COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2017) 1–21

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12555

When Cars Hit Trucks and Girls Hug Boys: The Effect of Animacy on Word Order in Gestural Language Creation

Annemarie Kocab, Hannah Lam, Jesse Snedeker

Department of Psychology, Harvard University

Received 1 September 2016; received in revised form 12 August 2017; accepted 22 August 2017

Abstract

A well-known typological observation is the dominance of subject-initial word orders, SOV and SVO, across the world's languages. Recent findings from gestural language creation paradigms offer possible explanations for the prevalence of SOV. When asked to gesture transitive events with an animate agent and inanimate patient, gesturers tend to produce SOV order, regardless of their native language biases. Interestingly, when the patient is animate, gesturers shift away from SOV to use of other orders, like SVO and OSV. Two competing hypotheses have been proposed for this switch: the noisy channel account (Gibson et al., 2013) and the role conflict account (Hall, Mayberry, & Ferreira, 2013). We set out to distinguish between these two hypotheses, disentangling event reversibility and patient animacy, by looking at gestural sequences for events with two inanimate participants (inanimate-inanimate, reversible). We replicated the previous findings of a preference for SOV order when describing animate-inanimate, irreversible events as well as a decrease in the use of SOV when presented with animate-animate, reversible events. Accompanying the drop in SOV, in a novel condition we observed an increase in the use of SVO and OSV orders when describing events involving two animate entities. In sum, we find that the observed avoidance of SOV order in gestural language creation paradigms when the event includes an animate agent and patient is driven by the animacy of the participants rather than the reversibility of the event. We suggest that findings from gestural creation paradigms are not automatically linkable to spoken language typology.

Keywords: Word order; SOV; SVO; Gesture; Language creation; Animacy; Transitive sentences

1. Introduction

If you overhear the following sentence in English, "Grandpa dropped the Waterford vase," you immediately know that the expensive vase was the thing that was dropped

Correspondence should be sent to Annemarie Kocab, Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138. E-mail: kocab@fas.harvard.edu

(the theme) and that Grandpa was the one who did the dropping (the agent). You know this for two reasons. First, this is by far the most plausible interpretation: Vases are inanimate and so they are not typically able to perform actions like dropping, but they can be easily dropped. Second, the order of the words in the sentence makes any other interpretation impossible, because in English, as in many other languages, *constituent order* reliably signals the participants' roles. Most of the world's languages are reported to have a dominant or default constituent order (1,188 out of 1,377 in the WALS sample; Dryer, 2013). But languages vary in which order they use. English uses subject-verb-object or SVO, and other orders, like SOV, are ungrammatical. In Turkish, Japanese, and Hindi, SOV is the default. For the three major constituents, S, O, and V, there are six logically possible word orders, but these orders differ substantially in their frequency.[1] SOV and SVO are most prevalent (48% and 41% of the 1,188, respectively), while the third most common word order, VSO, trails far behind at around 8% and the remaining three orders are rare (Dryer, 2005, 2013; Greenberg, 1963). A central question for cognitive science is why some orders are more common than others and what (if anything) this tells us about the human capacity for language.

The prevalence of subject-initial word orders (SOV and SVO) is often argued to stem from our conceptualization of entities and events. First, speakers may prefer for animates to appear first in a sentence and for animates to take the subject role (MacWhinney, 1977). Both preferences can be satisfied if the subject appears first. The appearance of the animate entity at the beginning of the sentence allows the speaker to describe events from the perspective of the agent, which most closely maps onto the speaker's own experience of and interactions with the world (MacWhinney, 1977). Second, speakers may have a preference for temporal isomorphism: to have participants or events that are earlier in the causal chain appear earlier in the sentence (Bever, 1970; Osgood, 1980). The subject is often the causal agent, the one who initiates the event (e.g., hitting, destroying, baking), regardless of animacy (e.g., "The hurricane destroyed the city"). Third, speakers may prefer to refer to entities in the order in which they come into existence. Agents, as opposed to patients, generally exist prior to and independent of the event (Dowty, 1991). For instance, in the sentence "John baked a cake," we can assume that John was present long before the baking began while the cake only emerged as a product of the baking. Finally, the subject-initial order bias could be a side-effect of a preference to begin a sentence with familiar information and move new information to the end (Halliday, 1967; Prince, 1981). The subject of a sentence is typically an entity that has appeared earlier in the discourse (perhaps because we follow the actions of animate, causal agents or perhaps because we tend to see events from the perspective of the entity we have been following). In contrast, the object is more likely to be new to the discourse (e.g., Lambrecht, 1996; Posse, 2000). To communicate effectively, we must begin with what the listener knows and build to what is new. Thus, subject-first word order could be a simple side-effect of this principle of good communication. In sum, while these accounts focus on different conceptual distinctions, they are similar in proposing that subject-initial word orders are prevalent because of the structure of human thought and communication and thus require no specific linguistic explanation.

There is far less consensus about what to make of the relative prevalence of SOV and SVO word orders. We could construe the situation as one where the two orders are of similar prevalence, suggesting that they are equally compatible with the cognitive and communicative constraints on language. We put this possibility aside and return to it in the discussion, because, strangely enough, it is seldom given serious attention in the literature. Most theorists have preferred to take the stronger position and propose that one word order is the natural state or initial state of human language while the other results from secondary historical or communicative processes. Early theorists typically argued that SVO was the natural order, perhaps because it was familiar to English-, French- and German-speaking scholars (see Newmeyer, 2005). More recent theories have tended to favor SOV order (see e.g., Gell-Mann & Ruhlen, 2011; Givón, 1979; Newmeyer, 2000a,b).

One set of clues comes from historical linguistics and typological studies. While languages that start out as SOV have diachronically changed to SVO, the reverse has not been observed (Gell-Mann & Ruhlen, 2011). Researchers have argued that the largely unidirectional shift from SOV to SVO suggests that SOV was the initial preferred word order for languages (Gell-Mann & Ruhlen, 2011; Givón, 1979; Newmeyer, 2000a,b).[2] Further evidence for a SOV bias comes from systems created naturally by deaf individuals. In homesign systems, gestural communication systems created by deaf children in the absence of linguistic input, two argument utterances tend to have SV or OV orders, rather than VS or VO orders (Goldin-Meadow, 2003; Goldin-Meadow & Mylander, 1998; Goldin-Meadow, Özyürek, Sancar, & Mylander, 2008). Similarly, Al-Sayyid Bedouin Sign Language, an emerging sign language, has SOV as its dominant word order, even though the Arabic dialect spoken by the hearing members of the community has a SVO order (Sandler, Meir, Padden, & Aronoff, 2005). Sign languages in general show a strong tendency to accept SOV orders. In their survey of 42 existing sign languages, Napoli and Sutton-Spence (2014) found that SOV order is permissible in all of them, while SVO is permissible in only some.

