The emergence of word order conventions: improvisation, interaction and transmission

Marieke Schouwstra Kenny Smith Simon Kirby

Manuscript in preparation (work in progress)

Abstract

When people improvise to convey information by using only gesture and no speech ('silent gesture'), they show language-independent word order preferences: SOV for extensional events (e.g., boy-ball-throw; Goldin-Meadow et al., 2008), but SVO for intensional events (e.g., boy-search-ball; Schouwstra & de Swart, 2014). Real languages tend not to condition word order on this kind of semantic distinction but instead use the same order irrespective of event type. Word order therefore exemplifies a contrast between naturalness in improvisation and conventionalised regularity in linguistic systems. Understanding the transition from naturalness to conventions is a major goal of language evolution research. We present a new approach to this challenge using a novel experimental paradigm in which silent gesture is both used for communication (Christensen et al., 2016) and culturally transmitted through artificial generations of lab participants (Smith et al., in prep). Experiment 1 implements improvisation, interaction and cultural transmission in a so-called gradual turnover setup. We show that under the influence of interaction and transmission, the natural word order observed in improvisation gradually disappears to be replaced by a simpler, conventionalised word order regime. Experiment 2 explores how this simplification process unfolds when *only* improvisation and interaction are implemented. We show that over time, word order usage becomes simpler in this case too. The trajectories of word order simplification in Experiments 1 and 2 look surprisingly similar. This demonstrates that the conventionalisation process operates similarly whether or not it is separated out over multiple generations of learners. Although 'lineage specificity' is shown for both experiments, the resulting dominant word order is mostly SVO, the order of the native language of all participants. In Experiment 3, we investigate what happens when the frequency of the different semantic event types is skewed, and show that under this condition it is possible for different word orders to emerge. Taken together, our experiments demonstrate that where pressures for naturalness and simplicity are in conflict, languages start natural, but naturalness will give way to regularity as signalling becomes conventionalised through repeated usage.

1 Introduction

All languages have conventionalised ways to describe who did what to whom, and many do this through Basic Word Order. Although all possible order-

ings of Subject, Object and Verb have been attested as dominant word orders in at least one language (Dryer, 2013), some word orders are more common than others For example, SVO (e.g. English) and SOV (e.g. Japanese) are common, whereas VOS (e.g. Malagasay) is only found in relatively few languages (Hawkins, 2014). Although evolutionary claims have been made about the emergence of this pattern of possible word orders in the languages of the world (Newmeyer, 2000), very little is known about the dominant forces in the evolution of conventionalised word order. A theme of much recent research is that cognitive biases operating at the level of the individual play an important role in shaping linguistic rules at the population level (Kirby, Tamariz, Cornish, & Smith, 2015; Culbertson, Smolensky, & Legendre, 2012). We propose that bias should not be taken to be a unitary and constant force. Rather, we see different biases at play at different stages in the formation of linguistic rules. We provide an account of the emergence of a systematic usage of word order that is the result of cognitive biases, combined with the dynamics of language usage and language learning.

This paper investigates the origins of basic word order by focusing on three mechanisms that play a role in cultural evolution: *improvisation, interaction* and *iterated learning*. Improvisation occurs when an individual has to convey information in the absence of rules from a conventional language. Examples of improvisation and its relation to interaction and transmission can be observed, for example, in emerging languages in the manual modality: homesign systems are largely based on improvisation (Goldin-Meadow, 2005). In places where multiple homesigners come together with no existing sign language in place, such as in the first cohort of Nicaraguan Sign Language, interactive principles come to play a dominant role (Kocab, Senghas, & Snedeker, 2016). Finally, the influence of iterated learning can be seen in subsequent cohorts of NSL (Senghas & Coppola, 2001; Flaherty, 2014).

The three mechanisms can also be studied in the laboratory, in silent gesture experiments in which naïve participants convey information using only gesture and no speech (Goldin-Meadow, So, Özyürek, & Mylander, 2008; Schouwstra & de Swart, 2014; Gibson et al., 2013; Hall, Mayberry, & Ferreira, 2013). By starting from improvisation in a silent gesture task, we will look at how communicative interaction and transmission to new learners affect the structure of emerging language. This allows us to investigate how individual humans improvise solutions to communicative challenges, how pairs of individuals create conventions through interaction, and how these conventions are transmitted over time through learning.

1.1 Silent gesture: improvisation and cognitive biases

When people are forced to communicate in a way they have not communicated before, without using language, they will not necessarily use the word order patterns of their native language. This has been observed in a number of studies of silent gesture, in which people are asked to convey information using only gesture and no speech. Previous research has shown that for transitive events that involve movement through space (such as 'boy tilts glass'), people show a language independent preference for SOV word order (Goldin-Meadow et al., 2008).

When other kinds of events are described, however, people will condition

the word order of their utterances on the semantic properties of the events they describe. For example, when events are reversible (typically, when a transitive event has two animate participants, such as in 'the boy kicks the fireman'), the preference for SOV order becomes less dominant. Some studies observe a preference for SVO for reversible events (Gibson et al., 2013; Futrell et al., 2015), others observe a move away from SOV in favour of SVO among other orders (Meir, Lifshitz, Ilkbasaran, & Padden, 2010; Hall et al., 2013; Meir et al., 2017; Kocab, Lam, & Snedeker, 2017; Hall, Ahn, Mayberry, & Ferreira, 2015).

Another contrast in event types that has an effect on word order relates to the referential properties of the verb. When events do not involve just movement through space, but can be categorised as *intensional* instead, this leads to a substantial, and cross-linguistic, change in word order preference in silent gesture (Schouwstra & de Swart, 2014). Intensional events are those events in which the direct object is dependent on the action (e.g., a creation event like 'man knits scarf'), possibly non-existent (e.g., 'girl looks for unicorn'), or non-specific (e.g., 'woman likes apple') (Forbes, 2013). For the interpretation of intensional sentences, the intension (i.e. meaning) of the direct object is more important than the object in the world it refers to. This can be contrasted with extensional events (such as the motion events used in the studies mentioned above), in which the specific objects in the world *are* relevant. The change in word order preference can be explained as a cognitive preference to convey specific and less relational information before abstract and more relational information: in order to describe an event in which a girl throws a ball, the information about the thrower (the girl) and the object thrown (the ball) is more specific than the throwing, which is a relation between the girl and the ball. On the other hand, for an event in which a girl dreams of a ball, the object dreamt of is not necessarily specific, and, in fact dependent on the action. (Schouwstra & de Swart, 2014; Schouwstra, 2017)[1].

The studies discussed in this section show that various cognitive preferences play a role in the ordering of constituents when people communicate in the absence of an existing conventional system. These findings reveal what individuals bring to the table in the earliest stages of a new language, but they immediately trigger the question: what happens when improvised utterances such as these are learned and used for interaction, and become part of a system of emerging linguistic rules? In other words, how are the products of improvisation subsequently transformed by interaction and transmission by iterated learning, creating a conventional linguistic system? This is the central focus of investigation in this paper. Before we present the experiments that target this question, we will first describe previous work that presented ways in which improvisation is different from language production in a fully conventionalised setting.

1.2 Comparing production and interpretation

One difference between silent gesture and conventionalised language is that in existing languages, people produce *and interpret* utterances. Recently, researchers have investigated what happens when people interpret silent gesture.

[1](Schouwstra, 2017) mainly focuses on word order preferences for complex events in which temporal information is included, but the study replicates the word order effects for intensional and extensional events.

A study that measured reaction times in a silent gesture matching task found that participants had lower reaction times for videos that used SOV order for extensional transitive events, even when the dominant order of their native language was SVO (Langus and Nespor, 2010). Thus, also in the *interpretation* of silent gesture, SOV order seems to have a special status, even when participants' native language is SVO. This implies that interpretation of silent gesture could work similarly to its production. Hall et al. (2015), on the other hand, claim that the factors that influence production are different from those that influence comprehension. They base their claim on two experiments, the first of which presented ambiguous gesture strings to participants from two language groups, English and Korean. All strings described reversible events, and thus described an action and two animate participants, in three different orders: Participant-Participant-Action, Participant-Action-Participant, and Action-Participant-Participant. Strings in all orders were robustly interpreted as presenting the Actor (Subject) first, both language groups. A second study, in which participants rated gesture strings (in SOV, SVO and OSV order) for reversible and irreversible events (events and gesture strings were presented as video clips alongside each other), found no difference between the two event types. Together, the two experiments suggest that in production (when a communicator is unconstrained by their audience), people have a preference to use a different word order for reversible events, but for interpretation they rely solely on an 'Agent-first' principle.

In an interpretation experiment with SOV and SVO ordered silent gesture strings with ambiguous action gestures, Thompson, Schouwstra, and Swart (2016) found that people use constituent order as a cue to distinguish between intensional and extensional events: participants were more likely to interpret SVO ordered gesture strings (than SOV ordered strings) as intensional events, but the strength of the effect was much less pronounced than for production. By fitting a Bayesian model to the experimental results, the authors infer that the biases underlying the interpretation process are skewed, but essentially defeasible. The inferred biases are shown to be consistent with the production data from Schouwstra and de Swart (2014), under the assumption that interpretation differs from production in that it takes uncertainty about the target meaning into account.

The studies mentioned so far focused on improvisation, and investigate its production and interpretation separately. One crucial difference between improvisation and fully established languages (even when it is rooted in an improvised setting), is that languages are used for communication, in which production and comprehension alternate in quick succession, and and, arguably, are often even relevant at the same time (Pickering & Garrod, 2013). In this paper, we will go beyond the study of interpretation and production as separate processes, and study interactive laboratory settings, in which both processes are relevant.

1.3 Mechanisms of cultural evolution

Apart from being used for communication, conventionalised languages have been repeatedly transmitted to new learners for many generations. A growing body of cultural evolution experiments has investigated these mechanisms in the lab, so that a general picture of different kinds of information transmission in different population types is now emerging (Galantucci, Garrod, & Roberts, 2012;

Tamariz & Kirby, 2016; Scott-Phillips & Kirby, 2010).

