The natural order of events: How speakers of different languages represent events nonverbally

Susan Goldin-Meadow*, Wing Chee So‡, Ashı Özyürek§¶, and Carolyn Mylander*

*Department of Psychology, University of Chicago, 5730 South Woodlawn Avenue, Chicago, IL 60637; ‡Department of Psychology, National University of Singapore, Republic of Singapore 119077; §Department of Linguistics, Radboud University Nijmegen, 6525 HT Nijmegen, The Netherlands; ¶Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands; and Department of Psychology, Koc University, 34450 Istanbul, Turkey

Edited by Rochel Gelman, Rutgers, The State University of New Jersey, Piscataway, NJ, and approved May 8, 2008 (received for review November 12, 2007)

To test whether the language we speak influences our behavior even when we are not speaking, we asked speakers of four languages differing in their predominant word orders (English, Turkish, Spanish, and Chinese) to perform two nonverbal tasks: a communicative task (describing an event by using gesture without speech) and a noncommunicative task (reconstructing an event with pictures). We found that the word orders speakers used in their everyday speech did not influence their nonverbal behavior. Surprisingly, speakers of all four languages used the same order and on both nonverbal tasks. This order, actor–patient–act, is analogous to the subject–object–verb pattern found in many languages of the world and, importantly, in newly developing gestural languages. The findings provide evidence for a natural order that we impose on events when describing and reconstructing them nonverbally and exploit when constructing language anew.

gesture | language genesis | sign language | word order

Consider a woman twisting a knob. When we watch this event, we see the elements of the event (woman, twisting, knob) simultaneously. But when we talk about the event, the elements are mentioned one at a time and, in most languages, in a consistent order. For example, English, Chinese, and Spanish speakers typically use the order woman–twist–knob [actor (Ar)–act (A)–patient (P)] to describe the event; Turkish speakers use woman–knob–twist (ArPA). The way we represent events in our language might be such a powerful tool that we naturally extend it to other representational formats. We might, for example, impose our language’s ordering pattern on an event when called on to represent the event in a nonverbal format (e.g., gestures or pictures). Alternatively, the way we represent events in our language may not be easily mapped onto other formats, leaving other orderings free to emerge.

Word order is one of the earliest properties of language learned by children (1) and displays systematic variation across the languages of the world (2, 3), including sign languages (4). Moreover, for many languages, word order does not vary freely and speakers must use marked forms if they want to avoid using canonical word order (5). If the ordering rules of language are easily mapped onto other, nonverbal representations, then the order in which speakers routinely produce words for particular elements in an event might be expected to influence the order in which those elements are represented nonverbally. Consequently, speakers of different languages would use different orderings when asked to represent events in a nonverbal format (the ordering rules of their respective languages). If, however, the ordering rules of language are not easily mapped onto nonverbal representations of events, speakers of different languages would be free to use orders that differ from the canonical orders found in their respective languages; in this event, the orderings they use might, or might not, converge on a single order. To explore this question, speakers of four languages differing in their predominant word orders were given two nonverbal tasks.

Gesture Task (6). Forty adults [10 English speakers, 10 Turkish speakers, 10 Spanish speakers, and 10 Chinese (Mandarin) speakers] were asked to describe vignettes displayed on a computer (some depicting interactions between real objects and people and others depicting animated toys) by using only their hands and not their mouths. The vignettes displayed 36 different motion events [see supporting information (SI) Table S1], chosen because events of this type are ones that children talk about in the early stages of language learning (1) and thus may have a special status with respect to early language. In addition, the events are typically described by using different word orders by speakers of the languages represented in our sample: (i) 20 events typically described by using intransitive sentences, 7 in which an entity performs an action in place (girl–waves) and 13 in which an entity transports itself across space (duck–moves–to wheelbarrow); and (ii) 16 events typically described by using transitive sentences, 8 in which an entity acts on an object in place (woman–twists–knob) and 8 in which an entity transfers an object across space (girl–gives–flower–to man). To determine the predominant speech orders speakers of the four languages use to describe these particular events, participants were also asked to describe the events in speech before describing them in gesture.**

