**Content-, conformity- and frequency-biased selection in the evolution of expressive forms in cross-signing**

Kangsuk Byun, Sean Roberts, Connie de Vos, Ulrike Zeshan, Stephen Levinson

**Abstract**

This paper studies communication among deaf sign language users with highly divergent linguistic backgrounds who have no signed or written language in common. It constitutes the earliest, least conventionalised stages of improvised communication, called ”cross - signing” (Zeshan 2015), as opposed to the semi - conventionalised contact language International Sign ( e.g. Supalla & Webb 1995). The specific focus here is on the evolution of the shared repertoire amongst signers over several weeks as they co - construct meaning across linguistic and cultural boundaries. We look at two possible factors influencing the selection of expressive forms (cf. Tamariz et al. 2014): content - bias (where the more iconically - motivated, and/or easily - articulated form is selected) and coordination - bias (where participants attempt to match each other’s usage). The data set consists of a 320 - minute corpus of first encounters between dyads of signers of Nepali Sign Language, Indian Sign Language, Jordanian Sign Language and Indonesian Sign Language. Recordings took place at the first meeting, after one week, and after three weeks. The participants vary naturally with regard to their linguistic and international experience as well as their age of sign language acquisition. In addition to spontaneous conversations, we collected structured dialogues using a Direct or - Matcher task. In this language elicitation game, the Director has the coloured images and the Matcher has identical but black and white images alongside a set of colour chips from which they need to select based on the Director’s descriptions. We coded and examined the various colour expressions exploited by the participants. The semantic field of colour was chosen for this investigation into the evolution of shared communication for two reasons: the visual domain of colour retains sufficient levels of abstraction while affording signers with iconic potential. Participants initially used a range of strategies, including pointing, articulating signs for common objects with that colour (e.g. referring to a common iconic sign for ‘tree’ and pointing to the base to mean ‘brown’), and their own native variants. However, three weeks later these individuals all start using the same forms, e.g. the Indian signer’s variant for ‘green’ and the Nepali signer’s improvised ‘tree - trunk’ variant for ‘brown’. The iconic motivation of the latter and the ease of articulation of the former suggest that the content - bias is in play. The coordination - bias also seems influential in the group’s eventual selection of one variant (cf. Tamariz et al. 2014). We explore these and further factors that may affect the two biases in the selection of forms within our data. We also consider participants’ meta-linguistic skills (Zeshan 2013) and fluency in multiple sign languages (Byun et al. in preparation).

1. **Introduction**

*Why cross-signing is interesting for research of cultural evolution*

*Previous research*

*This study why & how*

*Structure of the paper*

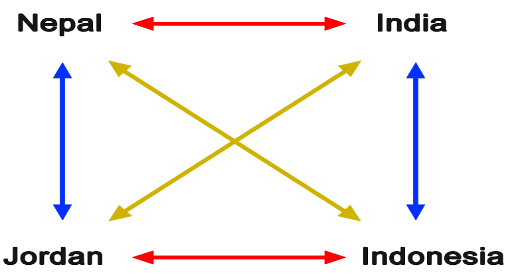
**(Sean’s note:** this paper addresses what Morin (2015) calls the Flop problem: what determines which variants are successfully taken up, and which are flops?)

1. **Data and participants**

**2.1 Data sets**

The data for this study come from two European Research Council-funded projects, one entitled Multilingual Behaviour in Sign Language Users (MULTISIGN), directed by Zeshan at the University of Central Lancashire in the UK, and the other called INTERACT, directed by Levinson at the Max Planck Institute for Psycholinguistics in the Netherlands. Both took place from 2011-2016. These projects analysed two types of data: conversational/spontaneous, and experimental/elicited. The latter is the focus of this paper.

The participants whose signing is represented in this corpus had no shared language at the commencement of data collection, and thus were compelled to devise and build communication in an ad hoc manner. As mentioned above and depicted in Figure 1, they were from Nepal, India, Jordan and Indonesia, and signed with each other in pairs, such that each person participated in three conversations. (The coloured arrows indicate the temporal simultaneity of the conversations; e.g. the two red pairs had their conversations at the same time.)



