

## **Language Change**

Eric Fuß

### POSTPRINT

#### **Abstract and Keywords**

The present chapter outlines a research program for historical linguistics based on the idea that the object of the formal study of language change should be defined as grammar change, that is, a set of discrete differences between the target grammar and the grammar acquired by the learner (Hale 2007). This approach is shown to offer new answers to some classical problems of historical linguistics (Weinreich et al. 1968), concerning, specifically, the actuation of changes and the observation that the transition from one historical state to another proceeds gradually. It is argued that learners are highly sensitive to small fluctuations in the linguistic input they receive, making change inevitable, while the impression of gradualness is linked to independent factors (diffusion in a speech community, and grammar competition). Special attention is paid to grammaticalization phenomena, which offer insights into the nature of functional categories, the building blocks of clause structure.

Keywords: actuation problem, transmission problem, language acquisition, logical problem of language change, grammaticalization, grammar competition, linguistic variation, functional categories, parameters, economy principles

## 18.1 Introduction

IN the Principles and Parameters framework, (syntactic) variation between grammatical systems is attributed to different settings for a limited number of parameters that are associated with invariable principles of Universal Grammar (UG) (Chomsky 1981a, 1986b, 1995b; and chapter 14). The latter identify the set of possible human languages and thus define the upper limits of linguistic diversity. The task of acquiring a given grammar then consists of filling in the gaps left open by the principles of UG, that is, detecting the parameter settings which are reflected by the linguistic input the child is confronted with (see chapters 11 and 12). The stimulus that serves to set a given parameter one way or other is usually called *trigger experience* or *cue* (see Lightfoot 1999 and note 9 on the latter notion). Apart from accounting for *synchronic* differences between a set of individual languages, this approach can also be used to describe *diachronic* differences between historical stages of a single language in terms of parameter values that vary over time (Lightfoot 1991, 1999; Kroch 2001; Roberts 2007). From this perspective, language change is constrained by the requirement that its outcome must meet the specifications imposed by UG. However, it seems that the pervasiveness of change cannot be directly attributed to properties of the human language faculty, in contrast to other universal properties of language. Today, most scholars agree that the possibility (and omnipresence) of change is to be attributed to another ‘external’ universal property of language, namely the fact that language transmission is necessarily discontinuous (see Paul 1880 and Meillet 1904/1905; for relevant statements in early generative work cf. Halle 1962; Klima 1964, 1965; Kiparsky 1965, 1968; see also Lightfoot 1979, Janda and Joseph 2003). During the process of first language (L1) acquisition, children do not have direct access to (abstract) properties of the target grammar; rather, they construct a ‘new’ grammar based on the linguistic input they receive. From this perspective, language change takes place when linguistic features fail to be transmitted in the course of language acquisition. However, at least in a monolingual, homogeneous speaker community, this (p. 460) explanation seems to lead to a paradox: how can a target grammar  $G_1$  produce an output that differs in significant ways from the input that led to the acquisition of  $G_1$  in the previous generation? This is sometimes called the ‘logical problem of language change’ (see Clark and Roberts 1993; Niyogi and Berwick 1998; Roberts 2007:230–231; see section 18.3 for discussion); it is also at the heart of what Weinreich, Labov, and Herzog (1968:102) call the ‘actuation problem,’ arguably the central explanandum in historical linguistics:

### *The actuation problem*

- (1) ‘What factors can account for the actuation of changes? Why do changes in a structural feature take place in a particular language at a given time, but not in other languages with the same feature, or in the same language at other times?’

An answer to the actuation problem presupposes a thorough grasp of the relationship between the linguistic experience presented to the child and the grammar constructed on the basis of this evidence. Unfortunately, our understanding of these matters is still quite limited, despite recent advances in the formal (and quantitative) study of L1 acquisition (see Guasti 2002, Clark 2009, and chapters 11 and 12 for overviews; Niyogi and Berwick 1997, 1998, Yang 2002, and Niyogi 2006 for formal/computational approaches to the relationship between language acquisition and change). It is generally assumed that under normal circumstances, L1 acquisition results in an accurate reproduction of the target grammar. However, the very fact of change shows that this assumption is in need of some qualification. In this context, cases of language change constitute a natural laboratory for the investigation of the relative weight of nature (the genetic endowment, i.e., UG) and nurture (the linguistic experience/environment). Diachronic variation can be taken to reflect the 'limits to attainable grammars' (Lightfoot 1991:172), in the sense that change reveals how certain linguistic choices which give rise to individual grammars are selected on the basis of the evidence available to the learner (Lightfoot 1979, 1991, 1999; Kroch 2001; Hale 2003, 2007; Roberts and Roussou 2003; Roberts 2007). In addition, the study of change might provide us with a better understanding of the role of general cognitive principles not specific to language (sometimes called 'third factors'), in particular '(a) principles of data analysis that might be used in language acquisition and other domains; (b) principles of structural architecture and developmental constraints ... including principles of efficient computation' (Chomsky 2005:6; see chapter 6). This chapter explores a couple of selected issues that arise when language change is studied from a UG-based perspective, focusing on the following two questions:

1. To what extent can a UG-based perspective contribute to our understanding of the phenomenon of language change?
2. What can the study of language change contribute to our understanding of properties of the human language faculty/UG?

The chapter is organized as follows. Section 18.2 is concerned with a set of general issues that arise when language change is studied from a UG-based perspective. In particular, (p. 461) it is argued that the proper object of a formal study of language change should be identified as *grammar change*, that is, a set of discrete differences between the target grammar and the grammar acquired by the learner (see Hale 2007). It also gives an outline of the generative approach to (syntactic) change in terms of a change in parameter values. Section 18.3 focuses on acquisition-based answers to the actuation problem and addresses the question of how the discontinuous nature of language transmission can be reconciled with the traditional observation of long-term, directional changes (drift, Sapir 1921). A related issue is dealt with in section 18.4, which discusses the fact that the transition from one historical language state to another typically involves linguistic variation and diachronic gradualness. Section 18.5 argues that the study of grammaticalization phenomena may offer insights into properties of UG and the theory of functional categories, in particular. Section 18.6 provides a concluding summary.

## 18.2 Basic Notions

### 18.2.1 'Language Change' vs. 'Grammar Change'

Generative linguists usually agree that the proper object of the scientific study of language is I-language (or simply 'grammar'), that is, a knowledge state in the mind of an individual (see Chomsky 1986b). This is usually contrasted with E-language, which is sometimes defined as the set of actual (or potential) expressions that are in use in a linguistic community. Thus, linguistic evidence consists of E-language data; it is the goal of the linguist (and the language learner) to detect properties of I-language by inspecting the linguistic behavior of individuals. Any principle or rule of grammar that is posited by the linguist is to be seen as a piece of I-language. *Universal Grammar* (UG) is then construed as a theory of formal universals of human language that identifies the set of possible I-languages/grammars (see Chomsky 1986b:23). Adopting this view for the study of language change implies that the proper object of formal historical linguistics must be I-language(s)—or grammar(s)—as well. In the following I will briefly review some consequences of this position (see Lightfoot 1999, Hale 2007, Roberts 2007 for in-depth discussion).

First of all, an I-language perspective on language change highlights the fact that language transmission between generations of speakers is necessarily discontinuous: 'a language is not some gradually and imperceptibly changing object which smoothly floats through space and time' (Kiparsky 1968:175; see also Lightfoot 1979:148). Rather, the grammar of a language—I-language—is always created anew in the mind of an individual child when he/she engages in the task of language acquisition. From this perspective, change takes place when the grammar acquired by the learner differs in some property from the target grammar(s) that generated the input the learner was exposed to. The proper scientific object of a *linguistic* theory of change should therefore be defined in terms of a set of discrete differences between I-languages, that is, between the target grammar and the grammar eventually acquired by the learner. Note that this (p. 462) perspective is at odds with views held in the traditional literature on language change where language change is often used to refer to a (sociolinguistic) process in which a given change spreads through a speech community. To avoid confusion with traditional uses of the term *language change*, Hale (1998, 2007) and Lightfoot (1999) introduce the notion of *grammar change* to refer to the proper scientific object of (generative) historical linguistics (for discussion see also Janda and Joseph 2003; Roberts 2007). Thus, we must differentiate between the following aspects of any given change:

- a. *Innovation* (grammar change, abrupt)
- (2) b. *Diffusion* (a grammar change gradually gaining a wider distribution in a<sup>1</sup> speech community, often perceived as 'language change')

As pointed out by Hale (2007:29), the fact that ‘some of the most hotly debated issues in traditional historical linguistics (is change “gradual” or “abrupt”?, is phonological change “regular”?, etc.) have arisen and remained obscure’ can ultimately be attributed to the failure to distinguish properly between these two aspects of diachronic variation. It is likely that the confusion of different notions of ‘change’ arises at least partially from the fact that what we perceive as ‘language change’ is normally the result of diffusion. In fact, there are presumably myriads of changes that never showed up in the records since they were confined to a single speaker and never spread to other speakers, let alone to the whole community of speakers. Still, it seems that we must focus on the first kind of change, that is, innovations, if we aim at developing a restrictive *linguistic* theory of language change.<sup>2</sup> This point can be illustrated with the following example adopted from Hale (2007:39):

- (3) a. Middle English *lutter* ‘pure’  
b. Modern English *pure* ‘pure’

(p. 463) Of course, the (lexical) change from *lutter* to *pure* cannot be explained in terms of a restrictive theory of possible sound changes (e.g., /l/ → /p/ is a very unlikely type of sound change). Rather, the change from (3a) to (3b) is an example of borrowing due to language contact with French. As a result, the original English word meaning ‘pure’ was replaced by the loanword *pure*. As has repeatedly been pointed out in the literature, there are presumably no linguistic constraints on borrowing (see, e.g., Thomason and Kaufman 1988, Harris and Campbell 1995, Curnow 2001; but also see chapter 16, section 16.8).<sup>3</sup> Furthermore, note that borrowing represents an instance of diffusion (the input contained both *pure* and *lutter*, and over time, more and more learners acquired *pure* instead of *lutter* as the realization of the concept ‘pure’). Thus, while we can formulate a constrained theory of possible sound changes (ruling out a sequence of grammar changes leading from (3a) to (3b)), it seems quite unlikely that we can develop a restrictive linguistic theory of possible diffusion/borrowing events. Any change can diffuse, and there are presumably no strong linguistic constraints at work here.

