

Chapter 9

Linguistic Frankenstein, or How to test universal constraints without real languages

Natalia Levshina

Leipzig University

The scarcity of diachronic data represents a serious problem when linguists try to explain a typological universal. To overcome this empirical bottleneck, one can simulate the process of language evolution in artificial language learning experiments. After a brief discussion of the main principles and findings of such experiments, this paper presents a case study of causative constructions showing that language users have a bias towards the efficient organisation of communication. They regularise their linguistic input such that more frequent causative situations are expressed by shorter forms, and less frequent situations are expressed by longer forms. This supports the economy-based explanation of the universal form-meaning mapping found in causative constructions of different languages.

1 Problems with testing functional explanations

Functional linguists have formulated many universal principles that are meant to explain the structure and use of human languages, such as the principles of economy, iconicity, cognitive complexity, minimization of domains, avoidance of identity, and so on (e.g. Haiman 1983; Rohdenburg 1996; 2003; Hawkins 2004; Haspelmath 2008). How can one decide which explanation is relevant for a certain cross-linguistic pattern and how does one make sure that the latter is not a result of a historical coincidence in the sense of Collins (2019 [this volume])? Ideally, we would need data from genealogically and geographically diverse languages over a large time span. Needless to say, this is unrealistic: as a rule, such



data are not available. The time depth and typological breadth of available diachronic data are very limited. Moreover, even in an ideal world where any kind of linguistic data is obtainable at the click of a button, this might still be insufficient. First, real language data are observational, which makes a causal interpretation of the correlational results rather difficult (this does not mean there are no successful attempts, e.g. Moscoso del Prado 2014). Second, real language is a battleground of various forces, many of which can be mutually exclusive, e.g. over- and underspecification, iconicity and economy-driven arbitrariness, and so on. Disentangling these factors in real ‘messy’ language data is not a trivial task. Moreover, as pointed out by Smith et al. (2017) in their discussion of the universal bias against free variation, transmission of language in populations can mask the biases of learners: the language in a population might retain variability even though every learner is biased against acquiring such variation. Unless the data contain meta-information about the speakers, these effects may go undetected.

These problems can be solved with the help of the artificial language learning paradigm, which has gained popularity recently. One can observe in real time how linguistic systems undergo change, revealing the cognitive and communicative biases of language users. One can control for some factors while testing those of interest, and study the behaviour of each individual speaker within a population. Similar to the protagonist of Mary Shelley’s gothic novel, Victor Frankenstein, who created a sentient living creature in his laboratory, a linguist can design a new language and watch it develop.

Moreover, there have been quite a few experiments that put to test typological universals, such as Greenberg’s Universal 18 about harmonic word order within the NP (Culbertson et al. 2012), the suffixing preference (St. Clair et al. 2009), definiteness hierarchy (Culbertson & Legendre 2011) or the bias towards consistency in head-dependent order (Christiansen 2000). In the present paper, however, I will focus on the experiments that demonstrate more abstract functional and learning biases, which, in their turn, can be used to explain language universals and language-specific phenomena. An overview of the main principles and discoveries of artificial language learning with human subjects is provided in §2. To illustrate the approach, I will also present the results of a recent study, which tests the principle of economy on artificial causative constructions (see §3). A brief summary and outlook are provided in §4.

2 The artificial language learning paradigm

2.1 Main types of artificial language learning experiments

There are several popular types of artificial language learning experiments (see Figure 1). First of all, learning can be iterated and non-iterated. In non-iterated learning, one can only study the individual process of acquisition. There is no further language transmission. In iterated learning, a subject learns a certain linguistic behaviour by observing the behaviour of one or more subjects who learnt it the same way, i.e. in the process of implicit induction and production (Kirby et al. 2014). The output of one generation of speakers serves as the input for the next one, similar to the transmission of real language and culture in general.

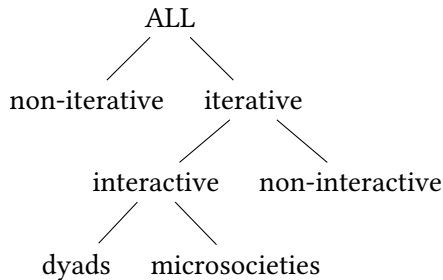


Figure 1: Main types of artificial language learning experiments.

Some communicative and learning biases may be strong enough to be detected in non-iterated learning. Sometimes even one generation is enough to radically change the language (Hudson Kam & Newport 2009). Weaker biases may require several generations in order to manifest themselves (e.g. Real & Griffiths 2009; Smith & Wonnacott 2010).

