April 15, 2025

Grammar across perceptual dimensions

Lyn Tieu, Philippe Schlenker, Jonathan Lamberton, Emmanuel Chemla

Abstract. Some grammatical properties are extremely stable across sign languages, and arise spontaneously in homesign systems. In particular, signers refer to objects and individuals by creating locations for them in signing space, and sometimes by turning or rotating to these locations to adopt their perspectives. Both the stability and spontaneous emergence of this location-based system suggest that it might be rooted in knowledge that is unlearned, and core to human cognition. We show that non-signers are sensitive to non-trivial constraints on these referential locations when they are realized in gestures (Experiments 1A and 1B), but also when they are realized using non-gestural means, including color, shape, and even sound (Experiments 2A-2D). This evidence suggests that the sophisticated location-based referential system of sign languages might have a cognitive source that is not specific to the visual modality. This raises the question of why spoken languages do not use equivalent systems, which are cognitively available in the auditory modality.

1. Introduction

1.1. Grammatical convergence in the visual modality

Human language exists with broadly similar abstract properties across modalities, including the spoken modality and the visual modality. Sign languages raise a key typological issue, namely why historically unrelated sign languages share numerous grammatical properties with each other, to the point that they are often considered a language family, even though it is one without common descent (e.g., Sandler and Lillo-Martin 2006, Schlenker et al. 2024). The puzzle is made sharper by the observation that the communication systems of homesigners, i.e. deaf individuals who grow up without access to a sign language and have to invent a gestural system to communicate, sometimes display grammatical properties of mature sign languages (e.g. Goldin-Meadow 2005, Coppola et al. 2013, Abner et al. 2015).

One possible answer has come from gestural research. It has long been accepted that sign languages are full-fledged languages with the same kind of grammar as spoken languages, and not merely gestural systems. But this doesn't imply that sign languages and gestural systems cannot share certain properties. Goldin-Meadow and Brentari (2017) convincingly argue that 'sign with iconicity' (sign languages with all their iconic potential) should be compared to 'speech with gesture' rather than to speech alone.

Empirically, it has been suggested that certain grammatical properties of sign languages can be adopted by non-signers, sometimes with little or even no prior exposure. For instance, Wilbur (2003, 2008) (and in subsequent work) shows that telicity, a property of verbal constructions that relates to whether the action or event has a natural endpoint ('telos' in Greek), is often marked in the lexical form of verbal signs. Hence, *decide* is classified as 'telic' because the action of deciding involves a natural endpoint, and its lexical sign involves a sharp ending in American Sign Language (ASL). By contrast, *think* is 'atelic', and its ASL sign does not involve a sharp ending. Similar mappings have been shown to exist across several sign languages (e.g., Croatian Sign Language, see Malaia & Milković 2021, Milković 2011, or Austrian Sign Language, see Schalber 2006). Strickland et al. (2015) show that non-signers indeed associate the sharp endings of signs that they do not know with telic meanings.

Moving beyond the level of lexical signs, Schlenker and Chemla (2018) show that non-signer English speakers are sensitive to a gestural version of rather subtle constraints on ASL object agreement verbs, whose endings target different positions depending on the object (e.g., *I tell you* would target the addressee position). In a hybrid sentence consisting of English words and a gesture replacing the verb, such as *You, I am going to SLAP-2*, the slapping gesture must target the addressee's position (represented here with -2). However, in a sentence with elided material such as *Your brother, I am going to SLAP-a, and then you too I am going to SLAP-a*, the slapping gesture can be directed towards a position *a* located, e.g., to the speaker's right and not towards the addressee. Strikingly, this mirrors the behavior of ASL object agreement verbs, whose agreement marker can be disregarded under ellipsis. In short, while English does not display object agreement on the verb, non-signing English speakers are nevertheless sensitive to a gestural counterpart of the constraints on object agreement observed in ASL.

How general are these findings, and what can we learn from them? Building on previous literature, Schlenker (2020) discusses several properties of sign language grammar (including those discussed below) that may be intuitively known by non-signers in the gestural case. Such systematic preferences and behaviors among non-signers might make the convergence of sign language grammars less puzzling. This convergence could follow if the relevant knowledge is one that is unlearned, or at least well-aligned with certain learner biases or preferences that humans may have. The next question then would be whether this convergence is form-specific, or modality-specific, and if not, whether it follows from *general* cognitive biases. We address these questions by testing the analogues of key grammatical phenomena in various forms and in various modalities. We present these phenomena in the following section.

1.2. Properties of interest

We focus on two major grammatical devices by which entities (and in particular individuals) are represented in signing space. One way to represent an entity is to establish a position for it in signing space, a 'locus', and to point to that locus to refer back to that entity. The other way involves a rotation on the part of the signer ('Role Shift') to signal that the signer is embodying another character. When a locus has been established to refer to this character, the rotation typically applies in an area corresponding to this locus. Both of these grammatical devices involve loci, both are constrained by rules, and both seem to be very common across sign languages (see, e.g., Sandler & Lillo-Martin 2006, Lillo-Martin 2012, Quer 2013, Steinbach 2021). The phenomena we will discuss now show these devices in action in sign languages, and we will subsequently test them outside of sign languages. For simplicity, we may at points continue to use sign language terminology, e.g., "locus", even when discussing potential counterparts outside of sign languages.