Various studies that investigate the emergence of language-like properties in novel communication paradigms or artificial languages have implemented mechanisms from cultural evolution in experimental setups that allow them to look at the influence of communicative interaction and cultural transmission to new learners. Transmission to novel learners is implemented in iterated learning experiments in which participants are trained on an initial artificial language, and the output they generate after learning is given as the input to a new 'generation' of learners, and so on for multiple generations. It was shown that this setup, when it starts with unstructured, holistic meaning-signal pairs, can produce systematic languages, that are increasingly learnable (Kirby, Cornish, & Smith, 2008). In a series of experiments that compared the contributions of interaction and transmission on artificially learned languages (Kirby et al., 2015), it was found that whereas learning alone favours simplicity and can lead to degenerate languages (i.e. languages in which many distinct meanings are expressed ambiguously with a single label), when learning is combined with communicative interaction, this forces a language to be both learnable and sufficiently informative. [XXX add more recent work? showing the circumstances under which systematic structure emerges in different types of meaning spaces, for instance, in a continuous meaning space (XXX cite Jon + cite Regier Science paper), in meaning spaces that show discontuities (XXX cite Amy), and in cases in which some information is contextually more salient than other information (XXX cite James / Cat).]

In contrast with these studies, which typically initialise the experiments with pre-existing meaning-structure mappings, there is a body of work that starts from scratch: with situations in which there are no signal-meaning mappings, and these emerge spontaneously.

1.4 Starting from scratch: improvisation and conventionalisation in the lab

Experiments that implement dyadic interaction in a graphical communication task found that participants' drawings become less complex, less iconic, and communicatively more successful over rounds of interaction (Garrod, Fay, Lee, Oberlander, & MacLeod, 2007). Theisen, Oberlander, and Kirby (2010) show that systematicity in a set of drawings in a graphical communication task emerges through the combination of increasingly arbitrary sub-parts of drawings. In a follow-up study, Theisen-White, Kirby, and Oberlander (2011) show that transmission to novel learners plays a crucial role in increasing the systematicity of graphical signs. A cultural transmission experiment that compared reproducing a drawing pattern to learning and reproducing found that learning leads to an increase in compressibility of the visual information transmitted (Tamariz & Kirby, 2015).

In a setup that introduces new learners each generation, that first observe the current system, and subsequently engage in the graphical communication task themselves, it was found that symbolic elements persist, even when the newly introduced learners have not observed the iconic roots of these elements (Caldwell & Smith, 2012).

Christensen, Fusaroli, and Tylén (2016) investigate which word orders are used for two kinds of events, manipulation events and creation events, a clas-

sification that is similar to that in Schouwstra and de Swart (2014); creation events are a subtype of intensional events, and all manipulation events are extensional. In an experiment in which silent gesture is used for communication (but role switches between production and interpretation occur infrequently), they replicate Schouwstra and de Swart (2014) and show that SOV word order is preferred for manipulation events, and SVO for creation events. In a second experiment (which was run as an additional communication round to a subset of the participants of Experiment 1), in which participants switch roles between production and interpretation every trial, they found that word order usage is influenced by interactive alignment, which they define as 'the tendency to re-use an interlocutor's previous choice of constituent order, thus potentially overriding affordances for iconicity'. In a third experiment, they show that the relative frequency of the event types (manipulation events vs. creation events) influences the word order frequencies used by the participants. However, structural iconicity remains the strongest determinant for word order in all these experiments.

Implementing both communicative interaction and transmission to new learners, Motamedi et al. (under review) show that when people use silent gesture to communicate about a set of related concepts (e.g. chef, frying pan, hairdresser, scissors), they develop systematically structured gestures. The study directly compared a communication condition, in which participants communicate repeatedly about the same concepts, to a turnover condition, in which the output of a communicating pair is transmitted to a new pair of participants, for multiple generations. They showed that systematic functional markers (gestures that indicate the meaning class a concept belongs to, such as a 'point-to-self' gesture to denote a person for a concept like 'chef') emerge to some extent in the communication condition, but these markers only become more widespread when new learners were introduced.

To wrap up, when people are engaged in silent gesture, and improvise in the absence of a conventional language system, they will use semantic cues to structure their utterances, and they will use utterance structure to obtain different semantic interpretations. This semantically conditioned word order regime can be seen as natural behaviour, and is rooted in cognitive biases (Schouwstra & de Swart, 2014; Thompson et al., 2016). However, although improvisation occurs in established languages, these languages are crucially different from silent gesture: they are used for communicative interaction and persist through cultural transmission. In this paper, we investigate the influence of these mechanisms of cultural evolution on this natural behaviour, in lab experiments that systematically combine silent gesture with interaction and transmission. Having the high level of control of a laboratory setting, will allow us, in addition, to study the emergence of linguistic conventions in closeup.

1.5 Global and local mechanisms

We will present a series of experiments that investigate how interaction and transmission affect the products of improvised, gestural communication. In the first experiment we investigate what happens to the semantic conditioning of word order on extensional and intensional meanings that occurs in improvisation, when improvisation is combined with interaction. We predict that this setup will lead to a simplification of the word order regime: conditioning of

word order on event type will decrease over generations.

In the second, by contrast, we implement only interaction, in an experiment in which two participants use silent gesture to interact with each other for an extended period. Given the findings by (Christensen et al., 2016), we predict to see some simplification in word order use, but given previously observed differences between interaction-only experiments and those that implement transmission to new learners (Motamedi, Schouwstra, Smith, Culbertson, & Kirby, Under revision; Kirby et al., 2015), we expect this to happen to a lesser extent than in Experiment 1.

In the third experiment we investigate whether the resulting word orders over interaction and transmission are dependent on the proportions of the event types. Our prediction is, following (Christensen et al., 2016), that skewing the proportions of intensional and extensional events will influence the usage of word order over the experiment.

With our experiments, we will be able to focus on the respective contributions of interaction and iteration on the products of improvisation. At the same time, the lab setup allows us to drill down into what exactly happens in interaction and iteration on the trial-by-trial level. An online process that has been known to influence how individuals use language is *alignment*: when individuals communicate, they adapt their behaviour to become more similar to their partner. The mechanism behind this process is *priming*: in conversation, a hearer activates mental representations during comprehension, and these are then more likely to be used in production (Pickering & Garrod, 2004). The particular type of priming we are interested in for this paper is *structural* priming. This type of priming has been shown to take place, crucially in contexts where multiple syntactic constructions are possible, such as in the case of the Double Object and the Prepositional Object construction: 'the girl gave the boy a book' versus 'the girl gave a book to the boy'. Participants were more likely to produce a particular structure after just being primed with one. Finding these priming effects from language comprehension affecting language production has been taken to be evidence for the fact that production and interpretation share mental representations of syntactic structures (Branigan, Pickering, & Cleland, 2000). Apart from experimental contexts, structural priming has been shown in natural dialog, either in elicitation situations (Levelt & Kelter, 1982), or in corpora (Gries, 2005). Recently, scripted and unscripted interaction was combined with artificial language learning. It was shown that priming was robustly present in interaction, which in turn showed a reduction in unpredictable variation (usage of two variants that is not conditioned on any other factor; (Smith & Wonnacott, 2010)) over the course of interaction (Fehér, Wonnacott, & Smith, 2016). Even though this study did not show conclusive evidence that priming is the driving mechanism behind reduction of variation, this line of work has taken an important step in exploring the connection between priming and language change. This means structural priming can not only be used to study speaker's linguistic knowledge, but extend its scope to processes that take place in the language itself (Pickering & Ferreira, 2008).

A difference between the contexts in which structural priming has been investigated so far and the situation under investigation in this paper, is that this experiment starts out with improvisation: there is no set of conventions, and participants just start from their personal preferences instead. Given our expectation that participants will naturally produce SOV word order for exten-

sional and SVO for intensional events, we are interested in whether structural priming could play a role in de-coupling the two word orders from the event classes they are initially paired with. This is the first study in which potential priming effects are analysed in a situation of emerging language conventions.

Together, our experiments will inform us about word order usage develops from being *natural*, i.e. dependent on semantic properties of individual events, to increasingly *conventional*, i.e. following a simpler word order regime.

2 Experiment 1: silent gesture, interaction and transmission

To investigate what happens to natural word order when silent gesture is used for interaction and transmitted to new generations of learners, we implemented silent gesture in a so-called gradual turnover design (Caldwell & Smith, 2012). Two communicators start in the first 'generation', by taking turns to be director and matcher, and using silent gesture to communicate. They are watched by a third participant, the observer. Each consecutive 'generation', the observer replaces one of the communicators (the one that has been in the experiment for the longest time), and a new observer enters the room.

We predicted that at the start of the experiment, word order would pattern like in previous silent gesture experiments (a preference for SOV for extensional events and SVO for intensional events), but the combination of interaction and cultural transmission would drive word order usage, over generations, to become less natural (i.e., less conditioned on semantic type) and more regular (i.e., implementing a simpler word order regime with a single word order dominating across items).

2.1 Setup

Materials

We created 64 line drawings of intensional and extensional events. They were organised in 8 groups of 8, such that each group of stimuli contained all 8 combinations of 2 actors, 2 patients, and 2 actions. Each group of 8 stimuli had one intensional and one extensional action. All actors were animate, and all patients were inanimate.[2] Two examples are provided in figure 1.

The stimuli were presented on iPads (one iPad per communicating participant), which ran interactive software that we developed (a combination of Python/web sockets for the server end, and HTML/JavaScript for the front end), in a full screen web browser. For the observer, a non-interactive screen (on a laptop) was set up, that for each trial gave the same information as the matcher. To avoid an effect of the order of the elements in the images, a mirror image of every stimulus was created, and in the experiment, originals or their mirror images were used at random (randomised each trial).

Director and matcher were video recorded using two Logitech webcams connected to Apple Macbook Air laptops running VideoBox, custom software we developed for recording video (Kirby, 2016).

[2]Only inanimate patients were used to rule out any possible effects of reversibility (see section 1).

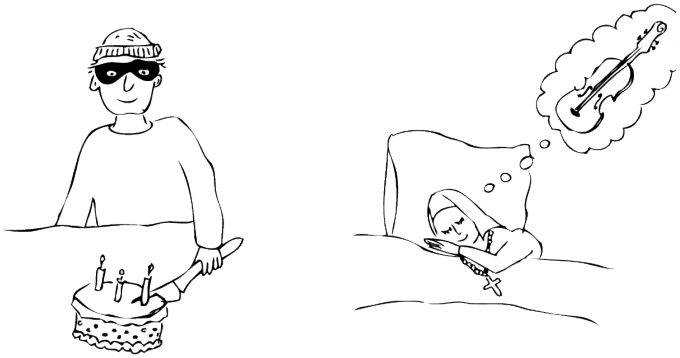

Figure 1: Two example stimuli: 'Burglar cuts cake' and 'Nun dreams of violin.' The former shows an extensional event, and the latter an intensional one. Previous work suggests that improvised orders should be SOV and SVO respectively.