From an I-language perspective on change, the relationship between language acquisition and change can be represented as follows (see Hale 2007:28):(4)

(4)

*A model of language acquisition and change*

$S_0$  (= the initial state of the learner; UG)



$S_1$



$S_2$



$S_3$



.

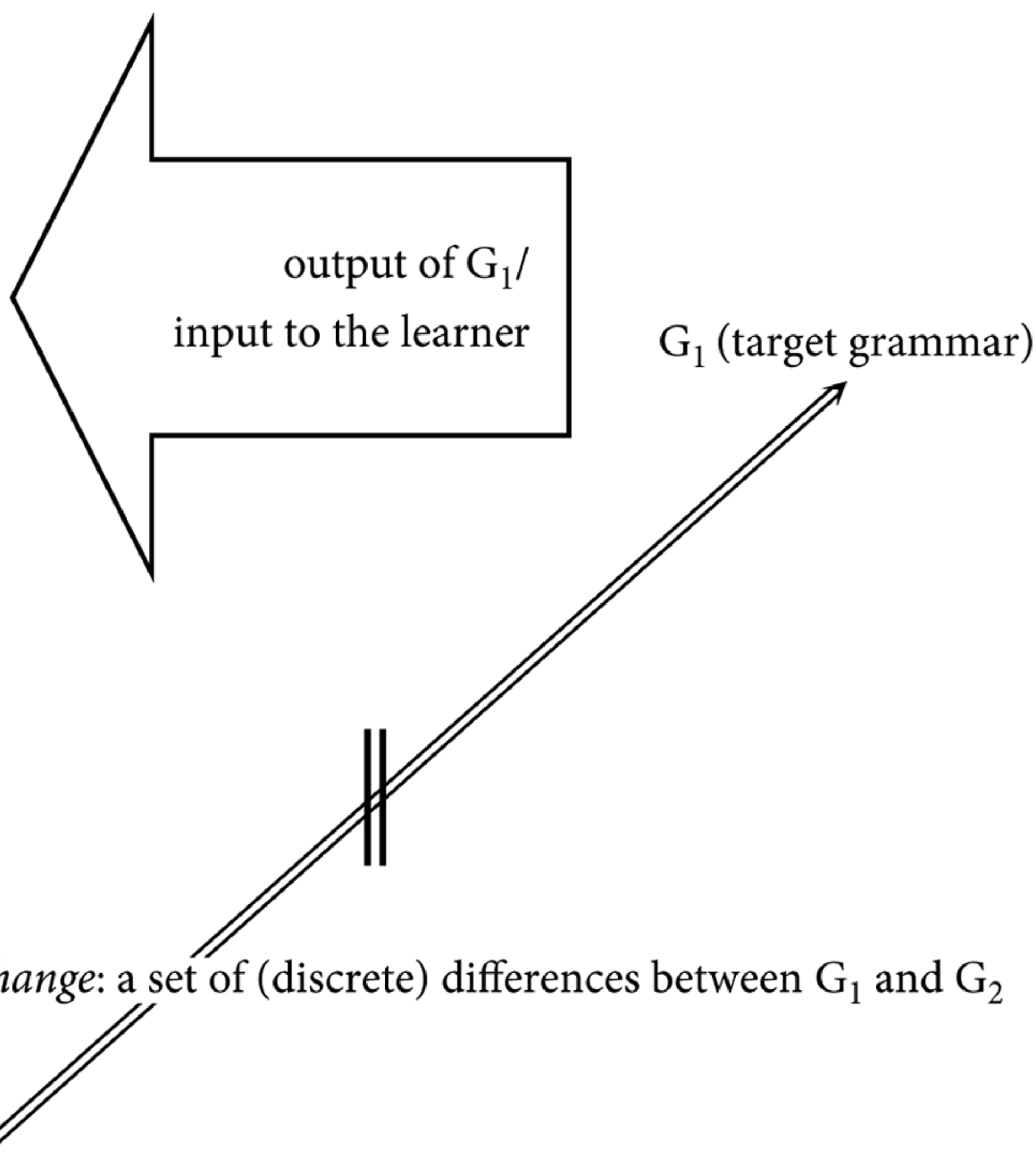
.

.

*Change: a set of (discrete) differences between  $G_1$  and  $G_2$*



$G_2$  (fixed knowledge state/grammar eventually acquired by the learner)



(p. 464) Starting out from  $S_0$  (the initial state of grammar, usually taken to be an expression of the genes, which can be modeled in terms of a system of abstract principles, UG), the learner constructs a number of intermediate knowledge stages during the acquisition process based on the evidence provided by the input (where  $S_n$  is revised to  $S_{n+1}$  if the learner becomes aware of the relevant evidence necessary to trigger a certain property of the grammar).<sup>4</sup> Eventually, the process of grammar construction leads to a fixed knowledge state which represents the grammar acquired by the learner in the course of language acquisition.<sup>5</sup>

### 18.2.2 Grammar Change as Parametric Change

The model represented in (4) highlights that changes resulting from misacquisition involve a set of discrete differences between the target grammar  $G_1$  and the acquirer's grammar  $G_2$ . As already noted in the introduction, these differences (i.e., grammar change) can be modeled in terms of shifting values of individual parameters.<sup>6</sup> However, the traditional notion that linguistic variation reflects parameterized principles of UG (e.g., in terms of two or more options with regard to some property) does not seem to sit comfortably with current minimalist theorizing (cf., e.g., Boeckx 2011a; see Roberts and Holmberg 2010, Holmberg 2010a, and chapter 14 for discussion).

First, the assumption of a richly structured UG with a fairly large number of principles (and associated parameters) does not seem to be compatible with evolutionary considerations: it is commonly assumed that language arose in humans rather recently, that is, during the last 100,000–200,000 years, which is a blink of an eye in evolutionary terms (see, e.g., Corballis 2003). On the assumption that complex systems can only develop via time-consuming natural selection, the rapid development of the human

(p. 465) language faculty (FL) seems to preclude the possibility that FL has a rich internal structure. Rather, it is more likely that language came into existence as a result of a small number of evolutionary innovations, with other properties of language determined by general cognitive structure and design principles (see, e.g., Hauser, Chomsky, and Fitch 2002; Chomsky 2005; Hornstein 2009).

Second, the traditional notion of parameter is at odds with the minimalist assumption that the computational system ( $C_{HL}$ ), that is, the mechanisms that build up hierarchical structures by combining individual lexical items, is universal and therefore cross-linguistically and diachronically invariable (Chomsky 1995b, 2000b, 2005). Thus, strictly speaking, there is no such thing as 'syntactic change': the properties of the syntactic component of grammar remain constant over time (see Hale 1998, 2007; Longobardi 2001a; Keenan 2002). On this assumption, linguistic variation is limited to the lexicon (the *Lexical Parametrization Hypothesis*, going back to Borer 1984, Manzini and Wexler 1987; for discussion see, e.g., Ouhalla 1991; Chomsky 1995b, 2000b; Roberts and Roussou 2003; Holmberg 2010a; Boeckx 2011a). More precisely, parametric variation is associated with lexical properties of a closed class of functional categories which trigger syntactic operations in order to license their (abstract) morphological content (including minimally

C, T,  $\nu$ , and D; see, e.g., Rizzi 1997 and Cinque 1999 for more elaborate inventories of functional categories; see section 18.5 for diachronic evidence that bears on this issue). On this assumption, syntactic change is to be identified with changes affecting the feature content of functional categories (e.g., via phonological erosion or grammaticalization processes, see Longobardi 2001a; Roberts and Roussou 2003; Roberts 2007). The fact that syntactic change is usually highly systematic (in contrast to other types of change, which may be sporadic, i.e., confined to a single lexical element) can then be accounted for if we adopt the assumption that the set of core functional categories must be present in every well-formed syntactic representation as the basic 'building blocks' or 'skeleton' of clause structure (see von Stechow 1995; Chomsky 1995b, 2000b; Hale 1998). Moreover, if overt inflectional morphology is taken to reflect the (abstract) feature content of functional categories, it is possible to construe a correlation between syntactic change and morphological change, that is, to provide a principled explanation for the traditional observation that changes affecting the inflectional morphology of a given language often go hand in hand with syntactic change (see, e.g., Sapir 1921; see also Lightfoot 1991, 1999; Roberts 1993a, 2007; Haeblerli 1999, 2004; and the papers in Kemenade and Vincent 1997 and Lightfoot 2002).<sup>7</sup>

From this perspective, grammar change takes place when there is an innovative value  $v_j$  of a parameter  $p_k$  (i.e., a lexical property/feature value linked to a certain functional category) which is for some reason more 'accessible' in the input data than the relevant value  $v_i$  which is part of the target grammar. However, as already noted, this gives rise to (p. 466) an apparent paradox, which is sometimes called 'the logical problem of language change' (see Clark and Roberts 1993, 1994; Niyogi and Berwick 1998; Roberts 2007:230–231):<sup>8</sup>

*The logical problem of language change*

'if the trigger experience of one generation, say  $g_1$ , permits members of  $g_1$  to set parameter  $p_k$  to value  $v_i$ , why is the trigger experience produced by  $g_1$  insufficient to cause the next generation to set  $p_k$  to  $v_i$ ?'

(Clark and Roberts 1994:12)

As already pointed out in section 18.1, this paradox is intimately connected to the question of what factors can account for the actuation of changes (Weinreich et al. 1968). In current theoretical approaches to language change, it is generally assumed that a solution to the actuation problem must be based on a closer inspection of the set of factors that may impede perfect transmission of linguistic features during L1 acquisition.