Iterated learning can be further subdivided into interactive and non-interactive (cf. Tamariz 2017). In non-interactive designs, one creates transmission chains where one subject's output is another subject's input. There is no actual interaction between the subjects. No common ground is created, and no feedback is given. Interactive experiments involve dyads of interacting users or even microsocieties, where everyone interacts with everyone else (Tamariz 2017). Language is transferred from one dyad to the following one, or from old members of a microsociety to the new ones. By using this approach, one can preserve common ground and feedback, which are crucial in everyday communication (Caldwell & Smith 2012).

The artificial language learning paradigm is very flexible, allowing for investigation of diverse forms: non-existent words, whistles, graphical scribbles. A language can also be fully artificial or semi-artificial. For instance, Smith & Wonnacott (2010) use some lexical items (nouns) from English, but novel verbs and plural noun markers. Usually, it is assumed that the results based on various media are comparable, although some recent studies suggest that the role of universal constraints (e.g. iconicity and compositionality) varies across different media (e.g. manual signs vs. sounds in Little et al. 2017).

Crucially, the studies based on artificial language learning share one fundamental assumption. Namely, those linguistic features that are easier to learn and use in communication will spread at the expense of less “fit” alternatives (Smith et al. 2017). By adjusting the linguistic input in a similar way, language users reveal their communicative and learning biases, which are so strikingly similar that one can speak about universal preferences.

2.2 Evidence of universal constraints from artificial language learning

The main results of recent studies can be concisely and non-exhaustively presented in a list of the following universal biases:

1. A bias towards **arbitrariness** (as opposed to iconicity), conventionalization and simplification of signs in interaction (e.g. Caldwell & Smith 2012). Simplified arbitrary signs are easier to select, produce and replicate than more complex iconic signs. At the same time, symbolic signs are more difficult to learn at first encounter, while iconicity seems to enhance the learnability of signs for new group members, as shown by Fay & Ellison (2013). They also found that the semiotic systems of larger populations reach a kind of a compromise: they favour simple iconic signs, i.e. those that are minimally complex and maximally informative/informativity.
2. A bias towards **combinatorial structure**, when meaningless elements (which serve as basic building blocks) are combined in higher-order units. This is also known as duality of patterning (Verhoef 2012).
3. A bias towards **compositional structure** of syntax (Kirby et al. 2008). During the process of iterative learning, language becomes more structured.
4. A bias towards **discrete structure** as opposed to holistic signals. For example, in an iterated language learning experiment with a language based on whistles, participants come up with categorical distinctions, rather than

paying attention to the precise acoustic realizations, e.g. in terms of pitch (Verhoef 2012).

5. A bias towards **regularity**. Languages exhibiting free variation become increasingly regular, revealing a strong bias towards regularity in adult learners (Smith & Wonnacott 2010). This bias may be obscured by so-called probability matching: in a language in which two forms are in free variation, adult learners have also been found to produce each variant in accordance with its relative frequency in the input (Hudson Kam & Newport 2009; Wonnacott & Newport 2005). The interplay between regularization and probability matching depends on the frequency distribution. The more forms with lower frequencies are used as free variants of the main form, the more scattered the pattern and the stronger the bias towards production of the main form (Hudson Kam & Newport 2009);
6. A bias towards **economy and communicative efficiency**, when more predictable information gets less formal coding, and less predictable information gets more formal coding. This bias has been observed in a study of differential case marking (Fedzechkina et al. 2012). The hypothesis is that a referential expression should be more likely to receive overt case marking when its intended grammatical function is less expected. The experiment shows that learners deviate from the initial input to make the language more communicatively efficient;
7. A bias towards **underspecification** of irrelevant conceptual dimensions. Silvey et al. (2015) have found that their artificial language, which was originally fully specified in the sense that it had a unique label for each object, became underspecified by losing contrasts across irrelevant dimensions, i.e. those that are not important for discriminating between the stimuli. In contrast, Tinitis et al. (2017) found a bias towards overspecification and redundancy in the contexts when the relevant dimensions were difficult to discern.

A key question is whether these biases are due to higher learnability or communicative advantages of the preferred features, or both. Using the terms from Haspelmath (2019 [this volume]), are we dealing with acquisitional or functional-adaptive constraints?

It is clear from the existing evidence that more learnable systems are not necessarily the ones that are also more usable, and the other way round. As was shown above, arbitrary signs, which are more usable in interaction, are less learnable

than more iconic ones. One can find a similar clash between learnability and usability with regard to regularization and underspecification. As found by Kirby et al. (2008), Verhoef (2012) and others, languages that are more regular and compositional are easier to learn and are more successfully passed from one generation to another. Their studies demonstrate that the learning errors decrease with time (number of generations), as compositionality and regularity increase. At the same time, such emerging systems also exhibit greater ambiguity because the number of lexical items drops. As a result, the languages become increasingly underspecified, which would reduce their usability.