Simple loci. In numerous sign languages, nominal expressions can create loci in arbitrary positions for entities which are not present in the extra-linguistic context, and pronouns are then realized by pointing towards these loci. When entities are present in the extra-linguistic situation, their locus is determined by their actual position and, for example, one points towards the signer or addressee to realize first or second person pronouns. Based on these cases, loci have been argued to be the overt realization of (unpronounced) logical variables, whose existence has long been posited on indirect grounds for spoken languages (Lillo-Martin and Klima, 1990). On this view, the English sentence in (1a) has two representations depending on whether *he* refers to Sarkozy (variable x) or to Obama (variable y), and similarly the French Sign Language (LSF) sentence in (1b) involves a pronoun (glossed as **IX** for 'index' or pointing) that targets the locus *b* or the locus *a* depending on whether it refers to Sarkozy or to Obama (here the loci *b* and *a* are introduced by classifiers, glossed as CL, associated with SARKOZY and OBAMA, respectively).

(1) a. Sarkozy_x told Obama_y that $\text{he}_{x/y}$ would win the election.
b. SARKOZY CL_b OBAMA CL_a b-TELL-a {IX-b / IX-a}WILL WIN ELECTION.
'Sarkozy told Obama that he, {Sarkozy/Obama}, would win the election.'
(LSF 4, 235; Schlenker et al. 2024b, https://youtu.be/4u8GoVjvn4g)

Dynamic loci. Loci bring even more to the table, and provide valuable indications about the nature of variables in natural languages. In standard ('first-order') logic, variables have rigid relations with quantifiers, constrained by structural considerations only. Such a logic is sufficient to account for the relation between *he* and its referent above. But pronouns in natural languages can enter into more permissive relations, as if a referent for them could be created on the fly as the discourse unfolds, and this led in the 1980s to the development of 'dynamic' theories (e.g., Kamp 1981, Heim 1982). The key examples involve indefinites, as in (2): he_x and she_y can depend on *[a man]*$_x$ and *[a woman]*$_y$ despite being in structural positions where this would be impossible in standard logic (technically: he_x and she_y are not 'in the scope of', or equivalently 'c-commanded by', their antecedents). Strikingly, this permissive (dynamic) relation between variables and indefinite antecedents is not only possible in ASL, but it is made fully transparent. As illustrated in the ASL example in (3), the loci allow us to interpret the pronouns (here blocking the reading '...he wonders who himself lives with'), and loci do so independently of any structural considerations.

(2) When [a man]$_x$ meets [a woman]$_y$, he_x wonders who she_y collaborates with.

(3) WHEN [FRENCH MAN]$_a$ a,b-MEET [FRENCH MAN]$_b$, IX-a WONDER WHO IX-b LIVE WITH.
'When a Frenchman meets a Frenchman, the former wonders who the latter lives with.'
(ASL, i P1040945; Schlenker 2011)

Attitude and Action Role Shift. In Role Shift, the signer shifts their body (and makes additional changes, e.g., to eyegaze, facial expression) to adopt the perspective of a person whose words or thoughts are reported ('Attitude Role Shift'), or whose actions are conveyed in a particularly vivid fashion ('Action Role Shift'). Importantly, in both cases, if the person is initially associated with a locus *a*, the signer must shift their body in a position corresponding to *a* to adopt that person's perspective. In addition, in Attitude Role Shift, after the signer has shifted their body, the first person pronoun (realized by a pointing sign towards the signer) typically denotes the person whose perspective is adopted rather than the signer themselves.

Loci with Locative Shift. Despite the similarities between loci and variables, there are differences as well. One of them, instantiated in ASL and likely far more broadly, has been termed 'Locative Shift' (e.g., Padden 1988, Emmorey 2002, Schlenker 2018). In a nutshell, when an individual is asserted to be located in a physical place associated with a locus *a*, one can optionally co-opt locus *a* to refer to the individual rather than to the place. This is illustrated in (4): John is associated with locus *b*, a French city is associated with locus *a*, and one can point towards *a* to refer to John-in-the-French-city. Similarly, an American city is associated with locus *c*, and one can point to *c* to refer to John-in-the-American-city. Locative Shift is merely optional though: one can still choose to point to locus *b* to refer to John.

(4) JOHN IX-b WORK [IX-a FRENCH CITY]$_a$ SAME WORK [IX-c AMERICA CITY]$_c$.
IX-a IX-1 1-HELP-a. IX-c IX-1 NOT 1-HELP-c.
'John works in a French city and also in an American city.
There [= in the French city] I help him. There [= in the American city] I don't help him.'
(ASL, 8, 1; Schlenker 2018)

Although a surprising property of ASL, Locative Shift is nevertheless a constrained one. First, its acceptability is degraded if it is denied rather than asserted that the person is at the relevant location (Schlenker et al. 2024a). Second, Locative Shift is degraded with first and second person pronouns (and possibly more generally for loci that correspond to individuals present in the extra-linguistic situation): pointing towards the signer or addressee appears to be non-negotiable in these cases (Schlenker 2018).

2. Study 1 (two experiments): Gestural loci

We tested whether analogous versions of sign language phenomena could be spontaneously recognized by non-signers in gestures. Spoken sentences were pronounced by a speaker who simultaneously produced gestures designed to follow the use of signing space loci and role shift.

2.1 Method

Task. On each trial, participants either watched a video of the speaker producing the relevant gestures (Exp. 1A) or read sentences that were accompanied by screen captures of the speaker producing the relevant gestures (Exp. 1B). They then had to rate the naturalness of the final gesture on a continuous slider scale that was labelled with the endpoints 'Completely unnatural' and 'Completely natural'. Exp. 1A (containing video recordings of the speaker) provided a more ecological setting, while Exp. 1B (containing auto-paced frames) provided a simpler counterpart of the videos. See Figures 1a-c for examples of conditions in the format of Exp. 1B (all materials are available in the online Supplementary Materials).

