Iconic Syntax:
Sign Language Classifier Predicates and Gesture Sequences*

Philippe Schlenker[a], Marion Bonnet[b], Jonathan Lamberton[c], Deanna Gagne[d], Mirko Santoro[e], Carlo Geraci[f]

Abstract. In several sign languages, the standard word order (e.g. SVO) gets turned into SOV (with pre-verbal arguments) when the predicate is a 'classifier', i.e. a distinguished construction with highly iconic properties (e.g. Pavlic 2016). In silent gestures, subjects also prefer an SOV order in extensional constructions, irrespective of the word order of the language they speak (Goldin-Meadow et al. 2008). But it was observed that in silent gestures and in Brazilian Sign Language (Libras), intensional constructions can override these SOV preferences, yielding SVO instead (Schouwstra and de Swart 2014, Napoli et al. 2017). This distinction was argued to be due to iconicity: arguments are expressed before the verb if they correspond to entities that are present before the action, otherwise they follow the verb. While agreeing with this intuition, we argue that the extensional/intensional distinction is neither empirically nor theoretically appropriate. In new data from American Sign Language (ASL), we replicate the distinction among extensional classifier predicates: for *x ate up the ball*, the ball is typically seen before the eating and a preposed object is preferred; but for *x spit out the ball*, the ball is typically seen after the spitting and a postposed object is preferred, although both *eat up* and *spit out* are used extensionally. We extend this finding to data involving pro-speech (= speech-replacing) gestures embedded in French sentences. We develop a new formal account within a pictorial semantics for visual animations (inspired by Greenberg and Abusch), one that derives the observed word order preferences, and also explains how the meaning of classifier predicates combines iconic and conventional properties. The result is a full integration of compositional and pictorial semantics, one that has non-trivial syntactic consequences.

Keywords: classifiers, iconicity, iconic syntax, silent gestures, pro-speech gestures, iconic semantics, pictorial semantics, visual narratives

---

* **Author contributions**: Schlenker designed the study and wrote the paper, with the exception of Section 4.5 and some footnotes. Geraci coordinated work on LIS and co-advised Bonnet's internship on the gesture study. Bonnet designed the gesture survey in collaboration with Schlenker and Geraci, collected the data, and computed the results. Lamberton was the native consultant in Schlenker's work on ASL, and helped construct examples; theoretical issues were discussed only after the data collection was complete (to avoid influencing judgments). Gagne provided an informal independent assessment of the ASL data, and wrote Section 4.5. Santoro constructed and assessed the LIS data in collaboration with Geraci.

Funding:
Schlenker: This research received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 788077, Orisem, PI: Schlenker). Research was conducted at DEC, Ecole Normale Supérieure - PSL Research University. DEC is supported by grant FrontCog ANR-17-EURE-0017.
Acknowledgments: For discussion and advice on the gesture survey, we are grateful to Nadine Bade, Emmanuel Chemla, Janek Guerrini, Léo Migotti, Luigi Rizzi and Anastasia Tsilia.

[a] Institut Jean-Nicod (ENS - EHESS - CNRS), Département d'Etudes Cognitives, Ecole Normale Supérieure, Paris, France; PSL Research University; New York University, New York.
[b] Université Paris-Diderot, Paris, France.
[c] Université Paris-Diderot, Paris, France.
[d] Gallaudet University, Washington, DC, USA.
[e] SFL, CNRS and Université Paris 8, Paris, France.
[f] Institut Jean-Nicod (ENS - EHESS - CNRS), Département d'Etudes Cognitives, Ecole Normale Supérieure, Paris, France; PSL Research University.

1 Introduction

1.1 Non-standard syntactic orders with iconic signs and gestures

On several occasions, a surprising syntactic property was unearthed in highly iconic constructions used in the visual modality. First, it was noted long ago that in several sign languages, the standard word order (e.g. SVO for ASL [American Sign Language]) can be modified when the predicate is a 'classifier', a distinguished construction with highly iconic properties: the arguments (and in particular the object) preferably appear in a preverbal position (e.g. SOV or sometimes OSV in ASL, Liddell 1980). Second, Goldin-Meadow et al. 2008 discovered that in some silent gestures produced by non-signers, subjects preferably use an SOV order, irrespective of the basic word order of the spoken language they speak.

These two lines of research were refined and partly unified by Schouwstra and de Swart 2014 and Napoli et al. 2017. First, Schouwstra and de Swart argued that Goldin-Meadow's result applies to extensional constructions (e.g. a gestural rendition of 'throw') but not to intensional ones (e.g. a gestural rendition of 'think of'): the arguments preferably appear before the gestural verb in extensional constructions, but the object appears *after* the verb in intensional constructions. Second, Napoli et al. 2017 proposed to extend this generalization to diverse verbal constructions (including classifier predicates) in Libras (Brazilian Sign Language): they argued that preverbal objects are preferred with extensional verbs, while postverbal objects are preferred with intensional verbs. They took this preference to be due to iconicity (the emphasis is ours):

> The extensional/intensional distinction is (…) spelled out in an iconic way. **Arguments that are present on the scene before an action takes place precede the V; those that are not follow the V.** This is not a vision issue per se, but a **visualization** issue. The preexisting arguments of an extensional event are already somewhere in our mental picture before the predicate is articulated. But in intensional events, arguments are brought into our mental picture only after the predicate is articulated because their existence depends upon that predicate." (Napoli et al. 2017)

1.2 Goals

This piece has two main goals. Our first goal is empirical. We argue that, in some cases at least, the extensional vs. intensional discussion isn't the right one. Within the class of extensional iconic constructions, we find a distinction between, say, *the crocodile ate up the ball,* which comes with arguments in preverbal position, and *the crocodile spit out the ball*, which preferably comes with a postverbal object. This is compatible with the spirit but not at all with the letter of Napoli et al.'s proposal. Both constructions are used extensionally (hence extensional vs. intensional couldn't be the crux of the matter), and what seems to matter is that in an event of eating up a ball, the ball is visible before the action takes place, whereas in an event of spitting out a ball, this is not the case (this is in line with Napoli et al.'s visualization intuition).

By focusing on repeated judgments of one Deaf native signer of ASL (corroborated in their essentials by a second native signer who is hearing), we obtain fine-grained paradigms in which this pattern (preverbal arguments for *eat up* but not *spit out*) is characteristic of predicate classifiers and does not arise in minimally different constructions, notably plain verbs and agreement verbs. We then extend our findings to pro-speech (= speech-replacing) gestures in French. In sequences of three gestures (involving for instance a crocodile, a ball, and an action of eating up / spitting out the ball) embedded in a linguistic environment, we show that preverbal objects are preferred for eat-up- but not spit-out-type gestural verbs.

Our second goal is theoretical. Neither Schouwstra and de Swart 2014 nor Napoli et al. 2017 derive their results from a formal semantic analysis. But one is available for visual representations, namely the semantics for pictures developed by Greenberg 2013, 2021 and extended to visual narratives by Abusch 2013. Abusch doesn't just apply Greenberg's semantics to sequences of pictures, however; she also enriches these with discourse referents (i.e. variables) so as to cross-identify objects that appear in different pictures (e.g. two cubes that look alike might or might not denote the same object; variables disambiguate). A simple modification of a broadly Abuschian analysis yields a semantics for visual animations. This analysis will explain why arguments are preferably preverbal when their denotations are visible before the action but not when they aren't. It can also be refined to apply to classifier

predicates: unlike standard iconic gestures, they have a conventional lexical form which need not be iconic, and thus only their position and movement in signing space is interpreted iconically.

As we will see, the result is an analysis that combines compositional and pictorial semantics, and thus explains the seamless integration of classifier predicates to the compositional core of sign language, as well as some surprising aspects of the syntax of iconic constructions in signs and gestures alike.

1.3 Structure

The rest of this article is organized as follows. In Section 2, we summarize traditional arguments for the preverbal position of arguments in sign language classifier predicates and silent gestures. In Section 3, we turn to Schouwstra and de Swart's and Napoli et al.'s argument for a distinction between extensional and intensional constructions, which we propose to recast in terms of the visibility of event participants. We make our empirical case for a distinction between eat-up- and spit-out-type constructions in two steps, first for ASL classifier predicates in Section 4, then for pro-speech gestures in Section 5. The theory is developed first for the purely iconic, gestural case in Section 6, before being refined for the mixed iconic case of classifier predicates in Section 7. A bridge between iconic and compositional semantics is defined in Section 8, thus explaining how iconic representations can be semantically integrated in sentences. We extend our investigation of classifier predicates to LIS (Italian Sign Language), which offers an argument for allowing some classifier predicates to behave like normal verbs; the corresponding theoretical refinement is developed in Section 10. Going full circle, we revisit the status of intensional constructions in Section 11, and draw some general conclusions in Section 12. (An Appendix refines our initial semantics for visual animations, and raw data on ASL and on gestures appear in the Supplementary Materials.)

2 The case for preverbal objects: initial findings

We start by summarizing the traditional case for preverbal arguments in iconic constructions, first in sign language classifier predicates, then in silent gestures. The literature has often been focused on SOV vs. SOV order, finding only rare instances of OSV. But the latter are present in our ASL and in our gestural data, and thus we set ourselves the limited goal of determining when the object is preferably pre- vs. post-verbal. The preference for SOV over OSV is left for future research.

2.1 Classifier predicates

Classifier predicates are sign language constructions in which a sign represents the position or movement of an entity in a highly iconic fashion; while the lexical form is conventional, its position or movement can be modulated at will and is interpreted iconically.[1] In Zucchi's (2017) terms, "classifier predicates are hybrid creatures: classifier handshapes are linguistic morphemes, movement and location of these handshapes (…) are not". We will take this combination of a conventional form with a free position or movement which is interpreted iconically to characterize classifier constructions across sign languages.

The gradient and iconic character of the information conveyed was displayed with experimental means in Emmorey and Herzig 2003. They studied a construction involving a classifier representing a small object (a sticker) relative to a handshape representing a flat object (a bar). Deaf signers were asked to provide a geometric representation of the scene. As the classifier's position was gradiently modified relative to the flat object handshape, so was the geometric representation of the scene: the signs were interpreted iconically. Zucchi 2011, 2017 discusses further examples involving the movement of an object, as in (1)a, which describes the movement of a car; it is paraphrased in (1)b.[2]

(1) a. CAR CL-vehicle-DRIVE-BY. (ASL, cited and illustrated from Valli and Lucas 2000, cited in Zucchi 2017)

[1] This paragraph follows the discussion of Schlenker 2021.

[2] For Zucchi 2011, 2017 and Davidson 2015, classifier predicates genuinely have a demonstrative component; we will not follow this aspect of their analysis, but discuss it in Section 7.4.

FIGURE 15.6 *Source: From Valli, C., & Lucas, C. (2000). Linguistics of American Sign Language: An introduction (3rd ed.). Gallaudet University Press.*

b. 'A car drove by *like this*', where the demonstration is produced by the movement of the classifier predicate in signing space (after Zucchi 2011)

In other words, the movement of the classifier predicate *CL-vehicle-DRIVE-BY* in signing space tracks in a gradient fashion the movement performed by the relevant car in real space.

Turning to the syntax of classifier predicates,[3] Liddell 1980 (pp. 94-95) discusses a detailed example in ASL in which the arguments (a fence and a cat) come before the classifier, representing a cat sitting on a fence. This is in sharp contrast with the basic SVO order of ASL (with further exceptions that Liddell discusses). The entire sequence is represented in (2).

(2) "A cat is lying on the fence" in ASL (Liddell 1980 p. 93)

As Liddell explains, the lexical sign for *FENCE* appears first, followed by a classifier (called 4-CL because it resembles the sign for 4) representing the fence, signed with the right hand; simultaneously, the lexical word for *CAT* is signed with the left hand. Then by putting another classifier, called V-CL (because it is a bent version of the sign for V) on the fence classifier, one iconically conveys that the cat was on the fence.[4] Liddell 1980 further conjectures that several other cases in which SVO order is overridden are due to iconicity.[5]

In a recent overview, Pavlič 2016 notes that the ASL pattern is by no means exceptional: in several sign languages, classifier predicates trigger changes of word order. He summarizes as follows cross-linguistic results that came after Liddell's initial observations:

Many subsequent studies on various sign languages have confirmed that classifier predicates may influence the constituent order of the sentence. In languages with a basic SVO word order such as Jordanian Sign Language (Hendriks 2007: LIU), Colombian Sign Language (Oviedo 2003: CoSL) Russian Sign Language (Kimmelman 2012: RSL), VGT (Vermeerbergen 2004) and Hong Kong Sign Language (Sze 2003: HKSL), they yield an SOV word order.

[3] See Section 20.2 of Sandler and Lillo-Martin 2006 for a discussion of other aspects of the syntax of classifier constructions.

[4] Liddell adds: "This is apparently the 'neutral' way to indicate that a cat is on a fence. That is, this configuration does not necessarily indicate the orientation of the cat with respect to the fence. However, the lack of motion of the bent-V classifier indicates that the cat was still rather than in motion. It is easy to see that there is a lot of information packed into the orientation, location, and (lack of) movement of the two classifiers."

[5] Liddell explicitly sets aside the issue of topicalization, which is independently known to trigger word movement effects in ASL. See also Liddell's discussion of the relation between iconicity and SOV (without topicalization) on pp. 89-90 of Liddell 1980.

2.2 Silent gestures

Goldin-Meadow et al. 2008 made a striking discovery: in a production task, speakers of syntactically very different languages (English, Mandarin, Spanish and Turkish) preferably used an SOV (= Actor Patient Action) word order when describing scenes in silent gestures; as they wrote, "the predominant gesture order was ArPA [= Actor Patient Action], which was identical to the predominant speech order for in-place and crossing-space actions in Turkish and for crossing-space actions in Chinese, but different from the predominant speech order for both types of actions in English and Spanish and for in-place actions in Chinese." The authors further replicated their results in a non-communicative task in which the subjects "were asked to reconstruct the same events by using sets of transparent pictures". Here the order in which they stacked the transparencies also reflected an SOV order. Langus and Nespor 2010 replicated these results in production and comprehension tasks involving Italian- and Turkish-speaking subjects (with importantly different results for more complex constructions, to which we return in Section 3.1).

Later research added important qualifications to these findings. As summarized in Hall et al. 2013, Meir et al. 2010 found that speakers of Hebrew (an SVO language) and Turkish (an SOV language) used an SOV order in pantomimes when the subject was animate and the object was inanimate; this agreed with the findings of Goldin-Meadow et al. 2008. But when both participants were animate, and the action was thus 'reversible' (involving for instance a woman, a boy, and pushing), an SVO gestural order arose in for both types of speakers. This finding was explained by the confusability of SOV when S and O are both animate and could thus equally have the roles of Actor and Patient. Gibson et al. 2013 gave an analysis of related results in terms of a rational 'noisy channel' theory of communication; its main tenet is that speakers choose the best signal to communicate the intended meaning conditional on the assumption that some of the message might be corrupted. In reversible actions, upon the loss of an argument, SVO would still be partly interpretable (yielding SV or VO), but SOV wouldn't be – e.g. *boy push* wouldn't yield information as to whether the boy was doing the pushing or was being pushed. With non-reversible actions, involving for instance a boy, a box and some opening, the advantage of the SVO order disappeared (as one could infer from *boy open* that the boy was the agent and from *box open* that the box was the patient).

In a pantomime-based task with English-speaking subjects, Hall et al. 2013 found a more nuanced picture: on the one hand, they confirmed the SOV preference for the representation of non-reversible events. On the other hand, they found that for the representation of reversible events, SOV was avoided and other orders emerged.

What all these studies have in common is the fact that in the representation of non-reversible events, an SOV order is preferred. But even in the non-reversible case, further distinctions are needed, as we will now see.

3 Refining the generalization: extensional vs. intensional constructions

Going beyond these results, several studies found that in silent gestures and in some sign language constructions, the choice between SOV and SVO in non-reversible constructions is determined by semantic considerations: extensional constructions give rise to an SOV preference, intensional constructions to an SVO preference.

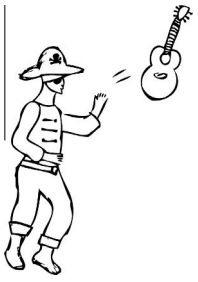
3.1 Schouwstra and de Swart 2014, Schouwstra et al. 2019 and Langus and Nespor 2010 on silent gestures

Schouwstra and de Swart 2014 argue that with extensional verbs, such as *throw*, an SOV order is indeed preferred in silent gesture production, in line with earlier research. But with intensional verbs such as *think of*, SVO order is preferred. Importantly, all the cases were 'non-reversible' and thus correspond to a case which, according to earlier generalizations, should have yielded robust SOV preferences across the board.

Elicitation of gestural sequences was conducted by way of vignettes such as those in (3).

(3) Two vignettes from Schouwstra and de Swart 2014

a. Extensional case



Example item: intensional event. 'Pirate throws guitar'.

. Example item: extensional event. 'Cook thinks of sock'.

Schouwstra and de Swart follow Forbes (e.g. 2020) in taking three features to characterize direct objects of intensional transitive verbs: "(1) resistance to substitution (i.e., Mary admires Mark Twain does not necessarily mean the same as Mary admires Samuel Clemens); (2) the possibility of an non-specific reading (such as in the sentence Mary is looking for a man, but not one in particular), or (3) existential neutrality (i.e., a sentence like John is looking for a unicorn is possible, in which the unicorn does not exist)."

One useful test for our purposes involves indefinites in object position: with extensional transitive verbs, they can be paraphrases with wide scope indefinites, as in (4)a-b, but with intensional transitive verbs, this typically fails: (4)b' doesn't follow from (4)a' (this corresponds to properties features (2) and (3) of Schouwstra and de Swart's characterization).

(4) Failure of existential exportation in a'.-b'. but not in a.-b.
a. The cook threw a saxophone.
b. There is a certain saxophone that the cook threw
a'. The cook wanted a saxophone.
b'. There is a certain saxophone that the cook wanted.

Schouwstra and de Swart 2014 provide the following justification for the distinction between SOV and SVO in extensional vs. intensional contexts:

> Direct objects that are arguments of extensional verbs refer to concrete objects that are identified as existing independently of the event, but intensional verbs take direct objects that are possibly non-specific or non-existent. This makes direct objects in intensional events more abstract and more dependent on the action than those in extensional events, and this is, we hypothesise, a reason to describe them after the verb.

As we discuss in Section 11, it is not entirely obvious why intensional objects should of necessity be construed as being "more dependent on the action" than extensional objects, but we will set this point aside for the time being.

