

# 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



# 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 Bower, *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.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/000>

© 2018, the authors

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/>

ISBN: no digital ISBN

no print ISBNs!

ISSN: 2363-877X

no DOI

ID not assigned!

Cover and concept of design: Ulrike Harbort

Fonts: Linux Libertine, Libertinus Math, Arimo, DejaVu Sans Mono

Typesetting software: Xe<sub>La</sub>T<sub>E</sub>X

Language Science Press

Unter den Linden 6

10099 Berlin, Germany

[langsci-press.org](http://langsci-press.org)

Storage and cataloguing done by FU Berlin

no logo

# Contents



# Introduction

Karsten Schmidtke-Bode

Leipzig University

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

- (1) With overwhelmingly greater than chance frequency, languages with normal SOV order are postpositional. (**Greenberg1963**)
- (2) A language never has more gender categories in nonsingular numbers than in the singular. (**Greenberg1963**)
- (3) If a language uses an overt inflection for the singular, then it also uses an overt inflection for the plural. (**Croft2003** 89, based on **Greenberg1966** 28)
- (4) In their historical evolution, languages are more likely to maintain and develop non-ergative case-marking systems (treating S and A alike) than ergative case-marking systems (splitting S and A). (**BickelEtAl2015** 5)

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



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

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

---

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



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

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

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

plural overtly signalled.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

## Acknowledgements

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



## Chapter 1

# Taking diachronic evidence seriously: Result-oriented vs. source-oriented explanations of typological universals

Sonia Cristofaro

University of Pavia

Classical explanations of typological universals are result-oriented, in that particular grammatical configurations are assumed to arise because of principles of optimization of grammatical structure that favor those configurations as opposed to others. These explanations, however, are based on the synchronic properties of individual configurations, not the actual diachronic processes that give rise to these configurations cross-linguistically. The paper argues that the available evidence about these processes challenges result-oriented explanations of typological universals in two major ways. First, individual grammatical configurations arise because of principles pertaining to the properties of particular source constructions and developmental mechanisms, rather than properties of the configuration in itself. Second, individual configurations arise through several distinct diachronic processes, which do not obviously reflect some general principle. These facts point to a new research agenda for typology, one focusing on what source constructions and developmental mechanisms play a role in the shaping of individual cross-linguistic patterns, rather than the synchronic properties of the pattern in itself.

## 1 Introduction

In the functional-typological approach that originated from the work of Joseph Greenberg, language universals (henceforth, typological universals) are skewed cross-linguistic distributional patterns whereby languages recurrently display certain grammatical configurations as opposed to others. Explanations for these patterns are usually result-oriented, in the sense that at least some of the relevant configurations are assumed to arise because of some postulated principle of



grammatical structure, which favors those particular configurations and disfavor other logically possible ones.

For example, a number of word order correlations have been explained by assuming that speakers will recurrently select particular word orders as opposed to others because these orders lead to syntactic structures that are easier to process (Hawkins2004 among others). Another case in point is provided by explanations of the use of explicit marking for different grammatical meanings, for example the use of overt marking for different number values, or that of dedicated case marking for different NP types occurring in particular argument roles. Cross-linguistically, explicit marking may be restricted to less frequent meanings, for example plural rather than singular, animate rather than inanimate P arguments, or inanimate rather than animate A arguments, but is usually not restricted to more frequent meanings. This has been assumed to reflect a principle of economy whereby speakers will tend to use explicit marking only when they really need to do so. Explicit marking can be restricted to less frequent meanings because more frequent ones are easier to identify, and hence less in need to be disambiguated (Greenberg1966; Corbett2000; TU2); Martinmarkedness; Haspelmath2008).

These explanations are based on the synchronic properties of the relevant distributional patterns, not the actual diachronic processes that shape these distributions from one language to another. For example, assumptions about the role of processing ease in determining particular word order correlations are based on the synchronic syntactic configurations produced by particular word orders, not the actual diachronic origins of these orders from one language to another. Similarly, the idea that the use of explicit marking reflects economy is based on the synchronic cross-linguistic distribution of particular constructions across different contexts (e.g. zero vs. overt marking across singular and plural, dedicated case marking across animate and inanimate A and P arguments), not the actual diachronic processes that give rise to this distribution in individual languages.

