Environmental constraints shaping constituent order in emerging communication systems: Structural iconicity, interactive alignment and conventionalization

Peer Christensen a,c,⇑, Riccardo Fusaroli b,c, Kristian Tylén b,c

a Centre for Languages and Literature, Lund University, Helgonabacken 12, 221 00 Lund, Sweden  
b Center for Semiotics, Department for Aesthetics and Communication, Aarhus University, Jens Chr. Skous Vej 2, 8000 Aarhus, Denmark  
c The Interacting Minds Centre, Aarhus University, Jens Chr. Skous Vej 4, 8000 Aarhus, Denmark

A R T I C L E   I N F O

Article history:
Received 4 August 2014  
Revised 23 July 2015  
Accepted 6 September 2015

Keywords:
Structural iconicity  
Interactive alignment  
Conventionalization  
Gesture  
Word order

A B S T R A C T

Where does linguistic structure come from? Recent gesture elicitation studies have indicated that constituent order (corresponding to for instance subject–verb–object, or SVO in English) may be heavily influenced by human cognitive biases constraining gesture production and transmission. Here we explore the alternative hypothesis that syntactic patterns are motivated by multiple environmental and social–interactional constraints that are external to the cognitive domain. In three experiments, we systematically investigate different motivations for structure in the gestural communication of simple transitive events. The first experiment indicates that, if participants communicate about different types of events, manipulation events (e.g. someone throwing a cake) and construction events (e.g. someone baking a cake), they spontaneously and systematically produce different constituent orders, SOV and SVO respectively, thus following the principle of structural iconicity. The second experiment shows that participants’ choice of constituent order is also reliably influenced by social–interactional forces of interactive alignment, that is, the tendency to re-use an interlocutor’s previous choice of constituent order, thus potentially overriding affordances for iconicity. Lastly, the third experiment finds that the relative frequency distribution of referent event types motivates the stabilization and conventionalization of a single constituent order for the communication of different types of events. Together, our results demonstrate that constituent order in emerging gestural communication systems is shaped and stabilized in response to multiple external environmental and social factors: structural iconicity, interactive alignment and distributional frequency.

© 2015 Elsevier B.V. All rights reserved.

“…in the syntax of every language there are logical icons of the kind that are aided by conventional rules…”  
[(C.S. Peirce, 1940:106)]

1. Introduction

Language structure is a highly complex phenomenon evolving in response to various potentially interacting pressures at several time scales, from online interaction to phylogenetic evolution (Beckner et al., 2009; Rączaszek-Leonardi, 2010). Consequently, it is a challenging task to reconstruct the evolutionary trajectories of existing languages and the forces that have shaped them (Tylén, Fusaroli, Bundgaard, & Østergaard, 2013). Prevalent approaches in the language sciences have pointed to a variety of motivations for linguistic structure. For instance, it has been suggested that syntactic structures are innate and genetically determined (Hauser, Chomsky, & Fitch, 2002; Nowak, Komarova, & Niyogi, 2001; Pinker & Bloom, 1990). Others have argued that linguistic structures are motivated by latent internal cognitive biases gradually amplified through iterated cultural transmissions: structures that are easier for human cognitive systems to learn and use are selected for and thus increasingly propagated through cultural history (Brighton, Smith, & Kirby, 2005; Christiansen & Chater, 2008; Deacon, 1997). Yet other approaches emphasize inherent semantic relations: for instance it is argued that subjects and objects are semantically primary and therefore tend to syntactically precede actions (Goldin-Meadow, So, Özyürek, & Mylander, 2008; Hall, Ferreira, & Mayberry, 2014). Despite different starting
points, these theories all assume that linguistic structures originate independently of the referent situations they encode and of their contextually and communicatively situated use. Rather than being an independent and closed system, we argue that language structure is intimately intertwined with its social and functional role in human life (Evans & Levinson, 2009; Tomasello, 2008). Language is first and foremost used to coordinate joint action and to communicate about the world: we share experiences, coordinate action, instruct each other, tell stories, gossip, make declarations and maintain social relations (Clark, 1996; Fusaroli, Gangopadhyay, & Tytelen, 2014; Tytelen, Weed, Wallentin, Roepstorff, & Frith, 2010). Consequently, language is continuously shaped by external constraints, such as structure in the external world as well as social dynamics (Bergmann, Dale, & Lupyan, 2013; Dale, Fusaroli, Duran, & Richardson, 2013; Fusaroli & Tytelen, 2012; Tytelen et al., 2013). Adopting such functional perspectives suggests a more articulated investigation of the way in which different aspects of language use provide resources for and pressures on evolving linguistic structure (Croft, 2001). In this paper, we investigate a plurality of environmental factors shaping constituent order (word order), a key component of linguistic structure.

Many languages, including English and Danish, have fixed constituent orders encoding participant roles in verbalizations of transitive events. Some events have only one plausible interpretation regarding who is performing an action and who, or what, is affected by it. This is the case, for instance, in the English sentence "John eats cake". The functional role of constituent order is more apparent when considering semantically ambiguous events, as in "John hits Mary". Verb-final languages (including Turkish and Japanese) often have case marking systems, which aid speakers in attributing participant roles (Greenberg, 1963: Universal 41; Bentz & Christiansen, 2013). However, in many other languages, such as English and Danish, the participant roles are often only decipherable based on constituent order. Thus, in the above case, the unmarked and fixed subject–verb–object (SVO) constituent order facilitates the correct interpretation of the ambiguous events. Together the SOV and SVO constituent orders account for 89% of the 1188 languages included in a survey featured in the World Atlas of Language Structures (Dryer, 2011). Given a total of six possible combinations of subject, verb and object orders, a particularly interesting and persistent question is why SOV and SVO are more common than other word orders. Moreover, there is converging evidence from different strands of linguistic research suggesting that the SOV order is predominant in less conventionalized languages and emerging sign systems (Napoli & Sutton-Spence, 2014). Though only few emerging sign languages have been studied, findings seem to support a unique status for SOV order. For instance, SOV is the dominant constituent order found in the Al-Sayyid bedouin sign language, a language surrounded by older spoken languages displaying SVO order (Sandler, Meir, Padden, & Aronoff, 2005). Another line of evidence indicates that SOV order is predominant in gestural communication systems spontaneously invented in early childhood by deaf children with hearing parents prior to exposure to conventionalized signed or spoken linguistic input (Goldin-Meadow & Mylander, 1998). More recently, researchers have developed inventive experimental approaches to investigate the cognitive underpinnings of novel sign systems. These studies used nonverbal gesture elicitation tasks to demonstrate that hearing adult non-signers, when asked to represent non-ambiguous transitive events (i.e., human agents performing actions on objects) using only gesture, have a strong preference for the SOV order, regardless of their linguistic backgrounds (Goldin-Meadow et al., 2008). The authors explain their observations by reference to inherent semantic relations: Entities (such as agents and patients) are argued to be cognitively more basic and less relational than actions, which might lead participants to specify entities before the more abstract actions. The results have subsequently been replicated by Langus and Nespov (2010) with speakers of Italian (SVO) and Turkish (SOV). Despite differences in the proposed underlying mechanisms, these findings have been interpreted as evidence for a universal, internal cognitive predisposition to conceptualize transitive events according to a specific order analogous to SOV structure in natural language.

