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

Series editors

Mark Dingemanse, *Max Planck Institute for Psycholinguistics*

N. J. Enfield, *University of Sydney*

Editorial board

Balthasar Bickel, *University of Zürich*, Claire Bowern, *Yale University*, Elizabeth Couper-Kuhlen, *University of Helsinki*, William Croft, *University of New Mexico*, Rose-Marie Déchaine, *University of British Columbia*, William A. Foley, *University of Sydney* , William F. Hanks, *University of California at Berkeley*, Paul Kockelman, *Yale University*, Keren Rice, *University of Toronto*, Sharon Rose, *University of California at San Diego*, Frederick J. Newmeyer, *University of Washington*, Wendy Sandler, *University of Haifa*, Dan Sperber *Central European University*

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*.

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

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.

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

Chapter 1

MPI-SHH Jena & Leipzig University

Martin Haspelmath

This paper addresses a recent trend in the study of language variation and universals, namely to attribute cross-linguistic patterns to diachrony, rather than to other causal factors. This is an interesting suggestion, and I try to make the basic concepts clearer, by distinguishing clearly between language-particular regularities, universal tendencies, and mere recurrent patterns, as well as four kinds of causal factors (preferences, constraints, restrictions). I make four claims: (i) Explanations may involve diachrony in different ways; (ii) for causal explanations of universal tendencies, one needs to invoke mutational constraints (change constraints); (iii) in addition to mutational constraints, we need functional-adaptive constraints as well, as is clear from cases of multi-convergence; and (iv) successful functional-adaptive explanations do not depend on understanding the precise pathways of change.

1 Language universals: Constraints on cross-linguistic distribution as explananda

Since **Greenberg1963**, it has been widely recognized that comparison of languages with world-wide scope can give us not only taxonomies (as in earlier typology, e.g. **Schlegel1808**; **Schleicher1850**: 5–10; **Sapir1921**), but intriguing limits on cross-linguistic distributions: Especially when one looks at several parameters simultaneously, not all logically possible types are attested, or some types are far more common and others far less common than would be expected by chance. We would like to know why – or in other words, we are looking for causal explanations.

Since **Chomsky1981** at the latest, many generative grammarians have also been interested in cross-linguistic regularities, and have often interpreted them

as following from innate principles of universal grammar (UG) and their parametric variation. Others have tended to prefer functional explanations of universals (e.g. **Comrie1989**; **Stassen1985**; **Dixon1994**; **Dik1997**; **Hawkins2014**), but these authors have likewise appealed primarily to general principles of language and sometimes have even adopted the term "universal grammar" (**KeenanComrie1977**; Foley & Van **Valin1984**; **Stassen1985**).

In contrast to these two dominant approaches of the 1970s–1990s, there is an alternative view, according to which the explanation for universals of language structure comes from diachrony. The first well-known author in this tradition is **Greenberg1969**, who stated that "[s]ynchronic regularities are merely the consequence of [diachronic] forces" (1969: 186). A straightforward example of the explanatory role of diachrony is the generalization that languages with prepositions, the possessor generally follows the possessed noun, while in languages with postpositions, it generally precedes it (**Greenberg1963**'s (**Greenberg1963**) Universal 2; **Dryer1992**). This can be explained on the basis of the diachronic regularity that new adpositions generally arise from possessed nouns in processes of grammaticalization (**Lehmann1982**[2015]: §3.4.1; **Bybee1988**: 353–354; **Collins2018** [this volume]; **Dryer2018** [this volume]). For example, English *because (of)* comes from *by + cause (of)*. Since the order of the elements remains stable in grammaticalization, it is explained that the possessed noun and the adposition tend to occur in the same position in languages.

The view that the explanation of language universals (at least sometimes) comes from diachrony has apparently been gaining ground over the last decade and a half. The early papers by Greenberg (1969; 1978) and **Bybee1988** represented minority views (though **Givón1979** and **Lehmann1982** discussed diachronic change extensively and contributed to giving it a prominent place in functional-typological linguistics). Prominent papers in this vein in more recent years are **Aristar1991**, Anderson (2005; 2008; 2016), Cristofaro (2012; 2013; 2014), **Creissels2008**, **GildeaZúñiga2016**, and in phonology, **Blevins2004** is a book-length study that adopts a similar approach (see also **Blevins2006**). The following are a few key quotations from some of these papers (and from some others):

(1) a. "The question for typology is perhaps not what kinds of system are possible, but what kinds of change are possible." (**Timberlake2003**)

    b. "recurrent synchronic sound patterns are a direct reflection of their diachronic origins, and, more specifically, … regular phonetically based sound change is the common source of recurrent sound patterns" (**Blevins2006**: 119–120)

  c. "statistical universals are not really synchronic in nature, but are rather the result of underlying diachronic mechanisms that cause languages to change in preferred or 'natural' ways" (**BickelEtAl2015**: 29)

  d. "there are no (or at least very few) substantive universals of language, and the regularities arise from common paths of diachronic change having their basis in factors outside of the defining properties of the set of cognitively accessible grammars" (**Anderson2016**)

This paper has two major goals: First, I would like to contribute to conceptual clarification, sorting out what kinds of claims have been made and what terms have been used for which kinds of phenomena (§2).

