The Evolution of Language: Grammaticalization and two kinds of Merge

Elly van Gelderen

Arizona State University

ellyvangelderen@asu.edu

29 December 2006

Abstract

It has been argued that the introduction of merge and recursion were crucial for the evolution of human language. Merge (more precisely external merge) made the efficient expression of already existing argument structure possible. In this paper, I make a case that grammaticalization is responsible for the second crucial development, namely the introduction of functional categories and morphology. Grammaticalization, I argue, ultimately follows from (cognitive) economy principles. Chomsky (2002: 113) sees the semantic component as expressing thematic as well as discourse information. If thematic structure was already present in proto-language (Bickerton 1990), the evolutionary change of Merge made them linguistic. What was added through grammaticalization is the second layer of semantic information, namely definiteness/specificity etc.

Keywords: evolution, internal merge, external merge, thematic, discourse, economy, grammaticalization.

1 Introduction

Estimates about the origin of modern human language range from 50,000 to150,000 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? Genetic and areal linguistics might be the places to start. Genetic linguistics provides insights into linguistic relationships and areal linguistics shows 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 and the new World Atlas of

Language Structures, Haspelmath et al 2006) show us more about recent trends than about original features. In this paper, I will therefore examine what (historical) syntax has to say. Recurrent processes in language change provide us insight into the original state of human language. Grammaticalization is such an easily observable process in the history of languages and can therefore be seen as involved in the evolution of language from its earliest stage to the present.

Hauser, Chomsky & Fitch (2002) argue that recursion sets human language apart from animal communication and Chomsky (2005: 11) specifies this further by saying that Merge, linking two elements, was the "`Great Leap Forward' in the evolution of humans". Likewise, Piattelli-Palmarini & 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 Merge brings with it certain relations such as Specifiers, Heads, Complements, and c-command. Heads, complements, and specifiers in turn define argument structure or thematic structure. The thematic layer is one aspect of evolution. 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 can be seen in terms of cognitive economy, e.g. heads are preferred over specifiers and semantically `lighter' elements over `heavier' ones.

The outline is as follows. In section two, I present a very general picture of the Minimalist Program which is elaborated on in section three, especially where Merge is concerned. In sections four and five, grammaticalization is the focus. I discuss how it follows from economy and how it is relevant to language evolution. Section six is a conclusion.

2 A Minimalist Picture

Within the Minimalist Program (Chomsky 1995; 2004 etc.), 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 2006: 10). Phonology and morphology are involved in the externalization and highly varied since there would be 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 2006: 9-10). Figure 1 shows the three components of human language. I will assume that Merge appears after SEM and PHON are already developed, but for the purpose of this paper nothing hinges on this.

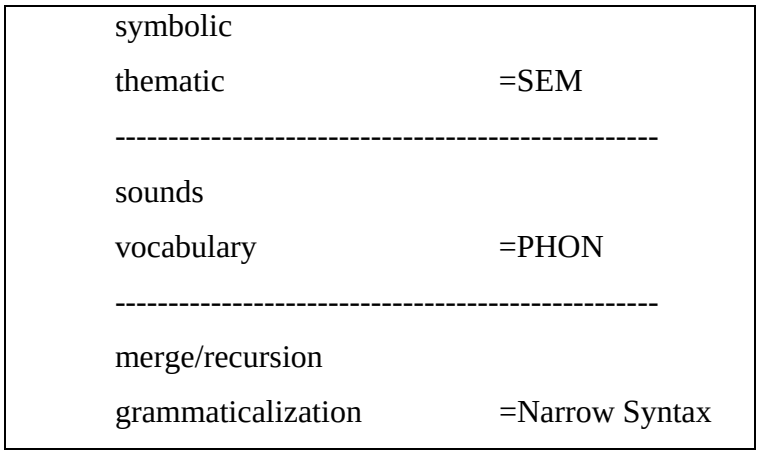


Figure 1: Three separate systems

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 into two kinds, internal and external merge. Chomsky (2005: 14) suggests that external merge is relevant to the argument structure, whereas internal merge

is relevant for scope and discourse phenomena, as in Figure 2. This means external merge may have been an early feature of language. The longest utterance that Nim, a chimpanzee trained by Terrace in the 1970s, uttered 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. Many languages use internal merge for this purpose. The differences between the two kinds of merge are listed in Figure 2.