More recently, however, Schouwstra and de Swart (2014) have extended these investigations showing that intentional events – a class comprising four distinct event subtypes (Forbes, 2010) – can revert this pattern and consistently motivate SVO gesture strings. Intentional events include unobservable mental events such as “thinking of x” or “wanting x”. Again, the authors interpret this as supporting the semantic origins of constituent order: Since direct objects in intentional events (e.g., “the pirate is thinking about a guitar”) are more abstract and relational, as they do not have a real extension in space, they are preceded by actions. Other studies have shown tendencies of participants to switch to SVO when gesturing about so-called semantically reversible events (Gibson et al., 2013; Hall, Mayberry, & Ferreira, 2013; Meir, Sandler, Padden, & Aronoff, 2010). Contrary to irreversible events, semantically reversible events have human agents and patients (e.g., “the baker hit the ballerina”), which potentially render their respective roles ambiguous. It has thus been speculated that SVO constituent order helps disambiguating agents and patients by maximally separating them by the verb phrase (see Hall et al., 2013 for a discussion).

Previous observations of the strong tendency for participants to produce gesture strings with SOV order seem indisputable and robust. However, we ask whether the proposed underlying mechanisms account exhaustively for the constituent order in gestural representations of transitive events. We suggest that additional factors may influence gesture order. In particular, we hypothesize that environmental and social–interactional factors not considered in previous studies might have a strong influence on gesture order. In the following, we report three experimental studies in which we systematically explore (i) whether constituent order is motivated by the structural organization of the real world referent events they denote, (ii) whether constituent order is shared through the human propensity to imitate and align linguistic behaviors, and (iii) whether, through frequency of exposure and use, these behaviors eventually conventionalize to create stable, consistent and cognitively economical patterns. Crucially, these are all environmental factors, that is, not intrinsic to language or internal cognitive systems, which may significantly contribute to the shaping of linguistic structure.

1.1. Environmental constraint 1: structural iconicity

Ferdinand de Saussure famously argued that linguistic signs are by definition arbitrary, that is, related to their referents by mere convention (De Saussure, 1972), a position that is still widely represented in linguistics and psychology (Levett, Roelofs, & Meyer, 1999; Nielsen & Rendall, 2011). However, many observations challenge the generality of this claim (Fischer & Nannya, 2001; Flumini, Ranzini, & Borghi, 2014; Perissi, Thompson, & Vigliocco, 2010). For instance, studies on sound symbolism have shown that many linguistic sound-meaning mappings are in fact non-arbitrarily motivated. When presented with antonyms in Thai, Kanarese, and Yoruba languages, English speaking participants were found to perform significantly above chance in mapping their referents, indicating that phonetic qualities of words provide cues revelatory of their meanings (Slobin, 1988). Similar observations have been made in a number of experiments requiring participants to evolve new communication systems online in response to joint
collaborative tasks (Fay, Garrod, & Swoboda, 2010; Galantucci, 2005; Garrod, Fay, Lee, Oberlander, & MacLeod, 2007). Again, participants were found to reliably employ strategies based on iconic representations, that is, they developed signs bearing resemblance to their referents, arguably to facilitate their addressees’ comprehension. While the evidence for iconic mappings between form and meaning in the lexical domain is widespread, less attention has been directed at iconic motivations in syntax (for an exception, see e.g. Haiman, 1985). Structural iconicity is a particular type of iconicity in which the structure of events or relations between referents is replicated in the syntax of a spoken or signed utterance. It can be defined as a non-arbitrary, motivated relationship between form and meaning, which is established when the arrangement of individual signs mirrors actual properties of the relations between their referents, i.e. in transitive events. Indeed, in some languages, the relative order between actual sequential events in the world is grammaticalized and reflected in speech (Itkonen, 2005). A well-known example is given by the famous quote “veni, vidi, vic” (I came, I saw, I conquered) from Julius Caesar’s letter to the Roman senate, which clearly illustrates how the sequential order of events is preserved in language, but the same kind of iconic relationship between language and the world is also evident in Mandarin Chinese (Tai, 1985). Beyond simple relations of resemblance, structural iconicity also covers more diagrammatic mapping relations (Fauconnier & Turner, 2002; Stjernfelt, 2007; Tyllén et al., in press). For instance, an apparently atemporal dependency relation (x depends on y) can be represented by a temporal relation (x precedes y). This is likely to be reflected in constituent orders as well. The constituent orders found in many of the world’s languages, underlying the linguistic expression of transitive events, might thus not be coincidental or driven by language- and cognition-internal factors alone. Constituent order might also be motivated by structurally iconic relations to the referent events.

1.2. Environmental constraint 2: interactive alignment

By itself, structural iconicity would lead speakers to rely exclusively on structural properties of the referent scene as motivation for representational structure in communicative exchanges posing questions for the many apparently more arbitrary relations also found in syntax. However, interactive alignment has been shown to be an important mechanism underlying online choices of structure in communicative utterances (Fusaroli & Tyllén, 2012; Pickering & Garrod, 2004b). Interactive alignment is the spontaneous propensity of interlocutors to flexibly adapt to each other in the course of conversations thus displaying increasing similarity in their way of speaking and referring to communicational topics. Such adaptations have been observed in many interactional behaviors (Fusaroli, Konvalinka, & Wallot, 2014): from subtle bodily sway (Shockley, Santana, & Fowler, 2003), to speech rate, utterance length and phonetic profile (Fusaroli & Tyllén, in press; Giles, Coupland, & Coupland, 1991), lexical (Fusaroli et al., 2012), and conceptual alignment (Angus, Watson, Smith, Gallois, & Wiles, 2012; Garrod & Anderson, 1987; Garrod & Doherty, 1994). Likewise, a number of studies show that interlocutors tend to align on their use of syntactic constructions beyond particular tokens of referent events: If a speaker uses a double object construction (“the pirate gives the chef an apple”) to refer to a ditransitive scene, there is a relatively higher probability that her interlocutor will spontaneously use the same construction to describe analogous but not identical scenes, even though the prepositional object construction (“the pirate gives an apple to the chef”) is an equally acceptable alternative (Branigan, Pickering, McLean, & Cleland, 2007; Branigan, Pickering, Stewart, & McLean, 2000; Hopkins, Vuill, & Keller, 2015; Reitter & Moore, 2014). Rather than purely relying on the referent event (structural iconicity), speakers widely rely on the linguistic structures offered by their interlocutor. The pressure for interactive alignment is argued to relate to the cognitive economy of communication itself: the alignment of linguistic representations is thought to facilitate the sharing of situation models (mutual understanding) and establishes parsimony between linguistic production and comprehension making dialogical interaction “easy” (Ferreira & Bock, 2006; Pickering & Garrod, 2009, 2013).

Albeit distinct constraints, structural iconicity and alignment can reinforce each other in shaping the structure of communication systems. This is the case when interlocutors repeatedly select and refer to the same aspect of their shared environment, reciprocally reinforcing structural iconicity through alignment. However, the two constraints may also compete: If interlocutors repeatedly encounter very different referent events, the inclination to imitate the structure of the referent and to imitate the interlocutor might be in conflict, possibly reducing the relative influence of one or both motivational pressures.