Interestingly, it has been claimed that children tend to regularize, or systematize more strongly than adults (Hudson Kam & Newport 2009). This finding has been attributed to children having less cognitive resources than adults – in particular, memory limitations. However, Smith et al. (2017) do not find this argument very convincing because, as they claim, memory limitations do not always lead to regularization. Alternatively, one may suppose that adult learners may be better at conforming to social expectations and norms. In general, there are important differences in the emergent languages depending on the social circumstances of communication. For instance, Perfors (2016) observes that adults regularize strongly when they believe that the variation is unpredictable (i.e. they are told that the previous person was under time pressure and might have made a few errors), than when they are asked to match an imaginary output of another person, who is believed to be performing the same naming task at the same time. When the participants believe that the variation is predictable (even if they do not know what it actually depends on), and their goal is to learn the language as closely as possible, they do probability matching more and regularize less. There is also evidence that speakers produce more regular language when they believe they are addressing a person, even though they are in fact communicating with a computer (Fehér et al. 2016). Apparently, speakers believe that producing a more regular language will facilitate communication with their human partner. Similarly, Little (2011) discovered that morphosyntactic complexity decreases when expert participants, who have been trained in an artificial language, interact with naïve ones who have little knowledge of the same language. This effect, however, was not observed when the experts interacted with other experts. Thus, the emergent language system depends on the social circumstances and pragmatic goals of the subjects. This relationship, which seems to be present already at the learning stage, makes a neat separation of acquisitional and functional-adaptive constraints a very challenging task.

In the remaining part of the paper, I will focus on the bias towards communicatively efficient, economical form-meaning mapping, using a non-iterative online experiment.

3 Case study: Frequency effects in causative constructions

3.1 Hypothesis: Economy and formal length

As was shown above, there is ample evidence that language learners are generally sensitive to frequency information. In this case study, I focus on the claim that more frequent situations are expressed by means of less coding material than less frequent ones. Such differences are predicted by the principle of economy and – broadly speaking – the principle of communicative efficiency. According to these principles, more predictable information needs less coding material than more predictable information. The experiment in Fedzechkina et al. (2012), which was mentioned above, demonstrated the effect of predictability based on semantic categories. In my own study, I want to focus on predictability based on frequency information. To the best of my knowledge, these effects have not been tested previously in artificial language learning experiments.

Causatives serve as convenient and well-studied material for testing the bias in question. There is a cross-linguistic correlation between form and meaning: more formally integrated causatives, such as lexical causatives *kill* or *break_{TR}*, tend to denote more integrated causing and caused events than less integrated forms, such as *cause to die* or *make break_{INTR}*. As Comrie (1981: 165) puts it, “the kind of formal distinction found across languages is identical: the continuum from analytic via morphological to lexical causative correlates with the continuum from less direct to more direct causation”. Consider the example in (1):

- (1) English (personal knowledge)
- a. John killed Bill in his mansion on Tuesday...
 - i. ...?? by shooting him in the forest on Monday.
 - ii. ... ?? by tampering with his gun.
 - b. John caused Bill to die in his mansion on Tuesday...
 - i. ... by shooting him in the forest on Monday.
 - ii. ... by tampering with his gun.

In this example, the lexical causative *kill* expresses direct causation with high spatiotemporal integration of the causing and caused events (John’s killing and Bill’s

dying, respectively) and with direct impact of the Causer (*John*) on the Causee (*Bill*), whereas the analytic causative (*cause to die*) expresses indirect causation without spatiotemporal integration of the events and without direct impact of the Causer. This correlation between conceptual and formal integration of events has also been found in a large typologically diverse sample of languages (Levshina 2018a).

Haspelmath (2008) suggests an alternative account of this correlation based on the principle of economy: more frequent forms are usually shorter, whereas less frequent ones tend to be longer. As my current corpus-based work shows (Levshina 2018b: Ch. 3), the frequencies of direct causation and related properties (e.g. lack of autonomy on the part of the Causee, implicative causation, factitive causation, etc.) are substantially higher than those of indirect causation. Thus, these parameters (conceptual integration, formal compactness and relative frequency) are intercorrelated: more compact causatives represent both more frequent situations and more integrated events, whereas less compact causatives represent less frequent situations and also less integrated events. This creates a situation in which it is very difficult to decide based on observational data alone which of the functional principles actually explains the cross-linguistic correlation between formal and conceptual integration, i.e. iconicity or economy. The purpose of the present study is to test whether the economy effect is still observed when the iconic correspondence is not present.