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

Figure 2: The two kinds of Merge

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*. This topic marker is a definiteness 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* (Actor, Patient, Instrument) and *sa* (Goal, Source, Location, 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)

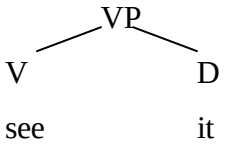
This picture suggests that syntax and morphology evolved independently. Some have argued that they are therefore separate systems, e.g. Jackendoff (2002: 260). Bobaljik (2006) too has agreement adding features after NS. I will argue that external merge emerged first, followed by internal merge. Grammaticalization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

3 How does merge work?

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 (hence 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; 2005; 2006), 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 (2005). Merge is one such operation that can be seen as a UG principle (Chomsky 2006: 4) but also as one possibly "appropriated from other systems" (Chomsky 2006: 5). I'll now turn to how a sentence is actually produced using a Minimalist approach.

There is a lexicon from which lexical items are chosen, after which Merge combines two items, e.g. *see* and *it* in (2), and one of the two heads projects, in this case V, to a higher VP:

(2)



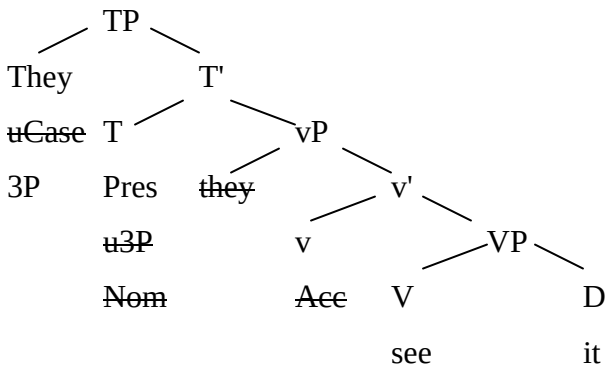
The items that merge are not arbitrarily selected. The head is searching to satisfy a feature, e.g. a thematic one in (2). The head is also the one projecting up, and this projection can then in its turn be selected. The VP domain is the thematic-layer, i.e. where the argument structure is determined.

Apart from merge, there are "atomic elements, lexical items LI, each a structured array of properties (*features*)" (Chomsky 2006: 4). Each language learner selects the features compatible with the input. Thus, the features are parameterized; not the syntax. The main kinds of features involve Case, agreement (also known as phi-features), and displacement to subject position. I will suggest these features were not present during the stage at which merge appeared originally. They were added as a way to incorporate distinctions having to do with specificity and topicality.

Features come in two kinds. Interpretable ones include number on nouns and are relevant at the Conceptual-Intentional interface. Uninterpretable features include agreement features on verbs and Case features on nouns. These features need to be valued and deleted. Continuing the derivation in (1) will make the function of features clearer.

After adding a (small) v and subject *they* to (1), functional categories such as T (and C) are merged to VP. Agree ensures that features in TP (and CP) find a noun or verb with matching (active) features to check agreement and Case. So, T has interpretable tense features but uninterpretable phi-features. It probes ('looks down the tree') for a nominal it c-commands to agree with. It finds this 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 (3) where the features that are not `struck through' are interpretable. The subject moves to Spec TP for language-specific reasons:

(3)



The derivation in (3) 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.