1.3. Environmental constraint 3: conventionalization

Importantly, structural iconicity and interactive alignment would potentially afford the evolution of a wide range of context-specific and highly variegated structural forms. Although interactive alignment motivates the spread and sharing of forms and representations, it cannot alone account for the gradual conventionalization of general communicative routines consistently observed in studies on emergent communication systems and for the stability observed in most well-established languages (Clark & Wilkes-Gibbs, 1986; Fay, Arbib, & Garrod, 2013; Galantucci & Garrod, 2010; Mills, 2014).

Again, we would argue for the importance of communicative constraints: for a communication system to be optimally functional and efficient, the encoding of relevant meaning differences is not enough. An efficient communication system should also be easy to remember, produce and comprehend (Fay, Garrod, & Roberts, 2008; Kirby, Cornish, & Smith, 2008). Having to master competing linguistic structures would require more cognitive resources. There could thus be strong motivations for stabilizing a single rather than several co-existing forms to refer to tokens of a set of related events (Kirby et al., 2008). But do all forms have an equal chance of becoming conventionalized? Other environmental factors might have an impact on which forms eventually outcompete others and stabilize. In some cases it might be a question of which form is more salient due to functional or sheer frequency factors. For instance, language learners have been shown to be sensitive to distributional frequencies of lexical and grammatical categories in their linguistic input (Reeder, Newport, & Aslin, 2013; Wonnacott, Newport, & Tanenhaus, 2008). Furthermore, languages relying solely on an absolute spatial frame of reference (e.g. north, south, east and west) tend to be associated with speech communities living in rural and open-terrain environments offering stable landmarks such as hilltops and rivers for these cardinal directions, while relative frames of reference (right, left, next to etc.) are more consistently found in languages spoken in areas with dense forests or urban environments (Levinson, 1996, 2003; Majid, Bowerman, Kita, Haun, & Levinson, 2004). It has also been found that the regional level of UV light can have implications for color categories found in the languages of the corresponding regions (Lindsey & Brown, 2002; Piewczyński et al., 2014). Together these communicational and environmental motivations continuously shape communication systems toward parsimony and elimination of structural redundancy (Tomaseello, 1999, 2008). For instance, most of the world’s languages have a single basic constituent order with which speakers predominantly communicate transitive events (Dryer, 2011; Greenberg, 1984).
Although such generalizations seem to potentially work against the general principles of iconicity suggested in previous sections, they facilitate communication by simplifying the procedures or rules for generating and comprehending utterances: It is cognitively and communicatively more economic to rely on a limited repertoire of general principles than a larger repertoire of highly contextualized rules.

1.4. Investigating the motivations for constituent order in a gestural communication task

In order to systematically address the impact and interaction of multiple environmental pressures on the development of new communication systems, we conducted a series of three experiments, where pairs of participants communicated about simple transitive events using only gesture. The first study is contingent on the very general observation that the world has structural properties available as resources for human communication. The study thus investigates how structure intrinsic to referent events in the stimuli may, by itself, shape the manner in which the stimuli are communicated. In contrast to the earlier emphasis on internal semantic relations (Goldin-Meadow et al., 2008), we suggest that the choice of constituent order might reflect the structural organization of the referent events themselves, through the principle of structural iconicity. In order to test this prediction, we contrasted simple transitive events of the type originally used by Goldin-Meadow et al. (2008) with a structurally different type of transitive event. The former type can be characterized as object manipulation events (e.g., “the doctor eats the cake”). In these events, the referents assuming agent and patient roles must be physically co-present before the action can be purposefully performed. In other words, in an object manipulation event, the patient logically precedes the action being executed: obviously, one cannot manipulate or act upon an object, which is not physically or symbolically already present. By contrast, in a different type of transitive event, which we henceforth call object construction events, agents perform actions that cause objects to come into existence. This type of event can be exemplified by sentences like “the doctor bakes a cake”. In these cases, actions precede objects that, in turn, are dependent on the performed actions. We predict that the stimulus events will motivate different constituent orders following the principle of structural iconicity: Participants will produce SOV gesture strings for object manipulation events and SVO for object construction events. Notice that although our category of construction events corresponds to one of the subcategories of Schouwstra and de Swart’s intensional events (2014) the contrastive manipulation in this experiment is motivated from the quite different perspective of structural iconicity.

In experiment 2, we investigate the effect of interactive alignment on the choice of constituent order. If interlocutors are inclined to map the structure of the stimulus event, and at the same time are primed to imitate the constituent order used by their partner, what will happen when these are in conflict? We predict that in such situations, the pressures will be competing possibly resulting in weakening any prior effects of structural iconicity.

In experiment 3, we investigate one of the possible pressures leading to the stabilization and conventionalization of a single constituent order for both manipulation and construction events. We predict that if the two types of stimulus events differ in frequency so that participants encounter one type much more frequently than the other, they will, over repeated trials, be inclined to generalize the constituent order used for the more frequent event type to the less frequent one, thus pointing toward conventionalization of one general constituent order.

Lastly, we examine how these different environmental pressures might contribute to and interact with each other in the evolution of communication systems for talking about transitive events.

2. Experiment 1: the impact of structural iconicity

2.1. Materials and methods

2.1.1. Participants

25 pairs of participants (n = 50, 13 m/37 f, mean age 23.9, SD 2.8) participated in the experiment in return for monetary compensation. Participants were recruited among students at Aarhus University. Pair members knew each other in advance. All participants were native speakers of Danish, a language with fixed SVO constituent order. None of the participants had any prior knowledge of sign languages or other forms of conventionalized gestural communication.

2.1.2. Design and procedure

The experiment was carried out as a two-condition within-pair contrast. Correspondingly, two sets of stimuli materials were employed – one for each of the two experimental conditions: object manipulation and object construction. Each set consisted of sixteen pictures (for stimulus examples, see Fig. 1). A full list of stimulus materials is provided in the Supplementary material). In the object manipulation condition, stimulus pictures featured human agents performing manipulative actions on objects. In the object construction condition, the stimuli depicted the same set of agents engaging in simple constructive actions. Importantly, the stimulus pictures were designed to form near-minimal pairs with a high recurrence of the same constituent elements across pictures. The combinatory nature of the stimuli was intended to promote the production of gesture strings containing all three semantic elements necessary to clearly elucidate whether participants produce signs using SOV or SVO order. Since pictures differed only minimally with other pictures, participants would have to include all semantic elements in their gesture strings in order to disambiguate individual pictures.

Participants were seated at a table facing each other. They were instructed to take turns in communicating and identifying the pictorial stimuli using only gesture. Speaking was thus not allowed. Each participant was provided with a list of stimulus numbers indicating which picture to communicate for each turn and a sheet of paper with a blank column in which they indicated the stimulus numbers communicated by their co-participant. To ensure that both participants within a pair had a shared understanding of the stimulus events before engaging in the task, they were invited to briefly look at the stimuli and ask questions. The experimental procedure consisted of two conditions each comprising two rounds of 32 trials (with participants each taking turns in communicating and identifying 16 stimulus pictures), for a total of 4 rounds. The order of the two experimental conditions was counterbalanced between pairs to control for potential order effects. Three cameras were used to document participants’ gesture behaviors. One camera recorded the sessions from a global perspective, while two smaller cameras were placed on the table in front of the participants in order to capture individual gesture production in greater detail (see Fig. 2 for experimental setup).