**Figure 1. Participant configurations of cross-signing participants**

[**https://youtu.be/jS\_pp2zhwGA**](https://youtu.be/jS_pp2zhwGA)

The four participants had never met before and all used different languages, although they each had very rudimentary knowledge of English. The data collection took place in India, where their spontaneous conversations were filmed (in the configurations shown in Figure 1) three times: upon their first meeting, at their second meeting which took place one week later, and at their third meeting which was three weeks later. The experimental data was filmed at the same intervals, though several days after each of the conversations, such that the participants had the chance to have ‘natural’ conversations before tackling the elicitation activities. The resulting experimental corpus totalled 5 hours and 25 minutes. In the data coding, we concentrated on the first and third meetings, for an annotated corpus of 2 hours 30min. The coding was carried out using the video annotation program ELAN (Sloetjes 2013). Further detail about this process is given below in section 3.2.

As mentioned, one part of the project data is made up of conversations that are spontaneous and ‘natural’, by which we mean that the participants were not being managed or directed in any way, apart from being asked to sign in pairs in front of the camera. The data set that we are concerned with here is based on experimental tasks with motivated materials that put the participants under pressure to communicate about a specific topic, in this case colours, and work toward completion of a set exercise. This enabled us to quantify, examine and compare the production of certain forms and utterances robustly to determine what had changed or developed across the three-week period.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **Dyad A**  **Signers’ nationalities** | **Dyad B**  **Signers’ nationalities** | **Recording length** |
| **1week** | **Session1** | **Jordan-India** | **Nepal-Indonesia** | **137min** |
| **Step2** | **Jordan-Indonesia** | **Nepal-India** |
| **Step3** | **Indonesia-India** | **Jordan-Nepal** |
| **2week** | **Step4** | **Jordan-Indonesia** | **Nepal-India** | **107min** |
| **Step5** | **Jordan-Nepal** | **Indonesia-India** |
| **Step6** | **Jordan-India** | **Nepal-Indonesia** |
| **3week** | **Step7** | **Indonesia-India** | **Jordan-Nepal** | **81min** |
| **Step8** | **Jordan-Indonesia** | **Nepal-India** |
| **Step9** | **Jordan-India** | **Nepal-Indonesia** |

**Table 1. Summary of data**

**total 5:25**

**1week : 2hour 17min (1week → 2week = 30min)**

**2week : 1hour 47min (2week → 3week = 26min )**

**3week : 1hour 21min (1week → 3week = 56min)**

[**https://youtu.be/bwsLT2lm00c**](https://youtu.be/bwsLT2lm00c)

Table 1 shows the total length of time taken by all the participants to complete the elicitation tasks in the first, second and third rounds. It is apparent that the time decreases precipitously from the first to the second round, and then decreases again slightly between the second and third round. This may be indicative of a rapid initial evolution in the cross-signers’ selection of variants. The fact that the spontaneous conversations preceded the elicitation tasks by several days in each instance [JMBW1] enabled the participants to develop a basic framework of communication without having to focus on a specific target area or activity, and without being put under pressure to achieve any particular outcome. This facilitated the experimental task because the participants had gained some familiarity with the tools they could use to interact with each other, thus making it feasible to concentrate on a shared exercise. This also made it more straightforward for the researchers, as many communicative problems that would have complicated the data were resolved prior to the experiments, and this enabled us to focus more explicitly on the selection of forms and linguistic evolution.

[JMBW1]Could a visual depiction be given for this, e.g. a small graph with columns/bars, e.g. red for conversations and blue for experiments? Showing the relative times and time lapses between all of the various sessions?

