1   Introduction

In the course of the last decades, a wealth of data has been published on the origin and development of grammatical forms. Part of these data, as they were available up to 2002, were published in the first edition of the *World Lexicon of Grammaticalization*. The main purpose of the present work is, first, to make some revisions to the first edition and, second, to provide an update on some of what has happened in grammaticalization studies since 2002. The update concerns, on the one hand, new, general developments that this field has experienced. To this end, some of the directions that appear to be particularly relevant to the field are pointed out in Section 1.3. On the other hand, the update concerns a wider database, considerably expanded with regard to the distinct language varieties (i.e. dialects included), which are over 1,000, and the processes discussed, which involve 544 grammaticalization paths.

Grammaticalization studies as we know them today are commonly assumed to have a history of hardly more than a century, even if their roots can be traced back to the nineteenth or even the eighteenth century. In this widely held assumption it is ignored that such studies have a much longer history in China, going back at least to the fourteenth century (Chappell and Peyraube 2011: 784). But presumably more than at any time in their history, these studies have experienced a boom in more recent times.

The framework adopted in the present book is concerned with language use across space and time; hence the findings presented may be of help for diachronic reconstruction, especially in areas where other tools available to the historical linguist, such as the comparative method, comparative typology, and internal reconstruction, do not yield appropriate results.

But the book should be of interest not only to the student of historical linguistics. The descriptive linguist will find information, for example, on how and why different grammatical meanings can be related to one another in a principled way. Thus, the data presented should be of help for a better understanding of notions such as polysemy, heterosemy, and homonymy, and why there are some regular correspondences between grammatical forms and the meanings expressed by them, or why certain linguistic forms have simultaneously lexical and grammatical functions. We may illustrate this potential

1

2   Introduction

with the following example. The English form *have* is part of a number of
different constructions.[1] On the one hand, it forms the predicate of a possessive
construction (e.g. *They have a new car*), where it can be said to be a lexical
item, that is, a possessive verb. On the other hand, it is also a grammatical item,
namely an aspect marker in a perfect construction (*They have come*), and it is
also a modal auxiliary (*You have to leave now*).

   As the reader will see in this lexicon, such different morphosyntactic and
semantic uses of one and the same form can be accounted for on the basis of
principles of grammaticalization. Since these principles do not only apply to
English but also to many other languages across the world, they are not only of
help in accounting for structural similarities and differences within a given
language but also across languages.

   While the book addresses mainly the needs of linguists it should also be of
interest to students of other disciplines. Anthropologists, sociologists, and
psychologists may discover that the kind of cognitive processes and other
forms of human behaviour held responsible for the evolution of grammatical
forms is not all that different from the kind of behaviour they observe in their
own fields of study.

   Conceived as a reference work, the book is based on a comparative typolo-
gical perspective, and as such it differs from relevant monographs on gramma-
ticalization (e.g. Lehmann 1982 [1995]; Heine and Reh 1984; Heine, Claudi,
and Hünnemeyer 1991a; Traugott and Heine 1991a; 1991b; Hopper and
Traugott 1993; 2003; Bybee, Perkins, and Pagliuca 1994; Pagliuca 1994;
Heine 1997b; Giacalone Ramat and Hopper 1998; Heine and Kuteva 2007).
Accordingly, an attempt was made to collect many data from as many different
languages as possible and to avoid theoretical and regional biases – as far as this
was possible and feasible. European languages are among the best documented
in the world, both with regard to their history and their present-day structure.
It is therefore hardly surprising that they have figured prominently in linguistic
reference works. Attempts were therefore made in this work to avoid any
Eurocentric bias and it is hoped that we were successful to some extent; for
example, hardly more than 5 per cent of the grammaticalization processes
discussed below are restricted to European languages. These attempts were
supported in particular by two factors. On the one hand, there has more recently
been a growing interest in the crosslinguistic and especially the areal dynamics
of grammaticalization, and this interest surfaced, e.g. in the following impor-
tant academic meetings: *Areal patterns of grammaticalization and cross-
linguistic variation in grammaticalization scenarios*, University of Mainz,

---

[1] The term "construction" is used here in a loose sense for recurring form–meaning pairings
consisting of more than one morphological unit, except when directly referring to the model of
Construction Grammar (see Section 1.3). This usage thus differs from that of Construction
Grammar (e.g. Goldberg 2006: 5) in not including "atomic constructions".