Second, I argue that there are two ways in which diachrony and universals may interact: Some cross-linguistic generalizations are due to change constraints, as envisaged by the authors in (1), but others are due to functional-adaptive constraints. More specifically, I want to make four points:

- The notion of "diachronic explanation" is too vague, because explanations may involve diachrony in rather different ways (§3).

- Universal tendencies cannot be explained by common pathways of change, only by change constraints, or what I call mutational constraints (§4).

- Multi-convergence clearly shows that functional-adaptive constraints are needed in order to explain at least some cross-linguistic regularities (§5).

- Functional-adaptive explanations do not depend on understanding the pathways of change, though knowing about the pathways illuminates the explanations (§6).

Before arguing for these four points, I will discuss some technical terms in the next section, because there is often confusion between terms for language-particular regularities (§2.1), cross-linguistic regularities (§2.2), and causal factors (§2.3).

2 Regularities and causal factors: Concepts and technical terms

General terms such as *restriction, constraint, preference, tendency, bias,* and *motivation* have been used in diverse and sometimes confusing ways by linguists. This

section clarifies how these terms are used in the present paper, noting along the way what other meanings some of them have been given and what other terms have been used for (roughly) the same concepts. I distinguish between terms for regularities and terms for causal factors, and within the terms for regularities, I distinguish between language-particular and cross-linguistic regularities.

2.1 Language-particular regularities

Regularities within a particular language can concern language use or the conventional language system. Regularities of language use are increasingly studied by corpus linguistics, and they are often thought to be at the root of system regularities, especially in what is often called a "usage-based" view (**Bybee2010**). However, regularities of use and system regularities are conceptually different, and linguists normally distinguish clearly between *parole* (language use) and *langue* (language system). In what follows, I focus on the systems of linguistic conventions.

For regularities within language systems, linguists normally use the general terms *rule* and *construction* (or *schema*). In addition, descriptive linguists use many other well-established class (or category) terms like *clause*, *noun phrase*, *suffix*, *dative case*, or terms for relations between constructions such as *alternation* or *derivation*. All of these relate to systems of particular languages.

The term *constraint* is sometimes applied to language-particular regularities, e.g. in constraint-based formalisms such as HPSG, and optimality theory of course also uses constraints for language-particular regularities. However, I will use this term exclusively for causal factors, as explained in §2.3 below.

Language-particular regularities can also be seen as "explanations", at least in the weak sense that they answer why-questions about lower-level regularities ("Why is there a Dative case on the object of this sentence? Because the verb's valency requires a Dative."). Statements of rules or constructions may thus be called "descriptive explanations" if one wishes. In this paper, however, I focus on causal explanations that help us explain the conventional systems of languages themselves.

2.2 Cross-linguistic regularities: Recurrent patters and universal tendencies

Cross-linguistic regularities are typically generalizations over language-particular regularities,[1] and I will distinguish two kinds of regularities here. On the one hand, similar phenomena may be found in different parts of the world, e.g. ejective consonants, or high vowel epenthesis, or optative mood forms, or functive markers (**Creissels2014**). These are called RECURRENT PATTERNS. On the other hand, some regularities are so strong that we call them UNIVERSALS, because they occur with much greater than chance frequency. I also often use the term UNIVERSAL TENDENCIES, because there is no claim that there are no exceptions.[2]

Recurrent patterns are not accidental similarities, in the sense that there must be something in the human condition that makes it possible for very similar linguistic categories to appear independently in languages that have no historical connection. However, the discovery of a recurrent pattern does not imply a claim about further languages.

By contrast, the discovery of a universal implies a claim about all other languages: If a universal holds (i.e. is found with much greater than chance frequency in a reasonably representative sample), it is claimed that it also holds in any other representative sample. Thus, universal tendencies are claims that can be tested by examining data from the world's languages, while recurrent patterns can only be strengthened by additional further observations, but neither confirmed nor disproven by additional data.

Universal tendencies need to be distinguished, in particular, from family-specific or region-specific trends, so they need to be based on a world-wide sample. A well-known example is the finding that in all major world regions, languages with OV order tend to have postpositions, and languages with VO order tend to have prepositions (**Greenberg1963**: Universal 2; **Dryer1992**: 83), even though many languages are exceptions. Another universal tendency is the limitation of nominal suppletion to the most frequent nouns (**Vafaeian2013**), even though many languages do not have nominal suppletion at all. We may even identify universal tendencies within patterns that are quite rare, e.g. universals of infixa-

---

[1] However, comparative corpus linguistics studies comparable corpora of language use, so there is no necessary connection between cross-linguistic comparison and the study of systems (as opposed to use).

[2] Another term for a cross-linguistic distribution is **Bickel2013**'s (**Bickel2013**) family bias, which means 'preponderance within a family'. Note that this use of *bias* is quite different from the more common use as in *cognitive bias* (e.g. **TverskyKahneman1974**); a term like *family tendency* would probably be more transparent.

tion (**Yu2007**), because universal tendencies can be implicational ("If a language has infixation, then…").

Recurrent patterns, by contrast, are not associated with any kind of quantitative claim, so they could be called *frequent patterns*, or *sporadic patterns*, depending on one's subjective assessment of their frequency. They are no doubt important for a complete account of human language, but they will be left aside in what follows, as it is not clear what causal factors might illuminate them.