This paper discusses various types of diachronic evidence about the cross-linguistic origins of two phenomena that have been described in terms of typological universals, the distribution of accusative vs. ergative case marking alignment across different NP types and that of zero vs. overt marking across singular and plural.

This evidence, it will be argued, challenges classical, result-oriented explanations of typological universals in two major ways. First, recurrent grammatical configurations cross-linguistically do not appear to arise because of principles that favor those particular configurations in themselves. This challenges the idea that these principles play a role in the emergence of the distributional



patterns described by the relevant universals. Second, individual configurations arise through several distinct diachronic processes, which do not obviously reflect some general principle. This challenges the idea that explanations for particular distributional patterns can be read off from the synchronic properties of the relevant grammatical configurations, because these properties can originate differently in different cases. These facts call for a source-oriented approach to typological universals, one in which the patterns described by individual universals are accounted for in terms of the actual diachronic processes that give rise to the pattern, rather than the synchronic properties of the pattern in itself.

## 2 The animacy/referential hierarchy: Some possible origins of alignment splits in case marking

One of the most famous typological universals is the animacy/referential hierarchy in (1):

- (1) 1st person pronouns > 2nd person pronouns > 3rd person pronouns >  
human > animate > inanimate (TU2 among others)

Among other phenomena, this hierarchy captures some recurrent splits in the distribution of accusative and ergative case marking alignment across different NP types. Accusative alignment can be restricted to a left end portion of the hierarchy (e.g. pronouns, pronouns and animate nouns), but is usually not restricted to a right end portion of the hierarchy (e.g. inanimate nouns, nouns as opposed to pronouns). Conversely, ergative alignment is sometimes restricted to a right end portion of the hierarchy (e.g. inanimate nouns, nouns as opposed to pronouns, nouns and 3rd person pronouns), but is usually not restricted to a left end portion of the hierarchy (1st/2nd person pronouns, pronouns as opposed to nouns, pronouns and animate nouns).

A classical result-oriented explanation for this distribution invokes the economy principle mentioned in Section 1. Speakers tend to use dedicated case marking only when it is really needed, that is, when some grammatical role is more in need of disambiguation. The NPs towards the right end of the hierarchy (inanimates, nouns as opposed to pronouns) are more likely to occur as P arguments, hence, when they do, the P role is relatively easy to identify, and hence less in need of disambiguation. Dedicated case marking for P arguments, leading to accusative alignment, may then be limited to the NPs towards the left end of the hierarchy (pronouns, animate nouns). By contrast, these NPs are more

likely to occur as A arguments, hence, when they do, the A role is less in need of disambiguation. Dedicated case marking for A arguments, leading to ergative alignment, may then be limited to the NPs towards the right end of the hierarchy (Silverstein1976; Dixonergativity; Dixon1994; Comrie2; DeLancey1981; Song2001; TU2).<sup>1</sup>

This explanation, however, is not supported by the available diachronic evidence about the origins of the relevant grammatical configurations cross-linguistically.

In many cases where accusative or ergative alignment is restricted to particular NP types, the relevant alignment pattern is a result of the development of an accusative or ergative marker through the reinterpretation of a pre-existing element with similar distributional restrictions. In some cases, for example, accusative markers restricted to pronominal, animate or definite direct objects are structurally identical to topic markers. This is illustrated in (2) for Kanuri.

(2) Kanuri (Nilo-Saharan; **Kanuri**)

- a. Músa shí-ga cúro.  
Musa 3SG-ACC saw  
'Musa saw him.'
- b. wú-ga  
1SG-as.for  
'as for me'

In such cases, the accusative marker plausibly originates from the topic marker in contexts where the latter refers to a P argument and is reinterpreted as a marker for this argument ('As for X' > 'X ACC': see, for example, Rohlfs1984 and Pensado1995 for Romance languages, and König2008 for several African languages). Topics are usually pronominal, animate and definite, so topic markers are mainly used in the same contexts as the resulting accusative markers.

Ergative markers not applying to first and second person pronouns have been shown to originate from various types of source elements not applying to these pronouns either. Sometimes, for example, the ergative marker is derived from an indexical element, such as a demonstrative or a third person pronoun, as illustrated in (3) for Bagandji. McGregor2006; McGregor2008 argues that in such cases the indexical element is originally used to emphasize the referent of the A argument, as this referent is a new or unexpected agent. This strategy does

---

<sup>1</sup>Following a standard practice in typology (see, for example, Comrie2 or Dixon1994), the labels A, P and S are used throughout the paper to refer to the two arguments of transitive verbs and the only argument of intransitive verbs.

not apply to first and second person pronouns because the referents of these pronouns are typically expected agents.

