THEORIA ET HISTORIA SCIENTIARUM, VOL. XVIII
Ed. Nicolaus Copernicus University 2021

DOI: http://dx.doi.org/10.12775/ths.2021.004

Anu Vastenius
Division for Cognitive Semiotics, Centre for Language and Literature, Lund University, Sweden
anu.vastenius@semiotik.lu.se

Jordan Zlatev (corresponding author)
Division for Cognitive Semiotics, Centre for Language and Literature, Lund University, Sweden
jordan.zlatev@semiotik.lu.se
https://orcid.org/0000-0002-6096-4763

Joost van de Weijer
Humanities Lab, Lund University, Sweden
joost.van_de_weijer@humlab.lu.se
https://orcid.org/0000-0001-9843-3143

Cognitive and semiotic determinants of sign order in gestural and pictorial event representations

Abstract. The way people order signs in non-verbal event representations of events has been a topic of extensive research in recent decades, with conflicting findings. Based on a literature review, we distinguish the following factors that have been argued to influence sign order: (a) diagrammatic iconicity, (b) manipulation vs. construction events, (c) reversibility, (d) most common word order of L1, (e) the semiotic system (gesture, depiction) used, and (f) a putative "natural" Agent–Patient–Act order. To test the role of these factors, we conducted a study where Swedish participants observed events that varied with respect to reversibility and construction/manipulation status, and then had to communicate them to an addressee using gestures for half the stimuli. For the other half, they used sequences of simple drawings of the event participants and the action. The results showed the huge influence of the semiotic system (e) on sign order. There was a role of reversibility (c) only when gestures were used, while L1 word order (d) only had an effect when sequences of pictures were used. The "ontological status" of the Patient (b) was shown to affect the representations in both

semiotic systems but was much stronger for gestures. Even the two most general factors (a) and (f) where shown to be sensitive to the type of semiotic system. Future studies of how such cognitive and semiotic determinants interact are needed to fully understand the phenomenon of sign order preference.

Keywords: cognitive semiotics; basic word order; diagrammatic iconicity; gestures; manipulation/construction events; non-verbal event representations; pictures; pantomime; reversibility; semiotic systems.

1. Introduction

Everyone who has played charades has probably experienced the dilemma of how to gesture a complex event so as to help the team members understand the message. In what order should its parts be expressed? Is it better to start by miming the "Agent" (the one who instigates the action), the "Patient" (the one affected), or perhaps the act itself? Or does it, for some reason, feel more "natural" to place the act always at the end? And what if the game is played with pictures instead of gestures? From a semiotic perspective, words, gestures and pictures are examples of different kinds of *signs*, consisting (minimally) of expression and denoted intentional object (an actual or imagined person, thing, property, event etc.).[1] Experimental research which effectively concerns *sign order preferences* during the past decade has shown that there is no single answer to such questions, as the preferred order of gestures and other non-verbal signs in event representations depends on a number of factors. We refer to these factors as *cognitive and semiotic determinants of sign order*, given that they concern various aspects of human cognition and sign use, including iconicity and the type of *semiotic system* (e.g., language, gesture, depiction) that is being used (Zlatev, 2019). The nature and the relative importance of these determinants is the topic of this article.

In an influential study, Goldin-Meadow, Mylander, So and Özyürek (2008) argued that irrespective of the preferred word order of their native language, participants tend to use the order Agent–Patient–Act, in both gestural and pictorial representations. The authors speculated that this corresponds to

[1] A formal definition is, for example, the following: "A sign <E, O> is used (produced or understood) by a subject S, if and only if: (a) S is made aware of an intentional object O by means of expression E, which can be perceived by the senses. (b) S is (or at least can be) aware of (a)" (Zlatev, Żywiczyński & Wacewicz, 2020, p. 160).

a general cognitive bias for placing the representation of the act last in the representation of the whole event. Additional support for this proposal can be found in emerging signed languages in deaf communities (Baker, van den Bogaerde, Pfau & Schermer, 2016; Sandler, 2012), and in so-called *home sign* systems, which develop among deaf children who are not exposed to an established signed language (Coppola & Newport, 2005).

In further research, however, other factors were found to influence the order of signs in event representation. One such determinant is if the action represented is "reversible" – which is the case when both the Agent and Patient are human (e.g., *Man–Kiss–Woman*) – or not (e.g., *Man–Kick–Ball*).[2] Rather than displaying the supposed "natural" order of Goldin-Meadow et al. (2008), participants showed a tendency to prefer Agent–Act–Patient order in such cases (Langus & Nespor, 2010; Meir et al., 2010; Hall et al., 2013). Another determinant was shown to be the "ontological status" of the Patient (Schouwstra, 2012; Christensen & Tylen, 2013). If, for example, the Patient was created by the act itself (e.g., *Painter–Draw–Picture*), participants again preferred Agent–Act–Patient order, while if the Patient was a physical object that existed even prior to the act, then the order Agent–Patient–Act was indeed more common.

But how such factors interact with one another and with the presumed "natural order" is something that remains to be explored. In addition, what do these factors themselves depend on? Underlying explanations that have been proposed include *iconicity* (Christensen et al., 2015), the risk of *confusing the agent* if both referential expressions are on the same side of the action (Gibson et al., 2013), and *role-conflict in production*, as producing the Act following the Patient can easily make the latter appear "agent-like" (Hall, Mayberry & Ferreira, 2013). Below follow brief explanations of these three factors.

Iconicity is, somewhat simplistically defined, the *resemblance between the expression and the content* of a message, which may consist of one or more signs (Jakobson, 1965). Such resemblance can either be more concrete, as in onomatopoeia (e.g., *Bam!* resembles a loud noise) or more abstract: plural nouns are in almost all languages "longer" than singular nouns, resembling the fact that many things are more than one. The first of these is called *imagistic*, and the second *diagrammatic* iconicity (Devylder, 2018). To help explain why construction events were represented differently from manipulation events,

[2] Here and bellow we use capitalized English terms without grammatical morphology to denote the non-linguistic constituents of events, rather than (English) sentences.