Figure 1(a). Example of the gestural "simple loci" condition, as presented in the frame-based experiment 1B.

Figure 1(b). Example of the gestural “locative shift” condition, as presented in the frame-based experiment 1B.

Figure 1(c). Example of the gestural “attitude role shift” condition, as presented in the frame-based experiment 1B.

Participants. A total of 100 participants were recruited through Prolific; 51 completed Exp. 1A and 49 completed Exp. 1B. Participants were pre-screened for first language (English), previous approval rate (90-100%), and geographic location (Canada, USA); people who had participated in pilot versions of the experiment were excluded. All participants provided informed consent before starting the study. Participants in Exp. 1A were paid £3.50 for the task (average completion time: 23m41s), while participants in Exp. 1B were paid £4.25 for the task (average completion time: 20m51s). In Exp. 1A, we additionally excluded two participants who in the post-experiment demographic survey did not report English as one of their native languages, and two others who completed the experiment in less than 11 minutes (the combined length of all videos appearing in the experiment). This left 47 participants for analysis in Exp. 1A, and all 49 participants in Exp. 1B.

Procedure. Participants completed an acceptability judgment task on Qualtrics. Participants were given three example items in the instructions (which involved describing an action towards a tall person: a congruent upwards-oriented gesture, a less congruent downwards-oriented gesture, and a neutral gesture, see Supplementary Materials). These three examples then appeared as the first three items of the experiment, before 40 experimental items presented in a completely randomized order.

2.2 Results

We will simultaneously present the conditions, materials, and results. Fig. 2 provides a summary of all results.

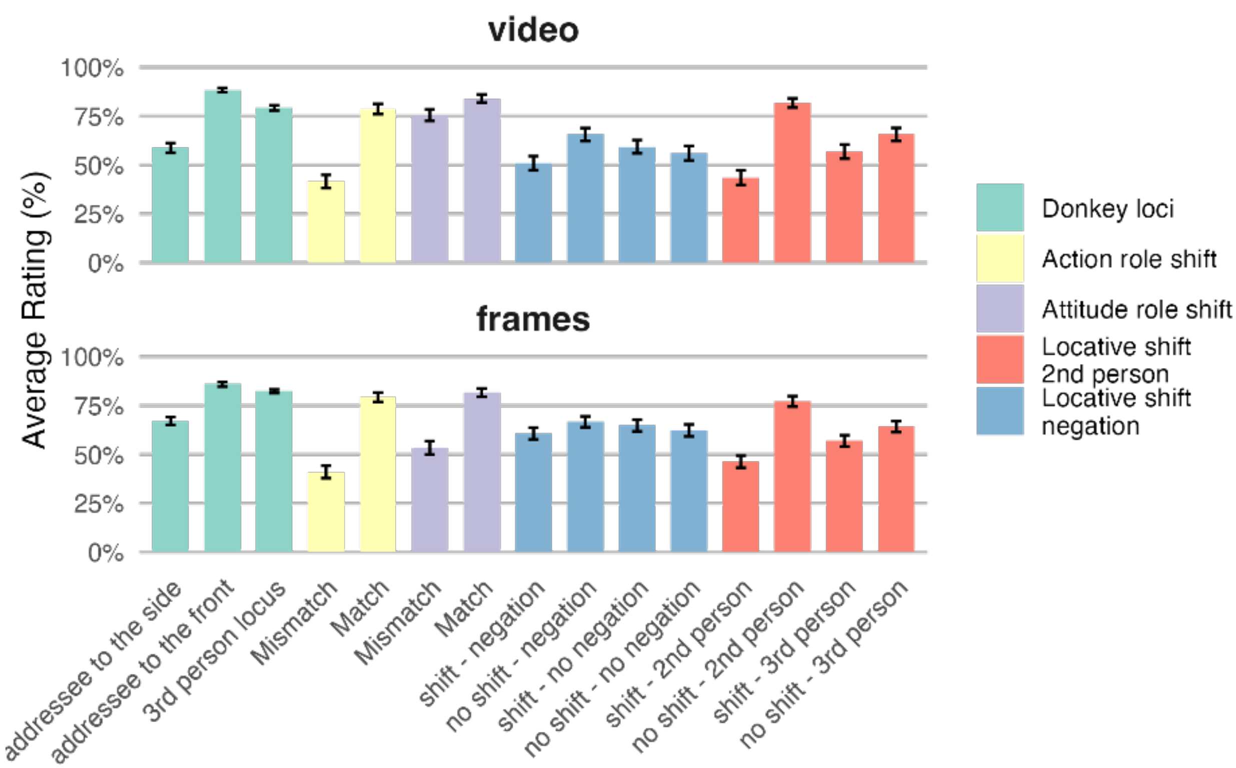

Figure 2. Average rating (expressed in percentage of the response bar filled in up to "Completely Natural") for the video and frame-based presentation of gestural counterparts of various signing space phenomena. The first bar for each of the phenomena (grouped by color) corresponds to what is predicted to be a degraded condition, and should therefore be significantly lower than the subsequent bars in the group.

Simple and dynamic loci. The experiment contained four paradigms similar to the one in (5) below. A first clause introduced two gestural "loci" using indefinites and hand gestures, one to the left (HAND-left), and one either to the right (HAND-right) or to the front of the speaker (HAND-front). Participants judged that (a), with the index finger pointing directly in front of the speaker (IX-front), was more acceptable than (b), in which the gestural locus for the second person was to the side rather than directly in front of the speaker (video: $\chi^2(1)=125$, $p<.001$; frames: $\chi^2(1)=58.7$, $p<.001$). Furthermore, (c) and (d), which introduced loci to the left and right and subsequently referred to them with pointing gestures, were also rated better than (b) (video: $\chi^2(1)=81.2$, $p<.001$; frames: $\chi^2(1)=51.4$, $p<.001$). Overall, non-signers accepted the use of loci and gestures pointing towards them, while recognizing the constraint on second person loci.