- (3) Bagandji (Australian: **Bagandji**)  
 Yaḍu-**ḍuru**      gāndi-d-uru-ana.  
 wind-DEM/ERG carry-FUT-3SG.SBJ-3SG.OBJ  
 ‘This wind will carry it along / The wind will carry it along.’

In other cases, the ergative marker is derived from a marker used to encode instruments in transitive sentences with no overt third person arguments. In these sentences, the instrument can be reinterpreted as an agent, thus evolving into the A argument of the sentence. As a result, the marker originally used for the instrument becomes an ergative marker. This process has been reconstructed by **Mithun2005** for Hanis Coos, illustrated in (4). Instruments are typically inanimate, so the relevant markers do not usually occur with first and second person pronouns.

- (4) Hanis Coos (Coosan; **Mithun2005**)  
 K’win-t    ɣ=mił:aqəṭš.  
 shoot-TR OBL/ERG=arrow  
 ‘An arrow shot (him).’ (from ‘(He) shot at him with an arrow’)

Restrictions in the distribution of particular alignment patterns across different NP types can also be a result of phonological processes targeting a subset of these NPs. In English, for example, accusative alignment became restricted to pronouns as a result of nouns losing the relevant inflectional distinctions due to sound change, as illustrated in Table 1.

Table 1: Pronominal and nominal declension in late Middle English (**Blake2001**)

	1st person	‘ name’
NOM	<i>ik</i>	<i>name</i>
ACC	<i>mē</i>	<i>name</i> (from <i>naman</i> )

In Louisiana Creole, A, S and P arguments were originally undifferentiated for both nouns and pronouns. Pronominal A and S forms, however, underwent phonological reduction, plausibly due to their high discourse frequency. As a result, as can be seen from Table 2, pronouns developed distinct forms for A and S arguments on the one hand and P arguments on the other, while nominal A, S, and P arguments remained undifferentiated. This led to an accusative

case marking alignment pattern restricted to pronouns (Haspelmath & the APiCS Consortium 2013).

Table 2: Pronominal declension in Louisiana Creole (Haspelmath & the APiCS Consortium 2013: 233)

		Subject	Object
Louisiana Creole	1SG	<b>mo</b>	<i>mwa</i>
	2SG	<b>to</b>	<i>twa</i>

These various processes do not appear to be triggered by the fact that, in the resulting grammatical configurations, dedicated case marking is restricted to roles more in need of disambiguation. In some cases, a pre-existing element is reinterpreted as a marker for a co-occurring argument. Topic markers are reinterpreted as markers for a co-occurring P argument, and demonstratives and third person pronouns are reinterpreted as markers for a co-occurring A argument. This is a metonymization process triggered by the contextual co-occurrence of the relevant elements. In other cases, a pre-existing element evolves into a case marker as a result of the reanalysis of the argument structure of the construction. Such processes are plausibly due to meaning similarities between the source construction and the resulting construction, for example, instruments can be reinterpreted as agents because of their role in the action being described, particularly in the absence of an overtly expressed agent. In yet other cases, an existing alignment pattern becomes restricted to particular NP types because other NPs, due to their phonological properties, lose their inflectional distinctions as a result of regular sound change. Finally, particular NPs may develop distinct forms for some argument roles as a result of the original forms undergoing phonological reduction due to their discourse frequency.

Restrictions in the distribution of accusative and ergative alignment, as determined by individual processes, directly follow from restrictions in the distribution of various source constructions, or in the domain of application of particular developmental mechanisms (such as particular phonological processes). These restrictions too, then, cannot actually be taken as evidence for principles that favor the resulting grammatical configurations independently of particular source constructions and developmental mechanisms. This is further supported by the fact that, when the source constructions or the developmental mechanisms involved are not subject to particular distributional restrictions, the distribution of accusative or ergative alignment does not display those restrictions either.

For example, accusative markers sometimes originate from ‘take’ verbs in constructions of the type ‘Take X and Verb (X)’, where the ‘take’ verb is reanalyzed

as a marker for its former P argument (Lord1993; Chappell2013 among several others). The P arguments of ‘take’ verbs can be pronominal, nominal, animate or inanimate (e.g. ‘take him, it, the child, the sword’), and the resulting accusative markers apply to all of these argument types. This is illustrated for Twi in (5), where the accusative marker *de*, derived from a ‘take’ verb, applies to both animate and inanimate P arguments.

