

## 12 Grammaticalization from a biolinguistic perspective

---

ELLY VAN GELDEREN

### 12.1 Introduction

Estimates about the origin of modern human language range from 50,000 to 150,000 years ago. These estimates are based on archeological findings, the presence of tools and beads in e.g. the Blombos cave at 70,000 years ago, and mutations in a gene connected to speech (FOXP2) at about 120,000 years ago. Genetics and archeology work well together and suggest a homeland for modern humans in Africa. What can linguistics contribute to this picture? In this paper, I show that a biolinguistic approach (e.g. Chomsky 2005b, 2007) has much to offer. In this introduction, I first briefly mention some other linguistic approaches that geneticists and others have often turned to, namely genetic and areal linguistics. In the remainder of the paper, I argue for a biolinguistic approach.

Genetic linguistics provides insights into linguistic relationships, and areal linguistics can show which features are typical for the areas where language is supposed to have started. Genetic groupings such as the four families in Africa and the three in the Americas are much contested (see the criticism Greenberg received throughout his life). Areal linguistics (e.g. Nichols 1992; Haspelmath et al. 2006) shows us more about recent trends than about original features. For instance, since Dryer's (1999) maps of Object—Verb and Verb—Object distributions show both orders in Central Africa and we know that these orders can change quickly, this approach cannot be used to give us insight into an earlier state.

Thanks to audiences in Oslo, Stellenbosch, and Montreal and especially to Harry Bracken, Terje Lohndal, and Eric Reuland for very useful discussion and comments on an earlier version.

In this paper, I will therefore examine what (historical) syntax has to say when couched within a biolinguistic framework. Hauser, Chomsky, and Fitch (2002) argue that recursion sets human language apart from animal communication and Chomsky (2005b: 11) specifies this further by saying that Merge, linking two elements, was the “‘Great Leap Forward’ in the evolution of humans.” Likewise, Piattelli-Palmarini and Uriagereka (2005) emphasize the role of recursion and Merge. Some principles follow for free from Merge and some from general cognitive principles. The emergence of a binary merge brings with it certain relations, such as heads and elements merged with heads into phrases, and c-command, as will be explained below. Heads and phrases, merged in binary fashion, in turn define argument structure or thematic structure. Organizing the thematic layer through Merge is one aspect relevant to the evolution of language. I will argue that grammaticalization was the other step responsible for markings in the grammatical layer. Typical grammaticalizations are prepositions starting to function as case markers, verbs as auxiliaries and affixes, and pronouns as agreement morphemes. These changes are cyclical and continue to occur in contemporary languages. These changes, I argue, can be seen in terms of cognitive economy of the syntactic derivation, e.g. semantically “lighter” elements are preferred over “heavier” ones. The evidence for these claims, and also for the nature of early language, comes from observable linguistic changes.

The outline is as follows. In section 12.2, I present a very general picture of the Minimalist Program, and in particular its biolinguistic focus. This framework is elaborated on in section 12.3, especially where the operation Merge is concerned. In sections 12.4 and 12.5, grammaticalization is the focus. I discuss how it follows from economy and how it is relevant to language evolution. Section 12.6 is a conclusion.

## 12.2 A Minimalist sketch of the language organ

Starting in the 1950s, Chomsky and the generative model he develops present an alternative to then current behaviorist and structuralist frameworks. Chomsky focuses not on the structures present in the language/outside world but on the mind of a language learner/user. The input to language learning is seen as poor (the “poverty of the stimulus” argument) since speakers know so much more than what they have evidence for

in their input. How do we know so much on the basis of such impoverished data? The answer to this problem, “Plato’s Problem” in Chomsky (1986), is universal grammar (henceforth UG), the initial state of the faculty of language, a biologically innate organ. UG helps the learner make sense of the data and build up an internal grammar.

Initially, many principles were attributed to UG but currently (e.g. Chomsky 2004, 2005b, 2007), there is an emphasis on principles not specific to the faculty of language, i.e. UG, but to “general properties of organic systems” (Chomsky 2004: 105), also called “third factor principles” in Chomsky (2005b). Merge is one such operation that can be seen as a UG principle but also as one possibly “appropriated from other systems” (Chomsky 2007: 7) and relevant to other systems. In this paper, I argue that economy principles are such general cognitive principles.