Participants were debriefed upon completion of the task. Before the purpose of the study was revealed, they were informally interviewed in order to assess their intuitions and dispositions concerning the task, and in particular, the order in which they gestured during the task.
2.2. Analysis

2.2.1. Gesture coding

Video data from the experiment were coded by two research assistants naïve to the purpose and hypotheses of the study. A subset of the data (approximately 12%) was coded by both and tested for inter-coder reliability to ensure that the coding was not influenced by research assistants’ potential individual biases. The procedure followed a strict coding scheme targeting the relative order of individual constituent signs in each gesture string used to communicate a stimulus event. Only the first spontaneously produced gesture string in each trial would be coded for further

Fig. 1. Examples of stimulus materials. Stimulus set 1, manipulation events: (a) a nun holding a paper airplane, (b) a ballerina throwing a cake, (c) a baker eating a banana. Stimulus set 2, construction events: (d) a nun conjuring a cake, (e) a ballerina painting a paper airplane, (f) a baker building a sand castle.

Fig. 2. Experimental setup. The two participants were seated at a table facing each other. Stimulus materials were presented on a cardboard stand in one side of the table. Each participant had exclusive visual access to instructions as to which stimulus number to gesture and to a form on which to report the number of the stimulus scene they thought their partner was gesturing.
analysis. That is, the research assistants were instructed to omit gesture strings containing repeated constituents and self-repair sequences. While participants were allowed (although not instructed) to provide gestural feedback (e.g. head nods or "thumbs up") to indicate understanding, these behaviors were not included in the later analysis. Participants occasionally produced gesture strings that were ambiguous with respect to their constituent orders, thus rendering coding impossible. Such ambiguity arose primarily when participants omitted individual constituents, merged objects and actions into a single sign (e.g. via a combination of hand shape and movement), or, in rare cases, produced other orders than SOV and SVO. Since our research questions only pertain to factors affecting the relative frequency of SOV and SVO constituent orders, these cases were excluded from further analysis and amounted to 34.9% of the coded material.\(^1\) The remaining material consisted of 2348 gesture strings (1114 for the object manipulation condition, 1234 for the object construction condition).

### 2.2.2. Accuracy

We calculated participants’ task performance as the percentage of matching accuracy. Since the study was designed to investigate spontaneous gesture string production, it was not intended as a difficult task. However, participants were instructed to be as accurate as possible. Differences in accuracy between conditions were assessed using a bootstrapped paired t-test stratified by individual, pair and round. Bootstrapping is a common statistical procedure used to estimate statistical descriptors (e.g. confidence intervals) without relying on distribution assumptions through the repeated sampling of random subsets of the data. Since fully random sampling could introduce unintentional biases in the estimate by e.g. oversampling one individual, pair or round, we stratified the bootstrapping so that the sampling would reflect the actual structure of the data.

### 2.2.3. Data analysis

Intercoder reliability was analyzed using percent agreement and unweighted Cohen’s Kappa (Cohen, 1960). In order to assess the impact of structural iconicity on gesture order, we employed a logistic regression with gesture constituent order (SOV or SVO) as dependent variable and event structure (manipulation or construction event) as independent variable. In order to ensure maximal statistical robustness of our results, we employed two additional statistical procedures on the logistic regression: (i) a 5-fold cross-validation and (ii) a Bayesian variational inference on the results. Cross-validation is a procedure widely employed in statistical learning to avoid overfitting statistical models to one’s data and ensure generalizability of the findings (Hastie, Tibshirani, & Friedman, 2009). Thus, the dataset was divided into five subsets (or folds) according to pairs, meaning that data from each pair was produced in a single fold (Rodriguez, Perez, & Lozano, 2010). Subsequently, the logistic regression was trained on four folds, that is, the coefficients of the model were optimally fitted to four fifths of the pairs. Then the model was tested on the remaining fold (the remaining one fifth of the pairs). In other words, we only calculated the statistical indices of variance explained, likelihood and signifi-

\(^1\) A percentage that is comparable to those found in analogous studies. For example, in the original 2008 study by Goldin-Meadow et al., only 23% of the data consisted of gesture strings with all three constituents. In the Langus and Nespor (2010) study, roughly 60% of the data contained all three constituents, even though participants were explicitly asked to produce three gestural signs for each picture.

\(^2\) Fifty-one percent of the excluded data points consisted of gesture strings with sign simultaneity (e.g. signs for objects were incorporated in the action gesture), coder uncertainty accounted for 2% of the excluded data; pointing or posture imitation: 3%; Other orders: 3%; two-gesture strings: 38% (mainly SOV: 32% and SV: 6%); four gestures or more: 4% (mainly SVOV: 3%).


cance on the data on which the model was not trained. This procedure ensures more conservative statistical estimates of the effects observed with a higher generalizability to new observations. Given the limited dataset (25 pairs), the validation procedure was repeated on each of the five folds, so that the full dataset could be used as testing material, increasing the statistical robustness of the results. Additionally, we controlled for potential artifacts due to imbalance in the data caused by the occasional omission of constituents or ambiguous gestures in the data materials. Bayesian variational inference is a method developed to overcome potential biases (or random effects) by accounting for within-pairs and between-pairs variance components and exploiting the available group data to optimally constrain inference in individual pairs (for the full mathematical details, cf. Brodersen et al., 2013).

Finally, the relative likelihood of the model produced was calculated using Bayesian Information Criterion (BIC, the lower the better likelihood, Schwarz, 1978).

To assess the relative impact of manipulation vs. construction events (beside the general impact of structural iconicity), we estimated the effect size and statistical significance of the difference in the percentage of structurally iconic gesture strings in the two conditions. In other words, we compared the percentage of SOV gestures produced in the manipulation condition with the percentage of SVO gestures produced in the construction condition. The analysis was performed using a bootstrapped paired t-test, stratified by individual, pair and round. To assess the impact of structural iconicity on accuracy we used a stratified (individual, pair and round) bootstrapped ANOVA, with structural iconicity and event type as independent factors. All analyses were run in Matlab 2014a (Mathworks Inc.), relying on the micp, bioinformatics and statistics toolboxes. The plots were generated with ggplot2 0.9.3.1 in R 3.1.1.

### 2.3. Results

#### 2.3.1. Intercoder reliability

Intercoder reliability was found to be 98%, Cohen’s k = .96, which is generally considered “perfect agreement” (Altman, 1990).

#### 2.3.2. Task performance

Participant pairs generally performed well in the task with a mean identification accuracy of 95.60% (CI: 94.76%, 96.26%). We observed no effects of structural iconicity or of the individual conditions – manipulation vs. construction events – on accuracy (p > 0.4).