Transparency Task (7). Another 40 adults (10 speakers of each of the same four languages) were asked to reconstruct the same events by using sets of transparent pictures. A black line drawing of the entities in each event (e.g., woman, knob) and a black cartoon drawing of the action in the event (e.g., an arrowed line representing the twisting motion) were placed on separate transparencies. Participants were asked to reconstruct the event by stacking the transparencies one by one onto a peg to form a single representation (see Fig. 3). Participants were given no indication that the order in which they stacked the transparencies was the focus of the study; in fact, the background of each transparency was clear so that the final product looked the same independent of the order in which the transparencies were stacked. The task was designed to test whether speakers would extend the ordering patterns of their languages not only to the pictorial modality (where, unlike gestures, all of the elements of an event are presented simultaneously in the final product), but also to a noncommunicative situation: the experimenter made it

Author contributions: S.G.-M. and A.O. designed research; W.C.S. and C.M. performed research; S.G.-M. and W.C.S. analyzed data; and S.G.-M. and A.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

†To whom correspondence should be addressed. E-mail: sgm@uchicago.edu.

**By having participants produce speech descriptions first, we recognize that we may be biasing them to use their speech orders in their gesture descriptions. The striking finding is that they did not.

This article contains supporting information online at www.pnas.org/cgi/content/full/0710060105/DCSupplemental.

© 2008 by The National Academy of Sciences of the USA

www.pnas.org/cgi/doi/10.1073/pnas.0710060105

PNAS | July 8, 2008 | vol. 105 | no. 27 | 9163–9168
clear that she was busy with another task and not paying attention when the participants stacked the transparencies.

**Results**

To ensure that speakers of the four languages did indeed use different word orders in speech, we first examined the speech they produced to describe the vignettes. We focused on the word order in the world’s languages (9, 10): actors (Ar) (which are typically subjects, S), patients (P) (typically objects, O), acts (A) (typically verbs, V). Speakers of all four languages consistently used ArA order when describing intransitive actions both in place and crossing space (Table 1). However, speakers used different orders to describe transitive events. Following the patterns of their respective languages, English and Spanish speakers used ArAP and Turkish speakers used ArPA to describe all transitive actions, both in place and crossing space. Chinese speakers used ArAP for in-place but ArPA for crossing-space actions.

**Table 1. Speech and gesture strings produced by Turkish, English, Spanish, and Chinese speakers categorized according to their fits to predominant orders**

<table>
<thead>
<tr>
<th>Types of actions described</th>
<th>Speech strings(^1)</th>
<th>Gesture strings(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Predominant speech order</td>
<td>Proportion consistent with speech order</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SE</td>
</tr>
<tr>
<td>Actors, acts (intransitive)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-place and crossing-space actions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turkish speakers</td>
<td>Ar A</td>
<td>1.00 (0.00)</td>
</tr>
<tr>
<td>Chinese speakers</td>
<td>Ar A</td>
<td>1.00 (0.00)</td>
</tr>
<tr>
<td>English speakers</td>
<td>Ar A</td>
<td>1.00 (0.00)</td>
</tr>
<tr>
<td>Spanish speakers</td>
<td>Ar A</td>
<td>0.94 (0.03)</td>
</tr>
<tr>
<td>Actors, patients, acts (transitive)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-place actions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turkish speakers</td>
<td>Ar P A</td>
<td>0.97 (0.02)</td>
</tr>
<tr>
<td>Chinese speakers</td>
<td>Ar A P</td>
<td>0.88 (0.04)</td>
</tr>
<tr>
<td>English speakers</td>
<td>Ar A P</td>
<td>0.98 (0.01)</td>
</tr>
<tr>
<td>Spanish speakers</td>
<td>Ar A P</td>
<td>0.92 (0.05)</td>
</tr>
<tr>
<td>Crossing-space actions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turkish speakers</td>
<td>Ar P A</td>
<td>0.93 (0.04)</td>
</tr>
<tr>
<td>Chinese speakers</td>
<td>Ar A P</td>
<td>0.80 (0.06)</td>
</tr>
<tr>
<td>English speakers</td>
<td>Ar A P</td>
<td>0.88 (0.10)</td>
</tr>
<tr>
<td>Spanish speakers</td>
<td>Ar A P</td>
<td>0.86 (0.05)</td>
</tr>
</tbody>
</table>

Gestures were produced in place of speech and thus were not accompanied by any speech at all. *P < 0.0001, comparing proportion of gesture strings consistent with gesture order vs. speech order.

