# *Introduction*

This book has two related goals. On the one hand, we wish to address the question of syntactic change in the context of the minimalist programme, by using (variants of ) some of the technical devices that have been proposed in order to provide a general analysis of a pervasive diachronic phenomenon, grammaticalization. On the other hand, we wish to address a deeper question raised by the nature of the minimalist programme itself. A central idea behind the minimalist programme is the idea that language is in some sense a perfect system (the strong minimalist thesis: see Chomsky (1995:1–10), (2000:96f.), (2001:1–2)). Now, perfect systems do not vary over time, so the very existence of syntactic change appears to be a challenge to this thesis. The existence of synchronic variation among grammatical systems also poses an apparent problem for the strong minimalist thesis. The account of grammaticalization that we develop will lead to what we believe to be an interesting response to this problem, and an explanation for the existence of apparent variation and change in syntactic systems which we believe to be consistent with the strong minimalist thesis.

The term *grammaticalization* was first introduced by Meillet (1912) to describe the development of new grammatical (functional) material out of 'autonomous' words. Since then the topic has received much attention in the literature on language change, especially amongst typologists (see the references and citations in Janda (2001), and the impressive compendium of examples of the phenomenon in Heine & Kuteva (2002)). As Hopper and Traugott (1993:1–2) point out, the term 'grammaticalisation' can be used to describe either the framework that considers 'how new grammatical forms and constructions arise' or 'the processes whereby items become more grammatical through time'. The primary empirical goal of this book is to provide a general characterization of the diachronic phenomenon of grammaticalization within a modified version of Chomsky's (1995, 2000) minimalist framework, combined with an approach to language change of the kind argued for by Lightfoot (1979, 1991, 1998), Clark and Roberts (1993) and Roberts (2001). We do this by

1

## 2    *Syntactic Change*

developing the basic idea that grammaticalization involves the creation of new
functional material, either through the reanalysis of existing functional material
or through the reanalysis of lexical material. Within the set of assumptions we
adopt, there is no need to treat grammaticalization as a separate framework, or
for that matter as a distinct process of the grammar. Thus our central claims are:
(a) that grammaticalization is a regular case of parameter change not funda-
mentally different from other such changes; (b) grammaticalization is therefore
epiphenomenal, as recently argued by Newmeyer (1998), Joseph (2001a). This
brings us to our main theoretical goal, which is to provide an understanding of
the nature of functional categories, using grammaticalization as our tool, since
it creates new functional material. We thus aim at providing a way of identify-
ing the inventory of potential functional categories in the grammar of natural
language.

We have just mentioned what we see as the basic nature of grammaticaliza-
tion. Let us expand on this idea a little more; in Chapters 2 to 4 we discuss
numerous cases, analysing them all in broadly the same way. In all these cases,
we argue that grammaticalization is the creation of new functional material. It
must, then, involve some sort of categorial reanalysis of lexical or functional
material. The main question is how this kind of change can be captured formally.
In what follows, we develop an account in terms of the central idea that such
a change always involves structural simplification. The idea can be illustrated
in simplified form with the development of the future expression in *tha* + VP,
where *tha* is the future marker, from *thélo* + subordinate clause, where *thélo* is
a verb, in the history of Greek (this change was discussed in Meillet (1912); and
is analysed in detail in 2.3). It is clear that the Post-Classical Greek construction
consisted of a biclausal structure, in that *thélo* heads a VP associated with the
full panoply of functional material, including a subordinate clause introduced
by the complementizer *hína* (cf. Joseph 1983, Horrocks 1997, Pappas & Joseph
2001). On the other hand, the Modern Greek construction with *tha* is standardly
analysed as monoclausal, since both the verb and *tha* occur in the same clause
(cf. Philippaki-Warburton 1992, Rivero 1994 among others). At the very least,
then, the grammaticalization of *thélo* to *tha* involves the associated structural
simplification in (1) (here CP is the clause, or Complementizer Phrase, see 1.2):

(1)        $[_{CP} \ldots [_{VP} \text{thélo CP}]] > [_{CP} [ \ldots \text{tha VP}]]$

This change is a structural simplification at least in that the earlier structure
contained two CP nodes, while the later one contains just a single CP. In Chapter
5 we will develop an account of language change which derives the fact that
structural simplification is a natural mechanism of change, and therefore the fact
that grammaticalization is a widespread and natural kind of change. Our general

characterization of grammaticalization, then, is that it is categorial reanalysis which creates new functional material, and that this reanalysis always involves structural simplification.