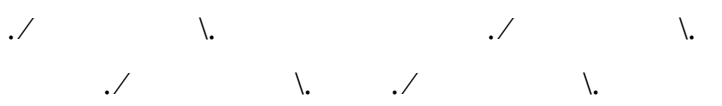
Structures made by Merge involve heads, complements, and specifiers. Merge, thus, automatically brings with it, the following UG Principles:

Principles connected with Merge

a. Merge involves projection, hence headedness, specifiers, and complements

b. The binary character of Merge results in either:

(i) (ii)



c. There is c-command of the specifier over (the Head and) the Complement, resulting in the special nature of the specifier.

Figure 3: Principles connected with Merge

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 (bi) in Figure 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 proto-language was SOV, i.e. (bii).

Turning to language evolution, languages closer to the proto-language will have Merge but there is no reason they would have Move and Agree as in (3) (though Newmeyer 2000: 385, n 4 suggests that proto-languages may have been inflectional). My focus on grammaticalization as a process responsible for morphology assumes that agreement and Case arise later.

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 binarity. 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 4.

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 & 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.

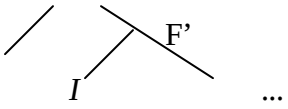
Two Economy Principles, provided as (4) and (7) below, are formulated in van Gelderen (2004). They are part of Universal Grammar and help learners construct a grammar. They are similar to principles such as c-command, in that they remain active in the internalized grammar and therefore also aid speakers in constructing sentences. They are different from absolute principles such as c-command because prescriptive and innovative tendencies can counteract them.

Principle (4) is a principle 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:

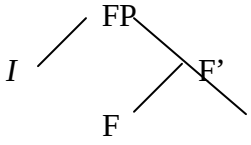
(4) **Head Preference Principle (HPP)**:

Be a head, rather than a phrase.

This means that a speaker will prefer to build structures such as (5a) rather than (5b). The FP stands for any functional category and the (first person) pronoun is merged in the head position in (5a), and in the specifier position in (5b):

(5) a. 

...

b. 

…

The speaker will only use (b) 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 <2>. 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, however, slow 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 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, the exact nature of the input needed for the change, the `cue', is not explored in this paper.

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. As expected, children acquiring their language obey this same economy principle.

Thus, according to Diessel (2004), young children produce only stranded constructions in English, as in (6):

(6) those little things **that** you play with (Adam 4:10, from Diessel 2004: 137).

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

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 1.

relative pronoun *that* to complementizer	Demonstrative to article
Negative adverb to negation marker	Adverb to aspect marker
Adverb to complementizer (e.g. *till*)	Full pronoun to agreement

Table 1: Examples of reanalysis due to the Head Preference Principle

Within Minimalism, there is a second economy principle that is relevant to grammaticalization. Combining lexical items to construct a sentence, i.e. Merge, "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. This could be formulated as in (7):

(7) **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 (7) is still valid since the special Merge, i.e. internal Merge, requires steps additional to the ones 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, (5) could still hold as an Economy Principle. As mentioned, Chomsky (2005: 14) suggests that 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 that is expressed in the LMP.

The Late Merge Principle works most clearly in the case of heads. 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 (8), and manner, where *according to* would be the modern equivalent:

(8) *Fand þa ðær inn æþelinga gedriht swefan* ***æfter*** *symble*
`found then in there (the) noble company sleeping after (their) feast'
(*Beowulf* 118-9)

Of the 41 instances of *after* in the relatively early *Beowulf*, only one occurs inside a fronted PP and in the *Christ* from the *Exeter Book*, there are none out of 15. The objects of these prepositions are full nouns or personal pronouns. None of these introduce a subordinate sentence.

