslan): An introduction to sign language linguistics*. Cambridge University Press.

Kim, J., Kang, J., Kim, J., & Nam, K. (2022). Temporal dynamics of form and meaning in morphologically complex word processing: An ERP study on Korean inflected verbs. *Journal of Neurolinguistics, 64*, Article 101098. https://doi.org/10.1016/j.jneuroling.2022.101098

Kimmelman, V., Pfau, R., & Aboh, E. O. (2020). Argument structure of classifier predicates in Russian Sign Language. *Natural Language & Linguistic Theory, 38*(2), 539–579. https://doi.org/10.1007/s11049-019-09448-9

Koenders, E. (2024). Noun classifiers in Hong Kong Sign Language. *Sign Language & Linguistics, 27*(1), 35–72. https://doi.org/10.1075/sll.22004.koe

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General, 140*(1), 14–38. https://doi.org/10.1037/a0021446

Kuchinke, L., Võ, M.-L.-H., Hofmann, M., & Jacobs, A. M. (2007). Pupillary responses during lexical decisions vary with word frequency but not emotional valence. *International Journal of Psychophysiology, 65*(2), 132–140.

Kyröläinen, A.-J., Porretta, V., & Järvikivi, J. (2015). Pupillometry as a window to real time processing of morphologically complex verbs. *New Developments in the Quantitative Study of Languages*, 45–46.

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science, 7*(1), 18–27.

Lehtonen, M., Cunillera, T., Rodríguez-Fornells, A., Hultén, A., Tuomainen, J., & Laine, M. (2007). Recognition of morphologically complex words in Finnish: Evidence from event-related potentials. *Brain Research, 1148*, 123–137. https://doi.org/10.1016/j.brainres.2007.02.026

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology, 6*, Article 1171. https://doi.org/10.3389/fpsyg.2015.01171

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition, 1*(1).

McDonald, B. H. (1982). Aspects of the American Sign Language predicate system [Doctoral dissertation, State University of New York at Buffalo]. ProQuest Dissertations and Theses Global.

McGarry, M. E., Massa, N., Mott, M., Midgley, K. J., Holcomb, P. J., & Emmorey, K. (2021). Matching pictures and signs: An ERP study of the effects of iconic structural alignment in American Sign Language. *Neuropsychologia, 162*, Article 108051. https://doi.org/10.1016/j.neuropsychologia.2021.108051

McGarry, M. E., Mott, M., Midgley, K. J., Holcomb, P. J., & Emmorey, K. (2021). Picture-naming in American Sign Language: An electrophysiological study of the effects of iconicity and structured alignment. *Language, Cognition and Neuroscience, 36*(2), 199–210. https://doi.org/10.1080/23273798.2020.1804601

Meir, I., Padden, C. A., Aronoff, M., & Sandler, W. (2007). Body as subject. *Journal of Linguistics, 43*(3), 531–563.

Meir, I., & Tkachman, O. (2018). Iconicity. In Oxford research encyclopedia of linguistics. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.343.

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

Mott, M., Midgley, K. J., Holcomb, P. J., & 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.

Nicklin, C., & Plonsky, L. (2020). Outliers in L2 research in applied linguistics: A synthesis and data re-analysis. *Annual Review of Applied Linguistics, 40*, 26–55.

Ortega, G. (2017). Iconicity and sign lexical acquisition: A review. *Frontiers in Psychology, 8*.

Padden, C., Hwang, S.-O., Lepic, R., & Seegers, S. (2015). Tools for language: Patterned iconicity in sign language nouns and verbs. *Topics in Cognitive Science, 7*(1), 81–94. https://doi.org/10.1111/tops.12121

Padden, C., Meir, I., Aronoff, M., & Sandler, W. (2010). The grammar of space in two new sign languages. In D. Brentari (Ed.), *Sign languages* (pp. 570–592). Cambridge University Press. https://doi.org/10.1017/CBO9780511712203.026.

Pavlič, M. (2016). The word order parameter in Slovenian Sign Language: Transitive, ditransitive, classifier and locative constructions [Doctoral dissertation, Ca' Foscari University of Venice].