## **18.3 Language Acquisition and the Actuation Problem**



Under the assumption that language acquisition is a deterministic process (that is, two different sets of input data give rise to two different grammars), the possibility of change can be related to shifts in the Primary Linguistic Data (PLD), that is the set of (partially parsed) linguistic signals on the basis of which the learner constructs a grammar: change occurs when learners fail to detect a trigger/cue for a certain property of the target grammar  $G_1$  in the linguistic input they are exposed to.<sup>9</sup> From this perspective, the actuation problem concerns the question of how and why the PLD generated by a target grammar  $G_1$  may trigger a grammar  $G_2$ , where  $G_1 \neq G_2$ . Moreover, answers to this apparent (p. 467) paradox must reconcile the phenomenon of language change with the standard assumption that language acquisition is highly accurate under normal circumstances. However, the omnipresence of change might be taken to suggest that the standard model of L1 acquisition is perhaps overly simplistic. To facilitate change between two grammars  $G_1$  and  $G_2$ , we must look for factors that might give rise to differences between  $PLD_1$  that led to the construction of  $G_1$  and  $PLD_2$  feeding the construction of  $G_2$ .<sup>10</sup>

One likely source of differences between  $PLD_1$  and  $PLD_2$  is grammar-external factors such as language contact or conscious changes adopted by adult speakers of  $G_1$ , for example the use of linguistic features associated with a prestige dialect, or the avoidance of features that are not part of the prestige dialect (see, e.g., Thomason and Kaufman 1988; for relevant examples in connection with standardization processes, see Weiß 2001 on the loss of multiple negation and doubly-filled Comp in Standard German, Hoeksema 1994 on the loss of verb projection raising in [written] Dutch, and Nevalainen and Tieken-Boon van Ostade 2006 on the history of English). As a result, the evidence for a certain property of the target grammar may be obscured by the fact that the child usually receives input from different speakers with possibly different grammars. The learner must determine which output string is generated by which grammar, a non-trivial task. It is at least conceivable that in such a situation, the child may mistakenly attribute a certain output string to the wrong grammar, which in turn may give rise to a new grammar with properties that differ from those of the target grammar (see Hale 2007:38-39 for discussion). A relevant example comes from Kroch and Taylor's (1997) analysis of the loss of V2 in the Middle English period. Kroch and Taylor attribute the loss of V2 to a mixed dialect situation where speakers of a northern V2 grammar came into contact with speakers of a southern variety in which subject pronouns regularly intervened between a fronted XP and the finite verb. According to Kroch and Taylor, the resulting mixed input (in particular, the systematic deviations from V2 generated by the southern grammars) led to the acquisition of a grammar that also generated V3 patterns, leading to the loss of V2 in the northern variety:

- (5) a. Output string generated by southern grammar: XP – pronoun –  $V_{fin}$   
 b. Output string generated by northern grammar: XP –  $V_{fin}$  – pronoun

The relevant grammar change would then result from a misanalysis in which learners mistakenly attributed output string (5a) to the northern grammar, which originally was a strict V2 grammar. However, one might argue that this outcome was actually not an

instance of grammar change, but rather the result of diffusion (see note 1). As already mentioned, there are presumably no strong linguistic constraints on contact-induced change. The same goes for changes triggered by sociolinguistic factors. In what follows, (p. 468) we will therefore focus on (internal) triggers of grammar change.<sup>11</sup> Relevant proposals typically assume that for some reason, the PLD the learner is confronted with differs from the PLD that gave rise to the target grammar, due to factors such as (stylistically motivated) shifts in the frequencies with which certain constructions are used, (morpho-)phonological erosion, linguistic variation inherent to speech production, or 'noise in the channel' that blurs the evidence for certain properties of the target grammar in the linguistic input the learner receives (see, e.g., Lightfoot 1979, 1991, 1999; Hale 1998, 2003, 2007; Roberts 2007).

It is a widespread assumption that stylistically motivated changes in language use may lead to significant changes in the make-up of the triggering experience (see Lightfoot 1991, 1999). As a result, the evidence necessary to trigger a certain property of the grammar may cease to be robustly expressed in the input data. When a certain threshold (e.g., in frequency) is crossed, this may lead to grammar change. Examples discussed in the literature include the reanalysis of surface VO patterns (in terms of underlying VO order) due to an overuse of NP postposition from an OV base (see Stockwell 1977; and van Kemenade 1987 on English), the rise of ergative/absolute case marking via a reanalysis of frequently used passive constructions (see Anderson 1977, 1980), or the loss of V2 due to an increased frequency of subject-initial clauses (see Lightfoot 1991, 1997; Roberts 1993a; Clark and Roberts 1993). Of course, this raises the question of what counts as robust quantitative evidence signaling (the absence of) a certain parameter value. Lightfoot (1997, 1999) links the loss of V2 patterns to the observation that in stable V2 languages such as German or Swedish, at least 30% of main clauses exhibit subject-verb inversion (XP-V<sub>fin</sub>-subj.). Lightfoot is led to conclude that a positive value for the V2 parameter can be acquired as long as the frequency of examples with inversion does not drop below the threshold of 30%.<sup>12</sup>

It is commonly assumed that fluctuations in the frequency with which certain patterns are attested in the PLD may give rise to reanalysis, that is, a process in which a (p. 469) given surface string is assigned an underlying structure that differs from the relevant structure in the target grammar.<sup>13</sup> While reanalyses do not alter the surface manifestation of the relevant syntactic patterns, they may give rise to other changes, in particular, the obsolescence of forms generated by earlier grammars (in the examples mentioned, the loss of OV orders, and the loss of subject-verb inversion [i.e., a V2 grammar], respectively). The idea that grammar change is triggered by shifts in E-language predicts that language change is a contingent and unpredictable process that depends on the availability (and frequency) of certain cues in the linguistic input the learner is exposed to (which may be subject to random fluctuations). This conclusion seems to be a logical consequence if we assume that incomplete transmission of linguistic features cannot be

attributed to built-in factors/tendencies (such as speaker/hearer economy) that drive change, but results solely from the interaction between properties of the PLD and UG.

However, purely frequency-based approaches have been criticized by Kroch (2001), who points out that usage frequencies may remain stable over long periods of time. Another set of problems comes from the traditional observation that language change does not seem to be random in the way expected. On the one hand, there seem to be cases of *drift* (see, e.g., Sapir 1921:160ff.), where languages are apparently subject to long-term, highly gradual directional changes (e.g., from the inflected to the isolating morphological type). On the other hand, it is a well-known fact that certain types of change are more likely than others (e.g., a change from basic OV order to VO seems to be much more frequent than the reverse change from VO to OV; see Kiparsky 1996, Roberts 2007) and that there are pathways of change that shape the historical development of languages cross-linguistically (e.g., in grammaticalization processes, Hopper and Traugott 2003). See section 18.3.1 on how these observations can possibly be captured by UG-based approaches to language change.

(p. 470) An alternative way to tackle the actuation problem is to assume that grammar change may be triggered by independently motivated changes in other parts of the grammar. This line of research has been (and continues to be) very productive as a means of explaining syntactic change by correlating it with (independently triggered) changes affecting properties of inflectional morphology (see Sapir 1921 for an early statement concerning the connection between the loss of case marking and the rise of SVO word order).<sup>14</sup> Relevant generative case studies include the impact of the loss of verbal inflection on the availability of verb movement (for the history of English, see Roberts 1993a; for the Scandinavian languages, see Platzack 1988 and Holmberg and Platzack 1995) and pro-drop (for the history of French, see Roberts 1993a, Vance 1997; for present-day varieties of French see Roberts, 2010a; for Swedish, Falk 1993; for the Scandinavian languages in general, see Holmberg and Platzack 1995; for English, Allen 1995, Haeberli 1999), the relation between the loss of nominal inflections (i.e., case) and changes affecting word order and the rise of Exceptional Case-Marking constructions (for the history of English, see Lightfoot 1979, 1991, 1999; van Kemenade 1987; Kiparsky 1996, 1997; Roberts 1997; Haeberli 1999, 2004; Biberauer and Roberts 2005, 2008), or changes affecting the inventory of C-related clausal particles and the rise of generalized V2 syntax in Germanic (see Ferraresi 2005 on Gothic; and Axel 2007 on Old High German).<sup>15,16</sup> Other causal connections that are posited to address the actuation problem include the rise of phonemic contrasts due to the loss of a segment that originally triggered a relevant phonological rule ('phonemicization,' see, e.g., Hyman 1976), the reanalysis of prosodically triggered ordering properties in terms of syntactic operations (see Stockwell and Minkova 1994, Dewey 2006 on the rise of V2 in Germanic), and the interaction of morphosyntactic change with shifts in the semantic/pragmatic function of expressions. The latter is usually based on the assumption that the semantic/pragmatic function of linguistic forms may become opaque over time (e.g., due to overuse or the development of competing, more 'expressive' forms), leading to a loss or reanalysis of forms (see, e.g., Givón 1976 on the rise of subject agreement via a reanalysis of clitics in

left-dislocation structures; Stockwell 1977, van Kemenade 1987 on the rise of VO orders via a reanalysis of NP postposition; Hinterhölzl 2004, 2009 on the change from OV to VO order; Roberts 2007:276–277 on a possible scenario for the rise of OV order via a reanalysis of discourse-driven leftward movement of objects; (p. 471) Fuß 2008, Hinterhölzl and Petrova 2009, Trips and Fuß 2009, and Walkden 2015 on the diachrony of V2 in Germanic).<sup>17</sup>

Attempts to link parametric change to independent changes in other parts of the grammar potentially offer an internal solution to the actuation problem; in addition they may reveal insights into the interaction between different components of the grammar (and thus properties of UG) which possibly cannot be reached from a purely synchronic perspective.<sup>18</sup> However, one might argue that this approach merely shifts the burden of explanation (with respect to actuation) to another part of the grammar, that is, we may ask why learners at some point failed to acquire certain morphological or phonological properties.