**2.2 Participant profile**

|  |  |
| --- | --- |
| **Participants** | **Language Background** |
| **Signer A**  **Family background: Hearing parents**  **International experience:**  **no experience** | **Fluent: Jordanian Sign Language**  **Intermediate: (written) Arabic**  **Minimal: (written) English** |
| **Signer B**  **Family background: Hearing parents**  **International experience:**  **no experience** | **Fluent: Indonesian Sign Language**  **Intermediate: (written) Bahasa Indonesia**  **Minimal: (written) English** |
| **Signer C**  **Family background: Deaf parents and siblings**  **International experience:**  **no experience** | **Fluent: Nepali Sign Language**  **Minimal: (written) Nepali, (Written) English** |
| **Signer D**  **Family background: Hearing parents**  **International experience:**  **no experience** | **Fluent: Indian Sign Language**  **Minimal: (written) ?, (Written) ??** |

**Table 2. Socio-linguistic backgrounds of the participants**

[**https://youtu.be/-QU1nAQoRCI**](https://youtu.be/-QU1nAQoRCI)

The four participants are all deaf and have all used sign language from a young age. They are all roughly 20 years old (see xxxxxx). The Nepali signer is the only female in the group and the only one who comes from a deaf family. None of the participants had any previous experience of international travel and this was in fact the first time they had met people from other countries. The Jordanian, Indonesian and Nepali signers were provided with in-country assistants or guides to help them become familiar with the setting in India (which was unnecessary for the Indian signer). As stated, their only commonality in terms of language background is a very minimal use of English, and hence they sometimes relied on fingerspelled English words, using the one-handed rather than the two-handed alphabet. They are each well-versed in their national spoken language but did not tend to fingerspell words from these languages.

This project represents their first time ever participating in cross-signing, although as discussed above, the data analysed for this paper was technically gathered upon their second meeting, since they had already had one round of spontaneous conversations by the time that they engaged in the first experiment. During the intervening week, the participants were also free to interact informally all day with each other and with the wider university community, though at night they were in university accommodation which was gender-segregated. This meant that the Nepali signer was separated from the other three in the evening, though in practice all of the signers spent more time interacting with local people and university students than with each other during the one-month study period. A summary of participants’ sociolinguistic backgrounds and language skills is provided in Table 2. Information on the studio environment and filming set-up is presented in section 3 below, because this is where we explain in full the Director-Matcher task targeting colour signs.

**3. Methodology**

In sections 3.X and 3.X, we aim at a detailed description of precisely how our experimental tasks were carried out and what they involved, including how many colours were used for each session, how many of them were the same across each session in the round, and so on. Readers who do not require this depth of detail at this point may find it useful to refer to section 3.X.

**3.1. Studio Environment** [**https://youtu.be/BrmFzfttaxc**](https://youtu.be/BrmFzfttaxc)

The culture and setting of India is by nature extremely visually colourful, and it is not surprising that the environment in and around the university was filled with an array of different colours. This made it challenging to secure a room in which the availability of coloured items was minimal so that participants could not rely on pointing to them when describing specific colours to each other in the experimental tasks. Eventually we were able to find a minimalist room with pale green walls, which can be seen in Figures 2, 3 and 4 below. Figure 2 also shows that the participants had solid-coloured dividers to shield their view of each other’s pictures and colour chips. In addition, they were able to see the tan-coloured chairs against the wall, and the view from the window, as illustrated in Figure 3. From the window they could see green trees and a white, often very cloudy sky. The use of colour was minimised by having the participants and coordinator coordinator wear black, as depicted in Figure 4.

Arranging for all participants to wear the same colour, and sourcing a room at the university that was one solid colour but was also large and suitable enough to set up lighting and cameras, was difficult in this context. Also, Lucknow often has power outages, which required us to vacate the building temporarily until the lights came back on. Problems with lighting can have a large effect on an experiment that relies on people identifying and describing specific colours, and it was a challenge to ensure that the lighting was correctly positioned so as not to cause glare, as well as to manage the data gathering process in the midst of frequent power cuts.