Christensen et al. (2015) appeal to diagrammatic iconicity (which they refer to as "structural iconicity"), arguing that there is an abstract form or resemblance between manipulation events and Agent–Patient–Act order, on the one hand, and between construction events and Agent–Act–Patient order, on the other. In a later paper, this is spelled out in more detail:

> Agent and patient roles must be physically co-present before the action can be purposefully performed. In other words, in an object manipulation event, the patient logically precedes the action being executed: obviously, one cannot manipulate or act upon an object, which is not physically or symbolically already present. By contrast, in a different type of transitive event, which we henceforth call object construction events, agents perform actions that cause objects to come into existence. […] In these cases, actions precede objects that, in turn, are dependent on the performed actions (Christensen et al., 2016, p. 70).

This kind of explanation may help with respect to "ontological status", but it does not address the factor of reversibility. At least two different explanations have been suggested with respect to this. One is based on the notion of a *noisy-channel*, which states that when Agent and Patient can potentially be confused, then it follows from information-theoretic considerations that it is more efficient to place them on "different sides" of the Act representation (Gibson et al., 2013). It should be noted that this proposal is neutral with respect to animacy: if the Agent and Patient roles are taken by an animate, often human, entity or not.

In conflict with this assumption, in the study with gestures by Hall et al. (2013) Agent–Act–Patient order was obtained more often when both the Agent and Patient were animate compared to when the Patient was inanimate. The two linked hypotheses centred on animacy and reversibility were contrasted by Kocab et al. (2018), who opposed examples like *Truck–Hit–Car* (reversible, not animate) and *Boy–Hit–Girl* (reversible, animate). This implies that the noisy-channel alone, without animacy, is not determinative for Agent–Act–Patient preference. In another recent study, the outcome and conclusion were essentially the same (Meir et al., 2017).

Still, the explanations suggested for why animacy is contributing to Agent–Act–Patient sign order are not fully clear. To our knowledge, the most convincing proposal is one by Hall et al. (2013): that the order of Agent–Act–Patient in reversible events may be preferred in the case of *bodily* representation of events, in both gesture and signed language, due to a strong tendency to represent

the Agent through the body of the gesturer, what they call the *body-as-agent phenomenon*:

> Our results from elicited pantomime illuminate the source of this pattern by demonstrating that a very similar phenomenon (body-as-agent) appears when naïve participants describe events in pantomime. That is, when they produce action gestures, *participants' own bodies take on the most prominent role in the event: the human agent*. This suggests that the roots of the body-as-subject pattern lie in aspects of cognition that are common across deaf and hearing individuals, both signers and non-signers, namely the body as the origin of action (ibid., p. 15).

This factor would favour either Agent–Act–Patient, or Agent–Patient–Agent–Act event representations, but notably, this would apply only to *gestures and signed languages*, and not carry over to events represented in other semiotic systems such as depiction.

This leads us to a determinant of sign order that has gone very much under the radar in the literature: inherent properties of the *semiotic system* used to represent the events themselves, with language, gesture and depiction constituting such independent, though interacting, semiotic systems (Stampoulidis, Bolognesi & Zlatev, 2019; Zlatev, 2019). As mentioned, Goldin-Meadow et al. (2008) claimed that using gestures or pictures did not affect the preferred sign order, and that neither was influenced by the basic word order of the languages spoken. But in the only study that systematically used sequences of pictures to represent events, there was an effect of the first language (L1) of the participants (Vastenius, van de Weijer & Zlatev, 2016). Speakers of Kurdish, which has Agent–Patient–Act as most common word order, had a tendency to place the Patient before the Action more often when using a sequence of pictures to describe events than speakers of Swedish, which prefers the Agent–Act–Patient word order.

In fact, there are good reasons to suppose that event representations in gesture and depiction would differ. Gestures use the body as means of expression and display (relatively) *rapid fading*: the representation disappears after performing it (Hockett, 1960). In contrast, pictures subsist once created, and when placed in a given conventional order, as in cartoons, resemble written language to a considerable extent (Vastenius et al., 2016). Thus, a stronger effect of language, especially among literate participants, on sign order in pictures rather than gesture can be expected.

To sum up, the following cognitive and semiotic determinants of sign order require further investigation: (a) diagrammatic (structural) iconicity, (b) manipulation vs. construction events, (c) reversibility, (d) most common word order of L1 and (e) the type of semiotic system – in addition to (f) the original proposal of a cognitive bias for an Agent–Patient–Act order. Based on the discussion in this section, we would expect (a), (b) and (f) to be *general*, in the sense of applying irrespective of L1 and semiotic system. On the other hand, if the *body-as-agent* explanation is correct, (c) would apply above all to gesture, while L1 word-order would apply above all to event representations in terms of sequences of pictures.

To address these questions and general predictions, we conducted a study where Swedish participants observed video-clips of events that varied with respect to reversibility and construction/manipulation status, and then had to communicate them to an addressee using gestures (pantomime) for half the stimuli. For the other half, they used simple drawings of the event participants and the action. We explain the methodology and our predictions in Section 2. The results are presented in Section 3, and we discuss the study's implications for understanding sign order determinants in Section 4. We conclude in Section 5.

2. Method

2.1. Elicitation of stimuli and materials

We recorded stimuli in the form of 32 short video clips (plus 6 training clips), where several key factors were manipulated (see Table 1). Colleagues and family friends participated as "actors". 16 events were *reversible*, with both Agent and Patient, or Goal (see below) being human, and 16 events were *non-reversible* so that Patient or Goal was an artefact or an animal.