3.2 Design and procedure

The participants of the experiment were asked to learn an alien language. At the beginning, they read an introduction:

In this experiment you will learn the lingua franca of a highly developed civilization that exists on a planet in a galaxy far, far away... The planet is called Atruur. Its only vegetation form is called 'grok'. It is similar to a cactus and is used by the Atruurians for food, as fuel for their flying vehicles and for entertainment. Because the Atruurians traditionally detest any form of physical activity, they have developed a technology for teleportation and telekinesis.

The introduction also mentioned that the word order is SV (for intransitives) or SOV (for transitives). To explain that to non-linguists, examples were provided, which are shown below for illustration:

- (2) Grok babum.
 cactus grow
 ‘A grok (cactus) grows.’
- (3) Sia grok hum.
 Atruurian cactus see
 ‘An Atruurian sees a grok (cactus).’

The subjects were first asked to learn the language by copying the sentences in Atruurian that describe situations shown in video clips. At first, they saw four situations: a cactus-like plant appears, disappears, grows and shrinks in size. The goal of that task was to introduce the basic vocabulary.

Next, the participants saw 32 causal situations, which represented a causal version of the same situations. In each of these causal situations, there was a flying saucer (sometimes with an alien inside) which hovered above the plant and flashed a yellow or blue light three times in a row. As a result, the plant either appeared, disappeared, grew or shrunk. Varying types of saucers were shown.

Crucially, the subjects saw two types of causing events. The first of them involved the saucer flashing a yellow light above the plant. The other one was when the saucer flashed a blue light from the left of the plant. The yellow-light causing event was three times as frequent as the blue-light causing event (i.e. 75% vs. 25%). The distribution of the four caused events was the same for each of the causing events. There were no reasons to assume that one type of causation is more or less direct than the other. The colour and the position of the Causer with regard to the Causee are not mentioned in the semantic parameters that are distinctive of different causative constructions in the languages of the world (Levshina 2018b: Ch. 3).

As for the artificial language, the most important thing is that each causing event is represented by two allomorphs. One of the causing events was associated with the forms *tere-* or *te-*, as in (4), and the other one was described by using the forms *gara-/ga-*.

- (4) a. Sia grok te-babum.
 Atruurian plant CAUS-grow
 ‘The Atruurian caused the plant to grow (by flashing with yellow light from above).’

- b. Sia grok tere-babum.
Atruurian plant CAUS-grow
'The Atruurian caused the plant to grow (by flashing with yellow light from above).'

Note that the form-meaning mapping varied across the subjects. That is, for some of them, *te-/tere-* denoted the causing event with yellow light flashed from above, whereas the *ga-/gara-* forms were used for the causing event with blue light flashed from the left of the plant. For the others, this was the other way round. The prefixes were evenly distributed among the stimuli, so that there was truly free variation. There was no condition in the experimental design that could explain the preference for the longer or the shorter form.

One should mention here that free variation is less exotic than it seems. It occurs in the language of late learners of a second language, e.g. hearing parents of a deaf child who learn to sign, or during the emergence of a new language, e.g. Tok Pisin and other pidgins and creoles (see an overview in Hudson Kam & Newport 2009). This is why the input language is not completely outlandish from a functional point of view. However, since language users have a bias towards regularization and against free variation, I expected that the subjects would regularize the free variation in the input, preferring the short allomorphs to convey the frequent causing events, and using the long allomorph to express the rare causing events.

After the training session, the subjects were asked to describe in Atruurian what is going on in video clips. The stimuli represented a selection from the previous stimuli: each of the caused events was presented with causing event A and causing event B. In total, there were eight test situations.

The experiment was performed online, using *Google Forms* with built-in *YouTube* videos. The latter were created with the help of *Adobe Animate CC* software by myself. Figure 2 demonstrates four fragments from one of the video clips, with the causing event A and the caused event of disappearance.