**Figure 2,3,4**

**3.2. Set-up** <https://youtu.be/Q5CYMsShGS8>

Figure 1 above indicated that two dyads at a time were filmed. It was initially unclear how the experiment coordinator would be able to manage two dyads simultaneously. However, our chosen studio was sizable enough to permit two dyads to work at the same time. Furthermore, the coordinator was able to orchestrate and devise a setup that enabled him to oversee all four of the participants at once, whilst ensuring that participants could see and manipulate items on their side of the desk but not their interlocutor’s side, and making certain that the camera could capture the signing and visual materials (see Figure 2).

**3.3. Director-Matcher task**

Director-Matcher task , Bohnemeyer, Jurgen (2001)

**3.3.1.** [**https://youtu.be/58zF\_sTIR7s**](https://youtu.be/58zF_sTIR7s)

The activity that we utilised to elicit the experimental data was a ‘director-matcher task’ (Bohnemeyer[GW1] , 2001). The task was adapted[GW2] to focus on the elicitation of colour signs (see section XX). For each day of the data collection, nine pictures were used in total across three rounds, so each participant had the chance to use three of the pictures. The same pictures were used for subsequent data collection days, but they were distributed randomly and had different arrangements of colours on them.

The same picture was given to both participants in the pair, and on each picture, five items were coloured. Of these five items, two were coloured the same across both pictures, and three were in different colours, for a total of eight colours targeted per game (see Figure 5). This meant that each participant had the opportunity to sign about 24 different colours (eight colours across three rounds) in the course of a day. With three rounds over three separate days of data collection, this permitted us to look at nine rounds of the game to examine the evolution of colour-related forms in cross-signing.

Table X summarises these permutations and lists the colours used in each round, which we identified by giving each one a number. The numbers highlighted in yellow were the colours that were the same across both pictures.

[GW1]Description from ref to be added

[GW2]More about this?



**Figure 5**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **Stimulus** | **Colours-1** | **Colours-2** |
| **Week1** | **First time**  **1** | **Picture 1** | **1, 9, 12,14, 4** | **1, 9,13, 22, 10** |
| **Picture 2** | **16, 24,20, 6, 2,** | **16, 24, 8, 5,7** |
| **Picture 3** | **18, 19,3, 21, 15** | **18, 19,17,11,23** |
| **First time**  **2** | **Picture 4** | **6, 7,16,19, 20** | **6, 7,23, 24,9** |
| **Picture 5** | **3, 11,10, 2, 21** | **3, 11,18, 8, 12** |
| **Picture 6** | **14, 22,4, 5, 1** | **14, 22,13, 15, 17** |
| **First time**  **3** | **Picture 7** | **12, 21,10, 18, 19** | **12, 21,15, 1, 7** |
| **Picture 8** | **2, 5, 14, 9, 17** | **2, 5,20, 22, 24** |
| **Picture 9** | **8, 23, 16, 3, 4** | **8, 23,6, 13, 11** |

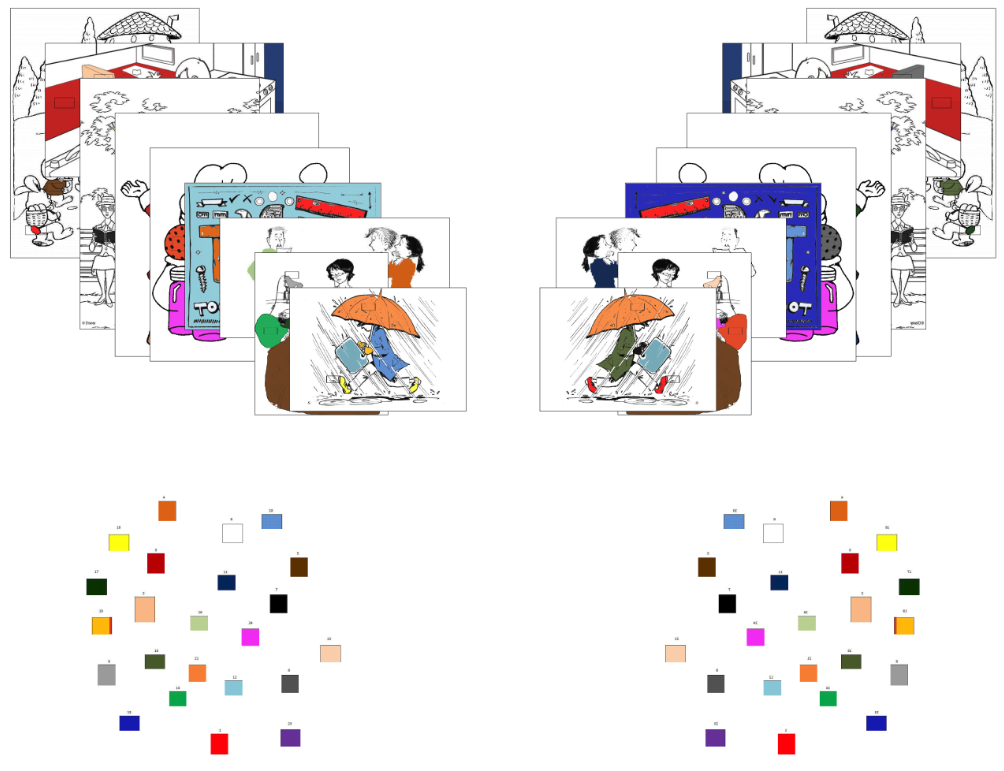
[**https://youtu.be/NJVUncrHNYI**](https://youtu.be/NJVUncrHNYI)