This kind of change is extremely prevalent across languages. Auxiliaries typically derive from 'full' verbs (the English modals are a much discussed and typical example, see 2.1 below). Determiners and pronouns often derive from demonstratives (all the Romance determiners and 3rd-person pronouns are reflexes of one of the two Latin demonstratives *ille* and *ipse*, see 4.1). Complementizers derive from demonstratives (English *that*), relative pronouns (Romance *que/che*), prepositions (English *for*), verbs of saying (e.g. Ewe *bé*, see Lord 1976), etc. (see Chapter 3). Heine and Kuteva (2002) provide an inventory of literally hundreds of cases of this kind of change.

In the typological literature where these and other changes have been documented, and the evidence for diachronic pathways has been accumulated (see, among others, Heine, Claudi & Hünnemeyer 1991, Heine & Reh 1994, Hopper & Traugott 1993, Lehmann 1985, Heine & Kuteva 2002), grammaticalization has been claimed to be universal. Indeed, Meillet (1912) pointed out that the process was one of only two ways in which new grammatical material could arise, the other being *analogy* (new paradigms developing by formal resemblance to existing paradigms). Our account of grammaticalization must thus explain this ubiquity. To this end we introduce the notion of markedness into the theory of parameters in Chapter 5.

The claim that grammaticalization follows a pathway of language change, that is, expresses a tendency, poses a challenge to the standard Principles and Parameters approach to syntactic change, whose exact nature we describe in detail directly. We thus see that in pursuing our analysis of grammaticalization, a number of theoretical issues that are of importance for both synchronic and diachronic syntax must be addressed. These questions all revolve around the familiar tension between empirical coverage and explanatory depth, which has been highlighted by Chomsky repeatedly since the 1960s. In a sense, then, the deeper purpose of the book is an attempt to deal with this tension in the diachronic domain, paying particular attention to the descriptive and explanatory potential of the notion of functional categories.

In the diachronic domain, the principal theoretical question that arises is whether language change is a deterministic process (as language acquisition is for example). In other words, the issue is whether we can identify clear pathways of change that make languages converge onto certain parametric settings. In the context of the theory of principles and parameters (whether in its minimalist manifestation or otherwise), we can view parameters as creating a space of possible variation within which grammatical systems are distributed.
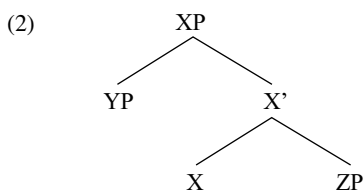
4    *Syntactic Change*

The natural view of synchronic variation among grammatical systems is to think of them as randomly scattered in this space; and the natural view of diachronic change is to see it as a random 'walk' around this space. The prediction is then that there are no pathways or 'drift' in diachronic change, a view which has been consistently and vigorously defended by Lightfoot (see in particular Lightfoot (1998) for a recent restatement). This view is theoretically cogent, and to the extent that seeing syntactic change as parametric change is the key to an explanatory account of syntactic change (as we believe, with Lightfoot, that it is), then we are led to the view that there cannot be pathways, drift or natural tendencies in syntactic change. However, the phenomena of grammaticalization provide ample evidence of just such pathways or tendencies, and thus syntactic change must – at the very least at the descriptive level – be in a significant sense non-random.

