Explanation in typology

Diachronic sources, functional motivations and the nature of the evidence

Edited by

Karsten Schmidtke-Bode

Natalia Levshina

Susanne Maria Michaelis

Ilja Seržant

Conceptual Foundations of Language Science

language science press

Conceptual Foundations of Language Science

No scientific work proceeds without conceptual foundations. In language science, our concepts about language determine our assumptions, direct our attention, and guide our hypotheses and our reasoning. Only with clarity about conceptual foundations can we pose coherent research questions, design critical experiments, and collect crucial data. This series publishes short and accessible books that explore well-defined topics in the conceptual foundations of language science. The series provides a venue for conceptual arguments and explorations that do not require the traditional book-length treatment, yet that demand more space than a typical journal article allows.

In this series:

1. N. J. Enfield. *Natural causes of language*.

ISSN: 2363-877X

Explanation in typology

Diachronic sources, functional motivations and the nature of the evidence

Edited by

Karsten Schmidtke-Bode

Natalia Levshina

Susanne Maria Michaelis

Ilja Seržant

language
science
press

Karsten Schmidtke-Bode , Natalia Levshina , Susanne Maria Michaelis & Ilja Seržant (eds.). 2018. *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence* (Conceptual Foundations of Language Science). Berlin: Language Science Press.

This title can be downloaded at:
http://langsci-press.org/catalog/book/000

ISBN: no digital ISBN
no print ISBNs!
ISSN: 2363-877X
no DOI
ID not assigned!

Cover and concept of design: Ulrike Harbort
Fonts: Linux Libertine, Libertinus Math, Arimo, DejaVu Sans Mono
Typesetting software: XeLaTeX

Language Science Press
Unter den Linden 6
10099 Berlin, Germany
langsci-press.org

Storage and cataloguing done by FU Berlin

no logo

Contents

Introduction

Karsten Schmidtke-Bode
Leipzig University

The present volume addresses a foundational issue in linguistic typology and language science more generally. It concerns the kinds of explanation that typologists provide for the cross-linguistic generalizations they uncover, i.e. for so-called universals of language. The universals at issue here are usually probabilistic statements about the distribution of specific structures, such as the classic Greenbergian generalizations about word order and morphological markedness patterns. Some examples are given in (1)–(4) below:

(1) With overwhelmingly greater than chance frequency, languages with normal SOV order are postpositional. (**Greenberg1963**)

(2) A language never has more gender categories in nonsingular numbers than in the singular. (**Greenberg1963**)

(3) If a language uses an overt inflection for the singular, then it also uses an overt inflection for the plural. (**Croft2003**: 89, based on **Greenberg1966**: 28)

(4) In their historical evolution, languages are more likely to maintain and develop non-ergative case-marking systems (treating S and A alike) than ergative case-marking systems (splitting S and A). (**BickelEtAl2015**: 5)

As can be seen from these examples, cross-linguistic generalizations of this kind may be formulated in terms of preferred types in synchronic samples or in terms of higher transition probabilities for these types in diachronic change (see also **Greenberg1978**; **Maslova2000**; **Cysouw2011**; **Bickel2013** for discussion of the latter approach). But this is, strictly speaking, independent of the question we are primarily concerned with here, namely how to best account for such generalizations once they have been established.

The most widespread typological approach to explanation is grounded in functional properties of the preferred structural types: For example, typical correlations in the ordering of different types of phrases (e.g. object–verb and NP–postposition) have been argued to allow efficient online processing (e.g. **Hawkins1994**; 2004). Markedness patterns in morphology (e.g. the distribution of zero expression in case, number or person systems) have been attributed to economy, i.e. the desire to leave the most frequent and hence most predictable constellations unexpressed, or rather to a competition between economy and the motivation to code all semantic distinctions explicitly (e.g. **Haiman1983**; **Comrie1989**; **Aissen2003**; **Croft2003**; **Haspelmath2008**; among many others). The general idea behind this approach is thus that speech communities around the world are subject to the same kinds of cognitive and communicative pressures, and that the languages they speak tend to develop structures that respond to these pressures accordingly, or, as **Bickel2014** puts it, "in such a way as to fit into the natural and social eco-system of speakers: that they are easy to process, that they map easily to patterns in nonlinguistic cognition, and that they match the social and communicative needs of speakers."

