

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

Susan Goldin-Meadow*[†], Wing Chee So[‡], Aslı 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.Ö. 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.Ö. 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

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

| Types of actions described | Speech strings [†] | | | | Gesture strings [‡] | | | | | |
|-------------------------------------|-----------------------------|----|---|--------|------------------------------|----|--|--------|--|--------|
| | Predominant speech order | | Proportion consistent with speech order | | Predominant gesture order | | Proportion consistent with gesture order | | Proportion consistent with speech order [§] | |
| | Mean | SE | Mean | SE | Mean | SE | Mean | SE | Mean | SE |
| Actors, acts (intransitive) | | | | | | | | | | |
| In-place and crossing-space actions | | | | | | | | | | |
| Turkish speakers | Ar | A | 1.00 | (0.00) | Ar | A | 0.85 | (0.10) | | |
| Chinese speakers | Ar | A | 1.00 | (0.00) | Ar | A | 0.98 | (0.01) | | |
| English speakers | Ar | A | 1.00 | (0.00) | Ar | A | 0.99 | (0.01) | | |
| Spanish speakers | Ar | A | 0.94 | (0.03) | Ar | A | 0.97 | (0.02) | | |
| Actors, patients, acts (transitive) | | | | | | | | | | |
| In-place actions | | | | | | | | | | |
| Turkish speakers | Ar | P | A | 0.97 | (0.02) | Ar | P | A | 1.00 | (0.00) |
| Chinese speakers | Ar | A | P | 0.88 | (0.04) | Ar | P | A | 0.84 | (0.06) |
| English speakers | Ar | A | P | 0.98 | (0.01) | Ar | P | A | 0.90 | (0.03) |
| Spanish speakers | Ar | A | P | 0.92 | (0.05) | Ar | P | A | 0.86 | (0.05) |
| Crossing-space actions | | | | | | | | | | |
| Turkish speakers | Ar | P | A | 0.93 | (0.04) | Ar | P | A | 0.69 | (0.13) |
| Chinese speakers | Ar | P | A | 0.80 | (0.06) | Ar | P | A | 0.90 | (0.08) |
| English speakers | Ar | A | P | 0.88 | (0.10) | Ar | P | A | 0.78 | (0.08) |
| Spanish speakers | Ar | A | P | 0.86 | (0.05) | Ar | P | A | 0.87 | (0.04) |

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.

[†]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.

[‡]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.

[§]A blank cell indicates that the predominant gesture order is identical to the predominant speech order for that language group.

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 position of semantic elements traditionally used to characterize 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 transitive actions.**

Participants used the same order, ArA, the order found in each of their spoken languages, in the gesture strings they produced to describe intransitive actions in place and crossing space (Table 1). But participants also used a single gesture order for transitive actions, even though their spoken languages used different orders to convey actions of this type. The predominant gesture order was ArPA, which was identical to the predominant speech order for in-place and crossing-space actions in Turkish and for crossing-space actions in Chinese, but different from the

predominant speech order for both types of actions in English and Spanish and for in-place actions in Chinese. We analyzed in-place and crossing-space actions separately by using ANOVAs with one within-subjects factor (order) and one between-subjects factor (language group). We found significant effects for order but not group in each analysis: gesture strings were significantly more likely to display the ArPA order than the ArAP order found in spoken Chinese, English, and Spanish for in-place actions [$F(1,26) = 63.18, P < 0.00001$] and in spoken English and Spanish for crossing-space actions [$F(1,16) = 49.42, P < 0.00001$]. Fig. 1 presents examples of ArPA gesture strings.

Fig. 2 presents the proportion of all gesture and speech strings describing transitive actions that were consistent with the ArPA order. We analyzed the data in Fig. 2 by using an ANOVA with one within-subjects factor (modality: gesture vs. speech) and one between-subjects factor (language group). We found an effect of modality. Proportions were significantly different for gesture vs. speech [$F(1,35) = 235.65, P < 0.00001$] and, as expected, the effect interacted with language group [$F(3,35) = 32.00, P < 0.00001$]. Gesture was significantly different from speech for English, Spanish, and Chinese ($P < 0.00004$, Scheffé), but not for Turkish. Importantly, there were no significant differences between any pairings of the four language groups for gesture ($P > 0.74$). Thus, participants did not display the order found in their spoken language in their gestures. Instead, the gestures all followed the same ArPA order.

A priori we might have guessed that gesturers would begin a string by producing a gesture for the action, as the action frames the event and establishes the roles that other elements can assume. Indeed, the glass and box gestures in Fig. 1 are very similar in form; the fact that these two gestures represent

**According to ref. 31, Chinese was originally an SOV language and became SVO; it is currently in the process of moving back to SOV and thus displays both orders.