12–14 March 2015, symposium on *Grammaticalization in Japanese and across languages*, Tokyo, 3–5 July 2015, National Institute for Japanese Language and Linguistics, and *Grammatikalisierung in interdisziplinärer Perspektive*, Munich, 6–8 July 2016.

On the other hand, this second edition of the lexicon has also been enriched with linguistic data from the languages of eastern Asia. Languages such as Chinese, Korean, and Japanese have at their disposal historical records some of which are older than those to be found in many European languages. Hence, these languages provide an almost ideal laboratory for the study of grammaticalization, as demonstrated by scholars from these countries, and this is also one of the reasons why this volume has an expanded authorship.

1.1   On Grammaticalization

Grammaticalization is defined as the development from lexical to grammatical forms and from grammatical to even more grammatical forms.[2] In accordance with this definition, grammaticalization concerns the genesis and development of grammatical forms. Its primary goal is to describe how grammatical forms and constructions arise and develop through space and time, and to explain why they are structured the way they are. Technically, grammaticalization involves four main interrelated mechanisms, typically though not necessarily applying in the order as listed below:[3]

(1)   **Mechanisms of grammaticalization**
    (a)   extension (or context generalization) – use in new contexts,
    (b)   desemanticization (or "semantic bleaching") – loss in meaning content,
    (c)   decategorialization – loss in morphosyntactic properties characteristic of lexical or other less grammaticalized forms, and
    (d)   erosion (or "phonetic reduction") – loss in phonetic substance.

While three of these mechanisms involve a loss in properties, there are also gains. In the same way that linguistic items undergoing grammaticalization lose semantic, morphosyntactic, and phonetic substance, they also gain properties characteristic of their uses in new contexts. Grammaticalization requires specific contexts to take place, and it can be, and has been, described as a product of context-induced reinterpretation (Heine 2002). Accordingly, context is a crucial factor in shaping the structure of grammatical forms – to the extent that they may express meanings that cannot immediately be derived from their respective source forms.

[2] The term "grammatical forms", or "grams", roughly corresponds to Ramat's (1999) "feature values", and to what in some traditions are referred to as "functional categories".
[3] For alternative catalogues of parameters, see Lehmann (1982 [1995]) and Hopper (1991).

It has been argued that grammaticalization is not a distinct process, since the four mechanisms listed in (1) can be observed to be at work also in other kinds of linguistic change (Newmeyer 1998: 248ff.).[4] There are a couple of reasons why we think that such a position is not justified. First, the main task of grammaticalization studies is to explain why grammatical forms and constructions are structured the way they are, and these four mechanisms, as opposed to many other conceivable mechanisms, have been found to be relevant to achieve such explanations. Thus, irrespective of how one wishes to define a "distinct process", one is led to conclude that these mechanisms are part of one and the same explanatory framework.

Second, grammaticalization, as conceived here, is above all a semantic process. This process is context dependent, and grammaticalization can therefore be described in terms of context-induced reinterpretation. Not every reinterpretation leads to the rise of grammatical meanings. Rather, it is only when forms for concrete (e.g. lexical) meanings are used to also express more abstract (grammatical) meanings that grammatical forms emerge; for example, when a form used for a visible object (e.g. the body part 'back') is used also to refer to a nonvisible item (the spatial notion 'behind'), or a form used for an action ('go to') is used also to refer to a grammatical concept (future tense). On account of its specific directionality, context-induced reinterpretation has been described in terms of metaphorical transfer, leading, for example, from the domain of concrete objects to that of space, from space to time, from ("real-world") space to discourse space, and so on.