Perniss, P., Thompson, R. L., & 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

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1651). https://doi.org/10.1098/rstb.2013.0300

Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes, 23*(7–8), 942–971. https://doi.org/10.1080/01690960802069730

Rastle, K., Lavric, A., Elchlepp, H., & Crepaldi, D. (2015). Processing differences across regular and irregular inflections revealed through ERPs. *Journal of Experimental Psychology: Human Perception and Performance, 41*(3), 747–760. https://doi.org/10.1037/a0039150

Rojas, C., Vega-Rodríguez, Y. E., Lagos, G., Cabrera-Miguieles, M. G., Sandoval, Y., & Crisosto-Alarcón, J. (2024). Applicability and usefulness of pupillometry in the study of lexical access: A scoping review of primary research. *Frontiers in Psychology, 15*. https://doi.org/10.3389/fpsyg.2024.1372912

Sandler, W., & Lillo-Martin, D. (2006). *Sign language and linguistic universals*. Cambridge University Press.

Schembri, A. C. (1996). *The structure and formation of signs in Auslan*. North Rocks Press.

Schembri, A. (2003). Rethinking "classifiers" in signed languages. In K. Emmorey (Ed.), *Perspectives on classifier constructions in sign languages* (pp. 3–34). Lawrence Erlbaum Associates.

Schlenker, P., Bonnet, M., Lamberton, J., Lamberton, J., Chemla, E., Santoro, M., & Geraci, C. (2024). Iconic syntax: Sign language classifier predicates and gesture sequences. *Linguistics and Philosophy, 47*(1), 77–147.

Schmidtke, J. (2018). Pupillometry in linguistic research: An introduction and review for second language researchers. *Studies in Second Language Acquisition, 40*(3), 529–549.

Solomyak, O., & Marantz, A. (2010). Evidence for early morphological decomposition in visual word recognition. *Journal of Cognitive Neuroscience, 22*(9), 2042–2057.

SR Research Ltd. (2020). *EyeLink 1000 Plus* [Apparatus]. SR Research Ltd. https://www.sr-research.com/eyelink-1000-plus/.

SR Research Ltd. (2024b). *EyeLink Data Viewer* (Version 5.1.1) [Computer software]. https://www.sr-research.com/data-viewer/.

SR Research Ltd. (2024a). *Experiment Builder* (Version 2.6.11) [Computer software]. https://www.sr-research.com/experiment-builder/.

SR-Support (2020, September 3). *FAQ: How can I convert pupil size to mm?* [Online forum post]. SR Support Forum. https://www.sr-research.com/support/thread-154.html.

Stockall, L., & Marantz, A. (2006). A single route, full decomposition model of morphological complexity: MEG evidence. *The Mental Lexicon, 1*(1), 85–123. https://doi.org/10.1075/ml.1.1.07sto

Supalla, T. (1986). The classifier system in American Sign Language. In C. Craig (Ed.), Noun classes and categorization: Proceedings of a symposium on categorization and noun classification, Eugene, Oregon, October 1983 (pp. 181–214). John Benjamins.

Supalla, T. (2003). Revisiting visual analogy in ASL classifier predicates. In K. Emmorey (Ed.), *Perspectives on classifier constructions in sign languages* (pp. 249–257). Lawrence Erlbaum Associates.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology Section A, 57*(4), 745–765. https://doi.org/10.1080/02724980343000477

Tang, G. (2007). *Hong Kong Sign Language: A trilingual dictionary with linguistic descriptions*. Chinese University Press.

Taub, S. F. (2001). *Language from the body: Iconicity and metaphor in American Sign Language*. Cambridge University Press.

