The cultural evolution of complex linguistic constructions in artificial sign languages

Yasamin Motamedi[1,2] **(Yasamin.Motamedi@ed.ac.uk)**
Marieke Schouwstra[2] **(Marieke.Schouwstra@ed.ac.uk)**
Jennifer Culbertson[2] **(Jennifer.Culbertson@ed.ac.uk)**
Kenny Smith[2] **(Kenny.Smith@ed.ac.uk)**
Simon Kirby[2] **(simon@ling.ed.ac.uk)**

[1]Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands
[2]Centre for Language Evolution, University of Edinburgh, UK

Abstract

Though most documented sign languages make use of space to denote relationships between predicate arguments, studies of emerging sign languages suggest that spatial reference does not emerge fully-formed but takes time to develop. We present an artificial sign language learning experiment that expands the cultural evolutionary framework to investigate complex linguistic constructions. Our results demonstrate the gradual emergence of consistent devices to distinguish between sentence arguments, some of which rely on iconic spatial contrasts. These findings mirror data from emerging sign languages and point to the cultural mechanisms that facilitate the evolution of complex linguistic structures.

Keywords: language; cultural evolution; learning; communication; sign language; gesture

Introduction

Sign languages, as manual-visual linguistic systems, are able to represent relationships between a predicate and its arguments using the space around the signer. Though there are differences in exactly how spatial reference is utilised, spatial modulation is attested across most sign languages (Mathur & Rathmann, 2012). Signers use indexed locations in space to refer to particular arguments, such that a deictic point to an arbitrary location can pronominally refer to the subject or object of a sentence. This mapping can iconically represent a real-word spatial relationship, such that references to arguments in a sentence reflect their orientation in relation to each other in the real world, but that is often not the case and the relationship between referenced locations is primarily grammatical. It has been suggested that the iconic potential of spatial mappings makes the use of space almost inevitable in sign languages (Aronoff, Meir, & Sandler, 2005), and the beginnings of spatial reference systems have been attested in several emerging sign languages (Senghas, Coppola, Newport, & Supalla, 1997; Padden, Meir, Aronoff, & Sandler, 2010).

However, studies on emerging sign languages also suggest that systematic spatial reference is not a property that emerges immediately in a linguistic system, but takes time to evolve over generations of a language (Padden et al., 2010). Spatial agreement systems are used to represent complex relationships between a predicate and its arguments, and as such pose problems for learners. Furthermore, the ability to represent animate agents using the signer's own body may interfere with the development of abstracted spatial reference. Referring to multiple participants in space requires abstraction away from the signer's body and may therefore take longer to evolve (Padden et al., 2010; Meir, Padden, Aronoff, & Sandler, 2007). Finally, the use of space is not the only grammatical tool used to denote who does what to whom, and in fact its use is often restricted to particular classes of verbs.

Using a novel experimental method, we ask how systematic spatial reference emerges in a linguistic system, and how the iconic affordances of the manual modality affect the evolution of such a system. Though emerging sign languages provide valuable natural evidence of the early evolution of languages, the experimental research we present here is able to test the particular factors that influence language with a greater degree of control and precision.

Previous experimental research has experimentally demonstrated the importance of cultural evolutionary processes in the emergence of linguistic structure, namely interaction between language users and transmission to new learners of a language (Kirby, Tamariz, Cornish, & Smith, 2015; Kirby, Cornish, & Smith, 2008). The gradual development of spatial agreement systems in naturally emerging sign languages similarly points to the impact of interaction and transmission on the evolution of spatial reference. Therefore, we propose a cultural evolutionary stance on the emergence of spatial agreement, and explore the impact of cultural evolutionary processes in a laboratory study. We present a study that investigates the effect of interaction and transmission on the emergence of signals that participants produce to signal complex events with multiple animate arguments. We place silent gesture (where hearing participants with no knowledge of sign language communicate using gesture; Goldin-Meadow, So, Ozyürek, and Mylander (2008); Schouwstra and de Swart (2014); So, Coppola, Licciardello, and Goldin-Meadow (2005)) within a cultural evolutionary framework that implements interaction between participants, and transmission with an iterated learning model. Pairs of participants communicate about a set of events using only gesture and the gestures they produce are used to train a new pair of participants, who then use what they have learnt to communicate with each other. We provide an experimental account of the evolution of linguistic structure that is informed by data on natural sign languages. The present study offers a more precise understanding of the cultural evolutionary mechanisms that facilitate the emergence of linguistic structure, and serves to elucidate how modality-specific factors affect the emergence of systematic structure in language.