At this point then we seem to reach a paradox: on the one hand we accept that parameter changes cannot follow pathways. On the other hand, by acknowledging the prevalence of grammaticalization we appear to also accept that syntactic change follows some pathways. One way of resolving this paradox is to deny that whatever we have identified as grammaticalization exists. This is a logically viable option, but at the same time it goes against the empirical evidence. What this apparent paradox boils down to is a deeper theoretical question, namely that of reconciling the clear evidence for pathways of change at the descriptive level with the fact that an explanatory account of change must involve parameter change. We will propose in Chapter 5 that this can be done in terms of a theory of markedness which has the effect of creating 'basins of attraction' (in the sense of complexity theory) within the parameter space. In this way, we retain the explanatory force of the notion of parametric change, but make it compatible with the evidence that changes tend naturally to go in certain directions and not others – grammaticalization being a prime example. More generally, we see this as a way of reconciling the tension between descriptive and explanatory adequacy in diachronic syntax (and possibly in language typology – see Roberts (2001) on this).

In the synchronic domain, our account raises two theoretical issues. The first is very straightforward and central to current concerns: what is the characterization of a possible functional category? To put it another way, what is the inventory of functional categories that Universal Grammar (UG) makes available? Despite the centrality of this question, there are few proposals concerning this matter in the literature, to our knowledge. Chomsky (1995, 4.10) imposes a constraint on possible functional categories in the context of certain technical assumptions (see 1.3 and 4.5 for more detailed discussion of this point). In particular, Chomsky denies the presence of functional categories whose sole

role is to create place holders for moved material. Although we endorse this idea, we will argue against Chomsky's postulation of non-interpretable features (see 1.3), offering a more restrictive alternative on conceptual grounds. The most detailed empirical discussion of functional categories to date is undoubtedly Cinque (1999), but Cinque proposes a purely extensional (and preliminary) characterization of the set of clausal functional categories (see 1.2 (18) for a list). No intensional characterization is offered. If grammaticalization indeed involves the creation of new functional material through natural processes of change, then we can see this process as a kind of natural laboratory for the investigation of the question of what makes functional heads what they are. In Chapter 5, we will tentatively suggest a semantic characterization of the class of possible functional categories based on von Fintel (1995), which is empirically supported by our investigations of grammaticalization. This characterization has the merit of making the existence of functional categories appear closer to the general minimalist desideratum of (virtual) conceptual necessity (see the Introduction to Chomsky (1995) for a discussion of this notion).

It is clear at this point that the central theoretical construct that we will make use of, and indeed try to develop, in this book is the notion of functional categories. The basic idea behind the postulation of functional categories in recent principles and parameters/minimalist syntax is to ascribe full configurational status to closed-class, grammatical categories. In section 2 of Chapter 1 we will try to provide a justification of this idea. Here we want to simply say a few words about what functional categories are. The standard functional categories are I (Inflection, roughly corresponding to the Aux node of earlier generative grammar), C (Complementizer) and D (Determiner). Each contains closed-class grammatical morphemes, purely morphological material (bound morphemes or morphophonological features) or perhaps no overtly realised material at all, depending on the language. The central assumption made in recent principles and parameters/minimalist research is that each functional category has a full phrase-structural status. One way to understand this is to postulate that functional categories conform to the basic schema of X'-theory, which, following essentially Kayne (1994), we take to be as in (2):

(2)

```
              XP
            /    \
          YP      X'
                 /  \
                X    ZP
```

6    *Syntactic Change*

YP is the specifier of X (SpecX) and ZP is the complement of X. We will illustrate how the main functional categories fit into the schema in (2) in 1.2.

Alternatively, we can adopt Chomsky's (1995) bare phrase-structure theory. This amounts to assuming for functional categories that grammatical features are able to merge autonomously with lexical items. This may or may not give rise to a canonical X'-schema like (2), depending on the exact nature of the bundle of features which constitutes the functional category. The essential idea in the context of bare phrase-structure theory, however, is that grammatical features are entered in the lexicon in the same way as 'full' lexical items.

The fundamental advantage of relating grammatical and morphological categories directly to phrase structure in this way is that it makes it possible to reduce two easily observed differences among languages to one theoretical construct. The two differences are: (i) languages differ in their inflectional morphology, and (ii) languages differ in which word orders are grammatical. These differences can be encoded as properties of functional heads. Difference (i) is directly encoded by the assumption that functional heads exist – so we say that languages can differ in that in one language a given functional head F has an overt exponent, while in the next language F has no exponent. Ultimately, this reduces to the fact that different languages have different lexica, in that sound-meaning pairs vary arbitrarily: the most fundamental and inescapable dimension of cross-linguistic variation. Difference (ii) is slightly more subtly handled: we assume, following Chomsky (1995, 2000), that functional categories are those which 'attract' other material, that is, they are the categories which act as triggers for movement. If we assume a universal base, following Kayne (1994), then the only way in which word-order differences can emerge is from differential triggering properties of functional categories.