There is a clear parallel to evolutionary biology here, in that languages are said to *converge* on similar structural solutions under the same functional pressures, just like unrelated species tend to develop similar morphological shapes in order to be optimally adapted to the specific environment they co-inhabit (**Deacon1997**; **Caldwell2008**; **EvansLevinson2009**; **Givón2010**). When applied to language, this line of explanation at least implicitly invokes what is known as "attractor states", i.e. patterns of structural organization that languages are drawn into in their course of development.[1] For this reason, one could also speak of a **result-oriented** approach to explanation.

There is, however, another way of looking at the same patterns, one that redirects attention from the functional properties to the diachronic origins of the linguistic structures in question. On this view, many universal tendencies of order and coding are seen as by-products, as it were, of recurrent processes of morphosyntactic change, notably grammaticalization, but without being adaptive in the above sense: There is no principled convergence on similar structural traits because these traits might be beneficial from the perspective of processing, iconicity or economical communicative behaviour. Instead, the current

[1] The term attractor state (or basin of attraction) is adopted from the theory of complex dynamic systems (e.g. **Cooper1999**; **HoweLewis2005**; **Holland2006**), which has become increasingly popular as a way of viewing linguistic systems as well (see **BecknerEtAl2009** and **Port2009** for general overviews, and **Haig2018** or **Nichols2018** for very recent applications to typological data).

synchronic distributions are argued to be long-term reflections of individual diachronic trajectories, in particular the diachronic sources from which the structures in question originate. **Givón1984** and **Aristar1991**, for example, suggested that certain word-order correlations may simply be a consequence of a given ordering pair (e.g. Gen–N & Rel–N, or V–O & Aux–V) being directly related diachronically: Auxiliaries normally grammaticalize from main verbs that take other verbs as complements, and since these complements follow the verb in VO languages, they also follow the auxiliary in the resulting Aux–V construction; the mirror-image pattern holds for OV languages (see also **Lehmann1986**: 12–13). If this line of reasoning extends to most other word-order pairs, there is no need to motivate the synchronic correlations in functional-adaptive terms, e.g. by saying that the correlations arise *in order to* facilitate efficient sentence processing.

In the domain of morphology, **Garrett1990** argued that patterns in case marking, specifically of differential ergative marking, are exhaustively explained by the properties of the source of the ergative marker: When ergative case arises from the reanalysis of instrumental case, the original characteristics of the latter, such as a restriction to inanimate referents, are directly bequeathed to the former. The result is a pattern in which animate A-arguments are left unmarked, but since this is a direct "persistence effect" (**Hopper1991**) of the history of the ergative marker, there is again no need for an additional functional-adaptive explanation in terms of other principles, such as a drive for economical coding patterns. Rather than being result-oriented, then, this way of explaining universals can be characterized as **source-oriented**.

Such source-oriented explanations thus move away from attractor states of grammatical organization and often emphasize the importance of "attractor trajectories" instead (**BybeeBeckner2015**: 185): In some domains of grammar, the patterns of reanalysis and ensuing grammaticalization are so strikingly similar across the world's languages that it is not surprising that they yield similar outcomes, such as strong correlations between V–O & Aux–V or V–O & P–NP ordering. In other cases, it is argued that many individual, and partly very different, diachronies are capable of producing a uniform result, but without any consistent functional force driving these trajectories. **Cristofaro2017**, for instance, claims that this is the case for plural markers: An initial system without number marking can develop an overt plural morpheme from many different sources – usually by contextual reanalysis – and thus ultimately come to contrast a zero singular with an overt plural, but these developments are neither triggered nor further orchestrated by a need for economical coding: They do not happen to keep the (generally more frequent) singular unmarked and the (generally less frequent)

plural overtly signalled.

In other words, whether the individual diachronic trajectories are highly similar or rather diverse, the premise of the source-oriented approach is that they can scale up to produce a predominant structural pattern in synchronic samples. Hence they obviate the need for highly general functional principles tying these patterns together.

While the source-oriented approach was still a more marginal position in previous volumes on explaining language universals (e.g. **Hawkins1988a**; **Good2008**), it has gained considerable ground over the last decade, notably in a series of articles by Cristofaro (e.g. **Cristofaro2012**; 2014; 2017) but also in other publications (e.g. **Creissels2008**; **GildeaZúñiga2016**). Moreover, while the basic thrust of the two explanatory approaches is straightforward, clarification is needed on a number of – equally fundamental – details. After all, both approaches are functionalist in nature, as they rely on domain-general mechanisms (**Bybee2010**) to explain the emergence of language structure and linguistic universals; and in both approaches, these mechanisms constrain how languages "evolve into the variation states to which implicational and distributional universals refer" (**Hawkins1988b**). But as **Plank2007** notes, "what is supposed to be the essence and force of diachronic constraints would merit livelier discussion." It is the goal of the present book to offer precisely a discussion of this kind.