Thierfelder, P., Cai, Z. G., Huang, S., & Lin, H. (2024). The Chinese lexicon of deaf readers: A database of character decisions and a comparison between deaf and hearing readers. *Behavior Research Methods, 56*(6), 5732–5753. https://doi.org/10.3758/s13428-023-02305-z

Thierfelder, P., Durantin, G., & Wigglesworth, G. (2020). The effect of word predictability on phonological activation in Cantonese reading: A study of eye-fixations and pupillary response. *Journal of Psycholinguistic Research, 49*(5), 779–801. https://doi.org/10.1007/s10936-020-09713-8

Thompson, B., Perlman, M., Lupyan, G., Sehyr, Z. S., & Emmorey, K. (2020). A data-driven approach to the semantics of iconicity in American Sign Language and English. *Language and Cognition, 12*(1), 182–202. https://doi.org/10.1017/langcog.2019.52

Thompson, R. L., Vinson, D. P., & Vigliocco, G. (2009). The link between form and meaning in American Sign Language: Lexical processing effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(2), 550–557. https://doi.org/10.1037/a0014547

Traxler, M. J., Pickering, M. J., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory and Language, 47*(4), 530–547. https://doi.org/10.1016/S0749-596X(02)00021-9

Valli, C., & Lucas, C. (2000). *Linguistics of American Sign Language: An introduction* (3rd ed.). Gallaudet University Press.

Van Engen, K. J., & McLaughlin, D. J. (2018). Eyes and ears: Using eye tracking and pupillometry to understand challenges to speech recognition. *Hearing Research, 369*, 56–66. https://doi.org/10.1016/j.heares.2018.04.013

Vergara-Martínez, M., Duñabeitia, J. A., Laka, I., & Carreiras, M. (2009). ERP correlates of inhibitory and facilitative effects of constituent frequency in compound word reading. *Brain Research, 1257*, 53–64. https://doi.org/10.1016/j.brainres.2008.12.040

Vinson, D. P., Cormier, K., Denmark, T., Schembri, A., & Vigliocco, G. (2008). The British Sign Language (BSL) norms for age of acquisition, familiarity, and iconicity. *Behavior Research Methods, 40*(4), 1079–1087. https://doi.org/10.3758/BRM.40.4.1079

Vinson, D. P., Thompson, R. L., Skinner, R., & Vigliocco, G. (2015). A faster path between meaning and form? Iconicity facilitates sign recognition and production in British Sign Language. *Journal of Memory and Language, 82*, 56–85. https://doi.org/10.1016/j.jml.2015.03.002

Wallin, L. (1996). Polysynthetic signs in Swedish Sign Language (D. Miller, Trans.). Department of Linguistics & Department of Sign Language, Stockholm University.

Weiland, H., Bambini, V., & Schumacher, P. B. (2014). The role of literal meaning in figurative language comprehension: Evidence from masked priming ERP. *Frontiers in Human Neuroscience, 8*. https://doi.org/10.3389/fnhum.2014.00583

Wilcox, S., Wilcox, P. P., & Jarque, M. J. (2003). Mappings in conceptual space: Metonymy, metaphor, and iconicity in two signed languages. *Jezikoslovlje, 4*(1), 139–156.

Winter, B., & Perlman, M. (2021). Iconicity ratings really do measure iconicity, and they open a new window onto the nature of language. *Linguistics Vanguard, 7*(1). https://doi.org/10.1515/lingvan-2020-0135

Zhang, X., Cao, H., & Li, H. (2023). Neurophysiological effects of frequency, length, phonological neighborhood density, and iconicity on sign recognition. *Neuroreport, 34*(17), 817–824. https://doi.org/10.1097/WNR.0000000000001959

Zwitserlood, I. (2003). Classifying hand configurations in Nederlandse Gebarentaal (Sign Language of the Netherlands) [Doctoral dissertation, Radboud University Nijmegen]. LOT.

Zwitserlood, I. (2008). Morphology below the level of the sign: Frozen forms and classifier predicates. In J. Quer (Ed.), *Signs of the time: Selected papers from TISLR 8* (pp. 251–272). Signum Verlag.

Zwitserlood, I. (2012). Classifiers. In R. Pfau, M. Steinbach, & B. Woll (Eds.), *Sign language: An international handbook* (pp. 158–186). De Gruyter Mouton.