Methods

Participants took part in an artificial sign language learning task where they learnt gestures produced by a previous participant in a training stage, before communicating with a partner during testing, using only gesture.

Participants

50 participants (15 male, 35 female, median age = 22) were recruited from the University of Edinburgh's Careers Hub website, and were compensated £7 for participation in the experiment, which took between 30 and 50 minutes to complete. Participants were self-reported right-handed native English speakers with no knowledge of any sign language.

Materials

Participants were asked to communicate events , presented orthographically as pairs of sentences in English. Sentences involved two arguments, *Hannah* and *Sarah*, who could either be the agent of the sentence, be the goal or recipient of the sentence, or who might not be present in the target sentence at all (see figure 1a for examples). Sentences were presented in pairs at each trial, and pairs of sentences were grouped into blocks of four, where each block comprised a sentence pair of one of four verb types: plain spatial verbs (e.g. to cycle), spatial locative verbs (e.g. to cycle to), physical transfer verbs (e.g. to kick a ball to), and non-physical transfer verbs (e.g. to help; all verbs are shown in table 1 in the appendix). There were four blocks of four pairs in total, giving a total of 16 sentence pairs, and 32 target sentences (figure 1 gives an example of sentences as they would be shown in pairs and blocks). Two blocks consisted of different-agent pairs, such that if *Hannah* was the agent in the first sentence of a target pair, *Sarah* would be the agent in the second sentence of that pair, and vice versa. The two remaining blocks consisted of same-agent sentence pairs, such that either *Hannah* or *Sarah* was the agent in both sentences in a target pair (e.g. *Hannah is walking to Sarah/Hannah is swimming to Sarah*). Order of presentation for target sentences was randomised within sentence pairs and within blocks.

Participants were placed in individual experiment booths; target sentences were presented on-screen using Psychopy (Peirce, 2007) and video recording and streaming between networked computers was enabled via custom software, Videobox (Kirby, 2016).

Procedure

Participants were organised into 5 transmission chains of 5 generations, where each generation was made up of a pair of participants, who communicated with each other during testing (see figure 1b). Participants in generations 2-5 took part first in a training stage, followed by a testing stage. Participants in the first generation of each chain only took part in the testing stage, and were therefore required to innovate gestures at the beginning of each chain.

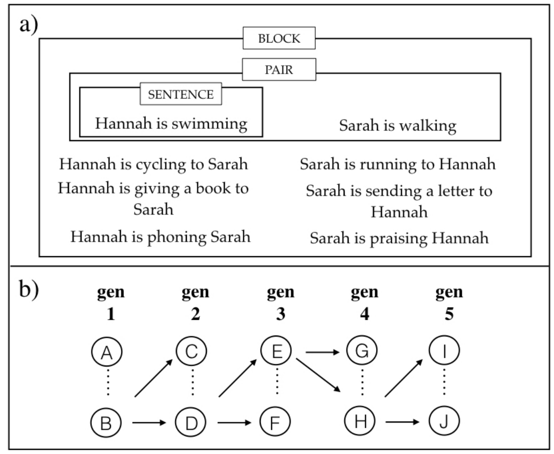

Figure 1: Examples of sentence pairs and blocks (a) and diagram illustrating diffusion chain structure (b). Pairs of participants interact at each generation (dotted lines); gestures produced by 1 participant in a pair are transmitted as training to the next generation (solid lines with arrows). A single chain is made up of 5 generations and there were 5 chains in total.

Training stage Participants in the training stage were trained on gestures produced by a participant in the previous generation of the chain. The training model was selected at random from one of the two participants in the previous generation, and the full set of gestures produced by that participant were used as training data. At each trial, the participant was shown a video of their model gesturing, and was asked to select the pair of target sentences they were trying to communicate. Whilst the video was playing, participants were shown an array of sentence pairs onscreen, from which to make their choice. The array of sentence pairs comprised the target pair, and three distractors. The distractors differed from the target pair on either the agent configuration, the verb, or both. For instance, a target pair that had *Hannah* as agent in the first sentence and *Sarah* as agent in the second, would have a distractor pair with the same verb construction, but as a same-agent pair, with *Hannah* as the agent of both sentences. Another distractor would keep the agent configuration, but would replace the verbs with other verbs from the same category, and a final distractor would present both different verbs and a different agent configuration. Building the arrays in this way required participants to specify who does what to whom, without necessarily having to describe Hannah and Sarah (i.e., the role is important, not the individual). The position of each pair on the screen was randomised at each trial, and participants could make their guess by pressing the 1, 2, 3 or 4 key, depending on the position on the screen.