Orthogonal to this, 16 events were simple *transitive* events, consisting of only Agent, Patient and Action, while the other 16 consisted of *translocative* motion events (Naidu et al., 2018; Zlatev, Blomberg & David, 2010), 8 of which were self-motion events, where the Agent moves to the Goal, and 8 where caused-motion events, where the Agent moves an object to a Goal (or in two cases, from a Source). Finally, and crucially, the 16 simple transitive events were balanced so that 8 were construction events, and 8 were manipulation events, crossed with the reversible/non-reversible dimension. The events were shown to participants on the screen of a portable computer, with size

19 × 34.5 cm, with resolution 1920 × 1080, in what was presented as a "guessing game", see below.

Table 1. The 32 events presented as video-stimuli, divided by type (transitive/translocative), reversibility, and "ontological status" (manipulation/construction)

	Simple transitive events		**Translocative motion events**	
	Manipulation events	**Constructions events**	**Self-motion events (Agent + Goal)**	**Caused motion events (Agent + Patient + Goal)**
Reversible	*01. Boy–Punch–Man*	*09. Girl–Create–Magician*	*17. Boy–Walk–To–Man*	*25. Boy–Throw–Orange–To–Girl*
	02. Girl–Hug–Woman	*10. Magician–Create–Girl*	*18. Girl–Run–To–Woman*	*26. Man–Push–Chair–To–Woman*
	03. Man–Lift–Girl	*11. Girl–Create–Boy (with a magic stick)*	*19. Man–Run–To–Boy*	*27. Man–Throw–Bottle–To–Boy*
	04. Woman–Push–Boy	*12. Boy–Create–Girl (with a magic stick)*	*20. Woman–Walk–To–Girl*	*28. Woman–Give–Apple–To–Man*
Non-reversible	*05. Girl–Eat–up–Chocolate*	*13. Boy–Draw–Face (on paper)*	*21. Boy–Run–To–Bicycle*	*29. Boy–Put–Orange–Into–Basket*
	06. Man–Break–Bottle	*14. Boy–Make–Paperplane*	*22. Girl–Run–Into–House*	*30. Boy–Throw–Bottle–Into–Bucket*
	07. Woman–Petting–Dog[3]	*15. Girl–Making–Cake*	*23. Man–Walk–From–Car*	*31. Girl–Place–Glass–On–Table*
	08. Boy–Bouncing–Ball	*16. Girl–Paint–Flower*	*24. Woman–Walk–Out of–Hut*	*32. Man–Drop–Mobile–On–Sofa*

Six stimulus videos, representing different event types were presented to allow the participants to practice, both before the gesture and the picture task, and they were not included in the data analysis. They were constructed so that

[3] In these verbal descriptions of the stimuli, the *present participle* is used in the four cases where the action was not completed, but rather ongoing during the entire clip.

the same Agent, Act or Patient was repeated in several stimuli to demonstrate to the participants that it was essential to show the details in the clips to avoid confusion, including videos of *Boy–Build–House (of blocks)*, *Girl–Build–House (of blocks)*, *Boy–Make–Snake (of clay)* etc.

The pictures to be used when representing events with a picture-sequence were sketches drawn by a hobby artist (see Figures 1–3, and Appendix A for all pictures). The drawings were schematic, with some three-dimensionality to increase comprehensibility, and all objects were displayed from the front or slightly from the side-front. Four pictures were schematic representations of people, which could appear in either of the roles of Agent, Patient and Goal As shown in Figure 1, they were depicted from the neck upwards. Images of the non-human Patients and Goals, such as Car, Apple, and Chocolate were drawn whole (Figure 2).

Figure 1. Picture cards for human beings: Woman and Man

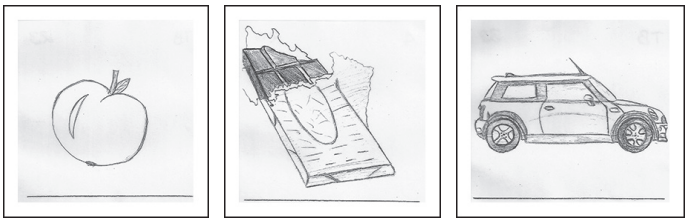
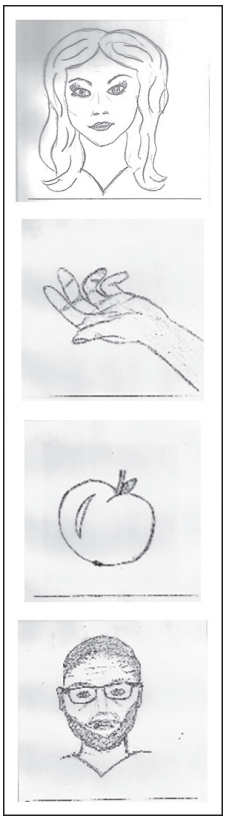


Figure 2. Examples of picture cards for non-human Patients and Goals: Apple, Chocolate, Car

In earlier studies, Acts were represented by abstract arrows (Goldin-Meadow et al., 2008; Vastenius et al., 2016). For this study more concrete depictions were created in order to make them more imagistic than diagrammatic (see Section 1), and more similar to the other kinds of pictures. Thus, as shown in Figure 3, Acts were displayed *metonymically* (i.e., using a part-for-whole strategy) showing hands/arms, feet/legs or a mouth (to represent Eating). Some depictions included *pictorial runes* (Forceville, 2011), which were used

Figure 3. Examples of picture cards for acts: Push, Put, Run, Paint and Hug



to represent motion, and when necessary, small arrows completed the act pictures to indicate the direction of the intended movement.