Desemanticization thus results from the use of forms for concrete meanings that are reinterpreted in specific contexts as more abstract, grammatical meanings. Having acquired grammatical meanings, these forms tend to become increasingly divergent from their old uses: they lose categorial properties characteristic of their old uses, hence undergoing decategorialization, and they tend to be used more frequently, to become more predictable in their occurrence, and, consequently, to lose in phonetic substance. Thus, the four mechanisms are not independent of one another. Typically, extension precedes or coincides with desemanticization, which again is followed by decategorialization and frequently also by erosion. There are a few cases where it has not yet been possible to establish that decategorialization really followed desemanticization in time, and we do not wish to exclude the possibility that in such cases the two may have occurred simultaneously. However, such cases appear to be less common: new grammatical meanings arise, and it usually takes quite some time before any corresponding morphological, syntactic, and/or phonetic changes can be observed. In many languages, prepositions unambiguously

---

[4] Newmeyer (1998: 260) refers to desemanticization as "appropriate semantic change", to decategorialization as "downgrading analysis," and to erosion as "phonetic reduction".

serving a grammatical function still have the morphosyntactic structure of their earlier uses as adverbial phrases (e.g. English *by means of, in front of, with respect to*) or verbal phrases (e.g. Chinese *zai* '(to be) at'; Alain Peyraube, p.c.), and tense or aspect auxiliaries may still behave morphosyntactically largely like lexical verbs even if they have lost their lexical semantics and serve exclusively as functional categories (e.g. English *be going to, used to, keep (doing)*, etc.). To conclude, there is evidence to suggest that grammaticalization can be defined as a distinct process.

It is sometimes assumed that grammaticalization invariably involves lexical categories; that is, that it is confined to the development from lexical to grammatical forms. This view tends to ignore the fact that such cases account for only part of what falls under the rubric of grammaticalization. Equally commonly, as we will see in the course of this work, items that are already part of the inventory of grammatical forms give rise to more strongly grammaticalized items. Prepositions often develop into conjunctions, temporal conjunctions tend to give rise to causal or concessive conjunctions, demonstrative determiners develop into definite articles or relative clause markers, verbal perfect inflections may become past tense markers, and so forth – all developments that take place within the domain of functional categories. Such developments are distinguished mainly from developments involving lexical categories by the difficulty of identifying and reconstructing them.

This raises the question of what constitutes a grammatical as opposed to a lexical meaning or category. Many proposals have been made on this issue, perhaps the most attractive one by Boye and Harder (2012) in their theory of discourse prominence. For example, depending on the kind of criteria that one may wish to adopt, demonstratives can be classified as lexical, grammatical, or something else (Diessel 2003; 2006). For Boye and Harder (2012: 20), in particular, demonstratives qualify as lexical, that is, nongrammatical expressions on the basis of tests of addressability and focus. In fact, like lexical expressions, demonstratives can be both addressed and focalized (e.g. *Look at that. – What?*) – that is, they are classified as discursively primary rather than secondary (Boye and Harder 2012: 13).

We have no problems with this conclusion to the extent that it is based on the theory of discourse prominence proposed by these authors. The perspective adopted here is slightly different, however, in that it is based on crosslinguistic principles of grammatical change, suggesting that the rise of new demonstrative markers is compatible with an interpretation in terms of grammaticalization, whereby lexical or less grammatical forms, such as verbs, may develop into (more) grammatical forms such as demonstratives – in accordance with paradigm parameters of grammatical change (Heine and Kuteva 2007: 33–46). As the examples provided in this book and in Heine et al. (2017) suggest,

demonstratives lack semantic, morphosyntactic, and frequently also phonolo-
gical features characterizing the verbs from which they are derived. At the same
time, demonstratives have acquired a schematic function that distinguishes
them from their lexical sources.

