Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Interpretability in Sign Language Animal Signs

Permalink
https://escholarship.org/uc/item/7mw795rq

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Tkachman, Oksana
Sadlier-Brown, Emily
Lo, Roger Yu-Hsiang
et al.

Publication Date
2023

Peer reviewed

Cross-Linguistic Frequency and Interpretability in Sign Language Animal Signs

Oksana Tkachman
o.tkachman@alumni.ubc.ca

Emily Sadlier-Brown
emily.sadlier-brown@ubc.ca

Roger Yu-Hsiang Lo
roger.lo@ubc.ca

Carla Hudson Kam
carla.hudsonkam@ubc.ca

Department of Linguistics
University of British Columbia
Vancouver, BC, Canada

Abstract

The meanings of iconic signs are usually not easily accessible to sign-naïve people. However, most previous studies asked participants to guess the meaning of iconic signs in isolation and without any context or cues. We ask whether signs whose form is based on more cross-linguistically common underlying motivations are easier to interpret than signs based on less common underlying motivations. Since recent research suggests that iconicity is a relationship of resemblance between the signifier and the signified that is instantiated contextually, we also provide participants with a prompt (in the form of a word). We find that interpretability of iconic signs does correlate with cross-linguistic frequency of the underlying motivation of the sign.

Keywords: iconicity; interpretability; sign language; cross-linguistic frequency; underlying motivation

Introduction

Iconicity is a relationship of resemblance between the form and the meaning of a linguistic signal (word, sign, etc.). It has long been noted that sign languages have a large number of lexical forms that appear to be highly iconic (Stokoe, Casterline, & Croneberg, 1965; Wescott, 1971). For example, the American Sign Language (ASL) signs DRINK and SMOKE (see Figure 1) have a very obvious connection to the actions they depict and their meaning is easily recognizable even by sign-naïve people.[1] At the same time, signs are conventionalized forms—they are not idiosyncratic mimetic depictions—and different sign languages employ different signs for the same meaning. Sometimes these differences are subtle differences in form (e.g., different numbers of fingers used to make a similar shape), other times they are differences in the nature of the iconic depiction itself (e.g., depicting a cat using whiskers vs. ears). We are particularly interested in the latter kinds of differences, which can be seen as different iconic motivations for representing a single meaning. Depictions are not evenly spread across languages: some are much more frequent cross-linguistically (e.g., for cat signs, whiskers is much more common than ears [Tkachman & Hudson Kam, 2019]). In this study we probe this frequency difference, asking whether cross-linguistically preferred depictions are also preferred on another, very different, measure– the preferences of sign-naïve perceivers.

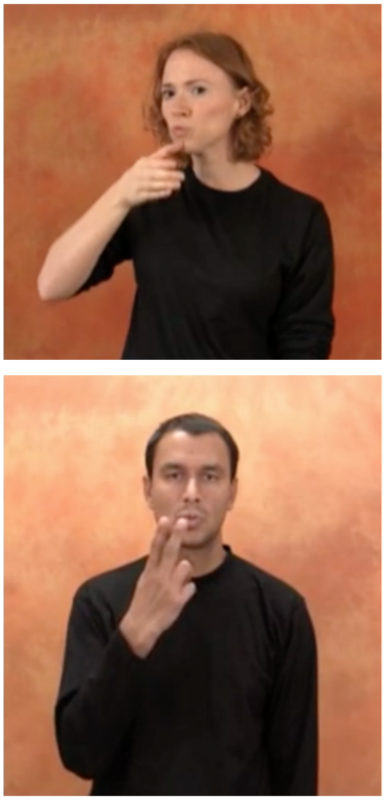

Figure 1: ASL signs for DRINK (top) and SMOKE (bottom). Images are from the online multilingual dictionary *Spread the Sign* (`www.spreadthesign.com`).

Extant studies have shown variation in the degree to which sign-naïve people can guess the meaning of different iconic signs, with some signs seeing near universal success and others much poorer performance (Grosso, 1993; Pizzuto & Volterra, 2000; Lai & Yang, 2009; Ortega, Schiefner, & Özyürek, 2019; Sehyr & Emmorey, 2019). These studies are important because they show that iconicity does not equal interpretability, reinforcing the fact that signs are conventionalized lexical items that function within a linguistic system. Moreover, they show that signs need not be transparently iconic to function perfectly well as signs within a language. Nevertheless, iconicity is a property of some signs, and we are interested in the constraints on this iconicity.