Schouwstra et al. 2019 confirm these generalizations (SOV for extensional objects, SVO for intensional ones), but from the perspective of interpretation. They show subjects ambiguous gesture sequences, e.g. ones in which the gestural verb+object sequence could mean 'to build a house' (with an intensional object) or 'to climb a house' (with an extensional object), and they find that SOV weakly favors the extensional reading while SVO weakly favors the intensional reading.[6]

Let us add that Schouwstra and de Swart's results subsume earlier results by Langus and Nespor 2010 on gestural descriptions of speech- and thought-acts: while Langus and Nespor replicated Goldin-Meadow et al.'s (2008) SOV preference for the gestural description of simple actions, for speech- and thought-acts, Italian and Turkish speakers alike went with SVO (an order that went against the syntax of Turkish). The authors took their result to argue for a split between a grammar-free (cognitively based) SOV word order for simple event descriptions and a grammatical SVO order for embedding-like

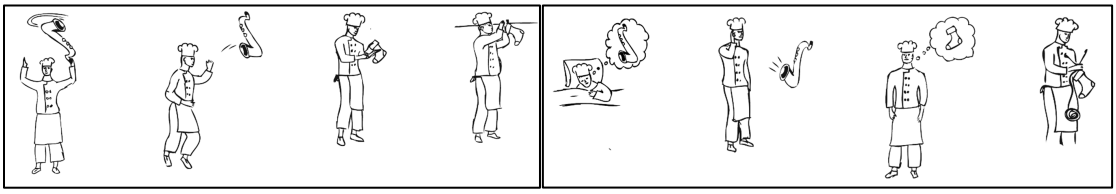
[6] In the authors' words, "in comparison to the production experiment", in comprehension "the effect of word order on meaning in interpretation is modest".

structures.[7] But an alternative is that SVO order emerged in attitude reports because these were intensional, as the (less complex) constructions in Schouwstra and de Swart's experiment.

3.2 Napoli, Spence and Quadros 2017 on Libras

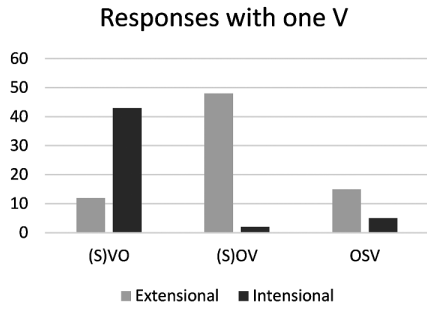
Napoli et al. 2017 proposed to apply to Libras (Brazilian Sign Language) the main ideas advocated by Schouwster and de Swart. As mentioned at the outset, Napoli et al. 2017 justified the import of the extensional vs. intensional distinction in terms of "visualization": "arguments that are present on the scene before an action takes place precede the V; those that are not follow the V". To test this intuition in Libras, they adopted Schouster and de Swart's vignette-based method to elicit descriptions, but now in signs rather than in gestures. Their conditions were as in (5)a, with illustrative vignettes as in (5)b.

(5) a. "The extensional events include the predicates 'swing', 'throw', 'climb', 'eat', 'carry', 'drop', 'paint' (that is, apply paint to a surface, rather than depict by painting), 'hang' (on a washing line), 'cut' (with scissors), and 'slice' (with knife or pizza wheel). The intensional events include the actions 'knit', 'want', 'look for', 'build', 'dream of', 'hear', 'sculpt', 'think of', 'see', and 'draw'."
b. Illustrative vignettes from Napoli et al. 2017's Figure 1
(i) Extensional vignettes (ii) Intensional vignettes



The results go in the authors' expected direction, as shown in (6) for productions that contained a single verb (= V; some productions contained several); subjects (= S) appear in parentheses because they could be omitted, but objects (= O) were not.

(6) Napoli et al. 2017: orderings found for extensional vs. intensional gestural constructions, productions with just one verb



Intensional gestural verbs clearly give rise to a preponderance of (S)VO order while extensional gestural go with (S)OV instead (it is worth observing that there are some OSV orders as well, a point that applies to our ASL and gestural data as well). Unfortunately, Napoli et al. note that their results are confounded by another distinction: verb-final word orders might be triggered by the fact that the verb is morphologically heavy rather than intensional. The problem is that the extensional predicates elicited 'heavy' verbs, i.e. verbs that are "morphologically or prosodically complex". As a result, "one might propose that all instances of OSV and SOV are produced via fronting of the O". It is thus essential to compare SVO and SOV orders *within the class of heavy verbs*; when this refinement is performed, their

[7] In Langus and Nespor's (2010) words: "SOV is the preferred constituent order in the direct interaction between the sensory-motor and the conceptual system; the SVO order is preferred by the computational system of grammar."

counts still suggest that extensional verbs usually prefer preverbal objects and intensional verbs post-verbal ones.[8]

An additional confound was that in extensional constructions the verb was "typically realized as a handling classifier";[9] in our study of ASL below, we will tease apart the specific role of classifier constructions in triggering word order changes, in line with Liddell 1980 and Pavlič 2016.[10]

3.3 Intermediate summary and proposal

In sum, in the classifier constructions of diverse sign languages and in the silent gestures of subjects speaking different spoken languages, preverbal objects are sometimes preferred. But Schouwstra and de Swart 2014 showed that in gestures, the preference for preverbal objects is determined by the semantics of the predicate: extensional constructions prefer SOV, intensional constructions prefer SVO order. Napoli et al. 2017 displayed a similar generalization in Libras, but without disentangling what is due to the semantics *per se* and to the form of classifier constructions. Both studies posited that complements of intensional constructions are preferably postverbal because in a sense their denotations do not exist 'before' the action described.

Importantly, there is an interesting discrepancy between Napoli et al.'s analysis and their stated generalization. In their view, the key "visualization": unlike the object of an intensional verb, "the preexisting arguments of an extensional event are already somewhere in our mental picture before the predicate is articulated". While it might make intuitive sense to assume that objects of intensional verbs do not "preexist" in this way (although a formal account is non-trivial), for extensional verbs the *visibility* of the object should depend on the construction under study. This is particularly important if the source of the phenomenon is visual iconicity, as one would expect in sign language and in silent gestures alike.

In the rest of this article, we will thus focus on minimal pairs involving a contrast between (say) *eating up a ball* and *spitting out a ball*. Both are extensional constructions, as seen for instance by possibility of exporting indefinites in both cases, as in (7).

(7) a. The crocodile ate up a ball.
=> there is a certain ball that the crocodile ate up
b. The crocodile spit out a ball.
=> there is a certain ball that the crocodile spit out

---

[8] We note that not all the situations described in (5) would lead one to expect an intensional behavior, at least according to (4): when one sees or hears a saxophone, there is a typically a certain saxophone that one sees or hears, unless one is under an illusion of some sort. Visually representing this as in (5)b(ii) (where the saxophone being heard is real) only strengthens the inference that there is a certain saxophone which the cook heard, and which was present in the seen before the action. And in fact, 'hear' and 'see' belong to the exceptions to the authors' generalization that the object should follow the verb, as seen in the cases of in the authors' Table A2; this might be because the vignettes failed to elicit an intensional construction (and also one whose object would typically be seen after the action).

[9] Specifically, Napoli et al. 2017 write: "…in all of our extensional materials, a person interacts with a concrete object (…) and many times the person moves that object ('swing', 'throw', 'carry', 'drop', 'hang') (…) which means that the verb is typically realized as a handling classifier. And even when the object is not moved, as with 'climb' in 'gnome climb tower', the verb is realized as a handling classifier since the hands curl to grasp the bricks in the tower. Also, with these extensional verbs the person usually affects that object ('eat', 'cut', 'slice', 'paint'), which again means that a classifier will typically appear. In many of these instances, the nondominant hand is a classifier for the O, and the dominant hand acts on it. So, for example, the witch will paint a wall, where, after signing both WITCH and WALL, the signer will articulate the predicate, and now the nondominant hand will be a classifier for the wall and the dominant hand will act on it (making the up-down movement of painting)."

[10] It should be added that Koraka 2021 (citing precedents in other sign languages such as BSL [Sutton-Spence and Woll 1999]) found related results in Greek Sign Language: with predicates such as *BUILD, MAKE* and *BAKE*, SVO order was preferred. (We note that these predicates are straightforward to analyze in terms of an iconic semantics as developed in this paper: the object of the creation is typically visible after the action and not before.)

In fact, the extensional nature of *spitting out* will be particularly clear in our examples: the context will explain that a crocodile ingested a ball before spitting it out. In such cases, it is clear that on all three criteria mentioned by Schouwstra and de Swart, the construction is extensional:
(i) It does not resist substitution: if the ball ingested is also Ann's most treasured possession, *the crocodile spit out the ball* has the same truth conditions as *the crocodile spit out Ann's most treasured possession*.
(ii) The construction does not allow for a non-specific reading: if the crocodile spit out a ball, then there is a particular ball that the crocodile spit out.
(iii) It is not existentially neutral: if the crocodile spit out a ball, then there is some existent thing that the crocodile spit out.
Crucially, however, one might expect the ball to be visible before the action in eat-up-type cases but after the action in spit-out-type cases, and thus in iconic terms we might expect spit-out-type constructions to display post-verbal objects.

We will show that in ASL paradigms involving classifier predicates as well as in pro-speech gestures, this expectation is met: arguments are preferably preverbal for eat-up-type verbs while the object is preferably postverbal for spit-out-type verbs.

The spirit of Napoli et al.'s account further leads one to expect that these word order preferences are tied to the iconic nature of the constructions. The role of iconicity will be seen in ASL: by investigating both classifier predicates and plain verbs with similar meanings, we will show (in line with Liddell and followers, but in a more controlled fashion) that classifier verbs are indeed the driving force behind the contrast, at least for our main ASL consultant.

Finally, Napoli et al. make reference to iconic principles but do not provide an explicit semantics for them.[11] We will develop one, based on a Greenberg/Abusch semantics for visual narratives, extended here to visual animations. This formal analysis will have three benefits. First, it will explain why the object is preferably realized before the predicate in eat-up-type constructions, but after the predicate in spit-out-type constructions. Second, it will capture the mixed character of predicate classifiers: their lexical form is conventional, but their position and movement in signing space is interpreted iconically. Third, our semantics will allow iconic representations to introduce variables in iconic representations, which will allow for anaphoric connections between gestures or classifiers and the sentences they are embedded in.

4 ASL classifier predicates

We will now show that, for our ASL consultant, classifier predicates trigger a change of word order preferences: SVO is the unmarked word order with non-classifier verbs (it can be overridden by considerations of information structure); SOV or OSV are preferred orders for classifier predicates of the eat-up-type, while SVO is preferred with spit-out-type predicates. (We will not replicate the strong preference for SOV over OSV from the earlier literature; and as we will see, for pro-speech gestures we will find some acceptable OSV orders as well.)

4.1 Elicitation methods and transcription conventions

The consultant (and co-author) is a Deaf, native signer of ASL (of Deaf, signing parents).[12] Elicitation was conducted using the 'playback method', described for instance in Schlenker et al. 2013, Schlenker and Lamberton 2019, to appear. It involved repeated quantitative acceptability judgments (1-7, with 7 = best), as well as inferential judgments. Since the acceptability facts are subtle, we maximized the chance of finding contrasts by presenting any given paradigm on a single video, and asking (a) how acceptable each sentence was, and (b) which member of a triple of sentences (in each group labelled (i) or (ii) below) was 'best'. (b) proved useful in case several sentences were tied on a 7-point scale.

[11] Without developing a formal account, they write: "Signing space may be likened to a canvas, with time as an added dimension. This iconicity is fundamental to understanding how sign languages work."
[12] We use the term *consultant* to refer to a collaborator that assesses sentences, including if this person is also a contributor to the article.

Notationally, we boxed the lettered reference (e.g. a., b.) of sentences taken to be 'best' in any given triple (if the consultant offered several options or made different choices in different judgment tasks, this is indicated). References such as *(ASL, 35, 1680; 3 judgments)* at the end of paradigms cross-reference the ASL video (here video 35, 1680) and indicate the number of iterated judgment tasks (on different days). For clarity, we provide links to anonymized versions of the source ASL videos, and specialists are invited to consult the raw judgments in the Supplementary Materials when relevant.

Transcription conventions are standard for sign language (similar ones will be used in our discussion of LIS in Section 9). Loci are alphabetized from dominant to non-dominant side (here: from right to left – typically with *a* to the signer's right and *b* to the signer's left). A suffixed locus, as in *WORD–i,* indicates that the word points towards locus i (a position of signing space associated with a discourse referent). *IX-i* (for 'index') is a pointing sign towards locus *i*. $WORD_i$ is used for a word associated with locus *i* by virtue of being signed in (rather than by pointing to) the corresponding area of signing space. Agreement verbs include loci in their realization – for instance the verb *a-HIT-b* starts out from locus *a* and targets locus *b*. We put *-cl* at the end of classifier predicates; in our examples, *-cl* co-occurs with loci because the predicates involve a movement from one locus to another (e.g. *a-SWIM-SWALLOW-cl-b*). In some examples cited from the earlier literature, ^ above a *WORD* is used to indicate that it co-occurs with raised eyebrows (= Brow Raise).

4.2 Eat-up-type verbs

For our consultant, a plain verb such as *EAT* preferably comes with an SVO order, as shown in (8)(ii)a, in line with the usual SVO preference of ASL. By contrast, when *EAT* is replaced with a classifier predicate representing a whale swimming and then swallowing a ball, as in (8)(i), our consultant prefers preverbal arguments – with a preference for OSV in this case, as in (8)(i)c.

Reminder: In each triple of sentences in (i) and (ii), we box the lettered reference of any sentence that was selected as 'best' in at least one judgment task. Note also that we provide still images for some examples for illustrative purposes, using SVO irrespective of whether this is the preferred order or not.

(8) (i) Classifier predicate
a. SVO: [6] YESTERDAY
$WHALE_a$ a-SWIM-SWALLOW-CL-b $PERSON_b$.




b. SOV: [7] YESTERDAY $WHALE_a$ $PERSON_b$ a-SWIM-SWALLOW-cl-b.
c. OSV: [7] YESTERDAY $PERSON_b$ $WHALE_a$ a-SWIM-SWALLOW-cl-b.
'Yesterday a whale swam and swallowed a person.'

(ii) Plain verb
a. SVO: [7] YESTERDAY WHALE EAT PERSON.
b. SOV: [4.3] YESTERDAY WHALE PERSON EAT.
c. OSV: [5] YESTERDAY PERSON WHALE EAT.
'Yesterday a whale ate a person.'
(ASL, C1 (35, 1664c); 3 judgments)
Anonymized video: https://www.dropbox.com/s/n85f9lf2rs4e8ml/1%20change%201664_processed.mov?dl=0

Similar but weaker contrasts were found in the paradigm in (9).

(9) (i) Classifier predicate
a. SVO: [6] YESTERDAY $CROCODILE_a$ a-CRAWL-SWALLOW-cl-b $BALL_b$.
b. SOV: [7] **YESTERDAY $CROCODILE_a$ $BALL_b$ a-CRAWL-SWALLOW-cl-b.** [preferred in 1/3 judgment]

c. OSV: [7] YESTERDAY $BALL_b$ $CROCODILE_a$ a-CRAWL-SWALLOW-cl-b. [preferred in 2/3 judgments][13]
'Yesterday a crocodile went over to a ball and ate it.'

(ii) Plain verb
a. SVO: [7] YESTERDAY CROCODILE EAT BALL.
b. SOV: [6] YESTERDAY CROCODILE BALL EAT.
c. OSV: [7] YESTERDAY BALL CROCODILE EAT.
(ASL, C4 (35, 1614c); 3 judgments)
Anonymized video: https://www.dropbox.com/s/k5ogo69ziblbpq9/4%20change%201614_processed.mov?dl=0

The 3-sentence paradigms in (8)(i) and (9)(i) involve a subject and object associated with loci, and classifier predicates that move between these loci. By contrast, the plain verb constructions in (8)(ii) and (9)(ii) involve a subject, and object and a verb that are all signed in a neutral locus. The reason we did not assign loci to the subject and object was that our consultant may dislike introducing loci without a reason, i.e. when they are optional and do not get used in later discourse: associating the subject and object with a locus could have artificially degraded the plain verb sentences. But a downside of this choice is that the comparison between the paradigms in (i) and (ii) isn't entirely minimal. We address this problem by considering agreement verbs, which are not classifier predicates but include loci in their realization.

4.3 Agreement verbs as controls

While subtle, the preceding paradigms suggest two initial results: (i) Preverbal arguments are somewhat dispreferred with plain verbs. (ii) SVO is somewhat dispreferred with classifier predicates. This raises two questions. First, could (i) be due to the absence of loci, which might make thematic role assignment unclear in non-canonical (non-SVO) configurations? Our consultant made explicit suggestions to precisely that effect (see the Supplementary Materials). Second, could (ii) be due to the presence of loci in our classifier constructions?

To address these questions, we investigate control paradigms that involve agreement verbs: unlike classifier predicates, their movement is conventionally specified and needn't be iconically interpreted; but like classifier predicates, they include loci in their realizations. These control paradigms will suggest that the dispreference for preverbal arguments with plain verbs is at least in part due to the absence of loci, as it largely disappears with agreement verbs. They will also suggest that the dispreference for SVO order in classifier predicates is not due to locus presence, at least for our main ASL consultant: agreement verbs do not display this dispreference.

Our classifier predicates involved a movement from one locus to another. For comparison, we investigated agreement verbs that display both subject and object agreement. For instance, *HIT* involves a movement from the subject locus to the object locus, as illustrated in (10)(i)a. In addition, *HIT* can also be used as plain verbs signed in a neutral position, an option displayed in (10)(ii)a, where the subject and object are also signed in a neutral locus. There are two main findings. (i) First, Preverbal arguments are more acceptable with the agreement version than with the plain verb version. This suggests that loci *per se* play a role, presumably by making explicit the thematic roles of the subject and object. (ii) Second, unlike what happened with classifier predicates, SVO order is not at all dispreferred in agreement verbs (nor in plain verbs). This establishes that, for our main consultant, the dispreference for SVO orders with classifier predicates is not due to loci, but to the use of a classifier predicate. This confirms earlier generalizations from the literature, but now by way of well-controlled examples

(10) (i) Agreement verb
a. SVO: [7] YESTERDAY
$WHALE_a$ a-HIT-b $PERSON_b$.

[13] In the sessions of 12.04.25 [by email], the consultant wrote: "(b) is my favorite here. (c) is equally good, but since BALL is inanimate, I think the most important thing here is the relative locations and the action, which is emphasized more in (b)."

(preferred in 3/3 judgments)
b. SOV: [6.3] YESTERDAY WHALE$_a$ PERSON$_b$ a-HIT-b.
c. OSV: [7] YESTERDAY PERSON$_b$ WHALE$_a$ a-HIT-b. (preferred in 1/3 judgment)
'Yesterday a/the whale hit a/the person.'

(ii) Plain verb
a. SVO: [7] YESTERDAY WHALE HIT PERSON.

(preferred in 3/3 judgments)
b. SOV: [5.7] YESTERDAY WHALE PERSON HIT.
c. OSV: [5.7] YESTERDAY PERSON WHALE HIT.
'Yesterday a/the whale hit a/the person
(ASL, 35, 1680; 3 judgments)
Anonymized video: https://www.dropbox.com/s/dojfrohs7n3524h/IMG-1680%20copy_processed.mov?dl=0

The same conclusions can be drawn on the basis of another paradigm involving *LOOK* rather than *HIT*, as illustrated in (11).

(11) *Context:* We are discussing a crocodile and a ball.
(i) Agreement verb
a. SVO: [7] YESTERDAY CROCODILE$_a$ a-LOOK-b BALL$_b$. (preferred in 1/4 judgments)
b. SOV: [6] YESTERDAY CROCODILE$_a$ BALL$_b$ a-LOOK-b.
c. OSV: [7] YESTERDAY BALL$_b$ CROCODILE$_a$ a-LOOK-b. (preferred in 3/4 judgments)
'Yesterday, the crocodile looked at the ball.'