The volume begins with a programmatic paper by **Martin Haspelmath** on what it means to explain a universal in diachronic terms. He aims to clarify how diachrony is involved in result-oriented and source-oriented accounts, respectively, and thus lays out a general conceptual framework for the explanation of universals. At the same time, Haspelmath opens the floor for debating the strengths and weaknesses of the two explanatory accounts at issue here. His own position is that, in many cases, current source-oriented explanations are ill-equipped to truly explain the phenomena they intend to account for, and hence cannot replace result-oriented motivations. Haspelmath's arguments for this position, as well as his terminological proposals, provide a frame of reference to which all other contributions respond in one way or another.

The lead article is followed by two endorsements of source-oriented explanations, articulated by **Sonia Cristofaro** and **Jeremy Collins**, respectively. They both describe the approach in widely accessible terms, allowing also readers outside of linguistic typology to appreciate the general argument as well as the specific examples discussed. The phenomena themselves involve domains that are particularly well-known for being explained in functional-adaptive terms, namely differential argument marking, number marking and word-order corre-

lations, and these are all argued to be best captured by persistence effects from their respective diachronic origins.

We then proceed to papers that allow for progressively more room for functional-adaptive motivations and, importantly, for methodological discussions on how to obtain evidence for such pressures. Accordingly, all of these papers adduce novel empirical data and discuss them in light of the present debate.

Matthew Dryer's paper is an immediate follow-up on Collins' discussion of word-order correlations. On the one hand, Dryer argues that the various correlates of adposition–noun ordering (e.g. OV and NP–P, and Gen–N and NP–P) are, indeed, best accounted for in source-oriented terms. In particular, only this approach proves capable of explaining the occurrence (and the individual semantic types) of both prepositions and postpositions in SVO languages. On the other hand, however, Dryer contends that there are some significant correlations for which a source-based account either fails to offer an explanation or else makes the opposite prediction of the patterns we find synchronically. Dryer concludes, therefore, that neither a purely source-based nor a purely result-based explanation is sufficient to deal with word-order correlations.

In a similar fashion as Dryer's paper, **Holger Diessel**'s article demonstrates that different aspects of the same grammatical domain – in this case adverbial clause combinations – are amenable to different types of explanation. Diessel focuses specifically on the structure and development of preposed adverbial clauses and argues that some of their typological characteristics, notably the properties of their subordinating morphemes, receive a satisfactory explanation in terms of the respective source construction(s), thereby supplanting earlier processing-based explanations. On the other hand, he proposes that the position of adverbial constructions (in general) is clearly subject to a number of functional-adaptive pressures, and that these may already have affected the diachronic sources from which the current preposed adverbial clauses have grammaticalized.

Karsten Schmidtke-Bode offers a review of Hawkins' (2004, 2014) research programme of "processing typology", examining the plausibility of Hawkins' functional-adaptive ideas in diachronic perspective. On a theoretical level, it is argued that a predilection for efficient information processing is operative mostly at the diffusion stage of language change, regardless of the source from which the respective constructions originate. On a methodological level, the paper proposes that the cross-linguistic predictions of Hawkins' programme can be tested more rigorously than hitherto by combining static and dynamic statistical models of large typological data sets; this is demonstrated in a case study on the distribution of article morphemes in VO- and OV-languages, respectively.

An important methodological point is also made by **Ilja A. Seržant**, who claims that certain functional-adaptive pressures may not actually surface in standard typological analysis because they are weak forces, clearly at work but also easily overridden by other, language-specific factors. Because of their weak nature, they may not be directly visible any-more in a synchronic type, but they can be detected in qualitative data from transition phases. Based on diachronic data from Russian, Seržant shows how the development of differential object marking was crucially influenced by considerations of ambiguity avoidance (and hence a classic functional-adaptive motivation), over and above the constraints inherited from the source construction. In the absence of such longitudinal data, transition phases can be identified on the basis of syn-chronic variability, and Seržant shows that a wide variety of languages currently exhibit variation in differential object marking that mirrors the diachronic findings from Russian, and that is not predictable from the source meaning of the marker in question.