These typological observations suggest that SOV may be a default for human languages, reflecting properties of either our conceptual system or our linguistic capacity. These observations align with empirical evidence from gestural language paradigms in the laboratory, which also show a bias for SOV order in communication. In such experiments, hearing non-signing participants are asked to describe various events using only gestures and no speech. Participants produce a variety of gestures, using their hands to represent entities (e.g., a closed fist to represent a ball) or using their own body to represent the attributes of or actions taken by an animate entity in the event (e.g., placing a crooked finger on one's upper lip to signify a mustached man, or putting up two hands and acting out pushing). Interestingly, gestural paradigms have been shown to be successful in shifting speakers away from their native language biases. Gestural descriptions do not necessarily follow the dominant or default order of participants' native languages and there is consistency across participants in the ordering of gestures. Across different language groups (Chinese, English, Spanish, and Turkish), which have different dominant word orders, gesturers primarily produce SOV, SV, and OV descriptions to describe events with two entities, like a man swinging a pail (Goldin-Meadow et al., 2008). This preference for SOV orders in gesture has been replicated by other laboratories (Gibson et al., 2013; Hall et al., 2013) and extended to speakers of Italian, Japanese, and Korean (Gibson et al., 2013; Langus & Nespor, 2010). Goldin-Meadow et al. (2008) attribute this pattern to the relative salience of entities (as compared to actions), which leads participants to mention them first. On this argument, the gestural language paradigm is seen as a method that forces participants to abandon their native language and its biases, pushing them to rely on their mental representation of events to organize their gestural descriptions.

If the SOV word order is a cognitive default, then the question arises of why the SVO order occurs as often as it does (and why languages would switch from SOV to SVO). Recent findings from gestural language studies suggest one possible answer. When an event would be unambiguous regardless of the word order used to describe it (GIRL BALL KICK, GIRL KICK BALL, BALL GIRL KICK), because the agent is animate and the patient is inanimate (the ball is not capable of kicking but can be kicked), participants predominately produce SOV order. However, when the agent and the patient are both animate (specifically, human) and capable of carrying out the action (GIRL BOY KICK, GIRL KICK BOY, BOY GIRL KICK), the event becomes reversible (both the girl and boy are capable of kicking) and participants are far less likely to gesture using SOV, producing SVO instead (Gibson et al., 2013).

Two hypotheses have been proposed to account for this shift in gestural paradigms from SOV to SVO when an animate patient is introduced. The first, *the noisy channel account*, is based on information theory and the notion that signals must be robustly recoverable despite the inevitable errors (*noise*) that occur in production and perception. This account proposes that the reversibility of an event, which may lead to confusability, drives the shift away from SOV order (Gibson et al., 2013). Word orders that have the subject and object separated by the verb, such as SVO, are argued to be more robust against information loss. For instance, if the producer uses SOV order to describe an event, such as BOY GIRL KICK, and the listener misses the beginning of the sentence, she would get GIRL KICK, which could be either an SV sentence (in which the girl is the agent) or an OV sentence (in which the girl is the patient). On the other hand, if the producer used SVO order, the listener would get KICK GIRL which, if the listener knows the word order constraints, could only be a VO fragment making it easier to recover the intended meaning, where the girl is the patient, the entity who was kicked.

The second hypothesis, *the role conflict account*, proposes that this change in word order is driven by the tendency to represent animate, human entities via embodied gestures (Hall et al., 2013) and thus might be specific to the manual modality. When participants embody two animate entities, as is the case in reversible events, having the verb come right after the object (as is the case with SOV order) results in a role conflict, since the gesturer takes on the role of the object (the patient of the event) immediately before describing the action. Take, for instance, an event involving a boy pushing a girl. If the producer uses a SOV word order, then the gesturer takes on the role of the boy, then the girl, and then enacts pushing, without any interruption between being the boy and acting out the role of the girl. To avoid this role conflict, Hall et al. (2013) propose that when describing events with two animate human entities, gesturers avoid SOV word order and switch to orders where the subject immediately precedes the verb. They point out that this can be achieved not only using the SVO order but also with orders such as OSV, SOSV, and so on. Critically, Hall and colleagues find that the introduction of an animate patient also leads to an increase in the frequency of these other word orders, suggesting that animacy has effects that go beyond those predicted by the reversibility hypothesis (Hall et al., 2013). Critically, this observation also calls into question the parallel that is made between word order in the gestural paradigm and historical changes in the prevalence of different word orders in spoken languages. To the best of our knowledge there is no strong evidence that OSV and SOSV orders, like SVO orders, have become more common over the millennia. We return to this point later in the discussion.

This study seeks to distinguish between these two hypotheses. The noisy channel account attributes the shift to SVO order to the reversibility of sentences with animate patients and agents. The role conflict hypothesis attributes this change directly to the animacy of the patient. Thus, to distinguish them, we disentangled reversibility and patient animacy, by looking at gestural sequences for events with two inanimate participants (e.g., a hitting event involving a car and a truck) that are reversible (a car can hit a truck, and a truck can hit a car). If the shift away from SOV word order for events with two animate participants is driven by the reversibility of the event (the noisy channel account), then we should see fewer SOV sequences (and more verb medial sequences) for events in which an inanimate agent acts on an inanimate patient since they are also reversible events. In contrast, if the shift away from SOV word order for animate-animate events is driven by the desire to avoid role conflict, then there should be no marked decrease in the use of SOV sequences for the inanimate-inanimate events since neither argument is likely to be expressed with an embodied gesture, making role conflict irrelevant.

2. Method

2.1. Participants

Forty monolingual native English speakers were recruited from the Harvard University Psychology Study Pool in exchange for partial course credit or payment. The age range of the participants was 14 to 40 years (M_{age}: 18 years). All participants reported no language-related impairments and no prior knowledge of any sign language.

2.2. Materials

Stimuli consisted of 48 short Flash animated clips depicting simple events of three types (see Fig. 1 for schema of events and Appendix for list of events). All clips were pretested for clarity by 50 participants on Amazon Mechanical Turk.

