Explanation in typology

Diachronic sources, functional motivations and the nature of the evidence

Edited by

Karsten Schmidtke-Bode

Natalia Levshina

Susanne Michaelis

Ilja Seržant

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 Michaelis

Ilja Seržant

Karsten Schmidtke-Bode , Natalia Levshina , Susanne 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 (cf. 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 adap-

---

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

tive 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 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 (cf. 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 (cf. **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 correlations, 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. Whereas some typological characteristics of the structure and position of adverbial clauses can successfully be accounted for by their source constructions (and thereby supplant earlier processing-based explanations), others are still best rendered in functional-adaptive (and hence result-oriented) terms.

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 functional-adaptive pressures may not actually surface in standard typological analyses but do become visible in qualitative data from transition phases. Based on diachronic data from Russian, he 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 synchronic 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

Attractor states and diachronic change in Hawkins' "Processing Typology"

Karsten Schmidtke-Bode

Leipzig University

This paper provides an assessment of John Hawkins' (2004; 2014) programme of explaining cross-linguistic regularities in terms of functional-adaptive principles of efficient information processing. In the first part of the paper, I systematize how such principles may possibly affect the diachronic development of languages, and I argue that evidence for efficient coding can be obtained primarily from the actualization process, rather than the innovation stage that is at the focus of purely source-based approaches to explaining universals. In the second part of the paper, I present a small case study on a specific prediction made in **Hawkins2014**, concerning the typology and diachrony of article morphemes. This will allow us to carve out both strengths and weaknesses of Hawkins' programme in its current manifestation.

1 Introduction

In debating the role of source- and result-oriented explanations in typology, a research programme that merits discussion is John Hawkins' approach to cross-linguistic variation, laid out most comprehensively in **Hawkins1994**, (2004) and (2014). The overarching hypothesis of these works is that many cross-linguistic generalizations about grammatical structure can be explained as adaptations to efficient information processing ("processing ty-pology", cf. **Hawkins2007**). In a nutshell, Hawkins argues that efficient in-formation processing can be achieved by (i) "minimizing domains" in which certain semantic and syntactic relations are processed, (ii) "minimizing forms" whenever their information content is recoverable from the context or long-term statistical knowledge, (iii) arranging elements in such a way that the ultimate message can be transmitted as rapidly

and accurately as possible, i.e. without delays, false predictions, backtracking, etc. These efficiency principles are thus attractors that are assumed to affect linguistic choices in usage events and ultimately also the conventionalized shapes of grammars.

Hawkins' programme is one of the most systematic attempts to ground typological data in psycholinguistic research and to link it to the arena of language use; in this spirit, it is similar, for example, to the work of Bybee (1985; 2010) and Croft (2001; 2003). Moreover, Hawkins' "performance-grammar-correspondence hypothesis", according to which grammatical rules are basically crystallized usage preferences, echoes one of the key tenets of the usage-based theory of language (cf. **Langacker1987**; **KemmerBarlow2000**). And some specific efficiency principles, such as the "minimization of forms" in proportion to their degree of predictability, even have exact parallels in other functional-typological works (e.g. **Haiman1983**; **Croft2003**; **Haspelmath2008**).

At the same time, however, Hawkins' work is not always received uncritically within usage-based linguistics. Among other things, it is couched in a formal phrase-structure architecture that appears to presuppose the existence of many grammatical categories (cf. **Diessel2016**); some of its principles have been criticized for not being truly domain-general but perhaps specific to language (such as a pressure for short constituent recognition domains, cf. **Bybee2010**); and crucially in the present context, Hawkins has also been criticized for neglecting or underestimating the diachronic dimension behind the phenomena he attempts to explain (e.g. **Cristofaro2017**; **Collins2018** [this volume]). But to the extent that Hawkins does make reference to historical developments, the nature and plausibility of his diachronic claims are worth investigating in more detail, which is precisely what the present contribution aims to do.

To this end, the first part of the paper develops a systematization, in the usage-based framework, of how Hawkins' functional-adaptive principles possibly affect the diachronic development of languages. I argue that there is solid evidence for efficient information processing in the moulding of grammar, suggesting that there is a place for result-oriented processes, beside source determination, in accounting for typological distributions. In the second part of the paper, I exemplarily focus on a diachronic prediction made in **Hawkins2014**, according to which languages of different word-order types show markedly different propensities for grammaticalizing definite articles (the prediction will be formulated more precisely as we go along). This miniature case study will not only serve as a testing ground for this specific efficiency-based hypothesis, but also allow us to identify some general merits and potential problems of processing typology.