Participants were given feedback about whether their guess was correct or incorrect, and shown the correct answer. Participants completed 16 training trials, one for each sentence pair. Participants completed the training stage individually, without any interaction with their partner.

Testing stage Participants in the testing stage communicated with a partner, taking turns to be director (the person gesturing) and matcher (the person interpreting). As director, a participant was shown a sentence pair on screen and asked to communicate it to their partner. A three-second countdown prepared them for streaming and recording, when they would produce their gesture to camera. The directing participant saw themselves onscreen whilst they gestured, with their image mirrored. An unmirrored image was streamed to their partner in another booth. Either participant could interrupt the video stream when they had finished their gesture or were ready to make a guess. The matcher's task was very similar to the training task; they watched their partner gesture on screen and had to select the correct target pair from an array of four. Once the matcher had made their guess, both participants were given full feedback, about both the correct answer and the matcher's selection. Participants switched roles after every block (every four trials) and directed and matched for the full set of 16 sentence pairs, giving a total of 32 testing trials. Order of the blocks and target sentence pairs was randomised for each participant.

Results

Gesture coding

Gestures were coded for the presence of an agent gesture, a goal (or recipient) gesture and a verb gesture. For each argument, the type of gesture was coded, as well as the location and path of the gesture. The goal or recipient of the target sentence was frequently omitted from gestures; as such, we focus on differentiation between agents across sentence pairs.

Differentiation strategies

We identified three main strategies participants employed to differentiate between agents in target sentences: the lexical strategy, the body strategy and the indexing strategy (exemplified in figure 2). All strategies make use of iconic representation, and both body and indexing strategies make use of the space around the gesturer to disambiguate target sentences.

Lexical strategies Two out of five chains differentiate sentence arguments based on the gesture type, using a 1- and 2-handshape to denote *Hannah* and *Sarah* in target sentences. Though this begins as a way to simply distinguish the first sentence in a target pair from the second, these handshapes come to represent individual arguments in later generations.

Body strategies A further two chains rely on differences in body orientation to signal differences between agents in target sentence pairs. Participants use an iconic spatial strategy to represent sentence arguments. For instance, in figure 2b, the participant orients their body to the right to represent Hannah, and to the left to represent Sarah.

Indexing strategy Finally, one chain developed a strategy in which locations in the space around the gesturer were indexed to refer to different sentence arguments. In the example shown in figure 2c, the participant points to her left to signal the agent, *Hannah*. In addition, her verb gesture moves between the indexed locations for agent and recipient.

Participants show a difference based on sentence type (whether same- or different-agent), producing different gestures to represent different agents, and showing divergence between sentence context over generations of the experiment. Figure 3 shows the proportion of gesture sequences that are different across two sentences in a target pair. Rows show the proportion of variation across different aspects: agent gesture type, location of the agent gesture, location of the verb gesture and path of the verb gesture. Variation across these aspects corresponds to different strategies. Chains 1 and 5 primarily vary gestures on agent type, using the lexical strategy (e.g. figure 2a). Chains 2 and 4 vary gestures based on agent location, as well as verb path and location, as they are implementing a body strategy, where agent and verb are simultaneously inferred through the participant's use of their own body (e.g. figure 2b). Finally, chain 3 shows the primary difference on the location of agent gestures, using an indexing strategy to place sentence arguments in difference locations (figure 2c).