This view is in accordance with the general line of change from lexical to
grammatical forms and from grammatical to even more grammatical forms as it
has been defined in earlier work (Kuryłowicz 1965: 69; Heine and Kuteva
2002a: 2; Hopper and Traugott 2003: 18). On this view then, there is reason to
argue that demonstratives are grammatical (or functional) categories, even if
they exhibit some features that distinguish them from other grammatical
categories, as appropriately pointed out by Diessel (1999a; 1999b; 2003;
2006) and others. Within a more general chain of grammaticalization, however,
demonstratives are distinctively less grammatical than categories they may
give rise to, such as definite articles or markers of relative and complement
clause subordination (see the examples provided in Chapter 3).

Grammaticalization is an essentially unidirectional process; that is, it leads
from less grammatical to more grammatical forms and constructions. However,
this process is not without exceptions. A number of examples contradicting the
unidirectionality principle have been found (see, e.g., Joseph and Janda 1988;
Campbell 1991; Ramat 1992; Frajzyngier 1996b; and especially Newmeyer
1998: 260ff.; see also the contributions to *Language Sciences* 23, 2–3;
Campbell and Janda 2001). A few additional examples are pointed out in
this book. These concern modality (see, for example, D-NECESSITY<>
D-POSSIBILITY), on the one hand, and participant marking (for example,
A-POSSESSIVE<> RECIPIENT), on the other.

As acknowledged by most of the scholars concerned, however, such exam-
ples are few compared to the large number of cases that conform to the
principle[5] (see Haspelmath 1999; 2000: 249). Furthermore, such cases can
frequently be accounted for with reference to alternative factors, and finally, no
instances of "complete reversals of grammaticalization" have been discovered
so far (see Newmeyer 1998: 263).

Grammaticalization begins with concrete, lexical forms and constructions
and ideally ends in zero – that is, grammatical forms increasingly lose in
semantic and phonetic content – and they may be replaced by new forms.
Grammaticalization has therefore been described as a cyclical process (Givón

---

[5] See, e.g., Harris and Campbell (1995: 338), who summarize this situation thus: "there is a strong
tendency for grammaticalization to proceed in one direction, though it is not strictly unidirec-
tional". Similarly, Joseph and Janda (1988: 198–200) claim that cases of demorphologization,
a process that would contradict the unidirectionality principle, are rare and not seldom con-
troversial. Finally, Newmeyer (1998: 275–6, 278) observes that cases conforming to the uni-
directionality principle ("downgradings") "have occurred at least ten times as often as
'upgradings'", and he concludes, "I suspect that, for whatever reason, there is a *general* direc-
tionality to the semantic changes observed in grammaticalization" (emphasis in original).

1979a; Heine and Reh 1984); see, e.g., the literature on the Jespersen Cycle
(Dahl 1979) or the Negative Existential Cycle (Dahl 1979; Croft 1991a; van
Gelderen 2008; 2009; Willis, Lucas, and Breitbath 2013a; Veselinova 2014).
While there is some evidence to support this assumption, we have to be aware
that, first, a grammaticalization process can stop at any point of development
and, second, "worn-out" grammatical forms are not necessarily replaced by
new forms. Thus, the metaphor of a grammatical cycle, though useful in certain
cases, should not be generalized since it often does not apply for some reason or
other.

1.2   Problems

Grammaticalization is a complex subject matter; it relates in much the same
way to diachronic and synchronic linguistics as to semantics, syntax, and
morphology, and it is rooted in cognition and pragmatics. Obviously, an
endeavour such as that found here is an ambitious one – one that has to take
care of a wide range of problems. In this section we deal with the most serious
of these problems in turn, to the extent that they relate to the subject matter of
the present book.