3.3 Participants

The participants were recruited via my personal network and LinguistList. Most of them had a background in linguistics or languages. After the experiment, they were asked about the aims of the experiment. None of them guessed the true purpose. Overall, I obtained responses from 84 participants. Some of the responses were removed. This was the case if the participants did not follow the training procedure instructions (e.g. a participant did not type in the training sentences),

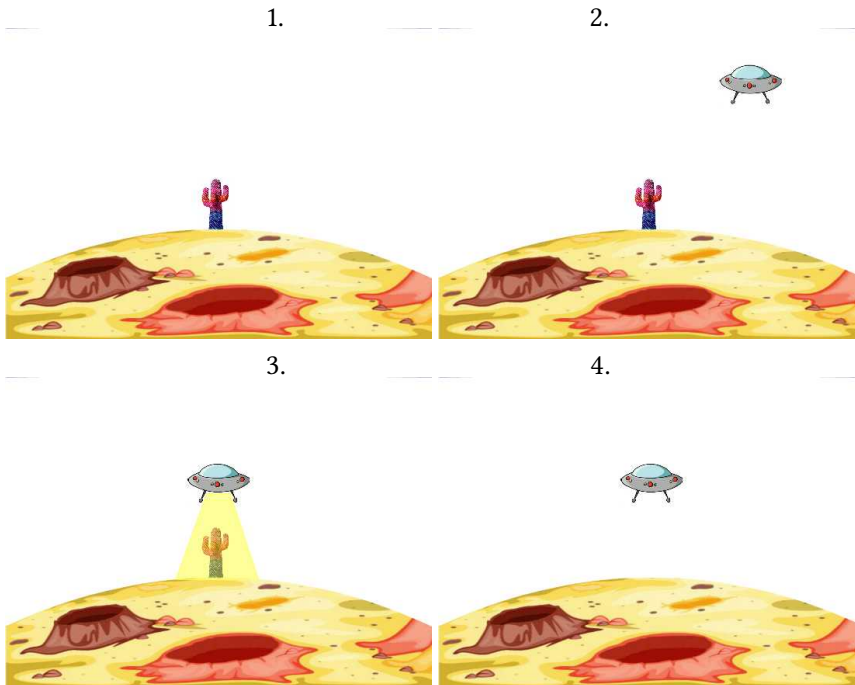


Figure 2: Fragments from one of the video clips.

or if the output was unanalysable. As a result, I had 554 valid data points from 70 participants.

The participants with valid responses had different L1s, but mostly had a Slavic and Germanic linguistic background. There were 40 native Czech speakers, 12 native German speakers, 7 native English speakers, 2 Dutch speakers, 2 Italian speakers, as well as native speakers of Brazilian Portuguese, Croatian, Danish, Polish, Russian, Slovak and Turkish. None of these languages has productive causative prefixes.

3.4 Results of the artificial language learning experiment

The counts aggregated across all participants are presented in Table 1. Lexical and spelling errors were ignored. Figure 3, which visualizes these counts, shows that there is a difference between the proportions of short and long forms expressing the frequent and rare causing events. The short forms are overall more preferred than the long ones, but the situations with the more frequent causing event are more frequently expressed by the short forms in comparison with the situations

that involve the rare causing event, where the proportions of the short and long forms are almost equal.

Table 1: The number of forms selected and their marginal sums.

	Frequent	Rare	Total
Short	168	137	305
Long	109	140	249
Total	277	277	554

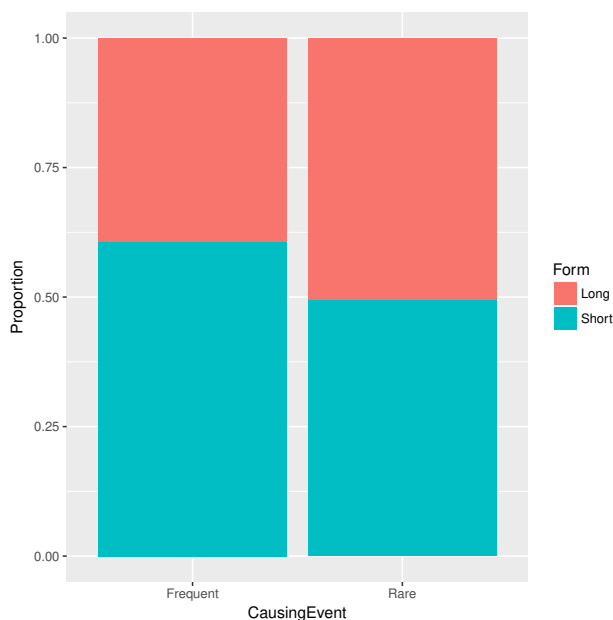


Figure 3: Proportions of short and long forms.

A closer look at the individual subjects' preferences reveals that most of them use both long and short forms. Seven subjects produced only the short forms. There were no subjects who always preferred the long forms. The distribution is shown in Figure 4.