2 The diachronic dimension in processing typology

In the usage-based approach, language change is conceived as a multi-step process (cf. **Croft2000**, 2006; **Aitchison2013**) that starts by breaking a convention in the form of a linguistic innovation ("altered replication"), followed by the spread of that innovation through both the linguistic system ("diffusion") and the speech community ("propagation"). Hawkins' publications contain a number of indications as to how his efficiency principles influence innovation, diffusion and propagation processes. I will tackle of each of them briefly, in reverse order, as this reflects an increasing degree of explicitness of the respective proposals.[1]

As for propagation processes, Hawkins is usually reticent with regard to the forces that implement efficient structures, even though the central diachronic mechanism in his programme is that of "selection": efficient variants are said to be selected relatively more frequently than their inefficient counterparts, until they may ultimately oust the inefficient ones completely. It is in this way, Hawkins argues, that preferred patterns in performance can conventionalize into grammatical rules, although he concedes in **Hawkins2014** that it is presently poorly understood how exactly this "translation from performance to grammar" works. Now, if one subscribes to the view that propagation is entirely driven by sociolinguistic forces like prestige, solidarity and the resulting accommodation (e.g. **Croft2000**; **Cristofaro2017**, 2018 [this volume]), it remains mysterious, indeed, how Hawkins' very idea of selection processes can fit in.

On the other hand, there are well-known accounts of language change in which propagation is not exclusively a social phenomenon: **Keller1994**'s (**Keller1994**) "Invisible Hand" theory, for example, leaves room for functional considerations in the selection process. Some of Keller's classic examples of invisible-hand processes, such as the emergence of a traffic jam or a short-cutting footpath do not, in fact, involve social motives: People follow a certain course of action because they primarily consider its functional advantages, regardless of the sociolinguistic profile of the person whose behaviour they adopt. **Cristofaro2017** claims that there is no empirical evidence at all for this scenario in linguistics, but this assessment is overly pessimistic: **Rosenbach2008**, in a detailed examination of evolutionary accounts of language change, concludes that "the evidence available does not speak for the *exclusive* role of social factors in the selection process" (**Rosenbach2008**:

---

[1]Although the present section is specifically about Hawkins' work, it actually applies to "functional-adaptive constraints" (**Haspelmath2018** [this volume]) more generally, not least because Hawkins' processing typology draws on and incorporates similarly-minded principles from many other functionalist typologists (e.g. Greenberg, Comrie, Keenan, Givón, Haiman, Croft, Haspelmath, etc.).

44; emphasis in original). Therefore, I currently see no reason to dismiss a priori a theory in which both social and functional selection pressures can be operative in propagation (cf. also **Haspelmath1999**; **Nettle1999**; **Enfield2014** for similar positions).[2] On this view, then, Hawkins' efficiency principles are relevant to, and hence at least partially drive, the propagation process, although empirical evidence that clearly disentangles functional and social selection processes is, of course, very hard to come by (cf. also **Seiler2006**).

The empirical picture is clearer, in my view, when it comes to DIFFUSION or ACTUALIZATION processes, i.e. the spread of an innovation through the linguistic system.[3] Although Hawkins himself does not speak of diffusion or actualization, the process is actually highly germane to his research, as many of the phenomena he discusses in support of his efficiency theory are cases of limited diffusion. In relativization, for example, a well-known pattern is for resumptive pronouns, once they have been innovated, not to spread across the entire range of relativization sites, but to be restricted to certain sections of Keenan & **Comrie1977**'s (**Comrie1977**) accessibility hierarchy (as in Hausa, Hebrew, Welsh and many other languages). Similarly, when object case markers develop and spread within the linguistic system, they typically end up being confined to animate, definite or pronominal objects, rather than being extended across the board (cf., e.g., **Sinnemäki2014** for a quantitative study). Many other cases of such differential marking are collected in **Haspelmath2008** and subsumed by Hawkins (2004; 2014) under his "Minimize Forms" principle: The marker in question is applied to those environments that require more processing effort, and is left out economically elsewhere. Processing effort, in turn, may be related to various factors, notably constraints on working memory (e.g. longer process-

---

[2]Note that recent mathematical models of language change (e.g. **BlytheCroft2012**) clearly show that selection as such is a crucial element of propagation processes, in as far as alternative models of propagation that do not rely on a weighting of linguistic variants (e.g. **Trudgill2004**) do not produce the empirical patterns of propagation that have been established in historical linguistics and sociolinguistics. However, **BlytheCroft2012** also concede that their model cannot distinguish between social and functional factors in selection, i.e. it leaves open which of these is more vital in the propagation process or how they possibly interact.