The preceding paragraph actually outlines our view of parametric variation. We take it that functional categories are present as features in the lexicon. Each such feature may be associated to a PF representation, that is, a lexical item (difference (i)), or may be marked by a diacritic as a trigger for movement (difference (ii)). Note that both properties are entirely arbitrary. For concreteness, we assume that there is a universal set of such features, and that languages do not 'select' among them. All functional features are present in all languages (although they may not all be realized, *pace* Giorgi and Pianesi 1997) – see Cinque (1999:133). There are a number of technical points underlying these assumptions, which we will deal with as appropriate (see Chapter 5). Language acquisition consists in correctly assigning lexical items or the triggering property to the functional features: the features themselves, an X'-schema of the type in (2) (or whatever abstract principles determine the nature of phrase structure),

and the nature of the movement operation are innately given as aspects of UG.
Language change consists of some change in the realization/attraction property
of functional heads, that is, a change in the lexicon. In this respect, we agree
with Longobardi (2001a) that no purely syntactic change is possible. Since our
focus is on grammaticalization, we are naturally more concerned with the pa-
rameters underlying difference (i). However, we will see that difference (ii) is
intimately implicated in such changes.

The second theoretical goal of this book is directly related to the diachronic
issue discussed above. As mentioned there, the natural supposition that arises
from the characterization of grammatical variation in terms of parameters is
that grammatical systems will be randomly distributed in the space of variation
defined by the parameters. However, even on fairly conservative assumptions
regarding the number of parameters, this space must be large enough to admit, in
principle, millions, if not billions, of systems. If so, then a random distribution of
grammars in such a large space should not give rise to any discernible groupings
of systems. But the results of four decades of work in language typology clearly
show that grammatical systems group together around certain constellations of
variant properties; in other words, certain parameter values tend to covary (this
point is developed in more detail in Roberts (2001)). This uneven distribution
of grammatical systems in the parametric space can be explained in terms of
markedness: the 'basins of attraction' in the space will, over time, pull grammars
into them. Over millennia, then, the parametric space will take on exactly the
kind of uneven distribution we observe.[1] Our account of the prevalence of
grammaticalization sketched above and presented in detail in Chapter 5 also
accounts for the existence of discernible language types.

The pioneering empirical work on grammaticalization has largely been done
by typologists, many of whom work in a 'functional' framework of one kind or
another (see Newmeyer (1998, Chapter 5) for a discussion of the alleged links
between grammaticalization and functional approaches to syntax). We hope
that linguists working in these frameworks will find our proposals interesting
and stimulating. We recognise that the formal approach that we adopt is not
necessarily compatible with other frameworks; however, one of the implicit

---

1. The explanation being offered for the uneven distribution of grammars is a historical one, but
not in the conventional sense. A conventional historical explanation would rely on genetic
relatedness to explain typological similarities. It is clear that such an explanation cannot work.
On the one hand, it is easy to observe that the languages of the Indo-European family show a range
of typological variation; just looking at basic word order, we find SOV (Indic), SVO (Romance)
and VSO (Celtic) languages. On the other hand, there is evidence for deep similarities among
historically unrelated (or at best very distantly related) languages, the shared VSO typology,
which appears to go well beyond the basic word order, of Celtic and Semitic is a case in point.

8   *Syntactic Change*

themes of this book is that the right kind of formal approach to grammatical-
ization can be quite revealing. We offer our ideas as an attempt to shed light
on a central and intriguing property of language, which is clearly of common
interest, from a novel perspective. We hope that this book will also be relevant
to those researchers who may not be interested in issues of syntactic change, but
are interested in theoretical questions such as the notion of functional categories
and the nature of parametric variation.