(ii) Plain verb
a. SVO: [7] YESTERDAY CROCODILE LOOK BALL.
b. SOV: [5.8] YESTERDAY CROCODILE BALL LOOK.
c. OSV: [6.8] YESTERDAY BALL CROCODILE LOOK.
'Yesterday, the crocodile looked at the ball.'
(ASL, 35, 1636; 4 judgments)
Anonymized video: https://www.dropbox.com/s/5gjt1g9mk52u1br/IMG-1636%20copy_processed.mov?dl=0

These results shouldn't be surprising: at this point, we have just replicated in a controlled and special case the observations of Liddell 1980. Although the contrasts are subtle, we find that SVO order is dispreferred by our consultant with eat-up-type classifier predicates but not with plain verbs or agreement verbs. Agreement verbs are particularly useful because, like the classifier predicates under study here, they involve a movement between two loci; but unlike classifier predicates, their movement isn't interpreted iconically (as it is lexically specified).[14]

[14] Two remarks should be added.
(i) As we briefly discuss below, our second ASL consultant prefers SOV order for *a-LOOK-b*, but not for *a-HIT-b*, hence partly disagreeing with our main consultant about *a-LOOK-b*.

We will now see a more radical argument that locus use *per se* isn't responsible for the preference for preverbal arguments for eat-up-type classifier verbs: with matched classifier verbs of the spit-out-type, SVO order re-emerges as the preferred word order.

4.4 Spit-out-type verbs

In spit-out-type classifier predicates, SVO becomes the preferred order to express the target meaning, of the form: *x spit out y*. This conclusion will have to be assessed not just in terms of acceptability but also in terms of inferential judgments: non-SVO orders may be acceptable but with an irrelevant meaning, such as *x spit on y*.

Our contexts are longer than for eat-up-type verbs because we had to explain why the whale or crocodile started out with a person or a ball in their mouth. In all cases, the context involves a plain verb *EAT* and a neutral locus for the nominals. Thus only the target clauses vary from one example to the next. The important observation in (12) does not pertain to acceptability but to meaning: with the predicate classifier in (12)(i), the intended meaning *the whale spit out the person* is obtained with SVO order but not with SOV or OSV order, as the latter two mean that the whale spit (or barked) on the person (the relevant parts of the raw data are highlighted in the Supplementary Materials). In other words, with a preverbal object, the person is presented as being visible before the action takes place, rather than being initially in the whale's mouth, as intended. No such issue arise with the plain verb *SPIT-OUT* in (12)(ii): all word orders give rise to the intended meaning (*the whale spit out the person*), with a preference, as in our other plain verb cases, for the SVO order.[15]

(12) DAY-BEFORE-YESTERDAY WHALE EAT PERSON.
'The day before yesterday a whale ate up a person.

(i) Classifier predicate
a. SVO: [6] YESTERDAY WHALE$_a$ a-SWIM-SPIT-OUT-cl-b PERSON$_b$.

(ii) There is also a possible discrepancy between our generalization (based on our main ASL consultant) and the findings of Fischer 1975, who argues that *when loci are used* preverbal arguments are more natural. As she writes (p. 19) about the English sentence *the girl kicked the boy,* "the most straightforward way" of translating it "*if the sentence occurs in isolation*" (her emphasis) is by way of (i), which presumably involves a neutral locus:

(i) GIRL KICK BOY

But "if the sentence occurs in a more extended discourse, especially if one will wish to refer to the boy and/or the girl again", the "most natural way" to sign the sentence would be with (ii)a, although reversal of the arguments is possible as in (ii)b; we take these correspond to contemporary glosses as in (iii)a-b respectively.

(ii) a. BOY (HERE) (right hand) GIRL (HERE) (left hand) SHE-KICK-HIM left "kicks" right, from direction of location of girl to direction of location of boy
b. GIRL (HERE) (left hand) BOY (HERE) (right hand) SHE-KICK-HIM left "kicks" right, from girl to boy)

(iii) a. BOY$_a$ GIRL$_b$ b-KICK-a
b. GIRL$_b$ BOY$_a$ b-KICK-a

[15] There are precedents for the story in (12), not all of which are of dubious reality; see for instance the Old Testament, Book of Jonah; and the BBC's story entitled "Humpback whale gulps and spits out Cape Cod lobsterman" (retrieved on June 13, 2021 at https://www.bbc.com/news/world-us-canada-57450685).

(most natural for the intended meaning in 3/3 judgments)
Yesterday, the whale **spit out the person**.'
b. SOV: [6.3] YESTERDAY WHALE$_a$ PERSON$_b$ a-SWIM-SPIT-OUT-cl-b.
Yesterday, the whale **spit/barked on the person**.'
c. OSV: [6.3] YESTERDAY PERSON$_b$ WHALE$_a$ a- SWIM-SPIT-OUT-cl-b. (most natural in 1/3 judgment)
Yesterday, the whale **spit/barked on the person**.'

(ii) Plain verb
a. SVO: [6.7] YESTERDAY WHALE SPIT-OUT PERSON.[16]
b. SOV: [4.7] YESTERDAY WHALE PERSON SPIT-OUT.
c. OSV: [5.7] YESTERDAY PERSON WHALE SPIT-OUT.
Yesterday, the whale spit out the person.'

(ASL, C2 (35, 1676c); 3 judgments)
Anonymized video: https://www.dropbox.com/s/3gmj13330djl269/2%20change%201676_processed.mov?dl=0

Let us turn to our crocodile-related example. As in our whale paradigm, the SVO order is the preferred one in the classifier predicate case to express the 'spit out' meaning. The SOV and OSV orders are a bit degraded, and yield in 3 judgments out of 3 a possible meaning of *spit on a/the ball;* unlike the whale-related example above, the consultant also gave in one judgment task a principal meaning of *spit out the ball* (this was not repeated in later judgment tasks).

(13) DAY-BEFORE-YESTERDAY CROCODILE EAT BALL. FINALLY
'The day before yesterday a crocodile ate up a ball. Finally,

(i) Classifier predicate
a. SVO: [6.3] YESTERDAY CROCODILE$_a$ a-CRAWL-SPIT-OUT-cl-b BALL$_b$.
Yesterday, the crocodile **spit out the ball**.'
b. SOV: [5.7] YESTERDAY CROCODILE$_a$ BALL$_b$ a-CRAWL-SPIT-OUT-cl-b.
Yesterday, the crocodile **spit out the ball** (1/3 judgments) / **spit on a/the ball** (3/3 judgments[17]).'
c. OSV: [6] YESTERDAY BALL$_b$ CROCODILE$_a$ a-CRAWL-SPIT-OUT-cl-b.
Yesterday, the crocodile **spit out the ball** (1/3 judgments) / **spit on a/the ball** (3/3 judgments).'

(ii) Plain verb
a'. SVO: [7] YESTERDAY CROCODILE SPIT-OUT BALL.
b'. SOV: [6.3] YESTERDAY CROCODILE SPIT-OUT RELEASE.
c'. OSV: [7] YESTERDAY BALL CROCODILE SPIT-OUT.
Yesterday, the crocodile **spit out the ball**.'

(ASL, C5 (35, 1632c); 3 judgments)
Anonymized video: https://www.dropbox.com/s/jlthdy3lye9dmi8/5%20change%201632_processed.mov?dl=0

In sum, we have found a difference between eat-up-type and spit-out-type verbs with respect to classifier predicates: SVO order is dispreferred for eat-up-type constructions, corresponding to situations in which the denoted object is usually visible before the action; SVO order is preferred in spit-out-type verbs if one wishes to express the target meaning, namely *x spit y out* (rather than: *x spit on y*): this corresponds to the case in which the denoted object is only visible after the action. By contrast, for plain verbs there is no clear distinction between the two cases: SVO order is always acceptable.

[16] The consultant forgot to explicitly select a 'best' sentence among (12)a',b',c' in the session of 21/04/25, but since the acceptability ratings were respectively 6, 4, 5, it is clear that a' was 'best'.
[17] In (13)b,c, in 2 judgments out of 3, 'spit on' was the only meaning obtained. In 1 judgment out of 3, a disjunction was given: "crocodile spit out the ball (or spit on another ball?)". See the Supplementary Materials for the raw data.

One remark should be added. In one judgment task (highlighted in the Supplementary Materials), our consultant noted that in the SVO order one might expect something additional to describe the spitting out. Specifically, he noted that (12)a should "have something showing the ejecting of the person", and that without this the sentence seems to be saying "the crocodile spit out there. There is a ball there now." Similarly, he noted that in (13)a "one would normally add another verb to describe the spitting out" (in both cases, he noted that without this addition, the timing of the verb and object needed to be fine-tuned).

4.5 Additional assessment

Our data have the advantage of yielding repeated and particularly fine-grained judgments from our consultant, but they are based on just one person. To start assessing their robustness, we asked another native signer (this time a hearing child of Deaf parents, and also a contributor to this article) to assess these paradigms. This independent review of the video sentences indicated that when the order was SOV with spit-out-type classifier predicates, the interpretation leaned toward "A whale spit at a person" rather than "spit out" a person. Further, our second consultant independently corroborated the initial consultant's comment that to best express the sense of a person being spit out by the whale, a depiction of the person being spit out would likely be added, even in SVO order with the spit-out-type classifier predicate. Finally, unlike our initial consultant, she preferred a preverbal object (SOV or OSV) over SVO order in the case of the agreement verb *a-LOOK-b* in (11)(i), but not for *a-HIT-b* (we have no account for these latter facts).

5 Gestural verbs

We turn to pro-speech gestures and extend our basic finding: in gestural sequences, arguments are preferably preverbal with eat-up-type verbs while the object is preferably postverbal in spit-out-type constructions.

For practical expediency, we focus on French, an SVO language. In order to integrate our investigation with recent work on gestural grammar and gestural semantics conducted within formal semantics (e.g. Schlenker 2019, 2020), we investigate our consultants' acceptability judgments on gestural sequences that are integrated in a linguistic environment. We thus depart from most earlier work on gestural sequences along three dimensions: we investigate pro-speech (= speech-replacing) gestural sequences embedded in a linguistic environment, rather than silent gestures devoid of a linguistic context; we assess acceptability and inferential judgments rather than production data[18]; and we work with consultants (10 of them) rather than with experimental subjects (an experimental extension could be conducted in future research). In this way, we compare consultant-based data from ASL with consultant-based data on pro-speech gestures, as part of paradigms (mixes of words and gestures) that are comparable to ones used in recent formal semantics.[19]

Our gestural sequences are modeled after our ASL classifier predicate constructions. The obvious difference is that classifier predicates are based on lexical forms, whereas this is not the case for pro-speech gestures. We sought to address two empirical questions. (i) What was the general acceptability of different sequences of pro-speech gestures? (ii) To what extent did they give rise to the target meanings?

As we will see, our results are in many respects similar to those obtained with classifier predicates in ASL. While SVO orders are relatively acceptable throughout (though not necessarily with the right meaning), possibly due to the pressure of French syntax, we find clear differences across example types. SOV order is preferred in eat-up-type verbs, although SVO and OSV are not impossible, and all give rise to the intended meaning ('x eats up y'). In spit-out-type verbs, by contrast, SVO order is strongly preferred, and yields the intended meaning ('x spits out y'). SOV and OSV order are

[18] Investigating the processing or interpretation of gesture sequences in perception isn't new, however; for instance, Langus and Nespor 2010, Hall et al. 2015 and Schouwstra et al. 2019 all make use of such data.

[19] If anything, integrating gestural sequences in a full-fledged linguistic environment might be expected to strengthen an SVO bias. It is thus interesting that *despite* this pressure, we found contrasts across the eat-up- and spit-out-type constructions.

dispreferred, and give rise equally to the intended ('x spits out y') and to the non-intended meanings ('x spits near y'), both at a fairly low level.

5.1 Elicitation methods and transcription conventions

We designed a survey and collected the judgments of 10 consultants who had prior experience giving acceptability judgments (several of them are linguists, none of them are signers).[20] Acceptability was assessed, as in ASL, on a 7-point scale. We also assessed inferential strength on a 7-point scale (with 7 = strong inference); this went beyond the methods used in ASL, where inferences were described in words by the consultant, without using a 7-point scale. Since a key finding in ASL was that spit-out-type classifier predicates gave rise to 'spit out y' or 'spit on y' meanings depending on whether the object was post- or pre-verbal, we tested two inferences – schematically: (i) *x spit out y*, and (ii) *x spit on y*. In order to ask parallel questions for eat-up-type constructions, we investigated two inferences as well – schematically: (i) *x ate up y*, and (ii) *x closed its mouth near y*. We only report average scores in the main text; the Supplementary Materials include the full raw data, as well as a link to the survey.

To justify the presence of pro-speech gestures, the beginning of the survey mentioned that the videos displayed someone "addressing a child with the goal of being particularly expressive". The linguistic contexts were stated in the historical present because we thought that past tense environments could make the gestural verbs a bit degraded, as they lacked the necessary tense marking.

To ensure that pro-speech gestures were properly understood, we initially introduce them as co-speech gestures, before using them as pro-speech gestures. For instance, the gesture for CROCODILE initially co-occurred with the French words *un crocodile* ('a crocodile'), before being used in the target gestural sequence to represent a crocodile. By analogy with the initial occurrence of the relevant lexical signs in ASL, the co-speech introduction was in a neutral position, whereas pro-speech uses were associated with gestural loci.[21]

Gestures are encoded with CAPITALS in a non-standard font. A co-speech gesture precedes the expression it co-occurs with, which is **boldfaced** and sometimes bracketed. A pro-speech gesture occurs in the linear position in which it is pronounced. As in ASL, loci are alphabetized from dominant to non-dominant side – here from right to left, typically with *a* on the speaker's right and *b* on the speaker's left. $WORD_i$ is used if the relevant gesture is produced in locus *i*. Absence of a locus indicates that the relevant gesture (here: co-speech gesture) is produced in neutral gestural space.

5.2 Eat-up-type verbs

We start by replicating the preference for preverbal arguments. As announced, we are interested both in acceptability and inferential judgments, and averages for each judgment type are thus reported in a table after each paradigm. Our first paradigm, in (14), involved a crocodile ingesting a ball.

(14)	Hier	en	me promenant	sur	les bords		du	lac,	je	vois
	Yesterday	*while*	*me take-a-walk*	*on*	*the shores*		*of-the*	*lake,*	*I*	*see*
	CROCODILE	**[un**	**crocodile]** et	BALL	**[une**	**balle]**	à	la dérive.		
	CROCODILE	***[a***	***crocodile]*** *and*	*BALL*	***[a***	***ball]***	*at*	*the drift.*		

[20] As noted in Schlenker 2020, applying standard elicitation methods to sentences with gestures is well motivated. First, Sprouse and Almeida, 2012, 2013 and Sprouse et al. 2013 argued for the general validity of introspective methods in standard linguistic judgments. Second, Tieu et al. 2017, 2018 largely confirmed with experimental means early semantic judgments on *co*-speech gestures that appeared in the literature (Schlenker 2018c). Third, Tieu et al. 2019 did the same thing for semantic judgments on *pro*-speech gestures (Schlenker 20019), as did Schlenker and Chemla 2018 for more grammatical judgments.

[21] Some consultants commented that they would have preferred for the co-speech gestures to be located, like the pro-speech gestures. One worry we had was that the subject NP component of the target pro-speech gestures might seem redundant if had already appeared as a located as a co-speech gestures, especially since the shape of the gestural verb incorporated that of the subject NP (e.g. the gestural predicate a-CROCODILE-MOVE-EAT-b includes a crocodile shape). Having the co-speech gesture appear in neutral gestural space guaranteed that the location of the subject NP wouldn't come across as redundant.

Quelques minutes plus tard, je m'approche un peu, et tout à coup je m'aperçois que
A-few minutes more late, I me come-near a little, and suddeny I me notice that

a. SVO
CROCODILE$_a$ a-CROCODILE-MOVE-EAT-b BALL$_b$.

Video: https://www.dropbox.com/s/56q1uxuy91li3cq/CROCODILE-EAT_SVO.mp4?dl=0
b. SOV
CROCODILE$_a$ BALL$_b$ a-CROCODILE-MOVE-EAT-b.
Video: https://www.dropbox.com/s/qfpnvd4e3ghl7cz/CROCODILE-EAT_SOV.mp4?dl=0
c. OSV
BALL$_b$ CROCODILE$_a$ a-CROCODILE-MOVE-EAT-b.
Video: https://www.dropbox.com/s/ls1gd41iudi21vs/CROCODILE-EAT_OSV.mp4?dl=0

'Yesterday, while taking a walk on the lake side, I see a crocodile and a ball floating. A few minutes later, I come a bit closer, and I notice that [GESTURE SEQUENCE].'

Eat up a ball	Acceptability	The crocodile swallowed the ball.	The crocodile closed its mouth near the ball.
a. SVO	5,1	6	3
b. SOV	6,2	6,5	2
c. OSV	5,9	6,1	1,8

We explored a structurally similar paradigm involving a shark eating up a dog.

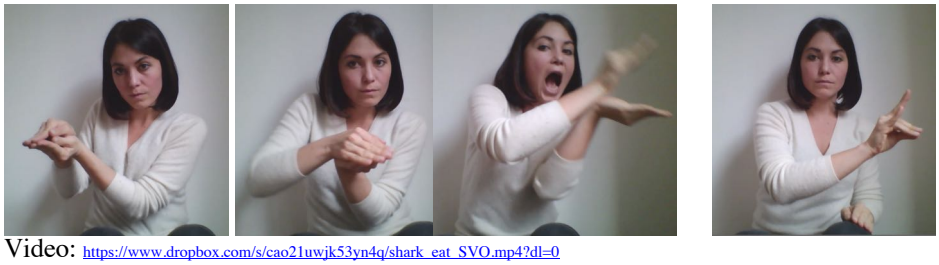
(15) Hier en arrivant à la plage, je vois
Yesterday while arriving at the beach, I see

SHARK **[un requin]** SHARK-SWIM **[qui chasse]** et
*SHARK **[a shark]*** *SHARK-SWIM **[that hunts]*** *and*

DOG **[un chien]** DOG-MOVE **[qui se promène]**.
*DOG **[a dog]*** *DOG-MOVE **[that itself takes-walk]***.

Quelques minutes plus tard, je m'approche un peu, et tout à coup je m'aperçois que
A-few minutes more late, I me come-near a little, and suddeny I me notice that

a. SVO
SHARK$_a$ a-SHARK-MOVE-EAT-b DOG$_b$.



Video: https://www.dropbox.com/s/cao21uwjk53yn4q/shark_eat_SVO.mp4?dl=0
b. SOV

SHARK$_a$ DOG$_b$ a-SHARK-MOVE-EAT-b.

Video: https://www.dropbox.com/s/nxl1242f88t73zk/shark_eat_SOV.mp4?dl=0

c. OSV

DOG$_b$ SHARK$_a$ a-SHARK-MOVE-EAT-b.

Video: https://www.dropbox.com/s/j31u29tamn4ecsc/shark_eat_OSV.mp4?dl=0

'Yesterday, arriving at the beach, I see a shark hunting and a dog walking around. A few minutes later, I come a bit closer, and I notice that [GESTURE SEQUENCE].'

