Gestural representation of event structure in dyadic interaction

Peer Christensen (hr.pchristensen@gmail.com)
Center for Semiotics, Aarhus University
Jens Chr. Skous vej 2, building 1485, 6th floor, 8000 Aarhus C, Denmark

Kristian Tylén (semkt@hum.au.dk)
Center for Semiotics, Aarhus University
Jens Chr. Skous vej 2, building 1485, 6th floor, 8000 Aarhus C, Denmark

Abstract

What are the underlying motivations for the conceptualization of events? Recent studies show that when people are asked to use nonverbal gestures to describe transitive events they prefer the semantic order Agent-Patient-Act, analogous to SOV in grammatical terms. The original explanation has been that this pattern reflects a cognitively "natural order" for the conceptualization of events. However, other types of transitive events have not been investigated in earlier studies. We report experimental findings from a referential game in which pairs of participants used gestures to match shared sets of stimuli depicting two types of transitive events: (i) *object manipulation events* and (ii) *construction events*. We argue that these event types have inherently different logical and sequential structure and, accordingly, will yield different gesture orders. Our findings confirm such predictions: manipulation events predominantly elicited gesture strings with SOV order, while construction events elicited SVO order. The results indicate that participants were highly sensitive to differences in event structure. Even with increased communicative pressure, pairs did not settle on a single order for the two types of events. We conclude that gesture order seems to be motivated by extralinguistic event structure rather than a cognitively "natural order".

Keywords: Event structure; gestural sign emergence; representation; conceptualization; communication system evolution; word order.

Introduction

The motivation for conceptual structure in language and cognition is a hotly debated topic. Gesture is a particularly interesting window into conceptualization due to their affordances for spatial, iconic and symbolic representation and how they unfold in time. Previous studies have indicated that people of different linguistic backgrounds tend to use one specific gesture order (SOV) when asked to describe transitive events by hand gestures alone (Goldin-Meadow, So, Özyürek & Mylander, 2008; Langus & Nespor, 2010). Some participants in these experiments spoke languages with SOV syntactic structure (e.g. Japanese), while others spoke SVO languages (e.g. Spanish). The findings thus suggest that acquired linguistic structure is bypassed in the production of nonverbal gestures.

Different explanations concerning the origin of this predisposition are proposed in the literature. For instance, Goldin-Meadow and colleagues (2008) consider it evidence for innateness of the conceptualization of events, thus transcending acquired linguistic structure. Langus and Nespor (2010) suggest that the SOV order emerges prelinguistically from interactions between modular cognitive systems. However, while replicating findings for "extensional" events (which are equivalent to transitive events in Goldin-Meadow et al., 2008), Schouwstra (2012) found a preference for the SVO-analogous order for "intensional" events exemplified by different types of verbs like "see", "search for" and "think of". This challenges previous innatist and modularist interpretations.

An explanation not considered in the literature is that gestural representations may rather reflect the order, or event structure of the referent situation itself. Some types of 'transitive' actions thus logically depend on the existence of the object. For instance, in *object manipulation* events, the agent and patient must necessarily be co-present for the performed action to relate them physically. In other events involving constructive actions, the agent and action logically precede the patient, which emerges as a product of the action itself.

To test such predictions we conducted an experiment in which participants engaged in a dyadic referential game – a paradigm frequently used in experimental semiotics (see e.g. Galantucci & Garrod, 2011 for a review). The game involved the matching of stimulus pictures using only gesture as a communicational medium. Participants were tested on two different stimulus sets: pictures depicting object manipulation events (e.g. 'a ballerina throwing a paper plane') and pictures depicting object construction events (e.g. 'a ballerina building a sand castle). We hypothesized that object manipulations would motivate a gesture order analogous to SOV in language while construction events would elicit SVO-type gesture strings.

Materials and Methods

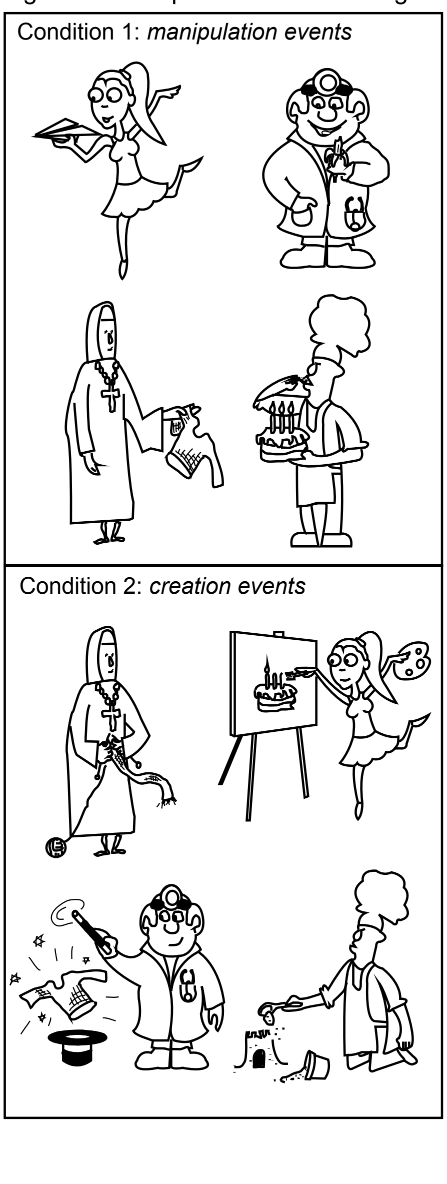
Participants: Twenty-five pairs of participants, all students at Aarhus University, were recruited for the experiment. As shyness and uneasiness were considered potential threats to participants' spontaneous and unconstrained production of nonverbal gestures, participants were recruited in pairs.