- (5) Twi (Niger-Congo; Lord1993)
- a. Wɔ-**de** no yɛɛ ɔsafohéne.  
they-ACC him make captain  
‘They made him captain.’
  - b. O- **de** afoa ce boha-m.  
he-ACC sword put scabbard-inside  
‘He put the sword into the scabbard.’

Accusative and ergative markers can also develop from the reanalysis of possessor or oblique markers used on the notional A or P arguments of various types of source constructions, for example, ‘X is occupied with the Verbing of Y’ > ‘X is Verbing Y ACC’, ‘To X it will be the Verbing of Y’ > ‘X ERG will Verb Y’, ‘Y is X’s Verbed thing’, ‘Y is Verbed by X’ > ‘X ERG Verbed Y’. These processes have been described for a wide variety of languages (see, for example, Harris & Campbell 1995; Bubenik1998 Gildea1998; Creissels2008). In such cases too, the relevant A and P arguments can be nominal, pronominal, animate or inanimate NPs (e.g. ‘The Verbing of you, of it, of the house’; ‘You are Verbed, the house is Verbed’). The markers used for these arguments, then, will be used with all of these NPs, and the resulting accusative or ergative markers are used with all of these NPs too. This is illustrated in (6) and (7), where accusative and ergative markers derived in this way are used, respectively, with nominal inanimate and pronominal animate arguments.

- (6) Wayana (Carib; Gildea1998) i-pakoro-**n** iri pək wai.  
1-house-ACC make occupied.with 1.be  
‘I’m (occupied with) making my house.’ (originally ‘I am occupied with my house’s making.’)
- (7) Cariña (Carib; Gildea1998) A-eena-ri i-’**wa**-ma.  
2-have-NMLZ 1-ERG-3.be  
‘I will have you.’ (from a nominalized construction ‘To me it will be your having.’)

On a similar note, loss of inflectional distinctions through sound change, leading to the loss of particular alignment patterns, targets specific forms because of their phonological properties. This process, then, can affect different NP types cross-linguistically, provided that the relevant forms have specific phonological properties. This leads to different distributional restrictions for particular alignment patterns. In English, as detailed earlier, the process affected nouns as opposed to pronouns, leading to accusative alignment becoming restricted to pronouns. In Nganasan, however, a combination of sound change and analogical levelling led to a loss of inflectional distinctions for pronouns, but not for nouns (Filimonova2005). As a result, as can be seen from (8), accusative alignment became restricted to nouns, even though this configuration should be disfavored in terms of economy, because nominal P arguments are less in need of disambiguation than pronominal ones.

(8) Nganasan (Uralic; Filimonova2005)

- a. **Mənə** nanuntə            mintəl'i-ʔə-ŋ.  
1SG    2SG.LOC-INSTR take-INDEF-2SG  
'You have taken me with you.' (pronominals originally had dedicated accusative forms, e.g. *mənə-m* '1SG-ACC')
- b. ɲüləʒə tundi-**m** tandarku-čü.  
wolf    fox-ACC chase-3SG.A  
'The wolf is chasing the fox.'

If there were principles that favor or disfavor particular distributional restrictions for accusative and ergative alignment because of properties of the resulting grammatical configurations, we would not expect the development of these restrictions to be tied to specific source constructions and developmental mechanisms.

Finally, individual distributional restrictions develop through several distinct processes, which are rather different in nature and provide independent motivations for the restriction. In some cases, particular restrictions arise as accusative and ergative case markers develop through processes of context-driven reinterpretation of various types of source elements, which, for different reasons, are restricted in the same way. In other cases, the restrictions reflect the domain of application of different phonological processes. To the extent that different diachronic processes provide different motivations for particular distributional restrictions, explanations for these restrictions cannot be read off from the restrictions in themselves, because these can originate differently in different cases.