(5) If ever I am able to hire HAND-left **[a mathematician]** …

(a) or HAND-front **[you]**, I will pick IX-front.	(c) or HAND-right **[a sociologist]**, I will pick IX-left.
(b) or HAND-right **[you]**, I will pick IX-right.	(d) or HAND-right **[a sociologist]**, I will pick IX-right.

Attitude Role Shift and Action Role Shift. Participants were also sensitive to role shifts: for both repetitions of the paradigms below, participants gave higher ratings to the target conditions, in which there was a match between the gestural "locus" of the referent (e.g., HAND-left) and the Role Shift (e.g., RS_{left}). This was the case for attitude role shift (6), in which the gesture involved the referent pointing towards themselves (video: $\chi^2(1)=5.35$, $p=0.021$; frames: $\chi^2(1)=66.2$, $p<.001$), and for action role shift (7), in which the speaker mimicked the action the referent took (video: $\chi^2(1)=99.1$, $p<.001$; frames: $\chi^2(1)=119$, $p<.001$).

(6) I was standing next to HAND-left **[little Robin]** and HAND-right **[little Francis]**, and I was holding a really yummy chocolate bar. And I asked: Who wants it? And so of course:

Target: HAND-left **[little Robin]** goes: RS_{left} IX-self.	*Target:* HAND-right **[little Francis]** goes: RS_{right} IX-self.
Control: HAND-left **[little Robin]** goes: RS_{right} IX-self.	*Control:* HAND-right **[little Francis]** goes: RS_{left} IX-self.

(7) I was standing next to HAND-left **[little Robin]** and HAND-right **[little Francis]**, and the situation between them seemed to be very tense. Suddenly,

Target: HAND-left **[little Robin]** goes: RS_{left} SLAP-right.	*Target:* HAND-right **[little Francis]** goes: RS_{right} PUNCH-left.
Control: HAND-left **[little Robin]** goes: RS_{right} SLAP-left.	*Control:* HAND-right **[little Francis]** goes: RS_{left} PUNCH-right.

Locative Shift. Using two repetitions of the paradigm illustrated in (8) below, we tested the acceptability of locative shift in cases where it was motivated (8a) compared to when it was unmotivated due to negation (8b). We compared these attempted locative shifts to examples without any shifts, as in (8c) and (8d). Participants judged Locative Shift to be *less* acceptable when it was unmotivated. We fit mixed effect linear regression models with Shift (Shift vs. No shift), Motivation (negation vs. no negation), and their interaction as fixed effects, and random by-participant intercepts. A comparison of the model with one without the interaction term revealed that Shift interacted significantly with Motivation for the video presentation (video: $\chi^2(1)=12$, $p<.001$): the cost of shifting was greater when the Locative Shift was unmotivated, compared to when it was motivated. This interaction was not significant for the frame-based experiment (frames: $\chi^2(1)=3.05$, $p=0.081$).

(8) John and Mary are my best employees. As you know, **John** HAND-bottomleft is in New York and **Mary** HAND-bottomright is in L.A. As requested, I'm not gonna transfer **John** HAND-bottomleft to **London** HAND-topleft, but I am gonna transfer **Mary** HAND-bottomright to **Paris** HAND-topright. And if the international expansion fails, you know who I'll fire first?

(a) IX-topright	*shift, no negation*	(c) IX-bottomright	*no shift, no negation*
(b) IX-topleft	*shift, negation*	(d) IX-bottomleft	*no shift, negation*

Using two repetitions of a similar paradigm (9), we also observed a statistically significant interaction between Shifting and Person, with Locative Shift being costlier for 2nd person referents compared to 3rd person referents (video: $\chi^2(1)=31.8$, $p<.001$; frames: $\chi^2(1)=23.2$, $p<.001$). In this case, the lack of motivation for the shifting was not due to the negation, but to the 2nd person referent.

(9) You and John are my best employees. As you know, **John** HAND-bottomleft is in New York and **you** HAND-front are in L.A. As requested, I'm gonna transfer **John** HAND-bottomleft to **London** HAND-topleft, and I'm gonna transfer **you** HAND-front to **Paris** HAND-topright. And if the international expansion fails, you know who I'll fire first?

(a) IX-topleft	*shift, 3rd person*	(c) IX-bottomleft	*no shift, 3rd person*
(b) IX-topright	*shift, 2nd person*	(d) IX-front	*no shift, 2nd person*

2.3 Discussion

The results of Exp. 1A and 1B suggest that non-signers know certain constraints on loci when these are realized in gestural "loci". The target constructions are fairly common in sign language, but are likely quite rare in gestures. If this is the case, the findings suggest that participants are accessing unlearned knowledge. In a second study, we investigated whether the same constraints would extend to even less familiar stimuli by testing loci that were realized using perceptual properties other than location.

3. Study 2 (4 experiments): Abstract "loci" across perceptual dimensions

The hypothesis that people are tapping into 'unlearned knowledge' opens the door to two broad possibilities. One is that non-signers' linguistic knowledge includes dedicated rules for the visual modality, perhaps even governing the specific use of locations as loci. Another possibility is that more general cognitive rules (not tied to a specific visual realization) drive their behavior.[1] The latter, but not the former, predicts that similar rules could drive the behavior of abstract "loci" that involve different visual realizations or even auditory ones. In Study 2, we tested this possibility through four entirely parallel experiments that used different cues for such versions of "loci": position (the most conservative one), shape and color (which cannot be realized with signs or gestures), and sound. (Beyond these abstract cases, see Patel-Grosz et al. 2019, 2022, for a related investigation of loci-like objects in dance; see Schlenker 2022 and Migotti Ramponi 2023 for arguments that 'musical variables' could exist.)