	Acceptability	The shark swallowed the dog.	The shark closed its mouth near the dog.
a. SVO	6,2	6,9	1,4
b. SOV	6,6	6,7	1,8
c. OSV	5,5	4,6	3,9

Overall, we find the following results. SOV order is preferred, while SVO and OSV order are still possible. All orders strongly give rise to the target inference, of the form: 'x swallows y', rather than 'x closes its mouth near y'. While these results might suggest that syntactically anything goes, and that semantically only plausibility matters, we will see that the results are very different in spit-out-type constructions.

5.3 Spit-out-type verbs

The beginning of our spit-out-type videos explained that a crocodile or shark had ingested a ball or dog. The discourse continued with a description of what happened next, with the use of a spit-out-type gestural verb. Besides issues of acceptability, the main semantic question was whether this gestural verb was interpreted as 'x spits out y' or 'x spit something near y'.

A small issue arose as we decided on the formulation of the inferential questions. In case the sequence was interpreted as 'x spit out y', it made good sense to use a definite for y (*the ball*, *the dog*), as the simplest assumption was that the animal spit out the very thing it had ingested before. But in the case of 'x spit on y', this made little sense: if the object had not been spit out and was thus still in the animal, how could the latter spit something on it? To avoid artificially creating an incoherence, we thus used an indefinite in this case (*a ball, a dog*).

(16) Hier en me promenant sur les bords du lac, je vois
Yesterday while me take-a-walk on the shores of-the lake, I see

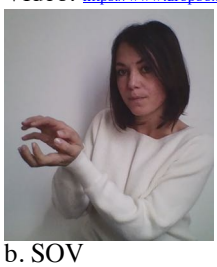
CROCODILE **[un crocodile]** avaler BALL **[une balle]**.
*CROCODILE **[a crocodile]** and BALL **[a ball].***

Maise heureusement deux minutes plus tard,
But fortunately two minutes more late,

a. SVO

CROCODILE$_a$ a-CROCODILE-MOVE-SPIT-b BALL$_b$.

Video: https://www.dropbox.com/s/7tf6weia26pyl0g/CROCODILE-SPIT_SVO.mp4?dl=0




b. SOV

CROCODILE$_a$ BALL$_b$ a-CROCODILE-MOVE-SPIT-b.

Video: https://www.dropbox.com/s/tjo27jc071jcf9t/CROCODILE-SPIT_SOV.mp4?dl=0

c. OSV

BALL$_b$ CROCODILE$_a$ a-CROCODILE-MOVE-SPIT-b.

Video: https://www.dropbox.com/s/tc1vx1wwu3dfr7v/CROCODILE-SPIT_OSV.mp4?dl=0

'Yesterday, while taking a walk on the lake side, I see a crocodile swallow a ball. Fortunately, two minutes later, [GESTURE SEQUENCE].'

Spit out the ball	Acceptability	The crocodile spit out the ball.	The crocodile spit out something near a ball.
a. SVO	6,3	6,8	1,1
b. SOV	4,5	4,4	3,1
c. OSV	3,9	3,7	4,2

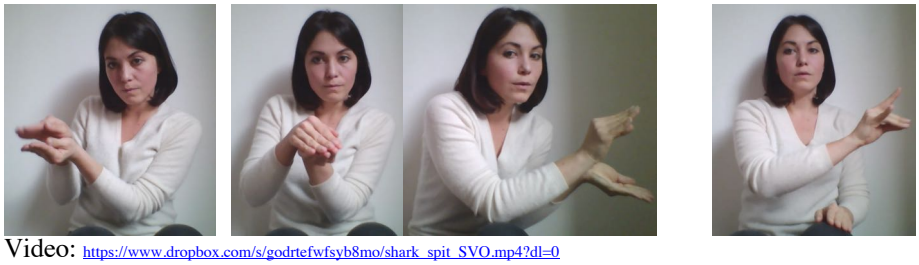
We turn to the paradigm involving a shark and a dog.

(17)

Hier	en	arrivant	à	la	plage,	je	vois		
Yesterday	*while*	*arriving*	*at*	*the*	*beach,*	*I*	*see*		
SHARK	**[un**	**requin]**	avaler	DOG	**[un**	**chien]**	qui	se	promenait.
SHARK	***[a***	***shark]***	*swallo*	*BALL*	***[a***	***dog]***	*who*	*itself*	*take-walk.*
Mais	heureusement		deux	minutes	plus	tard,			
But	*fortunately*		*two*	*minutes*	*more*	*late,*			

a. SVO

SHARK $_a$ a-SHARK-MOVE-SPIT-b DOG$_b$.



Video: https://www.dropbox.com/s/godrtefwfsyb8mo/shark_spit_SVO.mp4?dl=0

b. SOV

SHARK$_a$ DOG$_b$ a-SHARK-MOVE-SPIT-b.

Video: https://www.dropbox.com/s/xh46fzjqiyjhzat/shark_spit_SOV.mp4?dl=0

c. OSV

DOG$_b$ SHARK $_a$ a- SHARK-MOVE-SPIT-b.

Video: https://www.dropbox.com/s/u2pzube9988m317/shark_spit_OSV.mp4?dl=0

'Yesterday, as I arrive at the beach, I see a shark swallow a dog walking by. But fortunately, two mintues later, [GESTURE SEQUENCE].'

	Acceptability	The crocodile spit out the ball.	The crocodile spit out something near a ball.
a. SVO	5,8	6,5	2
b. SOV	4,4	3,6	4,3
c. OSV	3,8	4,1	3,8

While in eat-up-type constructions, SOV order was more acceptable than SVO order, in spit-out-type examples SVO order is rather clearly preferred over SOV and OSV orders. But there is also a clear semantic difference between SVO and SOV/OSV. The SVO order strongly gives rise to the intended meaning, namely that 'x spit out y'. By contrast, SOV and OSV give rise to approximately the same endorsement for the intended meaning (strongly favored by the context), and for the non-intended meaning: to a significant extent our consultants understood that 'x spit out something near y'; and both endorsements were rather low.

5.4 Summary of the gestural data and interim conclusion

Averaging over our two shark-related and crocodile-related paradigms, aggregate results appear in (18). The main findings appear in (18)(ii) (contrasting with (18)(i)): a pre-verbal object is degraded in spit-out-type examples, and it has difficulty yielding the target 'x spit out y' meaning (boldfaced numbers).

(18) Aggregate results: eat-up paradigms (= (14)-(15)) vs. spit-out paradigms (= (16)-(17))

(i) Eat-up paradigms	**Acceptability**	**Inference: x swallowed y**	**Inference: x closed its mouth near y**
a. SVO	5.65	6.5	2.2
b. SOV	6.4	6.6	1.9
c. OSV	5.7	5.4	2.9

(ii) Spit-out paradigms	**Acceptability**	**Inference: x spit out y**	**Inference: x spit out something near y**
a. SVO	6.05	6.7	1.5
b. SOV	**4.5**	**4**	3.8
c. OSV	**3.9**	**3.9**	4

In sum, the generalizations of and Napoli et al. 2017 ought to be refined. Their intuition was based on iconicity: entities corresponding to arguments of a verb are visualized as existing before the action when the verb is extensional but not when it is intensional. In line with the underlying intuition but not with the stated generalization, we argued that the visualization-based theory leads one to expect that some extensional iconic constructions should prefer SVO order in case the patient is presented as being visible after but not before the action. Unlike Napoli et al. 2017, our sign language data were controlled for verb type: eat-up-type verbs came in a classifier predicate version and in a plain verb version, as did spit-out-type verbs. The preference for preverbal arguments arose only in classifier predicates of the eat-up type: not in plain verbs, not in classifier predicates of the spit-out type. This suggests, more sharply than in the findings of Napoli et al. 2017, that our results are closely connected to iconicity, since classifier constructions differ from other verbs in being iconically interpreted (with respect to position and movement in signing space). We then extended our investigation to pro-speech gestures in French and found the same pattern: arguments were preferably preverbal in eat-up-type verbs (thus overriding the rigid SVO word order of French), but an SVO order was regained with spit-out-type verbs, and non-SVO orders had trouble yielding the target meaning ('x spit out y').

The next question is theoretical. Schouwstra and de Swart 2014 and Napoli et al. 2017 did not provide a formal semantic derivation of their generalization, let alone of the modified one we proposed (based on visibility, not intensionality). The first order of business is thus to offer a formal semantics for iconic constructions, and one that derives the visibility-based generalization. But this won't be enough: our target constructions are embedded in a fully linguistic (non-gestural) context, and thus an issue of integration will arise: how does an iconic semantics latch onto a standard compositional semantics?

We will develop the account in two stages. We start from the case of iconic gestural sequences and adapt the Greenberg/Abusch semantics for narrative sequences to visual animations. A salient property of the account, inherited from Abusch, is that animations come with variables, and these will help ensure integration with the linguistic context. We will then extend this account to the case of classifier predicates, which are more complex in that their form is conventional and only their position and movement in signing space is interpreted iconically.

6 Pictorial semantics I: Pure Iconic Case

6.1 Main idea: the importance of iconic semantics

Our main generalization is that in iconic constructions, preverbal arguments are preferred for eat-up-type verbs while SVO is preferred for spit-out-type verbs. We propose that this derives from an iconic semantics because our target constructions are construed as depicting three scenes: one corresponding to the subject with its spatial position, another corresponding to the object with its spatial position, and

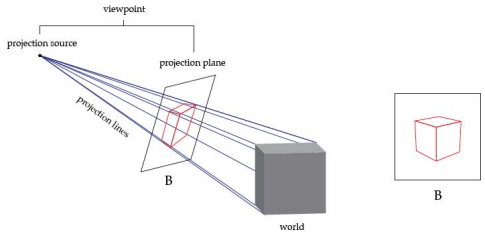
a third onecorresponding to the action (we will refine this idea in Section 10). The three-scene sequence is preferably construed as an animation, with the result that the order of the arguments reflects the temporal order in which the scenes are presented as being perceived. In eat-up-type constructions, the subject and object are presented as being visible before the action occurs. In spit-out-type constructions, the object (but not the subject) is presented as being visible only after the action occurs. Temporal iconicity is thus responsible for non-trivial facts pertaining to word order.[22]

We will develop an account with two important simplifications: (i) we will treat the 3-dimensional sign/gestural representations as if they were 2-dimensional; (ii) we will treat animations as being 'pseudo-continuous', in the sense that we analyze them as being made of a discrete sequence of pictures separated by a fixed temporal increment (an extension to continuous animations is defined in the Appendix). These simplifications will help draw a direct connection with existing work on visual narratives in the Greenberg/Abusch tradition, to which we now turn.

6.2 A Greenberg/Abusch semantics for pictorial sequences

What is the meaning of a picture? Greenberg 2013, 2021 starts from the intuition that, relative to a viewpoint, a picture is true of those worlds that can project onto the picture, as illustrated in (19) for the case of a system of perspectival projection (since our focus is not on how projections work, we will later leave out the reference to the system of projection). A viewpoint is made of a perspectival point or 'projection source' and a projection plane, as shown in (19).

(19) **An example of a projection method: perspective projection (Greenberg 2021)**



We will want to talk about visual sequences and thus it will be convenient to further relativize the semantics to times, as in (20). Here we define truth of a time; truth *simpliciter* (relative to some contextual parameters) will be regained later by existentially quantifying over times.[23]

(20) **Truth-of for a picture (modified from Greenberg 2021, adding times)**
Let π be a viewpoint, w a world, and S a system of projection. Then:
P is true of time t relative to π and w iff at t, w projects to P from π according to S, or in other words:
$\text{proj}_S(t, \pi, w) = P$.

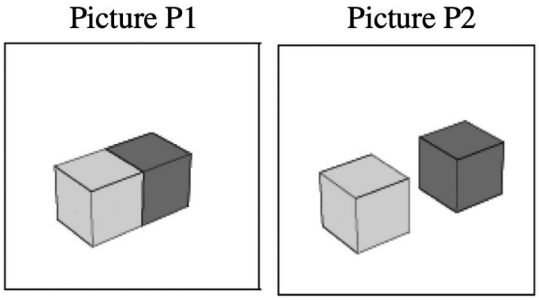
Following the spirit of Abusch's work, we can then extend Greenberg's notion of pictorial truth to temporally ordered sequences of pictures, as in the case of the 2-picture sequence in (21), from Abusch and Rooth 2017, which represents "a short comic of two cubes moving apart".

---

[22] In the framework of Cognitive Linguistics, temporal coherence is used as an iconic criterion to determine the order between main and subordinate adjunct clauses in spoken languages (Givón and Talmi 1993, Haiman 1983, Li 2016). In a sense, we are using related ideas to determine the order of predicates and arguments in highly iconic constructions in signs and gestures.

[23] Two remarks should be added.
(i) We could replace … *is true of time t*… with … *is true at time t*…, treating t as an additional parameter with respect to which truth is relativized. Nothing hinges on this notational point; we treat times separately because we need to quantify over them when we analyze visual narratives.
(ii) Schlenker, to appear, make use of eventualities to allow for a unification between pictorial semantics and music semantics, which in turn proved helpful to give a semantics for mixed sequences – such as cartoons accompanied with music. We are not concerned with the interaction between pictures and music in the present piece.

(21) Two cubes moving apart (Abusch and Rooth 2017)

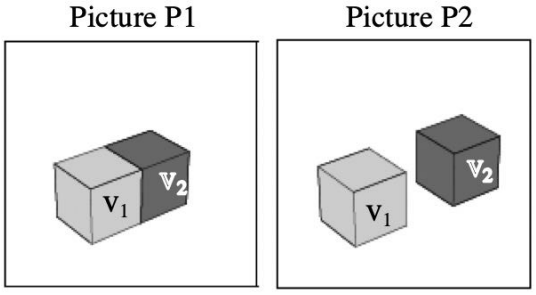


A simple notion of truth for n temporally ordered pictures can be given as in (22) (which is close to the implementation of Schlenker, to appear):

(22) **Truth-of for <u>discrete</u> pictorial sequences**
Let π be a viewpoint and w a world. Then:
A pictorial sequence of the form $<P_1, \ldots, P_n>$ is true of times $<t_1, \ldots, t_n>$ relative to π and w iff
(1) $t_1 < \ldots < t_n$, and
(2) relative to π and w, P_1 is true of t_1 and … and P_n is true of t_n.

Importantly, Abusch 2013, 2020 argues that a definition along the lines of (22) does not do justice to ambiguities that arise in visual narratives: in (21), the most natural interpretation is that the same two cubes appear in pictures P1 and P2 (hence: the two cubes move away from each other). But the semantics as given is equally compatible with a less plausible scenario in which one or both cubes were replaced with similar-looking ones between the time of P1 and that of P2. Abusch argues that this is a genuine ambiguity which should be accounted for by positing different anaphoric relations among variables present in the two pictures. On the most plausible understanding of the mini-narrative in (21), each of the two cubes is associated with a variable that remains constant across the two pictures, as depicted in (23).

(23)



Following the spirit of Abusch 2013, 2020 (but using an implementation close to Schlenker 2019b), we take a variable to be a distinguished part of a picture, with a requirement that the object the variable denotes (relative to an assignment function) should project onto that picture part. To be concrete, we can take the picture part in which the light cube appears in (23) to be a variable v_1, with the requirement that (relative to an assignment s) the object $s(v_1)$ denoted by v_1, namely the left-most cube, should in fact project onto that picture part. Having the same variable v_1 appear in P1 and in P2 in (23) will enforce coreference between the two light cubes; similar results are obtained for the cross-reference between the two dark cubes thanks to v_2.

Importantly, different tokens of one and the same variable type may correspond to different parts of different pictures: in (23), the first token of v_2 appears towards the center of P1, while the second token of v_2 appears further to the right in P2. If we wish to identify tokens of pictorial variables with picture parts, variables should really be taken to be *functions* from pictures to picture parts, as in (24):

(24) Variable v_2 in (23):

P1 → central area of P1
P2 → right-of-center area of P_2

The fact that pictorial variables 'move' from one picture to the next will be important for some linguistic applications discussed in Section 6-7.

To take variables into account, the definition of pictorial truth-of in (20) must be extended as in (25); the boldfaced condition in (25) now requires that the objects denoted by the variables project to the appropriate parts of the pictures.

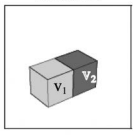
Notation: If a picture *P* contains variables among $v_1,\ldots, v_k$, we sometimes write $P[v_1,\ldots, v_k]$ to remind ourselves of this fact.

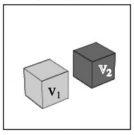
(25) **Truth-of relative to a viewpoint, a world and an assignment function for individual pictures**
Let π be a viewpoint, w a world, and s an assignment function, and let $P[v_1,\ldots, v_k]$ be a picture containing variables $v_1, \ldots, v_k$. Then:
$P[v_1,\ldots, v_k]$ is true of time t relative to π, w, s iff relative to π, at t: w projects to P and **$s(v_1), \ldots, s(v_k)$ respectively project to variables $v_1,\ldots, v_n$ of P**.

The definition of truth-of for pictorial sequences needs to be adapted to take into account this further relativization to assignment functions, as in (26).

(26) **Truth-of relative to a viewpoint, a world and an assignment function for <u>discrete</u> pictorial sequences**
Let π be a viewpoint, w a world, and s an assignment function. Then:
A pictorial sequence of the form $\langle P_1, \ldots, P_n\rangle$ (where $P_1, \ldots, P_n$ may contain variables) is true of times $\langle t_1, \ldots, t_n\rangle$ relative to π, w, s iff
(1) $t_1 <\ldots< t_n$, and
(2) relative to π, w and s, P_1 is true of t_1 and … and P_n is true of t_n.

We can then apply these definitions to the two-picture sequence in (23), as in (27). The boldfaced condition enforces coreference between the two cubes, as is desired.[24]

(27) For $\langle P_1, P_2\rangle = \langle$ 

, 

$\rangle$, $\langle P_1, P_2\rangle$ is true of times $\langle t_1, t_2\rangle$ relative to π, w, s iff
(1) temporally, $t_1 < t_2$, and
(2) relative to π, w, s, P_1 is true of t_1 and P_2 is true of t_2,
iff
(1) $t_1 < t_2$, and
(2) relative to π, at t_1 [w projects to P_1 and **$s(v_1)$ and $s(v_2)$ respectively project to variables v_1 and v_2 of P_1**], and at t_2 [w projects to P_2 and **$s(v_1)$ and $s(v_2)$ respectively project to variables v_1 and v_2 of P_2**].

It can already be seen that the variables ensure identity of reference across pictures; for instance $s(v_2)$ should be a dark cube that projects to part v_2 of P_1, and to part v_2 of P_2 (keeping in mind that these are different picture parts: the variable 'moved' from one picture to the next), and it should be the *same* object $s(v_2)$ that projects onto both picture parts.

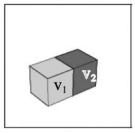
From the definition of truth-of for picture sequences, it is easy to derive a definition of truth *simpliciter* by existentially quantifying over times and assignment functions, as in (28):[25]

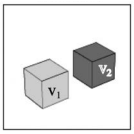
(28) **Truth relative to a viewpoint and a world for pictorial sequences**
Let π be a viewpoint and w a world. Then:
A pictorial sequence of the form $\langle P_1, \ldots, P_n\rangle$ (where $P_1, \ldots, P_n$ may contain variables) is true relative to π, w iff for some assignment function s, for some times $t_1, \ldots, t_n$, $\langle P_1, \ldots, P_n\rangle$ is true of $\langle t_1, \ldots, t_n\rangle$ relative to π, w, s.