Do some depictions work better for conveying a particular

---

[1] We follow the convention of representing sign glosses with small caps.

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

meaning than others? In the present study we ask whether depictions that are the most common cross-linguistically are also more interpretable by perceivers. Specifically, we conducted a forced-choice study to see whether sign-naïve people asked to choose a sign that is most suitable for a provided prompt are more likely to choose signs based on iconic depictions that show up more frequently cross-linguistically than signs based on iconic depictions that show up less frequently. When we talk about depictions we mean the core idea or feature of the meaning that is iconically represented, for example, ears for cat or the act of holding a cigarette for smoking, not the specific handshape or movements involved. Henceforth, we will refer to these core ideas or features as underlying motivations.

Methods

Participants

A total of 105 participants, who stemmed from a variety of language backgrounds, were recruited at a North American public university. Participants received a course credit for their participation.

Stimuli

All signs used as stimuli came from the cross-linguistic online dictionary *Spread the Sign* (`www.spreadthesign.com`). Entries in the dictionary consist of short video clips of a sign. At the time of stimuli preparation, 33 languages were included in the dictionary.[2] Our stimuli were drawn from signs from a list of 20 animals, the same list as used in Tkachman and Hudson Kam (2019). The animals were: BAT, BEAR, BEE, CAT, CATERPILLAR, DINOSAUR, DOG, GOLDFISH, FROG, GORILLA, HORSE, KANGAROO, LION, OSTRICH, MOUSE, ROBIN, GIRAFFE, ALLIGATOR, SNAKE, WHALE. The video clip of the sign for each of the 20 animals was extracted for each language (although at the time of stimuli preparation, not every language in the dictionary had an entry for every animal on the list; for each of 20 animals, 6.15 of the 33 sign languages did not have a token in the corpus, range 3-12).

The coding for what underlying motivation(s) were represented in signs was done for a previous study (Tkachman & Hudson Kam, 2019). In that study, if, for instance, a sign for 'cat' represented a cat's whiskers, the underlying motivation was coded as 'whiskers'. Signs using alternative representational forms (such as fingerspelling, a type of borrowing from a surrounding spoken language) were excluded from analysis. The underlying motivations were then counted to determine which were most frequent. In cases where a sign incorporated two or more underlying motivations (i.e., a compound using two motivations in succession such as wings+fangs for BAT in Spanish Sign Language), both motivations were included. For each animal, the underlying motivations were then ranked by frequency, with the top ranked underlying motivation being the one that was represented by the largest number of languages. For each animal's ranked list, we then focused on the most common motivation, the second most common motivation and the least common motivation. In selecting the least common motivation, however, we excluded motivations that were only used once (so as to exclude potentially idiosyncratic innovations).

The most common motivation for each animal was labeled *T1* ("Target 1"), the second most common motivation was labeled *T2* ("Target 2") and the least common motivation was labeled *T3* ("Target 3"). This was to facilitate comparisons in which participants chose between cross-linguistically more and less common motivations. We also included comparisons in which participants chose between an attested sign for an animal and an unattested sign (i.e., a sign for a different animal). To facilitate these comparisons each animal was randomly paired with another animal on the list. The resulting pairings were fixed in this study so they were the same across all participants and trials. A sign representing the *T1*, *T2* and *T3* for each animal was chosen from among the sign languages, with attention paid to achieve approximately equal representation of the sign languages in the final set of stimuli.

Procedure

The experiment was held online using jsPsych v6.3.0 (de Leeuw, 2015). Participants completed the experiment on their own devices. Participants saw the following on-screen instructions: *In this experiment, you will be presented with a series of words accompanied by two videos each. The videos show a sign from some sign language. Please pick the video that best suits the meaning of the word provided.* Participants saw a word for an animal displayed in the lower centre of the screen. This is the target animal (i.e., the target meaning). The target meaning was provided in order to check if the form-meaning association can be successfully triggered in participants. On the left side of the screen, a video clip of a sign played automatically, followed by a video clip on the right side of the screen. Participants used the keyboard to make their selection. They had 5 seconds from the end of the second video to make their choice before the experiment advanced to the next trial. Trials were counter-balanced for which side of the screen the expected answer appeared on.