We analysed the changes in agent distinctions over generations in the experiment, collapsing the measures shown in figure 3 across features to simply account for whether or not participants create a distinction between agents in the two sentences of a target sentence pair, investigating whether participants structure signals in similar ways across strategies. A binomial mixed effects model analysed the fixed effects of generation and sentence type on the proportion of different agent gestures, as well as their interaction. Chain, target and participant were included as random effects with random intercepts, and random slopes of generation and verb type were implemented for chain and target, respectively. The random effects structure for participant was nested within chains. Comparison of the model revealed a significantly better fit over a reduced model without the interaction term ($\chi^2 = 11.51, p < 0.001$). The model indicated a significant effect of the different-agent sentence type in comparison to same-agent sentence pairs ($\beta = 2.73, SE = 0.46, z = 5.91, p < 0.001$), as well as a significant interaction between generation and sentence type ($\beta = 0.69, SE = 0.22, z = 3.12, p = 0.002$), though no significant effect of generation ($\beta = 0.04, SE = 0.12, z = 0.29, p = 0.77$). Participants were more likely to produce gestures that differentiate between agents in different-agent contexts compared to same-agent contexts, and this contrast strengthens over generations in chains of participants.

We also analysed the effect of verb type on the distinctions participants made. Spatial reference in signed languages is not used across all verbs, but usually affects specific sets of verbs. As such, it is possible that participants in the experiment create distinctions based on semantic properties of the verbs they encounter. We ran a binomial mixed effects analysis that examined the effect of verb type on differences between agent gestures. The random effects structure described above was also implemented here. The model showed no improvement over the null model ($\chi^2 = 1.82, p = 0.61$), indicating that participants do not condition differences between sentence pairs based on verb type.

Discussion

Our results demonstrate the evolution of systematic agent distinctions, which emerge over generations of interacting participants. In addition, participants frequently use iconic spatial mappings to create those distinctions, which become increasingly contrastive over generations in the transmission chains. These findings are consistent with data from naturally emerging sign languages that suggest the gradual emergence of systematic spatial mappings.

Differentiation strategies reflect sign language structure

The three main strategies that participants employ in the present experiment all find comparable forms in natural sign languages: specifically, as lexical signs, role-shift, and spatial agreement. The latter two strategies make use of the space around the gesturer to create distinctions between agents in the target sentences. The body strategy involves movement of the participant's body to represent sentence arguments, and exhibits similarities to role-shift found in natural sign languages, and can be used in natural languages to distinguish between sentence arguments (Padden, 1986; Cormier, Fenlon, & Schembri, 2015).

The indexing strategy, however, is the strategy that most closely resembles sign language verb agreement, such that locations in the space around the signer are indexed to refer to different sentence arguments (Liddell, 2003; Padden, 1986; Lillo-Martin & Meier, 2011). The use of indexing in chain 3 begins on the axis perpendicular to the participant's body; for example, the participant points to themselves to denote *Hannah*, and points directly away from their body to denote *Sarah*. Over generations of the chain, the use of indices is abstracted away from the body and indexes are contrasted parallel to the gesturer's body, such that *Hannah* might be indexed to the left of the participant, and *Sarah* might be indexed to the right. This change mirrors development in two young sign languages, Al-Sayyid Bedouin Sign Language (ABSL) and Israeli Sign Language (ISL). In both ABSL and ISL, early generations of signers demonstrate greater preferences for spatial contrasts that centre around the signer's body, on the perpendicular axis (Padden et al., 2010). However, later generations show an increase in the use of the parallel axis, as demonstrated in the present study. Furthermore, participants in our study made no distinction between verb types, consistent with findings from ABSL that showed spatial mappings were not restricted to any class of verbs in early generations of the language (Padden et al., 2010). Our findings, consistent with natural language data, indicate that systematic spatial reference does not emerge wholesale, despite the iconic affordances of the modality, but takes time to emerge.

The evolution of complex constructions

The gestures participants produce support a gradual evolution of systematic linguistic structure, including the use of space; participants indicate a difference between sentence arguments from the first generation, but the mechanisms used to create these distinctions are neither consistent nor systematic early on. Instead, participants converge on particular strategies to make distinctions over generations, and participants show increasing divergence between the same- and different-agent sentence contexts. Participants' reliance on iconic, spatial gestures supports silent gesture research showing that hearing participants can use deictic indexing to track referents (So et al., 2005), though the present results further demonstrate how such a system evolves through use. The increased consistency of these systems supports previous iterated learning experiments (Reali & Griffiths, 2009; Smith & Wonnacott, 2010) that suggest learning leads to regularisation. Participants also demonstrate the negotiation of a system that is both expressive and learnable; they minimise the number of strategies used to convey differences in target sentences, settling on one strategy to use in the majority of trials, sufficient to express the differences in the meanings they are trying to convey. Consistent with previous experimental research, the systems participants produce maximise simplicity and informativeness (Kirby et al., 2015; Regier, Kemp, & Kay, 2015). Participants systematically signal differences between agents in target sentences, producing gestures that allow successful communication within their pair.