This changes in later texts in that the PPs are fronted more often and the object is a demonstrative. Texts such as the *Chronicle A* contain 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 (7.7%) by a demonstrative; the PP is preposed in 27% of the cases. In the later Chronicle (i.e. after 892), many of the objects of *after* are demonstratives, as in (9), 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:

(9) ***Æfter þysan*** *com Thomas to Cantwarebyri*
`After this, Thomas came to Canterbury'. (Chronicle A, entry for the year 1070)

The use of the demonstrative object indicates that the PP is starting to be seen as an adverb linking the sentence to another. This is confirmed by the frequent fronting of the PP (12 out of 22 =54.5%), as in (9) above, and (10). The fronting can be seen as a consequence of Late Merge:

(10) *7* ***æfter þon*** *forðferde Gyric mæsse preost*
and after that died Gyric the mass-priest (Chronicle A, entry for the year 963)

In table 3, the differences are summarized, with percentages rounded off.

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

Table 3: Percentages of demonstrative objects (Dem) with *after* and fronting

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

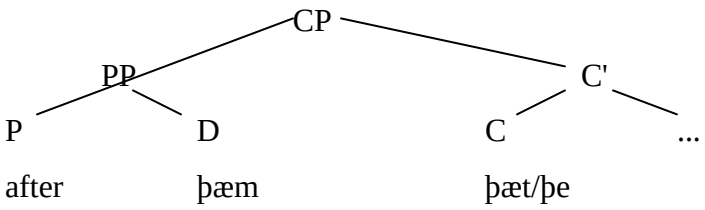
(11) *Witodlice* ***æfter þam þe ic of deaþe arise*** *ic cume to eow on galilee*
Surely after that that I of death arise I come to you in Galilee (c1000 *West Saxon* Matthew 26. 32 Hatton Ms).

(12) ***Efter þan þet*** *þe mon bið dead*
After that that the man is dead (c1175 *Lamb. Hom.* 51).

(13) ***Affterr þatt tatt*** *he wass dæd*
After that that he was dead
`After he was dead'. (c1200 *Ormulum* 7667)

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

(14)



Interestingly, the much earlier Lindisfarne glosses render the relevant part of (11) as (15), without the complementizer. The complementizer-less stage therefore seems to represent 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. Sentence (15) also switches the order of the CP-adverb and the C:

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

So far the development 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, and according to Rissanen (2006) this is common. 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 (c1360 Wyclif *De Dot. Eccl.* 22).
(17) After thei han slayn them (1366 Mandeville174).

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 & 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 *soþlice).* 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:

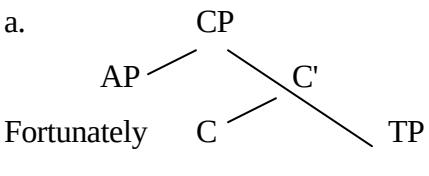
(19) Whan a man..clymbeth vp and wexeth **fortunat.** (OED, 1386, Chaucer)

(20) **Most fortunately**: he hath atchieu'd a Maid That paragons description, and wilde Fame: One that excels the quirkes of Blazoning pens (Shakespeare, Othello)

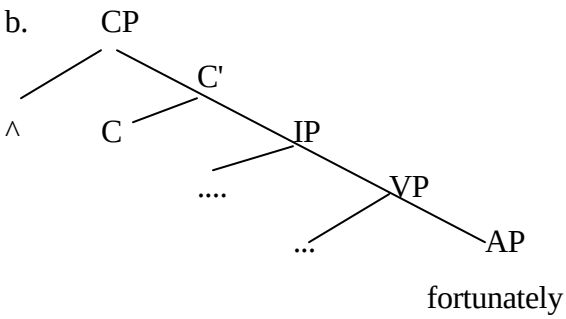
(21) **Fortunately**, Lord De la War..met them the day after they had sailed (OED, 1796).

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

(22) a.



b.



Other examples of the LMP are given in Table 2.

Like, from P > C (*like I said*)	Negative objects to negative markers
Modals: v > ASP > T	*To*: P > ASP > M > C

Table 2: Examples of reanalysis due to the Late Merge Principle

How exactly does Late Merge account for language change? If non-theta-marked elements can wait to merge outside the VP (Chomsky 1995: 314-5), 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 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 & Roussou (2003), Wu (2004), and Simpson & Wu (2002) also rely on some version of Late Merge.