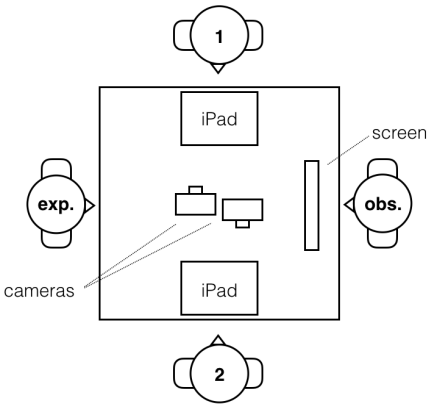

Figure 2: Overview of the experimental setup. Participants 1 and 2 interact using silent gesture, taking turns as director and matcher, with target stimuli presented on iPads. The observer, who has a clear view of participants 1 and 2, as well as the target stimuli (presented on a screen) will replace participant 1 in the next 'generation' to experimentally simulate a gradual turnover of population. The experimenter is also present during the experiment.

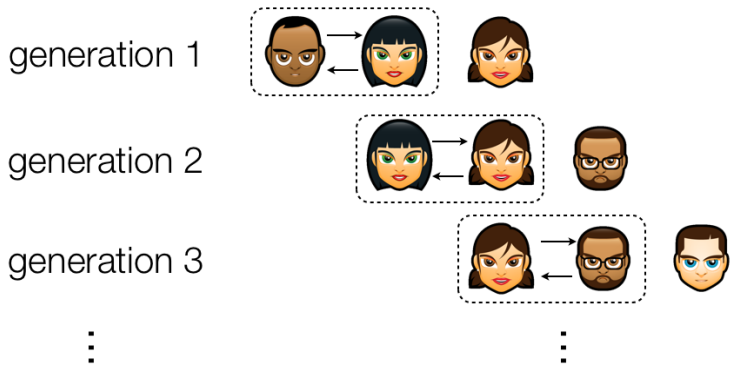

Figure 3: Gradual turnover design of Experiment 1: the experiment starts with two communicators and an observer. Each consecutive generation, the communicator who has been in the experiment the longest is replaced by the observer of the previous generation, and a new observer is brought into the lab.

Participants

Eight groups of participants (all right-handed, native speakers of English, with no previous knowledge of any sign language) took part in this study (58 participants in total; 20 male and 38 female; mean age 21.5).[3] All participants were recruited at the University of Edinburgh (three groups were recruited from a first year Linguistics and English Language course; the remaining 5 were recruited from the University of Edinburgh career platform). The participants were paid between £5-10, depending on how long they were in the experiment for.

Procedure

The setup of this experiment started in the first round with two communicating participants (taking turns to be director and matcher) and an observing participant. In every subsequent generation, one of the communicators left the room, the observer of the previous generation took the place of one of the communicators, and a new observer entered the room. The experiment covered 6 generations, and in each generations (32 trials), director and matcher switched roles every trial; see figure 3 for an illustration of the setup.

All participants were instructed simultaneously about the general procedure of the experiment, the software interface, and the feedback: participants were made familiar with the software by playing four trials of a communication game that used numbers (one through eight) instead of pictures. After the instruction phase, the first three participants remained in the lab, and the other participants waited in separate booths in an adjacent lab. The order in which participants would take part actively was assigned at random. After each generation, the next observer was brought into the lab. The waiting participants did not know in advance when they would start the experiment.

[3]When only seven participants showed up for the experiment, a confederate acted as the observer in the last round (note that the final observer had no active role in the experiment, and only watched the two communicators).

Each generation, the two communicators were seated at opposite ends of a table. The observer sat on one side of the table, and the experimenter on the other side, so that both had an unobstructed view of the communicators. In each generation, the observer was instructed to pay close attention to what the two communicators were doing, in order to be a good communicator in the next generation. The observers did not know which of the two communicating participants they would be replacing in the next round. The experiment lasted 6 generations of 32 trials each.

In each trial of the experiment, one of the participants—the director—saw one image on the iPad and conveyed what was on the screen using only their hands/upper body and no speech. The matcher saw eight images on their iPad, and they tapped the image that they thought was intended by the director. The matcher array consisted of the target picture, plus the seven alternatives from the same group of images (see above), making sure that the alternatives presented two Actors, two Patients, and two Actions in all possible combinations. This was done to force the director to provide all relevant elements in the target image.

The answer possibilities were visible throughout the trial, and there were no restrictions on when the matcher could tap their answer. The communicating participants switched roles every trial, and received full feedback immediately after each trial. Feedback was given by means of coloured borders around the relevant images on the iPads: director and matcher both saw a green border around the image when the matcher was correct. The director saw a red border plus an additional image of the selected answer when the matcher was incorrect. The matcher saw a red border around the selected answer, plus a green border around the intended answer if the answer was incorrect.

The observer was shown the same information on a screen in front of them that was given to the matcher of that particular trial (i.e., observers saw all answer possibilities, plus full feedback after the matcher's response).

Each generation consisted of 32 trials: 16 intensional and 16 extensional events (4 stimuli from each group of 8). Each communicating participant was the director for exactly 8 intensional and 8 extensional stimuli, and these were presented in random order. For each subsequent round, 16 'same' (stimuli that were used in the previous round) and 16 new stimuli were used, meaning that each item had been gestured at least once after the end of the third generation.

Participants were encouraged to communicate faster each generation. If the current generation was faster than the previous, they received a wrapped sweet as a reward. This was done to 'break the ice', to make the participants feel more comfortable, and to motivate them to communicate effectively. Everybody received a wrapped sweet after the first generation.

2.2 Analysis

The video recordings were coded for word order. To arrive at a word order characterisation, each gesture was determined as referring to the Subject, Object or Verb.[4] Many of the responses had multiple consecutive gestures that referred to one element (such as hat + nose for 'witch'); these were counted as one gesture.

[4]We are aware that the terms Subject, Object and Verb are syntactic terms, and it would be more correct to use semantic terms. However, in line with much of the literature, we will continue to describe the gestures in terms of S, O and V, to increase readability.

generation	communicator 1	communicator 2	observer
1	1	2	3
2	2	3	4
3	3	4	5
4	4	5	6
5	5	6	7
6	6	7	8

Table 1: Overview of which participant had which role in the six generations of the experiment. Note that participant 8 does not have an active role in the experiment, but was included since the presence of an observer may have an effect on the behaviour of communicators.

We kept a record of the lexical elements that were used (e.g., hat + nose), as well as the resulting word orders.

There were no time constraints on director and matcher during the task, and the matcher selected an answer whenever they were ready. Occasionally, the director would continue gesturing (or repeat some of their gestures), even when the matcher had stopped observing the director. Unobserved gestures were coded separately, but not included in the analysis. Similarly, sometimes, the matcher asked (using gesture) a clarification question, either by indicating that they did not get it, or by repeating one of the elements. Clarification questions and their (gestured) responses were not included in the final analysis.[5]

Of 1536 strings in total (recorded in 8 sessions of 6 rounds, of 32 trials each), 1100 were SVO, 205 were SOV, and 231 were other orders, coded as 'NA', and resulting in 15.0% omitted data points. The other orders included strings that omitted elements (such as in SV) or repeated elements non-consecutively (such as in SVOV). Note that there were no restrictions on the gestures that participants could use.

2.3 Results

Speed and accuracy

To assess changes in efficiency and effectiveness of communication, we looked at the mean number of constituents per trial over the course of the generations, and the number of correct responses per generation. To determine the number of constituents per trial, we looked at the coded strings and counted the number of elements in these.

A linear mixed effects model was performed on the number of constituents per trial, taking generation as fixed effect. The model took random slopes for generation on item and uncorrelated random intercepts and random slopes on chains.[6] The model revealed a significantly better fit than a reduced model which did not include generation as fixed effect (χ^2=7.06, p<.01). See figure 4.

[5]Because the participants engaged in face to face interaction, there were cases of repair. This is an interesting and rich topic, and repair sequences may contribute on the emergence of linguistic rules (Micklos, 2016; Byun, De Vos, Roberts, & Levinson, 2018), but for readability reasons, we did not include repair sequences in our analysis.

[6]The latter was done because the full model did not converge.

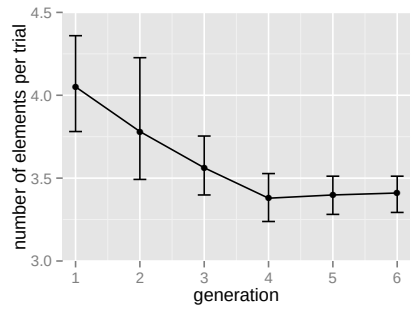

Figure 4: Number of elements per trial, by round. The minimum value on the y-axis is 3, because that is the fewest possible elements to convey the necessary information in one trial. Over time, gestures became more efficient.

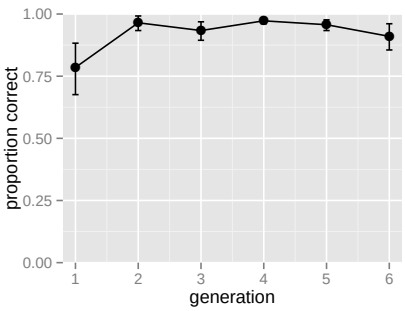

Figure 5: Proportion of correct answers per chain, by generation. Performance of matchers was at or near ceiling for most of the experiment, except the first generation.

A binomial mixed effects regression was performed on correctness of the response per trial, taking generation as fixed effect, and applying the same random effect structure as the previous model. The model revealed a high proportion of correct answers in the first generation, as reflected in the intercept ($\beta = 2.026$, $SE = 0.347$, $p < 0.001$) and a main effect of generation ($\beta = 0.305$, $SE = 0.154$, $p < 0.05$).