The main question, however, is whether the choice of forms is influenced by the type of causing event. In order to test this, I fit a generalized linear mixed-effects model with logit as the link function (R package *lme4*, function *glmer*,

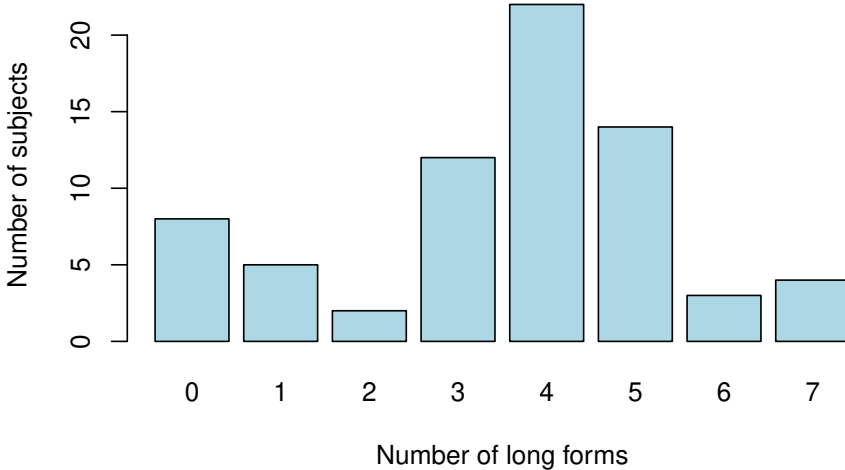


Figure 4: Individual preferences for the long and short forms

Bates et al. 2015). The type of prefix – long or short – was the response variable. The individual participants were treated as random effects (intercepts). There is a significant effect of the type of causing event: if the event is rare, the odds of the longer form to be chosen are 1.66 times greater than when the event is frequent (log-odds ratio $b = 0.501$, $p = 0.006$). This result supports the hypothesis that speakers have a bias towards the use of shorter forms to represent more frequent situations, and longer forms to represent less frequent situations. Random slopes, which represented individual differences in the effect of the predictor on the response, were tested as well, but they did not improve the explanatory power of the model.

The likelihood ratio test, a standard tool for variable selection and model comparison in regression analysis, demonstrates that the caused event does not have a significant effect on the choice of form ($p = 0.84$), and does not interact with the type of causing event ($p = 0.6$). This means that lexical conditioning can be excluded (cf. Smith & Wonnacott 2010).

4 Discussion

This paper has provided an overview of the applications of the artificial language learning paradigm in testing universal biases suggested by functional and cognitive linguists. One of them, known as the principle of economy, was tested in an online experiment. The results demonstrate that frequent causative situations are more commonly expressed by shorter forms, whereas the subjects are more tolerant of longer forms when expressing rarer causative situations. Therefore, the results of previous corpus-based studies, typological evidence and experimental approaches converge. The fact that the effect was detected in a non-iterative experiment with only one “generation” of language learners, suggests that the bias is very strong.

An important question remains about the nature of this bias and its place in Haspelmath’s (2019 [this volume]) classification. Can it be characterized as a functional-adaptive, acquisitional, mutational or maybe even representational constraint? The mutational type can be discarded because we do not have any qualitative changes in the constructions (e.g. possessed nouns becoming adpositions). As for representational constraints, they reflect the properties of the innate language faculty. Even if we accept that economy is an innate principle, since humans and other species are genetically programmed to gain maximal benefits from their behaviour at minimum costs (cf. Parker & Smith 1990), it represents a domain-general bias that is not restricted to human language only. There is evidence that the evolution of sense organs and brains is driven by the need to minimize the energy spent for each bit of information received from the environment (Stone 2015: 3). This is why linguistic economy is not a part of Universal Grammar in the generativist sense.

Thus, we are left with the functional-adaptive and acquisitional types. It is true that Haspelmath (2019 [this volume]) defines the latter as related to L1 by children only, but the overview presented in §2 demonstrates that learnability constraints can also be detected in artificial language learning by children and adults. §2 also showed that a clear distinction between communicative efficiency and learnability is often problematic within the artificial language learning paradigm. Similar to Slobin’s (1996) famous “thinking for speaking”, we can also speak about “learning for using”. This makes the task of distinguishing between these types very difficult. My preliminary answer is that we are dealing with a functional-adaptive constraint because it helps to optimize communication, even though there is no immediate interaction in the experiment. Obviously, more research with a clearer separation between the learning and communication stages is needed. The most

pertinent question at this stage is the following: Is it easier to learn a more communicatively efficient language, in which frequent meanings are expressed by shorter forms, and rare functions are expressed by longer forms, than a less efficient one, in which frequent meanings are expressed by longer forms and rare ones by shorter forms?