#### 2.3.3. The effect of structural iconicity on constituent order

Following our predictions, structural iconicity had a strong impact on gesture order with a balanced accuracy of 92.95% (CI: 91.17%, 94.57%), p < 0.0001, BIC = 1495.2. When presented with object manipulation events, participants produced SOV gestures in 84.66% (CI: 79.09%, 90.21%) of the trials. In contrast, when presented with object construction events, participants produced SVO gestures in 90.51% (CI: 85.33%, 94.09%) of the trials. We did not observe any statistically significant SOV or SVO bias in the production of structurally iconic gestures: Difference (in favor of SVO): 5.86% (CI: –0.06%, 12.19%), p = 0.08 (see Fig. 3).

### 2.4. Discussion

Participants generally found the task simple and intuitive as indicated by the high identification accuracy. Participants consistently produced gesture strings following the SOV order when communicating about object manipulation events, irrespective of the fact that their native language has fixed SVO order. This replicates the earlier findings by Goldin-Meadow et al. (2008) and Langus and Nespor (2010). However, participants showed an equally clear preference for SVO order when communicating about
Rather than language-internal semantic relations (or innate, organization of the events impacts constituent order through the SVO constituent order, we find it likely that the structural impact of iconicity and semantic abstractness in motivating the situation itself. We therefore introduced a second experimental iconicity. Such pressures, we believe, are found in the communicative structure that could potentially weaken the effects of structural iconicity further. Error bars represent 95% confidence intervals.

object construction events. Our findings are predicted by the principle of structural iconicity suggesting that people rely on structural relations in the referent events when ordering their communicative gestures. Similar to the way in which participants use iconic pantomime to refer to the depicted individual constituent elements (e.g., making a dancing gesture to refer to a ballerina or a praying gesture for a nun), the relative order of such elements is also iconically motivated by structure in the referent events. The results also replicate recent findings by Schouwstra and de Swart (2014) who treated construction events as a subcategory of intensional events. According to Schouwstra and de Swart, the representation of this category by means of SVO constituent order was not motivated by structural iconicity, but by the relative semantic abstractness of patients in the events. However, concerns have been raised (even by the researchers themselves, cf. Schouwstra, 2012) regarding whether construction events fit the general definition of ‘intensional’, since the patients in these events are both concrete and have ‘extensional’ properties (Parsons, 1990). While further studies are needed to decide the relative impact of iconicity and semantic abstractness in motivating SVO constituent order, we find it likely that the structural organization of the events impacts constituent order through the principle of structural iconicity. The implications are evident. Rather than language-internal semantic relations (or innate, cognitive modules, Langus & Nespor, 2010), participants used specific features of the stimulus events as a source providing structure to the emerging communication system. The principle of structural iconicity thus provides a simple but forceful mechanism for the emergence of shared structure in communication systems.

While structural iconicity accounts well for the results of experiment 1, it would also predict the systematic co-existence of both SOV and SVO in other communication systems, such as natural languages. To the authors’ knowledge, no known language has distinct, grammaticalized constituent orders for the two event types. Questions thus arise as to the additional pressures working on linguistic structure that could potentially weaken the effects of structural iconicity. Such pressures, we believe, are found in the communicative situation itself. We therefore introduced a second experimental manipulation: interactive alignment.

3. Experiment 2: the impact of interactive alignment

In experiment 1, the two experimental conditions were artificially separated in blocks of stimuli belonging to the same condition. Participants would thus consistently encounter and communicate about the same type of events within a condition. In everyday conversations, however, we frequently switch back and forth between conversational topics relating to different event types. This actualizes different constraints, such as interactive alignment – a well-documented propensity to conform to other interlocutors’ choice of linguistic structure in online interaction (Brannon et al., 2000, 2007; Fernández & Grimm, 2014; Fusaroli & Tylén, 2012; Reitter & Moore, 2014). To accommodate the potential impact of interactive alignment on choice of constituent order, we added a new experimental condition with mixed stimuli containing new tokens of both event types. In this third condition, participants performed the same basic referential task, however, a balanced set of stimulus pictures of both event types were presented in a randomized order. On the one hand, this minimal manipulation gives rise to situations where structural iconicity and interactive alignment are in agreement with respect to the predicted gesture order. That is, the constituent order afforded by the stimulus event for a given turn matches the order used by the interlocutor in the previous turn. These cases resemble the typical situation in experiment 1 where the two pressures work together and possibly reinforce each other. Importantly, however, the randomization of event types also creates situations in which structural iconicity and interactive alignment are in conflict. This would be the case when the constituent order afforded by the stimulus is different from the one used by the interlocutor in the previous turn. In such cases, the two pressures motivate different and competing constituent orders for communicating the current stimulus events. We predicted that, structural iconicity notwithstanding, participants would tend to replicate the gesture order of their partner observed immediately prior to each subsequent turn.

3.1. Materials and methods

3.1.1. Participants

This experiment was conducted as an additional condition following the two conditions in experiment 1. It relied on a subset of 13 participant pairs from experiment 1 (n = 26, 11 m/15 f, mean age 24.0, SD 3.3, see Section 2.1.1 for further details on the participants).

3.1.2. Design and procedure

In experiment 2, participants were engaged in a single session consisting of 32 trials, 16 turns per participant. They were asked to communicate and identify individual stimulus pictures within a novel set of 16 pictures. Half of the stimuli depicted object manipulation events, while the other half depicted construction events. With a few exceptions, the stimuli depicted the same agents, objects and actions as in experiment 1, but in new combinations. The order of the stimulus pictures was randomized and no cues indicated the association of individual pictures to the two conditions. Apart from these features, the task and experimental procedure were identical to those used in experiment 1.

3.2. Analysis

3.2.1. Gesture coding

Coding of gesture strings followed the same procedure as in experiment 1. Again, gesture strings that did not include signs for all three constituent elements or, in rare cases, did not comply with either SOV or SVO orders were excluded from further analysis. The excluded data amounted to 34.1% of the coded material. 3

3 Forty-seven percent of the excluded data points consisted of gesture strings with sign simultaneity. Two-gesture strings: 51% (mainly SO: 50%); four gestures or more: 2% (all SVOV).
The remaining material consisted of 274 gesture strings. Since the analysis of interactive alignment relies on comparisons between one interlocutor’s gesture order in trial t and the other interlocutor’s gesture order in trial t−1, only cases with available data from adjacent trials were considered. This amounted to a total of 206 gesture strings, of which approximately half (108) had event structures congruent with previous trials and the other half (98) had event structure incongruent with the previous trials.

3.2.2. Data analysis
To assess the relative impact of interactive alignment, we estimated the extent to which participants used the same gesture strings as their partner in the previous trial in cases where the stimulus event structure was either congruent or incongruent with respect to adjacent turns using a stratified (individual and pair) bootstrapped two-way ANOVA with congruency and condition (manipulation or construction events) as independent variables and structural iconicity as dependent variable. All other analyses (including task performance and impact of structural iconicity) followed procedures identical to those used in experiment 1.

3.3. Results

3.3.1. Intercoder reliability
The intercoder reliability was found to be 100%, Cohen’s k = 1.0 (perfect agreement).

3.3.2. Task performance
Again, we observe high task performance with a mean identification accuracy of 98.29% (CI: 96.65%, 99.28%).