The effects of iconicity on emergent structure

All participants rely on iconic representations to communicate target sentences to their partners. In particular, the use of the gesturer's body and the use of space around the gesturer allow for iconic representations of animate agents. The privileged status of the body is attested in natural sign languages (Meir et al., 2007), and use of the body is attested in the development of spatial grammatical devices (Meir et al., 2007; Padden et al., 2010; Kocab, Pyers, & Senghas, 2014). Further, consistent with research on emerging sign languages and experimental research (Padden et al., 2010; Theisen, Oberlander, & Kirby, 2010), our results suggest a movement away from iconic reliance on the body (e.g. the axis change in chain 3) as gestures become more consistent and regular across the system.

Conclusion

We have demonstrated the emergence of systematic signals to communicate complex events, through the cultural evolution of communicative signals, via interaction between users and transmission to new users. Participants make use of different representation tools, all of which have analogues in natural

Figure 2: Examples of differentiation strategies used in the experiment. (a) shows an example of the lexical strategy, in which the participant uses 1- and 2-handshapes to denote arguments. (b) shows a participant using body orientation to denote sentence arguments. (c) illustrates the indexing strategy, in which the participant indexes locations to refer to sentence arguments.

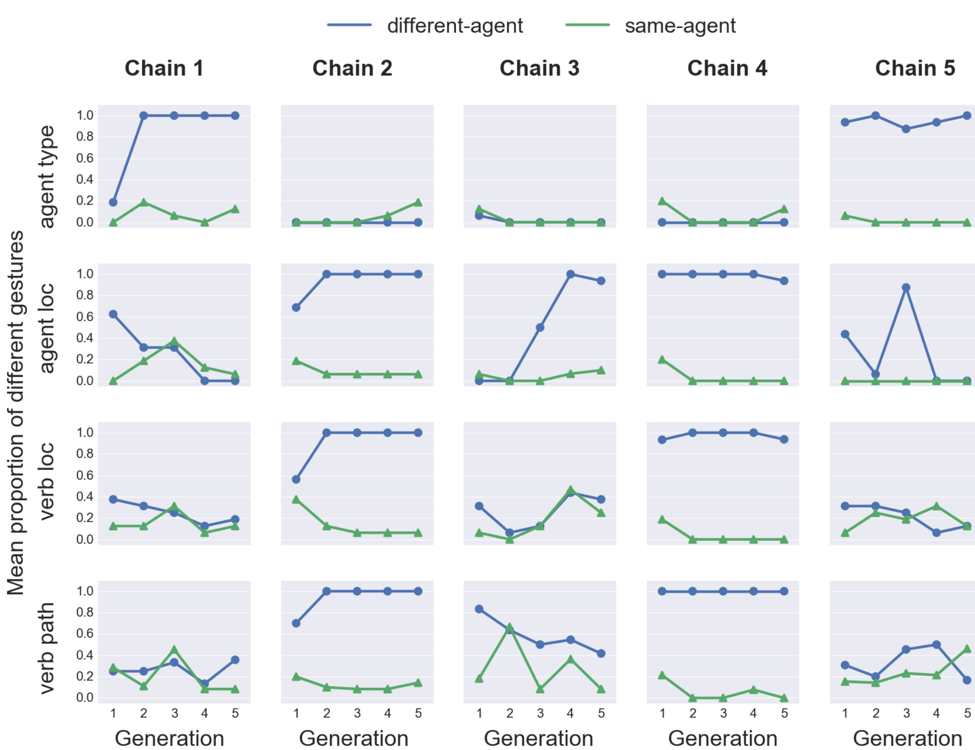

Figure 3: Proportion of gestures that differentiate agents in target sentences, based on which aspect of the gesture is varied (agent type, agent location, verb location). Coloured lines show proportions for different-agent (blue circles) and same-agent contexts (green triangles). Columns show the proportions for each chain, at each generation. All chains show differences based on context, though they make distinctions in different ways

sign languages. Our findings support data concerning the evolution of spatial reference in emerging sign languages, which suggest that the phenomenon takes time to emerge and systematise. Using an experimental method, we have been able to observe this gradual evolution in a controlled environment, to test more precisely the mechanisms that drive the emergence of spatial reference. Furthermore, these results shed light on modality-specific effects of iconicity, and their influence on the structure of emerging communication systems.