There were six comparison types in the study. For three comparison types, participants selected from two signs that were both attested forms for the target meaning. These are called target-target comparisons, since both signs on the

---

[2] The data came from the following 33 sign languages: American Sign Language; Argentinian Sign Language; Austrian Sign Language; Belorussian Sign Language; British Sign Language; Bulgarian Sign Language; Chilean Sign Language; Chinese Sign Language; Croatian Sign Language; Cuban Sign Language; Czech Sign Language; Estonian Sign Language; Finish Sign Language; German Sign Language; Greek Sign Language (Cyprus); Greek Sign Language (Greece); Icelandic Sign Language; Indian Sign Language; International Sign Language; Italian Sign Language; Japanese Sign Language; Latvian Sign Language; Lithuanian Sign Language; Mexican Sign Language; Polish Sign Language; Portuguese Sign Language; Romanian Sign Language; Russian Sign Language; Slovak Sign Language; Spanish Sign Language; Syrian Sign Language; Turkish Sign Language; Ukrainian Sign Language; Urdu Sign Language.

screen are attested forms for the target meaning. For the other three comparison types, participants selected from two signs, one of which was an attested sign for the target meaning, and one of which was a distractor, that is, a sign for a different animal (the randomly assigned pairing as described above). These are called target-distractor comparisons. All participants completed all trials; however, trials were blocked. Each block elicited only one target-target comparison and one target-distractor comparison, yielding three blocks. The block order was randomized by participant.

Predictions

Target-target comparisons allowed us to test whether underlying motivations that are more common cross-linguistically are also more commonly chosen as best suited for the word provided. In these trial types, there is no correct answer, since both signs are attested signs. In the two trial types which include *T1*, the expected choice is *T1*, the sign representing the most common underlying motivation cross-linguistically. Furthermore, we expect to see a higher preference rate in *T1 vs. T3* trials than in *T1 vs. T2* trials, because in *T1 vs. T2* trials we are comparing the most frequent and the second most frequent underlying motivations; that is, both motivations that, for whatever reason, resurface again and again in sign forms cross-linguistically. In *T1 vs. T3* trials, however, the comparison is between the most common and the least common ones, or, if commonality and participant preference have a shared source, the "best" choice of motivation and the "worst" choice. The starker comparison should lead to greater preference for T1, when the alternative is T3. In the remaining trial type, *T2 vs. T3*, the expected choice is *T2*, since this represents the second most common underlying motivation cross-linguistically (with *T3* forms being relatively rare, and therefore perhaps dispreferred by participants). We expected reaction times to follow this same pattern, with *T1 vs. T3* trials producing the quickest reaction times compared to *T1 vs. T2* trials and *T2 vs. T3* trials.

Target-distractor comparisons allowed us to see whether the frequency of underlying motivations, and the relationship between frequency and participants' choices in the *T vs. T* trials, are due to something about the adequacy of iconic representations. In these trials participants selected between attested animal signs and signs that referred to a different animal. For example, the word BEE is displayed and participants choose between an attested sign for BEE (either *T1*, *T2* or *T3*) and the most common sign for BEE's random pairing, for example the *T1* for BEAR. Because a sign for BEAR is not the intended target in this type of comparison, we labeled non-target signs (e.g., the BEAR sign when BEE was requested) *D1* ("Distractor 1") for these trial types. For these trial types, the expected choice is the attested sign for the target animal—*T1*, *T2*, or *T3*, depending on trial type. If all targets are good enough iconic representations, we would not expect any differences in performance across these *T vs. D* trial types. If, however, *T2* or *T3* are worse iconic representations (to the point that interpretability is affected), we might see performance differ between these three trial types, with, for example, *T1 vs. D1* trials yielding the highest proportion of expected choices, *T2 vs. D1* trials yielding the second highest proportion of expected choices, and *T3 vs. D1* trials yielding the lowest proportion of expected choices. Reaction times would follow this same pattern, with the quickest reaction times expected for *T1 vs. D1* and the slowest for *T3 vs. D1*.