[3]The term DIFFUSION is best-known in the context of "lexical diffusion" (**Wang1969**), which refers to the successive spread of a phonetic or morphosyntactic innovation to different lexical items (e.g. the Progressive construction to more and more lexical verbs, or final consonant devoicing to all relevant words). In the present paper, I am using the term diffusion in a broader sense, comprising also the application of an innovated grammatical marker or construction to a new morphosyntactic environment (e.g. the extension of *all but* in its historically younger sense 'almost' from adjectival uses (*This was all but remarkable*) to verbal environments (*He all but fell down*), cf. De **Smet2012**). Diffusion is thus synonymous with the term ACTUALIZATION (cf. **Timberlake1977**, **Andersen2001** and many others, most recently De **Smet2012**).

ing domains correlating with resumptive pronouns) and the relative unexpectedness ('surprisal') of a given configuration (cf. **NorcliffeEtAl2015** on memory- and expectation-based processing in cross-linguistic perspective). For example, discourse participants and animate entities are more likely to be subjects than objects, hence it is precisely these kinds of objects that are more surprising. Paired with the Hawkinsian assumption of efficiency on the part of the speaker, it is also only these objects that need to be marked overtly. A similar surprisal-based account is provided by **Haig2018** to explain "why differential object indexing is an attractor state" (**Haig2018**) in the grammaticalization of object pronouns.

In Hawkins' programme, then, all of these cases are amenable to an explanation in terms of efficient information processing. I believe that this account is presently superior to purely source-oriented typologies such as Cristofaro's, for the following reasons.

Firstly, there is solid evidence for efficiency where the occurrence of a particular marker is optional. This can be observed, for example, with variable relativizers that, other things being equal, show up less frequently when a relative clause is statistically expected given the previous co-text, and vice versa (cf. **WasowEtAl2011**). As **FoxThompson2007** observe, a sentence like

(1)   *This was the ugliest set of shoes* [*I ever saw in my life*].

would sound "quite awkward" (**WasowEtAl2011**: 181) if the relative clause were introduced by *that*; according to Wasow et al., this is precisely because a relative clause is expected in this context, which is in turn why relative *that* tends to be omitted efficiently. **Jaeger2010** shows that similar predictability effects account for a large portion of the variability of the English complementizer *that*.

Importantly, the same kinds of effect also show up in psycholinguistic experimentation, and in languages other than English. For example, recent studies have shown that optional case marking in Japanese, optional indexation in Yucatec Maya relative clauses or optional plural marking in an artificial language exhibit an efficient distribution in the participants' linguistic behaviour, other things being tightly controlled for (cf. **KurumadaJaeger2015**; **NorcliffeJaeger2016**; **KurumadaGrimm2017**). All of these synchronic effects are independent of the historical source of the respective marker. In other words, no matter how a particular relativizer emerges, it comes to be applied in ways that are consonant with Hawkins' efficiency predictions. And as, for example, Seržant (2018 [this volume]) shows, such optional marking can conventionalize into more fixed grammatical patterns over time.

Secondly, to the extent that the "Minimize Form" effects are typologically

sound (i.e. independent of geographical and genetic affiliations), they are in contrast with a powerful principle that we observe elsewhere in grammars, viz. the potent force of analogy (cf. **GentnerSmith2012**; **BlevinsBlevins2009**). Analogy is the driving force behind lexical diffusion, and where it runs to (near-)completion, the result is a productive grammatical rule in the traditional sense. Time and again, historical studies show just how sweeping analogical extension can be: By incremental diffusion processes, English has conventionalized a rule that every main clause requires an overt subject, and every lexical verb now needs *do*-support if it is to occur in an interrogative clause. In other languages, split alignment systems are gradually being eliminated in favour of unified marking: for example, younger speakers of Choctaw (Muskogean: USA) are in the process of re-shaping split intransitivity into a nominative-accusative system with consistent coding for the S argument (**Broadwell2006**); and **Creissels2018** argues more generally that there are strong analogical pressures on languages to regain consistent alignment patterns if these get disrupted by grammaticalization processes.[4]

In view of these analogical forces, one may wonder why systems of differential resumption, differential object marking or differential possessive marking exist quite pervasively, and in highly systematic ways. In Cristofaro's account, they are persistence effects, i.e. they are all due to the fact that the unmarked meanings are perceived as incompatible with the source construction. For example, when an object marker originates from a topic marker, it is expected to be restricted to object NPs whose properties are most closely associated with topicality (or topic-worthiness), such as pronominal, animate and definite entities, and not to apply elsewhere. In fact, **DalrympleNikolaeva2011** argue that such erstwhile topic markers are often extended to animate and/or definite objects, thus diffusing in principled ways to create DOM patterns that may plausibly be linked to the source construction. However, given the powers of analogy, why does diffusion stop there? If it is really the source construction pulling its weight here, one may wonder why it does not do so in many other cases.