Word order

We coded whether an order was natural or not, i.e. if SOV was used for an extensional event and SVO was used for an intensional event, and analysed naturalness in a logit mixed effects regression, entering generation as fixed effect, and random intercepts and slopes for generation on chains and items. Participants were more likely than not to produce natural orders in the first generation ($\beta = 2.472$, $SE = 0.641$, $p < 0.001$), and significant negative effect of generation was observed (β = -0.358, $SE = 0.100$, $p < 0.001$); see figure 6.

Looking at the proportion of SOV word order (using the same random effect structure as in the previous model, but uncorrelated), we found that participants were less likely to use SOV than to use SVO in the first round (β = -2.022, SE = 0.474, $p < 0.001$), and this proportion went further down over generations (β = -0.308, $SE = 0.148$, $p < 0.05$); see figure 7.

Regularisation

To measure if word order usage in the experiment became more regular over time, we calculated entropy scores for each chain, in each generation. This was done by calculating Shannon's entropy, i.e. using $-\sum p_i \log(p_i)$ for the proportions of SOV, SVO and other word orders. Generally, usage of different word orders relatively equally often results in high entropy, and using the same word order throughout results in an entropy of 0 bits.

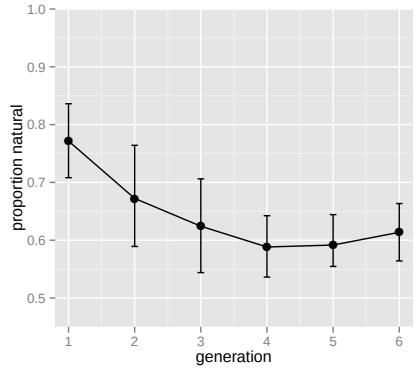

Figure 6: Proportion of natural gesture strings (SOV for extensional and SVO for intensional) plotted by generation. Note that a fully natural set of strings would obtain proportion 1, but a set of strings that are all in SOV order would score .5 on naturalness, because half of the stimuli are extensional events, for which SOV is the natural order. Early generations produced natural word orders, but this decreased over generations.

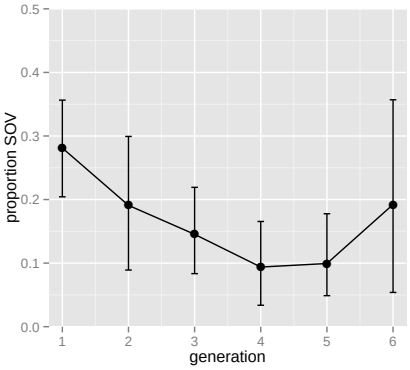

Figure 7: Proportion of strings in SOV order, by generation. SOV order was dispreferred by participants, and this effect increased over generations.

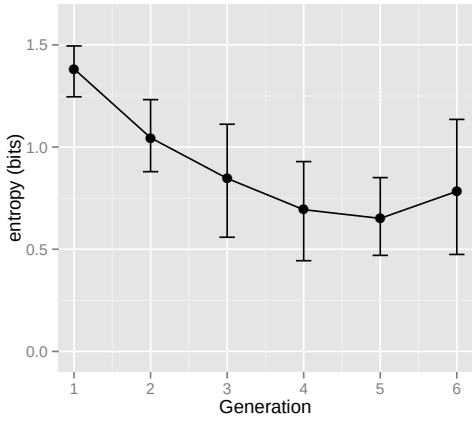

Figure 8: Entropy values plotted by generation. Error bars show 95% CI's. Entropy decreases significantly over generations.

We then ran a linear mixed effects model that took generation as fixed effect and a random effect for chain (entering random slopes on generation resulted in non-converge). The model revealed a significantly better fit than a reduced model which did not include generation as fixed effect (χ^2=16.03, p<0.001). See figure 8.

Lineage specificity

In order to assess if the change in word order happened uniformly, or rather in a way that was unique per chain, we looked at *lineage specificity* (Dunn, Greenhill, Levinson, & Gray, 2011). We compared the word order usage of participants in the same chain to that of participants in pseudo-chains, created by scrambling participants who were not in the same session. More specifically, for each participant in a given generation, we took counts of SOV strings, SVO strings, and orders categorised as 'other', and generated two tuples describing their word order usage: one for extensional and one for intensional events. For each two participants who participated in a pair, we calculated the euclidean distances for their word order use, by calculating the Euclidean distance (in three dimensional space) between the tuples:

$$dist = \sqrt{\left(\frac{\#\text{SOV}_{p1}}{\#\text{SOV}_{p2}}\right)^2 + \left(\frac{\#\text{SVO}_{p1}}{\#\text{SVO}_{p2}}\right)^2 + \left(\frac{\#\text{other}_{p1}}{\#\text{other}_{p2}}\right)^2} \quad (1)$$

We calculated separate distance scores for extensional and intensional events, and then summed the two, to calculate the Euclidean distance for each *veridical* pair of participants. We then ran a Monte Carlo simulation (1000 trials) in which we compared the veridical distances to those of pseudo-pairs (formed by pairing, for each generation, the first director with a director—from the same generation—in a different chain). We found that the veridical distances (mean = 9.72) were significantly smaller than the pseudo-distances (mean = 17.31,

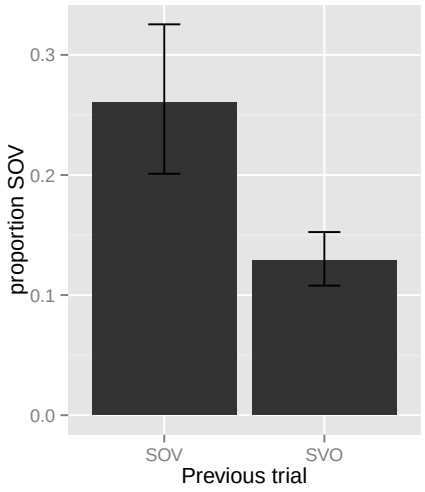

Figure 9: Comparing the proportions of SOV for two kinds of trials: when the word order in the previous trial was SOV, participants were more likely to use SOV than when the previous trial was SVO. Error bars represent 95% confidence intervals.

sd=2.58, z=3.006, p<.001), revealing that word order behaviour was specific to chains.

Underlying mechanisms for regularisation: priming

In order to assess if structural priming was a potential mechanism behind word order regularisation, we compared word orders in a given trial t to that of the previous trial t_{-1}. We compared cases in which t_{-1} had SOV word order to those in which it did not, and assessed the effect on trial t_0, by analysing current-SOV[7] in a logit mixed effects regression, entering previous-SOV as fixed effect, and and random intercepts and slopes for generation on chains and items (slope and intercept were uncorrelated for items, as the full model failed to converge). Generally, SOV was less likely than SVO to occur, as reflected by the intercept (β = -2.904, SE = 0.398, $p < 0.001$), but there was a main effect of previous-SOV (β =1.167, SE =0.427, $p < 0.01$). See figure 9.

2.4 Discussion

In this experiment we investigated how the usage of natural word order in silent gesture develops over generations when silent gesture is used for communication, and learned by novel learners. First of all, we found that over time, communication became more efficient and more successful. Focusing on word order, we found that natural word order (observed in the first generation) is not maintained, but increasingly replaced by a simpler word order regime, in which one

[7] A boolean value that gave TRUE if the trial had SOV order, FALSE if it had SVO order, and NA if it had any other order.

and the same word order is used, irrespective of the semantic properties of the event being communicated. Entropy measures confirmed that the word order regimes became increasingly regular over generations.

Further, we found that the trajectory from natural to more regular word order did not happen uniformly; the 8 different chains showed lineage specificity in their usage of word order. Thus, the chains changed their word order regime in different ways, and at different rates. However, SVO word order was found to become the dominant word order in nearly all of the chains. Because SVO is the dominant order of the native language of all participants (English), this might simply be a native language influence. Given this observation, we might ask whether it is *possible* for participants to converge on a different word order, i.e., one that is not the native word order. Experiment 3 below will investigate this.

Finally, we focused on priming as a potential mechanism behind word order change, and found that priming plays a role in the experiment: for a given trial, there was a higher chance to observe SOV order when the previous trial had that order than when it did not.

Our results contrast in interesting ways with those by (Christensen et al., 2016), who looked at the influence of social interaction on natural word order (they refer to this as structural iconicity), and found that although participants showed a statistical tendency to align with their communicative partners, natural word order was strongly represented throughout the experiment. Our experiment, on the other hand, shows a main effect of generation, indicating that the combination of transmission and interaction in our experiment was sufficient to make word order change to become significantly more regular (i.e. less meaning-dependent). A remaining question is, however, were both mechanisms *necessary* to make this process possible? In other words, was the addition of generations of new learners crucial for word order to become more regular in our experiment? Or could interaction alone have yielded the same, or a similar effect?

The next experiment investigates silent gesture in a setup that is nearly identical to Experiment 1, but instead of introducing a new participant each generation, the entire experiment is carried out by two participants (who interact over the course of 6 'generations').[8] In other words, we will investigate what happens to natural word order when it is used for communicative interaction, but not transmitted to new learners.

3 Experiment 2: silent gesture in communicative interaction

3.1 Setup

To test how word order changes when silent gesture is used for communication, we asked pairs of participants to use silent gesture to communicate about simple events, in a communication game in which they alternated the roles of director and matcher. The setup of this experiment was similar to Experiment 1, but the

[8]We will maintain the term generation, even though the experimental setup does not introduce new participants. This terminology was chosen to make comparison between the two experiments easier.

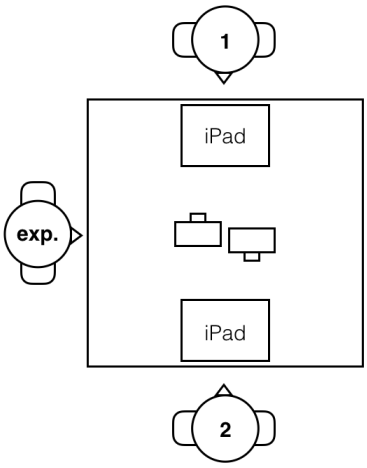

Figure 10: Overview of the experimental setup of Experiment 2. The setup was identical to Experiment 1, but there was no observer due to lack of generational turnover.

same two participants took part throughout the experiment, for six generations in total.