Analysis and Results

Given that the experiment was conducted online on the participant's computer, we applied a series of participant/trial exclusion steps to reduce the chance of noisy data being fed to statistical models. We first excluded participants who did not respond in more than half of the experiment trials. This step removed one participant ($1/105 = 1.0\%$ of all participants). Next, we discarded trials where no response was registered, resulting in 66 trials (0.5% of all trials) being removed. Finally, we eliminated trials the response time (RT) of which lies beyond three standard deviations from the mean RT calculated on the basis of individual participants' RT for each comparison type, following the recommendation put forth in Berger and Kiefer (2021). This criterion removed 77 trials (0.6%). Altogether, 12,337 trials entered the statistical analyses.

The included trials were analyzed in terms of RT and categorization outcomes. We present the model and results on categorization first before turning to RT. All analyses were performed in R (R Core Team, 2019), with Bayesian models fitted using `brms` (Bürkner, 2017). We modeled the binary categorization responses with a mixed-effects logistic regression that had **comparison type** (*T1 vs. T2*, *T1 vs. T3*, *T2 vs. T3*, *T1 vs. D1*, *T2 vs. D1*, *T3 vs. D1*; treatment-coded with *T2 vs. T3* as the reference level) as the fixed effect as well as by-participant and by-item random intercepts and slopes.[3] The default level for the response was when the video clicked did not match the expected option (i.e., matched responses were coded with 1, and mismatched responses were coded with 0), so a positive $\beta_{\textbf{comparison type}}$ for a pair means that that pair elicits more expected responses than the *T2 vs. T3* pair. The posterior predictions for an average participant are shown in Figure 2. All comparison types have proportion of expected choice above chance (*T1 vs. T2*: 95% CrI = [0.50, 0.74]; *T1 vs. T3*: 95% CrI = [0.56, 0.78]; *T2 vs. T3*: 95% CrI = [0.50, 0.73]; *T1 vs. D1*: 95% CrI = [0.94, 0.98]; *T2 vs. D1*: 95% CrI = [0.92, 0.97]; *T3 vs. D1*: 95% CrI = [0.85, 0.94]). There is clear evidence that pairs involving distractors bring out more expected choices than those involving only targets ($\text{mean}_{\text{dff in logit}} = 2.23$, 95% CrI = [1.84, 2.71]). Among the pairs that contrast two targets, the model provides little support for a difference in proportion of expected choice, despite

---

[3] The model was constructed with `brm(MATCHED ∼ COMP_TYPE + (1 + COMP_TYPE | PARTICIPANT) + (1 | ITEM), family = bernoulli(link = "logit"))`, and the priors were Normal(0, 2) for both the intercept and β coefficients, Exponential(1) for variances in the random structure, and LKJ(2) for correlation between random effects.

a trend towards higher proportions of expected choice in *T1 vs. T3* trials. However, among the target-distractor pairs, there is strong evidence that the *T1 vs. D1* pair and the *T3 vs. D1* pair are different in terms of the proportion of expected choice invoked ($\text{mean}_{\text{diff in logit}} = 0.92$, 95% CrI = [0.15, 1.76]).

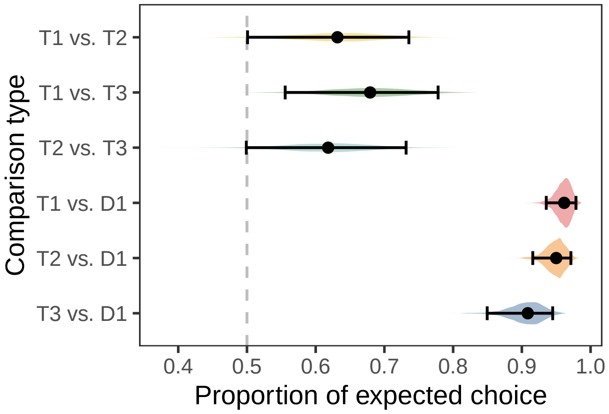

Figure 2: Posterior predictions of proportion of expected choice for different comparison types for an average participant. The dots represent the posterior mean, while the error bars span the 95% credible interval.