The application to (27) is straightforward; the boldfaced $\boldsymbol{d_1}$ and $\boldsymbol{d_2}$ highlight the effect of variables in enforcing the constant identity of each cube across the two pictures.

---

[24] Schlenker, to appear, discusses a similar example from Abusch, but involving more complex pictures.
[25] We could further existentially quantify over viewpoints to obtain a definition of truth relative to a world alone.

(29) For <P_1, P_2> = < 

, 

>, <P_1, P_2> is true relative to π, w
iff for some assignment function s, for some times t_1, t_2, <P_1, P_2> is true of <t_1, t_2> relative to π, w, s,
iff for some assignment function s, for some times t_1, t_2, relative to π and w,
(1) $t_1 < t_2$, and
(2) at t_1, [w projects to P_1 and s(v_1) projects to variable v_1 of P_1], and at t_2 [w projects to P_2 and s(v_1) projects to variable v_1 of P_2]
iff for some objects d_1, d_2, for some times t_1, t_2, relative to π and w,
(1) $t_1 < t_2$, and
(2) at t_1, [w projects to P_1 and **d_1** and **d_2** are objects that respectively project to variable v_1 and v_2 of P_1], and at t_2 [w projects to P_2 and **d_1** and **d_2** are objects that respectively project to variables v_1 and v_2 of P_2].

6.3 Pseudo-continuous pictorial sequences

The classifier verbs and gestural verbs under investigation are not pictures but rather continuous visual animations, and thus the analysis of the previous section must be extended. To facilitate comparison with the Greenberg/Abusch framework, we will assume for simplicity that a continuous pictorial sequence can be divided, like a film, into a discrete picture sequence, which we will call a frame sequence, with a fixed interval τ separating two frames – e.g. with τ = 16 ms for some videos. In effect, we treat continuous sequences as pseudo-continuous ones (= discrete frames with a fixed interval). The definition of truth should, as a first approximation, require that the temporal distance between the situations[26] depicted by successive frames should also be of τ. This would work well for simple films. But in the case of sign or gestural representations, it is obvious that there is at best a proportional relation between the interval separating two frames and the interval separating the corresponding situations: to represent an airplane flying from Boston to New York, a 2-second sign or gesture representing the plane definitely doesn't have to represent a 2-second flight!

We will thus adjust our original definition of truth-of for discrete sequences (in (26)) so it can apply to pseudo-continuous sequences. The adjustment lies in the boldfaced parts of the definition in (30), which leaves open what the multiplicative parameter μ is but requires that it should remain constant for an entire sequence. We index the frame sequence with the frame interval τ, which will allow us to distinguish formally between sequences that ought to be interpreted as discrete and those that ought to be interpreted as pseudo-continuous (classifier predicates and gestural verbs will be pseudo-continuous and will carry a frame interval as a subscript, but their arguments will add discrete pictorial representations without a fixed time interval separating two frames).

(30) **Truth-of relative to a viewpoint, a world and an assignment function for <u>(pseudo-)continuous</u> pictorial sequences**
Let π be a viewpoint, w a world, and s an assignment function. Then:
A pictorial sequence of the form <P_1, …, P_n>$_\tau$ (where P_1, …, P_n may contain variables) **with a constant interval τ between successive frames** is true of times <t_1, …, t_n> relative to π, w, s iff **for some multiplicative parameter μ,**
(1) $t_1 < \ldots < t_n$ and **there is a constant interval $\mu\tau$ between successive times,** and
(2) relative to π, w and s, P_1 is true of t_1 and … and P_n is true of t_n.

Finally, it will be useful to allow several discrete or pseudo-continuous sequences to be concatenated. In film, the viewpoint would typically change between different shots, and this might well apply in signs or in gestures as well, but for simplicity we will take the viewpoint to remain constant. With this assumption, the key idea is that concatenated sequences are themselves analyzed as discrete sequences, with the natural semantics given in (31).

Terminology and notation:
(i) A sequence of times T = <t_1, …, t_n> is temporally ordered if $t_1 < \ldots < t_n$.

[26] We use the term "situations" informally here, to mean: whatever is denoted by the pictures.

(ii) If S and S' are temporally ordered sequences of times, we write S < S' just in case the last member of S (i.e. the latest time of S) precedes the first member of S' (i.e. the earliest time of S').

(31) **Truth-of relative to a viewpoint, a world and an assignment function for sequences of pictorial sequences**
Let π be a viewpoint, w a world, and s an assignment function. Let $<S_1, \ldots, S_n>$ be a sequence of pictorial sequences (the latter may be (i) discrete sequences, including ones reduced to a single picture, and (ii) pseudo-continuous sequences). Then:
$<S_1, \ldots, S_n>$ is true of $<T_1, \ldots, T_n>$ relative to π, w, s iff $T_1, \ldots, T_n$ are temporally ordered sequences of times of length $|S_1|, \ldots, |S_n|$ respectively, and
(1) $T_1 <\ldots< T_n$,
(2) relative to π, w and s, S_1 is true of T_1 and … and S_n is true of T_n.

Here too, we can recover a notion of truth from that of truth-of by existentially quantifying over times – a straightforward extension.

We sketch in the Appendix an extension of this account to fully continuous animations, but for present purposes the assimilation of animations to pseudo-continuous ones will be enough.

6.4 Illustration with pro-speech gestures

We can now illustrate the workings of the system for the case of a gestural sequence involving a crocodile eating a ball. We take the three gestures to be entirely iconic, and for simplicity assimilate $BALL_a$ and $CROCODILE_b$ to a single picture each, which we call *B* and *C* respectively. We treat *a-MOVE-EAT-b* as a pseudo-continuous sequence S_τ, which represents a crocodile moving and swallowing a ball. The target gestural representation appears in (32)a, and it is analyzed in (32)b as the sequence <C[a], B[b], S_τ[a, b]>.

(32) a. CROCODILE_a BALL_b a-MOVE-EAT-b
b. <C[a], B[b], S_τ[a, b]>

Following the analysis of sign and gestural loci as variables, we treat the positions *a* and *b* as variables, and for notational simplicity we just call them *a*, *b* (it should be remembered that the variable *a* will 'move' from one picture to the next in the sequence S_τ). While the interaction of classifier predicates with loci is standard, a similar view for gestural verbs might initially seem more surprising. But Schlenker 2020a displayed several examples in which gestural pointing can depend on a quantifier, as is the cases with indefinites in (33) – an instance of 'donkey anaphora' in gestures.

(33) Whenever I can hire IX-hand-a **[a mathematician]** or IX-hand-b **[a sociologist]**, I pick IX-a.
Meaning: whenever I can hire a mathematician or a sociologist, I pick the former.
(Schlenker 2020a; Video 3927, 1st sentence https://youtu.be/nU5xdXLV43c)

Schlenker and Chemla 2018 further displayed cases in which gestural loci interact with gestural verbs in ways that are highly reminiscent of ASL agreement verbs. Combining these precedents with Abusch's view that pictorial representations can come with variables, it seems particularly natural to treat the sequence in (32) as including loci.

At this point, we must posit elementary truth conditions for C[a], B[b], S_τ[a, b]. For maximum simplicity, we will analyze S_τ[a, b] as a sequence of three pictures $<P_1[a, b], P_2[a, b], P_3[a, b]>_\tau$ with P_1 representing a crocodile leaving a position on the right, P_2 representing the same crocodile moving between the right and the left and P_3 representing it arriving at another position on the left and eating a ball (a more realistic analysis would involve many more frames, obviously). While the variable *b*, corresponding to the ball, plausibly stays in the same part of the three pictures, the variable *a* must be identified with different picture parts corresponding to the crocodile being initially on the right, then in the middle, then close to the ball it ends up swallowing on the left; this was precisely what we saw in (23), where the variable v_2 appeared in different picture parts in P1 and in P2.

Simplified truth-of conditions for the stationary crocodile and stationary ball appear in (34). For instance, instead of just saying *C[a] is true of time t relative to π, w, s,* we paraphrase things in terms of the presence of a crocodile or ball as displayed by the relevant pictures.

(34) a. C[a] is true of time t_1 relative to π, w, s iff relative to w and π, at t_1 [a crocodile being present projects to C[a] and s(a) projects to the variable *a* of C[a]],
corresponding roughly to: in w, at t_1, a crocodile s(a) is present as displayed by C[a].
b. B[b] is true of time t_2 relative to π, w, s iff relative to w and π, at t_2 [a ball being present projects to B[b] and s(b) projects to variable *b* of B[b]].
corresponding roughly to: in w, at t_2, a ball s(b) is present as displayed by B[b].

Simplified truth conditions for the three components of the gestural verb appear in (35). The import of the assignment function will be to ensure that the same intended ball and crocodile are depicted in (34) and in each of the components of (35).

(35) Truth conditions for the components of $<P_1[a, b], P_2[a, b], P_3[a, b]>_\tau$
a. $P_1[a, b]$ is true of time t relative to π, w, s iff relative to w and π, at t the following project to $P_1[a, b]$: a crocodile leaving from a position corresponding to *a*, a ball in a position corresponding to *b*, with s(a) projecting to *a* and s(b) projecting to *b*.
b. $P_2[a, b]$ is true of time t relative to π, w, s iff relative to w and π, at t the following project to $P_2[a, b]$: a crocodile in a position corresponding to *a*[27] moving towards a position corresponding to *b*, a ball in a position corresponding to *b*, with s(a) projecting to *a* and s(b) projecting to *b*.
c. $P_3[a, b]$ is true of time t relative to π, w, s iff relative to w and π, at t the following project to $P_3[a, b]$: a crocodile reaching and opening its mouth in a position corresponding to *b*, a ball in a position corresponding to *b*, with s(a) projecting to *a* and s(b) projecting to *b*.

Since our focus will not be on $S_\tau[a, b]$ (= $<P_1[a, b], P_2[a, b], P_3[a, b]>_\tau$), we can use the following shortcut:

(36) Let π be a viewpoint, w a world, and s an assignment function. $<P_1[a, b], P_2[a, b], P_3[a, b]>_\tau$ is true of $<t_3, t_4, t_5>$ iff **for some multiplicative parameter *μ*,**
(1) $t_3 < t_4 < t_5$ and **there is a constant interval *μ*τ between successive times,** and
(2) relative to π, w and s,
$P_1[a, b]$ is true of time t_3,
$P_2[a, b]$ is true of time t_4,
$P_3[a, b]$ is true of time t_5,
corresponding roughly to: relative to π and s, in w, at t_3, t_4, t_5, a crocodile s(a) is moving towards a ball s(b) and swallows it as displayed by $P_1[a, b]$, $P_2[a, b]$, $P_3[a, b]$.

Applying (31), we obtain for (32)b the truth conditions in (37).

(37) Let π be a viewpoint, w a world, and s an assignment function. If $<t_1>$, $<t_2>$, $<t_3, t_4, t_5>$ are ordered sequences of times (of length 1, 1 and 3 respectively), then
$<C[a], B[b], S_\tau[a, b]>$ is true of $<<t_1>, <t_2>, <t_3, t_4, t_5>>$ relative to π, w, s
iff
(1) $t_1 < t_2 < t_3$[28]
(2) relative to π, w and s, $<C[a]>$ is true of $<t_1>$ and $<B[b]>$ is true of $<t_2>$ and $S_\tau[a, b]$ is true of $<t_3, t_4, t_5>$,
corresponding roughly to:
(1) $t_1 < t_2 < t_3 < t_4 < t_5$,
(2) relative to π and s, in w,
at t_1, s(a) is a crocodile present as displayed by C[a],
at t_2, s(b) is a ball present as displayed by C[b],

[27] Here it must be remembered that *a* is not in the same position in P_1 (on the right) and in P_2 (towards the center because the crocodile moved).
[28] The requirement that $t_3 < t_4 < t_5$ is already stated because $<t_3, t_4, t_5>$ is defined from the start as an ordered sequence.

at t_3, t_4, t_5, the crocodile s(a) moves towards the ball s(b) and swallows it as displayed by $P_1[a, b]$, $P_2[a, b]$, $P_3[a, b]$.[29]

(Truth (rather than truth-of) conditions can be obtained by existentially quantifying over t_1, t_2, t_3, t_4, t_5.)

Now the key is that t_1 and t_2 must precede $<t_3, t_4, t_5>$ because the entire sequence is construed as a visual animation. In other words, the crocodile and the ball are presented as being seen before the swallowing action. This makes good sense in terms of what one would see in such a scene: if the pictures represent the focus of one's attention, one will first see the event participants, the crocodile and the ball, before seeing the crocodile swallowing the ball. But this conclusion no longer holds when one replaces an action of swallowing a ball with an action of spitting out a ball. In the latter case, the ball is expected to be seen after the action has been completed, hence the SVO order among the gestural elements.

Without going into details, the truth conditions of the SVO crocodile 'spitting out' sequence in (38)b can be obtained with the same kind of derivation as in (37), and will yield the truth conditions in (39).

(38) a. CROCODILE_a a-CROCODILE-MOVE-SPIT-b BALL_b.

b. $<C[a], S'_\tau[a, b], B[b]>$
with $S'_\tau[a, b] = <P'_1[a, b], P'_2[a, b], P'_3[a, b]>_\tau$.

(39) Let π be a viewpoint, w a world, and s an assignment function. If $<t_1>$, $<t_2, t_3, t_4>$, $<t_5>$, are ordered sequences of times (of length 1, 3 and 1 respectively), then$<C[a], S'_\tau[a, b], B[b]>$ is true of $<<t_1>, <t_2, t_3, t_4>, <t_5>>$ relative to π, w, s
iff
(1) $t_1 < t_2$ and $t_4 < t_5$
(2) relative to π, w and s, $<C[a]>$ is true of t_1 and $S'_\tau[a, b]$ is true of $<t_2, t_3, t_4>$ and $<B[b]>$ is true of $<t_5>$
corresponding roughly to:
(1) $t_1 < t_2 < t_3 < t_4 < t_5$,
(2) relative to π and s, in w,
at t_1, s(a) is a crocodile present as displayed by C[a],
at t_2, t_3, t_4, the crocodile s(a) moves from the left to the right and spits something out as displayed by $P'_1[a, b]$, $P'_2[a, b]$, $P'_3[a, b]$,
at t_5, a ball s(b) is present as displayed by B[b].

In sum, the use of preverbal arguments for eat-up-type gestures and of SVO for spit-out-type gestures is exactly what one expects on the basis of a detailed pictorial semantics in which these sequences are construed as visual animations.

7 Pictorial semantics II: Mixed Case

Having established how gestural sequences can be treated (in a simplified fashion) as visual animations, and how these can be integrated to sentential semantics, we turn to the more complex case of sign language classifier predicates. These require special measure because their lexical form is conventional, and only their position and movement is iconic.

7.1 Mixed representations

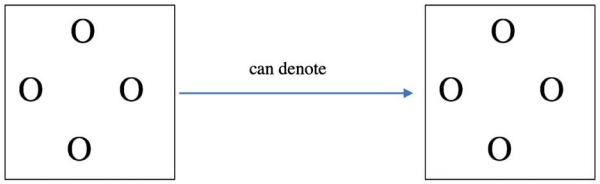
The formal problem we must address is this: we must allow for mixed representations in which some symbols are non-iconic, but their position and movement is.

A similar issue arose in a study of repetition-based plurals in ASL (Schlenker and Lamberton, to appear). In an ASL description of a drawing that involved 4 letters G arranged in a circular fashion, what appeared in signing space was (very roughly) as in (40)b on the left. These were four iterations of the manual letter  , which looks nothing like the Latin letter it denotes, G. Still, the inferential

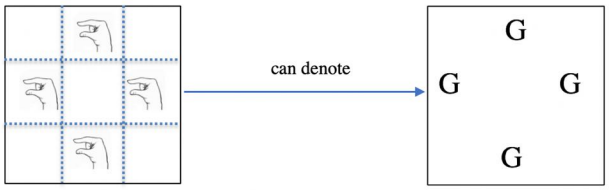
[29] In our informal paraphrase, we talk of *the* ball and *the* crocodile; the fact that these are the same ball and crocodile depicted by the earlier pictorial components is guaranteed by the assignment function.

judgments obtained suggested that the circular arrangement of the four manual letters 🤏 on the left-hand side tracked the arrangement of the G's in the drawing on the right-hand side.

(40) a. Purely iconic representation



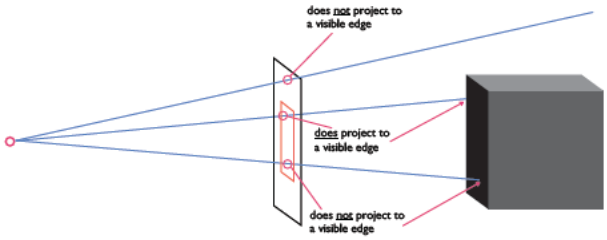
b. Semi-iconic and semi-symbolic representation



In a very simplified analysis, one can take the signing space to be divided into 9 elementary components or 'pixels', as shown in (40)b, with the rule that "a 🤏 will appear in a pixel just in case a *G* from the drawing projects onto it". By contrast, in a related case that involved a gestural representation of a little circle in an ASL sentence, a purely iconic rule could be followed: the four little circles appearing in signing space on the left-hand side of (40)a were taken to represent similar-looking shapes on the drawing, as displayed on the right-hand side of (40)a.

Schlenker and Lamberton, to appear, handled mixed cases such as (40)b by modifying the marking rules that determine what appears in an iconic representation. In a purely iconic projection, corresponding to a case discussed by Greenberg 2021, the features associated with any given element in the picture correspond to those features of the environment that project to that point. An illustration appears in (41): only visible edges are represented (as lines) on the picture (further possible marking conventions are discussed in Greenberg 2014).

(41) Illustration of a simple marking convention: only visible edges are marked (Greenberg 2014)



This simple marking rule can yield a simplified analysis of (40)a by way of the purely iconic rule in (42)a. To handle (40)b, by contrast, we need the mixed marking rule in (42)b.

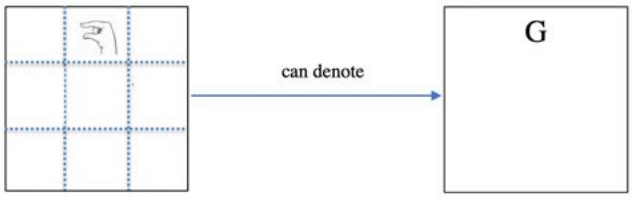
(42) a. Purely iconic marking rule: Assuming a method of projection, mark a pixel as 'black' if and only if its projection line meets the edge of an object.
b. Mixed marking rule: Assuming a method a projection, mark a pixel (which may be taken to be a large part of the 2-dimensional space) with a given word *W* if and only if its projection line meets an object that lies in the denotation of *W*.