A line was drawn at the bottom of each picture, to avoid possible ambiguity with respect to the way it is to be viewed. The pictures were printed on 12 × 12 cm sized cards and plasticized. When representing events through pictures, participants had to place them in a top-to-bottom row, as shown in Figure 4. To facilitate them in this, a 48 cm × 12 cm size "coaster" was provided.

In addition, 32 photos of the 32 video-clips were made, printed out, and used by the confederate for "guessing" the event that was being represented, as explained in the following sub-section.

Figure 4. An example of picture cards arranged on the coaster to display *Woman–Give–Apple–To–Man*

2.2. Participants and procedure

Sixteen adults (11 female), with mean age 23.6, were recruited for the experiment, which was conducted at the Humanities Laboratory of Lund University during the fall semester of 2019. All participants were native speakers of Swedish with no knowledge of languages with most common word order other than Agent–Act–Patient, which was controlled by asking them prior to recruitment. Each participant was tested individually after reading and signing an informed consent form and compensated with a cinema ticket upon completion of the experiment.

Apart from the main researcher, there was also a male "confederate" present in the room. The confederate was introduced as another participant, and the task was presented as a study in non-verbal communication: no speech was to be exchanged between the participant and the confederate. By playing the role of an addressee who had to "guess" the event that was being represented by the participant by picking out one of the 32 photos on his table, the confederate helped increase the motivation of the participants to perform the task with more precision.

The experiment started with a training session, where six clips were displayed on the computer monitor, and the participant was to represent three of these through gesture, and three with pictures, as shown in Figure 4. The confederate had six screenshots of the training clips printed on A4 sheet of paper and had to pick the one he thought the participant meant, and then show it to them. The participant then replied whether it was the right picture and was given the chance to repeat the non-verbal presentation until the confederate guessed right. The confederate was instructed to show (only during the training session) hesitation if the gestural or pictorial representation was too vague (for example, only consisting of an Act) so as to guide the participant to add details in the performance. After participants had the chance to ask questions, the training session concluded. The confederate was instructed to behave naturally in observing the gestural and pictorial representations of the participants and to try to choose the "correct" photo, but without worrying too much if he got this right or not. Importantly, he was not to behave differently in any of the conditions.

For the experiment proper, half of the participants represented the first 16 events with gestures and the second 16 events with pictures, and the other half of the participants did this in the reverse order. All video-clips were shown one by one in random order. For the gesture condition, the participants were given the following instructions (here translated from Swedish):

> *You are going to see some video-clips. Your task is to first watch them and after each clip to "tell", without talking, only with gestures, to the other participant what happened in it. He will then need to find a photo on his table that illustrates the clip. Do not use surrounding items or anything that you are wearing, only your body movements. The videos can be similar in their content, so be careful with your gestures.*

In this phase, the confederate did not give any feedback if he had "guessed" correctly. As shown in the instructions, participants were told not to use any surrounding items or anything that they were wearing, only body movements. They were allowed to use as many gestures as they wished.

When representing events with pictures, the procedure was the same, except that the pictures printed on cards were used instead of gestures. The decision to use pre-made cards rather than to ask participants to draw themselves was provoked by the need to ensure a degree of homogeneity of the representations, as drawing skills differ enormously across people. While there is also variation in how well people can represent events using gestures, this is much less pronounced, since pantomime is (arguably) our original, and more or less universal, communicative system (Zlatev et al., 2020).

The picture cards were given mixed and at once to the participant, who was instructed to spread them on the table in an optional order, but making sure that the "bottom line" was placed on the side close to them. The participant was given the following instructions (again, translated from Swedish):

> *You will get a bunch of pictures. Look through them and spread them with the bottom line downwards on the table in an optional order so that you can easily find each picture. You will see video-clips, your task is to first look at each clip, and then to use the pictures to "tell" without words what happened in the film to the other participant. He will then try to find the photo on his table that depicts the film among the alternatives in front of him. Place the pictures you need on the coaster from top to down to express the contents. The films can be similar so be careful with your choices.*

Each time after creating a representation, the pictures were returned to the pool of pictures on the table. The maximum number of pictures for one representation was four. All experiments were video-recorded with Sony HDR-CX360VE Digital HD video recorder and Panasonic 4K HC-VX980.

After both tasks were completed, there was a debriefing about how the participant had experienced the task in the form of a free interview, focusing on how the participants thought about the order in which they represented the different constituents of the events. We report some findings from this in Section 4, but importantly, no participant guessed that the study concerned the order of the signs, when asked what they thought its purpose was.

2.3. Data analysis

2.3.1. Coding

The video-recordings were analyzed with the help of ELAN 5.2.[4] For analyzing the gestural representations, the following five tiers were used (see Figure 5).

On the first tier (*Stimuli*), the representation of each event was separated, and the number and the name of the stimulus were marked. On the second tier (*Meaning*), the gestures used for each stimulus were separated and the contents of each gesture were annotated using simple English expressions, such as "beard", "run", "house", "short", "long hair". The list was not fixed in