We fitted a Bayesian mixed-effects model with a log-normal distribution for RT data. That is, RT (in ms) was assumed to follow a log-normal distribution, conditioned on **comparison type** (again with *T2 vs. T3* as the reference level) and maximally-specified by-participant and by-item random effects.[4] The posterior predictive mean RTs for an average participant across comparison types are plotted in Figure 3. Overall, the mean RT for the pairs involving two targets is longer than those consisting of a target and a distractor ($\text{mean}_{\text{diff}} = 233$, 95% credible interval (CrI) = [64, 394]). However, among similar comparison pairs (e.g., *T1 vs. D1*, *T2 vs. D1*, *T3 vs. D1*), there is little evidence suggesting meaningful differences (*T1 vs. T2* − *T1 vs. T3*: mean = −30, 95% CrI = [−347, 309]; *T1 vs. T2* − *T2 vs. T3*: mean = −84, 95% CrI = [−417, 226]; *T1 vs. T3* − *T2 vs. T3*: mean = −54, 95% CrI = [−380, 277]; *T1 vs. D1* − *T2 vs. D1*: mean = −140, 95% CrI = [−439, 147]; *T1 vs. D1* − *T3 vs. D1*: mean = −258, 95% CrI = [−559, 42]; *T2 vs. D1* − *T3 vs. D1*: mean = −118, 95% CrI = [−448, 174]), despite a clear trend in the pairs containing distractors.

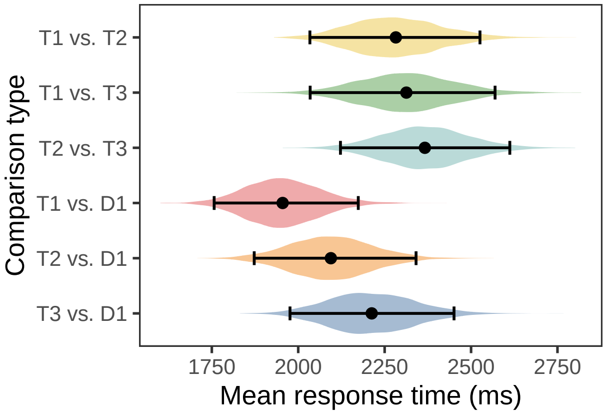

Figure 3: Posterior predictions of mean response time for different comparison types for an average participant. The dots represent the posterior mean, while the error bars span the 95% credible interval.

Summarizing the results from both RT and categorization analyses, participants responded to target-distractor pairs both faster and more in line with expectations. In addition, participants were more accurate in selecting a target over a distractor when the idea represented in the form of the target sign was cross-linguistically the most common than when the idea was cross-linguistically the least common.

Discussion

In the present study we ask whether iconic depictions that show up more frequently cross-linguistically are more interpretable, that is, more frequently chosen as "best suited" to a particular meaning. We found that participants indeed preferred cross-linguistically more common underlying motivations over cross-linguistically less common ones, as participants reliably chose the more common motivation in all three target-target comparisons. We also found that participants chose a target sign from a distractor depicting a different meaning with a particularly high (near ceiling) degree of accuracy as well as speed, and that accuracy was highest in trials where a cross-linguistically common motivation for an animal was compared to a non-attested motivation (that is, accuracy was highest on the trials we predicted would pit a maximally preferred motivation against a maximally dispreferred one). Together, our results show that cross-linguistic patterns of frequency are reflected in participants' patterns of preference, which suggests that distributional asymmetries in the two domains might share a common source.

In choosing between a target and a distractor, participants performed almost at ceiling demonstrating that when participants consider a specific target meaning (and not the entire lexicon), they really are capable of interpreting iconic signs. Indeed, it is reasonable to conclude that one of the reasons people did so well on this study as opposed to previous studies is because of our maximally narrowed context. The fact that people were so successful at picking target signs

---

[4] The model specification in `brms` was `brm(RT ∼ COMP_TYPE + (1 + COMP_TYPE | PARTICIPANT) + (1 | ITEM), family = lognormal())`. We used regularizing priors for model parameters: Normal(7.5, 2) for the intercept, Normal(0, 2) for coefficients, Normal(0, 2) for log(σ), Exponential(1) for all variances associated with random effects, and LKJ(2) for correlation between random effects.

over distractors, even when the target signs were based on the least cross-linguistically common underlying motivations, suggests that a variety of underlying motivations (even those that are uncommon cross-linguistically) are readily accepted by perceivers. This conclusion is further reinforced by the fact that participants performed fastest in target vs. distractor trials.