2.3 Causal factors: Preferences, constraints, restrictions

In addition to documenting language-particular systems and cross-linguistic distributions, we also want to know what might explain the distributions in causal terms. The explanatory devices are called *causal factors,* or *(system-external) motivations*, or *constraints*. Especially the latter term is short and relatively clear, so I will use it as the default term for a causal factor. (Two other terms that are used commonly as well, especially outside core linguistics, are *force* and *pressure*. It seems that all these terms are basically synonymous.)

If a constraint is very strong, it can also be called *restriction*, and if it is weaker, it can be called *preference*.[3] This seems to be in line with much current usage in linguistics. There is thus no objective difference between restrictions, constraints and preferences, and we could use one of the three terms for all types of constraints. (This situation is similar to the cases of sporadic and frequent patterns, which are subjective sub-cases of recurrent patterns.)

Depending on the way in which they affect cross-linguistic distributions, here I distinguish four types of constraints (or restrictions, or preferences), which can be briefly characterized as in (**??**).

(2) a. *functional-adaptive constraints*: what facilitates communication (including processing) for speakers and hearers

b. *representational constraints*: what is innately preferred or necessary in the cognitive representation of language

c. *mutational constraints*: what is preferred or necessary in language change (= change constraints)

---

[3]Another term for system-external causal factors is *bias*, which is used in particular by psychologists for cognitive preferences. Typical biases seem to be quite weak, so that even detecting them is an important part of research. By contrast, linguists' constraints are often very strong, and controversy concerns primarily the nature (functional-adaptive, representational, mutational) and the interaction of the constraints.

d. *acquisitional constraints*: what is preferred or necessary in acquisition by children

FUNCTIONAL-ADAPTIVE CONSTRAINTS are the kinds of factors that have been invoked by functionalists to explain cross-linguistic distributions (e.g. **Tomlin1986**; **Malchukov2008**; **Hawkins2014**; among many others). For example, phonological inventories favour five-vowel systems because these make the best use of the acoustic space (De **Boer2001**), and case systems favour overt ergatives for low-prominence nominals and overt accusatives for high-prominence nominals because of the association between roles and prominence status (**Dixon1994**). These constraints are called *functional-adaptive* rather than merely *functional* to emphasize their role in explaining systems, not usage (the functioning of language). Functional linguists often focus on understanding the functioning of language in usage, but here my interest is in explaining how systems come to have properties that facilitate communication.[4] **Good2008** uses the term "external explanation" in roughly this sense (cf. also **Newmeyer1998**: §3.4), but all four types of constraints are external in that they are not part of the system. ("System-internal explanation" is just another word for general regularities of language-particular systems, cf. §** above; I do not think that the notion of causality is relevant for such statements, so all causal explanatory factors are external.)

REPRESENTATIONAL CONSTRAINTS are the kinds of factors that have been invoked by generativists to explain grammatical universals, as noted in §1. In the principles and parameters framework (**Chomsky1981**), they were called the principles of universal grammar. For example, the principles of X-bar theory or binding theory have been regarded as representational constraints, as well as universal features and hierarchies of functional categories (e.g. **Cinque1999**). The general idea is that "the unattested patterns do not arise as they cannot be generated in a manner consistent with Universal Grammar" (**SmithEtAl2018**). Representational constraints are usually regarded as very strong, i.e. as restrictions (and thus universal grammar is said to be *restrictive*; cf. also **Haspelmath2014**).[5] However, there is no intrinsic reason why representational constraints could not be weaker preferences, e.g. why there could not be a weak innate preference to put elements

---

[4]Another term for functional-adaptive constraint is "naturalness parameter" (**DresslerEtAl1987**), and functional-adaptive changes have been called "natural changes".

[5]Cognitive linguists have also sometimes invoked representational constraints to explain universals, though these are not referred to as UG. An example might be the idea in **Croft1991** that all event types are modeled on the basic force-dynamic agent-patient event type. This is not very strong, i.e. it is a preference, but apparently a preference having to do with cognitive representations, not with communicative or processing preferences.

into a "Determiner" category (though this possibility is almost never considered by linguists). In **Good2008**'s (**Good2008**) survey, representational constraints are treated under the label of "structural explanations", but this term (like "system-internal explanations") is better reserved for general statements of regularities of language-particular systems.

MUTATIONAL CONSTRAINTS (or change constraints) are constraints on possible diachronic transitions or possible diachronic sources, which can have an effect on synchronic distributions. For example, if nasal vowels only ever arise from VN sequences, this explains that all languages with nasal vowels also have nasal consonants, and that nasal vowels are rarer than oral vowels in the lexicon (**Greenberg1978**). Likewise, if infixes only ever arise by metathesis from adfixes (= prefixes or suffixes), this explains that they only occur in peripheral position (**Plank2007**). And if adpositions only arise from nouns in possessor–noun constructions, this explains that their position correlates with the position of possessed nouns, as noted in §1. The notion of mutational constraints is not new (**Plank2007**: §2 calls them "diachronic laws"), but I introduce a new term here in order to make clear that the causal factor is located within the process of change, rather than diachronic change merely realizing a pattern that is driven by functional-adaptive constraints (see §3 below). One could also frame the contrast between mutational constraints and functional-adaptive constraints in terms of *source-oriented* vs. *result-oriented* factors (**Cristofaro2017**),[6] or one could say that mutational constraints locate the causal factors within the *mechanisms* of change (**Bybee2006**). These are just alternative ways of saying that cross-linguistic distributions are due to mutational constraints.