Materials

The experiment used the same stimuli materials as Experiment 1. Like in Experiment 1, they were presented to the participants on iPads (one iPad per participant). Because there was no observer in this experiment, no extra observation screen was needed.

Participants

Twelve pairs of participants took part in this study (24 participants in total; 13 male and 11 female; mean age 21.2). The participants were recruited at the University of Edinburgh recruitment web page. All participants were right handed, native speakers of English, with no knowledge of any conventional sign language; they were paid £7.50 for their participation.

Procedure

The two participants were seated opposite each other and each participant had an iPad in front of them, on which the stimuli and the feedback were presented. Like Experiment 1, this experiment lasted for 6 generations of 32 trials each. Participants were encouraged to speed up over the course of the experiment.

Data analysis

The video coding procedure was the same as that of Experiment 1. Of 2304 strings in total, 1552 were SVO, 398 were SOV, and 354 were other orders, coded as 'NA', and resulting in 15% omitted data points.

3.2 Results

Speed and accuracy

To test the influence of generation on the number of constituents used per gestured utterance, we ran a linear mixed effects model on the number of constituents per trial, taking generation as fixed effect. The model took random slopes for generation on item and chains. The model revealed a significantly better fit than a reduced model which did not include generation as fixed effect (χ^2=17.86, p<0.001).

In addition, we ran a model that combined the data of the two experiments, to assess if they behaved differently with respect to the number of constituents used over generations. We constructed a full model on the number of constituents, that included generation and experiments (and their interaction) as fixed effects. We compared the fit of the full model with three reduced models, respectively: first, the full model revealed a better fit than a model that took only experiment as a fixed effect (χ^2= 18.274, p<0.001), showing a significant effect of generation on number of constituents for the two data sets combined. Secondly, we compared the full model to a model that only included generation, and this revealed no significant difference in fit (χ^2= 0.416, p=0.812), indicating no observable difference between the two experiment with respect to the number of constituents overall. Thirdly, we compared the full model with a reduced model that included generation and experiment as fixed effects, but not their interaction, and again this revealed no significant difference in fit (χ^2= 0.4121, p=0.521), indicating no difference between the two experiments in how the number of constituents is affected by generation. See figure 11.

Correctness of the responses in the experiment was analysed in a binomial mixed effects regression, modelling correctness and taking generation as fixed effect. The model took random slopes for generation on chains and items. The model revealed a high proportion of correct responses in the first generation, as reflected in the intercept ($\beta = 1.386$, $SE = 0.173$, $p < 0.001$) and a main effect of generation ($\beta = 0.674$, $SE = 0.101$, $p < 0.001$).

To look at the differences between Experiments 1 and 2, we ran a model that included data from both, modelling correctness, and taking generation and experiment, and their interaction as fixed effects, applying the same random effect structure as above. Again, the model revealed a high proportion of correct responses initially ($\beta = 1.462$, $SE = 0.216$, $p < 0.001$), and a further increase over generations as reflected by a main effect of generation ($\beta = 0.5798$, $SE = 0.102$, $p < 0.001$). The model did not reveal a significant effect of experiment $\beta = 0.549$, $SE = 0.332$, $p > 0.05$), and no significant interaction between generation and experiment (β =-0.2792, SE=0.150, p>0.05). See figure 12.

Word order

We looked at word order, and coded the usage of SOV for extensional events, and SVO order for intensional events as *natural*. We predicted that the proportion of natural word orders would decrease over rounds. We ran a logit mixed effects regression entering generation as fixed effect, and uncorrelated intercepts and slopes on generation for chain and item.[9] Participants were more likely than not

[9]Running the full random effect structure resulted in intercept-slope correlations being estimated at 1.

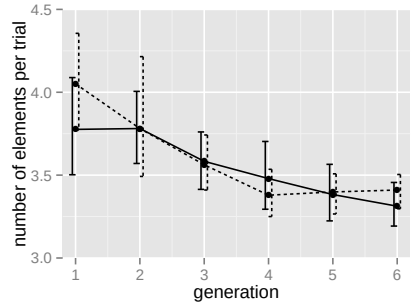

Figure 11: Number of elements per trial, by generation. For ease of comparison, the results of Experiment 1 are displayed as well; they are represented by the dashed line. The minimum value on the y-axis is 3, because that is the fewest possible elements to convey the necessary information in one trial. Participants became more efficient over time in a similar way to generations in Experiment 1.

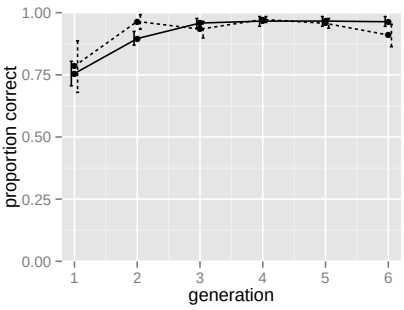

Figure 12: Proportion of correct answers per chain, by generation. Dashed line presents results of Experiment 1, for comparison. Participants were at or near ceiling throughout, except for the first generation, similarly to Experiment 1.

to produce natural orders in the first round, as reflected in the model intercept (β = 1.348, SE =0.301, p <0.001) and the proportion of natural orders went down over rounds, as reflected in a significant effect of round (β = -0.1233, SE =0.050, p <0.05).

In order to assess the differences between Experiment 1 and 2, we ran a combined model, focusing on proportions of natural word order, using generation and experiment as fixed effects, and the same random effect structure as above. We found a main effect of generation (β = -0.130, SE =0.050, p <0.01), no effect of experiment (β = 0.271, SE =0.288, p <0.347), and no interaction between generation and experiment (β = -0.098, SE =0.063, p <0.123); see figure 13.

To explore how the proportion of SOV word order developed over generations, we ran a logit mixed effects regression, entering round as a fixed effect, and random slopes for generation on items and chains (slope and intercept on chain were uncorrelated, because in the full model these were estimated as a correlation of 1). The model reveals that the proportion of SOV starts off lower than that of SVO in round 1, as reflected in the model intercept (β = -1.347, SE =0.354, p <0.001), and gradually diminishes further over time, reflected as a significant effect of generation (β = -0.299, SE =0.086, p <0.001). Again, we assessed the differences between Experiment 1 and 2 by running a combined model, entering generation and experiment as fixed effects, and using the same random effect structure as the model above. The model revealed a main effect of generation (β = -0.244, SE =0.096, p >0.05), no effect of experiment (β = -0.535, SE =0.470, p >0.05), and no interaction between experiment and generation (β = -0.046, SE =0.141, p >0.05). See figure 14.

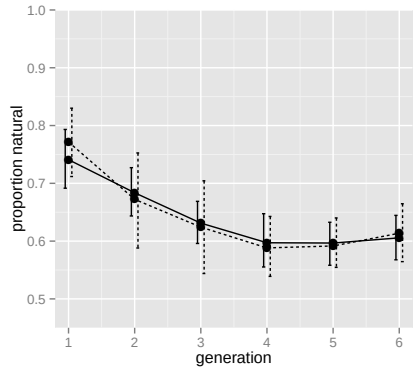

Figure 13: Proportion of natural gesture strings (SOV for extensional and SVO for intensional) plotted by generation. Participants started by producing natural word order, but this decreased over generations. Their behaviour is near identical to that in Experiment 1, showing that transmission has little effect on naturalness.

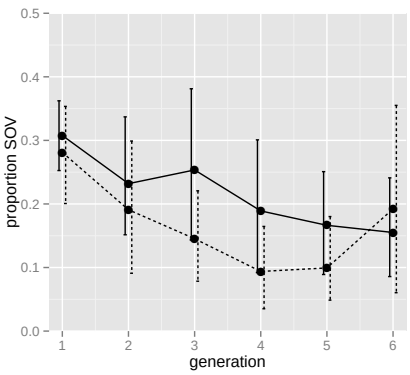

Figure 14: Proportion of strings in SOV order, by generation. Participants prefer SVO over SOV, and this preference increases over time, in a way similar to that in Experiment 1.

Regularisation

Like for Experiment 1, we calculated entropy scores per generation for each dyad, to measure whether there was regularisation of word order use over generations (see details above). We ran a linear mixed effects model that took generation as fixed effect and a random effect for chain (entering random slopes on generation resulted in non-converge). The model revealed a significantly better fit than a null model which did not include generation as fixed effect (χ^2=34.47, p<0.001). See figure 15.

To compare how entropy developed over generations between Experiment 1 and 2, we ran a combined model, entering experiment and generation as fixed effects, and chains as random effect. The model revealed a significantly better fit than a reduced model that only took experiment as a fixed effect (χ^2=47.772, p<0.001), but no significantly better fit than a reduced model that only took generation as a fixed effect (χ^2=0.9295, p=0.6283). See figure 15.

'Lineage specificity'

Like for Experiment 1, we were interested in whether we would find lineage specificity in our data. In this dyadic experiment, however, one cannot strictly speak of lineages, because there was no generational turnover. Instead, we calculated 'dyad specificity,' to assess if the pairs of participants who communicated together were more similar to each other than pseudo-pairs (formed by randomly combining the data of participants who were not in the same session). We calculated euclidean distances between participants on the basis of their proportions of SOV strings, SVO strings, and other orders in the same way as

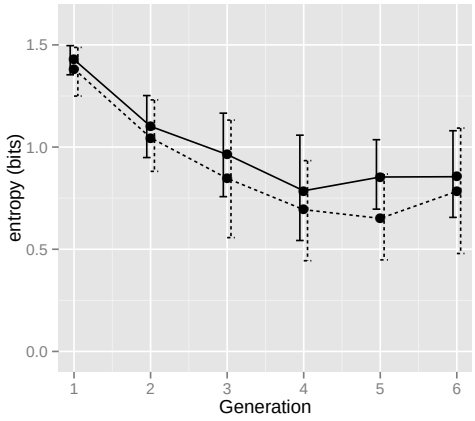

Figure 15: Entropy values plotted by generation. Error bars show 95% CI's. Entropy decreases significantly over generations.

we did for Experiment 1. We then ran a Monte Carlo simulation (1000 trials), in which we compared the actual distances to the distances of pseudo-pairs (participants were scrambled to form these pseudo-pairs), and found that the actual distances (mean = 14.50) were significantly smaller than the pseudo-pair distances (mean = 26.32, sd=3.61, z=7.205, p<.001), showing that there was indeed dyad specific behaviour in Experiment 2.