Thus, the members of each dyad knew each other beforehand.

All participants were native speakers of Danish and had no prior knowledge of any sign languages or other forms of conventionalized gestural communication. It should be noted that the Danish language has a fixed constituent order, which also includes verbal descriptions of both event types present in the stimuli.

Stimuli: Two sets of stimuli were employed - one for each of two experimental conditions; object manipulation and object construction. Each set consisted of sixteen pictures. In the object manipulation condition, stimulus pictures featured human agents performing manipulative actions with or on an object. In the object construction condition, the stimuli depicted the same set of agents engaging in simple constructive actions (see fig. 1 for examples of stimulus pictures).

Figure 1: examples of stimulus images



Procedure: The experiment was conducted as a referential game. Participants were seated at a table facing each other and took turns in matching and referring to picture stimuli using only gesture. Speaking was thus not allowed. Participants were randomly assigned to one of two condition orders. Then each participant was given a stimulus set and a list informing him/her which picture to communicate in each trial. They also received a sheet for reporting their comprehension of gestures communicated by their co-participants. Since it was important that both participants had the same understanding of the stimulus events before engaging in the task, they were invited to briefly look at the stimuli and ask questions in order to prevent confusion and resolve any possible ambiguities.
An average experimental session lasted approximately 50 minutes and included two conditions. Each condition had two rounds with 16 trials per participant corresponding to the number of stimulus pictures per set. All participants completed the task under both conditions, the order of which was balanced between pairs to control for potential effects related to the condition order. Three cameras were used to collect gesture data. One camera was used to document each session in their entirety. Two smaller cameras were placed on the table in front of the participants in order to capture the gesture strings produced by individual participants in greater detail. At the end of the session, the experimenter debriefed the participants.

Analysis: Videos were coded with respect to gesture order by a research assistant naïve to the hypotheses of the study. Codings subsequently entered a paired *t*-test, performed in Matlab. Due to a strict coding scheme, and since constituent elements (either agent, patient or act) were frequently omitted or produced simultaneously in the gesture strings, participants provided unequal amounts of data points. Therefore, to report the results as accurately as possible, all proportion means and standard errors are weighted with respect to the amount of data points (i.e. coded gesture strings) provided by each participant. Moreover, gesture strings that were ambiguous with respect to order were excluded from further analysis.

Results

As shown in table 1 below, for the object manipulation condition, 1140 gesture strings were coded as either SOV- or SVO-type data points, while the object construction condition yielded 1241 data points.

Table 1: The distribution of gesture strings.

	Object manipulation	Object construction
N gesture strings	1140	1241
w% SOV-type	80	11
w% SVO-type	20	89

In response to the events involving manipulative acts, participants produced a significantly larger proportion of SOV gestures (wM = .800, wSE = .0473) than they did for the stimuli presented in the object construction condition (wM = .113, wSE = .028), $t(49) = 5.375$, $p < .000$ (one-tailed), $r = .73$. The r-value (indicating effect size) suggests that the difference between the stimuli presented in the two conditions had a very large experimental effect.

Our findings indicate that participants prefer the SOV-analogous gesture order for object manipulation events and the SVO order for object construction events. Figure 2 illustrates the weighted mean proportions of SOV and SVO gesture strings produced in the two conditions.

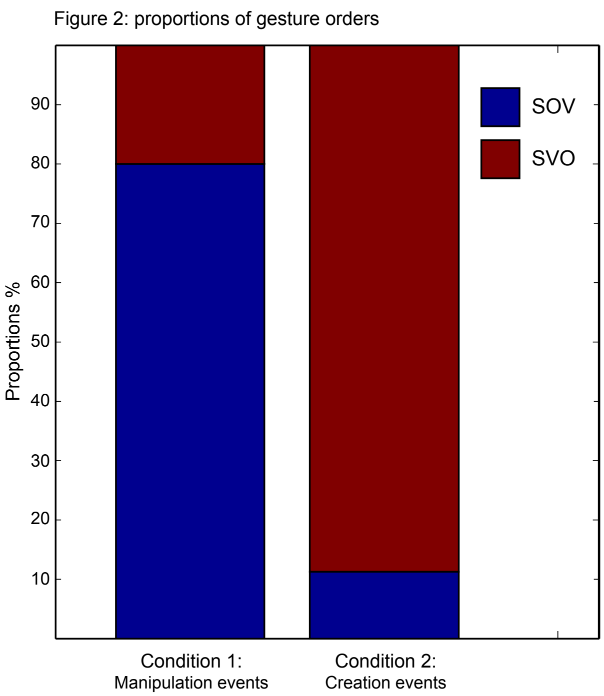


Figure 2: Proportion of SOV and SVO gesture strings used by participants to communicate transitive events

Discussion

The present study was designed to test the prevalence of SOV- versus SVO-type gesture strings in communicative representations of differently structured events. The results obtained from the object manipulation condition support the previous findings that participants produce SOV gesture strings for transitive events involving the manipulation of physical objects, even when participants are native speakers of an SVO-language (Goldin-Meadow et al., 2008; Langus & Nespor, 2010; Schouwstra, 2012). However, the present finding that participants predominantly produced gesture strings of the SVO-type to represent construction events severely challenges the notion that SOV order is an innate structuring principle for the non-verbal conceptualization of events. Thus, the interpretations of the findings offered in the cited papers fail to predict the outcome of the object construction condition. Here participants represented pictures portraying a different type of event, having an inherently logical and sequential order that is not found in the object manipulation events.

There are a number of competing interpretations offered regarding the present and previous findings. Goldin-Meadow et al. (2008) and Langus and Nespor (2010) both posit that SOV order will be predominant in gesture strings produced in response to transitive events. Goldin-Meadow et al. (2008) argue that SOV is cognitively more “natural” than other representational orders. They support this claim with converging evidence from studies of syntactic structure in emerging sign languages. Langus and Nespor (2010), on the other hand, explain their findings by proposing that two modular cognitive systems, a sensory-motor and a conceptual system, interact through a direct link to bypass a computational system for grammar. Consequently, this would lead to a preference for SOV order, according to the authors.

Schouwstra (2012) argues that differences in semantic structure, i.e. the ontological status of objects, leads to different orders in gestural representation. In a series of experiments, she contrasted “extensional” and “intensional” events and found that participants produce both SOV and SVO gesture strings, depending on the event type. For instance, for an intensional event like “thinking of an apple”, it is not a logical necessity that the object is present, or that it even exists. In such cases, objects are then less concrete than in extensional events like “eating an apple”. However, it should be noted that extensional events encompass four different categories of events. Essentially, Schouwstra argues that SOV and SVO orders are linearizations of events that arise in overt gestural representation. These linearizations are then taken to reflect a progression from conceptually concrete to increasingly abstract event elements.

So far, the apparent analogical relation between event structure, conceptualization and gestural representation has not been considered in the experimental literature. Most intensional events are mental and not easily observable in the physical world. Thus, they are devoid of overt event structure, at least experientially. Here, we contrasted two physical event types that differed in one important aspect, which we may refer to as the dependency relation between objects and actions. In object manipulation events, the object must logically be co-present with the agent. In other words, the object is a prerequisite for the action to be performed. Conversely, constructive actions must always precede the object’s presence in object construction events.

Our results suggest that there is a more direct route between the perceived structure of real-life events and how they are conceptualized. Participants thus converge on differential gestural representations of events, because the referent events have different logical and sequential structure that can readily be reproduced in gestural representation.

These experimental findings leave a number of interesting questions open for future investigation. Why don't any verbal languages (to the authors' knowledge) show the same differential word order for the two types of event structures? Might there be other motivational forces working against this differentiation? We might imagine that, for instance, increasing communicative pressures could challenge our findings and potentially make participants converge on one preferred gesture order for both event types. This could be the case, if stimuli from the two conditions were presented fully randomized, rather than in blocks. In other words, it would be interesting to see if participants would demonstrate the same sensitivity to structural differences on a trial-by-trial basis. This would entail sometimes working against forces of structural priming (Branigan et al., 2000). A follow-up experiment along these lines is pending.

Furthermore, besides the main objective of the experiment reported here, movement data was collected for twelve dyads using ActiGraph GT3X+ activity monitors. The monitors allow for fine-grained quantitative analyses of various components of participants' movements while producing communicative gestures in each session. For instance, these data provide the opportunity to investigate how individual signs for referents evolve in the course of each session.

Acknowledgments

We would like to thank the Interacting Minds Center at Aarhus University and the EUROcores project "Digging the Roots of Understanding" for providing funding for this project, and Karen Nissen Schriver for great coding work.

References

Branigan, H.P., Pickering, M.J., and Cleland, A.A. (2000). Syntactic co-ordination in dialogue. *Cognition* 75**,** 13-25

Galantucci, B., Garrod, S. (2011) Experimental semiotics: a review. *Frontiers in Human Neuroscience, 5*(11), 1-15.

Goldin-Meadow, S., So, W.-C., Ozyurek, A. & Mylander, C. (2008) The natural order of events: How speakers of different languages represent events nonverbally. *PNAS,* 105(27): 9163-9168.

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

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