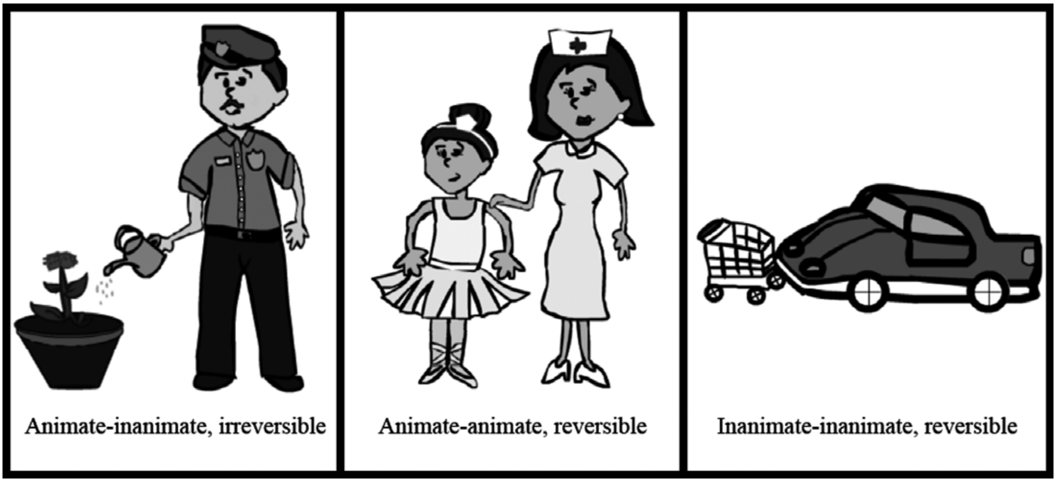

Fig. 1. Still image examples of the three types of animated stimuli events. The first panel depicts an animate-inanimate, irreversible event (a police officer watering a plant). The second panel depicts an animate-animate, reversible event (a nurse tapping a ballerina). The final panel depicts an inanimate-inanimate, reversible event (a shopping cart hitting a car).

The first type of event depicted a human agent performing a transitive action on a non-human patient (*animate-inanimate*, *irreversible*). The second type of event depicted a human agent performing a transitive action on a human patient (*animate-animate*, *reversible*). The third type of event depicted an inanimate agent performing a transitive action on an inanimate patient (*inanimate-inanimate*, *reversible*).

The position of the agent and patient on the screen (left or right side) was counterbalanced across all trials. The stimulus events were presented in a PowerPoint presentation on a Macintosh laptop.

2.3. Design

All participants saw eight clips of each event type. Half of the participants received one list, and half received the other list. Half of the animate-inanimate, irreversible events (e.g., a police officer watering a plant) appeared in one list and the other half in the other list. Reversible events, both animate-animate and inanimate-inanimate, were counterbalanced, so that one list contained one version of the event (e.g., a nurse tapping a ballerina, a shopping cart hitting a car) and the other list contained the reversed version (e.g., a ballerina tapping a nurse, a car hitting a shopping cart). The order of the clips was pseudo-random, with two constraints: Sequential clips only shared at most one entity in common and no more than two sequential clips portrayed the same event type.

The experiment was split into two blocks: Participants were asked to describe the events using gestures in the first block and then asked to verbally describe the same events in English in the second block. Verbal descriptions were used to confirm that our clips consistently elicited descriptions with SVO order.

2.4. Procedure

Participants were told that they would see short video clips, which they would first describe using gesture and no speech, and then describe using a single sentence in English. The experimenter asked participants to provide as much information as possible in their gestures, such that another person who did not see the clips would be able to understand the event, but, following Hall et al. (2013), she did not explicitly prompt participants to include information about the agent, patient, or action. Participants were also asked to refrain from pointing to objects in the room (e.g., "shoes" and "table"). Participants saw five practice trials and received feedback after each one. Feedback was normally in the form of reminding participants not to point at or use objects in the room (like their shoes or the table), but rather, to gesture them. After completing the practice trials, participants began the 24 test trials. The experimenter remained in the room but did not speak or provide any more feedback. Participants were allowed to watch the clips as many times as needed. When the participants finished describing an event, they pressed a key on the laptop to proceed to the next trial.

Once participants had described all the trials using gestures, they began the second block, where they described each event in a single sentence in English. These videos appeared in the same order as in the first block. Again, the participants described five practice events and received feedback from the experimenter after each one before they described the test events.

2.5. Coding

Gestures were coded offline, by the second author, using the ELAN video and audio annotation program (http://tla.mpi.nl/tools/tla-tools/elan/) developed by the Max Planck Institute for Psycholinguistics (Lausberg & Sloetjes, 2009; Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006). A second coder coded 10% of the descriptions. To be able to readily compare our findings with those in the previous literature, our coding scheme drew heavily on the coding schemes used in Hall et al. (2013) and Gibson et al. (2013). Specifically, gesture strings were coded for mention of the subject, object, verb, or other. Consecutive gestures that described the same referent (e.g., "TALL-MUSTACHE-HAT) were grouped together as one constituent (e.g., MAN). Gestures that provided scene information, such as a description of water in an event involving boats, were coded as other and were not used to determine the constituent order. Following Hall et al. (2013), a string of gestures was coded as a single phrase if participants did not drop their hands or if there were no pauses longer than 2 s. When participants produced multiple gestural phrases, we included only the final string following the pause or interruption in our analyses. Out of 960 trials, we observed only 30 instances of multiple gesture phrases (around 3% of the descriptions), indicating that multiples were rare.

In accordance with Hall et al. (2013), when the referent of the gesture was unclear, the trial was coded as ambiguous. If participants produced gestures that contained information about multiple constituents (e.g., one hand used to represent a boat and the other to represent a canoe, and moving together the two hands to represent collision), the trial was coded as simultaneous. Ambiguous and simultaneous trials were excluded from analysis. If the participant referenced an object by pointing to it in the room (e.g., pointing to a table or to their own shoes), the trial was also excluded. Based on these criteria, around 4% of the trials were excluded.

We coded the participants' use of embodiment to see whether participants only embodied animate entities, as we expected, or if they also embodied inanimate entities, such as shopping carts and cars. We categorized embodiment into three types: subject embodiment, object embodiment, and verb embodiment. Sequences were coded as exhibiting subject embodiment when participants used their body to represent the (animate) agent (e.g., raising their arms above their head to represent a ballerina) and/or represent an attribute of the agent (e.g., finger to upper lip to represent a mustache). Sequences were coded as exhibiting object embodiment when participants used their body to represent the patient, which could be animate or inanimate depending on the event type. Sequences were coded as exhibiting verb embodiment when participants used their body to represent an action (e.g., mimicking the action of opening a jar or kicking).

3. Results

3.1. English production

While participants gave English descriptions in the second block of the experiment, we report these results first. Overall, around 98% of the trials (949 of 960 trials) were described using canonical SVO English word order (the remaining were passive or coordinate sentences).

3.2. Gesture production

Within animate-inanimate events, SOV was the most common word order produced, comprising around 38% of the descriptions (see Fig. 2). The second most common word order, OV, made up approximately 14% of the descriptions, and SVO constituted only 5% of the utterances. For the animate-animate events, the use of SOV dropped to 8% and the use of SVO increased to 19%. The use of OSV also increased to 15%. For the inanimate-inanimate events, OSV was the most frequent word order observed, accounting for around 24% of the descriptions. The second most frequent order, SOSV, made up 10% of the gestural strings and SOV comprised around 8%. Inter-rater reliability was 95.8%.