The artificial language learning paradigm, as I tried to demonstrate in this paper, represents a valuable addition to the toolkit of typologists and functional linguists. However, there are a few caveats that need to be mentioned. First, the experiments involve very limited interaction, if any, in an artificial context. Second, the populations are extremely small. Third, even when a fully artificial language is used, one cannot exclude transfer effects from real language. For example, Goldberg (2013), in her critical evaluation of Culbertson et al.'s (2012) study of Greenberg's Universal 18, argues that the word orders Adj + N and N + Adj can be transferred either from English (e.g. *a blue bird* vs. *something red*, *all things nice*) or from the Spanish-type languages. Note, however, that Culbertson et al. studied a specific formal pattern. As we move to more abstract properties of language or communicative behaviour, such as compositionality or economy, it becomes more difficult to explain these properties by transfer from real languages, although one cannot exclude a possibility that these biases represent very abstract generalizations from the users' experience with language, and their intuitive expectations about what a "normal" human language should be like. This uncertainty will probably always loom until we find a real alien and have it learn an artificial language.

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° 670985). I'm extremely grateful to all colleagues who patiently participated in the experiment, helped to improve the design and to recruit the participants, especially to my good friends from Olomouc university, Michaela Martinkova and Marketa Janebova. All usual disclaimers apply.

References

- Bates, Douglas, Martin Maechler, Ben Bolker & Steven Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. DOI:10.18637/jss.v067.i01

- Caldwell, Christine A. & Kenny Smith. 2012. Cultural evolution and perpetuation of arbitrary communicative conventions in experimental microsocieties. *PLoS ONE* 7(8). e43807. DOI:10.1371/journal.pone.0043807
- Christiansen, Morten H. 2000. Using artificial language learning to study language evolution: Exploring the emergence of word order universals. In Jean-Louis Dessalles & Laleh Ghadakpour (eds.), *The evolution of language: 3rd international conference*, 45–48. Paris: École Nationale Supérieure des Télécommunications.
- Collins, Jeremy. 2019. Some language universals are historical accidents. In Karsten Schmidtke-Bode, Natalia Levshina, Susanne Maria Michaelis & Ilja A. Seržant (eds.), *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence*, 47–61. Berlin: Language Science Press. DOI:10.5281/zenodo.2583808
- Comrie, Bernard. 1981. *Language universals and linguistic typology: Syntax and morphology*. Chicago: University of Chicago Press.
- Culbertson, Jennifer & Géraldine Legendre. 2011. Investigating the evolution of agreement systems using an artificial language learning paradigm. In Dina Bailey & Victoria Teliga (eds.), *Proceedings of the 2010 Western Conference on Linguistics. Vol. 1*, 46–58. Fresno: Department of Linguistics, California State University.
- Culbertson, Jennifer, Paul Smolensky & Géraldine Legendre. 2012. Learning biases predict a word order universal. *Cognition* 122(3). 306–329. DOI:10.1016/j.cognition.2011.10.017
- Fay, Nicolas & T. Mark Ellison. 2013. The cultural evolution of human communication systems in different sized populations: Usability trumps learnability. *PloS ONE* 8(8). e71781. DOI:10.1371/journal.pone.0071781
- Fedzechkina, Maryia, T. Florian Jaeger & Elissa L. Newport. 2012. Language learners restructure their input to facilitate efficient communication. *PNAS* 109(44). 17897–17902. DOI:10.1073/pnas.1215776109
- Fehér, Olga, Elizabeth Wonnacott & Kenny Smith. 2016. Structural priming in artificial languages and the regularisation of unpredictable variation. *Journal of Memory and Language* 91. 158–180. DOI:10.1016/j.jml.2016.06.002
- Goldberg, Adele E. 2013. Substantive learning bias or an effect of familiarity? Comment on Culbertson, Smolensky, and Legendre (2012). *Cognition* 127(3). 420–426. DOI:10.1016/j.cognition.2013.02.017
- Haiman, John. 1983. Iconic and economic motivation. *Language* 59(4). 781–819. DOI:10.2307/413373