The biolinguistic angle also makes us ask (a) what the structure of the language organ is and (b) why it is this way. The answer to the first question may be that it uses Merge or recursion and the answer to the second question would lie in the relation between the Narrow Syntax and demands of other organs, e.g. the sensory motor one. Thus, there is a Narrow Syntax (with Merge) and mappings to two interfaces, the sensory-motor interface, PHON, and the conceptual-intentional one, SEM. As mentioned, Chomsky has suggested that some rewiring of the brain, a small mutation or the result of one, brought about Merge. Merge, linking two elements, was the “great leap forward” in the evolution of humans. “The individual so endowed [with Merge] would have the ability to think, plan, interpret, and so on.” Then, “[a]t some stage modes of externalization were contrived” (Chomsky 2007: 14). Phonology and morphology are involved in the externalization and are highly varied since there are no universal principles involved, unlike with Merge.

Work on animal communication has shown that animals use symbols. Bickerton (1990, 2000) has argued that animal communication probably uses thematic structure, i.e. SEM, but no recursion of structures, i.e. Merge. We know that some animals have an impressive set of sounds, so PHON, but not a large vocabulary. Chomsky entertains both the possibility that syntax was “inserted into already existing external systems,” namely the sensory-motor system, PHON, and system of thought, SEM (Chomsky 2002: 108), as well as the one where the externalization develops after Merge (Chomsky 2007: 14). Figure 12.1 shows the three components of human language. I will assume that Merge appears after SEM and PHON

symbolic	
thematic	=SEM
.....	
sounds	
vocabulary	=PHON
.....	
merge/recursion	
grammaticalization	=Narrow Syntax

FIG. 12.1. Three separate systems

are already developed, but for the purpose of this chapter nothing hinges on this.

The vocabulary develops after sounds are abundant enough (see Carstairs-McCarthy 1999). One could see this as a result of Merge as well.

Merge comes in two kinds, Internal and External Merge. Chomsky (2005b: 14) suggests that External Merge is relevant to the argument structure, whereas Internal Merge is relevant for scope and discourse phenomena, as in Figure 12.2. This means External Merge may have been an early feature of language. The longest utterance of Nim, a chimpanzee trained by Terrace in the 1970s, is apparently (1).

- (1) Give orange me give eat orange me eat orange give me eat orange give me you.

This sentence obviously has thematic structure, but this is not expressed in the hierarchical way that human language is. (External) Merge helped organize the thematic structure in human language. In many languages, marking the thematic positions is done through pure Merge (e.g. Chinese, English), but in some languages, inherent Case and adpositions mark thematic roles (e.g. Sanskrit, Latin, Malayalam, Japanese, Tagalog). This special marking has come about through grammaticalization of location and instrument markers to case markers. Definiteness and specificity are the second semantic aspect that needs to be marked. The differences between the two kinds of Merge are listed in Figure 12.2.

Austronesian languages that mark topic show a difference in morphology for the two systems of Merge. As is well known, Tagalog marks its topic through *ang*, as shown in (2). This topic marker is a definiteness

Grammaticalization from a biolinguistic perspective 229

Merge:	External Merge	-	Internal Merge
	=		=
	Theta		Discourse/Quantification
Grammaticalized through:			
	adpositions/inherent Case		definiteness/displacement

FIG. 12.2. The two kinds of Merge

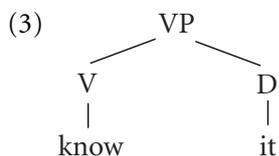
marker as well (technically, only *a-* is and *-ng* is a ligature linking the article to the noun, according to Frawley 1976). The other markers *ng* (for Actor, Patient, and Instrument) and *sa* (for Goal, Source, Location, and Benefactive) mark the theta-roles of the non-topics, and derive from location markers (Finney 1999):

- (2) b-um-ili ng kotse ang lalake Tagalog  
 AF-PF-buy P car TP man  
 “The man bought a car” (Frawley 1976: 106)

This picture suggests that syntax and morphology evolved independently. Some have argued that they are therefore separate systems, e.g. Jackendoff (2002: 260). Bobaljik (2006) too has agreement adding features after Narrow Syntax. I will make plausible that External merge emerged first, followed by Internal Merge. Grammaticalization affected both, however.

12.3 How does Merge work?

I will now turn to an abstract model of how a sentence is produced using a Minimalist approach. There is a lexicon from which lexical items are selected, after which Merge combines two items, e.g. *know* and *it* in (3), and one of the two heads projects, in this case V, to a higher VP:



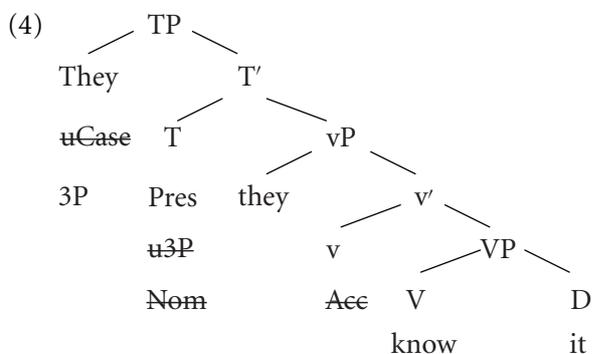
The items that merge are not arbitrarily selected. The head is searching to satisfy a feature, e.g. a thematic one in (3). The head is also the one projecting up, and this projection can then in its turn be selected. The VP domain is the thematic layer, i.e. where the argument structure is determined.

Apart from Merge, there are “atomic elements, lexical items LI, each a structured array of properties (*features*)” (Chomsky 2007: 6). Each language learner selects the features compatible with the input. Thus, the features are parameterized; not the syntax. Features come in two kinds. Interpretable ones include number on nouns and are interpreted by the conceptual-intentional interface, i.e. are relevant to the meaning of a sentence. Uninterpretable features are grammatical in nature, and include agreement features (also known as phi-features) on verbs and Case features on nouns. These uninterpretable features need to be valued and deleted since they are not relevant to the meaning. In Figure 12.1, the SEM level would have only interpretable features, whereas Narrow Syntax would work with both.

I will suggest that the uninterpretable features were not present during the stage at which Merge appeared originally. They were added through language change prompted by economy principles, grammaticalization, and ended up incorporating distinctions having to do with specificity, topicality, and quantification. Continuing the derivation in (3) will hopefully make the function of features clearer.

A sentence consists of three layers: a VP with basic information about the arguments; a TP where tense, mood, and aspect are expressed; and a CP where the sentence type is indicated, e.g. a question or assertion. In a language such as Modern English, T has interpretable tense features but uninterpretable phi-features. It probes (“looks down the tree”) for a nominal to agree with. It finds this nominal (“goal”) in *they* and each element values its uninterpretable features: the noun’s Case as nominative and the verb’s phi-features as third person plural. The final structure will look like (4), where the features that are not “struck through” are interpretable. The subject moves to Spec TP for language-specific reasons:<sup>1</sup>

<sup>1</sup> The derivation in (4) uses early lexical insertion, i.e. a lexicalist approach, as in Chomsky (1995, 2004). For the purposes of this paper, nothing hinges on this. Note that Merge is neutral as to where lexical insertion takes place; I will add lexical items in the tree for convenience.



The reason the effects of Merge are so far-reaching is that structures made by Merge involve heads, complements (or first-merged), and specifiers (or second-merged). Merge, thus, automatically brings with it the following structural relations.

A lot can of course be said about each of these. For instance, it has been argued that all languages are right-branching as in (i) in Figure 12.3. This would mean there are no headedness parameters. Pidgins and creoles are typically SVO, however, i.e. (bi), and this may also be the proto-order, though e.g. Newmeyer (2000) argues that the protolanguage was SOV, i.e. (ii).

Turning to language evolution, languages closer to the protolanguage will have Merge but there is no reason they would have Move (Internal Merge) and Agree/Probing as in (4) (though Newmeyer 2000: 385, n 4 suggests that protolanguages may have been inflectional). My focus on grammaticalization as a process responsible for morphology assumes that agreement and Case arise later (see Reuland, this volume, for another view).

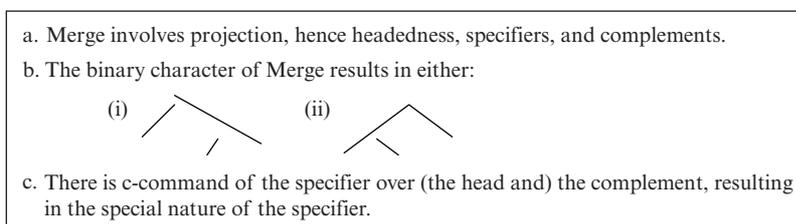


FIG. 12.3. Relations connected with Merge

So, the first step in the evolution of syntax is Merge. It brings with it notions of headedness (once you merge two elements, one determines the resulting label) and structural hierarchy. These notions also determine possible argument structures. The next step is for grammatical heads, such as auxiliaries and prepositions, to appear, as we will discuss in section 12.4.

### 12.4 Grammaticalization

As is well known, grammaticalization is a process whereby lexical items lose phonological weight and semantic specificity and gain grammatical functions.

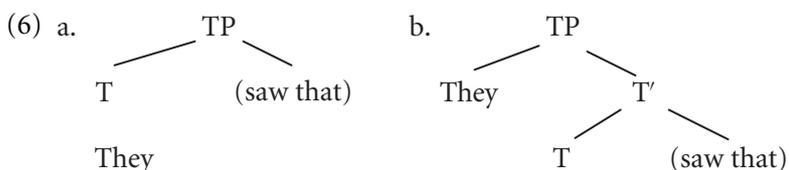
Grammaticalization has frequently been investigated in a functionalist framework. Recently, however, structural accounts have started to appear (e.g. Abraham 1993; Roberts and Roussou 2003; van Gelderen 2004) accounting for the cyclicity of the changes involved. Van Gelderen, for instance, uses economy principles that help the learner acquire a grammar that is more economical, and as a side effect more grammaticalized.

Two economy principles, provided as (5) and (8) below, are formulated in van Gelderen (2004). They are part of the cognitive system and help learners construct a grammar. Principle (5) is at work in the internalized grammar and holds for Merge (projection) as well as Move (checking). It is most likely not a principle specific to language but a property of organic systems, such as if you want to carry an object, you don't also carry the table on which the object is situated:

(5) **Head Preference Principle (HPP)**

Be a head, rather than a phrase.

This means that a speaker will prefer to build structures such as (6a) rather than (6b). The pronoun is merged in the head position in (6a), and in the specifier position in (6b). Specifier positions can accommodate entire phrases but require an additional merge:



*Grammaticalization from a biolinguistic perspective* 233

The speaker will only use (6b) for structures where a phrase is necessary, e.g. coordinates such as *you and I*. In some languages, there are prescriptive rules stopping this change (as there are in French, see Lambrecht 1981).

Under a Minimalist view of change, syntax is inert and doesn't change; it is the lexical items that are reanalyzed.<sup>2</sup> Pronouns are reanalyzed from emphatic full phrases to clitic pronouns to agreement markers, and negatives from full DPs to negative adverb phrases to heads. This change is slow, however, since a child learning the language will continue to have input of, for instance, a pronoun as both a phrase and a head. Lightfoot (1999) develops an approach as to how much input a child needs before it resets a parameter. In the case of pronouns changing to agreement markers, there will have to be a large input of structures that provide evidence to the child that the full phrase is no longer analyzed as that. This is already the case in French since the pronoun is always adjacent to the finite verb in spoken French. The child, therefore, always produces the pronoun in that position, even though regular subjects can precede or follow the verb (see Pierce 1992). However, rather than blame the change on a changed input (a cue), I blame it on factors internal to language and cognition.

Another instance of the HPP is the well-known fact that native speakers of English (and other languages) producing relative clauses prefer to use the head of the CP (the complementizer *that*) rather than the specifier (the relative pronoun *who*) by a ratio of 9:1 in speech. In addition, speakers strand prepositions in speech. For instance, children acquiring their language obey this same economy principle. Thus, according to Diessel (2004), young children produce only stranded constructions in English, as in (7a), using the head *that*, and not the version where the entire Preposition Phrase is fronted, as in (7b):

- (7) a. those little things **that** you play **with**  
       (Adam 4:10, from Diessel 2004: 137)  
       b. those little things **with which** you play

Once they become (young) adults, they are taught to take the preposition along and to disobey the HPP.

The Head Preference Principle is relevant to a number of historical changes: Whenever possible, a word is seen as a head rather than a phrase. Examples of changes predicted by the HPP are given in Table 12.1.

<sup>2</sup> The word "reanalysis" is used to emphasize that the language changes; of course the child analyzes.

TABLE 12.1. *Examples of reanalysis due to the Head Preference Principle*

Relative pronoun <i>that</i> to complementizer	Demonstrative to article
Negative adverb to negation marker	Adverb to aspect marker
Adverb to complementizer	Full pronoun to agreement

Within Minimalism, there is a second economy principle that is relevant to grammaticalization. Combining lexical items to construct a sentence, i.e. merging them, “comes ‘free’ in that it is required in some form for any recursive system” (Chomsky 2004: 108) and is “inescapable” (Chomsky 1995: 316, 378). Initially, a distinction was made between Merge and Move and it was less economical to merge early and then move than to wait as long as possible before merging. In van Gelderen (2004), this is formulated as in (8):

(8) **Late Merge Principle (LMP)**

Merge as late as possible

In later Minimalism, Merge is reformulated as External Merge and Move as Internal Merge, with no distinction in status. One could argue that (8) is still valid since the special Merge, i.e. Internal Merge, requires steps additional to those that Merge, i.e. External Merge, requires. The extra step is the inclusion in the numeration of copies in the case of Internal Merge. Traces are not allowed, since they would introduce new material into the derivation after the initial selection, and therefore copies of elements to be moved have to be included in the lexical selection. Move/Internal Merge is not just Move but “Copy, Merge, and Delete.” Since the numeration has to contain more copies of the lexical item to be internally merged, and since those copies have to be deleted in the case of traditional Move, (8) could still hold as an economy principle.

As mentioned, Chomsky (2005b: 14) suggests a real difference between the two kinds: External Merge is relevant to the argument structure, whereas Internal Merge is relevant for scope and discourse phenomena. This indicates a crucial difference between the two kinds of operations, and nicely expresses the intuition that External Merge is the Merge that made possible “the great leap” and that Internal Merge was brought about by the workings of the cognitive system. In section 12.6, I provide a way of reformulating the LMP: If the set of lexical items taken from the

*Grammaticalization from a biolinguistic perspective* 235

lexicon contains elements whose properties are less relevant to the thematic structure (External Merge), they can be used with fewer thematic, i.e. interpretable, features.

The Late Merge Principle works most clearly in the case of heads. Their grammaticalization path is always to higher functional categories. The history of *after* presents an interesting example. The preposition and adverb in Old English, according to the OED, indicate place (or order) or time, as in (9), and manner, where *according to* would be the modern equivalent:

- (9) Fand þa ðær inn æþelinga gedriht swefan æfter  
found then there in noble company sleeping after  
**symble**  
feast  
“He found therein a company of nobles sleeping after their feast”.

(*Beowulf* 118-119)<sup>3</sup>

Of the forty-one instances of *after* in the relatively early *Beowulf*, only one occurs inside a fronted PP and in the *Christ* from the *Exeter Book*, again an early text, there is none out of fifteen. The objects of these prepositions are full nouns or personal pronouns. None of these introduces a subordinate sentence.

This changes in later texts in that the PPs are fronted more often and the object of *after* is a demonstrative. The version of the *Anglo-Saxon Chronicle* known as *Chronicle A* contains entries that up to 891 are copied by Hand I but after 892 are entered for each year. Before 892, *after* is followed by a noun or pronoun and rarely (8%) by a demonstrative; the PP is preposed in 27% of the cases. In the later parts (i.e. entered after 892), many of the objects of *after* are demonstratives, as in (10), namely 17 out of 22 (= 77%). That the demonstrative preference starts so clearly with the entries of 892 might indicate that the language is closer to that which might have been spoken around 900:

- (10) Her forðferde Wulfstan diacon...7 æfter þon forðferde Gyric  
mæsse preost.  
“In this year died Wulfstan...and after that died Gyric the priest”.

(*Chronicle A*, entry for the year 963)

<sup>3</sup> References to *Beowulf* are from the Klaeber edition, to the *Anglo-Saxon Chronicle* from Thorpe, to the Old English Gospel from Skeat, and to the *Paston Letters* from Davis. Others are taken from the OED.

TABLE 12.2. Percentages of demonstrative (*Dem*) objects with after and fronting of the PP

	Beowulf	< 892 Chron A	> 892 Chron A
Dem object	0	2/26 = 8%	17/22 = 77%
Fronting	1/41 = 2%	7/26 = 27%	12/22 = 55%

The use of the demonstrative object in (10) indicates that the PP *æfter þon* is starting to be analyzed as an adverb linking the sentence to another. This is confirmed by the frequent fronting of the PP (12 out of 22 = 55%), as also shown by (10). The fronting can be seen as a consequence of Late Merge. In Table 12.2, the differences for *Beowulf*, the early part of *Chronicle A*, and the later part are summarized, with percentages rounded off.

The first instances of conjunctive use of *after* that the OED mentions involve sentences such as (11) to (13), with the dates given as in the OED. These are not conjunctions, i.e. heads, but full phrases (PPs) in the specifier of the CP. They indicate time so could be derived from a structure such as (14) and are different from (10) in that a complementizer follows the PP:

- (11) Witodlice **æfter þam þe ic of deaþe arise** ic cume to eow on galilee  
Surely after that that I of death arise I come to you in Galilee

(c. 1000 *West Saxon Matthew* 26. 32 Hatton Ms)

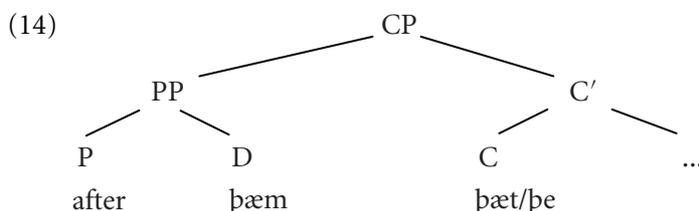
- (12) **Efter þan þet þe mon bið dead**  
After that that the man is dead

(c. 1175 *Lamb. Hom.* 51)

- (13) **Aftterr þatt tatt he wass dæd**  
After that that he was dead  
“After he was dead”

(c. 1200 *Ormulum* 7667)

The tree for (11) to (13) would be as in (14):



Interestingly, the much earlier Lindisfarne glosses render the relevant part of (11) as (15), without the complementizer. The complementizer-less stage represents an earlier variety. This is confirmed by data in Rissanen (2006), who examines the Helsinki Corpus Old English parts and finds an increase in complementizer following the PP:

- (15) *æfter ðon* uutedlice ic eft-ariso ic forlioro I iowih in galileam  
 “after that surely I arise-again I come before you in Galilee”  
 (Lindisfarne Matthew 26. 32)

The development explained so far has been that the PP with *after* gets fronted and that its object increasingly often is a demonstrative, not a full noun. The demonstratives are still inflected and cannot be “mistaken” for complementizers. This means the PP is still adverbial. The second stage is for a complementizer to follow the PP. The third stage is for the demonstrative to disappear and then for the preposition to be reanalyzed as a complementizer, as in (16) and (17):

- (16) *Aftir he hadde take þe hooli Goost*  
 (c. 1360 Wyclif *De Dot. Eccl.* 22)

- (17) *After thei han slayn them*  
 (1366 Mandeville 174)

The changes are indicated in (18):

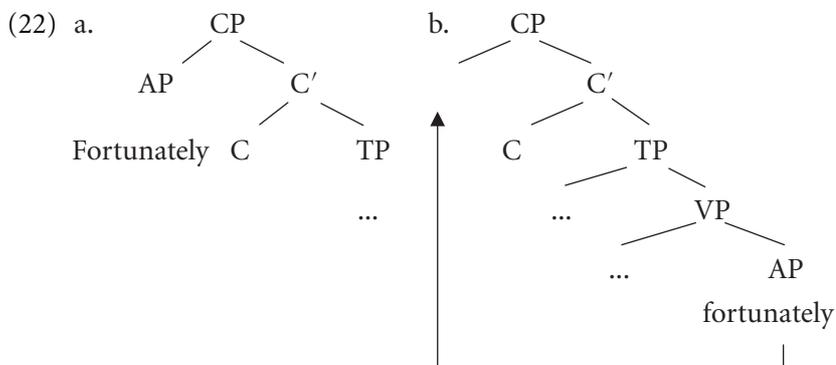
- (18) a. PP    ~~PP~~ 900 (Chronicle A) – present  
 b. PP C    1000 (West Saxon Gospel) – 1600  
 c. P C    1220 (Lambeth) – 1600 (OED 1611)  
 d. C    1360 (Wycliff) – present

This accounts for the change from lexical to functional **head** or from functional to higher functional head so frequently described in the grammaticalization literature (e.g. Heine and Kuteva 2002). Late Merge also

accounts for lexical phrases becoming base generated in the functional domain. An example is *fortunately* (which replaces Old English CP-adverbs such as *witodlice* and *soplice*). When it is first introduced into the English language from French in 1386, it is as adjective, as in (19), meaning “happy, successful, favored by fortune.” It then changes to a higher adverb, as in (20) and (21), initially by moving:

- (19) Whan a man . . . clymbeth vp and wexeth **fortunat**  
(OED, 1386, Chaucer)
- (20) **Most fortunately**: he hath atchieu'd a Maid That paragons descrip-  
tion, and wilde Fame: One that excels the quirkes of Blazoning  
pens  
(Shakespeare, *Othello* II. i. 61–63)
- (21) **Fortunately**, Lord De la War . . . met them the day after they had  
sailed  
(OED, 1796)

Structure (22a) shows the more recent structural representation and (22b) the earlier one. The preferred one under the LMP is (22a):



Other examples of the LMP are given in Table 12.3.

How exactly does Late Merge account for language change? If non-thematically marked elements can wait to merge outside the VP (Chomsky 1995: 314–15), they will do so. I will therefore argue that if, for instance, a preposition can be analyzed as having fewer semantic features and is less relevant to the argument structure (e.g. *to*, *after*, and *of* in ModE), it will

*Grammaticalization from a biolinguistic perspective* 239TABLE 12.3. *Examples of reanalysis due to the Late Merge Principle*

<i>for</i> , from P > C	negative objects to negative markers
modals: V > T	<i>to</i> : P > M(ood)

tend to merge higher (in TP or CP) rather than merge early (in VP) and then move. Like the Head Preference Principle, Late Merge is argued to be a motivating force of linguistic change, accounting for the change from specifier to higher specifier and head to higher head. Roberts and Roussou (2003), Wu (2004), and Simpson and Wu (2002) also rely on some version of Late Merge.

Concluding section 12.4, under the LMP as under the HPP, syntax is inert; it is the lexical items that are reanalyzed by the language learner. Two principles, the HPP and the LMP, provide an insight into what speakers do when they construct a sentence. In the next section, I will apply these to a scenario for language evolution.

## 12.5 Grammaticalization and language evolution

As argued above, Merge could have been the first step in creating syntax from a stage that consisted of either words or gestures (e.g. Corballis 2002), and as Traugott (2004: 134) puts it, as “an exaptation of thematic role structure.” The current section provides a scenario for the use of Merge and subsequent steps.

Once External Merge applies, certain structural and thematic relationships crystalize, unlike those in (1), uttered by Nim. Chomsky (2007: 11) talks about edge features as determining what merges externally, and at the VP level this is probably determined by thematic features. Thus a V selects a DP to merge and a (light verb) *v* merges with a VP (in which a DP with a specific theta role occurs). The *v*P, or VP-shell, represents the thematic level, and one that adult native speakers employ when they speak or write in “fragments”, as in (23a). Children first reach this thematic stage, as (23b) shows from Abe, before producing sentences with grammatical categories (though they understand grammatical categories before they produce them):

- (23) a. Gone for lunch  
b. like a cookie

(Abe, 3.7)

In many languages, thematic relations are additionally marked, namely by inherent case (Chomsky 1986: 193), e.g. dative to mark a Goal theta role in (24):

- (24) þæt he sæ-mannum onsacan mihte  
that he sailors-DAT strive-against might  
“that he might strive against the sailors”

(*Beowulf* 2954)

This inherent Case can be argued to be derived through grammaticalization of adpositions (e.g. Tauli 1958).

The next evolutionary step, which is an automatic result of Merge and economy (Narrow Syntax), is when Internal Merge (movement) arises, as well as grammatical elements relevant to specificity, definiteness, and quantification. Tense and aspect are relevant to specificity as well. A language where definiteness is expressed by preposing is Chinese, as well known from the work by Li and Thompson (1978). Cf. (25):

- (25) a. chi le fan Chinese  
eat PF rice  
“I ate some rice”  
b. fan chi le  
rice eat PF  
“I ate the rice”

(Yi Ting Chen p.c.)

Definite time adverbials also precede the verb whereas durative ones follow in Chinese. There are other languages, however, in which such grammatical functions are not purely done through movement but through grammatical elements. They develop when one feature of a lexical element is emphasized over others (hence the slight semantic loss). Similar data exist for other complementizers as well as sentence-adverbs and auxiliaries.

The two principles used above (HPP and LMP) make learners analyze lexical material that is already part of the structure and change the position of it. There are also a number of changes where a new element comes from outside of the sentence, e.g. a special pronoun being incorporated into the

CP to indicate subordination, and an emphatic topic pronoun becoming the subject (in Spec TP). This can be expressed by means of a principle that incorporates (innovative) topics and adverbials in the syntactic tree:

(26) **Specifier Incorporation Principle (SIP)**

When possible, be a specifier rather than an adjunct.

Sometimes, these “renewals” are innovations from inside the language, as in the case of the English negative nominal *na wiht* “no creature” (the root of *not*) to mark negation; but other times, these renewals are borrowed through contact with other languages. One such possible case is the introduction into English of the *wh*-relative. In Old English, there are a number of relative strategies, but by Early Middle English, the complementizers *þat* and *þe* are typical. This is predicted under the HPP since those forms are heads (see van Gelderen 2004: 83–87). By later Middle English, this form is competing with the *wh*-pronoun still present in present-day English (be it mainly in written English). Mustanoja cites Latin influence for the introduction of the *wh*-pronoun. Romaine (1982) shows that the introduction of the *wh*-pronouns was stylistically influenced, and Rydén (1983) shows both Latin and French influence. The first instances of *who* occur in epistolary idioms that are very similar to those in French letters of the same period. For instance, in many of the collections of letters from the fifteenth century, the same English and French formulaic constructions occur, such as in (27a) from Bekynton and (27b) from the *Paston Letters*:

(27) a. a laide de Dieu notre Seigneur, **Qui** vous douit bonne vie et  
longue

“With the help of God, our Lord, who gives us a good and long  
life”

(Bekynton, from Rydén 1983: 131)

b. be the grace of God, **who** haue yow in kepyng

“by the grace of God, who keeps you”

(*Paston Letters* 410)

The *wh*-pronoun is in the specifier position (since it can pied pipe a preposition and is inflected). This shows that, for creative reasons, speakers can start to use the specifier again.

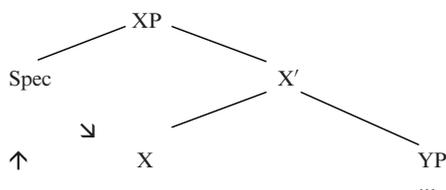


FIG. 12.4. The Linguistic Cycle

How are the three principles mentioned so far responsible for cyclical change? Let's see what happens when we combine the effects of the HPP and the LMP, as in Figure 12.4. The HPP will be responsible for the reanalysis, as a head, of the element in the specifier position; the LMP will ensure that new elements appear in the specifier position or in the head.

This scenario works perfectly for changes where a negative object such as Old English *na wiht* "no creature" becomes a spec(ifier) and subsequently a head *not* of a NegP, and for a locative adverb being reanalyzed as part of the higher CP. The SIP would enable the specifier position to be filled from outside of the clause, e.g. by a pronoun.

Givón (1979) and others have talked about topics that are later reanalyzed as subjects, and call this a shift from the pragmatic to the syntactic. What this means is that speakers tend to use the Phrase Structure rules, rather than loosely adjoined structures. With (25) added, typical changes can therefore be seen as (28):

- (28) a. Head → higher Head → 0 (= LMP)
- b. Adjunct → Spec → Head → 0 (= SIP/LMP and HPP)
- Phrase →

The change in (28a) is the one from lower head (either lexical or grammatical) to higher head, via LMP. The change in (28b) shows that either an adjunct (via SIP) or a lower phrase (via LMP) can be reanalyzed as specifiers, after which the specifier is reanalyzed as head (via HPP).

In this section, I have suggested that the emergence of syntax could have followed the path that current grammaticalization also follows. In particular, Merge brings with it a set of relations and a set of economy principles. These economy principles are responsible for what is traditionally called grammaticalization.

## 12.6 Conclusion and feature economy

I have examined two steps that are required in the evolution from pre-syntactic language to language as we currently know it. The one is Merge and the structural and thematic relations it entails to build a basic lexical layer (the VP). The others are the economy principles that enable learners to choose between different analyses. These two principles result in what is known as grammaticalization and build the non-lexical layers (the TP and CP). Lexical material is also incorporated into the syntax through a third principle, the SIP. This principle allows the speaker to creatively include new material, e.g. as negative reinforcement in special stylistic circumstances.

It is possible to formulate economy in terms of features: The computational load (in the Narrow Syntax) is less when semantic or interpretable features are not included in the derivation. Full phrases have more features (to check) and they are more likely to be interpretable. Apart from the preference for heads, there is also a preference for positions higher in the tree, i.e. merged later in the derivation. For instance, a PP base generated in the VP can come to be used as a sentence connector. These changes too can be accounted for through computational economy: The lower (externally merged) element in the tree has more semantic features whereas the grammatical/functional element has uninterpretable features (uF). Thus, this approach eliminates the “imperfection” of uF.

What does this tell us about the shape of the original language? The emergence of syntax followed the path that current change also follows, i.e. one that children take acquiring the language. Chomsky (2002: 113) sees the semantic component as expressing thematic as well as discourse information. If thematic structure was already present in protolanguage (Bickerton 1990), the evolutionary change of Merge made them linguistic. What was added through grammaticalization is the morphology, the second layer of semantic information.