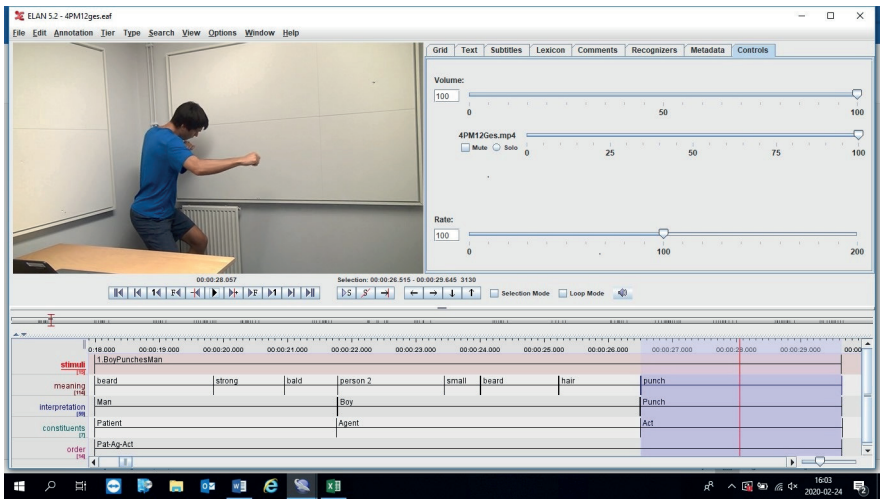


Figure 5. A screen-shot of the ELAN template used for the analysis of the participants' gestures

[4] ELAN (Version 5.2) [Computer software]. (2019). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan.

advance, and the choice of descriptors was made freely by the single coder (the first author of the paper). It was usual for the participants to present several elements in a single gesture, for example, "holding something round in hand and taking bites of it and chewing", and in that case, it would be annotated as "round object + eat". The participants sometimes took a certain position to express which person they meant, and this was coded as "person 1" or "person 2". On the third tier (*Interpretation*), the gestural meanings were interpreted to correspond to the actual constituents in the events, so that for instance "beard" would be coded as "Man", and "short" and "long hair" together would be coded as "Girl", on the basis of clear metonymic (part-whole) relation. On the fourth tier (*Constituents*), the interpretations were converted to the semantic roles of an event, by being designated as Agent, Act, Patient, Recipient or Landmark. For the subsequent statistical analyses, Recipient and Landmark were considered as Goal. Sometimes the participants showed one constituent several times (possibly in turns with other constituents), and in those cases, they were designated as many times as they appeared. No second coding was made, as the procedure was quite straightforward.

The coding of the pictorial event representations was even easier. The only tiers were *Stimuli* (the same as for gestures), *Images* (e.g., Woman, Apple, Give, Man) and *Constituents* (Agent, Patient, Act, Goal), since each picture card already had a predetermined role in an event. For instance, an apple would always be a Patient, and a woman in the particular stimulus clip would always be an Agent. Additionally, the number of pictures to be used for each stimulus event was restricted to four, which prevented the use of multiple cards.

Two patterns could be detected in the chains with multiple same constituents: (1) *Repeating*: the participant pantomimed a chain more than once, for instance Agent–Patient–Act–(pause)–Patient–Act. The chain was treated as Agent–Patient–Act, and (2) *Scene-setting*: a number of participants had a strategy of first representing what or who was present in the performance. They pantomimed Agent and Patient/Goal first, often with an eye contact with the confederate, and then went on to the actual action by repeating all constituents once more and adding the Act. For instance, Agent–Patient–(pause)–Agent–Act–Patient was treated as identical with Agent–Act–Patient.

2.3.2. Statistical analysis

The results were analyzed as mixed effects logistic regression models, with *participants* and *items* as random effects, and *semiotic system* (gestures or pictures), *event type* (construction or manipulation) and reversibility as fixed

effects. The dependent variable was *sign order* (Patient > Act or Act > Patient). All analyses were performed in the software R,[5] using the package lme4 (Bates, Maechler, Bolker & Walker, 2015). In the result section below, we provide the estimate of the effect (EST), the standard error of the estimate (SE), a *z*-value, and a *p*-value.

2.4. Hypotheses

On the basis of the theoretical background presented in the first section, and the operationalization provided by our methodology, we could formulate the following hypotheses:

- **H1.** The basic word order of Swedish (Act > Patient) will have a stronger effect *in the Pictures than the Gestures* condition, as the first is more similar to (written) language.
- **H2.** There will be a stronger tendency for Act > Patient order *in construction than in manipulation events*, irrespective of the semiotic system used (as the motivation behind this is general diagrammatic iconicity).
- **H3.** There will be an independent Act > Patient order effect in reversible *than non-reversible* events, but *only for the Gestures condition* (as the motivation for this is the body-as-agent phenomenon).
- **H4.** The Goal will be placed last in most cases, irrespective of the semiotic system (as the motivation behind this is general diagrammatic iconicity). Note: events 23 and 24 are here excluded, as they do not contain a Goal but a Source.

3. Results

As predicted in H1, the effect of L1 word order on the pictorial representations was strong. As shown in Table 2, when using Gestures in 94 (67.1%) of the cases the participants preferred the presumed "universal order", suggested by Golden-Meadow et al. (2008), and chose to present the Act last. However, in the Pictures condition, they placed the Patient last in the vast majority of the cases (87.9%), as in the dominant word order of their L1, Swedish. The effect

[5] R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

of semiotic system on constituent order was significant (EST = 3.970, SE = 0.492, z = 8.064, p = 0.000).

Table 2. The order of Patient and Act in the two semiotic systems

	Gestures	Pictures
Patient > Act	94 (67.1%)	21 (12.1%)
Act > Patient	46 (32.9%)	153 (87.9%)

Concerning H2, the effect of the "ontological status" of the Patient (manipulation vs. construction event) was clearly reflected in the Gesture condition, with Act > Patient order for construction events in 29 (76.3%) cases, and conversely, Patent > Act order preferred in manipulation events in 36 (76.6%) case. For the semiotic system of Pictures, there did not seem to be such a strong effect, as Act > Patient order was preferred for both types of events. However, this bias was stronger for construction events, as expected.