Our analyses focused on two questions. First, we looked explicitly at embodiment, coding whether participants embodied the entities in the event and whether they acted out the action while in the role of the agent (Section 3.2.1). Second, we looked at the word orders that participants used. Our initial analysis of word order (Section 3.2.2) focused narrowly on SVO and SOV utterances, because these orders have been central in prior discussions of the gestural language creation paradigm (see e.g., Gibson et al., 2013; Goldin-Meadow et al.,

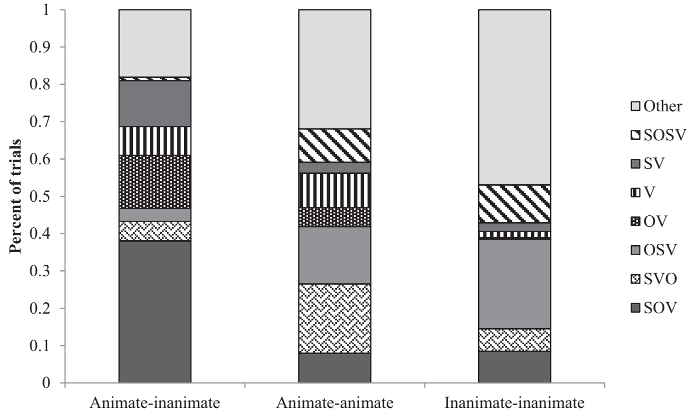

Fig. 2. The seven most frequent word orders produced by event type. All other orders and excluded descriptions are graphed as "Other."

2008) and because these are the orders for which the noisy channel and role conflict hypotheses make clearly contrasting predictions for the inanimate-inanimate trials. The noisy channel account predicts that inanimate-inanimate trials should pattern with the animate-animates trials since the events are reversible and thus use of non-verb-medial orders would be confusing. The role conflict account predicts that they should pattern with the animate-inanimate trials since no role conflict would result from the use of SOV. Our subsequent analyses of word order tested each of the two hypotheses more broadly. In Section 3.2.3, we tested the noisy channel hypothesis by looking at changes in the proportion of verb-medial orders across conditions (SVO, OVS, SVOS, OSVO, SOVS, SVOV). In Section 3.2.4, we tested the role conflict hypothesis by looking at changes in the proportion of order with OV sequences. All of the above analyses, with the exception of the analysis of embodied entities and actions, looked only at the gestural descriptions that contained at least one mention of each argument in other words, both the S and O (735 of 960 trials).

3.2.1. *Embodiment of animate and inanimate entities*

To test our prediction that participants would embody animate, but not inanimate, entities, we looked at participants' use of embodiment (of the subject, object, and verb) across the different event types (see Fig. 3). Using the R programming language, the data were submitted to logistic mixed-effects regression models with item and subject as random effects and use of each type of embodiment (describing the subject, object, and verb) entered as 1 and the absence as a 0. All reported analyses coded trial type using two dummy variables, with the animate-inanimate trials acting as the baseline (the intercept). Using the ANOVA function, we compared two models: one model with trial type as a predictor variable and one model without trial type as a predictor variable.

An example of our model specification (with trial type as a predictor variable) in the common glmer syntax is as follows: SubjectEmbodiment ~ TrialType + (1|Trial) + (1|Subject), family = binomial ("logit"). If the ANOVA comparison was significant, indicating

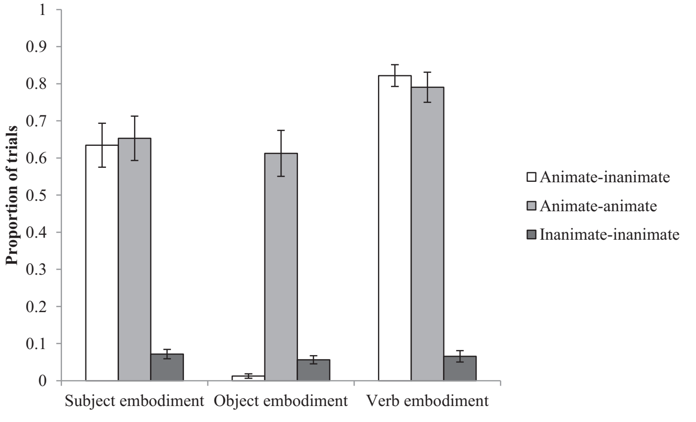

Fig. 3. Validating the effects of animacy on use of embodiment. Proportion of trials with subject, object, and verb embodiment. Error bars represent standard error.

Table 1
Results of the mixed-effects logistic regression predicting use of embodiment by trial type. Animate-inanimate trials represent the intercept

Embodiment	Predictor	β	Wald's *z*	*p*-value
Subject embodiment				
	Animate-inanimate	0.854	1.694	
	Animate-animate	0.181	0.373	.709
	Inanimate-inanimate	−5.281	−9.196	<.0001
Object embodiment				
	Animate-inanimate	−7.016	−6.165	
	Animate-animate	7.831	6.571	<.0001
	Inanimate-inanimate	2.110	1.699	.089
Verb embodiment				
	Animate-inanimate	2.755	6.618	
	Animate-animate	−0.741	−1.605	.108
	Inanimate-inanimate	−6.291	−11.753	<.0001

that the model with the variable of interest as a predictor was a better fit, we conducted follow-up comparisons to determine where the difference between groups lay.

There was a significant effect of trial type for all three types of embodiment: subject embodiment, object embodiment, and verb embodiment (see Table 1). Follow-up pairwise comparisons looking at subject embodiment showed that the difference between animate-inanimate and animate-animate trials was not significant ($p = .19$), but the difference between animate-inanimate and inanimate-inanimate was significant ($p<.001$) as was the difference between animate-animate and inanimate-inanimate ($p < .001$). For object embodiment, there was a significant difference between the animate-inanimate and animate-animate trials ($p < .001$), and between the animate-animate and inanimate-inanimate trials ($p < .001$). There was not a significant difference in the use of object embodiment between animate-inanimate and inanimate-inanimate trials (p = .095). Looking at verb embodiment, the difference between animate-inanimate and animate-animate was not significant (p = .58), but the difference between animate-inanimate and inanimate-inanimate was significant ($p < .001$), as well as the difference between animate-animate and inanimate-inanimate ($p < .001$). These results confirm our prediction that participants use embodied gestures to represent the agent when animate, the patient when animate, and the action when carried out by an animate agent. Participants use embodied gestures far less when the entity (agent or patient) is inanimate and when the action is performed by an inanimate agent.