Structural priming

We ran a linear mixed effects model, looking at whether a given trial was SOV, and taking the previous trial (previous-SOV) as a fixed effect (contrast coded) in a linear mixed effects regression, that took uncorrelated random slopes for previous-SOV on chains and items. The model revealed a preference for SVO overall, reflected by the model intercept (β = -2.102, SE =0.400, p <0.001), but no significant increase in SOV when the previous trial was SOV (β = 0.342, SE =0.326, p =0.29). See figure 16. We then combined the data of Experiment 1 and 2 into one model, investigating SOV, and entering previous-SOV and experiment as fixed effects (both contrast coded), and the same random slope structure as the previous model. This model revealed an overall preference for SVO (β =-2.152, SE =0.309, p <0.001), and a main effect of previous-SOV (β =0.634, SE =0.235, p <0.01), no effect of experiment (β =-0.245, SE =0.465, p >0.05), and no interaction between previous-SOV and experiment (β =0.585, SE =0.424, p >0.05).

3.3 Discussion

In this experiment, we found that communication between partners in the dyads became more efficient and successful over time. Comparing Experiments 1 and 2, we found that speed and communicative success, and the way this behaved over generations, were statistically similar in the two experiments. This is interesting, given the fact that in Experiment 1, new participants were introduced, whereas

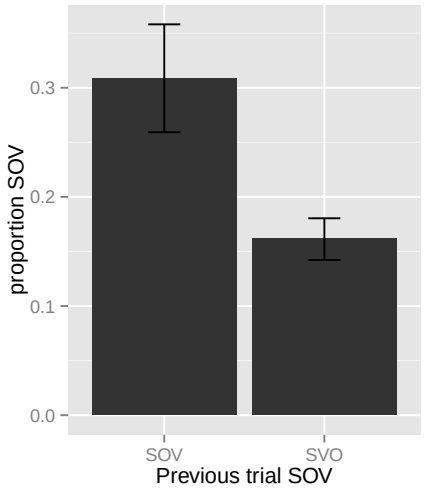

Figure 16: Comparing the proportions of SOV for two kinds of trials: when the word order in the previous trial was SOV, this resulted in a higher proportion of SOV compared to trials for which the previous order was SVO. However, this effect was not statistically significant. Error bars represent 95% confidence intervals.

in Experiment 2, the same two participants were interacting repeatedly. In other words, despite the actual individuals being replaced, and each individual consequently being involved for much less time, the population is behaving as if it were a single dyad in all respects.

Focusing on word order, we found that word order started out showing naturalness (SOV for extensional events, and SVO for intensional events), but this tendency decreased over time, in favour of a simpler word order regime (usage of the same word order irrespective of the semantic properties of an event), as confirmed in our entropy calculations over generation. Interestingly, when directly comparing these results to those of Experiment 1, we see no difference. This reveals the mechanisms involved in word order change. From Experiment 1, we might have concluded that interaction *and* transmission to new learners are mechanisms involved in word order simplification, but Experiment 2 tells us that transmission to new learners is not crucial in this process: when two participants are engaged in prolonged interaction, they will simplify their usage of word order in a similar way. We will discuss the comparison between Experiments 1 and 2 in further detail in the general discussion.

Further, we found that participants who communicated in a pair were more similar to each other (in their usage of SOV and SVO word orders) than pseudo-pairs (formed by scrambling the participants). This shows that the word order change was not simply the result of a universal preference to increase usage of SVO word order; participants who formed a dyad *did* influence each other's usage of word orders and different dyads changed their word order usage in different ways, and at different speeds (similarly to what was observed for chains in Experiment 1).

Zooming in on trial by trial behaviour of the participants allowed us to study the mechanisms behind word order change in interaction. Looking at structural priming in Experiment 2, we found no statistical priming effect, but when Experiment 1 and 2 were analysed together, there was a main priming effect, and no main effect of experiment was found. In order to analyse priming, and have enough statistical power to confirm priming behaviour statistically, the number of data points for both orders (SOV and SVO) needs to be sufficiently high. However, some of the participants in Experiment 1 and 2 converged relatively quickly on a word order regime with a high proportion of SVO, meaning that there wasn't a lot of word order alternation for certain chains or dyads. This may explain the fact that no statistical support for a priming effect was found in the Experiment 2 data alone, whereas it was found in the combined data set. We will discuss structural priming in further detail in the general discussion.

All in all, these two experiments show that several different preferences may play a role in the word order behaviour observed. First of all, there is a preference to use natural word orders at the initial stage of the experiment, but along the way this naturalness disappears. Apart from a pressure to use the same word order as the communicative partner (which was confirmed in priming effects), participants had a clear preference for SVO word order. This order is the dominant order of the native language of all participants (English), and to assess the strength of this preference, we conducted an additional experiment, in which the proportions of event types were skewed against SVO. We will discuss this in the next section.

4 Experiment 3: silent gesture and event frequency

In Experiment 1 and 2, the dominant word order at the end of the experiment was always SVO. There are two possible causes of this preference for SVO. The first is the fact that English was the native language of all participants; participants may simply tend to start using the dominant order of English (possibly because that is an order they know is shared between them). A second possible explanation comes from Marno et al. (2015) and ? (?). Both papers show that when participants have learnt a consistent lexicon of gestures, they prefer SVO order. If either of these accounts is true, one would expect SVO to become dominant, irrespective of the nature of the stimuli. However, Christensen et al. (2016) showed a general preference for SOV, in a dyadic interaction silent gesture experiment with native speakers of Danish (an SVO language). Christensen et al. (2016) further showed that changing the proportions of the stimuli changes the proportions of the word orders after interaction. They conclude that differences in frequency of referents may impact the structures that arise after interaction.

[XXX rewrite this bit]

We ran an experiment that investigated the issue of event frequency for the two mechanisms of cultural evolution that have been the focus of this paper: interaction and transmission. In our experiment we increase the relative proportion of extensional events, and because SOV is the preferred

4.1 Setup

4.1.1 Interaction and transmission: a skewed stimuli set

To assess how the proportion of word orders used in the interaction and transmission of silent gesture is affected by the type of information that is conveyed, we ran an additional experiment in which the stimuli set was skewed in favour of extensional events. Instead of equal proportions of intensional and extensional events, we used a majority of extensional (75%) and a minority of intensional events (25%): each generation consisted of 24 extensional and 8 intensional trials, and we made sure that each participant had the role of director for 12 and 4 of these respectively.

The experiment consisted of two conditions. The setup of the first condition (interaction&transmission) was the same as that of Experiment 1, running groups of 7 or 8 participants in a gradual turnover design, to implement interaction and transmission. Eight groups participants (all right-handed, native speakers of English, with no previous knowledge of any sign language) took part in this study (60 participants in total; mean age 20.6).[10] All participants were recruited at the University of Edinburgh (three groups were recruited from a first year Linguistics and English Language course; the remaining 5 were recruited from the University of Edinburgh career platform). The participants were paid between £5-10, depending on how long they were in the experiment for. The instructions and procedure of this first condition were identical those of Experiment 1.

The setup of the second condition (interaction only) was identical to that of Experiment 2 (except for the proportions of intensional and extensional events, which was identical to the first condition). Twelve pairs of participants (24 participants in total; mean age 21.5) were recruited at the University of Edinburgh recruitment web page, and paid £7.50 for their time.

4.2 Data analysis and results

Data analysis was done following the same procedure as that of Experiments 1 and 2. For the first condition, of 1536 trials in total (8 chains of 6 generations of 32 trials each) 760 were SOV, 595 were SVO, and the remaining 181 were other orders (resulting in 11.8% omitted data points). For the second condition, of 2304 trials in total (12 dyads of 6 generations of 32 trials each) 1061 were SVO, 955 were SOV, and 288 were other orders (resulting in 12.5% omitted data points).

Word order

To assess how the proportion of SOV was affected by generation, we ran a binomial mixed effects regression with generation and condition (the latter centered) as fixed effects, and random slopes for generation on chains and items. The model revealed no conclusive preference in the first generation ($\beta = 0.330$, $SE = 0.393$, $p > 0.05$) and a decrease of SOV over generations ($\beta = -0.314$, $SE = 0.111$, $p < 0.01$). See figure 17

[10] When only seven participants showed up for the experiment, a confederate acted as the observer in the last round (note that the observer had no active role in the experiment, and only watched the two communicators).

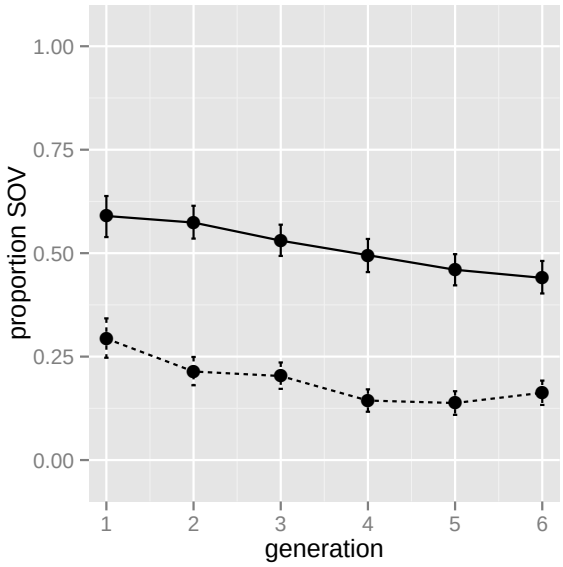

Figure 17: Proportion of SOV order over generations. Solid line shows Experiment 3, and dashed line shows Experiments 1 and 2. In experiment 3, SOV has a higher proportion throughout, but this proportion goes down over the generations, just like in Experiments 1 and 2.

Additionally, to compare how the proportion SOV develops over generation in the current experiment vs Experiments 1 and 2, we ran a binomial mixed effects regression on SOV, including generation, skewing (skewed stimuli set vs equal stimuli set), and mechanism (gradual turnover vs. dyadic interaction) and their interactions as fixed effects (with mechanism and skewing centered), and random slopes for generation on chains and items. The model revealed that there was an overall preference in favour of SVO initially (β =-0.683, SE =0.298, p <0.05), and that the proportion of SOV went further down over the generations (β =-0.275, SE =0.061, p <0.001). There was a main effect for skewing, indicating that the proportion SOV was higher overall in the skewed experiment (β =1.659, SE =0.475, p <0.001). See figure **??** for a visualisation of this effect. All other main effects and interactions (including that between skewing and SOV) were non-significant.