Susanne Maria Michaelis adds another source of data to the debate at hand. She argues that creole languages provide a unique window onto the relationship between synchronic grammatical patterns and their diachronic trajectories, as the latter are often relatively recent and also accelerated when compared to normal rates of grammatical change. The developments are, consequently, more directly accessible and less opaque than in many other cases. By inspecting creole data on possessive forms in attributive and referential function (e.g. *your* versus *yours*), Michaelis finds evidence for the development of the same kinds of coding asymmetries that this domain offers in non-contact languages around the world. She proposes that the data are indicative of result-oriented forces that drive diverse diachronic pathways towards the same synchronic outcome. This stance contrasts most explicitly with Cristofaro's, who interprets such situations in exactly the opposite way (i.e. as providing evidence *against* a unifying functional explanation).

Natalia Levshina, finally, adopts an entirely different methodological approach to illuminate the present discussion: In her paper, she showcases the paradigm of artificial language learning, which can be employed to inspect whether users of such newly acquired languages develop performance biases that are in keeping with hypothesized functional principles, such as an increasingly efficient distribution of morphological marking. Her case study clearly demonstrates such biases and discusses where they may ultimately come from, i.e. how they fit into the new conceptual framework of constraints offered by Haspelmath's position paper.

The volume is rounded off by a brief **epilogue** in which **Karsten Schmidtke-**

Bode and **Eitan Grossman** summarize and further contextualize the arguments put forward by the contributors.

Overall, the purpose of the present book is to provide a state-of-the-art overview of the general tension between source- and result-oriented explanations in linguistic typology, and specifically of the kinds of arguments and data sources that are (or can be) brought to bear on the issue. It should be made clear from the outset that the two types of explanation are framed as antagonistic here even though in most cases, an element of both will be needed in order to fully account for a given grammatical domain. As we emphasize in the epilogue, the diachronic source of a grammatical construction certainly constrains its further development, but the major issue at stake here is the extent to which result-oriented, functional-adaptive motivations enter these developments as well. By the end of the day, universals of language structure will thus differ in the *degree* to which they are shaped by such adaptive pressures.

Acknowledgements

The present volume originated in the context of the project *Form-frequency correspondences in grammar* at Leipzig University. The support of the European Research Council (ERC Advanced Grant 670985, Grammatical Universals) is gratefully acknowledged. An oral precursor to this volume was a workshop on the topic at the 49th Annual Conference of the Societas Linguistica Europaea in Naples in 2016, co-organized by the editors of this book. We would like to thank the participants and the audience of that workshop for insightful contributions and discussion. We would also like to thank Eitan Grossman and Mark Dingemanse for extensive feedback on all papers in the present volume. Finally, we are grateful to Sebastian Nordhoff and his team at Language Science Press as well as the participants of Language Science Press's community proofreading.

Chapter 1

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. **Haiman1983**; **Rohdenburg1996**; **Rohdenburg2003**; **Hawkins2004**; **Haspelmath2008**). 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 (2018 [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 **Prado2014**). 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 **SmithEtAl2017** 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 (**CulbertsonEtAl2012**), the suffixing preference (St. **ClaireEtAl2009**), definiteness hierarchy (**CulbertsonLegendre2011**) or the bias towards consistency in head-dependent order (**Christiansen2000**). 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 §**??** 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 §**??**). A brief summary and outlook are provided in §**??**

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 (**KirbyEtAl2014**). 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.

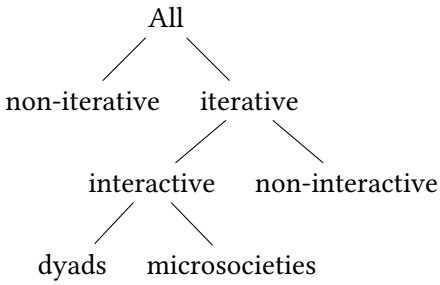

Figure 1: Main types of artificial language learning experiments.

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 **KamNewport2009**). Weaker biases may require several generations in order to manifest themselves (e.g. **RealiGriffiths2009**; **SmithWonnacott201**

Iterated learning can be further subdivided into interactive and non-interactive (cf. **Tamariz2016**). 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 (**Tamariz2016**). 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 (**CaldwellSmith2012**).