3.2.2. *Testing the noisy channel hypothesis: Distribution of SVO*

Next, while participants produced a large variety of word orders (see Fig. 2), to narrowly test the noisy channel hypothesis, we looked at the use of SVO as opposed to SOV across conditions. To do this, we restricted our sample to include only descriptions that were SOV or SVO and constructed logistic mixed-effects regression models with item and subject as random effects, looking at the use of SVO order (entered as 1, Fig. 4). We compared two models, one with trial type as a predictor and one without trial type as a predictor, with the animate-inanimate trials as the baseline. The model with trial type as a predictor performed significantly better than the model without (see Table 2), indicating that there were differences in the use of SVO order across the different trial types. We then conducted pairwise comparisons to determine where the difference lay. There were significant differences between all three trial types: animate-inanimate and animate-animate ($p < .001$), animate-inanimate and inanimate-inanimate ($p < .001$), and animate-animate and inanimate-inanimate ($p < .001$). In other words, following findings from previous literature, we observe a significant difference in the use of SVO when comparing animate-inanimate, irreversible events with animate-animate, reversible events. The new condition in this study, the inanimate-inanimate, reversible events, showed an

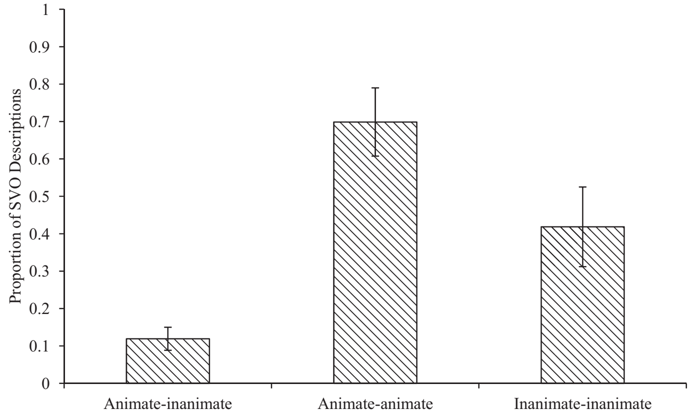

Fig. 4. Proportion of trials with SVO descriptions over SVO and SOV descriptions. Error bars represent standard error by subject.

Table 2
Results of the mixed-effects logistic regression predicting use of SVO by trial type. Animate-inanimate trials represent the intercept

Word order	Predictor	β	Wald's *z*	*p*-value
SVO				
	Animate-inanimate	−3.096	−6.538	
	Animate-animate	3.982	7.713	<.0001
	Inanimate-inanimate	2.226	4.108	<.0001

intermediate pattern with greater use of SVO than the irreversible events but less use of SVO than the animate-animate events.

In short, the analysis above provides some support for the role conflict hypothesis (the proportion of SVO descriptions is greater for reversible animate events than for reversible inanimate events), but it also provides some support for the noisy channel hypothesis (the proportion of SVO descriptions is greater for reversible inanimate events than for irreversible animate-inanimate events). However, as Fig. 2 shows, focusing solely on the SVO and SOV descriptions may be misleading, given the large number of responses that employed other orders and the increase in these responses for the reversible events. The noisy channel hypothesis makes predictions about these other sequences as well: Verb-medial word orders should be greater for reversible events regardless of animacy. The role conflict hypothesis predicts that use of orders with the O appearing immediately before the V should be lower when describing events with animate agents and patients (where gesturers are likely to use embodied gesturers, giving rise to a potential role conflict) compared to events where the patient or both entities are inanimate.

3.2.3. Testing the noisy channel hypothesis: Distribution of verb-medial orders

Our next analysis provided a broader test of the predictions of the noisy channel hypothesis. Each utterance was coded as either having verb-medial order (SVO, OVS, SVOS, OSVO, SOVS, SVOV) or not having a verb-medial order (Fig. 5). Logistic mixed-effects models were constructed, with item and subject as random effects. Again, we compared two models, one with trial type as a predictor and one without trial type as a predictor, with animate-inanimate trials as the baseline. Our analysis showed a significant effect of trial type on use of verb-medial orders (see Table 3). Pairwise comparisons revealed that the difference between animate-inanimate and animate-animate was significant ($p < .001$) as was the difference between animate-animate and inanimate-inanimate ($p < .001$), but the difference between animate-inanimate and inanimate-inanimate was not ($p = .58$).

These results are in clear conflict with the predictions of the noisy channel hypothesis, where reversibility is argued to be the key pressure behind participants' avoidance of SOV word order. Instead, we found that the animacy of the patient seems to determine whether verb-medial orders are used.

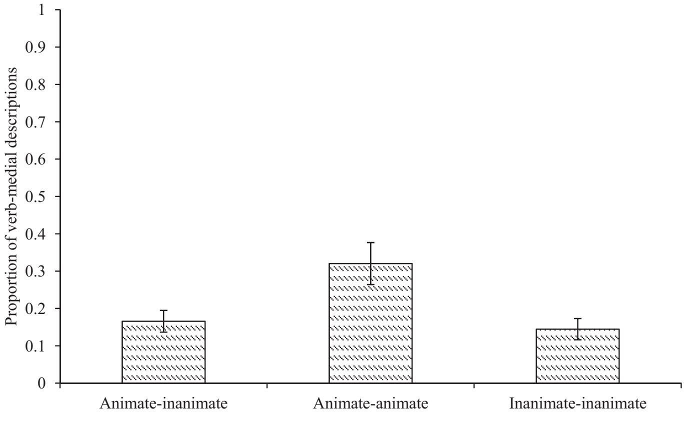

Fig. 5. Proportion of trials with verb-medial descriptions over all descriptions with at least one instance of each constituent (S, O, and V). Error bars represent standard error by subject.

Table 3
Results of the mixed-effects logistic regression predicting use of verb-medial orders by trial type. Animate-inanimate trials represent the intercept

Word Order	Predictor	β	Wald's *z*	*p*-value
Verb-medial orders				
	Animate-inanimate	−2.336	−7.301	
	Animate-animate	1.180	4.380	<.0001
	Inanimate-inanimate	−0.074	−0.260	.795

3.2.4. Testing the role conflict hypothesis: Distribution of orders with O appearing immediately before V

Turning to the role conflict hypothesis, we coded each of the utterances with all three constituents as either having an OV sequence or not having an OV sequence (Fig. 6). We constructed two logistic mixed-effect models, with item and subject as random effects, comparing one model with trial type as a predictor and one model without trial type as a predictor, with animate-inanimate trials as the baseline. There was a significant effect of trial type on use of descriptions with the O before the V (see Table 4). Pairwise comparisons showed that the difference between animate-inanimate and animate-animate was significant ($p < .001$), as well as the difference between animate-inanimate and inanimate-inanimate ($p < .001$), but the difference between animate-animate and inanimate-inanimate was not significant ($p = .88$). The avoidance of orders with the O preceding the V in the inanimate-inanimate trials is not expected under the role conflict hypothesis, where events with no animate entities should pattern more similarly to events with an animate agent and inanimate patient because there is no role conflict stemming from the use of embodied gestures.