Concluding section 3, 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.

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 2000), and as Traugott (2004: 134) puts it as "an exaptation of thematic role structure". The current section provides a scenario for subsequent steps.

Once External Merge applies, certain structural and thematic relationships crystalize, unlike those in (1), uttered by Nim. Chomsky (2006) 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 v a VP (in which a DP with a specific theta role occurs). The vP represents the thematic level, and one that adult native speakers employ when they speak or write in `fragments', as in (23a). Children reach this

stage too, as (23b) shows, though they understand grammatical categories before they produce them:

(23) a. Work in progress
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 as 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 is when internal merge (movement) arises, as well as grammatical elements relevant to specificity and definiteness. I will add that 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 & Thompson (1978). Cf. (25):

(25) a. *zei* *pao* *le*
thief run ASP
`The thief has run away'.
b. *pao* *le* *zei*
run ASP thief
`A thief has run away' (from Li & Thompson 1978: 228)

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 CP-adverbs and auxiliaries.

The two principles used above (HPP and LMP) take 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 DP *na wiht* `no creature' 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: 83-7). 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 us gives good life and long
`With the help of God, our Lord, who gives us a good and long life' (Bekynton, from Rydén, p. 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 pied a preposition and is inflected). This shows that, for creative reasons, speakers can start to use the specifier again.

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 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:

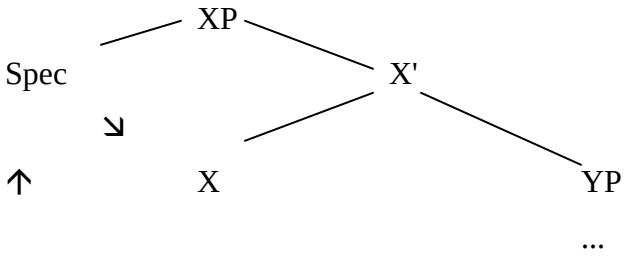


Figure 4: The Linguistic Cycle

This scenario works perfectly for changes where a negative object such as Old English *na wiht* `no creature' becomes a Spec 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 (a) is the one from lower head (either lexical or grammatical) to higher head, via LMP. The change in (b) 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.

6 Conclusion and future directions

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 other is Economy of Merge, the HPP and LMP, the 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 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 proto-language (Bickerton 1990), the evolutionary change of Merge made them linguistic. What was added through grammaticalization is the morphology, the second layer of semantic information.

Abbreviations

BNC	British National Corpus, see references.
HPP	Head Preference Principle
LMP	Late Merge Principle
OED	Oxford English Dictionary
SIP	Specifier Incorporation Principle
UG	Universal Grammar

References

Abraham, Werner 1993. "Grammatikalisierung und Reanalyse: einander ausschlieBende oder ergänzende Begriffe". *Folia Linguistica Historica*, 13: 7-26.

Bickerton, Derek 1990. *Language and Species*. Chicago: University of Chicago Press.

Bickerton, Derek. 2000. How proto-language became language. In Chris Knight, Michael Studdert-Kennedy and James R. Hurford, eds., *The Evolutionary Emergence of Language*, 264-284. Cambridge: UK: Cambridge University Press.

Bobaljik, Jonathan 2006.

British National Corpus, BNC, http://sara.natcorp.ox.ac.uk.

Carstairs-McCarthy, A., 1999. *Origins of complex language.* OUP.

Chomsky, Noam 1986. *Knowledge of Language*. New York: Praeger.

Chomsky, Noam 1995. *The Minimalist Program*. Cambridge: MIT Press.

Chomsky, Noam 2002. *On Nature and Language*. CUP.

Chomsky, Noam 2004. "Beyond Explanatory Adequacy". In Adriana Belletti (ed.), *Structures and Beyond*, 104-131 OUP.