\(^1\)Proportions were calculated by taking the number of spoken sentences a participant produced that were consistent with the predominant speech order and dividing that number by the total number of spoken sentences the participant produced to describe the target event.

\(^2\)Proportions were calculated by taking the number of gesture strings a participant produced that were consistent with the predominant gesture order or the predominant speech order and dividing that number by the total number of gesture strings the participant produced to describe the target event. Participants did not always produce gestures for all three elements when describing transitive actions (see Table S1). When ArA strings were produced for a transitive action, we counted those strings as consistent with the predominant order for gesture and speech and thus included them in the numerator for both proportions.

\(^3\)A blank cell indicates that the predominant gesture order is identical to the predominant speech order for that language group.

\(^*\)Significant at the P < 0.0001 level.

\(^*\)Significant at the P < 0.00001 level.

\(^\dagger\)Significant at the P < 0.000001 level.

\(^\ddagger\)Significant at the P < 0.0000001 level.

\(^\S\)Significant at the P < 0.00000001 level.

\(^\S\)Significant at the P < 0.000000001 level.

\(^\S\)Significant at the P < 0.0000000001 level.

\(^\S\)Significant at the P < 0.00000000001 level.

\(^\S\)Significant at the P < 0.000000000001 level.

\(^\S\)Significant at the P < 0.0000000000001 level.

\(^\S\)Significant at the P < 0.00000000000001 level.

\(^\S\)Significant at the P < 0.000000000000001 level.

\(^\S\)Significant at the P < 0.0000000000000001 level.

\(^\S\)Significant at the P < 0.00000000000000001 level.

\(^\S\)Significant at the P < 0.000000000000000001 level.

\(^\S\)Significant at the P < 0.0000000000000000001 level.

\(^\S\)Significant at the P < 0.00000000000000000001 level.

\(^\S\)Significant at the P < 0.000000000000000000001 level.
and transitive events by using ANOVAs with one within-subjects factor (order) and one between-subjects factor (language group). We found an effect of order but not group in each analysis. Participants stacked transparencies in the ArA order significantly more often than the AAr order for intransitive actions in place and crossing space \( F(1,36) = 203.02, P < 0.000001 \), and in the ArPA order significantly more often than the other possible orders for transitive actions in place \( F(4,144) = 85.01, P < 0.000001 \) and for transitive actions crossing space \( F(5,180) = 185.71, P < 0.000001 \).

Here again we found that participants did not display the orders of their spoken languages. The order in which participants selected transparencies was the same across all four language groups, and this order was identical to the predominant gesture order in the gesture task. Across all event types, 81% (SE = 6%) of 1,423 transparency trials followed the ArPA pattern, as did 90% (SE = 3%) of 614 gesture strings.††

Recall that half of the vignettes depicted crossing-space actions and thus contained endpoints (e.g., the scaffolding in Fig. 3). Here, too, the language that participants spoke did not inevitably determine their orderings on the gesture or transparency tasks. When describing the vignettes in speech, participants followed the patterns of their languages: Turkish speakers produced words for endpoints before words for actions in 97% of their spoken utterances; Chinese, Spanish, and English speakers produced words for endpoints after words for actions in 88%, 94%, and 100% of their spoken utterances, respectively. However, participants in all four groups performed similarly on the gesture and transparency tasks. Participants tended to place gestures for endpoints at either the beginning (Turkish 25%, Chinese 40%, Spanish 26%, and English 37%) or end (58%, 48%, 56%, and 48%, respectively) of their three-gesture strings.

††There were no ordering differences in vignettes portraying real vs. toy actors. In addition, actors preceded acts whether the actors were animate or inanimate, and preceded patients whether the patients were animate or inanimate, suggesting that participants’ orders were based on the semantic roles, not animacy, of the entities involved.

Fig. 2. Proportion of speech (Left) and gesture (Right) strings produced by speakers of Turkish, Chinese, English, and Spanish to describe transitive actions that were consistent with the ArPA order. Included are both in-place and crossing-space transitive actions.

Fig. 1. Examples of ArPA gesture strings produced by speakers of all four languages. (Top) The pictures show a Spanish speaker describing the boy tilts glass vignette. (Middle) The pictures show an English speaker and a Turkish speaker describing the captain swings pail vignette; note that the English speaker (Upper Middle) conveys the captain by producing a gesture for his cap, the Turkish speaker (Lower Middle) by pointing at the still picture of the captain. (Bottom) The pictures show a Chinese speaker describing the girl covers box vignette.