[1] In terms of the models discussed by Cohn and Schilperoord (2024), on the first hypothesis, sign languages and gestures would be able to make use of special affordances of the visual modality, whereas the second hypothesis would be more amodal.

3.1 Methods

The preregistration for this study can be found at: https://osf.io/mdc4r/?view_only=e42978fcec6749bdbe5f3d8d0ec7bb3e. More details about the statistics, methods, and materials can be found in the Supplementary Materials below and online at: https://osf.io/m4dxp/?view_only=66dc7157127d432c8b5d55bb8c55b2fa.

Participants. We recruited 240 participants through Prolific, pre-screened for first language (English) and vision (normal or corrected-to-normal vision), as well as participation in previous studies (no prior participation). We received 242 responses, due to some responses timing out in the Prolific system. We then excluded three participants who did not report English as one of their native languages in the post-experiment demographic survey (see R analysis script in the Supplementary Materials, R Core Team 2021). Included in the final data analysis were 59 participants in the Audio experiment, 61 in the Color experiment, 59 in the Position experiment, and 60 in the Shape experiment. Participants in each task were paid £2.50 (average completion time: 13m18s). All participants provided informed consent before starting the study.

Procedure. Through a Qualtrics survey, participants completed one training trial (which corresponded to the example item from the instructions), followed by 30 experimental trials presented in a randomized order. On each trial, participants watched a video unfold, with two possible endings presented side by side (i.e. with two competing sounds, two visual objects of different colors, in different positions, or of different shapes). The task was a binary forced choice task in which participants had to indicate their choice of the better ending (*Which version do you prefer?*) (see Marty et al. 2020 for evidence of a sensitivity advantage when conditions are presented jointly as opposed to individually).

Materials. Each participant completed only one of the four experiments, so they only saw one of the four perceptual dimensions. Each version of the experiment tested four phenomena: dynamic "loci" (3 repetitions, for a total of 6 trials), action role shift, (3 repetitions, for a total of 6 trials), attitude role shift (3 repetitions, for a total of 6 trials), and locative shift (4 repetitions of two tests, for a total of 16 trials - only the first test is presented in the main text). All stimuli creation was automated using a Python script: a single file contained the sentences with indications for loci and frame separations (see online Supplementary Materials), and from there a frame-by-frame version of each stimulus was created for each of the four dimensions of interest.

The stimuli were constructed from an abstract code, an example of which is given for each phenomenon in (10). This led to frame-by-frame videos, which we have represented using screen captures in Figure 3. For more details, see the Appendix and online Supplementary Materials.

In a nutshell, the "/" separates frames, and the first number between a set of brackets encodes the locus to be realized (subsequent information such as ":.2" indicates a 200ms delay between the display of the text and that of the locus-like information). In the final frame, the code indicates two numbers in that position, separated by a dash, specifying which two loci should be contrasted for the response choices. For the three visual dimensions, this led to a frame with two simultaneous subframes side-by-side showing the relevant loci; for the audio modality, the two subframes were also represented on each side, but one after the other so that the relevant sounds could be heard and associated with each response side.

Contrasts. In the *Dynamic loci* conditions, the contrast was, for a given noun, between a congruent ('matching') and an incongruent ('mismatching') locus that had previously been introduced. For *Action role shift* and *Attitude role shift* conditions, the contrast was between an action (e.g., an image of an explosion) and an attitude (e.g., the text "me!"), respectively, presented at a particular position (location), embedded within a particular shape (shape modality), within a particular color patch (color modality), or contingent with a specific sound (audio modality). For the *Locative shift* condition, the contrast was between the "shift, no negation" and the "shift, negation" conditions from before, where the shift was expected to only be appropriate in the latter case.

(10) a. **Dynamic loci (3 versions):**
If ever I can afford to hire/a lawyer[1:.1]/or a mathematician,[2:.3]/I will choose/the lawyer.[1-2:.1]

b. **Action role shift (3 versions):**
I was watching/Tom[1:.1]/and Jerry.[2:.3]/They were really upset./Suddenly,/Tom went:/[1-2:.1:kaboom]

c. **Attitude role shift (3 versions):**
I was standing next to/Jane[1:.1]/and Billy.[2:.3]/I asked:/Who wants my candy?/Jane goes:/me![1-2:.1]

d. **Locative shift (4 versions):**[2]
Next week,/John[10:.1]/will go to/<keep>France.[11:.1]/But Mary[20:.3]/will NOT go to/<keep>Denmark.[21:.1]/Who will get more work done?/[11-21:.1]

Perceptual dimensions. For *position*, the different "loci" were represented with black circles at different points on the screen. For *shape* and *color*, the different "loci" were represented by an element in the middle of the screen, which varied in shape and color, respectively. For *audio*, the "loci" corresponded to two distinct sounds. Some representative examples are given in Figure 2; all materials can be found in the online Supplementary Materials.

[2] A second test was used for these, looking at a full paradigm with four examples as in Study 1 (that is, including and comparing not only the shifted reference, whether motivated or not motivated, but also their corresponding unshifted versions as a baseline). It yielded similar results (for details, see the Appendix).