- Haspelmath, Martin. 2008. Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics* 19(1). 1–33. DOI:10.1515/COG.2008.001
- Haspelmath, Martin. 2019. Can cross-linguistic regularities be explained by constraints on change? In Karsten Schmidtke-Bode, Natalia Levshina, Susanne Maria Michaelis & Ilja A. Seržant (eds.), *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence*, 1–23. Berlin: Language Science Press. DOI:10.5281/zenodo.2583804
- Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press. DOI:10.1093/acprof:oso/9780199252695.001.0001
- Hudson Kam, Carla L. & Elissa L. Newport. 2009. Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology* 59(1). 30–66. DOI:10.1016/j.cogpsych.2009.01.001
- Kirby, Simon, Hannah Cornish & Kenny Smith. 2008. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *PNAS* 105(31). 10681–10686. DOI:10.1073/pnas.0707835105
- Kirby, Simon, Tom Griffiths & Kenny Smith. 2014. Iterated learning and the evolution of language. *Current Opinion in Neurobiology* 28. 108–114. DOI:10.1016/j.conb.2014.07.014
- Levshina, Natalia. 2018a. Finding the best fit for direct and indirect causation: A typological study. *Lingua Posnaniensis* 58(2). 65–83. DOI:10.1515/linpo-2016-0010
- Levshina, Natalia. 2018b. *Towards a theory of communicative efficiency in human languages*. Leipzig: Leipzig University (Habilitation thesis). DOI:10.5281/zenodo.1542857
- Little, Hannah. 2011. *Foreigner directed speech: Its role in the cultural transmission of language and the resulting effects on language typology*. Edinburgh: University of Edinburgh (MSc thesis).
- Little, Hannah, Kerem Eryilmaz & Bart De Boer. 2017. Signal dimensionality and the emergence of combinatorial structure. *Cognition* 168. 1–15. DOI:10.1016/j.cognition.2017.06.011
- Moscoso del Prado, Fermin. 2014. Grammatical change begins within the word: Causal modeling of the co-evolution of Icelandic morphology and syntax. In *Proceedings of the Cognitive Science Society*, vol. 36, 2657–2662.
- Parker, Geoffrey A. & J. Maynard Smith. 1990. Optimality theory in evolutionary biology. *Nature* 348(6296). 27–33. DOI:10.1038/348027a0
- Perfors, Amy. 2016. Adult regularization of inconsistent input depends on pragmatic factors. *Language Learning and Development* 12(2). 138–155. DOI:10.1080/15475441.2015.1052449

- Real, Florencia & Thomas L. Griffiths. 2009. The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition* 111(3). 317–328. DOI:10.1016/j.cognition.2009.02.012
- Rohdenburg, Günter. 1996. Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics* 7(2). 149–182. DOI:10.1515/cogl.1996.7.2.149
- Rohdenburg, Günter. 2003. Cognitive complexity and horror aequi as factors determining the use of interrogative clause linkers in English. In Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of grammatical variation in English*, 205–250. Berlin, New York: Mouton de Gruyter. DOI:10.1515/9783110900019.205
- Silvey, Catriona, Simon Kirby & Kenny Smith. 2015. Word meanings evolve to selectively preserve distinctions on salient dimensions. *Cognitive Science* 39(1). 212–226. DOI:10.1111/cogs.12150
- Slobin, Dan I. 1996. From “thought and language” to “thinking for speaking”. In John J. Gumperz & Stephen C. Levinson (eds.), *Rethinking linguistic relativity*, 70–96. Cambridge: Cambridge University Press.
- Smith, Kenny, Amy Perfors, Olga Fehér, Anna Samara, Kate Swoboda & Elizabeth Wonnacott. 2017. Language learning, language use and the evolution of linguistic variation. *Philosophical Transactions of the Royal Society B* 372(1711). 20160051. DOI:10.1098/rstb.2016.0051
- Smith, Kenny & Elizabeth Wonnacott. 2010. Eliminating unpredictable variation through iterated learning. *Cognition* 116(3). 444–449. DOI:10.1016/j.cognition.2010.06.004
- St. Clair, Michelle C., Padraic Monaghan & Michael Ramscar. 2009. Relationships between language structure and language learning: The suffixing preference and grammatical categorization. *Cognitive Science* 33(7). 1317–1329. DOI:10.1111/j.1551-6709.2009.01065.x
- Stone, James V. 2015. *Information theory. A tutorial introduction*. Wrocław: Septel Press. DOI:10.13140/2.1.1633.8240
- Tamariz, Monica. 2017. Experimental studies on the cultural evolution of language. *Annual Review of Linguistics* 3. 389–407. DOI:10.1146/annurev-linguistics-011516-033807
- Tinits, Peeter, Jonas Nölle & Stefan Hartmann. 2017. Usage context influences the evolution of overspecification in iterated learning. *Journal of Language Evolution* 2(2). 147–159. DOI:10.1093/jole/lzx011

- Verhoef, Tessa. 2012. The origins of duality of patterning in artificial whistled languages. *Language and Cognition* 4(4). 357–380. DOI:10.1515/langcog-2012-0019
- Wonnacott, Elizabeth & Elissa L. Newport. 2005. Novelty and regularization: The effect of novel instances on rule formation. In *BUCLD 29*, 663–673.