Table 3 also gives an indication of the kind of planning that was carried out prior to the experiments to ensure robust organisation and reliable use of the materials. The actual table we used is much longer, encompassing all of the dates and rounds, but for the sake of brevity only the part related to the first week of data-gathering is shown here. The numbers in yellow, which indicate colours that were the same across both pictures, are also not repeated after that one use in the week, i.e. 1 and 9 do not appear after the first row. This was done to maximise discussion about as large a range of colours as possible.

**3.3.2.**

[**https://youtu.be/RpJPQqiixjQ**](https://youtu.be/RpJPQqiixjQ)

The way in which the director-matcher task works is that one participant (the ‘director’) describes the colours in their picture while the other participant (the ‘matcher’) uses the descriptions to select the correct colours and place them in the relevant places in his or her black-and-white version of the same picture. There is no overt control by any participant or coordinator; the roles are flexible as both participants have coloured images to describe and they may take turns being the matcher and the director. This appears to promote a greater level of comfort for the participants and more natural utterances, including spontaneous try-marking and repair attempts. Explicit coordination by the researcher was avoided even where participants were producing stilted monologues or lists of the colours in their picture, which they tended to do at first, merely pausing between each item to give the matcher a chance to select the appropriate colour chip (e.g. ‘Okay, the first item is red. The second is blue. The third is light green… Good, your turn now.’) The participants were able to discern on their own that this was not a fruitful way of engaging with the task, and they shifted to shorter and more conversational turn-taking (e.g. ‘The colour on my picture is red. What colour do you have?). This allowed them to complete all components of the task satisfactorily by carefully going through their five coloured items together, rather than one participant listing her five and then the other listing his five. It was important at the end of the round for the researcher to capture on camera the final arrangement of colour chips on the black and white images, to see what colours the matcher thought were being referred to, and whether they corresponded to the colours on the director’s image.



**Figure6**

**3.4. Their language collect (add pic)**

<https://youtu.be/glyMj5VSRL0>

After the month of data collection, but before the participants departed from the setting, their respective in-country guides (see section X) interviewed them about what signs they use for various colours in their own sign language, using visual prompts. These conversations were recorded so that they could be cross-checked with the experimental data during the analyses. In each case, the guide made sure to inform the participant that if their language has more than one sign for that colour, or if they are aware of other signs for that colour from other languages, we would like them to tell us of all such variants as well.

These interviews were vital in securing all of the relevant signs from the signers’ languages, as they may have sometimes opted not to use a sign from their own language, e.g. when they felt it would not be understood as well as another strategy. Also, they helped us to distinguish instances when a signer already knew a particular sign (and selected it based on the knowledge they assumed their interlocutor to have), from instances of true assimilation or communicative evolution, i.e. when a signer began accommodating the usage of the interlocutor. (Example of Indonesia and Nepal.)