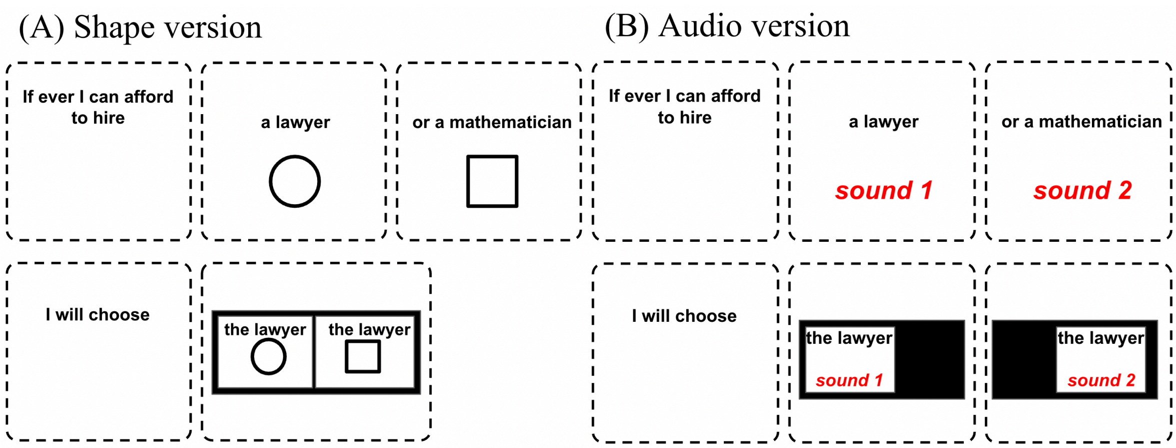

Figure 3. Representation of a test trial in Study 2 for (A) the *shape* modality and (B) the *audio* modality. Successive screens unfolded automatically, up until a contrast (involving a matching locus vs. a mismatching locus) for which participants were asked to report their preference. The contrast was presented on one screen for the visual modalities (left) and on two successive screens for the audio modality (to avoid superposition of the sounds and clear assignment of the left/right response choices).

3.2 Results

The results appear in Fig. 4. The plots display the percentage of 'matching' responses, corresponding to selections of the "locus" that matched the target referent in form (position, shape, color or audio). The results from the four dimensions were quite similar, with participants favoring congruent responses for the *Dynamic loci*, *Attitude role shift*, and *Action role shift* conditions (all z's > 2.95, p's <.0031) – but not for the *Locative shift* condition (with in fact negative z's, although p-values were all above 0.16).[3]

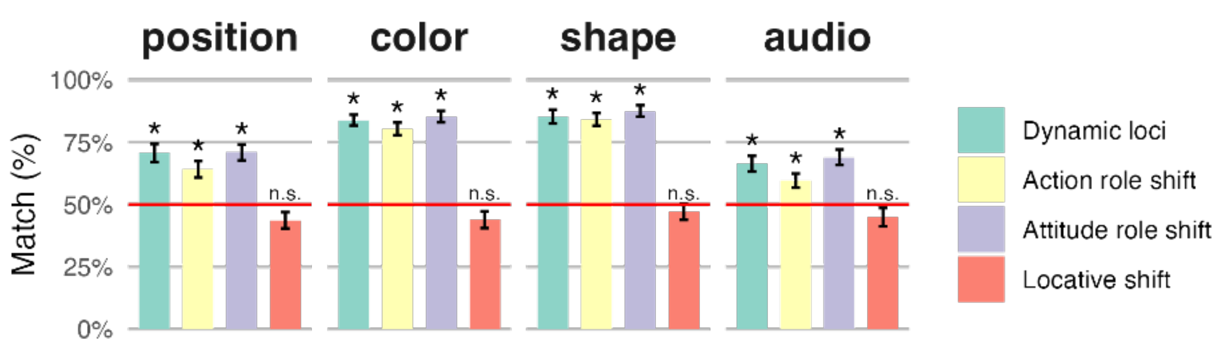

Figure 4. Percentage of matching responses for the conditions *Dynamic loci*, *Attitude role shift*, *Action role shift*, and *Locative shift* (all realized in non-gestural modalities). In all modalities, participants showed sensitivity to all phenomena except for *Locative shift*.

[3] We report results from models with only by-participant intercepts as a random structure, as in the pre-registration script. We noted that we would run the models with the maximal random structure that would converge. All versions yielded qualitatively similar results.

4. Discussion and conclusion

Our results suggest that, for three of our four target phenomena (locus-matching rules for dynamic loci, attitude role shift, and action role shift), participants are sensitive to abstract rules evidenced in sign language and in gestures, even when these are applied to stimuli that have a very different visual form – using shape and color rather than position, and even a non-visual form, in the case of the auditory stimuli. In other words, the spontaneous preferences that participants display in our tasks, tested across modalities, are aligned with features that are well conventionalized in sign languages. Although innocent looking, such preferences could potentially tell us about the kinds of interpretive and learning biases that humans may have, even in the absence of previous exposure to sign languages. These biases might even be at play in the absence of any sort of relevant exposure at all, since many of the stimuli types we tested are certainly absent from the participants' input, gestural or otherwise. This finding thus strengthens the hypothesis that participants here are tapping into unlearned knowledge (or natural preferences, if one prefers a graded version of the notion), and moreover that this unlearned knowledge is domain-general, or at least modality agnostic.

Turning to our fourth target phenomenon, locative shift, participants' intuitive knowledge of the rule in the gestural case fails to generalize to the other stimulus types. One possible explanation is that locative shift involves visual iconicity, in the sense that one needs to pictorially represent an individual as being in a location in order to co-opt the locus for that location to refer to the individual. But if this is so, it is a bit surprising that our most conservative non-gestural stimuli, which involve positions on a screen rather than in gestural space, fail to display the effect. As with all null results, sharper materials and more sensitive tasks could reveal the presence of this more subtle effect, and a key question then would be to see if the phenomenon surfaces similarly in every dimension tested.