The pictures show a Chinese speaker describing the captain. (girl Bottom cap, the Turkish speaker (Bottom) by pointing at the still picture of the girl.) The pictures show a Spanish speaker describing the captain. (Bottom) The pictures show an English speaker and a Turkish speaker describing the captain swings pail vignette; note that the English speaker (Upper Middle) conveys the captain by producing a gesture for his cap, the Turkish speaker (Lower Middle) by pointing at the still picture of the captain. (Bottom) The pictures show a Chinese speaker describing the girl covers box vignette.
containing endpoints and select the endpoint transparency primarily at the beginning (78%, 69%, 71%, and 83%, respectively) but also at the end (10%, 25%, 20%, and 12%, respectively) of the transparency stacks containing endpoints; the few remaining endpoints were placed in the middle of a gesture string or transparency stack. The interesting generalization is that gestures and transparency selections tended to be positioned outside of the semantic core: only 15% (SE = 5%) of gestures and 8% (SE = 2%) of transparencies for endpoints were placed between gestures or transparencies for Ar, P, and A, reinforcing the notion that events are built around these semantic elements, not endpoints. It is intriguing that languages also seem to privilege Ar, P, and A, typically encoding them as S, O, and V and relegating endpoints to the linguistic periphery as, for example, indirect (as opposed to direct) objects or objects of a preposition. This pattern may reflect another cognitive preference that languages co-opt and build on.

To summarize both the gesture and transparency tasks, we note four striking findings: (i) Participants adhered to a consistent ordering even though consistency was not demanded by either task. (ii) The ordering was the same across participants, both within and across language groups. (iii) The ordering was the same across tasks, both communicative (gesture) and non-communicative (transparency). (iv) The ordering was not necessarily the same as the ordering in the participants’ spoken language.

**Discussion**

What might account for the particular ordering we observed across language groups and tasks? On the basis of the gesture task alone, we might hypothesize that the participants arrive at the ArPA order because of communicative pressure; the ArPA order might, for example, be particularly easy for listeners to decode. But data from the transparency task weaken this argument. Actors and patients also preceded acts in the transparency task, even though the goal was not to communicate (and the medium was not gesture).

We therefore speculate that, rather than being an outgrowth of communicative efficiency or the manual modality, ArPA may reflect a natural sequencing for representing events. Entities are cognitively more basic and less relational than actions (9), which might lead participants to highlight entities involved in an action before focusing on the action itself, thus situating Ar and P before A. Moreover, there is a particularly close cognitive tie between objects and actious (10), which would link P to A, resulting in an ArPA order.

The ArPA order found in our participants’ gesture strings and transparency selections is analogous to SOV word order in spoken and signed languages. In principle, all six possible orderings of S, O, and V should be found equally often in the languages of the world. However, two orders predominate; the orders are about equally frequent and together account for ±90% of the world’s languages. SOV is one of those two orders (SVO is the other) (11, 12). In addition, although direction of change is difficult to assess over historical time, SOV has been hypothesized to predominate in the early stages of spoken (13, 14) and signed (15) languages. Even more relevant to our study,
SOV is the order currently emerging in a language created spontaneously without any apparent external influence. Al-Sayyid Bedouin Sign Language arose within the last 70 years in an isolated community with a high incidence of profound prelingual deafness. In the space of one generation, the language assumed grammatical structure, including SOV order (16). In addition, deaf children whose profound hearing losses prevent them from acquiring a spoken language and whose hearing parents have not exposed them to a conventional sign language invent their own gestures to communicate, and those gestures display a consistent OV order in both American (17) and Chinese (18) deaf children (the deaf children typically omit gestures for transitive actors, the S (10)).

If SOV is such a natural order for humans, why then aren’t all human languages SOV? Languages respond to a variety of pressures, for example, the need to be semantically clear, processed efficiently, or rhetorically interesting (19). We suggest that, initially, a developing language co-opts the ArPA order used in nonverbal representations and uses it as a default pattern, thus displaying SOV order, which may have the virtue of semantic clarity. But as a language community grows (20) and its functions become more complex (21), additional pressures may exert their influence on language form, in some cases pushing the linguistic order away from the semantically clear ArPA (SOV) order.