Table 3. Constituent order in manipulation and construction events in the two semiotic systems

	Gestures		Pictures	
	Manipulation event	Construction event	Manipulation event	Construction event
Patient > Act	36 (76.6%)	9 (23.7%)	6 (10.2%)	1 (1.8%)
Act > Patient	11 (23.5%)	29 (76.3%)	53 (89.8%)	55 (98.2%)

The interaction between semiotic system and event type was not significant (EST =-1.409, SE = 1.389, z =-1.015, p = 0.3102), but the main effects of both semiotic system (EST = 4.642, SE = 0.858, z = 5.410, p = 0.000), and event type were significant (EST = 3.412, SE = 0.902, z = 3.784, p = 0.000). In other words, the construction vs. manipulation factor was significant in general, but more so for Gestures than for Pictures, implying that H2 was supported.

H3 concerning the effect of reversibility of the events for the semiotic systems Gesture and Pictures was also confirmed. For Pictures, there was a minor difference between the reversible and the non-reversible events. But this was in the opposite direction to what could have been expected if the effect was

independent of semiotic system, with *less* Act > Patient order in reversible than in non-reversible events. On the other hand, for Gesture there was a clear effect of the reversibility: in comparison to non-reversible events, where the Act > Patient order appeared in 10 (26.3%) of cases, in reversible events, this order dominated with 30 cases (63.8%).

Table 4. The order of Patient and Act in reversible vs. non-reversible events

	Gestures		Pictures	
	Reversible event	Non-reversible event	Reversible event	Non-reversible event
Patient > Act	17 (36.2%)	28 (73.7%)	5 (8.8%)	2 (3.4%)
Act > Patient	30 (63.8%)	10 (26.3%)	52 (91.2%)	56 (96.6%)

The observed pattern was confirmed by the statistical analysis. Within Gesture, there was a significant difference between constituent order in reversible and non-reversible items (EST = 3.514, SE = 1.477, z = 2.379, p = 0.017), and within non-reversible items there was a difference between constituent order in Gestures and Pictures (EST = 7.958, SE = 1.650, z = 4.822, p = 0.000). Further, the difference between reversible and non-reversible events was significantly reduced in the Pictures condition, as indicated by the estimate of the interaction term in the model (EST-4.887, SE = 1.561, z =-3.131, p = 0.002).

On the other hand, the prediction of H4 of the placement of Goal was only partly supported. We expected that when represented at all, it would appear last in the sequence, due to diagrammatic iconicity, irrespectively of semiotic system. As can be seen in Table 5, when Goal indeed occurred (193 times in all), it was represented last in the Pictures condition in 88 cases (84.6%). However, it was almost as likely to be placed last as not last when using the semiotic system of Gestures. The statistical analysis showed that the difference in Goal position between Gestures and Pictures was significant (EST = 2.263, SE = 0.361, z = 6.268, p = 0.000).

Table 5. The placement of Goal in gestural and pictorial representations

Semiotic system	Gestures	Pictures
Goal not last	46 (51.7%)	16 (15.4%)
Goal last	43 (48.3%)	88 (84.6%)

4. Discussion

As shown in the previous section, the four hypotheses formulated in Section 2.4 were for the most part strongly supported. Concerning the first hypothesis (H1), the effect of L1 was found to be significant when using sequences of pictorial representations, and this was expected due to the similarity between using a set of pre-given pictures and the sequential order of language signs, and in particular in writing (Vastenius et al., 2016). This interpretation was confirmed in the debriefing session as most of the participants mentioned that they used the same order for pictures as "in language", placing the "verb" before the "object" (it was characteristic that they even used such grammatical terms). Some participants said that they silently articulated the corresponding sentence describing the event first and then placed the cards, while others stated that it felt "logical" to use the order as it is in (their) language. In contrast, when using gestures, the order was most often (but still in only in ca. 2/3 of the cases) Patient > Act, as predicted by a "natural order" analysis along the lines of Goldin-Meadow et al. (2008).

Turning to the third hypothesis (H3), it was also found to be supported and a significant difference between the two semiotic systems was found. In the Gesture condition, in line with the *body-as-agent phenomenon* (Hall et al. 2013), Act > Patient order was found to be dominant in reversible events, but not in non-reversible events. As observed in many previous studies where the participants use their own bodies to represent Agents and Acts, it is preferable to place these in contiguity with one another, since if the Act were placed after the Patient, for example in a representation of *Boy–Girl–Kiss*, the communicator will likely "enact" the girl with their whole body, and suggest the interpretation that it is the girl rather than the boy who is doing the kissing. Significantly, in the Picture condition, there was only a slight difference between the two event types, and furthermore in the other direction, with more Patient > Act order for reversible events. This strongly suggests that it is not the "noisy channel" or even only animacy but precisely the body-as-agent effect that explains the asymmetry between reversible and non-reversible event representations, and that this is specific for the semiotic system of gesture, as well as for signed languages.

Returning to the second hypothesis (H2), concerning the distinction between manipulation and construction events, and thus the "ontological status" of the Patient, the effect was found in both semiotic systems, gestures and pictures, even though it was much stronger in the case of gestures. As

expected, this could be attributed to general diagrammatic iconicity, given that in construction events the Patient does not yet exist prior to the Act. Also in the debriefing sessions many participants explained that they presented the events in a "chronological order" or even more clearly: "in the order the constituents appeared in the clips", as when in a creation-event the Patient appeared last. The much lower magnitude of the effect in the pictures condition can be possibly explained by the role of the strong Act > Pat bias due to the role of L1 word order, as shown by the results concerning H1.

Finally, the fourth hypothesis (H4) was the one that was least supported, as it was predicted due to diagrammatic iconicity that the representation of the Goal would be given last in both semiotic systems, but this was the case only for Pictures. In gestural representations, there was no clear preference of the Goal position. Thus, a more likely explanation of the final placement of Goal when the semiotic system of pictures was used was not diagrammatic iconicity, but rather the strong role of L1 word order, where the "indirect object" is most often placed sentence-final in Swedish.