3.3.3. The effect of structural iconicity on constituent order
Structural iconicity still had a strong impact on gesture order: Balanced accuracy: 80.48% (CI: 76.72%, 83.92%), p < 0.00001, BIC = 290.4. Participants produced SOV gestures in 82.74% (CI: 73.81%, 92.90%) of the trials when presented with object manipulation events and SVO gestures in 72.13% (CI: 62.52%, 81.69%) of the trials when presented with object construction events (see Fig. 4). We did not observe any statistically significant SOV or SVO bias: Difference (in favor of SOV): 10.86% (CI: 0%, 20.78%), p = 0.059.

3.3.4. The effect of interactive alignment on constituent order
Compared to structural iconicity alone, interactive alignment improves the model, that is, it statistically impacts the structure of the gesture strings produced, it increases our accuracy in predicting the data and it increases the model likelihood given the data: Balanced accuracy: 93.15% (CI: 88.92%, 98.63%), p < 0.0001, BIC = 103.79. We observed a main effect of alignment between previous gesture order and the structure of the current stimulus event: Participants followed the structure of the event when congruent with their interlocutor’s previous gesture order in 94.59% (CI: 88.88%, 99.17%) of the trials. Participants followed the structure of the current event when incongruent with their interlocutor’s previous gesture in 83.09% (CI: 74.96%, 90.41%) of the trials. Difference: 11.50% (CI: 2.11%, 20.62%), p = 0.017 (see Fig. 5). We observed no statistical interaction between congruency and condition (manipulation or construction events), p > 0.6.

3.4. Discussion
While structural iconicity remained the dominant factor, participants in experiment 2 displayed a statistical tendency to align gesture orders with their interlocutors, following the predictions of the principle of interactive alignment (Pickering & Garrod, 2004b). If the constituent order in their interlocutor’s previous gesture string were incongruent with the current event to be communicated, they displayed a statistically lower degree of structural iconicity than in cases of congruence. This happened even though participants, prior to experiment 2, were engaged in two full, blocked conditions of manipulation and construction events potentially creating strong priors to comply with the stimulus structure.

Both factors – structural iconicity and interactive alignment – thus have an impact on the local choice of constituent order for communicating an event and together enables us to better account for the constituent orders used by participants during interaction. Interestingly, though, the two factors originate from very different environmental pressures. While structural iconicity pertains to properties of the referent events, interactive alignment is related to the act of communication itself, seemingly irrespective of contents (however, see Fusaroli, Raczaszek-Leonardi, & Tynen, 2014; Fusaroli et al., 2012; Reitter & Moore, 2014). It has been suggested that the propensity of people to re-use each other’s expressive forms facilitates the establishment of common ground (Pickering & Garrod, 2004b). In effect, as interlocutors come to share
well-coordinated repertoires of signs, they depend to a lesser extent on information encoded and available in the signs themselves (e.g. their iconic similarity to their referent), but can rely on shared histories of interaction to scaffold understanding. In a series of drawing game-like experiments, Garrod and colleagues found alignment to work against iconicity (albeit on the “lexical” rather than “syntactic” level): through repeated trials, participants produced increasingly reduced and parsimonious signs, eventually appearing arbitrary to the casual bystander (Garrod et al., 2007). In the present experiment 2, we also observe that, in the presence of interactive alignment, structural iconicity seems to have a weaker effect. However, the effect is symmetric (equally strong for manipulation and construction events) and thus does not point toward a tendency to reduce structural properties in terms of, for instance, generalization and stabilization of one constituent order for both event types. This motivates questions as to which factors work to selectively conventionalize one of several possible forms. This is the topic of inquiry in experiment 3.

4. Experiment 3: the impact of conventionalization

In experiment 2, participants encountered and communicated about stimuli consisting of both manipulation and construction events in a randomized order. This weakened the impact of structural iconicity due to the propensity of participants to align with each other’s previously used constituent order. However, no statistically significant bias was observed in favor of a specific order. In experiment 3, we ask whether the general mechanism of interactive alignment can lead to the generalization, stabilization and conventionalization of a single constituent order for communicating both object manipulation events and object construction events. Studies have shown that speakers are sensitive to frequency distributions in their input when learning language structures (Reeder et al., 2013; Wonnacott et al., 2008). Compounding on these effects, different languages may realize different (yet otherwise equally motivated) linguistic conventions due to the salience or functional relevance of the individual factors in the cultural and environmental context. For instance, as mentioned earlier, generalization of absolute spatial frames of reference is mainly found in languages spoken in rural communities in areas with open terrain and stable cardinal landmarks, while languages associated with more urban or forestal environments primarily prefer relative frames of reference (Majid et al., 2004). Similarly, it is suggested that variability in the number of basic color terms in the languages of the world is a result of different functional needs and technological practices in the associated cultures (Berlin, 1991). The motivated selection and conventionalization of linguistic structures has the effect that these structures are generalized overriding less salient structures of motivation. We hypothesized that differences in relevance and frequency of reference might have an impact on the linguistic structures that are preferred, selected and thus conventionalized. In experiment 3, we manipulated the frequency by which participants would encounter events from the two conditions (manipulation and construction events). By presenting an 80/20 percent skewed distribution of stimulus events, we predicted that participants would be inclined to generalize the constituent order of the majority event type to the minority. Thus, if participants encountered a randomized set of 80% manipulation events and 20% construction events, we predicted that they would show a propensity to generalize the SOV constituent order to both types of events, while they would generalize the SVO order, if the majority of the stimuli were of the construction event type. Importantly, by this we do not intend to suggest that the actual distribution of SOV and SVO in the world’s languages is related to relative distributions of manipulation and construction events. Rather, the objective is to test the general prediction that simple distributional and frequency relations in the referent environment impact linguistic structure. We hypothesized alignment to play a central role in conventionalization processes. As seen in experiment 2, interactive alignment loosens the bond between stimuli and linguistic structure in favor of communicative facilitation (Garrod & Doherty, 1994; Garrod & Pickering, 2009) pushing the system toward parsimony and eliminating redundancies (Fay et al., 2008; Galantucci, 2005). We thus predicted that variance in individual pairs’ propensity to align with each other would correlate with the more general degree of conventionalization (i.e. generalization of the constituent order afforded by the majority event type spreading to the minority type).

4.1. Materials and methods

4.1.1. Participants

28 new pairs of participants \((n = 56, 28 \text{ m}/28 \text{ f}, \text{ mean age 22.5}, \text{ SD 6.1})\) were recruited for the experiment in return for a monetary compensation. Participants were recruited among students at Aarhus University. Pair members knew each other in advance. All participants were native speakers of Danish. None of the participants had any prior knowledge of sign languages or other forms of conventionalized gestural communication.

4.1.2. Design and procedure

In experiment 3, participants were engaged in 4 rounds of 30 trials, 15 turns per participant. Again, they were instructed to communicate and identify individual stimulus pictures within a set of 15 pictures. In one condition 80% of the events depicted in the stimuli consisted of object manipulation events, while the other 20% consisted of object construction events. In the other condition the frequency was inverted: 80% of the stimuli consisted of object construction events and 20% of object manipulation events. The order of the conditions was randomized and no cues indicated the association of individual pictures to the two conditions. Apart from these features, the task and experimental procedure was identical to that used in experiment 1 and 2.

4.2. Analysis

4.2.1. Gesture coding

Video data from the experiment were coded by two research assistants naïve to the purpose and hypotheses of the study, following procedures identical to those in experiment 1 and 2. A portion of the coded material amounting to 14.8% was excluded, as it did not allow for unambiguous categorization of constituent order. The remaining material consisted of 2557 gesture strings.

4.2.2. Data analysis

The analysis of the impact of structural iconicity on gesture order followed procedures identical to those used for experiment 1, while the analysis of interactive alignment followed procedures identical to those used for experiment 2. To assess the impact of conventionalization, we estimated the effect of structural iconicity on gesture strings as a function of the majority stimulus pattern and the interaction with event type using a stratified (individual, pair and round) bootstrapped two-way ANOVA. Finally we tested the hypothesis that interactive alignment would drive this conventionalization effect. In other words, that pairs with a stronger
tendency to reproduce each other’s constituent order, would also show a stronger tendency to systematically generalize the dominant constituent order (i.e., the order reflecting the most frequent event type). An index of interactive alignment was calculated for each pair in the following way: (i) First we calculated the reinforcement effects of each interlocutor on the other, that is, the tendency to reproduce the constituent order afforded by the stimulus event when it matched the order used by the interlocutor in the previous turn (positive reinforcement). (ii) Second, we calculated the competition effect that each interlocutor created on the other, that is, the tendency to reproduce the constituent order afforded by the stimulus event when it did not match the order used by the interlocutor (negative competition). (iii) Finally we subtracted negative competition from positive reinforcement, thus mapping the full extension of the changes induced by one’s interlocutor’s previous constituent order. In other words, interactive alignment quantified the impact of the constituent order used in the previous turn in both directions, either reinforcing or competing with the structural iconicity of the stimuli. An index of each pair’s conventionalization effects was calculated in the following way: (i) First we calculated the positive reinforcement impact of the most frequent type of events, that is, the tendency to reproduce the constituent order afforded by the most frequent event type when presented with such events. (ii) Second we calculated the competition effect of the most frequent type of event, that is, the tendency to reproduce the constituent order afforded by the less frequent event type when presented with such events. (iii) Finally we subtracted negative competition from positive reinforcement, thus mapping the full extension of the changes induced by the dominant event type. In other words, the conventionalization effects index quantified the impact of the dominant event type in both directions, either reinforcing or competing with the structural iconicity of the stimuli.

We then performed a bootstrapped correlation analysis between interactive alignment and conventionalization effects, stratified according to condition and pair identity.

4.3. Results

4.3.1. Intercoder reliability

The intercoder reliability was found to be 99%, Cohen’s $k = .97$ (perfect agreement).

4.3.2. The effect of structural iconicity on constituent order

Structural iconicity had a significant effect on gesture constituent order: Balanced accuracy = 86.01% (CI: 81.48%, 89.82%), $p < 0.00001$, BIC = 2272.5. Participants produced SOV gestures in 65.87% (CI: 55.37%, 75.93%) of the trials when presented with object manipulation events, and SVO gestures in 88.88% (CI: 82.69%, 94.09%) of the trials when presented with object construction events (see Fig. 6). Unlike in the first two experiments, we observed a statistical SVO bias: Difference (in favor of SVO): 23.01% (CI: 11.51%, 35.68%), $p < 0.0001$.

4.3.3. The effect of interactive alignment on constituent order

Compared to structural iconicity alone, interactive alignment did not significantly improve the model. Participants followed the structure of the current event in 83.16% (CI: 77.41%, 88.78%) of the trials when congruent with their interlocutor’s previous gesture, and in 75.58% (CI: 69.13%, 81.91%) of the trials when incongruent with their interlocutor’s previous gesture order. However, interactive alignment effects were not statistically significant. Difference: 7.57% (CI: −1.44%, 16.85%), $p = 0.155$. We also did not observe significant interactions with event type ($p = 0.99$).

4.3.4. The effect of conventionalization on constituent order

Compared to structural iconicity alone, conventionalization effects slightly improve the model, that is, it statistically impacts the structure of the gesture strings produced, it increases our accuracy in predicting the data and it increases the model likelihood given the data. Balanced accuracy: 90.85% (CI: 86.72%, 94.06%), $p < 0.0001$, BIC = 2118.6. On average, participants followed the structure of the current event in 85.73% (CI: 78.91%, 94.44%) of the trials when congruent with the majority event type, and in 69.50% (CI: 60.81%, 76.86%) of the trials when incongruent with the majority event type, with a difference in favor of the majority event type: 16.13% (CI: 3.84%, 28.85%), $p = 0.01$ (see Fig. 7). We observe an effect of event type, with SVO showing a greater propensity to generalize to minority events (46.90%) than SOV (8.41%), but no statistical interaction ($p = 0.1$) with type of event represented.

Finally, frequency effects and alignment effects were observed to be correlated: $R = 0.66$ (CI: 0.39 0.86), Adj$R^2 = 0.41$ (CI: 0.13 0.74), $p < 0.0001$. In other words, pairs showing a high tendency to align to each other also show a high tendency to conventionalize the constituent order characterizing the majority of events. This suggests that alignment might be driving conventionalization (see Fig. 8).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig6.png}
\caption{Effects of structural iconicity on gesture constituent order in experiment 3. The distribution of gesture constituent orders, SOV and SVO, in response to the two types of stimulus pictures, manipulation events and construction events, respectively. Error bars represent 95% confidence intervals.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig7.png}
\caption{The effect of stimulus frequency on structural iconicity. Bars represent the impact of structural iconicity on gesture constituent order for events of high frequency (80% of the stimulus events) versus low frequency (20% of the stimulus events). Error bars represent 95% confidence intervals.}
\end{figure}
The 95% confidence intervals of the line fit. The effect of alignment on conventionalization. Plot represents the correlation between participants' propensity to align to their partners' constituent order and their propensity to generalize the constituent order afforded by the more frequent stimulus events to less frequent stimulus events. Gray shading represents the 95% confidence intervals of the line fit.

4.4. Discussion

The purpose of experiment 3 was to test whether the salience of some referent events (operationalized as relative frequency) could drive the selection and conventionalization of a single constituent order for the communication of the two types of events. While we still observed an effect of structural iconicity on constituent order, the most frequent event type was more likely than the least frequent type to give rise to structurally iconic gestural sequences. In other words, altering the relative frequencies of the two event types had a significant impact on the communicative output toward conventionalization. Furthermore, the results suggest that alignment reinforces conventionalization: the more participants align with each other the stronger the conventionalization of gesture order. Conventionalization and alignment are both forces that push communicative exchange. This creates a trade-off between iconicity and alignment/ conventionalization. While alignment works locally to facilitate smooth and easy interaction, it also has the effect of loosening the bond to the referent stimulus (as seen in experiment 2). Over longer period of interaction, this can lead to routinization effects (Pickering & Garrod, 2004a): as a function of repeated interactions, interlocutors can rely on shared histories of successful interactions to create increasingly parsimonious and structured communication systems that – in effect – are easier to learn, remember and produce, while relaxing detailed iconic resemblances to referents (Fay et al., 2008; Fedzechkina, Jaeger, & Newport, 2012; Garrod et al., 2007; Kirby et al., 2008; Smith, Fehér, & Ritt, 2014).