Just consider what is perhaps the textbook example of a development that standardly overrides effects from the source construction, namely diffusion processes in grammaticalization. It is by analogical extension that the *going-to*-future has spread to inanimate subjects (*The icicle is going to break off.*), and that the French negative marker *pas* has been extended beyond contexts of directed motion (*Je ne vais pas* > *Je ne sais pas.*). In other words, if analogy works in many other

---

[4]For example, "many languages in which the grammaticalization of a new TAM form resulted in [tense-based split ergative alignment] have undergone a subsequent evolution that can be characterized as regularization under the pressure of analogy" (**Creissels2018**).

instances of grammaticalization, why not in those cases that involve differential marking? Siding with Hawkins (and **Haspelmath2018** [this volume]) here, I find it convincing that an attractor state of efficient coding shapes the development of grammatical systems, especially in light of the behavioural evidence cited above.

Thirdly, while I agree with Cristofaro (2018 [this volume]) that functional principles should be "visible" in the diachronic development of particular structures, I find her interpretation of this requirement too narrow: She demands that the alleged motivations be present at the innovation stage of a grammatical construction and hence directly influence its emergence. But as we have just seen, it is often during the actualization phase that functional-adaptive principles are operative, regardless of how or why a given marker originated in the first place (cf. also Seržant 2018 [this volume]).

Interestingly, while Joan Bybee is now often cited as a representative of source-oriented typology, her relevant publications (e.g. **Bybee1988**; 2008) reveal a broader perspective than Cristofaro's: "Identifying the causal mechanisms [that lead to typological generalizations] requires a detailed look at *all* the properties of a change – including its directionality, gradualness, spread through the community and through the lexicon" (**Bybee2008**). Crucially, it is precisely in lexical-diffusion processes that many of her well-known frequency effects apply: For example, Bybee's "conserving effect" of token frequency explains why highly entrenched main verbs like *speak, think* and *mean* resisted the innovative *do*-support in *wh*-questions for a long time (e.g. *What spekest thou?*, cf. **Ogura1993**), or why the change from *-th* to *-s* in the third-person of English verbs affected the most frequent verbs last (notably *hath* and *doth*, cf. van **Gelderen2014**: 172). In a similar vein, I would thus argue that the diffusion or actualization stage is highly relevant for the kinds of effects that lead to efficient typological marking patterns.

In conclusion, I consider Hawkins' account (and functional-adaptive motivations of similar kinds) capable of explaining why certain changes do *not* happen – particularly, why we find that analogical extensions are systematically brought to a halt even though they are so commonly carried through in other domains of grammar.[5]

Let us finally turn to the realm of INNOVATION, i.e. Hawkins' suggestions as

---

[5]See also **Smith2001** for a similar view: He investigates the diachronic development of agreement loss in Romance participles and argues that while parsing principles cannot be held responsible for the *rise* of participial agreement, they did play a role in its gradual disappearance. Specifically, Smith claims that agreement was retained longest in those environments where it was most beneficial for processing. Therefore, "functionality is here acting as a brake on actualization" (**Smith2001**), just as I argued more generally above.

to why, where and when certain grammatical structures emerge. A central concept here is that of correlated evolution: When a language changes in one part of the grammar, Hawkins often expects to see "ripple effects" (**Hawkins2014**) in domains that are linked to the changing subsystem by certain efficiency principles. For example, since Hawkins assumes that phrases of different types (VPs, PPs, NPs, etc.) show harmonic ordering patterns to allow efficient sentence processing, a change from OV to VO is predicted to engender innovations in PPs and NPs as well (cf. **DunnEtAl2011** and the papers in *Linguistic Typology* 15(2) for ample discussion of this issue). In the present context, perhaps the most interesting claim with regard to innovation is that efficiency principles can predict the occurrence of grammaticalization: While many grammaticalization paths are universal "attractor trajectories" (**BybeeBeckner2015**) – open to all languages with similar source constructions due to the same mechanisms of reanalysis –, Hawkins' efficiency principles predict under which structural conditions (e.g. in which language "types") particular events of grammaticalization are more or less likely to happen. In the remainder of this paper, I will briefly discuss a specific example of such a hypothesis developed by **Hawkins2014**.

3   A test case for processing typology

In his 2014 monograph, Hawkins examines the structure of noun phrases (NPs) from a processing perspective. Across the world's languages, NPs often contain elements in addition to the head noun that, in Hawkins' view, can function as processing cues to the recognition (or online "construction") of an NP, such as articles, classifiers and related morphemes.[6] Hawkins argues that such elements are more efficient in VO languages than in OV languages: As illustrated schematically in Figure **??**, an additional NP constructor C in a VO language can shorten the domain for the construction of the VP (V+NP), especially if N is delayed by intervening material (e.g. in AP-N sequences like *the very delicious meal*). In an

---

[6]Although Hawkins frames the idea of "online construction" in terms of syntactic trees, nodes and categories, the basic intuition behind it is functional in nature: Translated into the usage-based parlance of, e.g., **Croft2001**, **BecknerBybee2009** or **BatesMacWhinney1989**, Hawkins' idea is that a referential expression should be recognizable as such, based on reliable cues in the speech stream. Referential expressions (or NPs, for that matter) are arguably best cued by nouns and determiners, and the construction of an NP is thus facilitated by the early availability of such "constructing categories" within the string of units that ultimately belong to the NP. More generally, I believe that Hawkins is thus actually quite compatible with usage-based and construction-grammatical conceptions of processing, even though he uses terminology that is closely associated with generative syntax.