While our studies probed the origin of aspects of sign language grammars, our results raise a question for spoken languages, namely why they do not make use of auditory loci, given that they could be understood spontaneously. One can speculate that there are nonetheless comprehension constraints, e.g., some that make it hard to keep track of many sounds. But in practice sign language only makes use of a small number of loci in any given sentence, and a system could be imagined where one would pronounce loci related pronouns as "locus-one", "locus-two", etc. An alternative approach is to say that agreement markers sometimes serve that purpose. In fact, in example (2), gender on *he* and *she* plays the role of loci in disambiguating referents. But even in languages with a more general gender system as in Romance languages, the situation is much more impoverished than in sign languages (it is not possible to flexibly assign a gender or another agreement feature to a referent so as to later be able to pick it up with a corresponding pronoun, see Kuhn 2016 for an analysis of ASL loci as agreement markers, which explicitly points to departures from standard properties of agreement in spoken languages). In our view, there is thus no satisfying answer yet to the absence of loci in spoken languages.

In sum, on the basis of little or no prior input, non-signers spontaneously adhere to several non-trivial constraints on sign language loci when these are realized in gestures. This suggests that sign language constraints on loci are in part based on such preferences, and potentially rooted in unlearned knowledge. Strikingly, non-signers can also uncover most of these constraints when they are realized through very different perceptual properties such as color, shape, and sound. This unlearned knowledge is therefore not restricted to the visual modality of sign languages, and might therefore have deeper cognitive roots that spoken languages do not capitalize on.

References

Abner, N., Namboodiripad, S., Spaepen, E. and Goldin-Meadow, S. (2015). Morphology in Child Homesign: Evidence from Number Marking. (Slides of a talk given at the 2015 Annual Meeting of the Linguistics Society of America, Portland, Oregon)

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Cohn, N., & Schilperoord, J. (2024). A Multimodal Language Faculty : A Cognitive Framework for Human Communication. Bloomsbury Publishing.

Coppola, M., Spaepen, E., & Goldin-Meadow, S. (2013). Communicating about quantity without a language model: Number devices in homesign grammar. *Cognitive Psychology*, 67, 1–25.

Emmorey, Karen. 2002. *Language, cognition, and the brain: Insights from sign language research*. Mahwah, NJ: Erlbaum.

Goldin-Meadow, S. (2005). The resilience of language: What gesture creation in deaf children can tell us about how all children learn language. Psychology Press.

Goldin-Meadow, S. & Brentari, D. (2017). Gesture, sign and language: The coming of age of sign language and gesture studies. *Behavioral and Brain Sciences*, doi:10.1017/S0140525X15001247

Heim, I. (1982). *The Semantics of Definite and Indefinite Noun Phrases.* PhD Dissertation, University of Massachusetts, Amherst.

Kamp, H. (1981) A theory of truth and semantic representation, 277-322, JAG Groenendijk, TMV Janssen and MBJ Stokhof, eds.

Lillo-Martin, D., & Klima, E. S. (1990). Pointing out differences: ASL pronouns in syntactic theory. *Theoretical issues in sign language research*, *1*(1), 191-210.

Lillo-Martin, D. (2012). Utterance reports and constructed action. In R. Pfau, M. Steinbach, & B. Woll (Eds.), *Sign language: An international handbook* (pp. 365–387). Berlin: Mouton de Gruyter.

Malaia, E., & Milković, M. (2021). Aspect: Theoretical and experimental perspectives. In J. Quer, R. Pfau, & A. Herrmann (Eds.), The Routledge handbook of theoretical and experimental sign language research (pp. 194–212). Routledge.

Marty, P., Chemla, E. & Sprouse, J., (2020) "The effect of three basic task features on the sensitivity of acceptability judgment tasks", *Glossa: a journal of general linguistics* 5(1): 72. doi: https://doi.org/10.5334/gjgl.980

Milković, M. (2011). Verb classes in Croatian Sign Language (HZJ): Syntactic and semantic properties. Zagreb: University of Zagreb doctoral dissertation

Migotti Ramponi, L. (2023). Investigating musical meaning: Empirical evidence for a music semantics. PhD Dissertation, PSL University.

Padden, Carol A. 1988. Grammatical theory and signed languages. In Frederick J. Newmeyer (ed.), *Linguistics: The Cambridge survey 2*. 250–266. DOI: https://doi. org/10.1017/CBO9780511621055.014

Patel-Grosz, P., Grosz, P. G., Kelkar, T., & Jensenius, A. R. (2019). Coreference and disjoint reference in the semantics of narrative dance. Proceedings of Sinn Und Bedeutung, 22(2), 199–216. Retrieved from https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/78

Patel-Grosz, Pritty; Grosz, Patrick Georg; Kelkar, Tejaswinee; Jensenius, Alexander Refsum (2022). Steps towards a Semantics of Dance, *Journal of Semantics*, 39(4): 693–748 [DOI: https://doi.org/10.1093/jos/ffac009].

Quer, Josep. 2013. Attitude ascriptions in sign languages and role shift. *Proceedings of the Texas Linguistic Society (TLS) 13*. 12–28.

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

Sandler, W., & Lillo-Martin, D. (2006). *Sign Language and Linguistic Universals.* Cambridge University Press.

Schalber, K. (2006). Event visibility in Austrian Sign Language (ÖGS). Sign Language & Linguistics 9(1). 207-231.

Schlenker, P. (2011). Donkey Anaphora: the View from Sign Language (ASL and LSF). *Linguistics and Philosophy* **34**(4): 341-395.