This interpretation was also supported by the debriefing, as indicated above. Unlike in the case of the Pictures condition, where the main motivation was often said to be "the order in language", when events were represented through gesture, most participants found it difficult to motivate the sign order that they used. Some mentioned "chronological order", possibly referring to the order in which the constituents appeared (e.g., Patient appearing last in construction events), or to the order in which the focus was on each constituent, (e.g., in *Woman–Give–Apple–To–Man,* the focus was first on *Woman* and *Apple*, then on *Give* and finally on *Man*). One participant mentioned "trying to create the event as similar as possible as it was in the clip", i.e., creating as iconic representation as possible. Characteristically, L1 word order was never mentioned in this condition.

5. Conclusions

In this study, we aimed to investigate the role of, and the possible interaction between, different cognitive and semiotic determinants of sign order in event representations, and in particular: (a) the semiotic system used (gestures vs. pictures), (b) the most common word order in the first language of the participants, (c) the ontological status of the Patients ("manipulated" vs. "constructed"), (d) diagrammatic iconicity, (e) event reversibility, and (f) the

proposal for an Act-final "natural order". Importantly, while previous studies have started from (f), and moved up in this list, the way we presented them here represents – iconically – the degree to which we found these determinants to be supported in the study.

Our most important finding, underestimated since the study by Goldin-Meadow et al. (2008) who claimed to find no difference in this respect, was the huge role of semiotic system used in representing an event as a sequence of signs. This difference was reflected not only concerning the role of L1 in the order used, in support of the findings of Vastenius et al. (2016), but also concerning "reversibility". If further studies support our findings of *no* role of reversibility when Pictures (or other non-bodily expressions) are used to represent Agent and Patient, then this would strongly suggest that it is due to the body-as-agent–phenomenon. If so, reversibility should actually *not* be considered an independent determinant.

A reservation that needs to be made, however, is that there was an essential difference in how the two semiotic systems were operationalized in the study. As noted in Section 2.3, due to stark differences in drawing ability, the images in the Picture condition were pre-given, while gestures were, of course, created on the fly. So we need to emphasize that all we can conclude here about sign order preferences in picture use concerns pre-given pictures rather than spontaneously created pictures. Using the latter poses major methodological challenges (how, for example, to instruct participants to draw one event constituent at a time?), but is clearly something to be considered in further studies.

Finally, the placement of a Goal also varied in the two conditions (Gestures vs. Pictures), suggesting that diagrammatic iconicity is not in itself a potent determinant, but always interacting with other factors. And after the key role of semiotic system, we must acknowledge the role of (b) L1 word order, not in general, but when using representations that resemble writing. This is the only plausible explanation why participants overwhelmingly used the order Agent–Act–Patient–(Goal), in the Picture, but not in the Gesture condition.

The next in our order of cognitive-semiotic determinants is (c) ontological status, which clearly played a role in the case of gestures, and to some extent, even for pictures. Even the marginal degree to which it did in the latter case testifies to its potency, given that there are no differences between the linguistic orders (in Swedish) of construction and manipulation events. Further, we place this determinant before (d), as it is not clear if general diagrammatic iconicity helps explain the "ontological status" asymmetry, or only the order in constructed events. As noted above, it did not seem to play a role in the

order in which Goal was placed in the Gesture condition, which motivates our placement further down in our tentative hierarchy. Perhaps controversially, we place the determinant of Act-final "natural order" last in our list, not so much because the effect does not exist – we see it clearly, for example, in the results for Gesture concerning the first hypothesis – but because it is not clear that it offers any explanation, rather than just a description. Like the case of reversibility, it may very well be shown to be an artefact of other factors, such as diagrammatic iconicity.

In sum, many different studies have been conducted to examine what influences the order of using signs in emerging communication systems. We have contributed to the field by pointing to the rather underestimated role of different semiotic systems and the first language of the participants, and how these interact with each other, and with other factors like diagrammatic iconicity. We have elucidated some aspects of this interaction, but its exact nature must be the topic of further studies.

Acknowledgments
We would like to thank all the participants for taking part in the experiments and Mohamed Boudebza for drawing the pictures used in picture tasks and for acting as a confederate. We are also very grateful to an anonymous reviewer for detailed feedback to a previous version of the paper, allowing us to improve it considerably.

REFERENCES

Baker, A., van den Bogaerde, B., Pfau, R., & Schermer, T. (2016). *The linguistics of sign languages: An introduction*. Amsterdam: John Benjamins.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi:10.18637/jss.v067.i01.

Christensen, P., Fusaroli, R., & Tylén, K. (2016). Environmental constraints shaping constituent order in emerging communication systems: Structural iconicity, interactive alignment and conventionalization. *Cognition*, 146, 67–80.

Christensen, P., & Tylen, K. (2013). Gestural representation of event structure in dyadic interaction. Tilburg Gesture Research Meeting 2013. http://tiger.uvt.nl/list-of-accepted-papers.html.

Coppola, M., & Newport, E. L. (2005). Grammatical subjects in home sign: Abstract linguistic structure in adult primary gesture systems without linguistic input. *Proceedings of the National Academy of Sciences*, 102(52), 19249–19253.

Devylder, S. (2018). Diagrammatic iconicity explains assymetries in possessive conctructions: Evidence from Paamese. *Cognitive Linguistics*, 29(2), 313–348.

Forceville, C. (2011). Pictorial runes in Tintin and the Picaros. *Journal of Pragmatics*, 43(3), 875–890.

Galantucci B., Kroos, C., & Rhodes, T. (2010). The effects of rapidity of fading on communication systems. *Interaction Studies*, 11(1), 100–111

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*, 24(7), 1079–1088. https://doi.org/10.1177/0956797612463705.

Goldin-Meadow, S., Mylander, C. S., Wing C., & Özyürek, A. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *PNAS*, 105, 9163–9168.

Hall, M. L., Mayberry, R. I., & Ferreira, V. S. (2013). Cognitive constraints on constituent order: Evidence from elicited pantomime. *Cognition*, 129(1), 1–17.

Hockett, C. F. (1960). The origin of speech, *Scientific American*, 203, 88–111.

Jakobson, R. (1965). Quest for the essence of language. *Diogenes*, 13, 21–38.

Kocab, A., Lam, H., & Snedeker, J. (2018). When cars hit trucks and girls hug boys: The effect of animacy on word order in gestural language creation. *Cognitive Science*, 42, 918–938.

Langus, A., & Nespor, M. (2010). Cognitive systems struggling for word order. *Cognitive Psychology*, 60, 291–318.

McNeill, D. (2005). *Gesture and thought*. Chicago: The University of Chicago Press.

Meir, I., Aronoff, M., Börstell, C., Ilkbasaran, D., Kastner I., Lepic R., Lifshitz Ben-Basat, A., Padden C., & Sandler, W. (2017). The effect of being human and the basis of grammatical word order: Insights from novel communication systems and young sign languages. *Cognition*, 158, 189–207.

Meir, I., Lifshitz, A., Ilkbasaran, D., & Padden, C. (2010). The interaction of animacy and word order in human languages: a study of strategies in a novel communication task. *The Evolution of Language: Proceedings of the 8th International Conference*, 455–456.

Naidu, V., Zlatev, J., Duggirala, V., Van De Weijer, J., Devylder, S., & Blomberg, J. (2018). Holistic spatial semantics and post-Talmian motion event typology: A case study of Thai and Telugu. *Cognitive Semiotics*, 11(2), 1–27.

Peirce, C. S. (1958). *Collected Papers of Charles Sanders Peirce*. Vols. 1–6, C. Hartshorne, & P. Weiss (eds.), Vols. 7–8, A. W. Burks (ed.). Cambridge, MA: Harvard University Press.

Sandler, W. (2012). Dedicated gestures, and the emergence of sign language. *Gesture*, 12, 265–307.

Schouwstra, M. (2012). *Semantic structures, communicative strategies and the emergence of language*. LOT publications 312. PhD thesis, Utrecht Institute for Linguistics. OTS.

Schouwstra, M., & de Swart, H. (2014). The semantic origins of word order. *Cognition*, 131, 431–436.

Stampoulidis, G., Bolognesi, M., & Zlatev, J. (2019). A cognitive semiotic exploration of metaphors in Greek street art. *Cognitive Semiotics*, 12(1), 1–20.

Vastenius, A., van de Weijer, J., & Zlatev, J. (2016). The influence of native language word order and cognitive biases in pictorial event representations. *Cognitive Semiotics*, 9(1), 45–77.

Zlatev, J. (2019). Mimesis theory, learning and polysemiotic communication. In: M. Pe-

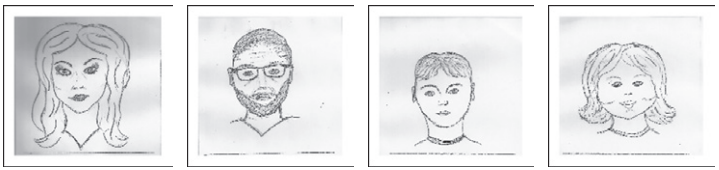
ters (Ed.), *Encyclopedia of Educational Philosophy and Theory*. Dordrecht: Springer.

Zlatev, J., Blomberg, J., & David, C. (2010). Translocation, language and the categorization of experience. In: V. Evans & P. Chilton (eds.), *Space in language and cognition: The state of the art and new directions*, 389–418. London: Equinox.

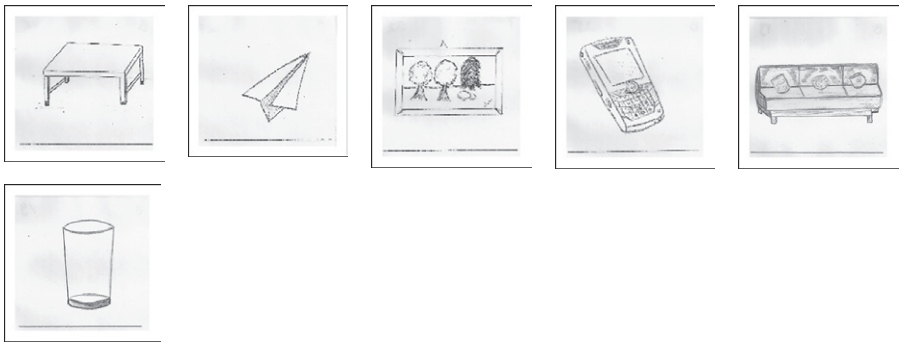
Zlatev, J., Żywiczyński, P., & Wacewicz, S. (2020). Pantomime as the original human-specific communicative system. *Journal of Language Evolution*, 5(2), 1–19.

APPENDIX A
Picture cards used in the picture task

a) Pictures of human participants: Woman, Man, Boy, Girl



b) Pictures of non-human participants: Bag, Bottle, Orange, House, Drums, Clay-snake Chair, Apple, Basket, Bicycle, Ball, Cake, Car, Cat, Chocolate, Hut, Dog, Face, Flower, Garbage bin, Table, Paper-plane, Picture, Phone, Glass, Couch



c) Pictures of acts: Give, Hug, Lift, Make (a cake), Make (a clay-snake/paper plane)/Build (a house), Paint, Pet, Punch, Push, Put, Run, Create (magically with a magic wand), Create (magically with hands), Drop, Eat, Throw, Walk, Bounce (a ball), Break (with hammer)