OV language, by contrast, additional NP constructors lengthen this dependency domain, no matter where they occur in the NP:

Figure 1: V-NP processing in VO- and OV-languages (adapted from **Hawkins2014**: 125)

From these considerations, one might derive the following prediction:

(2)   While all languages have source constructions for articles (notably demonstratives for definite articles and the numeral 'one' for indefinite articles), the grammaticalization of these sources into more general NP markers should be a more productive historical process in VO languages than in OV languages. As a result, the synchronic typological distribution of articles is significantly different in the two language types.

As a matter of fact, **Hawkins2014**' (**Hawkins2014**) prediction is narrower in scope: He applies it only to definite articles, and only to independent definite articles (i.e. words and clitics rather than affixes). The following examples illustrate the language types that are expected to be frequent according to (2):

(3)   a.   VO with definite article
            Maori (Austronesian, Oceanic; **Bauer1993**: 256)
            I   kite ia i   te   whare.
            T/A see 3SG OBJ DET house
            'She saw the house.'

       b.   OV without definite article
            Lezgian (Nakh-Dagestanian, Lezgic; **Haspelmath1993**: 343)
            Ada-z   balk'an aku-na.
            he-DAT horse   see-AOR
            'He saw the horse.'

In support of the hypotheses in (3), Hawkins cites some of **Dryer2005 WALS**'s (**Dryer2005 WALS**) data, which show, indeed, that free-standing definite articles are relatively more frequent in VO-languages (more on the data below).

Hawkins' approach, as illustrated by this specific example, has a number of assets: For instance, it emphasizes the importance of the linear dimension of language, which tightly constrains production and parsing processes but which tended to be neglected by (at least early) cognitive-linguistic and construction-based approaches to grammar (cf. also **Diessel2011** for a similar critique). Hawkins' work is clearly pioneering here, and in the recent usage-based literature, related

notions like contextual predictability, informativity and projective links have come to take a highly prominent place (cf., e.g., **GahlGarnsey2004**; **Levy2008**; **Auer2009**). Furthermore, Hawkins' diachronic thinking adds a new dimension to classic research in grammaticalization. As **Good2008** points out, work on grammaticalization typically offers "permissive explanations [...], that is, it focuses on particular grammaticalization paths without, in general, accounting for what factors will cause one language, but not another, to instantiate those paths." Hawkins' approach elevates this "permissive" nature of explanation to what Good (ibid.) calls a "probabilistic" one: It attempts to explain why certain grammaticalization processes are set in motion only (or preferably) in certain language types or at certain points in time (cf. also **Hawkins1986**, 2012 for representative work along these lines).

But just how convincing are such claims and the empirical support that Hawkins provides for them? In the present case, I have a number of reservations about the picture drawn in **Hawkins2014**.

To begin with, I do not quite see why the hypothesis is restricted to the development of definite articles, as indefinite articles should qualify equally well as NP constructors. Similarly, Hawkins' preoccupation with word-based processing (which is prominent throughout his 2014 book), to the neglect of affixes with identical functions, is not sufficiently motivated. In addition to the problem that free and bound markers are very hard to distinguish consistently for cross-linguistic comparison (**Haspelmath2011**), it remains unclear if there is a measurable psycholinguistic difference between word- and affix-processing. As long as there is no evidence for the view that free and bound definiteness markers are parsed in fundamentally different ways, we should rather take a more embracing approach to the data and ask whether VO- and OV-languages differ in their propensity to grammaticalize article morphemes from their respective source constructions.

With these considerations in mind, the first step of the empirical assessment is, just like in **Hawkins2014**, to examine the typological distribution of article morphemes. Dryer's *WALS* data, in their most recent version, are set out in Table 1:

For the purposes of testing our revised version of Hawkins' hypothesis, we need to discard the languages without a dominant order of V and O ("ndo"), and we basically conflate the figures in the first four rows of Table 1 and contrast them with those in the final row. In other words, (i) we consider both free and bound definiteness morphemes; (ii) we include those languages which are beginning to use a demonstrative like an article (row 3, cf. **Dryer2013a** for details) – thus incorporating cases of incipient grammaticalization; (iii) we include languages

Table 1: Distribution of articles in different word-order types (**Dryer2013**)

	VO	OV	ndo	Totals	
Distinct ART word	144	52	14	210	
DEF affix	49	33	6	88	ART
DEM used as ART	30	33	5	68	
Only INDEF ART	20	24	0	44	
No ART	70	111	14	195	NO ART
Totals	**313**	**253**	**39**	**605**	

with indefinite articles only.[7] The conflated form of the data thus looks like in Table 2:

Table 2: Distribution of articles in different word-order types (reorganized)

	VO	OV	Totals
ART morph	243	142	395
No ART morph	70	111	181
Totals	**313**	**253**	**566**

The distribution in Table 2 looks conspicuously skewed, but of course these are raw data that are not controlled for genetic and areal effects.[8] Therefore, what **Hawkins2014**' (**Hawkins2014**) analysis clearly needs to be augmented with (in this case as well as virtually all others in his book) is proper statistical modelling according to contemporary standards (cf., e.g., **Bickel2011**). To this end, I am seeking converging evidence from two complementary quantitative approaches to the data, namely mixed-effects logistic regression (cf. also **Cysouw2010**; **JaegerEtAl2011**) and Bickel's (2011; 2013) Family Bias Method (which is particularly suitable to testing hypotheses formulated in diachronic terms). In the supplementary ma-

---

[7]Some readers may object to this way of grouping the data. For example, one might reasonably argue that languages in which demonstratives are used with some article-like functions should *not* be said to have "proper" articles (yet). However, even when such languages are classified differently for statistical purposes, the results remain the same in many respects (cf. supplementary material SM3.2).

[8]For similar raw data, cf. also **Dryer2009**, who endorses Hawkins' processing explanation.

terials to this paper[9], I offer a more detailed, non-technical introduction to the Family Bias Method (SM1), as well as the statistical properties of all models (SM2–5). For reasons of space, I here confine myself to describing some major results of the analyses.[10]

Figure 2 shows that there is a significant effect of word order on the occurrence of articles in a mixed-effects regression model (β = -0.73, p < 0.001).[11] Although the model is not particularly good overall, probably missing important further predictors (R^2_c = 0.14, C = 0.72), Hawkins' hypothesized effect is clearly present, as the probability of *not* having articles (y-axis) increases significantly as we go from VO to OV (x-axis):

In the Family Bias estimations, too, it turns out that, among those families that do not just show a chance distribution of articles, VO families are about 2.6 times more likely to develop articles than OV families. This is illustrated in Table 3, and Figure ?? shows that this effect is stable (i.e. never reversed) across all six macro areas:

Table 3: (Rounded) family biases for articles in different word-order types (N_{total} = 217 genetic units, 99 of which are estimated to be "biased" (as opposed to internally diverse); Fisher exact test, p = 0.039)

	VO	OV	Totals
ART morph	50	19	69
No ART morph	15	15	30
Totals	**65**	**34**	**99**

In sum, the global typological picture is consistent with Hawkins' processing account, even when tested against a more comprehensive data set and with more rigorous modes of examination.

Recall, however, that a second prediction of this account is that articles are especially useful in those VO languages that have modifiers before the head noun

---

[9]Cf. http://www.kschmidtkebode.de/publications. #ZENODO

[10]All statistical analyses were performed in *R* 3.3.1 (*R* Development Core **Team2016**). I am grateful to Taras Zakharko and Balthasar Bickel for making their Family Bias algorithm freely available (**ZakharkoBickel2011**+).

[11]All regression analyses I performed are based on generalized linear mixed-effects models that include genealogical and macro-areal dependencies as random effects (cf. SM3). The model in Figure 2, for example, contains by-family and by-area random intercepts for the distribution of articles, while a by-area random slope for the word-order effect did not improve the model significantly and was hence excluded from the final model.
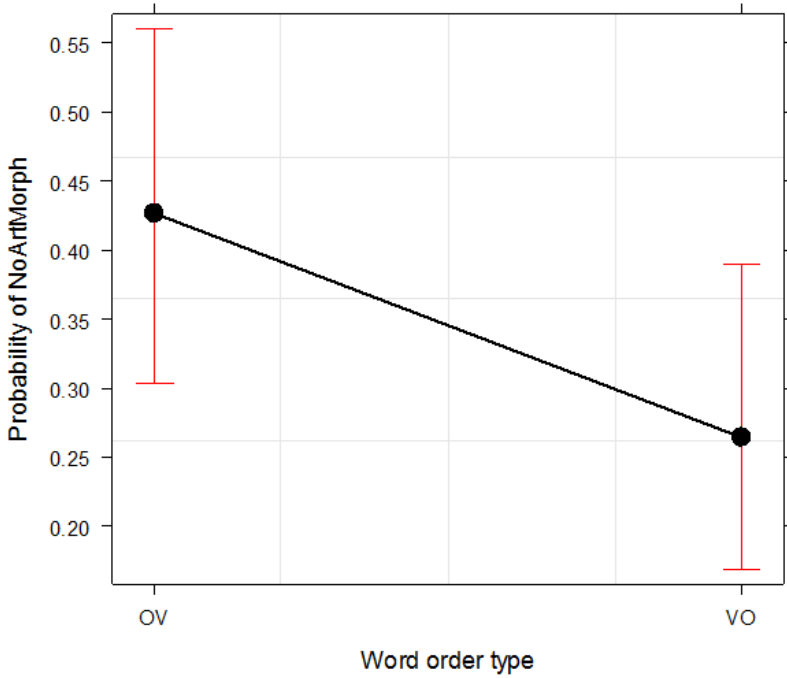
Figure 2: Effect of word order type on the probability of (not) having articles in a mixed-effects model (cf. SM3 for details)