Acknowledgments

Yasamin Motamedi was funded by the Carnegie Trust for the Universities of Scotland.

References

Aronoff, M., Meir, I., & Sandler, W. (2005). The Paradox of Sign Language Morphology. *Language*, *81*(2), 301–344.

Cormier, K., Fenlon, J., & Schembri, A. (2015). Indicating verbs in British Sign Language favour motivated use of space. *Open Linguistics*, *1*, 684–707.

Goldin-Meadow, S., So, W. C., Ozyürek, A., & Mylander, C. (2008). The natural order of events: how speakers of different languages represent events nonverbally. *PNAS*, *105*(27), 9163–8.

Kirby, S. (2016). *VideoBox.* Edinburgh: University of Edinburgh. Retrieved from `http://edin.ac/2haREUz`

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *PNAS*, *105*(31), 10681–6.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and Communication in the Cultural Evolution of Linguistic Structure linguistic structure. *Cognition*, *141*, 87–102.

Kocab, A., Pyers, J., & Senghas, A. (2014). Referential Shift In Nicaraguan Sign Language: A Transition From Lexical To Spatial Devices. *Frontiers in Psychology*, *5*.

Liddell, S. K. (2003). *Grammar, gesture and meaning in American Sign Language*. Cambridge: Cambridge University Press.

Lillo-Martin, D., & Meier, R. P. (2011). On the linguistic status of 'agreement' in sign languages. *Theoretical Linguistics*, *37*(3-4), 95–141.

Mathur, G., & Rathmann, C. (2012). Verb agreement. In R. Pfau, M. Steinbach, & B. Woll (Eds.), *Sign language: An international handbook* (pp. 136–157). Berlin: De Gruyter Mouton.

Meir, I., Padden, C., Aronoff, M., & Sandler, W. (2007). Body as subject. *Journal of Linguistics*, *43*(03), 531–563.

Padden, C. (1986). Verbs and role-shifting in American Sign Language. In C. Padden (Ed.), *Proceedings of the fourth national symposium on sign language research and teaching* (pp. 44–57). Washington DC: The National Association for the Deaf.

Padden, C., Meir, I., Aronoff, M., & Sandler, W. (2010). The grammar of space in two new sign languages. *Sign Languages: A Cambridge Language Survey. Cambridge University Press, Cambridge, UK*, 570–592.

Peirce, J. W. (2007). PsychoPy–Psychophysics software in Python. *Journal of neuroscience methods*, *162*(1-2), 8–13.

Reali, F., & Griffiths, T. L. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, *111*(3), 317–328.

Regier, T., Kemp, C., & Kay, P. (2015). Word meanings across languages support efficient communication. In B. McWhinney & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 237–263). Hoboken,NJ: Wiley-Blackwell.

Schouwstra, M., & de Swart, H. (2014). The semantic origins of word order. *Cognition*, *131*(3), 431–6.

Senghas, A., Coppola, M., Newport, E. L., & Supalla, T. (1997). Argument structure in Nicaraguan Sign Language: the emergence of grammatical devices. In E. Hughes, M. Hughes, & A. Greenhill (Eds.), *Proceedings of the boston university conference on language development* (pp. 550–561). Boston: Cascadilla Press.

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*(3), 444–9.

So, W. C., Coppola, M., Licciardello, V., & Goldin-Meadow, S. (2005). The seeds of spatial grammar in the manual modality. *Cognitive science*, *29*(6), 1029–43.

Theisen, C. A., Oberlander, J., & Kirby, S. (2010). Systematicity and arbitrariness in novel communication systems. *Interaction Studies*, *11*(1), 14–32.

Appendix

Table 1 shows the verbs in each category used in target sentences in the experiment.

Verb type	Verbs
plain spatial verbs	to cycle to run to swim to walk
spatial locative verbs	to walk to P to run to P to swim to P to walk to P
physical transfer verbs	to kick a ball to R to give a book to R to send a letter to R to throw a hat to R
non-physical transfer verbs	to help R to phone R to praise R to scold R

Table 1: Verbs used in target sentences