7.2 Application to singular nouns

Our mixed marking rule was motivated by iconically interpreted repetition-based plurals, as in (40)a. But an immediate consequence of the mixed marking rule is that, in principle at least, an unrepeated sign, for instance the manual 🤏, could be interpreted semi-iconically as well, in the sense that its position in signing space could provide information about the position of its denotation, as schematized in the (imaginary) example below, where 🤏 is signed high to suggest that its denotation (G in the

relevant drawing) appears high as well. In other words, although the sign is conventional, its position in signing space is interpreted iconically.

(43) An imaginary case in which the position of a single manual  is interpreted by the mixed marking rule in (42)b

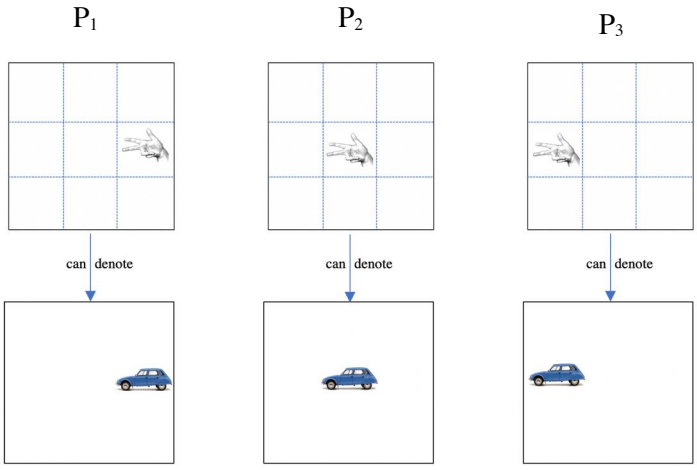


This option will prove important when we discuss the arguments of classifier predicates: even a conventional, non-iconic sign could be interpreted iconically with respect to its position in signing space. This possibility will be particularly salient for arguments of classifier predicates: their position and movement are interpreted iconically, and thus it is natural to also interpret iconically the position of their arguments in signing space.

7.3 Application to classifier predicates

To apply these ideas to classifier predicates, it is best to consider a maximally simple case, for instance one in which a vehicle classifier is used to describe the movement of a car in a cartoon or in a comic.[30] We can make use once again of the 9-pixel square in (40), now featuring a vehicle classifier in three different positions to represent three stages of a movement (in a less simplified analysis, the classifier movement would be made continuous or pseudo-continuous).

(44) **Illustration of a mixed marking rule for the movement of a vehicle classifier** (the shape of  is conventional but its movement is iconic[31])



Our earlier rules for pictorial semantics will continue to apply, but with the proviso that a mixed marking rule must be used. For instance, if we assume that the vehicle classifier in picture P_1 in (44) is associated with locus/variable *a*, it will give rise to truth conditions akin to (45); they are identical to what was delivered by (25), except that the marking rule is the mixed one described above, hence the

[30] In the cases we consider, the predicate classifier is not just conventional but also morphologically simple. Not all examples are of this type, however. For instance, in several sign languages two fingers can be used to represent two individuals, three to represent three, and these might be cases of number incorporation within the classifier. There might even be cases in which the classifier is itself the result of a syntactic operation. Our semantic analysis is neutral on this point, although it would have to take into account the semantics of these conventional forms – e.g. a classifier representing two individuals should of course have a different conventional specification than one representing one person, and the conventional part might be derived by compositional rules applying below or above the word level.

[31] The car image is from https://www.cocktail-distribution.com/p/26617-citroen-dyane-6-bleue-118-3663506002267.html (retrieved 04/09/21).

informal truth conditions we end up with (we write $P_1[a]$ because the picture includes a locus/variable *a*).

(45) $P_1[a]$ is true of time t relative to π, w, s iff relative to π, at t: w projects (according to the mixed marking rule) to P_1 and s(a) projects to variable *a* of P_1,
which rougly corresponds to: relative to π, at t: in w there is a blue car s(a) in a position corresponding to the right-most square of the middle line.

With these adjustments, we can give the same general analysis to the crocodile-related classifier construction as we did to the crocodile-related gestural case discussed above: the difference solely lies in the marking rule used for classifiers, as the crocodile-denoting manual shape is in principle allowed to have a conventional rather than purely iconic form (of course nothing prevents a classifier lexical form from being to some extent iconic, and this is in fact the case of our *CROCODILE* example; the point is that nothing in the analysis requires this to be the case, and this is important if the theory is to apply to the vehicle classifier in (1)). As in the gestural case, the SOV order in (46)a is associated with the meaning that one first saw the crocodile, then the ball, then the crocodile eating the ball. The OSV order in (47) comes with the meaning that one first saw the ball, then the crocodile, and then the crocodile eating the ball. An SVO order would yield the slightly odd meaning that one saw a crocodile, then the crocodile eating the ball, then the ball.

(46) a. CROCODILE_a BALL_b a-CRAWL-SWALLOW-b.
b. $\langle C[a], B[b], S_\tau[a, b]\rangle$

(47) a. BALL_b CROCODILE_a a-CRAWL-SWALLOW-b.
b. $\langle B[b], C[a], S_\tau[a, b]\rangle$

The ordering implications are the same when one considers the case of the crocodile spitting out a ball, as in (13), but now it is far more natural to present the ball as becoming visibly present after it was spit out.

But at this point we should go back to an observation made by both of our consultants about spit-out-type classifier constructions: while the SVO order yields the desired meaning, it has something missing. As mentioned, our main consultant once noted that (12)a should "have something showing the ejecting of the person", and that without this the sentence seems to be saying "the crocodile spit out there. There is a ball there now." Our second consultant initially remarked that "if someone really wanted to express that the whale spit out a person, there would be more information (perhaps a depiction) of the person being spit out", and later confirmed this initial impression. From the present perspective, these observations make complete sense: construed as an iconic animation, the SVO order is compatible with the intended meaning of 'spitting out', but it doesn't quite present all the stages of the ejection. The iconic representation literally yields the meaning described by our main consultant, namely: "The crocodile spit out there. There is a ball there now." And the reason is that the SVO order does not reflect grammatical composition in a logical semantics, but rather sequencing in an iconic semantics.[32]

7.4 The broader debate about classifier predicates

Following Zucchi 2017, the theoretical debate about classifier predicates can be summarized as follows. Supalla 1982 proposed that "classifier predicates are combinations of morphemes that are simultaneously articulated", hence a discrete analysis with no gradient component at all. Zucchi, along with others, took this view to have been made difficult by various experimental results (including Emmorey and Herzig 2003, mentioned above) which display clear gradience in the information conveyed by classifier predicates. As Zucchi further explains, accounts that make provisions for a gradient component are of three types. For Cogill-Koez 2000 (p. 155), "classifier predicates are best modeled, not as linguistic, but as systems of schematized visual representation created on the hands". For Liddell (e.g. 2003), classifier predicates have a morphological component, but also an analogical

[32] For reasons we come to below, we can't exclude that there is *also* a way to treat classifier predicates as standard verbs, but this is not the preferred option for our main consultant.

component. Both views are broadly compatible with the approach adopted here, with nuances. Due to the mixed marking rule we posited for classifier predicates, their lexical form need not be interpreted iconically, although their position and movement in signing space must be. The present analysis is fully compatible with the view that some classifier predicates may be morphologically complex. One important methodological point is that the present analysis, unlike those of Cogill-Koez and Liddell, is fully integrated in a formal semantics for pictures.

Zucchi 2017 also outlines a view (due to Zucchi 2011, 2012 and Davidson 2015) according to which classifier predicates are 'demonstrative predicates', in line with the paraphrase given in (1)b to the effect that "a car drove by like this", where *like this* makes reference to the classifier movement. There are several objections to this analysis. One is methodological: it does not provide an explicit semantics for the construction, unlike the projection-based analysis developed here. We could of course analyze the meaning of *like this* by making reference to this projection-based semantics, but once the latter is posited, the demonstrative component becomes unnecessary. The second objection lies in the results of the present study. Taking classifier predicates to yield visual animations (albeit with a mixed marking rule) has the advantage of explaining why signers are induced to treat not just the classifier itself but also its arguments as part of the animation, and thus as interpreting the sequence of arguments in terms of a sequence of visual scenes. It is unclear to us how this result could be matched by the demonstrative analysis. A third objection was developed in a different context in Schlenker 2018b. In a nutshell, some modulations of a helicopter path (involving an ASL classifier) were argued in Schlenker 2021 to trigger different varieties of presuppositions. But the same study included controls with explicit modifiers somewhat comparable to *like this*, and these failed to trigger the same presuppositions.[33]

A remaining question is whether the iconic analysis proposed here (broadly in the spirit of Cogill-Koez and Liddell) can explain how classifier predicates can be fully integrated with standard compositional meaning. Zucchi's and Davidson's "like this" analysis might initially seem to be at an advantage, but as we will now see, our analysis inspired by Greenberg/Abusch can seamlessly be integrated with standard compositional semantics.

8 Integrating sentential and iconic semantics

8.1 Defining a bridge

For sentences, we start from a simple intensional framework with evaluation relative to a context c, an assignment function s, a time t and a world w. We define a 'bridge' between sentence semantics and picture semantics as in (48) (for the case of a pictorial sequence) and (49) (for the case of a sequence of pictorial sequences). The key is that we can provide truth conditions for pictorial sequences relative to the very same parameters – c, s, t and w – that we use for sentential semantics. Following the Greenberg/Abusch tradition, pictures are evaluated relative to worlds and assignment functions. The time parameter will be used (in the boldfaced conditions below) as the time of evaluation of the very first picture in the narrative sequence. In addition, we stipulate that the parameters c, t, w determine a (salient) viewpoint. There will no doubt be interesting constraints to state on the relation between these parameters and viewpoints, and there will likely be cases in which viewpoints are existentially quantified, but for present purposes we will stick to the simplifying assumption that a triple c, t, w determines a viewpoint $\pi_{c, t, w}$.

(48) Intensional semantics for a pictorial sequence
Let c be a context, let t be a time, let w be a world, let s be an assignment function, and let $\pi_{c, t, w}$ be a viewpoint determined by c, t, w. Let $<P_1, \ldots, P_n>_\tau$ (where $P_1, \ldots, P_n$ may contain variables) be a pictorial sequence with a constant interval τ between successive frames. Then:

[33] Schlenker 2018b concluded: " Now our controls do not literally involve a 'like this' modifier, but rather display the relevant path after the word *WITH*. One could of course test closer analogues of 'like this' in ASL. But in any event the presuppositional or cosupposition behavior of some classifier predicates is not expected under the current *like this* analysis."

$[[<P_1, \ldots, P_n>_\tau]]^{c, s\, t, w}$ = 1 iff for some times $t_1, \ldots, t_n$ **with t = t_1**, $<P_1, \ldots, P_n>_\tau$ is true of $<t_1, \ldots, t_n>$ relative to $\pi_{c,t,w}$, w, s (according to (30)).

(49) Intensional semantics for a sequence of pictorial sequences
Let c be a context that makes available a viewpoint π_c, let w be a world, let t be a time, and let s be an assignment function. Let $<S_1, \ldots, S_n>$ be a sequence of picture sequences (which may be (i) discrete sequences, including ones reduced to a single picture, and (ii) pseudo-continuous sequences). Then:

$[[<S_1, \ldots, S_n>]]^{c, s\, t, w}$ = 1 iff for some temporally ordered sequences of times $T_1, \ldots, T_n$ **with t = the first member of T_1**, of length $|S_1|, \ldots, |S_n|$ respectively, $<S_1, \ldots, S_n>$ is true of $<T_1, \ldots, T_n>$ relative to $\pi_{c,t,w}$, w, s (according to (31)).

Our bridge between sentential and pictorial semantics makes it possible to embed animations under operators. In (16), the gestural animation is embedded under 'two minutes later'. This expression can be given a very simple analysis, on which it changes the value of the time parameter, as in (50).

(50) Let c, s, t, and w be a context, an assignment function, a time and a world, and let t+2 refer to the time that is two minutes after t. For any formula *F*,
$[[\text{two-minutes-later F}]]^{c, s\, t, w} = [[F]]^{c, s\, t+2, w}$

We can apply this semantics to (51)a, analyzed as in (51)b, with the truth conditions in (51)c.

(51) a. Two minutes later, CROCODILE$_a$ a-CROCODILE-MOVE-SPIT-b BALL$_b$.

b. two-minutes-later $<C[a], S'_\tau[a, b], B[b]>$

c. $[[(b)]]^{c, s\, t, w} = [[<C[a], S'_\tau[a, b], B[b]>]]^{c, s\, t+2, w}$
= 1 iff for some temporally ordered sequences of times T_1, T_2, T_3 with t+2 = the first member of T_1, of length 1, 3 and 1 respectively, $<C[a], S'_\tau[a, b], B[b]>$ is true of $<T_1, T_2, T_3>$ relative to $\pi_{c,s,t,w}$, w, s,
iff for some times t_1, t_2, t_3, t_4, t_5 with $t+2 = t_1$,
(1) $t_1 < t_2$ and $t_4 < t_5$
(2) relative to π, w and s, $<C[a]>$ is true of t_1 and $S'_\tau[a, b]$ is true of $<t_2, t_3, t_4>$ and $<B[b]>$ is true of $<t_5>$
Making use of the related derivation in (39), this roughly corresponds to:
for some times t_1, t_2, t_3, t_4, t_5,
(1) $t+2 = t_1 < t_2 < t_3 < t_4 < t_5$,
(2) relative to π and s, in w,
at t_1, s(a) is a crocodile present as displayed by C[a],
at t_2, t_3, t_4, the crocodile s(a) moves from the left to the right and spits something out as displayed by $P'_1[a, b], P'_2[a, b], P'_3[a, b]$,
at t_5, a ball s(b) is present as displayed by B[b].

8.2 Pictorial variables and anaphora I: gestures

Our analysis with pictorial variables comes with an immediate benefit: it yields a straightforward account of cases in which a part of an iconic representation is made available for further anaphoric uptake.[34]

We introduced a non-iconic case of gestural loci used as variables in (33). Turning to the iconic case, it is easy to see that gestural animations make available discourse referents as well. In (52) (which was not part of our formal survey), a crocodile repeatedly opening its mouth appears on the speaker's right, in gestural locus *a*. Pointing towards that gestural locus makes it possible to refer to the crocodile, as in: *I was in fear of IX-a.*

(52) a. Hier, je m' aperçois que j'ai perdu mon BALL **[ballon]** près du lac, qui
Yesterday I me notice that I have lost my BALL [ball] near the lake, which

est infesté de CROCODILE **[crocodiles]**. Je pars à sa recherche, et tout a à coup je vois ceci:
is infested of CROCODILE [crocodiles]. I leave to its search, and suddenlyI see this:

[34] See Schlenker and Lamberton 2019 for a different case in which some discourse referents are created by iconic representations (at the edge of repetitions).

BALL$_b$ CROCODILE$_a$ a-OPEN-MOUTH-REPEATEDLY.

Mais	je	ne	me	suis pas approché:	j'avais	bien	trop peur de IX-a.
But	*I*	*NE*	*me*	*was not approached:*	*I had*	*far*	*too fear of IX-a.*

'Yesterday, I notice that I lost my BALL **[ball]** near the lake, which is infested with CROCODILE **[crocodiles]**. I go looking for it, and suddenly I see this: BALL$_b$ CROCODILE$_a$ a-OPEN-MOUTH-REPEATEDLY. But I didn't come near: I was far too afraid of IX-a [i.e. of the crocodile].'

b. Je	reste là	un	moment, et	voici ce que je vois		ensuite:
I	*stay there*	*a*	*moment, and*	*here this that I see*		*then:*

BALL$_b$ CROCODILE$_a$ a-MOVE-EAT-b.

Là	je n'	ai	plus	eu	le	choix:
There	*I NE*	*have*	*no-longer*	*had*	*the*	*choice:*

j'	ai	dû	appeler	le	vétérinaire	pour	sauver	IX-b.
I	*have*	*had*	*to-call*	*the*	*vet*	*to*	*save*	*IX-b.*

'I stayed for a while, and here is what I saw next: BALL$_b$ CROCODILE$_a$ a-MOVE-EAT-b. At that point I had no choice: I had to call a vet to save IX-b [i.e. the crocodile].'

In (52)b, which is a continuation of (52)a, things get more interesting: the gestural verb involves a movement of the crocodile from position *a* to position *b*, and for this reason it becomes natural to point towards the crocodile's new position to refer to the animal. In other words, it is natural to use the expression *save IX-b* to mean 'save the [moved] crocodile'. This is as is expected given the present framework. As we saw in (23), one and the same variable – for instance v_2 in the case of the moving cubes – corresponds to different parts of different pictures: v_2 appears towards the center of P1 and towards the right of P2. The same thing happens in (52)b: the variable representing the crocodile is initially on the right, but gradually moves towards the left from one picture to the next.

8.3 Pictorial variables and anaphora II: signs

It is uncontroversial that classifier predicates are associated with loci, so the argument developed in the preceding section for gestures does not bear repeating for signs (nor is this interaction surprising in view of the complete integration of classifier predicates to the grammar of sign language). On the other hand, the phenomenon whereby loci can 'move' in signing space is widely attested empirically, but remains a bit of a mystery theoretically. This phenomenon has sometimes been called 'Locative Shift', and it is illustrated by an example that involve no obvious iconicity and no classifier predicates in (53) (from Schlenker 2013; see also Emmorey and Falgiers 2004, and Schlenker 2018a for references). *JOHN* is signed in a neutral position, while locus *a* is associated with the French city and locus *c* with the American city. While one can point towards the neutral position (in the middle) to refer to John, it is also possible (and in this case preferred) to point towards *a* to refer to John-in-the-French-city and towards *b* to refer to John-in-the-American-city.[35]

(53) [7] JOHN [WORK IX-a FRENCH CITY]$_a$ SAME [WORK IX-c AMERICA CITY]$_c$.
^ ^
IX-a IX-1a HELP **IX-a+**, IX-c IX-1c NOT HELP **IX-c+**.

[35] We have slightly adjusted the transcription conventions, putting loci introduced by expressions to their right rather than to their left. Schlenker 2013 wrote *a+* and *c+* to indicate that pointing is towards a position slightly higher than loci *a* and *c* (a notation we have preserved here). This might serve to distinguish between the person who is at the location – namely John – and the location itself. But cases of clear ambiguity are described in the literature (see Schlenker 2018a for examples and references).

'John does business in a French city and he does business in an American city. There [= in the French city] I help him. There [= in the American city] I don't help him.'
(ASL, 4, 66; Schlenker 2013: 2 judgments)

Locative Shift clearly arises in our scenarios as well. An example appears in (54). It was formed by simply adding to (9)b the sentence: *FINISHED IX-b SICK*, where the pointing sign *IX-b* indexes the position of the ball, not the original position of the crocodile. The example is acceptable because the sentence asserts that the crocodile moved towards the position of the ball, hence the latter can be coopted to refer to the (moved) crocodile.

(54) YESTERDAY CROCODILE$_a$ BALL$_b$ a-CRAWL-SWALLOW-cl-b. FINISHED IX-b SICK.
'Yesterday a crocodile went to a ball and swallowed it. And the crocodile was sick.'
(ASL, informally elicited from Lamberton[36])

In line with our paraphrase pertaining to (53), (*John-in-the-French-city*, *John-in-the-American-city*), Schlenker 2018a takes Locative Shift to be due to the fact that pronouns can optionally refer to situation stages of individuals rather than just to individuals. But for the iconic analysis discussed here, no such measures are needed: the pictorial semantics we have developed has the built-in requirement that loci-qua-pictorial-variables should move in animations, as was highlighted above. This raises a question for future research: can this case be unified with instances of Locative Shift that do not have an (obviously) iconic character?[37]

9 Extension to Italian Sign Language (LIS)

In this section, we discuss initial findings about LIS classifier predicates that are maximally similar to our ASL examples. The investigation is motivated by a syntactic difference: the basic word order of ASL is SVO, that of LIS is SOV (Cecchetto et al. 2006). In ASL, eat-up-type classifier predicates go against the basic word order of the language and prefer preverbal arguments, whereas spit-out-type classifier predicates go back to SVO. The question is what happens in LIS. Eat-up-type classifier predicates are expected to go with SOV, as this is both the basic word order of the language and the word order that ought to be preferred on iconic grounds. But what about spit-out-type classifier predicates? The findings are interesting: in line with what is expected on iconic grounds, SOV fails to yield the target 'x spit out y' meaning in sentences that parallel our ASL examples, ones in which the subject and object are assigned loci. While SVO order does yield the target meaning, however, this order is degraded (a point that parallels our second consultant's judgments for ASL). But there is an additional finding: when the object appears in a neutral locus, SOV order is regained, with the target 'x spit out y' meaning.

Since this part of our work was exploratory, we used more informal methods than in our investigation of ASL. The data were constructed jointly by two native signers (and contributors to this piece): one is the Deaf child of Deaf, signing parents, the other is the hearing child of Deaf, signing parents. We provide below the judgments on a 7-point scale of the Deaf consultant/co-author, who also signed the examples that appear on the relevant videos.

Two limitations should be noted at the outset. First, we did not consider OSV order, for reasons of simplicity, but also because it did not seem available unless markers of topicalization were used (none are present in the examples that we do discuss). Second, in order to come as close as possible to our ASL example, the LIS signers had to use a combination of two classifier predicates, one involving movement and one involving swallowing or spitting out; we do not currently have evidence that our ASL classifier predicates should be divided in this way. Importantly, eat-up-type and spit-out-type classifier constructions were still matched with each other, so whatever departure from ASL was found in one case was found in the other as well, and could not account for *differences* we found between the eat-up and spit-out cases. Third, the lexical form of the subject, *WHALE*, requires that it be signed in a

[36] Unlike the other ASL examples discussed in this piece, this one was discussed informally, as the paper was being written and after the elicitation phase.
[37] Schlenker 2018a, 2020 discusses counterparts of Locative Shift in non-iconic examples involving gestural loci. They should be revisited as part of this broader debate.

neutral locus, and as a result we did not provide a locus in the transcription. On the other hand, the object *BALL* could be localized (= assigned a locus) or not, and the difference mattered.

9.1 Eat-up-type classifier predicates

Since the basic word order of LIS is SOV, eat-up-type classifier predicates are expected to display SOV order whether they behave like LIS plain verbs or like ASL classifier predicates. This is indeed the case: with an eat-up-type classifier construction, SOV order as in (55)b is preferred over SVO as in (55)a. (We do not include control examples with plain verb *EAT* because this verb was degraded, probably for lexical reasons. Since there is little doubt that the basic word order of LIS is SOV, these examples do not play a crucial role anyway.)

(55) Classifier predicate
a. SVO: [5] WHALE a-SWIM-cl SWALLOW-cl-b BALL-b
(LIS, 17-06-21 12.48 #1)
Video: https://www.dropbox.com/s/yiwiq3i2xng86et/MS%20whale%20swallow%2017-06-21%2012.48%20%231.mov?dl=0

b. SOV: [7] WHALE BALL-b a-SWIM-cl SWALLOW-cl-b
Video: https://www.dropbox.com/s/dinyc2gkz58s27g/MS%20whale%20swallow%2017-06-21%2011.50%20%231.mov?dl=0
(LIS, 17-06-21 11.50 #1)

9.2 Spit-out-type classifier predicates

With spit-out-type classifier constructions, the SOV order is only slightly degraded, but crucially it does not yield the target meaning 'x spit out y', but rather 'x spit something on y', as seen in (56)b. This is indicative of an iconic interpretation: despite the basic word order of the language, the SOV sequence is interpreted as implying that the arguments are visible before the action. Importantly, however, our CODA consultant does get the target meaning with SOV: this is the main point of disagreement among our two consultants. For our Deaf consultant, the SVO order does yield the target meaning, but it is somewhat degraded, as seen in (56)a.

(56) Classifier predicate (localized object)
a. SVO: [5] WHALE a-SWIM-cl SPIT-OUT-cl-b BALL-b.
Available meaning: 'The whale spit out the ball.'
'Very difficult to access': 'The whale spit (something) on the ball.'
(LIS, 17-06-21 12.04 #2)
Video: https://www.dropbox.com/s/1a1pkv2etvbuirv/MS%20whale%20spit%2017-06-21%2012.04%20%232.mov?dl=0

b. SOV: [6] WHALE BALL-b a-SWIM-cl SPIT-OUT-cl-b.
Available meaning: 'The whale spit (something) on the ball.'
'Very very difficult to access': 'The whale spit out the ball.'
(LIS, 17-06-21 12.03 #3)
Video: https://www.dropbox.com/s/9a696ldkhq8orgp/MS%20whale%20spit%2017-06-21%2012.03%20%233.mov?dl=0

Strikingly, the target meaning is regained with an SOV order when the object is signed in a neutral locus, as seen in (57)b. SVO order yields the target meaning as well but is more degraded, as seen in (57)a.

(57) Classifier predicate (object in a neutral locus)
a. SVO: [5] WHALE a-SWIM-cl SPIT-OUT-cl-b BALL
(LIS, 17-06-21 12.04 #1)
Available meaning: 'The whale spit out the ball.'
'Impossible': 'The whale spit (something) at the ball.'
Video: https://www.dropbox.com/s/z7y2f8lw0w4izyo/MS%20whale%20spit%2017-06-21%2012.04%20%231.mov?dl=0

b. SOV: [6] WHALE BALL a-SWIM-cl SPIT-OUT-cl-b.
Meaning: 'The whale spit out the ball.'
(LIS, 17-06-21 12.03 #2)
Video: https://www.dropbox.com/s/ncn0arinhb5kkeo/MS%20whale%20spit%2017-06-21%2012.03%20%232.mov?dl=0

9.3 Interim conclusion

While a more detailed investigation should be conducted in the future, three findings are worth highlighting.
(i) First, and unlike what we found in ASL, a non-canonical word order is always somewhat dispreferred with our LIS classifier predicates: we might have expected SVO order to be perfect in spit-out-type examples, but in the cases we studied, it was degraded.
(ii) Second, this does not mean that classifier constructions can behave like normal verbs in our Deaf consultant's LIS: the target meaning 'x spit out y' is very hard to get with SOV order, which is precisely what we expect if temporal iconicity makes itself felt. (For our CODA consultant, it seems that the classifier construction has at least the option of behaving like a normal verb, and of yielding the target meaning.)
(iii) Third, when the object is signed in a neutral locus, SOV order *can* yield the target meaning. In other words, the effect of temporal iconicity seen in other cases is not just tied to the presence of the classifier predicate, but also to the use of non-neutral loci.

Why is SVO order degraded in spit-out-type constructions? One might posit that LIS just doesn't 'like' non-canonical word orders, but this begs the question of why ASL does (since its eat-up-type classifier predicates go against the basic SVO order of the language). Alternatively, one might notice that the combination of (i) and (ii) above (dispreferred SVO order for 'spit-out', combined with the absence of the target meaning for SOV order) is exactly what we found in our CODA consultant's ASL. Her remark was that a more articulated depiction would be needed to obtain the target meaning with SVO (our main ASL consultant made a similar remark). But so far, we have not managed to make the SVO order more acceptable in LIS by adding a further depiction. We leave this question for future research.

Why does SOV order yield the target meaning when the object appears in a neutral locus (= (iii) above)? This suggests that in this case the predicate classifier composes with its arguments like a normal verb rather than by iconic composition, since SOV is not the iconically meaningful order in this case. But this raises two questions. First, how can a classifier predicate be composed with its arguments in a non-iconic fashion? The question is even more pressing for our LIS CODA consultant, who even gets the target 'x spit out y' meaning with SOV order and an object in a *non*-neutral locus. To find an answer, we will need to explain why in some cases classifier predicates can behave like normal verbs – and why they can't always do so. Second, how is the choice between iconic and non-iconic composition effected? We will posit a principle of maximal iconicity according to which classifier predicates are composed iconically with their arguments if these appear in an iconically meaningful position (hence when they are properly localized), but not otherwise.[38]

10 The limits of temporal iconicity

In view of the previous section, one would like to say that under some conditions LIS has the *option* of treating a classifier predicate as a normal verb. There are two further cases that might justify such a refinement. First, in our survey, pro-speech gestures in French generally allowed for SVO order (on top of iconically-driven orders as well). While this might simply be because our consultants optionally treat gestures as codes for words, in which case they indirectly provide judgments about gesture-free sentences, an alternative is that they marginally allow the gestural verb to behave like a normal verb. Second, our main ASL consultant only assigns weak deviance to SVO orders in eat-up-type predicates,

[38] The LIS case with SOV yielding the target 'x spit out y' meaning when the object is in a neutral locus raises two questions about ASL. The first is about spit-out-type classifier predicates: Does SOV also yield the target 'x spit out y' meaning when the object of a classifier is signed in a neutral locus? Unless the object is moved (e.g. by topicalization), we expect that the answer is 'no' because the basic word order of ASL is SVO, not SOV (in our account of LIS, it is the possibility of treating the classifier predicate as a normal verb that is responsible for the availability of SOV in this case). The second question is about eat-up-type classifier predicates: Does SVO order improve when the object is signed in a neutral locus? By analogy with the LIS case, we would expect that the answer is 'yes': signing the object in a neutral locus should make it easier to obviate the need for an iconic composition between the classifier and its arguments. We leave these issues for future research.

which might suggest that he can treat the predicate as a normal verb as well, although this is not his preferred option.

While intuitive, this idea that classifier predicates can optionally behave like normal verbs does not follow from our current analysis.

10.1 Adding λ-abstraction

The reason a refinement is needed is that the semantics in (48), repeated in its essentials in (58), assigns a propositional meaning to the classifier predicate: it yields a truth value when evaluated relative to a context, assignment function, time and world. This entails that the classifier predicate cannot be composed with its arguments by standard composition rules such as function application: it has the wrong semantic type to do so.

(58) $[[<P_1, \ldots, P_n>_\tau]]^{c, s\, t, w}$ = 1 iff for some times $t_1, \ldots, t_n$ **with t = t_1**, $<P_1, \ldots, P_n>_\tau$ is true of $<t_1, \ldots, t_n>$ relative to $\pi_{c,t,w}$, w, s (according to (30)).

A natural solution to allow for composition is to make use of λ-abstraction. Since the rule in (48)/(58) fully integrates the iconic semantics of classifier predicates with standard intensional semantics, we can just apply the standard rule of λ-abstraction in (59), yielding (60). This can be applied to classifier predicates or gestural verbs alike (keeping in mind that the former must be interpreted with a mixed marking rule due to their conventional character).

(59) Standard interpretation of λ-abstraction
Let *v* be a variable and *F* an expression of any type (e.g. x),
$[[\lambda v\ F]]^{c, s, t, w} = \lambda x\ [[F]]^{c, s[v \rightarrow x], t, w}$

(60) Applying λ-abstraction to a three-picture sequence $<P_1[a, b], P_2[a, b], P_3[a, b]>_\tau$, abbreviated as $S_\tau[a, b]$ (see (35) and (46))
$[[\lambda b\ \lambda a\ S_\tau[a, b]]]^{c, s, t, w} = \lambda y\ [[\lambda a\ S_\tau[a, b]]]^{c, s[b \rightarrow y], t, w}$
$= \lambda y\ \lambda x\ [[S_\tau[a, b]]]^{c, s[b \rightarrow y][a \rightarrow x], t, w}$

After application of λ-abstraction, classifier predicates and gestural verbs have the semantic type of a transitive verb. Semantically, this allows them to compose with their arguments like any other transitive verb.[39]

λ-abstraction needn't be the only way to combine a classifier predicate with its arguments, however. The propositional meaning obtained for the classifier predicate in (46) is relativized to an assignment function s, and $S_\tau[a, b]$ contains the variables *a* and *b*. Thus another way of combining the predicate with its arguments would be by whatever means allows for variable binding or variable coindexation, as in the following case of double topicalization in French: the main clause is made of a transitive verb and two clitics, and their denotations are given by the left-most DPs which presumably bind them.

(61) Pierre, mes parents, il les vénère.
Pierre, my parents, he-clitic them venerates
'Pierre idolizes my parents.'

Importantly, in LIS, where we have our strongest argument that the classifier predicate can behave like a normal verb, our examples have none of the typical prosodic markers of topicalization (Conte et al 2010): eyebrow positioning, pause distribution, etc. all display unmarked settings. This suggests that, in these cases at least, λ-abstraction might be needed.

---

[39] Presuppositions will have be taken into account to develop a more adequate semantics. We conjecture that *a-SWIM-EAT-b* presupposes (rather than asserts) that *a* denotes an animal with a large mouth (because of the classifier shape), and that *b* denotes a large object (because of the size of the crocodile's open mouth). If so, the analysis would have to be developed within a trivalent account that can capture presupposition and presupposition projection. See also Schlenker 2019 and Tieu et al. 2019 for an argument that iconic representations (including gestures and classifier predicates) productively trigger presuppositions.

10.2 Iconic vs. non-iconic composition

If λ-abstraction can be freely applied to classifier predicates and gestural verbs, we predict that these should display all the syntactic and semantic possibilities of normal verbs (they might have further options as well due to the additional possibility of iconic composition). While this generalization might be correct for our CODA LIS consultant, for our other data this seems excessively liberal: in ASL, preverbal arguments are preferred in eat-up-type constructions, and the contrast with plain verbs was relatively clear in this respect. In LIS, for our Deaf consultant, SOV order failed to yield the target 'x spit out y' meaning when the object was localized. And a French pro-speech gesture paradigm displayed a related effect: preverbal objects were rather strongly preferred over SVO in (14), despite the likely pressure of the word order of French syntax.

We must thus restrict the application of λ-abstraction. A natural idea is that, *whenever this is possible,* iconic composition of a classifier predicates or gestural verb with its arguments (by way of the formation of a visual animation) is preferred over non-iconic composition (with λ-abstraction). In the case of LIS objects signed in a neutral locus, the object might not be in an appropriate iconic position, and this might block iconic composition and thus allow for λ-abstraction. Whether this rule favoring maximally iconic structures is on the right track remains to be seen.[40]

11 Intensional constructions revisited

One important question remains: why do intensional constructions tend to give rise to SVO orders in Schouwstra and de Swart's (2014) and Langus and Nespor's (2010) gestural data, as well as in Napoli et al.'s (2017) Libras data? The present account cannot offer a full answer because it is unclear how thoughts and spoken words should be analyzed in a projection-based iconic semantics.

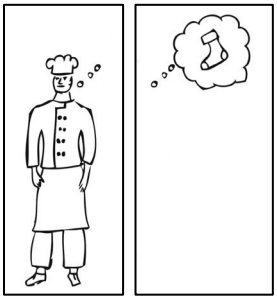
Still, there seem to be independent reasons to follow Schouwstra and de Swart 2014, who argue that "direct objects in intensional events" are "more abstract and more dependent on the action than those in extensional events". The independent motivation is that in narrative sequences that have nothing to do with gestures, the content of a thought or speech act seems to be preferably represented after the thought or speech act. The point can be made by starting from one of Schouwstra and de Swart's own stimuli, such as (62), and by artificially turning it into a mini-visual narrative made of two pictures, as in (63).

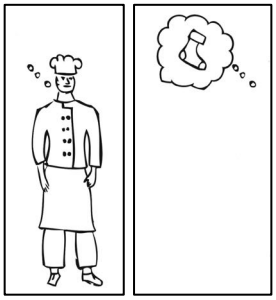
(62) Stimulus corresponding to a cook thinking of a sock (Schouwstra and de Swart 2014, Fig. 2)

(63) Transformation of (62) into a simple visual narrative made of 2 pictures
a. Thinking event before thought content
(i) **Original version** (ii) Mirror image version

[40] A rule of 'maximal iconicity' as in (i) below was posited in the analysis of Role Shift in Schlenker 2017. But it pertained to the interpretation of certain structures, not to the competition between iconic and non-iconic structures (unlike the competition rule we tentatively posit in this paragraph).

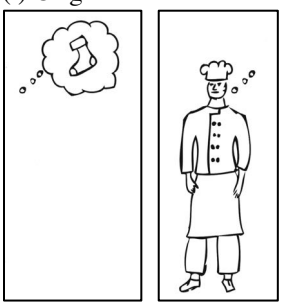
(i) Maximal Iconicity: In ASL and LSF Action Role Shift, expressions that can be interpreted iconically must be so interpreted. (Schlenker 2017)



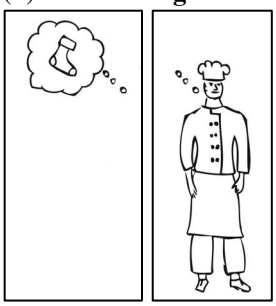


b. Thought content before thinking event

(i) Original version



(ii) **Mirror image version**



It seems to us more natural to represent the thinking event and then the thought, as in (63)a, rather than the other way around, as in (63)b. A complication comes from the fact that the bubble points towards the character in (63)a(i) but not in (63)b(i), and this may introduce an irrelevant bias. To address this problem, we can compare (63)a(i) to the mirror image version of (63)b(i) found in (63)b(ii): the bubble points towards the character in the latter picture sequence, but the order is still (we think) less natural than in (63)a(i).

The same transformation can be applied to Langus and Nespor's stimuli, turning (64) into (65).

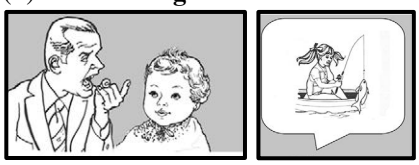
(64) Complex vignette used to elicit gestural embeddings (Langus and Nespor 2010, Appendix)

(65) Transformation of (62) into a simple visual narrative made of 2 pictures

a. Speech event before speech content

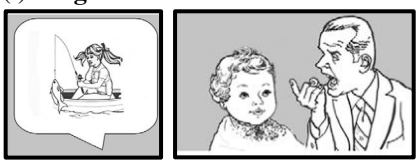
(i) Original version

(ii) **Mirror image version**



b. Speech content before speech event

(i) **Original version**



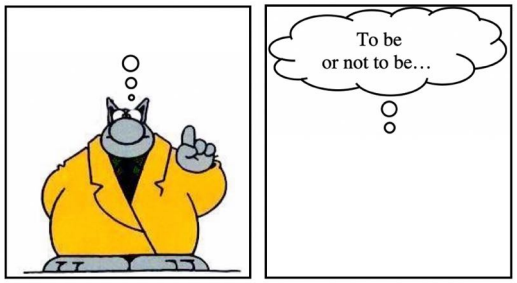
(ii) Mirror image version

To ensure that the bubble points towards the character, it is best to compare (64)a(ii) to (64)b(i). It seems to us that the former (with the speech event before the speech content) is more natural than the latter (opposite order).