Solutions to the latter problem often capitalize on the fact that the acquisition of morpho-phonological properties is a complex and difficult task due to the often messy and chaotic character of the acoustic input the learner receives. This is what Hale (2007:53ff.) calls ‘noise in the channel.’ As is well known, the output of the articulatory system may be highly variable, even for a single individual (due to random factors such as speed of pronunciation, a cold, etc.). Moreover, it can be shown (e.g., by methods of instrumental phonetics such as spectrographic analyses) that even a single speaker seldom realizes one and the same linguistic sign (sounds, morphemes, words, sentences) in exactly the same way. The variation inherent in the target realization may obscure properties of the target grammar  $G_1$ . If the range of variation in the phonological realization of a given underlying target structure crosses a certain threshold, this may cause the learner to posit an underlying form that differs from the relevant structure in the target grammar (see Hale 2003, 2007 and in particular the work of John Ohala [Ohala 2003 for an overview] for relevant considerations concerning phonetic aspects of phonological change; see also Blevins 2004 and chapter 8 for discussion).<sup>19</sup> In this way, the messy character of the incoming data may give rise (p. 472) to misparses/misanalyses that lead to wrong conclusions concerning properties of the target grammar. If these conclusions fail to be corrected (e.g., by further evidence to the contrary), they will eventually become part of the acquirer’s steady state grammar, an instance of grammar change.

Under the assumption that language acquisition is a deterministic process (see chapters 11 and 12), it is fairly clear that the complex mapping from  $G_1$  to properties of  $G_2$  introduces quite a number of random factors (‘noise in the channel’) that may prevent a flawless transmission of features from  $G_1$  to  $G_2$ . Moreover, from the fact that the make-up of the input seems to be different for each individual speaker we may conclude that change is not a rare phenomenon (as suggested by the *logical problem of language change*), but rather a necessary consequence of the way human languages are transmitted over time. From this perspective, the linguistic diversity we find in the world today (according to some estimations at least 6,500 different languages, belonging to at

least 250 larger families, see Nettle 1999:1) can be directly attributed to the discontinuous nature of language transmission: imperfect replication of grammars gives rise to small-scale inter-speaker variation which might grow into dialects or even mutually unintelligible languages as a result of social selection and geographical isolation (see Nettle 1999 for in-depth discussion). At this point, it is interesting to note that the wide range of linguistic diversity that we can observe today (not to mention the languages that have vanished before now) developed in a comparatively short period of time (100,000–200,000 years, or 4,000–8,000 generations if we assume a generation time of twenty-five years; see also chapter 14, section 14.4.2). The massive diversification of languages is a testimony to the pervasiveness and rapidity of linguistic change; it also highlights that the evolution of languages and the evolution of biological species operate at different speeds, a finding which can be traced back to the fact that only the latter is based on a direct transmission of properties (via inheritance).<sup>20</sup>

We have seen that the likelihood and directionality of (different types of) sound change may be attributed to purely phonetic factors (see, e.g., Hale 2007). However, it is still unclear how UG-based approaches can handle directionality effects and phenomena such as drift in other domains of the grammar.

### 18.3.1 Drift, Trajectories of Change, and the Role of ‘Third Factors’

In general, UG-based approaches to the actuation problem should lead us to expect that language change is basically a random phenomenon, that is, a ‘random walk through the range of possibilities defined by UG’ as Roberts (2007:348) puts it. Again, this (p. 473) expectation can be ultimately traced back to the discontinuous nature of language transmission, compare the following quote from David Lightfoot:

Languages are learned and grammars constructed by the individuals of each generation. They do not have racial memories such that they know in some sense that their language has gradually been developing from, say, an SOV and towards an SVO type, and that it must continue along that path. After all, if there were a prescribed hierarchy of changes to be performed, how could a child, confronted with a language exactly half-way along this hierarchy, know whether the language was changing from type x to type y, or vice versa?

(Lightfoot 1979:391)

However, this prediction is clearly at odds with the bulk of evidence accumulated in historical linguistics in support of the existence of (universal) pathways of change, including probable, possible, and impossible sound changes (see, e.g., Campbell 2004; Blevins 2004), grammaticalization clines (i.e., the observation that the development of functional categories from former lexical elements proceeds via identical stages across time and languages, e.g., full verb → modal verb → auxiliary → inflection, see, e.g., Heine and Kuteva 2002; Hopper and Traugott 2003), or the observation that basic OV order is often replaced by VO while the reverse change seems to be quite rare (Faarlund 1990:50; Kiparsky 1996:140).<sup>21</sup> Attempts to reconcile the observation of diachronic pathways with the notion of parametric change often converge with another line of thinking to solve the actuation problem, namely the assumption that there are properties of the acquisition device that may promote changes or determine the direction of change in case the evidence contained in the PLD is ambiguous or insufficient (see, e.g., Roberts and Roussou 2003:3ff. and Roberts 2007:345ff. for discussion).<sup>22</sup> It is usually assumed that these factors (p. 474) are not specific to the faculty of language. Hence, they are third factors in linguistic variation according to the typology introduced by Chomsky (2005:6; see section 18.1 and chapter 6). From this perspective, ambiguity in the PLD regarding the expression of certain parameter values is only a necessary condition for a change to take place, while the actuation of a change is attributed to learning strategies that select the most economic variant compatible with the input (relevant early generative proposals include Lightfoot’s 1979 *Transparency Principle*, or the *Least Effort Strategy* proposed in Clark and Roberts 1993).<sup>23</sup> In general, these economy considerations come in two varieties. First, following early generative work on L1 acquisition (see Wexler and Culicover 1980, Berwick 1985 for discussion), it is often assumed that there are marked and unmarked (or default) parameter values and that the learner assigns a given parameter the unmarked value if no decision can be made based on the evidence

available in the input.<sup>24</sup> Alternatively, the decision in question is assumed to be sensitive to the notion of derivational/representational economy (see, e.g., Clark and Roberts 1993), in the sense that the learner assigns a given input string the most economical representation/derivation that is compatible with the evidence. Economy-driven approaches to change typically focus on the diachronic loss of movement operations (see, e.g., Clark and Roberts 1993; Roberts 1993a, 1993b), while markedness considerations can be used to capture other changes such as the loss of pro-drop. More recently, Roberts and Roussou (2003) and Roberts (2007) have proposed a synthesis of these two approaches that is based on the idea that marked parameter values give rise to more complex syntactic derivations. Under the assumption that marked parametric choices correspond to the presence of formal features (associated with functional heads) which drive syntactic operations (Move, Agree), this idea can be formalized by the following simplicity metric (Roberts 2007:235; see also Longobardi 2001a):

- (6)** Given two structural representations R and R' for a substring of input text S, R is simpler than R' if R contains fewer formal features than R'.

(p. 475) (See also the discussion of Feature Economy in chapter 14, section 14.9). (6) ensures that in cases where the PLD underdetermines the target grammar, learners will generally prefer grammars that contain fewer formal features and thus yield simpler representations/derivations with a smaller number of movement operations. This can then be used to explain the phenomenon of drift in terms of a series of discrete and independent changes that successively reduce the inventory of (marked) formal features present in the grammar and in this way may give rise to the impression of generation-spanning, directional change (see Biberauer and Roberts 2005, 2006 for the change from OV to VO in the history of English as a combination of the loss of (i) EPP features and (ii) pied-piping of verbal projections; see Roberts 2007 for further discussion). In a similar vein, grammaticalization clines can be explained as reflecting a series of reanalyses in which movement operations linking positions in a fixed hierarchical sequence of functional heads (see, e.g., Cinque 1999) are converted into Merge operations that target the head of the former movement chain (Roberts and Roussou 2003). As a result, grammaticalization is expected to proceed upwards along the clausal spine, as exemplified by the cross-linguistically widespread development of future markers from lexical verbs, where the latter first turn into modals expressing obligation or necessity ( $\text{Mod}_{\text{necessity}}^0/\text{Mod}_{\text{obligation}}^0$ ) which are then reanalyzed as a realization of  $\text{T}_{\text{future}}^0$ .<sup>25</sup> In this way, the study of grammaticalization phenomena can also provide insights into the make-up of the clausal architecture and the nature of functional heads (see Roberts and Roussou 2003, ch. 5; Roberts 2010b; see section 18.5 for further discussion).

## 18.4 Diachronic Gradualness, Real and Imagined: The Transition Problem

The transition from one historical state to another is often described as a gradual process that is typically accompanied by a degree of linguistic variation which is not found in stable languages (see also Labov 1994 on the connection between variation and change). Changes do not manifest themselves instantaneously in the historical records. Rather, the replacement of old forms by new forms over time follows an S-shaped curve (which corresponds to the logistic function, which is also used in population biology, e.g., to model population growth). Thus, changes typically start slowly, gather momentum (up to a point where the growth is approximately exponential) before they lose speed and tail off slowly to completion (see Bailey 1973; Kroch 1989). These observations are at the heart of (p. 476) the so-called 'transition problem' highlighted by Weinreich et al. (1968:153), which concerns 'the route by which a linguistic change is proceeding to completion.' To reconcile the apparently gradual character of language change with the notion that (grammar) change is necessarily abrupt (involving a set of discrete differences between the target grammar and the grammar acquired by the learner), it is often assumed that the impression of gradualness arises from a set of independent factors which traditional approaches failed to recognize (Kroch 1989, 2001; Hale 1998, 2007; Lightfoot 1999:ch. 4; Roberts 2007:ch. 4):

- (i) the distinction between grammar change and its diffusion (Hale 2007; see section 18.2.1);
- (ii) the possibility of intra-speaker variation, which can be modeled as a form of bilingualism (*grammar competition*, Kroch 1989, 2001); and
- (iii) the use of a more fine-grained system of parametric choices ('microparameters,' Kayne 2000, 2005a), see Roberts 2007, 2010b; and chapters 14 and 16;
- (iv) 'true' optionality of syntactic operations (Biberauer and Roberts 2005).

First of all, the impression of gradualness may arise from not distinguishing properly between the actual change and its (highly gradual) diffusion within a population/speech community. As pointed out by Hale (1998:3), the actual (parametric) change 'has no temporal properties—it is a *set of differences*' (original emphasis) governed solely by the interaction of cognitive processes (including properties of UG) with the linguistic experience. In contrast, the gradual diffusion of a change through a speech community necessarily has a temporal dimension. It is governed by sociolinguistic factors (prestige, register choice, etc.) and can be modeled by mathematical methods of population biology (see, e.g., Niyogi and Berwick 1998; Niyogi 2006).

However, work by Anthony Kroch and his collaborators has shown that linguistic change is typically accompanied by another kind of variation which cannot be attributed to (incomplete) diffusion/inter-speaker variation. Relevant examples involve linguistic variation within one and the same text, compare the relative order of verbs and their complements in the following passage taken from the (northern) Early Middle English *Ormulum*:



Forr þatt I wolde bliþelig þatt all Ennglisshe lede wipþ ære sholde  
 For that I would gladly that all English people with ear should  
 lissstenn itt, wipþ herte sholde itt trowwenn, wipþ tunge sholde  
 (7) listen it, with heart should it trust, with tongue should  
 spellenn itt, wipþ dede sholde itt follghenn.  
 spell it, with deed should it follow.  
 (CMORM,DED.L113.33; Trips 2002:112)

According to Kroch, cases such as (7) represent genuine instances of intra-speaker variation, where speakers have command over two (or more) internalized grammars (p. 477) which differ with respect to a small set of parametric choices, giving rise to a wider range of linguistic variation (so-called ‘grammar competition,’ see, e.g., Kroch 1989, Han and Kroch 2000, and Ecay 2015 on the rise of *do*-support in English; Pintzuk 1999, Pintzuk and Taylor 2006 on word order change in the history of English; Santorini 1992 on Yiddish; Taylor 1994 on Ancient Greek; see also Kroch 1994, 2001 and Yang 2000; see Hale 2007:172ff. for critical discussion).<sup>26</sup> Thus, it is claimed that language change typically proceeds via a stage of ‘internal diglossia’ up to a point where one grammar (or parametric choice) eventually wins out over the other.<sup>27</sup> The replacement of one grammar/parametric option by another is taken to follow the same S-curve which characterizes change in a population; moreover, it is assumed that different surface manifestations of a single underlying parameter change replace competing (older) forms at the same rate in all contexts (the so-called Constant Rate Effect, CRE). Under this assumption, estimations of the rate of change can also be used as diagnostic to identify clusterings of surface changes that can be attributed to a single underlying change (see Kroch 1989, and more recently Ecay 2014; see Fruehwald et al. 2013 on Constant Rate Effects in phonological rule change).<sup>28,29</sup>

(p. 478) In addition, the impression of gradualness (and variation) might be due to grammar-internal factors. Roberts (2007:300ff., 2010b) discusses the possibility that the phenomenon of gradualness can be reconciled with the notion of parametric change if we replace the traditional notions of ‘parameter’ and ‘syntactic category’ with a system of more fine-grained microparametric choices (either involving a multitude of functional heads or a more elaborate system of morphosyntactic features), allowing for the possibility of ‘lexical diffusion’<sup>30</sup> of formal features (such as EPP-features) through a rich system of functional heads.<sup>31</sup> Under this assumption, ‘a series of discrete changes to the formal features of a set of functional categories taking place over a long period [may give] the impression of a single, large, gradual change’ (Roberts 2007:300). An alternative grammar-internal source of variation is true optionality of syntactic operations (see Biberauer and Richards 2006; for a relevant proposal to account for word order variation in Old and Middle English, see Biberauer and Roberts 2005).

We can therefore conclude that apparent diachronic gradualness does not impose a challenge for UG-based approaches to linguistic change. Rather, change between grammars is necessarily abrupt and instantaneous, with the impression of gradualness arising from a set of independent grammar-external and grammar-internal factors. The

next section discusses the phenomenon of grammaticalization, the prime example of an allegedly gradual change, and shows how this diachronic process can not only be subsumed under the model of language change outlined so far, but may also provide interesting evidence bearing on the nature of functional categories.

## 18.5 Grammaticalization and the Nature of Functional Categories

It is a well-known observation (going back at least to 19th-century grammarians such as Franz Bopp 1816) that grammatical categories like determiners, inflections, conjunctions, or auxiliaries evolve historically from formerly (free) substantial lexical categories such as nouns or verbs.<sup>32</sup> Studies of grammaticalization phenomena deal (p. 479) with (i) the historical pathways along which lexical elements develop into grammatically functional elements and then further into 'more grammaticalized' functional elements and (ii) the (pragmatic, morphosyntactic) contexts where this development takes place. Descriptive/typological studies often emphasize that grammaticalization is a separate type of change (in addition to well-known types such as reanalysis, or analogy/extension) in which the syntactic category of a lexical item changes gradually and irreversibly<sup>33</sup> along universal (historical) pathways, called *grammaticalization clines* or *paths* (see Hopper and Traugott 2003; Heine 2003), compare the following cline that characterizes the development of agreement markers (over several generations of speakers):

- (8) independent pronoun → weak pronoun → clitic pronoun → affixal (agglutinative)  
agreement marker → fused agreement marker → ∅

According to some researchers, the particular properties of grammaticalization phenomena pose a challenge to theoretical views widely held among generative linguists. For example, it has been argued that the seemingly gradual transition from one syntactic category to another is not compatible with the assumption of discrete syntactic categories (see Hopper and Traugott 2003, Heine 2003; but see Newmeyer 1998 for an opposing view). Moreover, the long-term, unidirectional character of clines seems to be incompatible with the fact that language transmission is necessarily discontinuous: learners do not have any historical knowledge that tells them that they currently take part in a generation-spanning grammaticalization process and that they must proceed further down the cline. Rather, the new grammar is created anew in the mind of the child, solely on the basis of (i) the linguistic experience the child is exposed to, and (ii) properties of the human language faculty. The challenge posed by grammaticalization phenomena has also been taken up in generative work (see note 32 for references). For example, Roberts and Roussou (2003) argue forcefully that the properties of grammaticalization can be accounted for in more formal terms if we adopt the following assumption (see also von Stechow 1995; Roberts 2007, 2010b):<sup>34</sup>

**(9) Grammaticalization involves the reanalysis of substantial lexical elements as phonological exponents of (higher) functional categories/heads.**

(p. 480) From this perspective, grammaticalization is taken to involve the reanalysis of a movement dependency where the phonological features of an element moved to a higher functional head F (or to the specifier of F<sup>35</sup>) are analyzed as the phonological realization of F. Accordingly, grammaticalization prototypically proceeds in an upwards fashion (Roberts and Roussou 2003): the phonological exponent [π] of a hierarchically lower lexical category is reanalyzed as the exponent of a higher functional head. The unidirectional character of this type of change can then be attributed to economy principles (i.e., third factors in the sense of Chomsky 2005) that favor less complex structures and derivations and drive grammaticalization processes in case the input data is ambiguous (see Roberts 1993b; Roberts and Roussou 1999, 2003; van Gelderen 2004).<sup>36</sup>

It can be shown that under this view, grammaticalization processes are not only fully compatible with standard generative assumptions, but can also serve to deepen our understanding of the nature of functional categories, in particular concerning (a) their featural make-up and (ii) the hierarchical organization of functional projections in the structure of the clause. The following discussion focuses on a set of general properties of grammaticalization phenomena that have been described in the literature (see, e.g., Lehmann 2002; Hopper and Traugott 2003), including the gradual loss of phonological and semantic content, the development from free into bound forms, the loss of variation concerning the use and placement of forms as well as the rise of selectional restrictions and properties characteristic of elements that are organized into (inflectional) paradigms.

The reduction of phonological substance can be characterized as the loss of segments or of marked phonological features (e.g., the reduction of Latin *ille* to Romance *le*, see Vincent 1997). Phonological reduction is usually accompanied by a loss of prosodic independence which proceeds along the following pathway:

**(10) content item > function word > clitic > inflectional affix > phoneme > ∅**

Adopting the basic assumption in (9), phonological attrition can be attributed to independently motivated, universal properties of functional categories. It is a well-known fact that functional elements such as complementizers, determiners, auxiliaries, and inflections are prosodically/phonologically deficient. In contrast to lexical categories, they cannot bear stress and are preferably monosyllabic/moraic elements (see Kenstowicz 1994 on English; Vogel 1999 on Italian).<sup>37</sup> Thus, we might suppose that the (p. 481) absence of stress and a reduced segmental make-up is presumably a necessary precondition for the reanalysis as a functional element (see Roberts and Roussou 2003 for discussion). After the initial reanalysis, it is predicted that the element is prone to undergo further phonological reduction due to the deficient phonological nature of functional categories, accounting for the pathway described in (10). The inherent prosodic deficiency of functional categories can also be used to account for the development of bound forms from formerly free forms (sometimes referred to as

'coalescence,' see Lehmann 2002). Again, this clearly parallels the behavior of functional categories (e.g., inflections), which are often realized by affixes or clitic elements that require the presence of a (lexical) host they can attach to (clearly a consequence of the general phonological deficiency of functional categories, henceforth FCs).<sup>38</sup>

Typically, phonological reduction is accompanied by the loss of semantic content, sometimes labeled *bleaching*. Informally, the notion of semantic bleaching refers to the erosion of substantial lexical meaning (see Eckhardt 2006 for in-depth discussion and a thorough critique of the notion of 'bleaching'). For example, we can observe that nouns evolving into nominalizing affixes, and pronouns evolving into agreement markers lose their referential (and descriptive) potential while verbs evolving into auxiliaries lose their thematic (predicational) properties:

- (11)** German *-hett* < Old High German *hett*, Gothic *haidus* 'person, nature, form, rank'  
Finnish
- (12)** a. 1pl AGR *-me* < 1pl pronoun *me*  
b. 2pl AGR *-te* < 2pl pronoun *te*  
English *shall*: future marker < '(root) modal' < OE *sceal*, full verb meaning
- (13)** 'subject has to pay an amount of money or has to return something to somebody'  
(Lehmann 2002:114)

However, it has been pointed out that the semantic changes characteristic of grammaticalization do not affect the semantic properties of a given lexical item in a random fashion (von Stechow 1995; Roberts and Roussou 2003; Eckhardt 2006). First, grammaticalization typically involves the loss of substantial lexical meanings, in the sense of, for example, 'predicative' or 'descriptive' content:

- a. Verbs lose their argument structure (or their thematic properties);
- (14)** b. Nouns lose their referential (and descriptive) content;
- c. Prepositions lose their capacity to designate spatial relations.

(p. 482)

Second, another set of semantic properties is preserved or added to the formerly substantial lexical category as a result of the grammaticalization process. This set of semantic properties involves abstract 'logical meanings':

- a. modals/auxiliaries preserve the modal content of the relevant former full verbs (as in the case of Old English *sceal* noted in (13));
- b. addition of the feature [ $\pm$ definiteness] in the case of determiners evolving from nouns or numerals;
- (15)** c. addition/preservation of abstract meanings typically expressed by an operator-variable relationship (development of (I) *wh*-pronouns from former Indefinites, (II) quantifiers such as English *many* from former OE adjective *manig*).

Thus, semantic changes typical of grammaticalization are not adequately described by the term semantic bleaching, since they show a set of clearly structured properties and, even more important, the retention or even addition of *logical* semantic content. In this way, the study of grammaticalization phenomena can inform us about semantic properties of functional categories. For example, Roberts and Roussou (2003) argue that elements subject to grammaticalization typically develop quantificational properties which they assume to be a characteristic of functional categories.<sup>40</sup> Moreover, if FCs are confined to a certain kind of semantic content, then it is expected that the transition from lexical to functional category requires (i) the loss of those semantic properties that are not compatible with the universal make-up of FCs, that is, nonlogical (i.e., predicative or descriptive) content and (ii) the retention or addition of logical content.

The loss of semantic content is typically accompanied by changes affecting selectional properties, such as a simplified subcategorization frame and a development from syntactic selection towards morphological selection (labeled *condensation* by Lehmann 2002). An often-discussed example comes from the development of the Romance synthetic future (Roberts 1993b; Roberts and Roussou 2003):

- a. Main verb (*habere*) that took a nonfinite clause as its complement →
- (16) b. Future-indicating auxiliary selecting a verbal projection →
- c. Inflectional suffix *m*-selecting a verb stem.

(p. 483) Again, this feature of grammaticalization processes can tell us something about the universal profile of functional categories (and the distinction between functional and lexical elements with respect to selection). For example, we might conclude that diachronic data support the idea that only functional categories can select for lexical categories, whereas lexical categories cannot select for other lexical categories (Roberts and Roussou 1999).

Furthermore, the study of grammaticalization phenomena may contribute new evidence to the debate on whether the inventory of functional categories is subject to language-specific parameterization or the same set of functional categories is universally present in all languages (see, e.g., Thráinsson 1996; Bobaljik and Thráinsson 1998). In this context, the observation that forms that were previously optional tend to become used obligatorily when subject to grammaticalization is potentially relevant. A well-known example comes from the rise of determiners, which cross-linguistically (e.g., in Romance [Vincent 1997; Giusti 2001] or Germanic [van Gelderen 2007]) develop from former demonstratives. While demonstratives are typically restricted to contexts where they add an indexical meaning to nouns, the resulting determiners are always obligatory. This follows if (at least a set of core) functional categories are taken to be the building blocks of syntax, which are universally present (either with phonological content or zero). Thus, if a determiner ( $D^0$ ) has a PF-realization / $\pi$ /, then / $\pi$ / will show up in all DPs, since  $D^0$  is a necessary component of a nominal expression (see von Stechow 1995; but see van Gelderen 2004 for the claim that grammaticalization processes may enrich the inventory of non-core functional categories). In a similar vein, the fact that grammaticalized elements tend

to occur in a fixed position (which Lehmann 2002 terms *fixation*) can be attributed to properties of FCs if we assume (i) that the order/hierarchical position of functional categories is universally fixed (Cinque 1999; Chomsky 2000b) and (ii) that the relative order of inflectional affixes is determined by the syntactic structure (see, e.g., the Mirror Principle, Baker 1988). As already pointed out in section 18.3.1, on the additional assumption that the relevant reanalyses convert movement dependencies into Merge operations targeting the head of the former movement chain, grammaticalization clines may then be taken to reflect the (universal) hierarchy of functional categories in the structure of the clause (see Roberts 2010b).

Finally, grammaticalization processes may also provide insights into differences between lexical and functional categories at the interface between syntax and morpho-phonology. Grammaticalization typically involves the development of open class lexical elements into elements that belong to a closed class of functional words or are integrated into a morphological paradigm (typically exhibiting a certain degree of formal, functional and semantic homogeneity). This is what Lehmann (2002) calls *paradigmatization*. Fuß (2005) argues that paradigmaticity can be attributed to universal properties of FCs under the assumption that paradigmaticity results from the way FCs are supplied with phonological content in the morphological component of grammar. More specifically, the phenomenon of paradigmaticity is taken to reflect a crucial difference between the phonological realization of lexical and functional categories. Lexical (p. 484) material typically involves an arbitrary, idiomatic pairing of form and meaning. This can be modeled either by assuming early, presyntactic insertion of lexical roots that include phonological features (Harley and Noyer 1999), or by assuming that substantial lexical heads (N, V, A, P) can be realized by any Vocabulary item that matches the category specification (e.g., in English, N may be realized by /kæt/, /dæg/, /bʊʃ/, etc., dependent on the choice of the speaker; see Marantz 1995, 1997). In contrast, the phonological realization of functional categories is non-arbitrary: it involves a competition between Vocabulary items that are specified for a (common) subset of the inflectional features contained in the functional head. The item realizing the greatest number of inflectional features is then chosen for insertion (the Subset Principle, Halle 1997). Under these assumptions, the traditional notion of a paradigm can be reconstructed in more formal terms by assuming that the insertion procedure for functional categories differs significantly from the way Vocabulary Insertion proceeds in the case of lexical heads. The impression that certain items 'belong to the same paradigm' arises from the fact that they are specified for an identical subset of morphosyntactic features (and thus form a natural class). In this way, increasing paradigmaticity (as a result of grammaticalization) can be systematically linked to the reanalysis of exponents of lexical items as exponents of functional categories.

## 18.6 Concluding Summary

It is a truism that the data set available to the historical linguist is very small if we compare it with the empirical sources available to linguists working on present-day languages. There are only a restricted number of historical records, and we do not have access to speaker judgments, or any kind of negative evidence. Thus, no matter how carefully we make use of the evidence available to us, we still have to face the fact that there are major gaps and discontinuities in the historical records. Given the state of the empirical evidence, a key question for historical linguistics is how to bridge the gaps in our knowledge of the past. Going back at least to Jakobson's (1931) *Prinzipien der historischen Phonologie* [Principles of historical phonology] an influential line of thinking suggests that formal approaches to language and language change which are based on abstract theoretical notions offer ways to (partially) fill in the gaps left by the historical evidence via formulating precise diachronic analyses and, ideally, a restrictive theory of linguistic change which delimits the set of transmission failures that can occur in a language with a given set of structural properties (this is what Weinreich et al. 1968 call the 'constraints problem').<sup>41</sup> In this chapter, I have argued that generative, (p. 485) UG-based approaches not only sharpen our understanding of the past; in addition, the study of language change can inform us about properties of the human language faculty as well, via making available evidence that cannot be gathered by purely synchronic investigations.

Section 18.2 introduced the notion of grammar change (a set of discrete differences between the target grammar and the grammar acquired by the learner) as the proper object of a restrictive approach to language change, arguing that only from this perspective can we hope to discover restrictions on possible changes imposed by properties of UG and the workings of language acquisition. In addition, it was shown how grammar change can be modeled in terms of parameter change, that is, diachronic variation in lexical properties of a closed class of functional categories.

In section 18.3, I addressed the actuation problem, suggesting that upon closer inspection, change is not a rare and paradoxical phenomenon, but rather a not unlikely outcome of the process of language acquisition if learners are highly sensitive to small fluctuations in the linguistic input they receive. To reconcile this perspective with the fact that there appear to be recurrent pathways of change that can be observed in many different languages, it has been suggested that the acquisition process is shaped by general cognitive principles of data analysis and efficient computation ('third factors' in the sense of Chomsky 2005; see chapter 6) that tip the scales in favor of less complex structures in case the evidence contained in the PLD is not sufficient to determine the value of a given parameter.

Section 18.4 discussed what Weinreich et al. (1968) call the 'transition problem,' that is, the question of how a language moves from one state to a succeeding state, focusing on the problems posed by diachronic gradualness and linguistic variation. I have argued that the impression of gradualness arises from a set of independent factors including inter-speaker (due to the gradual diffusion of a change in a speech community) and intra-

speaker variation (i.e., grammar competition), microparametric change, and the possibility of true formal optionality of syntactic operations.

Section 18.5 was concerned with the formal analysis of grammaticalization phenomena, arguing that from the perspective of a restrictive theory, this type of change may potentially reveal insights into the nature and (possibly universal) inventory of functional categories, which constitute major topics of interest in current research into the make-up of the human language faculty/UG. In this way, the study of language change appears as an integrated part of the generative enterprise, which hopefully will become even more productive when future historical linguists are no longer confined to written records of the past, but have access to a richer data set made available by large corpora and long-term investigations of developments in (present-day) colloquial speech.<sup>42</sup>

## Notes:

(<sup>1</sup>) According to Hale (1996:16), the difference between innovation (i.e. grammar change) and diffusion can be defined in somewhat more formal terms as follows (see also Hale 2007:36):

(i) *Innovation*

- a. The target grammar that generates the PLD has properties X, Y, Z.
- b. The grammar acquired by the learner has properties X, Y, W.

(ii) *Diffusion*

- a. There is a 'mixed' PLD generated by a grammar with properties X, Y, Z and another grammar with properties A, Y, W.
- b. The grammar acquired by the learner has properties X, Y, W.

Thus, cases of diffusion in fact do not represent instances of change, since the learner 'has *accurately* adopted a linguistic feature from some speaker' (Hale 2007:36, original emphasis).

(<sup>2</sup>) Of course, the study of diffusion can also reveal important insights in that it tells us something about social aspects of language, for example the factors that govern the diffusion of forms, the social stratification of speech communities, social factors that govern linguistic variation, etc. Crucially, however, it does not tell us much about language, that is, grammar, itself. See Niyogi and Berwick (1997, 1998), Niyogi (2006) for formal approaches that model the diffusion of a change through a speech community in terms of dynamical systems theory.

(<sup>3</sup>) However, see Heine and Kuteva (2005, 2008) for an opposing view.



(<sup>4</sup>) The timing of the intermediate stages is usually taken to be shaped by processes of cognitive maturation that control certain aspects of language development (see, e.g., Borer and Wexler 1987, 1992; Guasti 2002; and chapter 12). The idea that linguistic competence matures in the course of language acquisition can be used to explain certain differences between the child language and the target grammar such as the availability of so-called Root Infinitives (see Wexler 1994, Rizzi 1994, Roberts 2007 for discussion). Still, it is generally assumed that each of the intermediate stages must represent a possible human grammar, that is, 'there are no dead ends in language acquisition' (Chomsky 2002: 130-131).

(<sup>5</sup>) As pointed out by Hale (2007), conflicting evidence that is encountered after L1 acquisition has terminated does not lead to revisions of G2, but rather triggers the acquisition of an additional grammar, leading to multilingualism. Another possibility not discussed by Hale is to assume that conflicting evidence that cannot be associated with an additional grammar is simply ignored by the learner (e.g., if it is not frequent or systematic enough; see chapters 11 and 12).

(<sup>6</sup>) This does not amount to saying that there are no changes affecting the syntax of a given language apart from parametric change. For example, there might be changes that have to do with the frequency of a given structural choice, or properties of individual, open-class lexical items (e.g., changes affecting the subcategorization frames of verbs, adpositions, etc.). Importantly for our purposes, however, it is only parametric changes (i.e., differences in the featural properties of a closed class of *functional* categories, as will be argued shortly) that are of theoretical interest and significance, in the sense that only they can reveal something about the structure and the workings of the language faculty/UG; see Lightfoot (1991, 1999), Roberts (2007), and Hale (2007) for discussion.

(<sup>7</sup>) Ideally, a small change in the featural properties of (core) functional categories should yield a number of (at times dramatic) distinct changes on the syntactic surface, which is a hallmark of parametric change (see Lightfoot 1979, 1991, 1999). See Holmberg (2010) for a minimalist attempt to account for the traditional assumption that a single parameter determines a complex of syntactic properties; see also chapters 14 and 16.

(<sup>8</sup>) This paradox is intensified by the belief widely held in recent language-acquisition studies that children set parameters correctly very early and acquire basic properties of the target grammar in an almost flawless fashion (see, e.g., Wexler 1999), which has led some generative linguists to assume that language change cannot be explained in terms of L1 acquisition (see, e.g., Weerman 2008). However, see Yang (2010) for critical discussion and Cournane (2014) on lexical mapping errors in the acquisition of English modals which suggest that learners do play an active role in processes of language change (further reports on delayed/non-targetlike acquisition of morphosyntactic phenomena include, e.g., Fritzenschaft et al. 1990 on verb placement in German; Håkansson, and Dooley Collberg 1994 and Waldmann 2014 on non-targetlike embedded

V-Neg patterns in Swedish; and Bohnacker 2004 on determiner use in Swedish); see again chapters 11 and 12.

(<sup>9</sup>) The notion of ‘cue-based acquisition’ is developed in Lightfoot (1999) based on earlier proposals developed by Dresher and Kaye (1990) and Dresher (1999) to model the acquisition of phonological properties. The basic assumption is that UG contains not only a set of parameters, but also specifies for each parameter a *cue* that serves to switch the parameter one way or the other (see Fodor 1998a for a related approach; and chapter 11 for discussion). If the learner detects a cue that is attested robustly, this will activate a given parameter or syntactic operation in the learner’s grammar. Language change results either if a given linguistic feature fails to be cued or if it starts to be cued, in contrast to the target grammar.

(<sup>10</sup>) Note that this raises a number of further questions, in particular concerning the way the language acquisition device (LAD) converts information conveyed by the PLD into a grammar *G* with a set of properties  $\{P_1 \dots P_n\}$ , which cannot be addressed here in detail (but see recent work by Charles Yang and the discussion in chapter 11).

(<sup>11</sup>) For contact-induced change, see Thomason and Kaufman (1988) and the bulk of work put together by Anthony Kroch and his collaborators on changes in the history of English (see, e.g., Kroch and Taylor 1997, 2000; Pintzuk 1999; Kroch, Taylor, and Ringe 2000; Yang 2000; Kroch 2001; Trips 2002; see also chapter 17 for a discussion of this kind of change in relation to the diachrony of Haitian and other Creoles).

(<sup>12</sup>) Yang (2002, 2010, 2015) develops a formal model of the relationship between language acquisition and language change (the so-called ‘variational learning model’) that pays attention to the frequencies with which linguistic expressions are attested in the input data. Roughly put, Yang shows that there is a close connection between relevant frequencies and the speed (and robustness) with which a certain grammatical property is acquired. In Yang’s model, the dynamics of learning are formalized in a way that resembles the mathematical treatment of the dynamics of selection in an evolutionary system (see also Clark and Roberts 1993). From this perspective, L1 acquisition is based on a process of selection that successively reduces the number of grammatical hypotheses which are not compatible with the linguistic environment. According to Yang (2010), this acquisition strategy is sensitive to the traditional distinction between core and periphery: selecting properties of the ‘core’ linguistic system from a narrow range of options (the set of parameter values provided by UG) ‘is sensitive to *token* frequencies of specific linguistic data that are quite far removed from surface level patterns’ (Yang 2010 :20). For example, Yang argues that the option of topic drop in early child English is gradually driven out of the grammar by the small number (1.2%) of expletive subjects in child-directed speech. In contrast, the acquisition of language specific generalizations such as productive phonological rules (e.g., umlaut, glottalization, final devoicing, etc.) and morphological properties (e.g., plural, past tense, agreement, etc.) seems to be sensitive to *type* frequency (e.g., the number of English verbs that form the past tense with a suffixed /-d/). In addition, the acquisition of the periphery is taken to be

constrained by third factors (Chomsky 2005) such as the Elsewhere Condition (see also section 18.3.1). A question of some interest which is not addressed by Yang concerns the possible relationship between core and peripheral properties (e.g., inflection) in acquisition and change. See Heycock and Wallenberg (2013) for an application of Yang's learning model to the loss of V-to-T movement in the Scandinavian languages. See chapter 11 for further discussion of Yang's proposals and related matters.

(<sup>13</sup>) Reanalysis is an instance of what Andersen (1973) calls 'abductive change.' The concept of abductive change goes back to the notion of abduction, a method of logical inference introduced by Charles Sanders Peirce, where an inquirer makes (learned) guesses about possible explanations for a set of (seemingly connected) empirical facts. Thus, abductive change occurs when learners entertain a hypothesis concerning the grammar underlying the linguistic evidence they encounter that deviates from the actual target grammar. Note that approaches to change based solely on abduction (e.g., without a theory of what counts as robust evidence in the PLD) do not provide a satisfying answer to the actuation problem. This has to do with the fact that abduction is all about *facilitating* change (by introducing an inherent instability into the process of language transmission), but it does not have much to offer when it comes to *explaining* change. In particular, it does not address the question of *why* a given change takes place at a certain point in the historical development of a language, which is central to the actuation problem. See Kroch (2001) and Roberts (2007:123-124) for discussion.

(<sup>14</sup>) Fischer (2010) puts forward a differing proposal according to which word order change may set off grammaticalization processes (e.g., the rise of periphrastic constructions), which are then followed by the loss of inflectional morphology.

(<sup>15</sup>) See section 18.5 below on the reverse change, that is, ways in which inflectional morphology arises historically via processes of grammaticalization.

(<sup>16</sup>) Under the assumption (Kayne 1994) that OV order is derived by leftward movement of the verb's complements (and other VP-internal material) for licensing purposes, the loss of surface OV orders can be connected to independently motivated lexical or morphological changes (e.g., the erosion of the formerly rich system of inflections in the ME period, see, e.g., Roberts 1997, 2007; van der Wurff 1997, 1999; Fischer et al. 2000; Ingham 2002; Biberauer and Roberts 2005, 2006, 2008; see Hróarsdóttir 1996, 2000 on the loss of OV orders in the history of Icelandic).

(<sup>17</sup>) A different stance is taken by Simpson (2004), who argues that movement operations are never lost from the grammar. Rather, EPP features are always available for the language learner as a formal means to cope with movement operations encountered in the input where the original semantic/pragmatic trigger has been lost in the course of time (e.g., due to overuse of a certain construction). As a consequence, movement operations are not lost if the original trigger disappears, but rather they are converted

into fossilized, purely syntactic movement (see Fuß 2008 for a relevant analysis of the development of generalized V2 in the history of German).

(<sup>18</sup>) Though it might be that the link between, for instance, morphophonological changes and syntactic change is less direct than is often assumed. For example, it is conceivable that speakers tend to avoid ambiguity in their speech production after the loss of case distinctions by using word order as a means to distinguish between different arguments of the verb. This in turn leads to shifts in the PLD which at some point may give rise to parametric change (cf. Kroch 2001; see Bobaljik 2002, Alexiadou and Fanselow 2002, and Hróarsdóttir et al. 2006 for a critical view on the supposed link between rich verbal inflection and verb movement; but see Koenen and Zeijlstra 2014 for a recent defense of the claim that V-to-I movement is conditioned by rich verbal agreement).

(<sup>19</sup>) For example, see Ohala (1983) and Solé (2009) on cases of spontaneous nasalization which can be attributed to a tendency for hearers/learners to posit nasals (absent in the target grammar) in front of (final) voiced stops (see Ohala and Busà 1995 for an inverse change in which listeners fail to detect nasals before voiceless fricatives).

(<sup>20</sup>) Another factor (in addition to discontinuous transmission) that contributes to the swiftness of language change is that parents may pass on to their offspring linguistic features that they have acquired during their lifetime (i.e., the historical development of language(s) exhibits Lamarckian properties, for discussion see, e.g., Nettle 1999; Mufwene 2002; Andersen 2006).

(<sup>21</sup>) In the history of many languages, we can observe a change from basic OV to VO order (see, e.g., van Kemenade 1987, Lightfoot 1991, Pintzuk 1999 on English; Delsing 2000 on Swedish; Rögnvaldsson 1996, Hróarsdóttir 2000 on Icelandic; Grewendorf and Poletto 2005 on German language isolates in Northern Italy; Gerritsen 1984 and Kiparsky 1996 on the Germanic languages in general; Taylor 1994 on Ancient Greek; Devine and Stephens 2006 on Latin/Vulgar Latin/Romance; Rinke 2007 on Portuguese; Williamson 1986, Gensler 1994 on the Niger-Congo languages; Kiparsky 1996 on the Finno-Ugric languages [Finnish and Estonian, in particular]). In contrast, the reverse change from VO to OV seems to be rather rare, and relevant examples cited in the literature can mostly be attributed to language contact (see, e.g., Lehmann 1978 on Sinhalese [under contact with Dravidian languages]; Leslau 1945 and Biberauer, Newton, and Sheehan 2009 on South/Ethiopian Semitic languages such as Tigre and Tigrinya [under contact with Cushitic]; Givón 2001 on Akkadian [under contact with Sumerian]; Ratcliffe 2005 on Bukhara Arabic [under contact with Tajik and Uzbek]; Ross 2001 on Oceanic languages of North-West Melanesia [under contact with Papuan languages]; rare examples of apparent endogenous change from VO to OV include Georgian [Harris 2000], and the Mande languages [Claudi 1994]). As noted by Newmeyer (2000), the apparent rarity of VO-to-OV leads to an interesting paradox, given that the majority of the world's languages are SOV.

(<sup>22</sup>) Biberauer, Newton, and Sheehan (2009) argue that the set of possible diachronic pathways is also shaped by general properties of phrase structure. More specifically, they explore diachronic consequences of Holmberg's (2000:124) Final-Over-Final Constraint (FOFC), which rules out configurations where a head-initial projection is embedded under a head-final projection. For example, FOFC predicts (correctly, it seems, see, e.g., Kiparsky 1996) that the often-observed change from OV to VO must proceed in a 'top-down' fashion (i.e., a change in headedness must first affect higher functional projections such as TP/IP before it can affect the basic order in the VP). In contrast, the (much rarer) change from VO to OV is expected to proceed in a 'bottom-up' fashion, first affecting VP, before it can reverse the headedness of higher functional categories.

(<sup>23</sup>) Alternatively one might entertain a weaker position, claiming that principles of economy actually do not trigger changes, but merely restrict the set of possible changes. Note that from this point of view, economy considerations do not provide an answer to the actuation problem, but are relevant for the so-called 'constraints problem' concerning 'the set of possible changes and possible conditions for changes which can take place in a structure of a given type' (Weinreich et al. 1968:101).

(<sup>24</sup>) A related idea lies behind the Subset Principle (Berwick 1985:37; Manzini and Wexler 1987:61), which ensures that if there are two possible values A and B for a given parameter, and A generates a subset of the sentences generated by B, the learner will acquire the more restrictive setting A in the absence of decisive evidence for setting B. In this sense, A can be viewed as the unmarked setting (see Biberauer and Roberts 2009 who invoke the Subset Principle to account for the loss of optional word order patterns in the history of English). Note that an approach in terms of default settings is not always compatible with the Subset Principle. For example, Hyams (1986) claims that the default setting for the pro-drop parameter is [+pro-drop], which is clearly in conflict with the Subset Principle, since '[+pro-drop] permits a set of grammatical sentences that includes the set generated by the setting [-pro-drop]' (see O'Grady 1997 for discussion; and chapter 11).

(<sup>25</sup>) An interesting question that comes up here concerns the mechanisms that underlie the traditional observation (see, e.g., Jespersen 1917) that language change (and grammaticalization, in particular) often proceeds in a cyclic fashion. Thus, the reduction (and loss) of linguistic forms by phonological erosion and analogical leveling is often compensated for by grammaticalization processes that create new phonological exponents, which are subsequently again subject to erosion (for discussion see, e.g., Jäger 2008; and the contributions in van Gelderen 2009 and Willis et al. 2013).

(<sup>26</sup>) It is generally assumed that grammar competition is triggered in cases where the PLD contains robust evidence for conflicting parametric choices that cannot be part of a single grammar (see, e.g., Kroch 1994; Lightfoot 1999:92). For example, Haeberli (2004) argues that grammar competition is triggered by mismatches between morphology and syntax (e.g., a grammar with syntactic evidence for verb movement but without rich verbal inflection), giving rise to a second internalized grammar where the mismatch is

resolved. Over time, the more harmonic or economic variant may then eliminate its competitor. However, as pointed out by Hale (1998, 2007), this scenario (and the idea of grammar competition in general) seems to be at odds with the fact that speakers confronted with incompatible triggers/cues in the PLD either ignore a subset of the input data (see Anderson 1980, Bobaljik 2002, and Koenenman and Zeijlstra 2014 on relevant examples and the relative significance of syntactic and morphological triggers/cues) or acquire two fully distinct grammars, the use of which is typically linked to different registers (which arguably cannot be used to account for examples like (7)).

(<sup>27</sup>) Kroch (1994) argues that blocking effects imposed by UG restrict the co-existence of grammars that differ only minimally with respect to a set of parameter doublets (i.e., coexisting competing values for one parameter), thereby guaranteeing that one grammar will eventually win out over its competitors (see Wallenberg and Fruehwald 2013 for further discussion and the claim that the notion of competing grammars can also be used to account for [stable] syntactic optionality). However, as pointed out by Hale (2007:173), it is not entirely clear how morphological blocking, which is usually considered a grammar-internal process, can effect the loss of parametric variants that are part of different grammars. This seems to suggest that it is actually more appropriate to assume that grammar competition involves competing parametric options within a single— instead of multiple—grammars.

(<sup>28</sup>) But see Janda and Joseph (2003:140-141) for a critique of the CRE which is based on the conviction

that the order in which changes appear in written language need not reflect the order in which they first appeared in colloquial speech. In particular, we believe that novel patterns which arise individually in spoken language may cumulate for a long period of time before they jointly achieve a breakthrough, as a set, into writing.

(<sup>29</sup>) More recently, the research program initiated by Anthony Kroch has sparked a number of new studies that combine a generative approach to language change with advanced corpus linguistic and statistical methods (see, e.g., Wallenberg 2009; Fruehwald et al. 2013; Wallenberg and Fruehwald 2013; Bacovcin 2013; and Ecay 2014, 2015).

(<sup>30</sup>) The notion of 'lexical diffusion' is commonly used to refer to the idea that changes may spread through the lexicon gradually, affecting one lexical item at a time.

(<sup>31</sup>) See Guardiano and Longobardi (2005), Longobardi and Guardiano (2009), and Longobardi et al. (2013) for a new method of measuring historical relatedness and genetic affiliation of languages based on a fine-grained system of microparameters (relating to the syntax of noun phrases/DPs; this approach is presented in detail in chapter 16). See also Walkden (2014) for a generative approach to syntactic reconstruction.

(<sup>32</sup>) From the 1980s on, the study of grammaticalization processes has become a main focus of descriptive diachronic/typological linguistics, leading to a wealth of new data and a set of generalizations on the course and defining properties of grammaticalization (see,

e.g., Heine and Kuteva 2002; Lehmann 2002; Hopper and Traugott 2003; Heine 2003; and Narrog and Heine 2011. For a generative perspective on relevant phenomena see, e.g., Roberts 1993b; von Stechow 1995; Newmeyer 1998; Roberts and Roussou 1999, 2003; and van Gelderen 2004, 2011).

(<sup>33</sup>) In other words, it is claimed that grammaticalization is a unidirectional process. However, it has been pointed out that there exist exceptions to the purported irreversibility of grammaticalization processes, involving, e.g., changes where former inflections (such as case affixes) turn into clitics (see Campbell 1991; Newmeyer 1998; Allen 2003; Willis 2007; Norde 2009; and most recently Jung and Migdalski 2015).

(<sup>34</sup>) In some cases, the reanalysis may also affect functional elements that turn into other functional elements, as in the case of clitics (presumably realizations of D<sup>0</sup>) that turn into agreement markers (see Roberts and Roussou 2003; Fuß 2005).

(<sup>35</sup>) Note that changes in which elements occupying the specifier of a phrase (usually via movement) are reanalyzed as the head of that phrase are quite frequent (demonstratives > determiners, relative pronouns > complementizers, negative adverbs > sentential negation, etc.; see van Gelderen 2004 for detailed discussion).

(<sup>36</sup>) Fuß (2005) argues that grammaticalization phenomena are in addition shaped by morphological third factors (i.e., blocking effects induced by some form of the Elsewhere Condition) which restrict grammaticalization to cases where the newly coined exponents signal more featural distinctions than competing older forms.

(<sup>37</sup>) Note, moreover, that it is generally assumed that many functional heads (such as T in finite clauses without an auxiliary English or D in languages without overt determiners such as Tagalog) are phonologically empty which is the endpoint of phonological deficiency.

(<sup>38</sup>) The existence of free tense, aspect, and mood markers (as in many Creole languages for example; see chapter 17) indicates that the development of bound forms is not a necessary consequence of the reanalysis of lexical material as exponents of functional heads, but rather a tendency, caused by the phonological deficiency of functional categories.

(<sup>39</sup>) According to von Stechow (1995:185), the quantifying determiner *many* developed from a former adjective (OE *manig*) which originally combined with plural nouns to “[identify] those sets/groups that have many members” (examples like *the many Englishes spoken in the world today* show that *many* can still be used as an adjectival element). Cases where ‘many’+NP could be associated with existential readings (probably involving a phonologically empty determiner) eventually gave rise to a reanalysis in which *many* turned into a new phonological realization of D<sup>0</sup> which combined existential force with the selectional requirements (plural) of the adjective ‘many.’

(<sup>40</sup>) Note that on these assumptions, the category 'agreement' seemingly does not to qualify as a separate functional category in the same sense as D, T or C, since it is not quantificational (see Chomsky 1995b for a related conclusion).

(<sup>41</sup>) For example, Jakobson (1931) demonstrated that the (abstract) structuralist concept of the phoneme makes available an approach to sound change and linguistic reconstruction which is empirically and conceptually superior to the purely 'phonetic' version of the comparative method developed by the Neogrammarians.

(<sup>42</sup>) The work on sound change in (Northeastern) American English conducted by William Labov and his colleagues (see Labov et al. 2013 for a recent overview) is an impressive demonstration that the study of change in progress offers a level of detail unmatched by research based on purely historical data.