Regularisation

Like for the previous experiments, we calculated entropy scores per generation for each chain/dyad, to measure whether there was regularisation of word order use over generations. We ran a linear mixed effects model that took generation as fixed effect and a random effect for chain (entering random slopes on generation resulted in non-converge). The model revealed a significantly better fit than a null model which did not include generation as fixed effect (χ^2=39.49, p<0.001).

4.3 Discussion: event frequency and word order conventionalisation

In Experiment 3, there was a greater proportion of extensional events than in the previous two experiments, and we ran this experiment to test how word order regimes used by interacting and learning participants are affected by event frequency. Because SOV word order is preferred for extensional events in improvisation, we expected to see a higher proportion of SOV, and this was confirmed in the analysis. The way in which the proportion SOV developed over generations, was only marginally different from Experiments 1 and 2, however. This means we cannot conclude with certainty that having different proportions of events affects conventionalisation differently: this effect is overpowered by a general tendency to move towards SVO word order.

5 General Discussion

When people communicate in the absence of a shared system of linguistic rules, they will structure their utterances according to their semantic properties (Meir et al., 2010; Hall et al., 2013; Gibson et al., 2013; Schouwstra, 2017). Specifically, they prefer SOV word order for extensional, and SVO for intensional events, and this patterning emerges, probably, as a result of cognitive biases of the communicator, and represent natural ordering (Schouwstra & de Swart, 2014).

Conventional languages, on the other hand, do not condition word order on the type of event in the way silent gesture does. Instead they apply word order regimes that are more regular, and simpler. A crucial difference between improvised utterances and conventional languages is that the latter are the result of repeated communicative interaction, and transmission to new learners over many generations. We investigated the influence of these cultural evolutionary mechanisms by combining silent gesture with interaction and transmission in three laboratory experiments. This allowed us to assess the influence of these cultural mechanisms on word order change, and to look at the online mechanisms driving it in closeup.

5.1 The emergence of word order conventions: improvisation, interaction and transmission

We investigated what happens to the natural word order that comes up in improvisation, when silent gesture is used for communicative interaction, and transmitted to new learners. Experiment 1 implemented both interaction and transmission and showed that under these circumstances, word order becomes less natural over generations, leading to a general preference for SVO (but at the same time showing lineage specificity), and showing an effect of structural priming between participants. This finding contrasts with (Christensen et al., 2016), which presents an experiment in which participants use silent gesture for communicative interaction (see the description of this paper in section 1.3), and shows a statistical influence of structural alignment, this effect is only quite weak, and their observations 'do not point towards generalization and stabilization of one constituent order for both event types' (Christensen et al., 2016, p. 75). One might thus conclude that a combination of interaction and trans-

mission to new learners boosts conventionalisation of word order simplification. Something along similar lines is suggested by (Christensen et al., 2016, p. 77): *the frequent change of task-partners within a 'speech community' can radically enhance the conventionalization effect.* This conclusion would be consistent with previous observations that a combination of interaction and transmission to new learnerns—but not these factors individually—lead to language systems that are both expressive and compressible (Kirby et al., 2015), and exhibit systematic structure (Motamedi et al., Under revision).

To be sure there actually *was* an additional contribution of transmission on top of communicative interaction, we ran a second experiment, wich, like Experiment 1, involved communicative interaction of silent gesture, but with transmission taken out. Experiment 2 investigated what happened to natural word order when silent gesture was used in communicative interaction between two participants. This study did not introduce new learners; instead, two participants communicated with each other over 6 'generations'. The results of this experiment were remarkably similar to those in Experiment 1: word order usage changed at a similar pace, from natural (varying with meaning type) to more consistent over time.

In other words, as a system goes from being natural, and grounded in improvisation, it rapidly becomes more conventional and more regular through communicative interaction. The specific way in which naturalness disappears is a cultural process: the emerging conventions for word order are specific to the individuals creating them (as was shown by the lineage specificity result in Experiments 1 and 2). However, they are transferrable at the same time: novel learners can take over the process of conventionalisation (if they get a chance to observe the emerging system) and it will continue to take place.

Our finding that the pace and level of word order regularisation was the same for Experiments 1 and 2 is surprising, given previous results that indicate that a combination of interaction and transmission to new learners affects an emerging system differently from interaction alone. Especially, the contrast with the result described by (Motamedi et al., Under revision) is remarkable: in that study, a combination of interaction and transmission led to an increase in the systematic usage of lexical markers for different types of nouns (e.g. a person marker), while interaction alone led to some usage of markers, but less systematically. This finding and our finding combined suggest that different aspects of an emerging language are affected differently by cultural mechanisms like interaction and transmission. Specifically, word order regimes may be affected by interaction and less so by transmission, but the structure of the lexicon may need transmission to new learners in order to become fully systematic. This idea, that language is not a uniform phenomenon, and has different aspects that develop and evolve in different ways, should be studied in further detail in targeted experiments investigating the emergence of lexical and structural linguistic rules.

5.2 The emergence of word order conventions in closeup: structural priming

[Traditionally, work on structural priming focuses on language use in fully conventionalised languages. XXX expand this a bit?] For instance, Fehér et al. (2016) observe that 'from the point of view of communicative interac-

tion, sentence structure is somewhat independent of the propositional content of utterances—languages typically provide a number of structurally distinct means of conveying a given idea.' This situation, one in which semantic content can be conveyed in multiple structurally distinct ways, is seen as one that establishes the possibility of priming. In other words, the phenomenon of structural priming has been defined primarily in terms of an established system of linguistic conventions. This study shows that priming can play a role even when there are no conventions yet. This finding has consequences for the priming literature [XXX which?], but it offers opportunities for further investigation of priming as a driving force behind language emergence.

5.3 Predictions for and analogues in existing languages

It is clear from the results sketched here that indeed, usage of word order changes over time, and some pairs or chains of participants really end up using only one word order throughout the experiment, but this does not happen for all participants. In fact, averaging over the dyads and chains, we see that some level of naturalness still remains even at the end of the experiment. This observation suggests a pattern to be observed in newly emerged languages like Nicaraguan Sign Language and other sign languages: that we may find traces of natural word order (i.e., SVO order for intensional events, and SOV order for extensional events) in them, even in languages that do have conventions for word order.

Research into word order in signed languages is not without problems, e.g. in determining a so-called unmarked word order (Leeson & Saeed, 2012), and the potential influence of theoretical considerations on the coding type (Johnston, Vermeerbergen, Schembri, & Leeson, 2007), but spoken languages, to some extent, face these problems too, and fortunately, there has been a recent surge in collecting data on syntactic structure and word order information in signed languages (Napoli & Sutton-Spence, 2014). With a growing body of evidence, it has become possible to further investigate the precise circumstances of word order variation. Recently, word order patterns for intensional and extensional events have been investigated in two languages. (Napoli, Spence, & Quadros, 2017) studied LIBRAS (Brazilian Sign Language), and found that, even though LIBRAS has been classified as an SVO dominant language before, extensional events are likely to be described in SOV order (while for intensional events, SVO remains dominant).

A study of the same phenomenon in Nicaraguan Sign Language shows a slightly different, but essentially consistent picture (Flaherty, Schouwstra, & Goldin-Meadow, 2018). Nicaraguan Sign Language can be categorised as a strongly Verb-final language, with SOV being a dominant order, alongside with SVOV (Flaherty, 2014). NSL signers showed this verb-final preference quite consistently for both extensional and intensional events (i.e., very few SVO strings were observed). However, many strings contained repetitions of O and V, and when sub-strings were analysed, there turned out to be a significant difference between intensional and extensional events: V was more likely to be directly followed by an O for intensional events than for extensional events. Interestingly, this pattern was similar for first, second and third cohort NSL signers. In other words, NSL word order conventions seem not to have changed much, in this respect, over the course of generations, suggesting that transmission to new generations is not an essential driver of change for this aspect of language.

Observations from these two languages show that traces of natural word order behaviour can continue to exist in a fully conventionalised language; the case of NSL illustrates that there may be other word order conventions (in this case, strong V-finalness) that the natural pattern may interact with. Apart from the two languages mentioned here, we do not know whether or how languages (signed or spoken) encode extensional and intensional events differently.

6 Conclusion

The experiments in this paper offer a testing ground for investigating the emergence of linguistic conventions in the laboratory. Using the case of basic word order as a case study, we observed that participants behave according to a bias for naturalness initially, and alternate their word order by event type. Communicative interaction is a mechanism that drives this natural behaviour towards simpler word order regimes, and our experimental method enabled trial by trial analysis of word order usage, singling out structural priming as a potential driving mechanism behind word order simplification.

Surprisingly, transmission to new learners does not drive this process further. Our experiments thus show that different biases can play a different role, depending on the mechanisms that are dominant in a given situation (i.e. natural word order for improvisation, but a transition towards simpler word order regimes under the influence of interaction).

Appendix I

The line drawings that were used as stimuli in all experiments were tested for clarity on a crowdsourcing platform, on which 640 participants were each asked to give a single sentence description of one picture. We chose this setup, asking each participant to only describe one picture, to assess how well the pictures were understood under the least favourable circumstances, i.e. without the context of other, similar, pictures to help them along. We realised that in the actual experiment, the circumstances were like that only for the very first trial.

Those pictures that were clearly misinterpreted by more than 3 out of 10 participants in our picture description task were adapted and re-tested. For ten adapted pictures ten extra judgements were collected, obtaining at least 7 out of 10 correct responses for each.[11]

Appendix II

[Not sure I want this in the actual paper, but these graphs are informative.] Here I'm plotting the naturalness graphs with separate lines per dyad/chain, for Experiments 1 and 2 (in figure 18) and Experiment 3 (figure 19).

[11] Some of the actions were not described entirely accurately by all participants. For instance, some participants described the target action 'swing above head' as 'hold up', or 'want' as 'think of'. We included these slight misinterpretations as correct, as long as extensional events were interpreted as extensional, and intensional events as intensional.