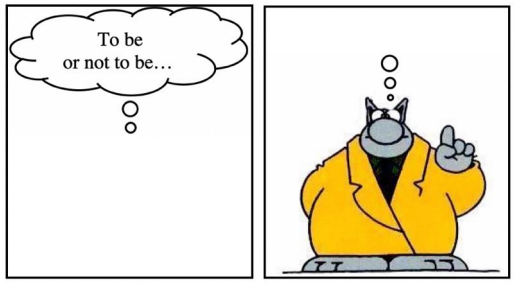
The point can be made more simply (without a directionality issue raised by pointing bubbles) by contrasting the order in (66)a (thinking event before thought content) with that in (66)b (thought content before the thinking event): both are a bit contrived (why would one separate the thought bubble from the character?), but the former less so than the latter, we believe.

(66) An existentialist cat[41]

a. Thinking event before thought content



b. Thought content before thinking event



It remains to explain, of course, why there is a preference for iconically representing thoughts as 'visible' after the corresponding thought or speech acts. This does not follow from a standard semantics for intensional verbs; for instance, in (67)a, roughly corresponding to the content of the picture in (64), the value of the embedded clause is standardly taken to be a proposition, i.e. a function from worlds to truth values, which is as intemporal as numbers are. The truth conditions just end up establishing (at a time t and in a world w) a relation between the subject denotation and this proposition, as illustrated in (67)b.

(67) a. The man tells the child that the girl is fishing.

b. $[\![(a)]\!]^{c,s,t,w} = \text{tell}'_{t,w}([\![\text{the man}]\!], \lambda w' [\![\text{the girl was fishing}]\!]^{c,s,t,w'})$

writing *tell'* for the semantic value of *tell*.

It is plausible, however, that we think of thoughts and spoken words as (metaphorically) visible only to the extent that they *represented* in some way. As a result, in visual narratives we preferably depict a thought after it is first formulated by salient individuals, as stated in (68), which is just another formulation of Schouwstra and de Swart's initial intuition.

[41] This is Geluck's Le Chat, here modified from: https://www.pinterest.fr/pin/402227810477055597/, accessed on April 20, 2021.

(68) Visual depiction of thoughts
In a visual narrative, a thought or spoken word should be depicted as being visible after and not before the time at which it is first formulated by an individual salient in the discourse.

There might well be more complex cases in which the prior representation of a thought or spoken word (by other people, or in written form) is salient enough that the thought can be depicted as existing before the action (this could for instance be the case for a thought or phrase which is repeated from individual to individual). As noted by E. Chemla (p.c.), we would predict that in such cases gestural sequences and ASL classifier predicates allow for pre-verbal intenstional objects. But since the scenarios and examples are complex to construct, we leave a test of this prediction for future research.

12 Conclusion

We have argued that in highly iconic constructions in gestures and in signs, some syntactic preferences ought to be derived from a semantics for visual narratives: an object preferably appears before or after the verb depending on the most natural position it would have in a visual animation. This followed intuitions developed in Schouwstra and de Swart 2014 for gestures and Napoli et al. 2017 for signs, but we departed from their generalizations as well as from their analysis: both groups of authors took the main distinction to be between intensional and extensional constructions, but we focused our argument on constructions that are uniformly extensional; and unlike these authors, we derived the generalization from a precise iconic semantics.

Our analysis showed how a visual narrative can be fully integrated with a standard compositional semantics. On a technical level, we developed a pictorial semantics (building on proposals by Greenberg and Abusch) in which visual narratives can be embedded in full-fledged sentences because they yield the same meaning types and can be linked to the rest of the sentence by shared discourse referents. If our analysis is on the right track, it is this extraordinary integration that is responsible for some word order effects in the cases under investigation.

Several extensions are left open. On an empirical level, both the sign language and gestural data could profitably be extended with experimental means, and a larger class of constructions (beyond eat-up and spit-out) should be investigated; Napoli et al.'s (2017) word-order effects in *non*-iconic constructions of Libras should also be revisited. On a theoretical level, our pictorial semantics should be made far more realistic, the principle of 'maximal iconicity' we tentatively posited to limit the scope of λ-abstraction should be further developed, and the analysis of intensional constructions should be further unified with the cases discussed in this paper.[42]

[42] Two further questions are left for future research.
(i) On the gestural side, we wrote as if the gestures we discussed are fully iconic, but this is probably a simplification, for two reasons: first, some gestures are conventionalized or semi-conventionalized; second, in our target sentences we took the precaution of first introducing the gestures as co-speech elements, thus clarifying their meaning and possibly giving them a near-conventional meaning for the purposes of the discourse. If so, the semantics of these gestures should be brought closer to that of classifier predicates.
(ii) On the sign side, we wrote as if the movement of a classifier is interpreted fully iconically, but it might be that things are more complicated and that some movements are 'unmarked' and do not convey as precise iconic information as others (the same issue might arise with gestures).

Appendix. From pseudo-continuous to continuous animations

In the main text, we only provided truth-of conditions for pseudo-continuous picture sequences, as in (30), copied in (69), not for genuinely continuous ones.

(69) **Truth-of relative to a viewpoint, a world and an assignment function for (pseudo-)continuous pictorial sequences**
Let π be a viewpoint, w a world, and s an assignment function. Then:
A pictorial sequence of the form $\langle P_1, \ldots, P_n \rangle_\tau$ (where $P_1, \ldots, P_n$ may contain variables) **with a constant interval τ between successive frames** is true of times $\langle t_1, \ldots, t_n \rangle$ relative to π, w, s iff **for some multiplicative parameter μ,**
(1) $t_1 < \ldots < t_n$ and **there is a constant interval $\mu\tau$ between successive times,** and
(2) relative to π, w and s, P_1 is true of t_1 and … and P_n is true of t_n.

A discrete pictorial sequence $\langle P_1, \ldots, P_n \rangle$ is in effect a function from an interval of integers (here: [[1, n]]) to pictures. A continuous sequence can be viewed as a function from an interval of real numbers (e.g. [t, t+9] for a 9-second snippet) to pictures.

Setting aside the issue of a multiplicative parameter, the definition in (69) can straightforwardly be extended to the continuous case as in (70).

(70) **Truth-of relative to a viewpoint, a world and an assignment function for continuous pictorial sequences - no multiplicative parameter**
Let π be a viewpoint, w a world, and s an assignment function. Then:
A visual animation viewed as a function A from a closed interval of real numbers [t, t+d] into a set of pictures P is true of a closed time interval of the same size [t', t'+d] relative to π, w, s iff relative to π, w and g, for every i in [0, d], A(t+i) is true of t'+i.

Adding a multiplicative parameter μ can be done as well, as in (71):

(71) **Truth-of relative to a viewpoint, a world and an assignment function for continuous pictorial sequences - with a multiplicative parameter**
Let π be a viewpoint, w a world, and s an assignment function. Then:
A visual animation viewed as a function A from a closed interval of real numbers [t, t+d] to a set of pictures P is true of a closed time interval [t', t'+μd] (with $\mu > 0$) relative to π, w, s iff relative to π, w and s, for every i in [0, d], A(t+i) is true of t'+μi.[43]

[43] Thanks to E. Chemla (p.c.) for providing feedback on this Appendix.

Supplementary Materials. Raw Data.

ASL raw data

https://drive.google.com/file/d/1PbHuL3neagPY-RoOGdW_NbGWGk_qVVUo/view?usp=sharing

Gesture survey

Survey: https://docs.google.com/forms/d/14_OQplxrropcTT05sdYVuBoHOMgqyOzZTOY9VQqvMiY/prefill

Pdf of the survey: https://drive.google.com/file/d/1tT00yv0GVPd2CS5ESZNMli6Kzdt2yiWm/view?usp=sharing

Survey results (June 23, 2021):[44] https://www.dropbox.com/s/gclimld0fy9ce1x/results_survey-21.06.23-MB-PS.xlsx?dl=0

[44] Numerical results were copied from the Google survey page presenting the overall results for the 10 participants and double-checked 3 times. Means are computed in the attached Excel document.

References

Abusch, Dorit: 2013, Applying discourse semantics and pragmatics to co-reference in picture sequences. In E. Chemla, V. Homer, G. Winterstein (eds), *Proceedings of Sinn und Bedeutung* 17: 9–25.

Abusch, Dorit: 2020, Possible worlds semantics for pictures. In L. Matthewson, C. Meier, H. Rullmann, and T. E. Zimmermann (eds.), *The Wiley Blackwell Companion to Semantics*, Wiley.

Abusch, Dorit and Rooth, Mats: 2017, The formal semantics of free perception in pictorial narratives. In Alexandre Cremers, Thom van Gessel & Floris Roelofsen (eds), *Proceedings of the 21st Amsterdam Colloquium*.

Cecchetto, Carlo; Geraci, Carlo; Zucchi, Sandro: 2006, Strategies of relativization in Italian Sign Language. *Natural Language & Linguistic Theory*, 24(4), 945-975.

Cogill-Koez, Dorothy: 2000, Signed Language Classifier Predicates: Linguistic Structures or Schematic Visual Representation? *Sign Language & Linguistics*, 3(2):153-207.

Conte, Genny; Santoro, Mirko; Geraci, Carlo; Cardinaletti, Anna: 2010. Why are you raising your eyebrows? In *LREC 2010 Conference proceedings*, 53–56.

Davidson, Kathryn: 2015, Quotation, Demonstration, and Iconicity. *Linguistics & Philosophy* 38(6), 477-520.

Emmorey, Karen and Falgier, Brenda: 2004, Conceptual Locations and Pronominal Reference in American Sign Language. *Journal of Psycholinguistic Research*, 33(4). 321-331.

Emmorey, Karen and Herzig, Melissa: 2003. Categorical versus Gradient Properties of Classifier Constructions in ASL. In K. Emmorey (Ed.), *Perspectives on Classifier Constructions in Signed Languages*, Lawrence Erlbaum Associates, Mahwah NJ, pp. 222–246

Forbes, Graham: 2020, Intensional Transitive Verbs. *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/intensional-trans-verbs/

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of crosslinguistic word order variation. *Psychological Science*. http://dx.doi.org/10.1177/0956797612463705.

Givón, Talmy: 1993, English Grammar: A Function–based Introduction. Vol. I. Amsterdam/ Philadelphia: John Benjamins

Greenberg, Gabriel: 2013. Beyond Resemblance. *Philosophical Review* 122:2, 2013

Greenberg, Gabriel: 2014, Reference and Predication in Pictorial Representation. Handout of a talk given at the London Aesthetics Forum (February 19, 2014).

Greenberg, Gabriel: 2019, Tagging: Semantics at the Iconic/Symbolic Interface. *Proceedings of the Amsterdam Colloquium 2019*.

Greenberg, Gabriel: 2021, Semantics of Pictorial Space. *Review of Philosophy and Psychology*. https://doi.org/10.1007/s13164-020-00513-6

Goldin-Meadow, Susan, So, Wing Chee, Özyürek, Asli and Mylander, Carolyn: 2008, The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences* 105(27): 9163-9168.

Haiman, John: 1983, Iconic and Economic Motivation. *Language* 59: 781–819.

Language, 81, 16–33.

Hall, Matthew L.; Mayberry, Rachel I.; Ferreira, Victor S.: 2013, Cognitive Constraints on Constituent Order: Evidence from Elicited Pantomime. *Cognition* 129, 1:1-17.

Hall, Matthew L.; Mayberry, Rachel I.; Ferreira, Victor S.: 2015, Production and Comprehension show Divergent Constituent Order Preferences: Evidence from Elicited Pantomime. *Journal of Memory and Language*, 81, 16–33.

Hendriks, B.: 2007, Negation in Jordanian Sign Language: A Cross-linguistic Perspective. In P. Perniss, R. Pfau, and M. Steinbach (Eds.), *Visible variation: Comparative Studies on Sign Language Structure*, pp. 103–128. Berlin: Mouton de Gruyter.

Kimmelman, Vadim: 2012, Word Order in Russian Sign Language. In *Linguistics in Amsterdam*, pp. 1–56.

Koraka, Marianthi: 2021, On word order in Greek Sign Language. Poster, FEAST (Formal and Experimental Advances in Sign language Theory), June 2, 2021.

Langus, Alan and Nespor, Marina: 2010, Cognitive systems struggling for word order. *Cognitive Psychology* 60(4), 291–318.
Li, Xijiang: 2016, The Ordering Distribution of Temporal Adverbial Clauses of Chinese EFL Learners and L1 Conceptual Transfer. *Modern Foreign Languages* (5): 682–692.
Liddell, S. K.: 1980, *American Sign Language Syntax*. Hague: Mouton Publishers.
Liddell, Scott: 2003, *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge Cambridge University Press.
Meir, I., Lifshitz, A., Ilkbasaran, D., & Padden, C. A. : 2010, The interaction of animacy and word order in human languages: A study of strategies in a novel communication task. Paper presented at the 8th international conference on the evolution of language, Utrecht, Germany.
Napoli, Donna Jo, Rachel Sutton-Spence & Ronice Müller de Quadros: 2017, Influence of predicate sense on word order in sign languages: Intensional and extensional verbs. *Language* 93(3). 641–670. DOI: https://doi.org/10.1353/lan.2017.0039
Oviedo, A.: 2003, *Apuntes para una Gramática de la Lengua de Señas Colombiana*. Cali: INSOR/Universidad del Valle.
Pavlič, Matic: 2016, The word order parameter in Slovenian Sign Language Transitive, ditransitive, classifier and locative constructions. PhD dissertation, University of Venice.
Sandler, Wendy and Lillo-Martin, Diane: 2006, *Sign Language and Linguistic Universals*. Cambridge University Press.
Schlenker, Philippe. 2013. Temporal and Modal Anaphora in Sign Language (ASL). *Natural Language and Linguistic Theory* 31(1). 207-234.
Schlenker, Philippe: 2017, Super Monsters II: Role Shift, Iconicity and Quotation in Sign Language. *Semantics & Pragmatics*. Volume 10, 2017.
Schlenker, Philippe: 2018a, Locative Shift. *Glossa: A Journal of General Linguistics*, 3(1), 115. DOI: http://doi.org/10.5334/gjgl.561
Schlenker, Philippe: 2018b, Sign Language Semantics: Problems and Prospects [replies to peer commentaries]. *Theoretical Linguistics* 44(3-4): 295–353.
Schlenker, Philippe: 2018c, Gesture Projection and Cosuppositions. *Linguistics & Philosophy* 41, 3:295–365.
Schlenker, Philippe: 2019, Gestural Semantics: Replicating the typology of linguistic inferences with pro- and post-speech gestures. *Natural Language & Linguistic Theory* 37, 2: 735–784.
Schlenker, Philippe: 2020, Gestural Grammar. *Natural Language & Linguistic Theory* 38:887–936
Schlenker, Philippe: 2021, Iconic Presuppositions. *Natural Language & Linguistic Theory* 39:215–289.
Schlenker, Philippe: to appear, Musical Meaning within Super Semantics. *Linguistics & Philosophy*
Schlenker, Philippe and Lamberton, Jonathan: 2019, Iconic Plurality. *Linguistics & Philosophy* 42: 45-108
Schlenker, Philippe and Lamberton, Jonathan: to appear, Meaningful Blurs: the Sources of Repetition-based Plurals in ASL. *Linguistics & Philosophy*
Schlenker, Philippe, Lamberton, Jonathan and Santoro, Mirko: 2013, Iconic Variables. *Linguistics & Philosophy* 36, 2, 91-149.
Schlenker, Philippe, and Chemla, Emmanuel: 2018, Gestural Agreement. *Natural Language & Linguistic Theory*. 36, 2: 87–625587. https://doi.org/10.1007/s11049-017-9378-8
Schouwstra, Marieke and de Swart, Henriëtte: 2014, The Semantic Origins of Word Order. *Cognition* 131(3), 431–6.
Schouwstra, Marieke; de Swart, Henriëtte; Thompson, Bill: 2019, Interpreting Silent Gesture: Cognitive Biases and Rational Inference in Emerging Language Systems. *Cognitive Science* 43, 7, e12732.
Sprouse, Jon and Almeida, Diogo: 2012, Assessing the reliability of textbook data in syntax: Adger's Core Syntax. *Journal of Linguistics* 48: 609-652.
Sprouse, Jon and Almeida, Diogo: 2013, The empirical status of data in syntax: A reply to Gibson and Fedorenko. *Language and Cognitive Processes*. 28: 222-228. [DOI: 10.1080/01690965.2012.703782]
Sprouse, Jon; Schütze, Carson T.; Almeida, Diogo: 2013, A comparison of informal and formal acceptability judgments using a random sample from Linguistic Inquiry 2001-2010. *Lingua* 134:

219-248. [DOI: 10.1016/j.lingua.2013.07.002]

Supalla, Ted: 1982, *Structure and Acquisition of Verbs of Motion and Location in American Sign Language*. Ph. D. thesis. University of California at San Diego.

Sutton-Spence, Rachel and Woll, Bencie: 1999, *The Linguistics of British Sign Language: An Introduction*. Cambridge: Cambridge University Press.

Sze, F. Y. B.: 2003, Word Order of Hong Kong Sign Language. In A. E. Baker, B. van den Bogaerde, and O. A. Crasborn (Eds.), *Cross-linguistic Perspectives in Sign Language Research*. Selected papers from TISLR 2000, Hamburg, pp. 163–192. Signum.

Tieu, Lyn;Pasternak, Robert; Schlenker, Philippe; Chemla, Emmanuel: 2017. Co-speech gesture projection: Evidence from truth-value judgment and picture selection tasks. *Glossa: A journal of general linguistics* 2(1).

Tieu, Lyn; Pasternak, Robert; Schlenker, Philippe; Chemla, Emmanuel: 2018, Co-speech gesture projection: Evidence from inferential judgments. *Glossa* 3(1), 109. DOI: http://doi.org/10.5334/gjgl.580

Tieu, Lyn; Schlenker, Philippe; Chemla, Emmanuel: 2019, Linguistic Inferences Without Words. *Proceedings of the National Academy of Sciences* 116 (20) 9796-9801.

Valli, Clayton, and Lucas, Ceil: 2000. *Linguistics of American Sign Language: an Introduction*. 3rd ed. Washington, DC: Gallaudet University Press.

Vermeerbergen, M.: 2004, The Quest for Basic Word Order in Flemish Sign Language. In A.-M. Bertonneau and G. Dal (Eds.), *La linguistique de la LSF: recherches actuelles*. Silexicales 4, pp. 257–267. Villeneuve d'Ascq: Université de Lille 3.

Zucchi, Sandro: 2011, Event descriptions and classifier predicates in sign languages. Presentation at FEAST (Formal and Experimental Advances in Sign language Theory) in Venice, June 21, 2011.

Zucchi, Sandro: 2012, Formal semantics of sign languages. *Language and Linguistics Compass* 6 (11): 719-734.

Zucchi, Sandro: 2017, Event Categorization in Sign Languages. In: Cohen H. and Lefebvre C. (eds.) *The Handbook of Categorization in Cognitive Science*, Elsevier. 2017, pp. 377-396