The artificial language learning paradigm is very flexible, allowing for investigation of diverse forms: non-existent words, whistles, graphical scribbles. A lan-

guage can also be fully artificial or semi-artificial. For instance, **SmithWonnacott2010** 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 **LittleEtAl2017**).

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 (**SmithEtAl2017**). 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. **CaldwellSmith2012**). 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 **FayEllison2013**. 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.

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 (**Verhoef2012**).

3. A bias towards **compositional structure** of syntax (**KirbyEtAl2008**). 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 (**Verhoef2012**).

5. A bias towards **regularity**. Languages exhibiting free variation become increasingly regular, revealing a strong bias towards regularity in adult learners (**SmithWonnacott2010**). 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 **KamNewport2009**; **WonnacottNewport2005**). 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 **KamNewport2009**);

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 (**FedzechkinaEtAl2012**). 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. **SilveyEtAl2015** 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, **TinitsEtAl2017** 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 (2018 [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

KirbyEtAl2008, **Verhoef2012** 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 **KamNewport2009**). This finding has been attributed to children having less cognitive resources than adults – in particular, memory limitations. However, **SmithEtAl2017** 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, **Perfors2016** 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 (**FehérEtAl2016**). Apparently, speakers believe that producing a more regular language will facilitate communication with their human partner. Similarly, **Little2011** 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 **FedzechkinaEtAl2012**, 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 $break_{\text{TR}}$, tend to denote more integrated causing and caused events than less integrated forms, such as *cause to die* or *make* $break_{\text{INTR}}$. As **Comrie1981** 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) 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 (**Levshina2017**).

Haspelmath2008 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 (**Levshina2016**), 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) Atruurian (artificial language;)
Grok babum.
cactus grow
"A grok (cactus) grows."

Atruurian

(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 (**Levshina2017**).

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 **KamNewport2009**). 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 videoclips. 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), 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,

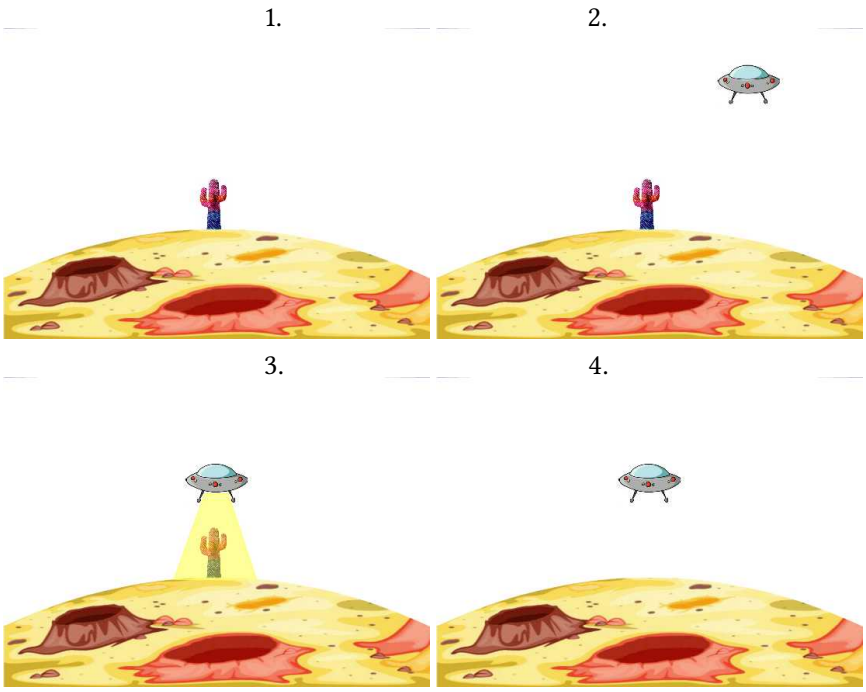

Figure 2: Fragments from one of the video clips.

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 ??. 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.

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.