The guides also watched the data clips with the participants to explicitly verify the origin and intent of the particular signs they used, e.g. ‘Was that sign from your own sign language, or hers?’ The resulting information was placed into a table for ease of reference during the analysis.

**Figure7**

**3.5 ELAN transcription**

**week1 week3**

Predictions:

The predictions are not easy to make. One the one hand, using a variant in an explicit teaching context, or checking or try-marking may draw attention to the variant or reinforce the memory of the variant for the interlocutors. This would predict that the presence of these factors would cause an increase in the frequency of a given variant. However, these contexts are often employed when there is some kind of trouble in communication (e.g. the signers have different native signs for the concept or one interlocutor does not understand the sign). Therefore, if a variant appears in these contexts it could be an indication that it is not a fit variant (e.g. its meaning or referent is not transparent, or it causes confusion of some kind). In this case, one would predict that the presence of these factors would be an effect of poor suitability and predict a decrease in frequency. Cause and effect are often difficult to tell apart, but at least in this case they make different predictions about the direction of the correlation.

**Statistical analysis [SEAN]**

For each unique variant, several variables were extracted from the ELAN transcription using pympi (Lubbers & Torreira, 2014). For the final week, this included the frequency the variant for the given target colour. For the first week, the variables included: the indexicality of the variant (not indexical, indexical of the body or indexical of something else); whether the variant had been explicitly taught at any point; the number of times the variant had appeared as try marked; whether the variant had been used in a checking context; the total frequency of the variant across all colour contexts; the average length of the variant in milliseconds. A mixed effects analysis framework (XXXX) was used to predict the frequency of each variant in the final week from variables in the first week, with a random effect for each target colour (see supporting information).

1. **Results**

There was a significant main effect of indexicality ( indexical β = 1.5, body-indexical β = 0.85, log likelihood difference = 21, df = 2, χ2 = 42.69, p < 0.001 ). Indexical variants were more frequent in the final week.

There was a significant main effect of teaching ( β = 4.2, log likelihood difference = 6.4, df = 1, χ2 = 12.76, p < 0.001 ). Teaching increased the frequency of variants in the final week.

There was a significant main effect of try marking ( β = 1.3, log likelihood difference = 23, df = 1, χ2 = 45.5, p < 0.001 ). Variants that were try marked more often were more frequent in the final week.

There was a significant main effect of checking ( β = 1.7, log likelihood difference = 5, df = 1, χ2 = 10.01, p = 0.002 ). Variants used in a checking context were more frequent in the final week.

There was no main effect of frequency in week 1 ( β = -0.012, log likelihood difference = 1.7, df = 1, χ2 = 3.42, p = 0.065 ). On its own, more frequent variants in week 1 are also more frequent in the final week (r = 0.27). However, when considering this variable with the other variables the relationship disappears.

There was no significant main effect of variant length ( β = -0.061, log likelihood difference = 0.48, df = 1,χ2 = 0.95, p = 0.33 ). There was no significant main effect of first user ( log likelihood difference = 1.9, df = 3, χ2 = 3.9, p = 0.27 ).

There was a significant interaction between try marking and teaching ( β = -0.98, log likelihood difference= 9.9, df = 1, χ2 = 19.89, p < 0.001 ). The effect of teaching was bigger when the variant was also often try marked (see graphs below).

There was a significant interaction between teaching and checking ( β = -3.4, log likelihood difference =11, df = 1, χ2 = 21.37, p < 0.001 , see graphs below).

Model comparison also showed that the final model was improved by adding indexicality as a random slope. This suggests that indexicality is more important for black and less important for green.

1. **Discussion**
   1. **Frequency bias**

*Quantitative results*

*Qualitative discussion*

There was no main effect of frequency in the statistical model. An explanation may be the following: A poor variant may be repeated many times before it is understood, while a good variant only needs to be used once. That is, frequent use in the first week may be an indication of communication problems. The sequential measures (teaching, try marking,checking) and indexicality are better measures of whether a variant is problematic, so the predictive power of frequency is small in comparison.