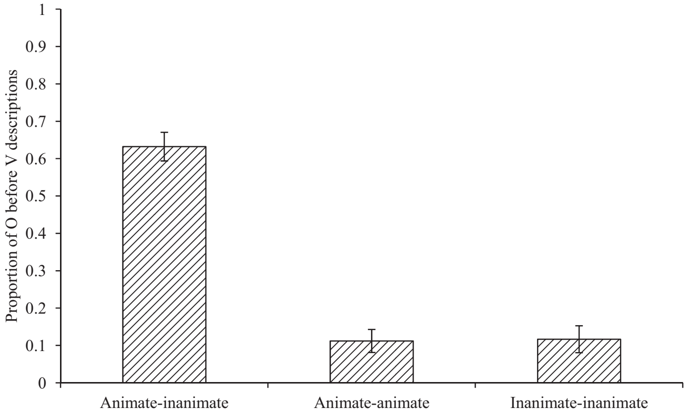

Fig. 6. Proportion of descriptions with the O placed immediately before the V out of all descriptions with at least one instance of each constituent (S, O, and V). Error bars represent standard error by subject.

Table 4
Results of the mixed-effects logistic regression predicting use of orders with OV sequences by trial type. Animate-inanimate trials represent the intercept

Word Order	Predictor	β	Wald's *z*	*p*-value
OV sequences				
	Animate-inanimate	0.791	3.183	
	Animate-animate	−3.134	−10.062	<.0001
	Inanimate-inanimate	−3.102	−10.639	<.0001

4. Discussion

Our experiment set out to test two competing hypotheses for a prior, robust finding in the gestural language creation studies: the avoidance of SOV order (and the apparent preference for SVO) when describing events involving animate agents and patients. The first account, the *noisy channel hypothesis*, posits that the reversibility of an event involving two animate entities drives participants to shift away from the SOV word order toward an order in the verb separates the two entities making the sentence more robust to information loss. This constraint and a bias for agent-first word orders would lead to a preference for the SVO order. The alternative account, the *role conflict hypothesis*, suggests that when using a modality where animate entities can be, and often are, embodied, gesturers run into a role conflict when an animate patient immediately precedes the verb, as it does in the SOV order. To avoid this role conflict, participants employ other orders, such as SVO, but also OSV and SOSV. Our results are not predicted by either hypothesis, in their current form. Nevertheless, four clear findings emerged from this work, which will constrain future theorizing. We discuss each one in turn.

First, we verified a critical assumption of the role conflict hypothesis (Section 3.2.1). We found that people frequently embodied animate agents and patients but rarely embodied inanimate agents or patients. Participants also used an embodied enactment of the verb when the agent was animate but not when it was inanimate. These analyses confirmed that under the role conflict hypothesis, we should expect the inanimate-inanimate sequences to group with the animate-inanimate ones.

Second, we replicated the previous findings of a preference for SOV order when describing animate-inanimate, irreversible events as well as a decrease in the use of SOV when presented with animate-animate, reversible events. The drop in SOV was accompanied by an increase in the use of SVO and OSV orders.

Third, we found no evidence that the drop in use of SOV is driven by the reversibility of the event. While we noted a decrease in the use of SOV in the inanimate-inanimate events, similar to what was observed with the animate-animate events, we also saw significantly *less* use of SVO in the inanimate-inanimate events, compared to the animate-animate events, in favor of OSV and other orders. As a result, our initial, and most narrow analysis, offered only partial support for the noisy channel hypothesis (Section 3.2.2). The use of SVO in the inanimate-inanimate condition was intermediate between the other two conditions, rather than patterning with the animate-animate condition. When we tested the noisy channel hypothesis more directly, comparing the proportion of verb medial sequences in conditions (Section 3.2.3), we observed that the inanimate-inanimate events were rarely described with verb-medial orders, like the irreversible animate-inanimate events. On the face of it, these findings might appear to support the role conflict hypothesis, since the word order usage in the three conditions grouped together on the basis of object animacy rather than reversibility. But critically, the role conflict hypothesis makes no claims about verb-medial constructions per se; rather, it makes predictions only about the appearance (or absence) of OV sequences.

Fourth, we found no clear evidence that the drop in use of SOV is driven by attempts to avoid role conflict. Recall that on the role conflict hypothesis we should expect a decrease in OV sequences whenever the object is animate and can be embodied. Thus, in our initial narrow analysis (Section 3.2.2), the hypothesis predicts that inanimate-inanimate sequences should pattern with the animate-inanimate ones, and not the animate-animate ones. The fact that they land neatly between the other two could be taken as evidence that role conflict is half the story. But this interpretation is undermined by the results of the analysis that targets the role conflict hypothesis most directly and most broadly (Section 3.2.4): OV sequences were as rare in the inanimate-inanimate condition as they were in the animate-animate condition. While, at first glance the results of this analysis might seem to support the noisy channel hypothesis (since the conditions pattern on the basis of reversibility), that hypothesis makes no predictions about OV sequences per se. While neither hypothesis offers a comprehensive explanation for our results, they articulate some of the possible pressures, animacy, and reversibility, acting on word order preferences in these paradigms.

What then can we make of this set of findings? Our results do not neatly align with the existing hypotheses for word order shifts across different types of events. The key to understanding these findings comes from close examination of the specific word orders that were produced in each of the three conditions. There are two striking differences between the novel inanimate-inanimate condition and the other two conditions. First, word order is generally more variable (with 50% of the utterances classified as "other," see Fig. 2). Second, there is a considerable increase in the use of OSV sequences. Neither of these effects is predicted by the existing hypotheses. On the noisy channel account, OSV is a poor word order for a reversible event. On the role conflict account, OSV orders might be favored for events with animate objects, but not for events with inanimate objects. While this rise in OSV orders was not expected, we suspect that this finding may reflect more fine-grained constraints on how events are depicted in the relatively simple systems that arise in these short laboratory studies.

First, for inanimate-inanimate events participants almost never embodied the agent, the patient, or the verb. Instead, they often used one hand to represent each entity. When using two hands to depict motion and contact events (which the majority of our inanimate-inanimate trials were), gesturers may prefer to have moving entity, the subject, appear right before the action. While the subject appears before the action in SVO, this order may be dispreferred because its use would require the gesturer to move the hand representing the subject (S) through space to represent the action (V) and then suddenly make the patient appear (O). This sequence may imply an event where the action creates the object, rather than one where the preexisting object is affected by the subject (see Schouwstra & de Swart, 2014, for data consistent with this hypothesis). Thus, the only option remaining is the OSV sequence.

The move from embodied action to using the hands may also explain the large number of longer sequences in the inanimate-inanimate condition (the sequences classified as "others" in Fig. 2). Gesturers may need initially to use iconic gestures to adequately identify the inanimate subject and object (since embodiment is not available as a strategy), before they re-represent them in the enactment of the event. Indeed, for the inanimate-inanimate trials that were classified as "other," 93 of 139 had just one instance of the verb. Of those 93 trials, 44 had the single verb in the final position (example orders: OSOSV, SOSOSV). For the 49 trials where the verb did not appear at the end of the sequence, at least one mention of both the S and O preceded the V in all but one trial (example order: SOSVO). For the trials that had multiple instances of the verb in the sequence, a little over half had at least one mention of the S and O preceding the first verb (26 of 46). Thus, it seems that the increase in "other" sequences in the inanimate-inanimate trials is driven by the use of sequences with additional mentions of the S and/or the O (but less often the V), generally at the beginning of the sequence.