### 3 Some possible origins of zero vs. overt marking for singular and plural

Another well-known typological universal pertains to the use of zero vs. overt marking for singular and plural. Languages can use overt marking for plural and zero marking for singular, but usually not the other way round. A classical, result-oriented explanation for this pattern, as mentioned in Section 1, is in terms of economy. Speakers tend to use overt marking only for meanings that are more in need of disambiguation, and plural is more in need of disambiguation than singular due to its lower discourse frequency. As a result, overt marking can be limited to plural, whereas it will not be limited to singular (Greenberg1966; TU2 Haspelmath2008). This explanation, however, is not supported by a number of diachronic processes that lead languages to have zero marked singulars and overtly marked plurals.

Often, in languages which make no distinction between singular and plural, an overt plural marker evolves through the reinterpretation of pre-existing expressions, whereas singulars retain zero marking. Sometimes, some expression takes on a plural meaning originally associated with a co-occurring expression. For example, in partitive constructions involving plural quantifiers ('many of them' > 'they PL'), the partitive marker can take on the meaning of plurality associated with the quantifier as the latter is lost. This process took place in Bengali, as illustrated in (9).

(9) Bengali (Indo-Aryan; Chatterji)

- a. āmhā-rā sâbâ  
we-GEN all  
'all of us' (14th century)
- b. chēlē-rā  
child-GEN  
'children' (15th century)

In other cases, plurality becomes the central meaning of expressions inherently or contextually associated with this notion but originally used to encode other meanings, for example, distributive expressions ('house here and there') or expressions of multitude ('all', 'people'). This is illustrated in (10) and (11).

- (10) Southern Paiute (Uto-Aztec; Sapir 1930–1931: 258) qa'nɪ /  
house /

qəŋqə'nɪ  
house.DISTR  
'house, houses'

- (11) Maithili (Indo-Aryan: **Maithili**1997) jən səb  
laborer all  
'laborers'

Another process that leads languages to have zero marking for singular and overt marking for plural is the elimination of an overt singular marker through regular sound change in a situation where both singular and plural are originally overtly marked. This was, for example, the case in English, where singular and plural were both originally overtly marked in most cases. The current configuration with zero marked singulars and -s marked plurals resulted from a series of sound changes that led to the elimination of all inflectional endings except genitive singular -s and plural -es (**Mosse**).

These various processes do not appear to be triggered by the higher need to disambiguate plural as opposed to singular. In some cases, an overt plural marker arises as a result of a metonymization process whereby plural meaning is transferred from a quantifier to some other component of a complex expression. This is plausibly due to the co-occurrence of the two. In other cases, some expressions evolves into a plural marker because it is contextually or inherently associated with the notion of plurality, and this notion becomes the central meaning of the expression as some other meaning component is bleached. In yet other cases, a pre-existing overt singular marker is eliminated due to regular sound changes driven by the phonological properties of the marker.

The end result of the various processes, the use of overt marking for plural rather than singular, is directly motivated in terms of the properties of particular source constructions and developmental mechanisms. In many cases, an overt marker is used for plural because the source construction is one associated with the notion of plurality. Alternatively, sound changes leading to the elimination of an overt marker target singular rather than plural markers due to the phonological properties of the former. The fact that overt marking is restricted to plural, then, cannot be taken as evidence for principles that favor this particular configuration independently of particular source constructions and developmental mechanisms. As in the case of accusative and ergative case marking alignment, this point is further supported by the fact that other source constructions and developmental mechanisms give rise to different configurations, that is, overt marking for both singular and plural, or just for singular.



A case in point is provided by Kxoe, illustrated in Table 3 below. This language has gender markers derived from third person pronouns **Heine1982**. As the pronouns have overt singular and plural forms, the resulting gender markers also encode singular and plural, so that the language has overt marking not only for plural, but also for singular.

Table 3: Gender/number markers and third person pronouns in Kxoe (Khoisan; **Heine1982**)

		SG	PL	
Nouns	M	/õa-mà	/õa-//u‘a	‘boy’
	F	/õa-hè	/õa-dji	‘girl’
	C	/õa-(‘à), /õa-dji	õa-nà	‘child’
Pronouns	M	xà-á, á-mà, i-mà	xà-//uá, á-//uá, í-//uá	‘he’
	F	xà-hè, á-hè, i-hè	xà-djí, á-djí, í-djí	‘she’
	C	(xa-‘à)	xà-nà, á-nà, í-nà	‘it’

Also, as described above, partitive case markers can evolve into plural markers by taking on the plural meaning associated with a co-occurring plural quantifier. In expressions where the quantifier is singular (‘one of them’), however, this same process can lead to the development of singular markers, sometimes leading to a configuration where only singular is overtly marked. This was the case in Imonda, which has zero marked plurals, but developed an overt non-plural (singular and dual) marker from a source case marker used in partitive constructions (**Imonda**).