Schlenker, P. (2018), Locative Shift. *Glossa: A Journal of General Linguistics*, 3(1), 115. DOI: http://doi.org/10.5334/gjgl.561

Schlenker, P. (2020). Gestural Grammar. *Natural Language & Linguistic Theory* 38:887–936

Schlenker, P. (2022), Musical Meaning within Super Semantics. *Linguistics & Philosophy* 45:795–872. https://doi.org/10.1007/s10988-021-09329-8

Schlenker, P., & Chemla, E. (2018). Gestural semantics: Comparing sign language and co-speech gestures. *Natural Language & Linguistic Theory, 36*(3), 877–935.

Schlenker, P., Bonnet, M., Lamberton, Jon, Lamberton, Jason, Chemla, E., Santoro, M., Geraci, C. (2024a), Iconic Syntax: Sign Language Classifier Predicates and Gesture Sequences. *Linguistics & Philosophy* 47:77–147.

Schlenker, P.; Lamberton, J.; Kuhn, J. (2024b), Sign Language Semantics. *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/sign-language-semantics/

Shravan Vasishth. (2021, August 8). *A Common Mistake in Data Analysis (in Psychology/Linguistics): Subsetting Data to Carry Out Nested Analyses (Part 1 of 2).* Retrieved from https://vasishth-statistics.blogspot.com/2021/08/a-common-mistake-in-psychology-and.html.

Steinbach, Markus "Role shift: Theoretical Perspectives." *The Routledge Handbook of Theoretical and Experimental Sign Language Research*, edited by Josep Quer, Roland Pfau, Annika Herrmann, Routledge, 2021, pp. 351-377. doi: 10.4324/9781315754499-16.

Strickland, B., Geraci, C., Chemla, E. Schlenker, P., Kelepir, M., & Pfau, R. (2015). Event representations constrain the structure of language: Sign language as a window into universally accessible linguistic biases. *Proceedings of the National Academy of Sciences*, 112(19), 5968-5973.

Wickham, Hadley (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wickham H, François R, Henry L, Müller K, Vaughan D (2023). _dplyr: A Grammar of Data Manipulation_. R package version 1.1.4, <https://CRAN.R-project.org/package=dplyr>.

Wilbur, Ronnie (2003). Representations of telicity in ASL. *Chicago Linguistic Society* **39** : 354-368.

Wilbur, Ronnie (2008). Complex Predicates Involving Events, Time and Aspect: Is This Why Sign Languages Look So Similar? In: Quer, J. (ed.) Signs of the Time. Selected Papers from TISLR 2004, 217-250. Hamburg: Signum Press.

Supplementary Materials

All experimental materials and analysis files can be found in the following online repository: https://osf.io/m4dxp/?view_only=66dc7157127d432c8b5d55bb8c55b2fa.

A. Details for Study 1 and Study 2

Details. All materials, code, and analyses can be found online at: https://osf.io/m4dxp/?view_only=66dc7157127d432c8b5d55bb8c55b2fa

Procedure. In all of the reported experiments across the two studies, participants first saw a consent form, followed by the instructions. Instructions contained examples unrelated to the target phenomena but covering the range of possible responses. These examples from the instructions re-appeared as the first items in the experiments. They were then followed by all experimental items presented in random order.

Statistical analyses. Following Vasishth (2021), we fit all the data of all target items in each experiment with a single mixed-effect model. We report the result of an ANOVA (χ^2-test) between this model and one in which the relevant distinction is dropped (for simple two-way comparison, collapsing the two relevant levels, for statistical interactions, dropping the interaction term). Models reported in the main text have a very simple random structure with an intercept for participants. More complex random structures yielded occasional singular fits (specially when per item random effects were included, certainly due to the low number of repetitions within each condition), but results were qualitatively similar in terms of inferred significance. All analyses were formally registered online for Study 2, and an entirely similar scheme was followed for Study 1. We used R packages ggplot2 (Wickham, 2016), lme4 (Bates et al. 2015) and dplyr (Wickham, 2023). All files required to run and replicate the analyses can be found here: https://osf.io/m4dxp/?view_only=66dc7157127d432c8b5d55bb8c55b2fa

B. Second locative shift test for Study 2

The possibility of locative shift was tested in two ways in Study 2. As described in the main text, one condition as exemplified again below, tested whether participants would prefer to refer to John with 11, the "locus" corresponding to France, or would prefer to refer to Mary with 21, the "locus" corresponding to Denmark. Because of the negation ("Mary will NOT go to Denmark"), we expected that there would be a preference for the former. These results are reported in the main text; participants' preferences were not significantly different from chance. We also included in the experiment items testing contrasts with baseline conditions (no shift). Specifically, we had items contrasting direct reference to John versus to Paris (10 vs. 11), and items contrasting direct reference to Mary versus to Denmark (20 vs. 21). We expected that the choice would be more neutral in the former case, while negation would make the direct reference to Mary preferable to reference to Denmark. We tested this by comparing the responses to the two conditions (and not only to chance level, as for all other tests in this study) and found no significant difference, except marginally in the case of shape but this would not take into account multiple comparisons (audio: $\chi^2(1) = 1.91$, $p = 0.17$; color: $\chi^2(1) = 0.678$, $p = 0.41$; position: $\chi^2(1) = 0.0903$, $p = 0.76$; shape: $\chi^2(1) = 3.91$, $p = 0.048$).

> Next week, / John[10:.1] / will go to /<keep> France.[11:.1] / But Mary[20:.3] / will NOT go to /<keep> Denmark.[21:.1] / Who will get more work done? / [11-21:.1]