SOV is the order currently emerging in a language created spontaneously without any apparent external influence. Al-Sayid 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, in Chicago for English speakers, and in Madrid for Spanish speakers. Participants were drawn from urban universities in each city for both gesture and transparency tasks. Each participant was initially given consent forms and a language interview. Only monolingual speakers were included in the study. None of the participants was conversant in a conventional sign language.

Participants in both tasks were shown 36 vignettes on a computer. A quarter of the vignettes depicted interactions between real people and objects; the remaining vignettes were animations involving toys representing objects and people. Table S1 lists the events depicted in the vignettes, the types of objects playing the actor, patient, or endpoint roles in each vignette, and whether or not the vignette depicted real objects and people or toys.

Gesture Task. Protocol. The participant and experimenter were positioned in a natural conversational grouping with a laptop computer between them. The entire session was videotaped, and the camera was positioned so that it had

a good view of both the participant and the computer screen. Participants were told that they would see a series of short videotaped vignettes and, after each, were to tell the experimenter what happened in the vignette. For the participants to get a better at look at the entities pictured in the vignettes, still pictures of the initial scene of the event including all of the entities involved in the event (actors, patients, endpoints) were provided for each vignette; participants occasionally referred to the pictures when describing the vignettes. Two practice vignettes were run before the participant began the set of 36 vignettes. After describing all of the vignettes in speech, participants were told that they would see the same vignettes again but, this time, they were to tell the experimenter what happened using only their hands and not speech. The still pictures were shown during the gesture alone descriptions and participants occasionally pointed at the picture to refer to one of the elements (see Fig. 2).

Coding and analysis. Speech was transcribed and coded by native speakers of each language. Gesture was described in terms of hand shape, palm orientation, motion, placement (e.g., neutral space at chest level, on the body, near an object), articulator (e.g., right hand, left hand, head), and size of motion (e.g., <2 inches, 2–5 inches, >5 inches). Gestures were divided into strings by using motoric criteria; string breaks were coded when participants relaxed their hands between gestures (30). Gestures were classified as either pointing or iconic gestures and given a meaning gloss. Pointing gestures were rarely used but when they were, they either indicated an entity on the computer screen or the still pictures (e.g., point at the man on screen, glossed as *man*) or an entity in the room that was similar to or had an interpretable relation to one in the vignette (e.g., point at self when referring to the girl in the vignette, glossed as *girl*; point behind self when referring to the chicken who stood behind the captain in the vignette, glossed as *chicken*). Iconic gestures were pantomimes used to represent either the action in the vignette (e.g., two fists moved away from the chest in a line, glossed as *push* in the "man pushes garbage can to motorcycle man" vignette) or an entity (e.g., two fists held at chest level rotated as if revving a motorcycle, glossed as *motorcycle man* in this same vignette). Other types of gestures (e.g., beats, nods) were transcribed but did not represent vignette actions or entities and thus were not included in the analyses. Reliability for gesture coding ranged between 87% and 91% agreement between coders depending on the category.

We analyzed only those strings that described the target event, i.e., strings containing a word or gesture for the action displayed in the vignette. In addition to containing a word or gesture for the action, the string had to include at least one other word or gesture representing an entity playing an actor or patient role in the action. For example, a *cap* gesture referring to the captain (actor) in the *captain swings pail* vignette produced in sequence with a *swing* gesture (action) would be included in the analyses; however, either gesture produced on its own would not. Moreover, if a string contained a gesture for the captain (actor) and one for the pail (patient), but no gesture for the swinging motion (act), the string would not be included in the analyses simply because, without the *swing* gesture, we could not be certain that the participant meant to be referring to the captain and the pail in their roles as actor and patient, respectively. When describing events in speech, participants typically mentioned all of the relevant elements in the event: two elements (actor, act) in intransitive actions and three (actor, act, patient) in transitive actions. When describing events in gesture alone, participants often omitted elements; they produced 501 gesture strings containing two relevant elements and 113 containing three, all of which were included in the analyses. We classified strings that did not have gestures for the full complement of elements according to the elements that were present in gesture. Thus, for gesture strings conveying transitive actions, we classified three types of strings as consistent with an ArPA pattern (ArA, PA, ArPA) and seven as inconsistent (AAr, AP, ArAP, AArP, APAr, PArA, PAAr). In addition, we conducted separate analyses of gesture strings containing endpoints. Participants produced 97 strings containing gestures for the endpoint and action and 121 containing gestures for the endpoint, action, and one or two other elements (actor and/or patient).

Occasionally participants produced strings in speech or gesture describing an action that did not match the intended action displayed in the vignette (e.g., "the garbage can moved to the motorcycle man" rather than "the man pushed the garbage can to the motorcycle man"). In those cases, we reclassified the string to fit the participant's interpretation and analyzed it along with strings conveying the action it described (in this case, as an intransitive action rather than a transitive action). Degrees of freedom vary for statistical analyses of some of the events because participants who produced the reclassified strings did not then have a data point for the originally intended action. If reclassified strings are omitted from the analyses, the results are unchanged. Proportions were submitted to an arcsine transformation before statistical analysis.

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 alerted the experimenter who then collected them and 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.).

- Brown R (1973) *A First Language* (Harvard Univ Press, Cambridge, MA).
- Greenberg JH (1963) Some universals of grammar with particular reference to the order of meaningful elements. *Universals of Language*, ed Greenberg JH (MIT Press, Cambridge, MA), pp 73–113.
- Haspelmath M, Dryer MS, Gil D, Comrie B (2005) *The World Atlas of Language Structures* (Oxford Univ Press, Oxford).
- Perniss P, Pfau R, Steinbach M (2007) Can't you see the difference? Sources of variation in sign language structure. *Visible Variation: Cross-Linguistic Studies on Sign Language Structure*, eds Perniss P, Pfau R, Steinbach M (Mouton, Berlin), pp 1–34.
- Steele S (1978) Word order variation: A typological study. *Universals of Human Language*, eds Greenberg JH, Ferguson CA, Moravcsik EA (Stanford Univ Press, Stanford, CA), pp 585–623.
- Goldin-Meadow S, McNeill D, Singleton J (1996) Silence is liberating: Removing the handcuffs on grammatical expression in the manual modality. *Psych Rev* 103:34–55.
- Gershkoff-Stowe L, Goldin-Meadow S (2002) Is there a natural order for expressing semantic relations? *Cognit Psychol* 45:375–412.
- Liddell S (1980) *American Sign Language Syntax* (Mouton, The Hague, The Netherlands).
- Gentner D, Boroditsky L (2001) Individuation, relativity and early word learning. *Language Acquisition and Conceptual Development*, eds Bowerman M, Levinson SC (Cambridge Univ Press, New York), pp 215–256.
- Goldin-Meadow S (2003) *Resilience of Language* (Psychology Press, New York).
- Baker MC (2001) *The Atoms of Language* (Basic Books, New York).
- Dryer M (2005) Order of subject, object and verb. *The World Atlas of Language Structures*, eds Haspelmath M, Dryer MS, Gil D, Comrie B (Oxford Univ Press, Oxford), pp 330–333.
- Newmeyer FJ (2000) On the reconstruction of "proto-world" word order. *The Evolutionary Emergence of Language*, eds Knight C, Studdert-Kennedy M, Hurford JR (Cambridge Univ Press, New York), pp 372–388.
- Givón T (1979) *On Understanding Grammar* (Academic, New York).
- Fisher S (1975) Influences on word order change in American Sign Language. *Word Order and Word Order Change*, ed Li CN (University of Texas, Austin), pp 1–25.
- Sandler W, Meir I, Padden C, Aronoff M (2005) The emergence of grammar: Systematic structure in a new language. *Proc Natl Acad Sci USA* 102:2661–2665.
- Goldin-Meadow S, Feldman H (1977) The development of language-like communication without a language model. *Science* 197:401–403.
- Goldin-Meadow S, Mylander C (1998) Spontaneous sign systems created by deaf children in two cultures. *Nature* 391:279–281.
- Slobin DI (1977) Language change in childhood and history. *Language Learning and Thought*, ed Macnamara J (Academic, New York), pp 185–214.
- Senghas A, Kita S, Ozyurek A (2004) Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science* 305:1779–1782.
- Sankoff G, Laberge S (1973) On the acquisition of native speakers by a language. *Kivung* 6:32–47.
- Sapir E (1921) *Language* (Harcourt, Brace, & World, New York).
- Whorf BL (1956) *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, ed Carroll JB (MIT Press, Cambridge, MA).
- Gumperz JJ, Levinson, SC, eds (1996) *Rethinking Linguistic Relativity* (Cambridge Univ Press, New York).
- Pederson E, et al. (1998) Semantic typology and spatial conceptualization. *Language* 74:557–589.
- Gentner D, Goldin-Meadow S, eds (2003) *Language in Mind: Advances in the Study of Language and Thought* (MIT Press, Cambridge, MA).
- Lucy J (1992) *Grammatical Categories and Cognition* (Cambridge Univ Press, New York).
- Papafragou A, Massey C, Gleitman L (2002) Shake, rattle, 'n' roll: The representation of motion in language and cognition. *Cognition* 84:189–219.
- Gilbert A, Regier T, Kay P, Ivry R (2006) Whorf hypothesis is supported in the right visual field but not the left. *Proc Natl Acad Sci USA* 103:489–494.
- Goldin-Meadow S, Mylander C (1984) Gestural communication in deaf children. *Monogr Soc Res Child Dev* 49:1–151.
- Li CN, Thompson SA (1981) *Mandarin Chinese: A Functional Reference Grammar* (University of California Press, Berkeley).