Finally, ACQUISITIONAL CONSTRAINTS are factors that impact the acquisition of language and that have an effect on cross-linguistic distributions. Such constraints are briefly discussed by **Anderson2016**, but they do not seem to play a big role in linguistics (but cf. **Levshina2018** [this volume] for discussion). Generative linguists who are concerned with learnability issues generally assume that what can be represented can also be learned, so that there is no distinction between representational constraints and what can be learned. This type of constraint is mentioned here only in passing, for the sake of completeness. It will play no role in what follows.

---

[6]Informally, instead of talking about "result-oriented factors", one could also say that functional-adaptive constraints are "pull forces" that attract the variable development into a certain preferred state.

3 Two ways in which causal explanations involve diachrony

The peculiar term *mutational constraint* that I adopt here may raise questions: Is it necessary to use a new term for something that is very straightforward?

The reason I am using this term here is that the possible alternatives "diachronic constraint" or "diachronic explanation" are not fully clear. First of all, diachronic explanations may simply be explanations of diachronic changes, but here we are concerned with causal factors leading to universals. Second, "diachronic" and "historical" are often used interchangeably (cf. **Good2008**'s (**Good2008**) term "historical explanation" for what I call mutational explanations), and when we speak about historical explanations, we often mean contemporary idiosyncrasies that are better understood if one knows their origins (e.g. the vowel alternation in *foot/feet* finds a historical explanation in the earlier productive pattern of vowel fronting conditioned by a high vowel in the following syllable). But all of this is irrelevant in the present context, where we are concerned with possible and impossible pathways (and sources) of change.

Most importantly, the term *mutational constraint* is necessary because there are two ways in which causal explanations involve diachrony: synchronic distributions may be diachronically DETERMINED, or they may come about by the diachronic REALIZATION of preferred outcomes. The term *mutational constraint* highlights the fact that change is seen as a causal factor here, not merely the way in which the cross-linguistic distributions arise. By contrast, when universal tendencies are explained by functional-adaptive constraints, diachronic change merely serves to realize the adaptation. It plays an important role, indeed a crucial role, because functional adaptation is impossible without change. In this sense, functional-adaptive explanations are also diachronic (cf. **Haspelmath1999a**). But functional-adaptive change is not the cause of the adaptation – the cause is the facilitation of communication for speaker and hearer. Mutational constraints are situations where the causal factor resides in the change itself.

Two types of mutational constraints may be distinguished: Source constraints and directionality constraints. Most of the diachronic regularities discussed by **Cristofaro2017** concern constraints on possible sources. The best-known directionality constraint is the irreversibility of grammaticalization (**Haspelmath1999b**; 2004).[7]

---

[7]Mutational constraints are themselves in need of explanation, of course. I say nothing about this in the current paper, because it is already long and complicated enough. Their explanation could itself be "functional" in some sense (to be made more precise), but it cannot be functional-

Another reason for avoiding the terms "diachronic constraint" or "diachronic explanation" is that they invite a contrast with "synchronic constraint" and "synchronic explanation". But these terms are themselves very problematic, because they seem to conceive of explanation in noncausal terms. The term "synchrony" has a clear application with reference to an abstract, idealized language system (de Saussure's *langue*), but in §2.1 I noted that language-particular system regularities should be described in terms of constructions or rules, and that causal constraints cannot play any role in them.[8]

Instead of "mutational constraint", one could use the term "constraint on change" (as in the title of this paper), but the new term "mutational" is more salient (it can be found more easily in automatic text searches), and since it is more specific, it can be used in new combinations like "mutational explanation" (an explanation in terms of a mutational constraint) or "mutational approach".

4 Universals are not explained by recurrent pathways of change, only by constraints on change

It has long been known that there are recurrent kinds of changes in phonology (lenition of consonants between vowels, diphthongization of long vowels, assimilation, etc.), and over the last few decades, recurrent changes in morphosyntax have become prominent as well, especially changes falling under the broad category of grammaticalization (**Lehmann1982**; **HeineEtAl1991**; **BybeeEtAl1994**; and much related work).

Bybee2006 highlights recurrent or common pathways of change in the tense-aspect domain (perfectives coming from anteriors and ultimately from completive, resultative or movement constructions; imperfectives coming from progressives and ultimately from locational or reduplicative constructions; and futures coming from volitional or movement constructions), and makes the claim that "the **true universals** of language are the **mechanisms of change** that propel the constant creation and recreation of grammar" (**Bybee2006**: 179–180).

But she does not distinguish clearly between recurrent pathways of change and constraints on possible changes. There is no doubt that the tense-aspect

---

adaptive, because the latter type of explanation (as I understand it here) by definition applies only to language systems, not to changes.

[8]Of course, in practice linguists often use the terms "synchronic explanation" and "synchronic constraint", but what they mean is either (i) very general language-particular statements ("descriptive explanations", §2.1), or (ii) representational constraints. The latter are biological limitations, which can hardly be labeled felicitously with the Saussurean term *synchronic*.

changes that she discusses are widespread and significant developments, but nobody knows how widespread they are, compared to other possible changes. There are many perfective, imperfective and future markers about whose sources we know little, or markers whose sources do not fit into any of Bybee's categories. It is true that the recurrence of the changes makes it virtually certain that the similarities are not accidental, but we do not know enough about tense-aspect developments to assert with confidence that no other sources are possible or likely, nor even that these sources are clearly predominant over other possibilities.