Participants were more likely to choose a sign based on more cross-linguistically common underlying motivation over a sign based on a less common one, but the proportion of choices between the two targets was not as different from each other as in target vs. distractor trials. Interestingly, since the instructions asked the participants to pick the sign that best suited the meaning of the word provided, they might have been aware of the fact that both signs provided probably meant the same thing, and the difficulty thus was in choosing the best fitting sign. This suggestion is supported by the fact that *T1 vs. T3* trials showed a trend towards producing the highest proportion of *T1* choices. That is, when the two signs compared had the maximal difference in terms of cross-linguistic frequency, participants were more likely to pick the most cross-linguistically common motivation over the least common motivation. This result suggests that there may be something conceptual that drives cross-linguistic frequency of certain iconic signs; but crucially still shows that participants can see iconic associations in less common signs as well (since *T3* motivations were still accepted at higher rates than distractors). In addition, the fact that signs based on more common motivations (*T1* and *T2*) were more frequently chosen over signs based on least common underlying motivations (*T3*) suggests that participants trying to choose a sign best fitting the word meaning were sensitive to something about *T3* signs being less of "the best fit" for the meaning provided. This suggestion is further supported by the fact that *T1 vs. T3* trials had a higher proportion of expected results then the other two within-target comparisons.

It is important to acknowledge the limitations of this study. While we find correlation between the cross-linguistic frequency of the underlying motivation of the iconic sign and its interpretability, we cannot say, on the basis of these data, why some signs are seen as better depictions of the provided meanings than other signs by sign-naïve people. We are continuing to explore this question in other studies.

In conclusion, cross-linguistic patterns in frequency of underlying motivations appear to be reflected in people's views about what motivations best suit a meaning: some motivations seem to work better, for languages and for individual perceivers. In the present study, we are establishing the phenomenon of the correlation between perception and cross-linguistic frequency, but we cannot speculate as to what causes this correlation quite yet. One possibility is that there may be a conceptual reason why some features are preferred, or perhaps more salient, than others, and we continue to explore this option in ongoing research.

References

Berger, A., & Kiefer, M. (2021). Comparison of different response time outlier exclusion methods: A simulation study. *Frontiers in Psychology*, *12*, 675558. doi: 10.3389/fpsyg.2021.675558

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. doi: 10.18637/jss.v080.i01

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. doi: 10.3758/s13428-014-0458-y

Grosso, B. (1993). *Iconicity and arbitrariness in Italian Sign Language: An experimental study.* Unpublished doctoral dissertation, University of Padua, Padua.

Lai, Y., & Yang, L. (2009). Iconicity and arbitrariness in Taiwan Sign Language: A psycholinguistic account. *Mingdao Journal*, *5*(2), 159–187. doi: 10.6953/MJ.200912.0159

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

Pizzuto, E., & Volterra, V. (2000). Iconicity and transparency in sign languages: A cross-linguistic cross-cultural view. In K. Emmorey & H. L. Lane (Eds.), *The signs of language revisited: An anthology to honor Ursula Bellugi and Edward Klima* (pp. 261–286). Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

R Core Team. (2019). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

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

Stokoe, W., Casterline, D., & Croneberg, C. (1965). *Dictionary of American Sign Language on linguistic principles.* Silver Spring, MD: Linstok Press.

Tkachman, O., & Hudson Kam, C. (2019). Semantic salience as iconic motivation of signs in artificial and natural sign languages. In *Iconicity in language workshop, DGfS2019: 41st Annual Conference of the German Linguistic Society.* doi: 10.13140/RG.2.2.17509.04325

Wescott, R. W. (1971). Linguistic iconism. *Language*, *47*(2), 416–428. doi: 10.2307/412089