Figure 3: Family biases by macro area (cf. SM2 for details)

in NPs (*a very delicious meal*). One would thus expect, for example, that the grammaticalization of articles is particularly productive in VO languages with ADJ-N order, and, from an efficiency perspective, less so in those with N-ADJ order. I tested this by examining the order of nouns and adjectives (**Dryer2013b**) in all VO languages in the same sample as above (N_{total} = 278 languages). Across several different statistical models (and operationalizations of the hypothesis, cf. SM4), I did not find support for Hawkins' efficiency hypothesis. In one analysis, for example, I probed whether free-standing article words are more likely in VO languages with ADJ-N order than in those with N-ADJ order. Figure 4 shows that this is neither the case for definite articles nor for articles in general.[12]

---

[12] Moreover, if we look at VO languages which are beginning to use a demonstrative as a definite article (N = 26 in **Dryer2013a**), Hawkins' account would lead us to expect that such incipient grammaticalization is particularly frequent in the constellation DEM-N and ADJ-N (and again less frequent if N precedes both the ADJ and the DEM). Now, of the 26 languages in question,
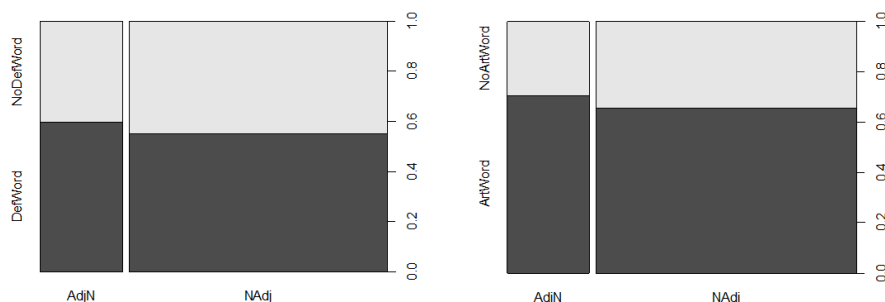
Figure 4: Occurrence of articles in VO languages depending on the position of adjectives (left plot: definiteness words only; right plot: all article-like words; for the corresponding mixed-effects models, cf. SM4)

Clearly, this picture does not speak for a critical processing pressure being at work. And the same conclusion actually carries over to OV languages: It is true that, to the extent that these languages show a reduced propensity for developing articles, they manage to keep NP processing domains slightly shorter; but there are several indications that this pressure cannot be particularly strong.

First, our Family Bias calculations show, in addition to the findings from above, that *none* of the large OV families in the sample actually exhibits a significant bias (towards or against articles) in the first place; they are all internally diverse, i.e. with no more than chance distributions of articles (Table 4).[13]

Second, from a more qualitative perspective, there is suggestive evidence that a potential efficiency motivation in OV languages is easily overridden by other factors. For example, **Ross2001** discusses an interesting case of an intense contact situation in which the Austronesian language Takia adapted its VO syntax to the OV structure of its Papuan contact language Waskia. Ross argues that, in the wake of this restructuring, Takia speakers must have shed the prenominal article word in NPs (cf. (4a)), which would be fully in line with Hawkins' predic-

---

22 are N-DEM and four are DEM-N. It is the latter type that is interesting here, and we find that two of these four languages are ADJ-N and the other two N-ADJ. Again, no clear pattern along Hawkinsian lines can be detected here.

[13]**Bickel2013** suggests that the (minimum) strength of a universal pressure can be calculated on the basis of the proportion of biased families k among all families of a particular kind n (here: OV families): $\hat{s} = (k+1)/(n+2)$. Based on the figures in Table 4, we obtain $\hat{s}_{(OV)} = (0+1)/(11+2) = 0.077$. This estimate is so small in magnitude that one is forced to conclude that there is no particular pressure at all on OV languages with regard to the development of articles.

Table 4: Distribution of biases (for or against) articles among large families in the sample (N_{total} = 29 genetic units)

	VO	OV	Totals
significantly biased	12	0	12
internally diverse	6	11	17
Totals	**18**	**11**	**29**

tion. At the same time, however, the degree of linguistic accommodation was so intense that Takia speakers also did something else: They grammaticalized a postnominal deictic element into a postnominal demonstrative with some article-like functions, reproducing exactly the article pattern in Waskia (cf. (4b-c)).