The findings presented in this work are meant to highlight processes of
human behaviour that can be observed across cultures; yet these findings are
based on data from a fraction of the world's languages. One may therefore
wonder what justification there may be to call this work a "world lexicon". Our
main reason is this: underlying human behaviour there appears to be a strategy
of linguistic processing whereby more abstract functions are in appropriate
contexts expressed in terms of forms for concrete concepts. We expect, for
example, that in some unknown language there are ways of expressing tem-
poral concepts in terms of spatial ones, spatial relations in terms of forms for
concrete concepts (such as body parts or salient landmarks), aspectual contours
of events in terms of forms for actions and motions, discourse procedural
functions in terms of propositional semantics, or functions concerning the
organization of texts in terms of linguistic forms for spatial or temporal deixis.
Languages differ considerably in the way and the extent to which this strategy
has given rise to grammaticalized constructions; nevertheless, we expect the
effects of the strategy to be essentially the same across languages, including
languages that are still undocumented, or extinct.

Throughout this work we are concerned with the relation between two kinds
of concepts, which we refer to as the "source" and "target" concepts of
grammaticalization. In this account, the impression is conveyed that there is
always a unidirectional development leading from one distinct concept to
another. But this is not only a simplified account; it is also at variance with
much of what we have argued for elsewhere, namely that, rather than being

a development in discrete steps, grammaticalization must be conceived of as a gradual process. To achieve the goal of having a treatment of grammaticalization as a gradual process in the form of a lexicon, we were forced to reduce the process to two salient uses of forms, viz., source and target uses.

We are fully aware that this procedure rests on a gross simplification of the facts. The nature of the process has been described by means of terms such as "cline", "continuum", "chain", "pathway", "scale", etc. (see Hopper and Traugott 2003: 6–7 for a discussion) and, depending on the perspective that one may wish to adopt, each of the terms has its justification. In the present book the perspective preferred is one according to which the process exhibits an interlocking pattern where uses of a grammaticalizing item share features with earlier, less grammaticalized uses of the same item. These uses, which have been described with reference to an "A>A/B>B" scenario (Heine et al. 1991a), exhibit a chain-like structure which cannot easily be divided into more or less separable "points" (see Heine 1992). Accordingly, rather than"cline" or "scale" the term preferred here is that of "chain of grammaticalization".

Furthermore, target uses are not all of the same kind. In one language, a target use can be represented as a fully fledged, conventionalized grammatical category. In another language, by contrast, this use may surface only in the form of a usage pattern of the source concept that is restricted to one particular context. Most of the data presented below are of the former kind, but there are also cases where the information available does not allow us to determine which of the two is involved, or where we decided for specific reasons to also include instances of the latter kind.

Another problem can be illustrated with an example concerning the evolution of aspect and tense categories, where two or more different linguistic forms may simultaneously be involved: an auxiliary (e.g. *be* or *have*), a nonfinite marker (e.g. an infinitival, participial, or gerundival marker), and perhaps also a locative marker. Tense and aspect constructions in a number of languages worldwide do not uncommonly involve three distinct morphological elements, the English future marker *be going to* being a case in point. Another European example is the Latin verb *habēre* 'to have', which in the Romance languages has given rise to perfect markers on the one hand and to future markers on the other. What accounts for this divergent development? The verb *habēre* was not itself grammaticalized; rather grammaticalization involved entire periphrastic constructions, or event schemata: the construction *habēre* + perfect passive participle gave rise to perfect expressions, while *habēre* + infinitive periphrasis was responsible for the development of future constructions. In a lexicon project like the present one, such propositional structures had to be reduced to the salient segments of the constructions concerned, such as the *habēre*-markers figuring in the expression of future tenses in Romance languages; we will return to this issue in Section 1.3.

A related problem that we encountered concerns what one may call "complex grammaticalization": a more complex linguistic structure can assume a grammatical function without involving the grammaticalization of any particular item figuring in this structure; see under Construction Grammar (Section 1.3). What exactly should the lexicon entry be that takes care of this grammaticalization? Take the following example. One widespread way of developing expressions for the grammatical concept of a comparative of inequality is to juxtapose two propositions that are in a polar contrast – one expresses the standard of comparison and the other the comparative notion. This opposition may be either antonymic, as in (2), or marked by the distinction of positive versus negative, as in (3).

(2)   *Cayapo (Ge-Pano-Carib; Stassen 1985: 184)*
      Gan ga   prik, *bubanne ba i pri.*
      You you big   but   I   I small
      'You are bigger than I am.'