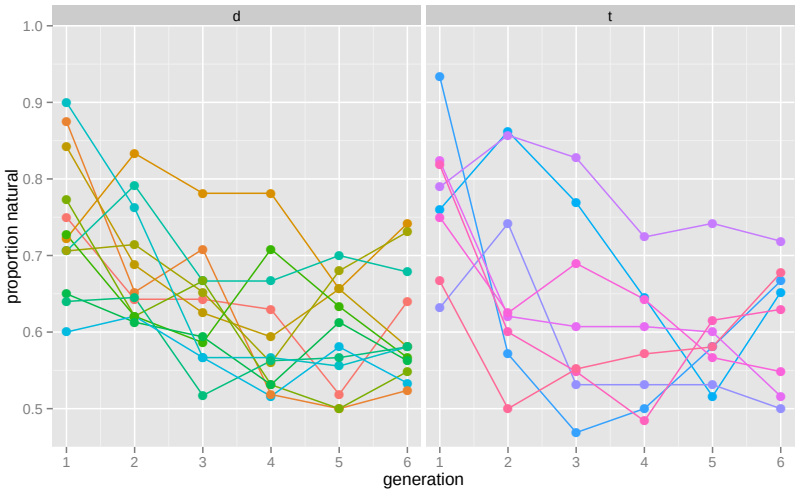

Figure 18: Naturalness over generations for Experiments 1 and 2, with separate lines for dyads/chains. Dyads on the left, chains on the right.

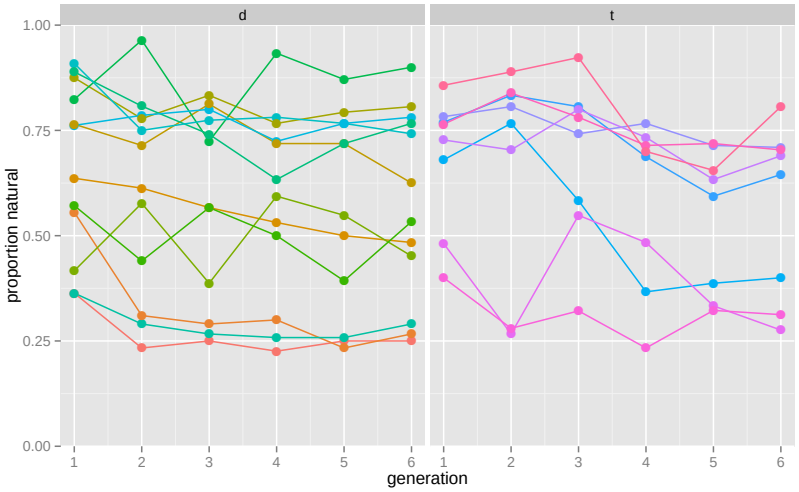

Figure 19: Naturalness over generations for Experiment 3, again, with separate lines for dyads/chains. Dyads on the left, chains on the right.

References

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75*(2), B13–B25.

Byun, K.-S., De Vos, C., Roberts, S. G., & Levinson, S. C. (2018). Interactive sequences modulate the selection of expressive forms in cross-signing. In *the 12th international conference on the evolution of language:(evolang xii)*.

Caldwell, C. A., & Smith, K. (2012). Cultural evolution and perpetuation of arbitrary communicative conventions in experimental microsocieties. *PloS one*, *7*(8), e43807.

Christensen, P., Fusaroli, R., & Tylén, K. (2016). Environmental constraints shaping constituent order in emerging communication systems: Structural iconicity, interactive alignment and conventionalization. *Cognition*, *146*, 67–80.

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, *122*(3), 306–329.

Dryer, M. S. (2013). Order of subject, object and verb. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online.* Leipzig: Max Planck Institute for Evolutionary Anthropology.

Dunn, M., Greenhill, S. J., Levinson, S. C., & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, *473*(7345), 79.

Fehér, O., Wonnacott, E., & Smith, K. (2016). Structural priming in artificial languages and the regularisation of unpredictable variation. *Journal of Memory and Language*, *91*, 158 - 180. (New Approaches to Structural Priming)

Flaherty, M., Schouwstra, M., & Goldin-Meadow, S. (2018). Do we see word order patterns from silent gesture studies in a new natural language? In C. Cuskley, M. Flaherty, H. Little, L. McCrohon, A. Ravignani, & T. Verhoef (Eds.), *The evolution of language: Proceedings of the 12th international conference (evolangxii).* NCU Press.

Flaherty, M. E. (2014). *The emergence of argument structural devices in nicaraguan sign language.* The University of Chicago.

Forbes, G. (2013). Intensional transitive verbs. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Fall 2013 ed.). Metaphysics Research Lab, Stanford University.

Futrell, R., Hickey, T., Lee, A., Lim, E., Luchkina, E., & Gibson, E. (2015). Cross-linguistic gestures reflect typological universals: A subject-initial, verb-final bias in speakers of diverse languages. *Cognition*, *136*, 215–221.

Galantucci, B., Garrod, S., & Roberts, G. (2012). Experimental semiotics. *Language and Linguistics Compass*, *6*(8), 477–493.

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of representation: where might graphical symbol systems come from? *Cognitive Science*, *31*, 961–987.

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of crosslinguistic word-order variation. *Psychological science*, 0956797612463705.

Goldin-Meadow, S. (2005). *The resilience of language: What gesture creation in deaf children can tell us about how all children learn language.* Psychology

Press.
Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *PNAS*, *105*(27), 9163–9168.
Gries, S. T. (2005). Syntactic priming: A corpus-based approach. *Journal of psycholinguistic research*, *34*(4), 365–399.
Hall, M. L., Ahn, Y. D., Mayberry, R. I., & Ferreira, V. S. (2015). Production and comprehension show divergent constituent order preferences: Evidence from elicited pantomime. *Journal of memory and language*, *81*, 16–33.
Hall, M. L., Mayberry, R. I., & Ferreira, V. S. (2013). Cognitive constraints on constituent order: Evidence from elicited pantomime. *Cognition*, *129*(1), 1–17.
Hawkins, J. A. (2014). *Word order universals.* Elsevier.
Johnston, T., Vermeerbergen, M., Schembri, A., & Leeson, L. (2007). Real data are messy': Considering cross-linguistic analysis of constituent ordering in auslan, vgt, and isl. *TRENDS IN LINGUISTICS STUDIES AND MONOGRAPHS*, *188*, 163.
Kirby, S. (2016). *Videobox: video recording, streaming and mirroring for experiments [computer software]* (Tech. Rep.). Centre for Language Evolution, University of Edinburgh.
Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory. *PNAS*, *105*(31).
Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.
Kocab, A., Lam, H., & Snedeker, J. (2017). When cars hit trucks and girls hug boys: The effect of animacy on word order in gestural language creation. *Cognitive science*.
Kocab, A., Senghas, A., & Snedeker, J. (2016). The emergence of temporal language in nicaraguan sign language. *Cognition*, *156*, 147–163.
Leeson, L., & Saeed, J. (2012). *Irish sign language: A cognitive linguistic approach.* Edinburgh University Press Edinburgh, UK.
Levelt, W. J., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive psychology*, *14*(1), 78–106.
Marno, H., Langus, A., Omidbeigi, M., Asaadi, S., Seyed-Allaei, S., & Nespor, M. (2015). A new perspective on word order preferences: the availability of a lexicon triggers the use of svo word order. *Frontiers in psychology*, *6*.
Meir, I., Aronoff, M., Börstell, C., Hwang, S.-O., Ilkbasaran, D., Kastner, I., Lepic, R., Ben-Basat, A. L., Padden, C., & Sandler, W. (2017). The effect of being human and the basis of grammatical word order: Insights from novel communication systems and young sign languages. *Cognition*, *158*, 189–207.
Meir, I., Lifshitz, A., Ilkbasaran, D., & Padden, C. (2010). The interaction of animacy and word order in human languages. In A. Smith, M. Schouwstra, B. de Boer, & K. Smith (Eds.), *The evolution of language* (pp. 455–456). Singapore: World Scientific.
Micklos, A. (2016). Interaction for facilitating conventionalization: Negotiating the silent gesture communication of noun-verb pairs. In *11th international conference on the evolution of language (evolang xi).*

Motamedi, Y., Schouwstra, M., Smith, K., Culbertson, J., & Kirby, S. (Under revision). *Evolving artificial sign languages in the lab: from improvised gesture to systematic sign.* (Cognition)

Napoli, D. J., Spence, R. S., & Quadros, R. M. de. (2017). Influence of predicate sense on word order in sign languages: Intensional and extensional verbs. *Language*, *93*(3), 641–670.

Napoli, D. J., & Sutton-Spence, R. (2014). Order of the major constituents in sign languages: Implications for all language. *Frontiers in psychology*, *5*, 376.

Newmeyer, F. J. (2000). On the reconstruction of 'proto world' word order. In C. Knight, J. R. Hurford, & M. Studdert-Kennedy (Eds.), *The evolutionary emergence of language.* Cambridge: Cambridge University Press.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: a critical review. *Psychological bulletin*, *134*(3), 427.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(02), 169–190.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), 329–347.

Schouwstra, M. (2017). Temporal structure in emerging language: From natural data to silent gesture. *Cognitive Science*, *41*(S4), 928-940.

Schouwstra, M., & de Swart, H. (2014). The semantic origins of word order. *Cognition*, *131*(3), 431–436.

Scott-Phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. *Trends in Cognitive Sciences*, *14*(9), 411–417.

Senghas, A., & Coppola, M. (2001). Children creating language: How nicaraguan sign language acquired a spatial grammar. *Psychological science*, *12*(4), 323–328.

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*(3), 444–449.

Tamariz, M., & Kirby, S. (2015). Culture: copying, compression, and conventionality. *Cognitive science*, *39*(1), 171–183.

Tamariz, M., & Kirby, S. (2016). The cultural evolution of language. *Current Opinion in Psychology*, *8*, 37–43.

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

Theisen-White, C., Kirby, S., & Oberlander, J. (2011). Integrating the horizontal and vertical cultural transmission of novel communication systems. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 33).

Thompson, B., Schouwstra, M., & Swart, H. de. (2016). Interpreting silent gesture. In S. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Fehér, & T. Verhoef (Eds.), *The evolution of language: Proceedings of the 11th international conference (evolangx11).* Online at `http://evolang.org/neworleans/papers/94.html`.