# 1 *Parameters, functional heads and language change*

## 1.1 *Introduction: the logical problem of language change*

In the Principles and Parameters framework cross-linguistic variation is accounted for by means of assigning different values to a finite set of options, called parameters, that are provided by Universal Grammar (UG). In Chomsky (1981, 1986a) parametric options are associated with the principles of UG. To take an example, consider the Extended Projection Principle (EPP), which basically requires that all clauses have a subject. A parameter then determines whether this subject, when pronominal, is always overtly realised (in finite contexts at least). It is in English; it does not have to be in Italian. This is the 'pro-drop' – or null-subject – parameter; its effects are illustrated with the Italian and English examples in (1a) and (1b) respectively:

(1)    a. Parla italiano.
            'He/she speaks Italian.'
       b. *Speaks Italian

In this model, the task of language acquirers is to set the right parametric values on the basis of the input they are exposed to. Thus UG along with the appropriate trigger experience yields a particular grammar. The task of the linguist, on the other hand, is first to identify the UG principles, and second to define the class of associated parameters. It is clear that the simplest possibility is that parameters are restricted to just two values; this desideratum has been largely followed in the literature.

   Although this approach to parameterization seems to work for cases like the 'pro-drop' parameter in (1), it turns out to be insufficient once a wider range of parameters is taken into account. Consider, for example, Binding Theory, and in particular Binding Principle A, which states that an anaphor must be bound in its local domain. As Wexler and Manzini (1987) show, the notion of the local domain can be defined as the category that contains the anaphor and one of the following: (i) a subject, (ii) Inflection, (iii) Tense, (iv), indicative Tense, or finally (v) a root Tense. In other words, Binding Principle A is

9

subject to a five-valued parameter. Moreover, it is possible to find languages
that make use of more than one value, depending on the type of anaphors
they possess. Dutch is an example, as it has two types of reflexives, namely
*zich* and *zichzelf*, which have distinct distributional properties. In particular,
*zich* accepts a long-distance antecedent, while *zichzelf* behaves more like the
English reflexive *himself/herself*, thus requiring a local antecedent (parame-
ter (a) in the Wexler and Manzini (1987) system). This is illustrated in (2a)
and (2b) respectively (cf. Koster and Reuland 1991 for an overview of the
data):

(2)      a. $Max_i$ bewondert $zichzelf_i$/*$zich_i$.
            '$Max_i$ admires $himself_i$.'
         b. $Jan_i$ liet   mij voor $zich_i$/$zichzelf_i$ werken.
            John made me for   him          work
            '$John_i$ made me work for $him_i$/*$himself_i$.'

Wexler and Manzini (1987) concluded that parameters must be associated with
lexical items, offering further support for Borer's (1984) original claim. Regard-
ing (2) then, the choice of the antecedent is a lexical property of the elements
*zich* and *zichzelf*, and as Pica (1987) showed, it correlates with the internal
structure of the reflexives. Attributing the parameter to the lexical properties of
the anaphors allows us to maintain Binding Principle A as a non-parameterized
principle, which states that anaphors must be locally bound. Parametric varia-
tion with respect to what counts as local is associated with the relevant lexical
items.

The idea that parameterization is restricted to the lexicon has been suc-
cessfully pursued in subsequent research, which has further limited the set
of parameterized lexical items to functional categories (see Chomsky (1995,
2000) for a recent statement). Language acquisition is still seen as the process
of parameter setting, albeit as specifically fixing the values associated with
functional categories. It is uncontroversial that the lexicon has to be learned,
and, on this view, parameter setting reduces to a facet of lexical learning. We
can now view the initial state of UG as consisting of a number of principles
and of open parametric options; the latter are associated with a specific set of
lexical items, the functional categories. To illustrate this, let us reconsider the
'pro-drop' parameter: the EPP is not parameterized, but the inflectional cate-
gory responsible for subject agreement, call it AgrS, is. In particular, if AgrS is
in some sense rich enough, that is, has the right properties, to license and iden-
tify an empty pronominal subject, we have the Italian setting, yielding (1a);
if not, then we have the English setting, predicting the ungrammaticality of