Our findings bear on the Sapir-Whorf hypothesis (22, 23), the hypothesis that the language we speak can affect the way we interpret the world even when we are not speaking. This hypothesis has been tested with respect to a variety of linguistic constructions with mixed results (24–29), but has not been tested with respect to word order. Our data suggest that the ordering we use when representing events in a nonverbal format is not highly susceptible to language’s influence. Rather, there appears to be a natural order that humans (regardless of the language they speak) use when asked to represent events nonverbally. Indeed, the influence may well go in the other direction; the ordering seen in our nonverbal tasks may shape language in its emerging stages.

In sum, we have shown that speakers of languages that differ in their predominant word orders do not invoke these orders when asked to describe or reconstruct events without speaking. Thus, the ordering found in a speaker’s habitual talk does not inevitably influence that speaker’s nonverbal behavior. Moreover, the ordering found in nonverbal tasks appears to be more robust than the ordering found in language; speakers of four different languages used different orders in their spoken sentences, yet all displayed the same order on two different nonverbal tasks. This order is the one found in the earliest stages of newly evolving gestural languages and thus may reflect a natural disposition that humans exploit not only when asked to represent events nonverbally, but also when creating language anew.

Materials and Methods
Data were collected in Istanbul for Turkish speakers, in Beijing for Chinese speakers, and in Madrid for Spanish speakers. Participants were drawn from urban universities in each city for both gesture and verbal tasks. This order is the one found in the earliest stages of different languages used different orders in their spoken sentences and thus may reflect a natural order that humans (regardless of the language they speak) use when asked to represent events nonverbally.

Verbal orders have been tested with respect to a variety of linguistic constructions and nonverbal tasks. This hypothesis has been tested with respect to a variety of linguistic constructions with mixed results (24–29), but has not been tested with respect to word order. Our data suggest that the ordering we use when representing events in a nonverbal format is not highly susceptible to language’s influence. Rather, there appears to be a natural order that humans (regardless of the language they speak) use when asked to represent events nonverbally. Indeed, the influence may well go in the other direction; the ordering seen in our nonverbal tasks may shape language in its emerging stages.

In sum, we have shown that speakers of languages that differ in their predominant word orders do not invoke these orders when asked to describe or reconstruct events without speaking. Thus, the ordering found in a speaker’s habitual talk does not inevitably influence that speaker’s nonverbal behavior. Moreover, the ordering found in nonverbal tasks appears to be more robust than the ordering found in language; speakers of four different languages used different orders in their spoken sentences, yet all displayed the same order on two different nonverbal tasks. This order is the one found in the earliest stages of newly evolving gestural languages and thus may reflect a natural disposition that humans exploit not only when asked to represent events nonverbally, but also when creating language anew.
Transparency Task: Protocol and Coding. Participants were told that they would be watching a series of video clips and that, after each vignette, they were to reconstruct the event by putting a set of transparencies on a peg one at a time. They were told that if they needed to see the vignette again, they could click on the repeat button on the computer. After two practice trials (the experimenter did not model the task for the participant), the experimenter played the first vignette on the computer and then placed the transparencies for that vignette on the table in a triangular configuration, beginning at the participant’s right and ending at his or her left. For example, the four transparencies presented to the participant for the man carries chicken to scaffolding vignette were placed on the table in the following order (Fig. 3): 1) arrow denoting the trajectory of the moving action, 2) chicken, 3) scaffolding, and 4) man. Transparencies were laid down in the same order for each participant, and orders were randomized over the 36 vignettes.

Participants were discouraged from talking during the study. To encourage them to treat the task as noncommunicative, the experimenter occupied herself with another task and did not pay attention as participants picked up the transparencies. When participants finished stacking the transparencies on the peg, they were notified by the experimenter and then started the next vignette on the computer.

Transcribers watched the videotapes of each session and recorded the order in which the participant placed the transparencies on the peg for each vignette.

ACKNOWLEDGMENTS. We thank S. Beilock, L. Gleitman, P. Hagoort, P. Indefrey, E. Newport, H. Nusbaum, S. Levine, and T. Regier for comments and A. S. Acuna, T. Asano, O. E. Demir, T. Göksun, Z. Johnson, W. Xue, and E. Yalabik for help with data collection and coding. This research was supported by National Institute on Deafness and Other Communication Disorders Grant R01 DC00491 (to S.G.-M.).