Chomsky, Noam 2005. "Three factors in Language design". *Linguistic Inquiry* 36.1: 1-22.

Chomsky, Noam 2006. "Approaching UG from below". ms.

Corballis, Michael 2002. "Did Language evolve from Manual Gestures?". In Alison Wray (ed.), *The Transition to Language*, 163-180. OUP.

Finney. Joseph 1999. "General Diachronic Course of Proto-Austronesian casemarkers". SEALS 9. Tempe: ASU Monograph Series.

Frawley, William 1976. *Comparative Syntax in Austronesian*. UCB, PhD.

Gelderen, Elly van 2004. *Grammaticalization as Economy*. Amsterdam: John Benjamins.

Givón, Tom 1979. "From discourse to syntax". *Syntax & Semantics* 12, 81-112. New York: Academic Press.

Hauser, Marc, Noam Chomsky, & Tecumseh Fitch 2002. The Faculty of Language: what is it, who has it, and how did it evolve? *Science*: 298: 1569-79.

Heine, Bernd & Tania Kuteva 2002. *World Lexicon of Grammaticalization.* Cambridge: Cambridge University Press.

Hopper, Mike & Elizabeth Traugott 1993. *Grammaticalization*. Cambridge: Cambridge University Press.

Jackendoff, Ray 2002. *Foundations of Language*. OUP.

Jelinek, Eloise 1998. "Voice and transitivity as Functional Projections in Yaqui". In Miriam Butt et al (eds), *The Projection of Arguments*, 195-224. Stanford: CSLI.

Jespersen, Otto. 1922. *Language*. London: Allen & Unwin.

Lambrecht, Knut 1981. *Topic, Antitopic, and Verb Agreement in Non Standard French*. Amsterdam: John Benjamins.

Li, Charles & Sandra Thompson 1978. "An Exploration of Mandarin Chinese". In W.A. Lehmann (ed.), *Syntactic Typology*, 223-265. Austin: University of Texas Press.

Lightfoot David 1999. *The development of Language*. Malden: Blackwell.

Piattelli-Palmarini, Massimo & Juan Uriagereka 2005. The Evolution of the Narrow Faculty of Language. *Lingue e Linguaggio*, 1-52.

Mustanoja, Tauno 1960. *A Middle English Syntax*. Helsinki.

Newmeyer, Frederick 2000. "On the Reconstruction of 'Proto-World' Word Order". In Chris Knight et al (eds) *The Evolutionary Emergence of Language*, 372-388. CUP.

Nichols, Johanna 1992. *Linguistic Diversity in Space and Time*. University of Chicago Press.

Pierce, Amy 1992. *Language Acquisition and Syntactic Theory*. Dordrecht: Kluwer.

Rissanen, Matti. 2006. ICEHL talk.

Roberts, Ian & Anna Roussou 2003. *Syntactic Change*. Cambridge: Cambridge University Press.

Romaine, Suzanne 1982. *Socio-historical Linguistics*. Cambridge: Cambridge University Press.

Rydén, Mats 1983. "The Emergence of *who* as relativizer". *Studia Linguistica* 37: 126-134.

Simpson, Andrew & Xiu-Zhi Zoe Wu 2002. "From D to T - Determiner Incorporation and the creation of tense". *Journal of East Asian Linguistics* 11: 169-202.

Tauli, Valter 1958. *The Structural Tendencies of Languages*. Helsinki.

Traugott, Elizabeth 2004. "Exaptation and Grammaticalization". Minoji Akimoto (ed.) *Linguistic Studies based on Corpora*, 133-56. Tokyo: Hituzi Syobo.

Wu, Zoe 2004. *Grammaticalization and Language Change in Chinese*. London: RoutledgeCurzon.

Notes

<1>Thanks to audiences in Oslo and Stellenbosch and especially to Terje Lohndal for very helpful discussions.

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