(4)   a.   Proto-Western Oceanic (**Ross2001**)

       a   tam^{w}ata a-ña

       DET man   that-3sg

       'that man'

   b.   Takia (Austronesian, Oceanic; **Ross2001**: 140)

       Waskia   tamol **an**

       Waskian man   **that**

       'that Waskia man'

   c.   Waskia (Nuclear Trans New Guinea, Madang; **Ross2001**: 140)

       Waskia kadi **mu**

       Waskia man **that**

       'that Waskia man'

In other words, Takia speakers chose precisely the diachronic route that Hawkins would predict to be disfavoured, which goes to show that the alleged processing pressure cannot have been very strong, after all. In this connection, one may also recall that our regression model from above, while bringing out a significant global effect from word order type, did not provide a particularly good fit to the data. The substantial amount of variation in the data that it cannot account for must thus be attributed to other, possibly stronger factors.[14]

---

[14]Diachronic research has actually put forward a number of plausible candidates for such factors. A prominent one since at least **Vennemann1975** is the loss of a case system and the concomitant rigidification of constituent order, which favours the development of articles to express

We conclude, then, that **Hawkins2014** correctly predicts a global difference between OV- and VO-languages in the development of articles. But the present analysis also revealed some challenges for this account. Therefore, it still needs to be established by future research whether the global correlation between word order type and the absence of articles really reflects a *causal* connection between these two phenomena, and whether this could be attributed to efficient information processing. If it turns out that Hawkins is correct, the findings in the present section suggest that we would be dealing with a 'weak universal pressure' in the sense of Seržant (2018 [this volume]) or a "weak cognitive bias" with "significant population-level consequences" (**ThompsonEtAl2016**: 4530).

4   Concluding remarks

The present contribution has taken a closer look at John Hawkins' "processing typology", a research programme that thoroughly subscribes to functional-adaptive motivations for grammatical structure. In the first part of the paper, I summarized where such motivations are possibly operative in diachronic change. In my view, a case can be made for Hawkins' efficiency considerations in the process of actualization, i.e. when a linguistic innovation comes to be extended to a principled, cross-linguistically similar subset of potential application sites (as in differential flagging and indexing, relativization, etc.). In this respect, I consider Hawkins' account as superior to purely source-oriented explanations of grammatical patterns. Of course, this does not deny that persistence accounts are relevant to typological patterns – they clearly are; but it argues against persistence as the sole or perhaps even the dominant explanatory principle for grammatical universals.

A more ambitious but also undoubtedly more problematic move is to link parsing and efficiency to certain innovation processes, such as when a particular grammaticalization channel is predicted to be set in motion only un-der specific structural conditions. In the brief case study presented here, we saw that Hawkins' NP processing hypothesis provides a neat match to the global typo-

---

information-structural distinctions that were previously coded by a more flexible word order (cf. also **Hawkins2004**; **HewsonBubenik2006**; **Fischer2010**; **CarlierLamiroy2014**). Another possible factor is the loss of an aspectual system (cf. **Abraham1997**; **Leiss2000**, 2007). However, especially the former type of explanation is often viewed critically (e.g. **Selig1992** on Romance; McColl **Millar2000** on English; **Leiss2000** on Germanic), and there is currently no proposal as to how various factors may conspire to explain the synchronic distribution of articles (cf. also **Lüdtke1991**). For some further information and preliminary typological analyses of these factors, interested readers are kindly referred to SM5.

logical data, even when these are analyzed in more rigorous and hence more appropriate ways than in **Hawkins2014**. On the other hand, the details of neither the typological picture nor individual diachronic studies produce evidence for a strong pressure on languages to develop into the predicted directions. Therefore, the hypothesis that speakers of OV languages are significantly less inclined than speakers of VO languages to grammaticalize additional NP constructors, remains plausible but currently rather weakly substantiated.

What we would need to see to make it more convincing is a triangulation of (i) typological data that are large enough to take several alternative predictor variables from the literature into account (e.g. case and aspect systems, the presence of other NP constructors such as classifiers), (ii) diachronic data from languages that have undergone (or are in the process of undergoing) changes in basic word order, (iii) behavioural evidence, such as psycholinguistic experimentation with artificial languages (e.g. along the lines of **CulbertsonEtAl2012**; cf. also **Levshina2018** [this volume]). As a matter of fact, a particularly strong aspect of Hawkins' work (especially in **Hawkins2004**; 2014) is that it generally attempts precisely this kind of methodological cross-fertilization; but for the domain at issue here, such an approach has yet to be fleshed out in sufficient detail.

Abbreviations

The paper follows the *Leipzig Glossing Rules*.
Additional abbreviation: T/A = tense/aspect marker

Acknowledgements

Did you like this book?

This book was brought to you for free

Explanation in typology