In one passage Bybee asserts that "the diachronic paths present much stronger cross-linguistic patterns than any comparison based solely on synchronic grammars" (2006: 180; see also **Bybee2008**: 169). But her evidence is not sufficient to show this, at least for tense and aspect, where the pathways of change are highly diverse, and few people would venture a claim that certain kinds of change are impossible or highly unlikely.

In order to explain universal tendencies, one needs to appeal to something that is stronger than "recurrent (or common) pathways of change", namely **mutational constraints, of the type mentioned earlier. Such constraints allow causal explanations of synchronic cross-linguistic distributions, just like functional-adaptive constraints.** In phonological change, also discussed by Bybee, some common pathways may indeed qualify as mutational constraints: It could well be that changes involving [h] are highly uniform (especially [s]/[x] > [h] > Ø), so that we are dealing with a mutational constraint, not just a recurrent pathway.[9] **Since such mutational constraints entail certain synchronic distributions, they qualify as true explanations, and if a synchronic distribution can be explained by a change constraint, it is not "accidental" (as Collins2018 [this volume] calls the universal that adposition order correlates with verb–object order).**[10]

At this point it is reasonable to ask how one can distinguish in practice between recurrent pathways and mutational constraints. The way to distinguish between synchronic cross-linguistic regularities and recurrent patterns is by gathering representative world-wide data samples (§??), and in principle,

---

[9]It is true, of course, that there are some really interesting constraints on morphosyntactic change, notably the constraint that grammaticalization cannot be reversed (**Haspelmath1999b**). However, such mutational constraints need not give rise to synchronic universal tendencies. Grammaticalization as such does not result in any universal tendencies, and Bybee (2006: §8) is apparently right that the lenition of [s] or [x] via [h] to Ø does not give rise to any synchronic universals either.

[10]It could be that Collins thinks that only representational or functional-adaptive constraints can explain synchronic universals, or it could be that he does not think that the sources of adpositions are sufficiently constrained. See the next paragraph for more on that possibility.

one would have to do the diachronic counterpart in order to establish a mutational constraint. As Collins (2018: # [this volume]) puts it, "we need large databases of **attested grammaticalisation pathways**". This is not very practical, however, as there are few solidly attested cases of grammaticalization, mostly from European (and a few Asian) languages, and most of what we think we know about general change patterns is based on indirect inferences and cannot be subjected to statistical testing the way this is possible with synchronic patterns. Thus, in practice, linguists rely on their general experience when making judgments, or they cite a range of examples to persuade their colleagues. This method is much less rigorous than the study of synchronic regularities, but it seems to be uncontroversial to assert that in general, both types of diachronic regularities exist: mutational constraints (where a particular outcome has no other possible source), and recurrent changes. This is all I want to argue for in this paper, and I make no strong claims about particular instances (e.g. whether adpositions are constrained to arise only from possessed nouns and transitive verbs, or whether these are merely recurrent sources).

5 Multi-convergence can only be explained by functional-adaptive constraints

Since mutational constraints are one possible source of synchronic universals, it could be that in fact all synchronic universals are due to mutational constraints of one kind or another, and that functional-adaptive and representational constraints are not needed. This is a fairly radical position, but **Cristofaro2017** comes close to adopting it.

Perhaps the strongest reason to believe that we also need functional-adaptive explanation is that there are many cases of multi-convergence, i.e. situations in which a uniform result comes about through diverse pathways of change that yield a very similar result. For example, I note in **Haspelmath2017** that inalienable adpossessive constructions tend to have shorter coding or zero, whereas alienable adpossessive constructions have overt or longer coding, and I also observe that these patterns can come about in different ways. The inalienable pattern may be shorter because of special shortening, or it may be shorter because only the alienable pattern got a special new marker. **Kiparsky2008** makes a very similar argument against **Garrett1990**'s (**Garrett1990**) explanation of split ergativity in mutational terms, noting that "[Garrett's] historical account is insufficiently general […] because the phenomenon to be explained has several historical sources".

Interestingly, two of the advocates of mutational explanations of universal tendencies observe the heterogeneity of the pathways themselves. **Anderson2016** is concerned with case-marking patterns in perfective and imperfective aspects across languages, and **Cristofaro2017** is concerned with the coding asymmetry of zero singulars and overt plurals:

> As it happens, common sources for a new perfective, on the one hand, and for a new imperfective, on the other, converge on similar patterns of split ergativity, although they are quite unrelated to each other. (**Anderson2016**: 23; cf. also **Anderson1977**)

> Different instances of the same configuration can also be a result of very different processes. For example, phonological erosion, meaning transfer from a quantifier to an accompanying element, and the grammaticalization of distributives into plural markers can all give rise to a configuration with zero marking for singular and overt marking for plural, yet they do not obviously have anything in common. (**Cristofaro2017**:18–19)

Anderson and Cristofaro are thus aware of the multi-convergence patterns, but for some reason they do not draw the conclusion that we need an additional causal factor to explain the convergence – and as far as I can see, this factor can only be a functional-adaptive constraint.[11]

The convergence of diverse processes on a uniform result could conceivably be accidental, but in this case it could not explain a universal tendency, because a universal tendency is by definition non-accidental. A universal tendency still holds if more and more languages are looked at, whereas accidental similarities of the results of diverse processes would not be repeated if more phenomena were considered. On the analogy of biological usage, where "convergent evolution" refers to the independent development of similar traits for adaptive reasons, one should probably avoid the term "convergence" if one thinks that the similarities are accidental and will not be confirmed by a larger sample. Thus, Anderson and Cristofaro should think of their observations in terms of coincidental similarity rather than convergence.

---

[11]In principle, it could also be a representational constraint (i.e. universal grammar), but since the patterns involve implicational universals, this would be difficult to argue for. In general, implicational universals cannot be easily explained by representational constraints.

6 Functional-adaptive explanations need not specify pathways of change

One point of criticism of functional-adaptive explanations is that they do not say how the change comes about. Especially Bybee and Cristofaro have argued that for a functional explanation of cross-linguistic regularities to be accepted, it must be shown how the functional motivation plays a role in the way in which the resulting patterns comes about.

I agree that the functional motivation must play a role in the way in which the pattern comes about, but I do not agree that the manner in which it influences the change must be identified for a successful explanation. Below are two relevant quotations.

> [I]n language universals, causal factors are linguistic changes that create particular synchronic states, and the existence of massive cross-language similarity in synchronic states implies powerful parallels in linguistic change. … the validity of a principle as explanatory can only be maintained if it can be shown that the same principle that generalizes over the data also plays a role in the establishment of the conventions described by the generalization. (**Bybee1988**)

> These [functional] explanations … have mainly been proposed based on the synchronic distribution of the relevant grammatical phenomena, not the actual diachronic processes that give rise to this distribution in individual languages. In what follows, it will be argued that many such processes do not provide evidence for the postulated dependencies between grammatical phenomena, and suggest alternative ways to look at implicational universals in general. (**Cristofaro2017**)

The problem with Bybee's claim is that the changes are seen as causal factors themselves: Bybee does not seem to envisage the possibility of a "pull force" that increases the probability of change toward a particular kind of outcome, without determining the way in which the change comes about. Moreover, she formulates the requirement that one should be able to DEMONSTRATE that the functional-adaptive principle plays a role in the change, but this requirement is too strong. In general, we do not know much about language change and how and why it happens. The primary evidence for functional-adaptive explanations is the fit between the causal factor and the observed outcome. If there is a good fit, e.g. if languages overwhelmingly prefer the kinds of word orders that allow easy

parsing (**Hawkins2014**), or if they tend to show economical coding of grammatical categories (**Haspelmath2008**), the best explanation is in functional-adaptive terms, as long as there is a way for languages to acquire these properties. The latter requirement is always met, as there are no synchronic states which cannot have arisen from other states. Thus, we may not know how exactly the zero singulars and overt plurals in Hebrew (e.g. *sus* 'horse', *sus-im* 'horses') may have come about, as they are found in much the same way in Proto-Semitic, but we know various ways in which plurals can arise (**Cristofaro2013**: §4), so there is no problem in assuming that the functional motivation of economical coding of the singular played a role in the development of the contrast.

Cristofaro is right that when we look at the changes that give rise to apparently functionally motivated distributions, we do not (necessarily) find evidence that the changes were driven by the need to obey the functional constraints, but finding such evidence is not necessary for a successful explanation. The evidence for the functional motivation does not come from the manner in which the change happened, but from the fit between the motivation and the observed outcomes. If there is a universal tendency, and it can be explained by a universal motivating factor, then that explanation should be accepted unless a better explanation becomes available.

Explanations of regularities in the world-wide distribution of cultural traits often appeal to functional-adaptive factors in adjacent fields as well. For example, anthropologists sometimes explain religion by prosociality, or monogamy by group-beneficial effects (e.g. **PaciottiEtAl2012**; **HenrichEtAl2012**). The issue here is whether better explanations are available, not whether there is a way for religion or marriage to develop. We know little about how religion and marriage first arose or generally arise in societies, and it is very difficult to study the diachronic developments. But we can try to correlate structural traits of human societies with other traits and draw inferences about possible causal factors. There is no perceived need in this literature to show that the mechanisms by which religion or monogamy arise must be of a particular type.[12] Basically, when the result is preferred, any kind of change can give rise to the result, and we do not need to understand the nature of the change, let alone show that the change was motivated by the result.

Another striking example from linguistics is the shortness of frequent words,

---

[12]The same is true for adaptive explanations in evolutionary biology: The fact that wings are adaptive can be inferred from the way wings are used by animals, and we do not expect that wings arise in uniform ways (wings of birds, bats and insects have diverse origins and arose by diverse paths of change, whose nature is not relevant to the adaptive explanation).

which is surely adaptive. But there are quite diverse paths to shortness. According to **Zipf1935**, shorter words are shorter because they underwent clipping processes (e.g. *laboratory > lab*), and according to **Bybee2007**, short words are short because "high-frequency words undergo reductive changes at a faster rate than low-frequency words […] the major mechanism is gradual phonetic reduction". But actually in most cases, rarer words are longer because they are (originally) complex elements, consisting of multiple morphs, e.g. *horse vs. hippopotamus, car vs. cabriolet, church vs. cathedral.* Drastic shortening of longer words seems to occur primarily in the modern age with its large number of technical and bureaucratic innovations, but even here, clipping is only one of many possibilities; for example, Ronneberger-**Sibold2014** discusses a number of fairly diverse of "shortening techniques" in German. What unites all of these processes is only one feature: the outcomes of the changes, which are functionally adapted.

When **Cristofaro2014** writes that "any model of the principles that lead to the use of particular constructions […] should take into account the diachronic development of these constructions, rather than just their synchronic distribution", I certainly agree, because I think that the diachronic developments can illuminate the functional adaptation, and a close study of whatever we can learn about diachrony can tell us whether any mutational constraints might play a role. But when there is strong evidence for a universal tendency and there is a good functional-adaptive explanation available, the diachronic evidence is not strictly speaking necessary.

7 A cost scale of constraints

What are we to do when there are several possible explanations, using different kinds of causal factors? For example, what do we do when word-order correlations can be explained either by functional adaptation (processing efficiency, **Hawkins2014**) or by mutational constraints? Or when case-marking splits can be explained either by universal grammar (**Kiparsky2008**: §??) or by efficiency of coding?

The answer is that there is a COST SCALE of constraints:

(3)   less costly < – – – – – – – – – – – – – > more costly
       mutational constraints > functional-adaptive constraints >
       representational constraints

The "cheapest" type of explanation is the mutational mode, because language change can be observed, and if we find that certain changes simply do not occur

(for whatever reason), we do not need to make more far-reaching claims. Thus, **Bybee2010** discusses the Greenbergian word order correlations and notes that "grammaticalization gives us the correct orders for free" – a formulation that reflects the assessment that mutational constraints do not involve any additional "cost".[13]

The next type of explanation on the scale appeals to functional-adaptive constraints. These are more costly because we cannot observe their effects directly and have to rely heavily on inference. But they are less costly than representational constraints, because they are far more general, applying also in other domains of cognitive processing and communication, often also in nonhuman animals. Again, this is not really controversial: In his chapter on universal grammar, **Jackendoff2002** says that "we should be conservative in how much linguistic structure we ascribe to an innate UG. We should welcome explanations of linguistic universals on more general cognitive grounds."

It is only when we observe a cross-linguistic regularity that cannot be explained either by mutational constraints or by functional-adaptive constraints that we need to appeal to representational constraints. These involve the most specific (and thus most costly) mechanism, which should only be invoked as a last resort.

8 Conclusion

In this paper I have argued that cross-linguistic regularities may be explained either by mutational constraints or by functional-adaptive constraints (or perhaps by representational constraints, as in generative grammar) (§??). Both kinds of explanations involve diachrony, but in different ways: Mutational constraints are constraints on possible sources or pathways of change, while functional-adaptive constraints influence the results of changes (§??). In order to explain a universal tendency, we need to appeal to mutational constraints; merely noting a frequent pathway of change is not enough (§??). We can be sure that a cross-linguistic regularity is due to a functional-adaptive rather than a mutational constraint if there are diverse pathways of change which converge on a single result (§??). The functional-adaptive constraint must influence language change in such a way that change in a particular direction becomes more likely, but this need not

---

[13]Cf. also the similar argumentation in **Kiparsky2008**, in connection with a different phenomenon (involving reflexives): "That is really all that needs to be said [...] The historical explanation covers the data perfectly." I completely agree with Bybee and Kiparsky in this respect.

be visible in the change itself (§**??**). But when we have good reasons to think that there is a mutational constraint, it takes precedence over functional-adaptive and representational explanations (§**??**).

Thus, the answer to the question in the title of this paper ("Can cross-linguistic regularities be explained by constraints on change?") is: Yes, some regularities can apparently be explained in this way, but clearly not all of them. There remains an important role for functional-adaptive constraints in explaining language universals.

Acknowledgement

Chapter 2

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 Linguistic Frankenstein

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.
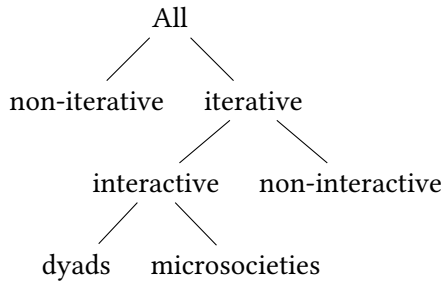

Figure 1: Main types of artificial language learning experiments.

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 gener-
ally 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 seman-
tic categories. In my own study, I want to focus on predictability based on fre-
quency 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 **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 cau-
sation without spatiotemporal integration of the events and without direct im-
pact 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;)
Atruurian **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 (**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,
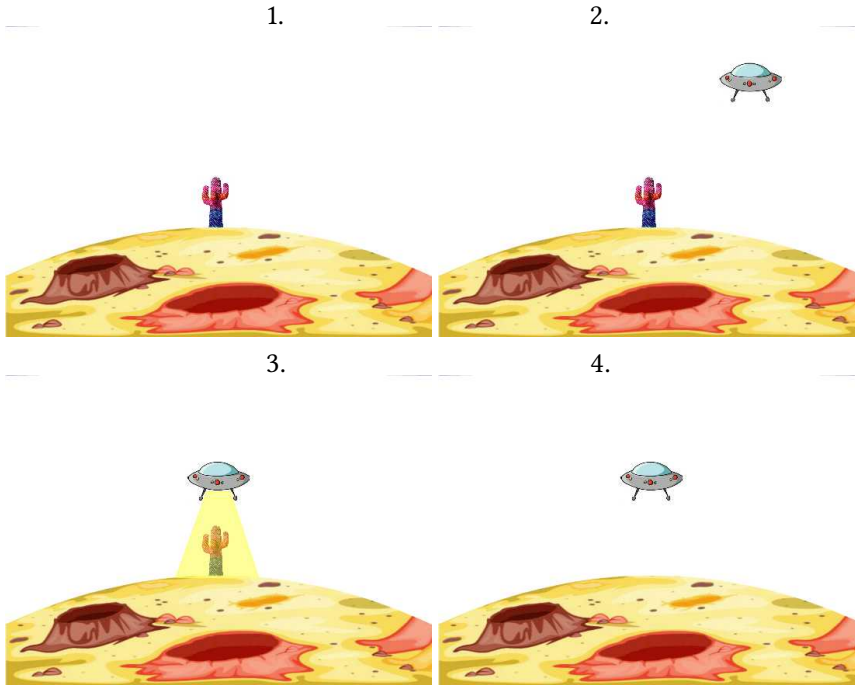
1.   2.

3.   4.

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.

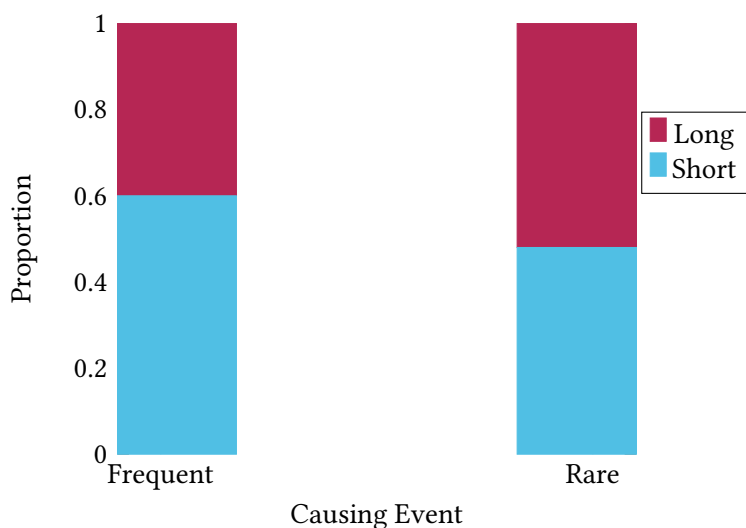|	Frequent	Rare	Total
Short	168	137	305
Long	109	140	249
Total	277	277	554

Figure 3: Proportions of short and long forms.

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
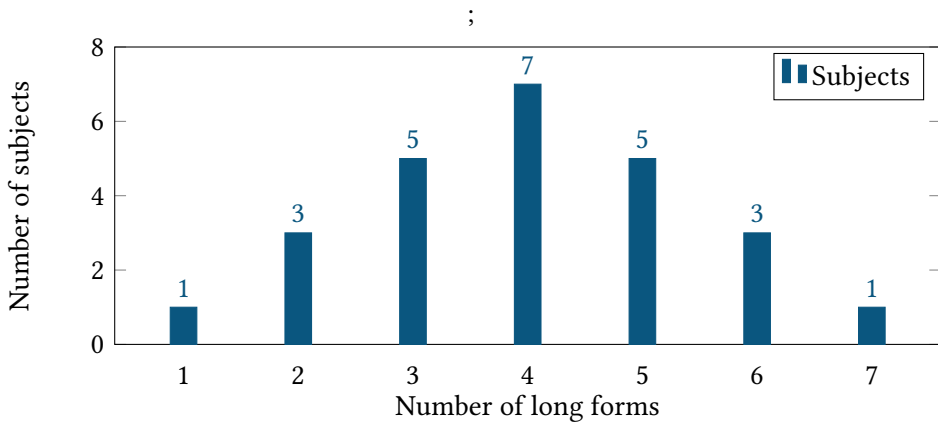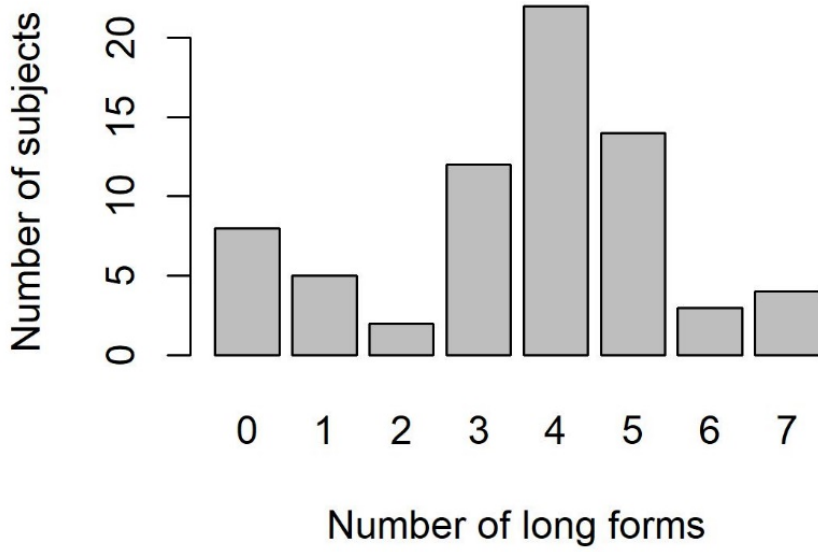
Figure 4: Individual preferences for the long and short forms

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 learnabil-

ity 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

Natalia Levshina

Michaela Martinkova and Marketa Janebova. All usual disclaimers apply.

Did you like this book?

This book was brought to you for free

Explanation in typology