(3)   *Abipon (Ge-Pano-Carib; Stassen 1985: 184)*
      Negetink chik naâ, *oagan nihirenak la   naâ.*
      Dog   not   bad   yet   tiger   already bad
      'A tiger is more ferocious (lit.: 'bad') than a dog.'

What is grammaticalized in such constructions is not a specific element but rather some propositional relation, viz., *be big* versus *be small*, or *be bad* versus *not be bad*. In a treatment like this book, which is concerned with segmentable linguistic forms, functions expressed by means of pragmatic or syntactic relations between forms without involving morphological segments of necessity had to be excluded.

Another problem concerns morphosyntactic mechanisms such as derivation and reduplication, which frequently are responsible for changes in the grammatical status of a linguistic form. An English adjective can be turned into an adverb by adding the suffix *-ly*. Similarly, category shift is achieved in quite a number of languages by means of reduplication; a Turkish adjective like *derin* 'deep' turns into an adverb (*derin derin* 'deeply') or a Malay verb (*diam* 'to be silent') into an adverb (*diam-diam* 'silently') when reduplicated (Ramat 2011: 499). While raising interesting questions for the student of grammaticalization, such mechanisms as well will have to be ignored in the present work. On the other hand, our concern will be with the question of how, for example, derivational forms like English *-ly* came to be what they are today. Accordingly, the fact that Modern English *-ly* is the result of a regular grammaticalization process whereby a Proto-Germanic noun *likom 'body, form' developed via Old English *līc* from a noun into an affix is of central interest to the present study.

The sentence in (3) raises yet another question: at which point can we say that grammaticalization has been concluded? Can we really say that (2) and (3) are suggestive of a completed process of grammaticalization, or do they merely represent contextually induced interpretations that are irrelevant for the grammatical structures of the languages concerned? A number of tests have been proposed in grammaticalization studies to deal with this question; frequently, however, the information available on a given language is not sufficient to allow for a successful application of these tests. In such cases we decided to adopt the solution proposed by the author(s) dealing with that language.

In some cases we rely on comparative findings to determine whether a grammaticalization process has been concluded. For example, one of our entries has the form ONE > INDEFINITE, according to which the cardinal numeral for 'one' may grammaticalize to indefinite articles. Now, it has been argued for languages like French (*un(e)*) or German (*ein(e)*), for example, that the two, numeral and indefinite article, are the same, their difference being due to contextual or other factors; that is, that the relevant entry is not an instance of grammaticalization. That the two meanings are in fact different is suggested by comparative observations. Thus, there are languages where a given linguistic item serves as an indefinite marker but not as a numeral, and, conversely, there are many languages where a given item denotes the numeral 'one' but not indefinite reference. We take such observations as evidence that ONE and INDEFINITE are in fact different concepts, even if in some languages the same or a similar word is used for both.

Another problem concerns the directionality of grammaticalization and how to achieve historical reconstruction. How do we know that INDEFINITE is historically derived from ONE rather than the other way around? In this case, there is diachronic evidence to give an answer: in some languages, including a number of European ones, there is a marker that is used for both the numeral 'one' and the indefinite article, and by using historical records it is possible to establish that at some earlier stage in the development of these languages the item only served as the numeral expression before its use was extended to also designate indefinite reference. Now, since grammaticalization is essentially unidirectional, we are led to assume that in languages where no historical records are available the evolution was the same.

Even in the absence of historical documents it is possible to reconstruct directionality of change by using the mechanisms sketched in the preceding section. For example, decategorialization has the effect that the element concerned loses morphosyntactic properties characteristic of its less grammaticalized (e.g. lexical) source, such as the ability to take modifiers or inflections, and it shifts from a form class having many members (e.g. an open class) to one having only few members (a closed class). Erosion again means that that element tends to become shorter and/or phonetically less complex, to lose the ability to receive distinct stress or tone, and so on. Thus, if we find two different