* 1. **Content bias**

*Quantitative results*

*Qualitative discussion*

Indexical variants were preferred over non-indexical variants. However, the strength of this effect varied by colour. The selection of variations for black was strongly predicted by indexicality, while the selection of variants for green was weakly predicted by indexicality. This could be due to the better availability of black in the environment which could act as a ground for indexing. For example, black hair was always available to the participants, while an easily indexed green object was not.

* 1. **Conformity bias**

*Quantitative results*

*Qualitative discussion*

[explanation of interaction effects]

The interaction between teaching and checking may be explained in the following way. Using a variant in an explicit teaching or checking context may reinforce the variant for the interlocutors. However, if both strategies are used, then this might indicate that there is a problem with the variant. Alternatively, perhaps the difference comes from differences in strategies used by the different signers **[Did different signers use different amounts of teaching/checking?]**.

1. **Conclusion**

* *relevance for cultural and language evolution*
* *...*

Straightforward models of cultural selection suggest that the most frequent variants will increase their relative frequency. However, although frequency of a variant in the first week was correlated with its frequency in the final week, the measures of sequential context were stronger predictors. This suggests that it’s not just what is said, but *when* it’s said that matters. A poorly motivated variant may be repeated several times between two pairs because it is causing problems. In contrast, an easily understood variant may only need to be used once before the pair can move on.

How should we relate the findings of this study to the theory of cultural evolution? In the broad models of genetic evolution proposed by Dawkins (1989, 1982) and Hull (1988, 2001), the gene is a replicator and an organism is an interactor (or vehicle) that interacts with the environment to cause the replicator to replicate. Croft’s model of language evolution argues that this model also applies to language: the word or phrase is a replicator and the individual speaker is the interactor (see Croft, 2000; Croft, 2008). In our study, the different signs are replicators, and the signers are interactors. However, we find that the type of sequence in which a form is used matters more than the simple frequency of a form. Hull notes that in biology there are many levels of interactor including the cell, the organism and the species. Therefore, one interpretation of our study is that conversational sequences are interactors (as well as signers being interactors). The type of sequence promotes (or inhibits) the replication of the signs within it, just as a particular type of cell supports the replication of the genes within it.

**References**

De Vos, C. (2011). Kata Kolok color terms and the emergence of lexical signs in rural signing communities. The Senses & Society, 6(1), 68-76. doi:10.2752/174589311X12893982233795.

Dingemanse, M. (2009) The selective advantage of body-pady-part term. Journal of Pragmatics 41:2130-2136

Dingemanse, M. & Enfield, N.J. (2015) Other-initiated repair across languages: Towards a typology of conversational structures. *Open Linguistics* 1: 96–118.

Hurford, J. (2011). Linguistics from an evolutionary point of view. In: Kempson, R., Fernando, T. & Asher, N. (eds.), *Handbook of the Philosophy of Science. Volume 14: Philosophy of Linguistics.* Amsterdam: Elsevier, pp. 473-498.

Lubbers, M. and Torreira, F. (2014), pympi python library version 1.1, *Available online* https:// github.com/dopefishh/pympi

Moerman, M. (1988). *Talking culture: Ethnography and conversation analysis.* Philadelphia: University of Pennsylvania Press. doi: 10.9783/9780812200355

Palfreyman, N. (2016) Colour terms in Indonesian sign language varieties: A preliminary study. In Zeshan, U. & Sagara, K. (eds.) *Semantic Fields in Sign Languages*. Berlin: de Gruyter/Ishara Press, pp. 267-297.

Scott-Phillips, T.C. & Kirby, S. (2010). Language evolution in the laboratory. *Trends in Cognitive Sciences* 14: 411-417.

Tamariz, M., Ellison, T.M., Barr, D.J., & Fay, N. (2014) Cultural selection drives the evolution of human communication systems. *Proceedings of the Royal Society B: Biological Sciences* 281(1788).