(12) Imonda (Border; **Imonda**)

- a. Agõ-ianèi-m      ainam    fa-i-kõhõ.  
                                  women-NONPL-GL    quickly CLF-LNK-go  
                                  ‘He grabbed the woman.’
- b. mag-m ad-ianèi-m  
                                  one-GL boy-SRC-GL  
                                  ‘to one of the boys’

Similar observations apply to loss of number markers through sound change. This process can affect either singular or plural markers depending on the phonological properties of the marker. From one language to another, then, the process may lead either to zero marked singulars and overtly marked plurals, as detailed above for English, or to the opposite configuration. In the Indo-Aryan language

Sinhala, for example, some inanimate nouns have overtly marked singulars and zero marked plurals (e.g. *pot-a/ pot* ‘book-SG/ book.PL’). This was a result of sound changes leading to the loss of the plural ending of a specific inflectional class in the ancestor language Nitz & Nordhoff 2010: 250–6. Similarly, in Nchanti, a Beoid language, nouns in classes 3/4 have overt marking in the singular and zero marking in the plural, e.g. *kʷāŋ/ kāŋ* ‘firewood.SG/ firewood.PL, *kʷēē/ kēē* ‘moon.SG/ moon.PL’. Originally, both singular and plural were marked overtly through the two prefixes *\*u-* and *\*i-* respectively. As these were eliminated, the singular prefix led to the labialization of the initial consonant of the stem, while the plural prefix left no trace **Hombert1980**

Finally, just like distributional restrictions for accusative and ergative case marking alignment, the fact that a language uses zero marking for singular and overt marking for plural can be a result of a variety of diachronic processes, which lead to this particular configuration for different reasons. In many cases, both singular and plural are originally zero marked (i.e., the language makes no distinction between the two), but zero marking becomes restricted to singular because different expressions, for different reasons, evolve into plural markers. In other cases, both singular and plural are originally overtly marked, and sound change leads to the loss of singular markers due to their phonological properties. Explanations for why overt marking is restricted to plural, then, cannot be read off from this configuration in itself, because it can originate differently in different cases.

## 4 Rare grammatical configurations and result-oriented explanations

Result-oriented explanations of typological universals are crucially based on the fact that certain logically possible grammatical configurations are significantly rarer than others in the world’s languages. This is usually accounted for by postulating principles that both disfavor those configurations and favor some of the other configurations. For example, the rarity of configurations where singular is overtly marked and plural is zero marked is assumed to be due to the fact that these configurations are disfavored by economy, and hence will usually not occur in the world’s languages. This same principle is assumed to also favor the opposite configuration, zero marking for singular and overt marking for plural, thus providing a motivation for the occurrence of this configuration.

Haspelmath (this volume) uses this line of reasoning to claim that result-oriented explanations should be invoked even in cases where the development of some

grammatical configuration is accounted for by the properties of particular source constructions or developmental mechanisms, rather than synchronic properties of the configuration in itself. Haspelmath concedes that, in such cases, there is no direct evidence that the occurrence of the configuration is motivated by principles pertaining to its synchronic properties (functional-adaptive principles, in his terminology). He argues, however, that this hypothesis is supported by two types of indirect evidence: the fact that other logically possible configurations are significantly rarer, and what he calls multi-convergence, the fact that different diachronic processes all lead to that particular configuration. According to Haspelmath, these facts can only be accounted for by assuming that the occurrence of the configuration is ultimately motivated by principles that favor that configuration independently of the diachronic processes that give rise to it. Haspelmath draws a parallel with the notion of adaptiveness in evolutionary biology (and other domains): The development of particular traits is independent of the fact that those traits are adaptive to the environment, in the sense of conferring an evolutionary advantage to the organisms carrying them, but adaptiveness provides the ultimate explanation for their spread and survival in a population.