These findings, and the explanations we have given, raise questions about the scope of the gestural language creation paradigms, the degree to which the results generalize across predicate types, and their relevance for understanding the factors that shape sign and spoken languages. The gestures that are created in this setting are highly iconic, and embodiment as a strategy seems to be readily available to individuals with no prior sign language experience. Additionally, the circumstances of the communication context do not provide a strong incentive to establish a stable lexicon: The experimental session is short, the listener is unknown, and no feedback is present to confirm that a mapping has been made. In contrast, while emerging sign languages, like Nicaraguan Sign Language

(NSL), make use of embodiment and have iconic signs, their users quickly converge on stable lexicons (e.g., Richie, Yang, & Coppola, 2014) and the language has arbitrary mappings and rich grammatical marking (e.g., Pyers, Senghas, Plançon, Zola, & Reynoso, 2016; Senghas, 2003). To what extent will the factors that drive performance in these laboratory studies, such as embodiment and the need for ad hoc iconicity, be relevant to the development of sign languages once stable lexical items are present? If modality constraints are central to this paradigm, as they appear to be, then we face the question of whether we can use this method to generate data about the communicative pressures that have shaped spoken languages.

To date, the link between gestural language creation and structure of spoken languages has largely rested on some broad parallels between typological patterns and the observed gestural sequences. Specifically, (a) SOV is the most common order of languages in the world and the most common order in the initial studies using canonical animate-inanimate events; (b) SVO is nearly as common as SOV and it appears in gestural studies for animate-animate events; (c) and, historically, there is often a shift from SOV to SVO. These three facts have lead some to suggest that SOV might be a cognitive default and that SVO might arise over historical time as animate-animate events nudge a language toward another structure. While this is a bold and intriguing conjecture, it is a potentially problematic proposal for at least three reasons.

First, a possible prediction that emerges from the gestural language studies is that these pressures could be resolved by adopting a mixed word order system in which SVO is used for animate-animate events and SOV for animate-inanimate events.[3] However, WALS notes that only a few languages (67 of 1,377) have two systematic defaults (Dryer, 2013). Just 29 languages have the predicted pattern (SVO and SOV) and the examples that are given suggest that the choice typically depends on syntactic factors, such as whether the verb is in a main or subordinate clause or whether there is an auxiliary verb (German and Dutch are typical examples).

Second, it is difficult to reconcile the radically different time scales of the two phenomena. If languages were initially SOV and change occurs in one direction only, then it would seem that it has taken at least 60 million years for half of the SOV languages to become SVO languages.[4] This possibility seems unlikely given the consistent presence, and social salience, of events that involve animates acting upon each other. While a clear advantage of laboratory experiments is that a controlled environment may be used to identify and isolate some factors that shape languages, such as predicate types or modality, a disadvantage is that we may underestimate the effects of other factors, particularly those that are less amendable to being directly tested.

Relatedly, third, historical patterns suggest that changes in word order from SOV to SVO first begin with a loss of case, which has not been robustly demonstrated in gestural paradigms (see Hall et al., 2013). Languages that have rich case systems typically have freer word order (e.g., Kiparsky, 1996; Kroch, 2001); swapping the order of arguments does not change their grammatical role, in contrast to fixed order languages like English. Over time, through language contact and trade, such languages often lose their case system and move to a more fixed word order, where constituent order indicates thematic roles. In contrast, languages that begin with no case system typically have fixed word order (SVO) and rarely move to more flexible word order.

Indeed, these observations about case marking offer a complementary explanation for the undirectionality of diachronic change from SOV to SVO. A switch from a default word order to another word order may not be driven just by properties of the events or modality concerns. Change may additionally result from the cascading effects of the dependencies between features of a language (e.g., case marking and more flexible word orders tend to go together). Note that this suggestion is distinct from proposing that the semantic properties of events or constraints on the production or comprehension of language do not have an effect on why certain properties bundle together. Indeed, it is likely not a historical accident that case marking is more often observed in languages with freer word order, and less often in languages with fixed word order. But in the case of large-scale historical language change, a multitude of considerations enter the picture.

As an example of external pressures acting on language change, one possible factor underlying diachronic word order change is the effects of second language learning. Morphology, particularly case marking, appears to be difficult for adult L2 learners to acquire. Even if their L1 language has case marking, L2 learners frequently substitute or omit case (e.g., Gürel, 2000; Papadopoulou et al., 2011; Parodi, Schwartz, & Clahsen, 2004). Using mixed-effect models to analyze a sample of 66 languages, Bentz and Winter (2013) observed that on average, languages with nominal case have around 16% L2 users in the population, while languages without case have around 44%. This raises the intriguing possibility that the difficulty of acquiring case marking for second language learners may gradually reduce the use or accuracy of case marking, resulting in the disappearance of individual case markers or even morphological case (Bentz & Winter, 2013). While based only on statistical association, this finding illustrates the importance of considering the multiple factors hypothesized to underlie language change and which are testable in the laboratory. If case marking systems are more fragile, as the above and other studies suggest, they may be the first to erode over time, particularly in cases of language contact. With the loss of case marking, other pressures may feature more prominently, driving a shift to a more fixed word order, where constituent order disambiguates the roles of participants in events. On this hypothesis, there is little reason to think that recent patterns of historical change (from SOV to SVO) are indicative of the longer term patterns in the history of language. The period for which we have good historical records is also the period over during which wide-scale trade was established and first large-scale empires were formed. These factors could presumably increase the number of late L2 learners and their social importance. While laboratory testing allows us to isolate some factors that may affect language change, other factors, as the above example illustrates, may not be testable in the laboratory, resulting in an incomplete picture of historical change.

Our results underscore the need for future work with non-manual language creation paradigms. When use of embodied gestures is possible and viable in a manual communication paradigm (events with animate agents), participants use this as a strategy. When both entities are inanimate, embodiment is not available and participants shift strategies.

The potential for iconicity offers a different set of constraints on what types of communication systems can emerge and how they may change, compared to the more linear speech stream of spoken languages. Thus, while gestural language communication paradigms have yielded important insights into shared mental construal of events and how gestural communication systems are shaped by the events to be communicated, such paradigms seem to have a limited implication for spoken language typology. Learning biases, such as the difference between child and adult learners, as well as the effect of second language learning, may also shape languages in a way that has not yet been captured in the laboratory.

In summary, we suggest that drawing a parallel between historical language change and the shifting word order preferences in gestural creation paradigms rests on several assumptions, some more probable than others. While communicative pressures, such as the presence of animate objects, may influence word order, a combination of many factors are likely behind a language's change from SOV to SVO. Indeed, despite the ubiquity of communicating about actions on other animates, SOV remains the most frequent word order observed in the world's languages. This observation suggests two different hypotheses. Perhaps SOV truly is the primordial default for spoken language and the pressures that lead to a shift away from this default are either very weak (resulting in a half-life of 60,000 years) or more recent. Or perhaps, as is suggested by the distributional patterns in the world, SOV and SVO both are equally natural, easily learnable, and cognitively transparent languages and the directionality of language change is a recent phenomenon.

Acknowledgments

We thank members of the Snedeker Laboratory for helpful feedback, and Melissa Kline and Ann Senghas for insightful conversation. This research was supported by an NSF Graduate Research Fellowship to A.K. and a Harvard BLISS Fellowship to H.L.

Notes

1. Following the conventions of the typological literature and gestural language creation experiments, we use the S, O, and V notation in a semantic sense, referring to agents, patients, and actions, respectively, of events. In our discussion of dominant or preferred word orders, we consider only transitive utterances (those with a verb and two arguments), excluding intransitive ones.
2. Note that these claims should not be treated as exhaustive. It is not the case that these proposals suggest that all languages were only SOV at one point in time, or that there was only a single factor that precipitated a shift from SOV to SVO.
3. We thank an anonymous reviewer for raising this point.

4. This estimate is based on the assumption that the universal cognitive or historical biases underlying SOV word order must have been present in all *Homo sapiens* at the point when we spread out from Africa (otherwise they would not be part of the universal substrate of human language). (If the original SOV language was in place prior to this time, the time frame only increases.) We also assume that the original SOV language was present shortly after that time.

References

Bentz, C., & Winter, B. (2013). Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change*, *3*, 1–27.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language*. New York: Wiley.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*, 547–619.

Dryer, M. S. (2005). The order of subject, object and verb. In M. Haspelmath, M. S. Dryer, D. Gil, & B. Comrie (Eds.), *The world atlas of language structures* (pp. 330–333). Oxford, UK: Oxford University Press.

Dryer, M. S. (2013). Order of Subject, Object and Verb. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online. Leipzig: Max planck institute for evolutionary anthropology*. Available at http://wals.info/chapter/81. Accessed February 8, 2017.

Gell-Mann, M., & Ruhlen, M. (2011). The origin and evolution of word order. *Proceedings of the National Academy of Sciences*, *108*, 17290–17295.

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*, 1079–1088.

Givón, T. (1979). *On understanding grammar*. New York: Academic Press.

Goldin-Meadow, S. (2003). *The resilience of language: What gesture creation in deaf children can tell us about how all children learn language*. New York: Psychology Press.

Goldin-Meadow, S., & Mylander, C. (1998). Spontaneous sign systems created by deaf children in two cultures. *Nature*, *391*, 279–281.

Goldin-Meadow, S., Özyürek, A., Sancar, B., & Mylander, C. (2008). Making language around the globe: A crosslinguistic study of homesign in the United States, China, and Turkey. In J. Guo, E. Lieven, N. Budwig, & S. Ervin-Tripp (Eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 27–39). New York: Psychology Press.

Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.), *Universals of language* (pp. 73–113). Cambridge, MA: MIT Press.

Gürel, A. (2000). Missing case inflection: Implications for second language acquisition. In C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th annual boston university conference on language development* (vol. 45, pp. 379–390). Somerville, MA: Cascadilla Press.

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

Halliday, M. A. K. (1967). Notes on transitivity and theme in English: Part 2. *Journal of Linguistics*, *3*, 199–244.

Kiparsky, P. (1996). The shift to head-initial VP in Germanic. In S. Epstein, H. Thrainsson, & S. Peters (Eds.), *Studies in comparative germanic syntax* (vol. 2, pp. 140–79). Dordrecht, the Netherlands: Kluwer.

Kroch, A. S. (2001). Syntactic change. In M. Baltin, & C. Collins (Eds.), *The handbook of contemporary syntactic theory* (pp. 699–729). Oxford, UK: Blackwell Publishers Ltd.

Lambrecht, K. (1996). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*, vol. 71. Cambridge, UK: Cambridge University Press.

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

Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, Instruments, & Computers*, *41*(3), 841–849.

MacWhinney, B. (1977). Starting points. *Language*, *53*, 152–168.

Napoli, D., & Sutton-Spence, R. (2014). Order of the major constituents in sign languages: Implications for all language. *Frontiers in Psychology*, *5*, 1–18.

Newmeyer, F. J. (2000a). *Language form and language function*. Cambridge, MA: MIT Press.

Newmeyer, F. J. (2000b). On the reconstruction of "proto-world" word order. In C. Knight, M. Studdert-Kennedy, & J. R. Hurford (Eds.), *The evolutionary emergence of language: Social function and the origins of linguistic form* (pp. 372–388). Cambridge, UK: Cambridge University Press.

Newmeyer, F. J. (2005). *Possible and probable languages: A generative perspective on linguistic typology*. New York, NY: Oxford University Press.

Osgood, C. E. (1980). *Lectures on language performance*. New York: Springer-Verlag.

Papadopoulou, D., Varlokosta, S., Spyropoulous, V., Kaili, H., Prokou, S., & Revithiadou, A. (2011). Case morphology and word order in second language Turkish: Evidence from Greek learners. *Second Language Research*, *27*, 173–205.

Parodi, T., Schwartz, B. D., & Clahsen, H. (2004). On the L2 acquisition of the morphosyntax of German nominals. *Linguistics*, *42*, 669–705.

Posse, E. S. (2000). The passive as an object foregrounding device in Early Modern English. In R. Bermudez-Otero, D. Denison, R. M. Hogg, & C. B. McCully (Eds.), *Generative theory and corpus studies: A dialogue from 10 ICEHL* (pp. 211–32). Berlin: Mouton de Gruyter.

Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223–255). New York: Academic.

Pyers, J., Senghas, A., Plançon, E., Zola, C., & Reynoso, N. (2016). *The differential role of iconicity in the creation and maintenance of an emerging sign language lexicon*. Washington, DC: Poster presented at the Linguistics Society of America Annual Meeting.

Richie, R., Yang, C., & Coppola, M. (2014). Modeling the emergence of lexicons in homesign systems. *Topics in Cognitive Science*, *6*, 183–195.

Sandler, W., Meir, I., Padden, C., & Aronoff, M. (2005). The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences*, *102*, 2661–2665.

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

Senghas, A. (2003). Intergenerational influence and ontogenetic development in the emergence of spatial grammar in Nicaraguan Sign Language. *Cognitive Development*, *18*, 511–531.

Wittenberg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. *Paper presented at Fifth International Conference on Language Resources and Evaluation (LREC)*. Paris: ELRA.