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

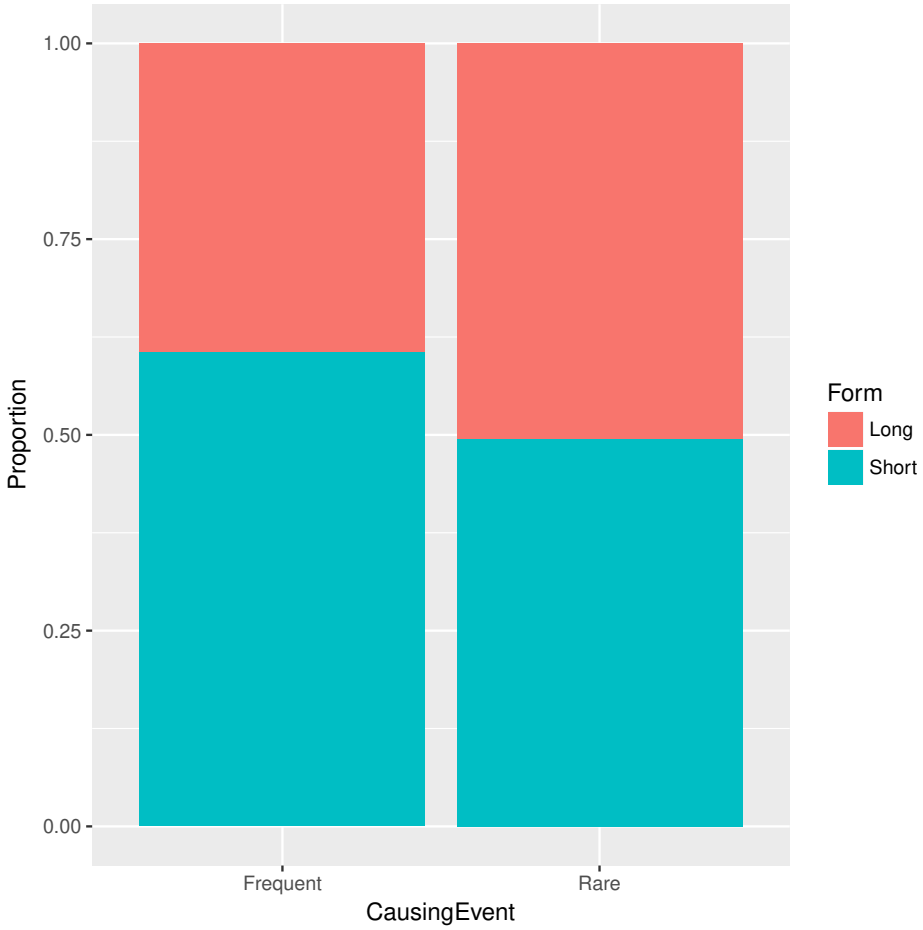

Figure 3: Proportions of short and long forms.

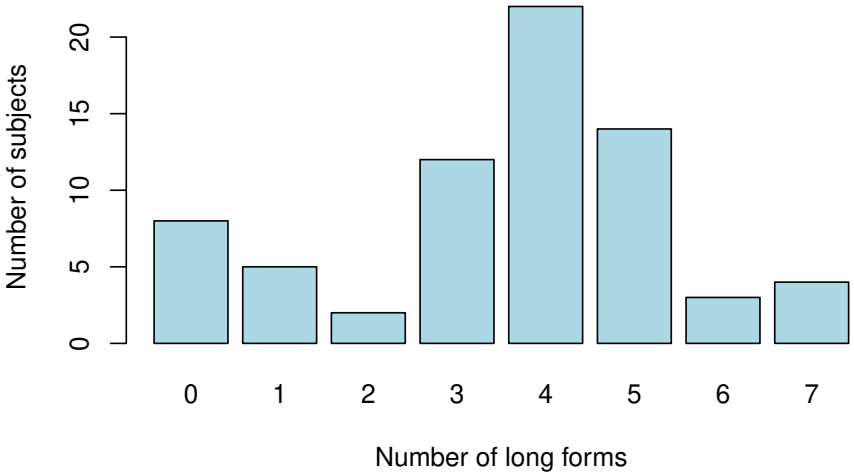

Figure 4: Individual preferences for the long and short forms

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*, **BatesEtAl2015**). 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. **SmithWonnacott2010**).

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 (2018 [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. **ParkerSmith1990**), 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 (**Stone2015**). 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. Although Haspelmath (2018 [this volume]) defines the latter as related to L1 by children only, the overview presented in §**??** demonstrates that learnability constraints can also be detected in artificial language learning by children and adults. §**??** also showed that a clear distinction between communicative efficiency and learnability is often problematic within the artificial language learning paradigm. Similar to **Slobin1996**'s (**Slobin1996**) 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, **Goldberg2013**, 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 **Horizon2020** 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.

Explanation in typology