There is, however, a logical fallacy in the idea that, if some principle motivates the non-occurrence of some configuration (and hence its rarity), then the occurrence of some other configuration is motivated by the same principle. The fact that some principle A provides the motivation for some phenomenon X can be framed as a logical implication,  $X \rightarrow A$  (because X will always involve A, unless other motivations for X are also postulated). This implication means, however, that the absence of A will lead to phenomena different than X, that is,  $\sim A \rightarrow \sim X$ , not that phenomena different from X are also motivated by A. This would be a distinct logical implication,  $\sim X \rightarrow A$ , with a different truth table. For example, if the non-occurrence of configurations where singular is overtly marked and plural is zero marked (X) is assumed to be due to economy (A), this means that principles other than economy ( $\sim A$ ) will lead to the occurrence of other configurations ( $\sim X$ ), not that the latter phenomenon is also due to economy. This undermines the general logic of result-oriented explanations, including Haspelmath's argument: From the fact that some principle provides a motivation for the non-occurrence of some configuration, we cannot conclude that it also provides a motivation for the occurrence of other configurations.

As for the multi-convergence argument, this ignores the fact that different diachronic processes can all lead to the same synchronic output for different reasons, as detailed in Sections 2 and 3. If the same synchronic output is motivated differently in different cases, multi-convergence cannot be taken as evidence for

principles that favor that output independently of the individual processes that give rise to it. Instead, to the extent that the various processes recurrently take place in different languages, the cross-linguistic distribution of the output will be a combined result of the effects of each process.

From a logical point of view, source-oriented explanations do not rule out that the cross-linguistic distribution of particular grammatical configurations may ultimately also be determined by properties pertaining to the synchronic properties of the configuration, as assumed by Haspelmath. For example, these factors could play a role in the transmission of the configuration from one speaker to another, or its retention across different generations of speakers. This would be the equivalent of the notion of adaptive evolution through natural selection in evolutionary biology: particular genetic traits do not develop because they confer an evolutionary advantage to the organisms carrying them, but this provides the ultimate explanation for their distribution in a population.<sup>2</sup>

In evolutionary biology, however, there is direct evidence for adaptiveness, in that particular genetic traits make it demonstrably more likely for the organisms carrying them to survive and pass them on to their descendants. For languages, on the other hand, there is generally no evidence that the fact that some grammatical configuration conforms to the principles postulated in result-oriented explanations, for example economy, makes it more likely for that configuration to spread and survive in a speech community. In fact, there is a long tradition of linguistic thought in which the propagation of individual constructions within a speech community is entirely determined by social factors independent of particular inherent properties of the construction (see, for example, **McMahon1994** and **BillLC** for reviews of the relevant issues and literature).

In principle, there is another sense in which particular grammatical configurations could be adaptive. While individual configurations directly reflect the properties of particular source constructions or developmental mechanisms, it could be the case that the specific diachronic processes that give rise to the configuration are ultimately favored by principles pertaining to its synchronic properties. For example, different processes of context-driven reinterpretation leading to overt marking for less frequent types of argument roles could be favored by the need to give overt expression to these roles. Similarly, different processes lead-

---

<sup>2</sup> A referee suggests that this is similar to Lass's **Lass1990** use of the notion of exaptation: particular grammatical traits may lose their original function, but they are retained in the language because they are deployed for novel functions. This, however, is meant to account for why particular traits survive in a language despite losing their original function, not why they are selected over others, as is the case with result-oriented explanations of typological universals and explanations of biological evolution in terms of adaptiveness through natural selection.

ing to zero marking for singulars (zero marking becoming restricted to singular due to the development of an overt plural marker, phonological erosion of an existing overt singular marker) could be favored by the lower need to give overt expression to singular as opposed to plural.

These assumptions, however, are not part of any standard account of the relevant processes in studies of language change (see Bybee et al. 1994: 298–300 and **Slobin2002** for an explicit rejection of this view in regard to grammaticalization, as well as **Otadependencies** for more discussion). In fact, diachrony provides specific evidence against the idea that particular grammatical configurations develop both because of properties of particular source constructions or developmental mechanisms and because of principles that favor the configuration in itself. As detailed in Sections 2 and 3, different source constructions and developmental mechanisms give rise to different grammatical configurations, even when this goes against some postulated principle that favors some of these configurations as opposed to the others. This is not what one would expect if there were principles favoring particular grammatical configurations independently of the specific source constructions or developmental mechanisms that give rise to them.