Two additional interesting phenomena were observed: only moderate conventionalization of order and asymmetrical effects with stronger dominance of SVO. Although we found significant conventionalization effects over the course of repeated experimental trials, we did not find convergence on a single constituent order reflecting only the dominant event type. Full conventionalization of a single order might require a larger time scale with more trials, and perhaps additional concomitant factors. For instance, the frequent change of task-partners within a 'speech community' can radically enhance the conventionalization effect as indicated in previous semiotic experiments (Garrod & Doherty, 1994; Garrod, Fay, Rogers, Walker, & Swoboda, 2010) and agent-based simulations (Baronchelli & Díaz-Guilera, 2012; Baronchelli, Gong, Puglisi, & Loreto, 2010; Loreto & Steels, 2007; Puglisi, Baronchelli, & Loreto, 2008; Steels, 2011).

As to the asymmetry effects, in experiment 1 and 2 we found no statistical biases in structural iconicity for manipulation event and construction events. However, in experiment 3 we see an asymmetry: When the most frequent event type consisted of construction events, the conventionalization effect (SVO) was stronger than when the majority of the stimulus events were manipulation events (SOV). We speculate that this effect might reflect the fact that all participants were native speakers of Danish – a language with SVO constituent order. This would nonetheless be surprising considering the well-established and consistent finding that participants producing nonverbal gestures are seemingly not influenced by their native language syntax (Goldin-Meadow et al., 2008; Langus & Nespor, 2010). However, other gesture elicitation studies investigating constituent order in cases of semantic reversibility have found language-specific effects (Gibson et al., 2013; Hall et al., 2013; Meir et al., 2010). While SVO-speaking participants generally used SVO order for manipulation events and SVO for reversible events, these studies have also shown a tendency for SOV-language speaking participants to produce gesture strings that followed their acquired basic constituent order across the two event types. Presumably, linguistic bias might be expressed only in experiment 3 as other constraints compete and mutually weaken their effects. Further studies will be needed to properly investigate and address the potential impact of participants' native languages and the conditions under which these may impact gesture order in nonverbal communicative tasks.

5. General discussion

The three experiments presented here point to the strong impact of diverse environmental and communicational constraints in shaping linguistic structure, such as constituent order, in novel communication systems. As such the findings can also potentially inform ongoing discussions on the underlying motivations driving language evolution. In all experiments, we found a clear effect of event structure on constituent order, pointing to the prominent role of structural iconicity. Participants produced gesture strings highly motivated by structure inherent in the referent stimulus events. In addition to structural iconicity, in the second experiment we introduced communicative pressures, which occasioned a significant (although comparatively smaller) interactive alignment effect indicating sensitivity to co-participants' gestural output. In other words, participants were less inclined to follow the structure of the stimulus event when this was incompatible with their interlocutor's prior gesture order. In the third experiment, we tested whether this propensity for interactive alignment to occasionally override structural iconicity can potentially lead to the conventionalization of a single constituent order for both event types, given that one event type is more salient in the environment. While structural iconicity still remained the stronger predictor of participants' gesture order, a clear tendency toward conventionalization of a single constituent order was observed which positively correlated with participants' propensity to align.

Interestingly, in none of the experiments were participants able to account for their gesture behaviors when asked during debriefing. In fact, they were generally neither aware of switching between SOV and SVO, nor of the inclusion of two different event types.

5.1. Multiple constraints motivating linguistic structure

Most natural spoken languages have either SOV or SVO as the basic constituent order, with a slight prevalence for SOV...
for the understanding of human language and communication. Experimental manipulation of a large array of constraints is crucial for the study of linguistic structure and conventionalization processes. In line with this idea, Fay et al. (2013) found that participants when deprived of the use of verbal language, were much more successful in communicating about different sets of concepts using hand gestures than non-verbal vocalizations. Results are interpreted in favor of the hypothesis that language evolved from manual modes of communication rather than from vocalizations (Tomaseillo, 2008).

An intriguing observation emerging from these lines of research concern the nature of linguistic motivation. Prominent cases often discussed in the literature typically concern sound symbolism and structural iconicity in syntax treating linguistic motivation as synonymous with mapping relations between referents and linguistic forms. Motivation is thus contrasted with arbitrariness that is often associated with ideas about innateness (Hauser et al., 2002) or strong forms of cultural relativism (De Saussure, 1972). However, our results suggest that strong environmental pressures associated with the communicational situation itself can push linguistic structure toward selection and conventionalization of mappings that might appear more arbitrary (as when SVO is generalized to manipulation events) (cf. also Fay et al., 2010; Garrod et al., 2007; Perfors & Navarro, 2014; Rastier, 2001). Indeed, not only linguistic structures that display iconic derivatives of the referent are ‘motivated’; also more abstract, conventional and even arbitrary aspects of language can potentially be traced back to different environmental constraints associated with the communicational usage situation.

6. Conclusions

Various approaches in the language sciences have searched for sources and motivations for linguistic structure in language- and cognition-internal processes, either in terms of cognitive biases, inherent semantic relations, or innate structure. Complementing these lines of research, our studies provide experimental evidence suggesting that various environmental and communicative factors are effective sources of motivation for linguistic structure.

Acknowledgements

The authors would like to thank Nicolas Fay for constructive and insightful feedback in early phases of this research project and Karen Nissen Schriver, Ditte Sofie Hylander Poulsen, Mette Christine Hein Christensen and Jonas Nölle for their great help coding the video material. This work was funded by The Danish Council for Independent Research’s project Joint Diagrammatical Reasoning in Language, the ESF EUROcores program Digging the Roots of Understanding and a seed grant from The Interacting Minds Centre, Aarhus University.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2015.09.004.

References


Sémantique et recherches cognitives


Plewczynski, D., Łukasik, M., Kurdej, K., Zubek, J., Rakowski, F., & Račaszek-Leonardi,


validation in prediction error estimation. IEEE Transactions on Pattern Analysis

and Machine Intelligence, 32(3), 569–575.

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 of the United States of America, 102(7), 2661–2665.


emergence of language. Utrecht: LOT.


461–464.


Journal of Personality and Social Psychology, 10(3), 301.


Stjernfelt, F. (2007). Diagrammatology: An investigation on the borderslines of


Tylén, K., Fusaroli, R., Bjøndahl, J. S., Rączaszek-Leonardi, J., Östergaard, S., &

Stjernfelt, F. (in press). Diagrammatic reasoning: Abstraction, interaction, and


together: A dynamical account of linguistic meaning making. Semiotic, 194,

39–62.


verb argument structure: Distributional learning in a miniature language.


Zlatev, J. (2008). From proto-mimesis to language: Evidence from primatology and