All this means that, to the extent that a principled source-oriented explanation is available for the occurrence of particular grammatical configurations, explanations in terms of the synchronic properties of the configuration are redundant, because we don't have either direct or indirect evidence for these explanations (see **Blevins2004** for similar arguments in phonology, and **Newmeyer2002**; **Newmeyer2004** for an application of this line of reasoning to Optimality Theory models of typological universals). Of course, one still needs to account for the fact that certain logically possible grammatical configurations are significantly rarer than others. This phenomenon, however, is logically independent of the possible motivations for the occurrence of the more frequent configurations, as detailed above. To the extent that individual grammatical configurations arise due to properties of particular source constructions or developmental mechanisms, any differences in the frequency of particular configurations will reflect differences in the frequency of the source constructions or developmental mechanisms that give rise to those configurations. The higher frequency of particular configurations will then be a result of the higher frequency of the source constructions and developmental mechanisms that give rise to them, while the rarity of other configurations will due to the rarity of possible source constructions or developmental mechanisms for those configurations (see **Harris2008** for an earlier formulation of this point in regard to tripartite case marking alignment). Frequency differences in the occurrence of particular source constructions or de-

developmental mechanisms need to be accounted for, but they need not be related to any properties of the resulting configurations, so they should be investigated independently.

## 5 Concluding remarks

Source-oriented explanations of typological universals are in line with classical views of language change held within grammaticalization studies and historical linguistics in general. These views are manifested, for example, in accounts of the development of tense, aspect and mood systems, or alignment patterns (BybeeEtAl1994; HarrisCampbell1995; Gildea1998; TraugottDasher 2005 among others). In these accounts, grammatical change is usually not related to synchronic properties of the resulting constructions, for example the fact that the use of these constructions complies with some postulated principle of optimization of grammatical structure. Rather, grammatical change is usually a result of the properties of particular source constructions and the contexts in which they are used. In particular, new grammatical constructions recurrently emerge through processes of context-induced reinterpretation of pre-existing ones, and their distribution originally reflects the distribution of the source constructions.

In source-oriented explanations, the patterns captured by typological universals originate from several distinct diachronic processes, which involve different source constructions and developmental mechanisms. These processes recurrently take place in different languages, and are plausibly motivated by the same factors from one language to another. Individual patterns, however, are a combined result of the cross-linguistic frequencies of the various processes, rather than a result of some overarching principle independent of these processes.

While this scenario is more complex and less homogeneous than those assumed in result-oriented explanations, it is consistent with what is known about the actual origins of the relevant grammatical configurations in individual languages, and it makes it possible to address several facts not accounted for in these explanations.

For example, the patterns captured by typological universals usually have exceptions. This is in contrast with the assumption that these patterns reflect principles of optimization of grammatical structure that are valid for all languages, because in this case one has to account for why these principles are violated in some languages. Also, individual principles invoked in result-oriented explanations are often in contrast with some of the grammatical configurations captured by individual universals. For example, the idea that zero marking of more fre-

quent meanings is motivated by economy is in contrast with the fact that these meanings are overtly marked in many languages.

These facts have sometimes been dealt with in terms of competing motivations models, but a general problem with this approach is that it may lead to a proliferation of explanatory principles for which no independent evidence is available (Newmeyer1998 *Otacompetingmotivations* among others). If the patterns captured by typological universals reflect the properties of different source constructions and developmental mechanisms, however, then it is natural that they should have exceptions, because not all languages will have the same source constructions, nor will particular developmental mechanisms be activated in all languages. Principles pertaining to the synchronic properties of the pattern will fail to account for all of the relevant grammatical configurations because the pattern is not actually motivated by those principles.

Over the past decades, several linguists have emphasized the need for source-oriented explanations of typological universals (Bybee1988; Bybee2006; Bybee2008; Aristar1991; Gildea1998; Creissels2008; *Otareferential*; *Otacompetingmotivations*; *Otadependencies*; Anderson2016). This view, however, has not really made its way into the actual typological practice, despite the close integration between typology and studies of language change (a fully fledged research approach along these lines is, on the other hand, the Evolutionary Phonology framework developed in Blevins2004). While diachronic evidence about the origins of the patterns captured by individual universals is much scantier and less systematic than the synchronic evidence about these patterns, it poses specific foundational problems for existing result-oriented explanations of these universals. These problems point to a new research agenda for typology, one focusing on what source constructions and developmental mechanisms play a role in the shaping of individual cross-linguistic patterns, as well as why certain source constructions or developmental mechanisms are rarer than others.

## Abbreviations

The paper conforms to the Leipzig Glossing Rules. Additional abbreviations include:











