This book examines cyclical change and shows how that change provides a unique perspective on the language faculty. According to one definition in the *Oxford English Dictionary*, a cycle is a "period in which a certain round of events or phenomena is completed." Towards the end of the cycle, similar events start again, but they are (slightly) different and happen at a different pace. The changes are therefore unidirectional.

With the exception of the negative cycle, generative linguists have not seriously examined the idea that language change is cyclical\(^1\). However, the emphasis within the Minimalist Program on principles not specific to language, in conjunction with universal grammar, has prompted generative linguists to look for economy principles. I claim that cyclical change provides insight into the principles governing the language faculty: economy principles are the reason for linguistic cycles.

Early descriptions of cycles or cyclical change can be found in de Condillac (1746), Tooke (1786-1805), Bopp (1816), and von Humboldt (1822) but, apart from work by Tauli (1958), Hodge (1970), Greenberg (1978), Givón (1978), and Katz (1996) not much recent research has been done on linguistic cycles. Thus, the current state of research is not much better than in 1972, when Robin Lakoff wrote that "there is no mechanism within the present theory of transformational grammar that would allow an explanation" (1972: 173-4). There is even outright rejection of the idea of linguistic cycles. For instance, Newmeyer (1998: 263-275; 2001) dismisses unidirectional change as does Lightfoot (e.g. 2006a: 38). Others, such as Traugott & Dasher (2002: 87), claim that the number of counterexamples to unidirectionality is small and not systematic. I take this approach and have not dealt with the occasional counterexample.

---
\(^1\) More work on the negative cycle has started to appear. For instance, in 2008 and 2009, one-day events on the negative cycle took place in Birmingham ([http://www.lhds.bcu.ac.uk/english/cycles-of-grammaticalization](http://www.lhds.bcu.ac.uk/english/cycles-of-grammaticalization)) but other cycles have not been given as much attention.
In this book, I claim that cycles are the result of reanalysis by the language learner and I attempt to provide an explanation of this phenomenon within generative grammar, i.e. within the Minimalist framework. I argue that the real sources of change are internal principles that bias the learner towards certain structures. This is very different from models such as Lightfoot's (2006) and Westergaard’s (2009) that examine how much input a child needs to reset a parameter. According to Lightfoot, "children scan their linguistic environment for structural cues" (2006: 32); therefore, change comes from the outside, i.e. it is triggered by variable data. The challenge for this type of an account is to determine the external triggers. I, on the other hand, argue that change mainly comes from the inside (allowing of course changes in the input due to external factors as well).

I have four goals in this book. One major objective is to provide examples of linguistic cycles: the head marking cycle (from subject and object pronoun to subject and object agreement) in Chapters 2 and 3 and the dependent marking cycle in Chapters 5 and 6. Changes in the copula are related to the agreement cycle and will be discussed in Chapter 4. These are major cycles, or macro-cycles, in that they change the overall typology of a language. Other cycles are the future and aspect cycles (Chapter 7) and the negative cycle (Chapter 8). These are minor cycles, or micro-cycles, in that they do not change the typological characterization of a language. For all cycles, I explore what the typical steps in the cycles are, where they start and how they renew themselves.

Another objective is to show that Economy Principles (present in the initial cognitive system or Universal Grammar of the child) can account for parts of linguistic cycles. In van Gelderen (2004), I apply this approach to some instances of grammaticalization. In the current book, I examine full cycles. I argue that Economy Principles, in particular Feature Economy, are responsible for the various stages of linguistic change. Loss of semantic features occurs when full verbs such as Old English will with features such as [volition, expectation, future] are reanalyzed as having only the feature [future] in Middle English. The features can then be considered grammatical rather than semantic. The grammatical features come in two kinds, features that are interpretable at the Conceptual-Intentional Interface and those that are uninterpretable at that interface but functioning to link two positions. I use Feature Economy to explain this change: semantic features are not economical in the computation since they make the
elements to be combined inert. Interpretable features are slightly more economical in their interactions since they can value uninterpretable features. Uninterpretable features act as probes and are the most economical in keeping the derivation going. Hicks (2008: 220) characterizes Feature Economy as "establish[ing] dependencies where possible". I see Feature Economy as the main force behind the linguistic cycle, semantic features become grammatical which in turn need semantic ones again.

A third objective of this book, pursued throughout but summarized in Chapter 9, is to argue that some of the cycles (e.g. the agreement cycle) are relevant in the typological classification of languages and others (e.g. the negative cycles) are not. Heine et al. (1991: 246) argue that there is “more justification to apply the notion of a linguistic cycle to individual linguistic developments” rather than to changes from analytic to synthetic and back to analytic. Synthetic morphology includes both head-marking and dependent-marking. Nichols (1992: 46-96) uses a complex typology in which languages are assigned points for head or dependent-marking. She considers head and dependent marking in the clause, the NP and the PP and finds a consistency (if dependents are marked in the NP, they are also marked in the clause and the PP). One of her other conclusions is that head marking and dependent marking “are about equally frequent overall” (1992: 95) even though certain geographic areas prefer one or the other. These conclusions are important for cyclical change but are outside the scope of this book. Thus, I have not looked at the marking of grammatical relations inside the NP or PP, and will not be looking at cycles involving head-marking to dependent marking and back.

Baker (2001) has suggested macro-parameters. Following Sapir, he argues that a language has a basic character. Thus, the choice of polysynthesis, for example, implies that the language will have many other characteristics. Some of the changes discussed in this book show that Baker's macro-parameters are not valid since a change in the polysynthesis 'parameter' does not necessarily trigger a change in other parameters, as would be expected in Baker’s approach. I try to find a compromise between Baker’s (2001 and 2008a) approach and a parametric approach involving only features, as in much recent work by Chomsky.

The final objective of the book is to argue that research into language change can provide insight into the shape of the earliest human language and how it evolved. Typical
answers to the question are put as either adaptive/gradual evolution (Pinker & Bloom 1990, Givón 2009) or as gapped evolution (Chomsky 2005). Chomsky asks two questions: Why is there language at all? and Why are there so many languages? The answer to the first question is that a major evolutionary shift occurred, enabling the operation Merge. Currently, that is the main component of language: "the core principle of language, unbounded Merge, must have arisen from some rewiring of the brain" (Chomsky 2008: 9). To answer the second question, externalization may have developed later and may not have involved a genomic change: "the reason might be that the problem of externalization can be solved in many different and independent ways, either before or after the dispersal of the original population" (Chomsky 2008: 10). I will argue that cycles may show us how languages develop. Chapter 10 is devoted to language evolution.

The outline of this introductory chapter is as follows. In section one, I provide some background on grammaticalization and cycles as well as a list of cyclical changes. Section two introduces the Minimalist Program and Economy Principles. Section three examines language acquisition data that may provide evidence for Economy Principles. Sociolinguistic factors interfere with Economy, as I show in section four (though I will not focus on this in the remainder of the book). In section five, I discuss the relevance of Economy to language typology and in section six methodological issues. Section seven provides an outline of the rest of the book.

1. Grammaticalization and Cyclical Change

1.1 Grammaticalization

As is well-known, grammaticalization is a process whereby lexical items lose phonological weight and semantic specificity and gain grammatical functions. The best-known examples of lexical elements changing to grammatical ones are verbs being reanalyzed as auxiliaries and prepositions as complementizers. There are also
grammatical elements that are reanalyzed into more grammatical ones.\textsuperscript{2} These changes necessitate renewal and the entire process is sometimes referred to as a \textit{linguistic cycle}. As Mithun (2000: 232) says, “morpheme order often does reflect the sequence of grammaticalization of affixes: those affixes closest to the root are indeed the oldest, and those on the periphery of words can be seen to be more recent additions”.

Grammaticalization was identified early on but was established as a term only in 1912 by Meillet. Works such as Lehmann (1982) and Traugott \& Heine (1991) have inspired many linguists to pay closer attention to this phenomenon, especially in a functionalist framework. Recently, there have been structural accounts of the cyclicity of the changes involved. Van Gelderen (2004; 2008abe; 2009b), for instance, discusses Economy Principles that help the learner acquire a grammar that is more economical and, therefore, more grammaticalized. Wu (2004), Simpson \& Wu (2002ab), Roberts \& Roussou (2003), Eckardt (2006), and Roberts (2007) provide formal accounts of grammaticalization, especially change ‘up the tree.’ Roberts \& Roussou use “upwards reanalysis” (2003: 205). Fuß (2005) argues for a morphological reanalysis when existing inflection is defective. The phonology of grammaticalization has been scrutinized by Schiering (2006) who shows that the phonology is not always reduced.

Grammaticalization is a descriptive term and I use \textit{reanalysis} to emphasize the role of the child acquiring the language. Technically, it is not \textit{re}analysis since a child hears language and analyzes the linguistic input in the most economical way. This may result in a grammar different from that of an earlier generation, which leads linguists to refer to the phenomenon as reanalysis. Grammaticalization is thus seen as following from the innate properties of the language faculty and the task of the linguist is to unearth the principles. Examining unidirectional language change provides a unique window on the principles of the language faculty.

The changes involved in grammaticalization can be schematized as in (1), were (1a) represents the morphosyntactic changes and (1b) the changes in argument status.\textsuperscript{3}

\begin{equation}
(1) \quad \text{a. phrase} > \text{word/head} > \text{clitic} > \text{affix} > 0
\end{equation}

\textsuperscript{2} Andersen (2008) distinguishes lexical elements becoming grammatical, which he terms \textit{grammatization}, from grammatical elements becoming other grammatical element, which he calls \textit{regrammatization}.

\textsuperscript{3} See Siewierska (2004: 261-2) for slightly different clines.
Once the change reaches the right side of (1), renewal and borrowing bring new words and phrases into the language, starting the grammaticalization chain all over again. Table 1.1 lists some well-known English examples of lexical elements reanalyzed as grammatical ones.

<table>
<thead>
<tr>
<th>V&gt;AUX</th>
<th>P&gt;AUX</th>
<th>P&gt;C</th>
</tr>
</thead>
<tbody>
<tr>
<td>go</td>
<td>motion &gt; future</td>
<td>to direction&gt;mood</td>
</tr>
<tr>
<td>have</td>
<td>possession&gt;perfect</td>
<td>on location&gt;aspect</td>
</tr>
</tbody>
</table>

Table 1.1: Instances of Grammaticalization

In (2) to (7), I provide examples of the six changes listed in Table 1.1: (a) exemplifies the earlier lexical use and (b) the later grammatical use. Most of these are cases where the lexical and grammatical items occur during the same stage of the language: Modern English has *go*, *have*, and *to* as both lexical and grammatical categories, as shown in (2), (3) and (4), but the lexical category is the earlier one. The grammaticalization of *on* in (5) resulted in its loss, so the grammaticalized form (5b) disappeared. In the case of *for* (6) and *after* (7), the lexical use in (6a) and (7a) has disappeared (renewed by *in front of* and *behind/following*), but the grammatical use remains.

(2) a. I told Cowslip we were **going** before I left the burrow. (BNC-EWC 3181)
   b. Anne can HAVE her Mini...Cause I's **gonna** get me a BMW
      (http://www.inkycircus.com/jargon/2006/09/anne_can_have_h.html)

(3) a. I **have** a garden.
   b. I **have** seen the garden.

(4) a. The highway **to** Phoenix
   b. I didn't expect **to** find you here. (BNC FPM 899)

(5) a. Above them **on** the balcony terrace, Alina Petrovna stood ... 
      (BNC-FYY 1799)
   b. and iuunden þene king. þær he wes **an** slaeting
and found the king there he was on hunting
‘and they found the king where he was hunting’ (Layamon, Caligula 6139)

(6) a.  
\[ hlynode \quad \textit{for} \quad hlawe \]
made-noise before mound
‘It made noise before/around the gravehill' (Beowulf 1120).

b. I would prefer \textit{for} John to stay in the 250 class. (BNC-ED2 626)

(7) a.  
\[ Ercenberht \textit{rixode after} his fæder \]
‘E. ruled after/following his father' (Chron A, entry for the year 640)

b. \textit{After} she'd hung up, she went through into the kitchen. (BNC GWO1402)

Once an element reaches the right side of (1), renewal takes place and we have cyclical change.

1.2 \textit{The Linguistic Cycle}

Hodge (1970: 3) calls the cyclical phenomenon where "one man's morphology was an earlier man's syntax" the ‘Linguistic Cycle.’ \textit{Spiral} is another term for cycle \(^4\) (see von der Gabelentz 1901: 256; Hagège 1993: 147); it emphasizes the unidirectionality of the changes: languages do not reverse earlier change but may end up in a stage typologically similar to an earlier one. Jespersen (1922: chapter 21.9) uses spirals when he criticizes the concept of cyclical change. His criticism is based on his views that languages move towards flexionless stages in a unidirectional manner. Jespersen's views are not correct: languages and families such as Finnish, Altaic, and Athabascan increase in morphological complexity.

The changes in (2) to (7) represent small steps in certain cycles, the auxiliary cycle in (2) to (5) and the clausal one in (6) and (7). This is true for other cycles as well. The negative cycle is well-established for Indo-European, e.g. Jespersen (1917), though some of these changes had been identified early on by Gardiner (1904) for Egyptian (see van der Auwera & de Vogelaer 2008). The article or definite cycle has been discussed by

\(^4\) \textit{Cycle} is also a technical term in a generative derivation. Currently, \textit{phase} is more commonly used in that context.
Tauli (1958) and more recently by Lyons (1999). Cycles have also been argued to affect morphological type. Hodge (1970) examines the rise and fall of overt morphology in Egyptian. He argues that a cycle occurred in Egyptian: Old Egyptian morphological complexity (synthetic stage) turned into Middle Egyptian syntactic structures (analytic stage) and then back into morphological complexity in Coptic. This cycle is discussed here as the agreement cycle (see particularly Chapter 2). Tauli also considered the changes involving agreement cyclical. Table 1.2 lists full cycles up to the point where they are renewed by an element similar to the left most in the cline.

<table>
<thead>
<tr>
<th>Subject Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>demonstrative/emphatic/noun &gt; pronoun &gt; agreement &gt; zero</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Object Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>demonstrative/pronoun &gt; agreement &gt; zero</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Copula Cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>demonstrative &gt; copula &gt; zero</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case or Definiteness or DP</th>
</tr>
</thead>
<tbody>
<tr>
<td>demonstrative &gt; definite article &gt; ‘Case’ &gt; zero</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>a negative argument &gt; negative adverb &gt; negative particle &gt; zero</td>
</tr>
<tr>
<td>b verb &gt; aspect &gt; negative &gt; C</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Future and Aspect Auxiliary</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/P &gt; M &gt; T &gt; C</td>
</tr>
</tbody>
</table>

Table 1.2: Cycles

In the next section, I will provide some background on the syntactical framework used. This will enable us to examine grammaticalization and the cycle in a structured, explanatory way.

2. Minimalism, Economy, and Cycles
In this section, I will review the Minimalist Program, the basic clausal structure, Economy Principles, and cycles.

2.1 The Framework

In the 1950s, Chomsky’s generative model offered an alternative to behaviorist and structuralist frameworks. Chomsky focuses not on the structures present in the language/outside world but on the mind of the language learner/user. The input to language learning is seen as poor (the 'poverty of the stimulus' argument): speakers know so much more than what they have evidence for in the input. How is this possible? The answer to this problem, Plato's problem in Chomsky (1986a), is Universal Grammar (hence UG), the initial state of the language faculty, a biologically innate organ. UG helps the learner make sense of the data and build an internal grammar. In the 1980s, UG is seen as consisting of Principles (true in all languages) and Parameters (choices to be made depending on the language).

Currently, the role of parameters and of UG in general is considered much less important than it was in the 1980s. Parameters now (Chomsky 2004; 2007) consist of choices of feature specifications as the child acquires a lexicon. All parameters are lexical and determine linearization; therefore, they account for the variety of languages. Baker, while disagreeing with this view of parameters, calls this the Borer-Chomsky-Conjecture (2008a: 156, 2008b: 3): "All parameters of variation are attributable to differences in the features of particular items (e.g., the functional heads) in the lexicon." I briefly discuss parameters in section five.

Initially, many principles were also attributed to Universal Grammar (UG). At the moment, however, the emphasis is on principles not specific to the language faculty (UG), but to "general properties of organic systems" (Chomsky 2004: 105), ‘third factor principles' in Chomsky (2005; 2007). Chomsky (2007: 3) identifies three factors crucial in the development of language.

(1) genetic endowment, which sets limits on the attainable languages, thereby making language acquisition possible; (2) external data, converted to the experience that selects one or another language within a narrow range; (3) principles not specific to FL. Some of the third factor principles have the flavor of
the constraints that enter into all facets of growth and evolution.... Among these are principles of efficient computation.

The third factor is divided into several types, including principles of efficient computation, which are "of particular significance in determining the nature of attainable languages" (Chomsky 2005: 6). The Economy Principles discussed later, and reformulated in terms of Feature Economy, are probably also part of more general cognitive principles, thus reducing the role UG plays. If children use economy principles in building their internalized grammars, there should be some evidence.

2.2 Basic clausal structures

In a pre-Minimalist system (e.g. Chomsky 1986b), a clause consists of an outer layer (the Complementizer Phrase, or CP), an inner layer (the Tense Phrase, or TP), and a thematic layer that contains the verb and its arguments (the traditional Verb Phrase, or VP). The outer layer is responsible for encoding discourse information and linking one clause to another; the inner layer is involved in the marking of tense and agreement through morphology or auxiliaries; and the lowest layer determines the thematic roles. Each layer can be expanded: when the sentence is negative, a Neg(ative)P is added.

In this framework, syntactic structures are built by using general rules, such as that each phrase consists of a head (X), and a complement (ZP) and specifier (YP).

(8)  XP
     /   \
    YP    X'
         /   \  
        X    ZP

This holds for both lexical (N, V, A) and grammatical categories (C, D, T).

In the Minimalist Program (Chomsky 1995; 2001; 2004; 2007), phrase structures are abandoned in favor of a general rule Merge. Merge combines two bundles of features; from Merge, the relations in (8) follow automatically. As Boeckx (2008: 75) explains,
following Chomsky (1995), the three levels in (8) can be seen as follows: "a minimal projection [X in (8)] is a lexical item selected from the lexicon, ...a maximal projection [XP in (8)] is a lexical item that doesn't project any further, ....an intermediate projection [X' in (8)] is ... neither minimal not maximal." For convenience, I continue to use the levels in (8) as well as specifier, head, and complement in what follows.

In a Minimalist approach, a Modern English derivation proceeds in four steps. First, items are selected from the lexicon. Chomsky (2006: 6) suggests the lexicon has "atomic elements, lexical items LI, each a structured array of properties." Abstracting away from features, a lexical array could be \{saw, it, T, Martians\}. Second, the elements are merged, e.g. saw and it in (9), and one of the heads (in this case V) projects to a higher VP.

(9)  
\[
\begin{array}{c}
\text{VP} \\
\text{V}  \\
\text{D}  \\
\text{see}  \\
\text{it}
\end{array}
\]

There is some debate as to whether there is an additional labeling operation and which elements are selected for merge. The labels in (9) are added for convenience only (Chomsky 2004, but see Lohndal 2006). Most likely, it is the theta-requirements that make sure saw and it merge and not saw and T.

Third, after adding a (small) v and subject Martians to (9), as in (10), functional categories such as T (and C) are merged to vP. Agree ensures that features on v and T find a noun with matching (active) features to check agreement. When agreement is checked so is what is often referred to as Case (at least in many languages, but see Baker 2009 for languages without this connection). Following Pesetsky & Torrego (2001), I will call this tense rather than nominative and aspect rather than accusative. So, v and T have interpretable tense and aspect features but uninterpretable phi-features. They probe (search)

---

5 Chomsky (2005: 13; 2007: 11) suggests that certain heads have edge features that drive merge, both external merge, as in (9), and internal merge. These edge features overlap with uninterpretable features and EPP-features, and I ignore them in this book.

6 Williams (1994: 11-12) is perhaps the first to argue that an NP has tense: “Nominative NPs are simply tensed NPs.”
for a nominal that they c-command to agree with. The v finds this nominal in *it* and T finds this nominal, or goal, in *Martians* and each element values its uninterpretable features which then delete.

The final structure looks like (10) where the features that are not ‘struck through’ are interpretable and not subject to elimination. The subject moves to Spec TP: it is merged from an internal position for language-specific reasons (EPP or OCC\(^7\)).

\[
(10) \quad \text{TP} \\
\quad \text{Martians} \quad T' \\
\quad \[u-T\] \\
\quad \[3P\] \quad T \quad vP \\
\quad \text{PST} \\
\quad \[u-\phi\] \quad \text{Martians} \quad v' \\
\quad \[iT\] \\
\quad v \quad \text{VP} \\
\quad \[i-\text{ASP}\] \\
\quad \[u-\phi\] \quad V \quad D \\
\quad \text{saw} \quad \text{it} \\
\quad \[3S\] \\
\quad \[u-\text{ASP}\]
\]

A difference from e.g. Chomsky (1995: 277-8) is that I see the case features on T and v as interpretable. This means that at LF they are interpreted, as finite and transitive respectively. The derivation in (10) uses early lexical insertion, i.e. a lexicalist approach, as in Chomsky (1995; 2004). In this book, nothing hinges on this. The CP layer is not indicated in (10).

\(^7\) Many have indicated that EPP or OCC features are non-Minimalist since they are not interpretable at the interfaces; see e.g. Stroik (2009: 7). I assume they will at some point turn out to be relevant to the CI interface.
At some point, the derivation has to be handed over to the Sensorimotor (SM) and Conceptual-Intentional (CI) systems external to the syntax. This is done through the interfaces PHON and SEM, corresponding to PF and LF in older frameworks. This emphasis on language-external systems is formulated as (11).

(11) Strong Minimalist Thesis
Language is a perfect solution to interface conditions (Chomsky 2007: 3)

Thus, the syntax has to satisfy the requirements of the external systems, which is achieved via the interfaces PHON and SEM. The former is responsible for linearization and externalization, e.g. what is spelled-out in (10) (see Nunes 2004 for an account on how to decide which copies to spell out). For the CI system, two aspects are relevant: the theta-structure (determined in English through position but in other languages through inherent Case) and the discourse information. The topic and focus can be determined through aspect together with case (Abraham 1997; Philippi 1997), through definiteness markers, or through position, as in Chinese where indefinite objects and subjects appear towards the end of the sentence, as (12) and (13) show.

(12) a. \textit{chi le fan}  \hspace{1cm} \text{Chinese}
    \begin{align*}
    & \text{eat PF rice} \\
    & \text{‘I ate some rice.’}
    
    \end{align*}

b. \textit{fan chi le}  \\
    \begin{align*}
    & \text{rice eat PF} \\
    & \text{‘I ate the rice.’}
    
    \end{align*}

(13) \textit{Lai le yi ge ren}  \hspace{1cm} \text{Chinese}
    \begin{align*}
    & \text{come PF one CL man} \\
    & \text{‘A man came.’} (\text{Li \& Thompson 1981: 20; Yi Ting Chen p.c.})
    
    \end{align*}

As Chomsky (2002: 113; 2008) points out, the semantic component expresses thematic as well as discourse information:
In “what John is eating what,” the phrase “what” appears in two positions, and in fact those two positions are required for semantic interpretation: the original position provides the information that “what” is understood to be the direct object of “eat,” and the new position, at the edge, is interpreted as a quantifier ranging over a variable, so that the expression means something like “for which thing x, John is eating the thing x.” (Chomsky 2008: 8)

Two mechanisms are responsible for this: external and internal Merge, respectively. Merge is essential, and it is the core of the derivational system. Through Merge, binary and hierarchical relationships between the merged elements form. We refer to the merged constituents as heads, complements, and specifiers. Merge, thus, brings with it the following Principles.

(14) **Principles connected with Merge**

a. Merge involves projection, hence headedness, and heads and phrases
b. The binary character of Merge results in either:
   
   ![Diagram](i) ![Diagram](ii)
   
   c. There is c-command of the specifier over (the Head and) the Complement, resulting in the special nature of the specifier.

d. There is recursion.

A lot can be said about each of these principles. For instance, it has been argued in Kayne (1994) that all languages are right-branching as in (bi). This means that there are no headedness parameters.

Within Minimalism, Principles of Economy have played an important role. For instance, Rizzi (2004: 224) argues that Relativized Minimality is an Economy Principle "that appears to be a natural principle of mental computation." ‘Fewest steps’, ‘last resort’, ‘least effort' are all relevant in syntactic derivations (see Chomsky 1995; Zwart 1996; Collins 1997; Fox 2000). This means that, in building derivations, there are ways to
resolve ambiguous structures. Economy is part of the syntax, the I-language, and not the processing system, the E-language, as in Hawkins (2004: 31). In the remainder of this section, I outline a few Economy Principles.

2.3 The Head Preference, Late Merge, and Feature Economy Principles

Lightfoot (1979: 121) introduces an Economy Principle, the Transparency Principle, that “requires derivations to be minimally complex”. His focus is on the child’s postulating of underlying structures that are close to their surface structures. In current Minimalism, the emphasis is not on simplifying the distance between numeration and Sensory-Motor Interface, but the spirit of Lightfoot’s proposal is simplicity of representation and this will be relevant in the principles discussed here. Van Gelderen (2004) justifies principle (15), 8 which is at work in the internalized grammar either due to UG or due to general cognitive principles. This principle holds for external merge (projection) as well as internal merge (movement).

(15) **Head Preference Principle (HPP):**
Be a head, rather than a phrase.

In accordance with the HPP, a speaker will build (16b) rather than (16a), if given evidence compatible with either. The FP stands for any functional category; a pronoun (as well as an adverb or a preposition) is merged in the head position in (16b), but occupies the specifier position in (16a).

(16)  

\[
\begin{array}{ll}
\text{a.} & \text{FP} \\
& \text{FP} \\
& \text{pronoun} \ F' \\
& \text{F} \ldots \\
\text{b.} & \text{FP} \\
& \text{F} \ldots \\
& \text{pronoun} \ F' \\
\end{array}
\]

---

8 Within a Distributed Morphology framework, Embick & Marantz (2006: 25) discuss the hypothetical principle Lexical Preference which says “use a word rather than a phrase if they express the same features.”
Hawkins' (2004) efficiency principle has a Minimize Forms, a less specific principle than the HPP. Besides, Minimize Forms is a performance principle. Optimality Theory has economy principles as well: STAY ('do not move') and TELEGRAPH ('do not spell out FCs). These OT constraints are ordered differently cross-linguistically, unlike the Head Preference and Late Merge ones.

The Head Preference Principle is relevant to a number of historical changes (Table 1.3): whenever possible, a word is seen as a head rather than a phrase. In this way, pronouns change from emphatic full phrases to clitics to agreement markers and negatives change from full DPs to negative adverb phrases to heads. This change is slow since a child learning the language will continue to encounter a pronoun as both a phrase and a head. For instance, coordinated pronouns are phrases as are emphatic pronouns. If they remain in the input, phrases will continue to be triggered in the child's grammar. In the case of pronouns changing to agreement markers, the child will initially assume the unmarked head option, unless there is substantial evidence that the pronoun is a full phrase.

<table>
<thead>
<tr>
<th>Specifier &gt; Head</th>
<th>Demonstrative pronoun <em>that</em> to complementizer</th>
<th>Negative adverb to negation marker</th>
<th>Adverb to complementizers (e.g. <em>till</em>)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specifier &gt; Head</td>
<td>Demonstrative pronoun to article</td>
<td>Adverb to aspect marker</td>
<td>Full pronoun to agreement</td>
</tr>
</tbody>
</table>

Table 1.3: Examples of the HPP

A practical issue here is how to distinguish between specifiers and heads. Sometimes, this is difficult and that is why they are reanalyzable by the language learner. Specifiers are full phrases and can be modified, coordinated, and occur in certain positions; a coordinated or modified element is never a head and head movement is usually recognizable. Table 1.4 summarizes this.

<table>
<thead>
<tr>
<th>Coordination</th>
<th>Specification/XP</th>
<th>Head/X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coordination</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Modification</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Movement</td>
<td>quite free</td>
<td>to head positions such as T and C</td>
</tr>
</tbody>
</table>
There is a second economy principle in early Minimalism (e.g. Chomsky 1995: 348). To construct a sentence, we need to select lexical items from the lexicon, put them together, or merge them, and move them. In Early Minimalism, Merge "comes ‘free’ in that it is required in some form for any recursive system" (Chomsky 2001: 3) and is "inescapable" (Chomsky 1995: 316; 378). Move, on the other hand, requires additional assumptions. This means that it is less economical to merge early and then move than to wait as long as possible before merging. This is expressed in (17).

(17) **Late Merge Principle (LMP):**
Merge as late as possible.

Principle (17) operates most clearly in heads. Thus, under Late Merge, (18a), with the auxiliary base generated in T, is preferable to (18b) with the auxiliary in a lower position and moving to T. See also Kayne (1999). The LMP accounts for the change from lexical to functional head or from functional to higher functional head frequently described in the grammaticalization literature (e.g. Heine & Kuteva 2002).

```
(18) a. TP
    T
    vP
    might

    v'
    ...

    v
    ...

    v'
    might

(18) b. TP
    T
    vP
    might
```

Late Merge also accounts for lexical phrases becoming base generated in the functional domain, e.g. certain Prepositional Phrases in Middle English as well as certain
Adverb Phrases. Van Gelderen (2009a) develops this idea in detail: from an original VP-adverbial PP, as in (19), to a frequent preposing, as in (20), to a reanalysis in the specifier of the higher CP in (21).

(19) *hu hit Hringdene æfter beorþege gebun hæfdon*
how it Ring-Danes after drinking lived had
‘how the Ring-Danes were doing after their drinking.’ (*Beowulf* 116-7)

(20) *Æfter pysan com Thomas to Cantwarebyri*
‘After this, Thomas came to Canterbury.’
(Chronicle A, Thorpe edition, entry for the year 1070)

(21) *for [æfterþan þet þe mon bið dead] me leið þene licome in þere þruh*
Because after-that that the man is dead they lay the body in the tomb
‘After the man is dead, they put the body in the tomb.’ (*Lambeth Homilies* 51: 4-5, Morris 1868)

Structure (22a) shows the more recent representation and (22b) the earlier one. Under LMP, (22a) is preferable.

---

9 For instance, there is a temporal change involved as well that is not immediately relevant here and therefore left out.
In Roberts & Roussou (2003), if a functional category chooses merge over move ($F^\text{merge}$ over $F^\text{move}$), the element that used to satisfy $F^\text{move}$ does not have to satisfy $F^\text{merge}$. It could be another lexical item. In my proposal, it is important that a lexical item has actually moved to the functional category. This is different from Simpson & Wu's (2002ab) Lateral Grammaticalization, where a reanalyzed item does not have to have been internally merged (i.e. moved).

Certain lexical items are ‘prone’ to a reanalysis under the LMP. For instance, non-theta-marked elements can wait to merge outside the VP (Chomsky 1995: 314-5). I will therefore argue that if, for instance, a preposition can be analyzed as having fewer semantic features and being less relevant to the argument structure, it will tend to merge higher (in TP or CP) rather than merging early (in VP) and then moving. How does the LMP work in practice? Assuming a lexicalist hypothesis in which a lexical entry "contains three collections of features: phonological ... semantic ..., and formal" (Chomsky 1995: 230), a LI such as the light verb *go* might have the semantic features [motion, future, location]. If *go* occurs with another verb, e.g. *bring*, one of the semantic features of *go*, in this case [future], need only be activated. Thus, a bi-clausal structure can be avoided.

The examples of grammaticalization given in Table 1.1 are all instance of Late Merge. I have added a few others in Table 1.5.

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>On, from P to ASP</td>
<td>VP Adverbials &gt; TP/CP Adverbials</td>
</tr>
<tr>
<td>Like, from P &gt; C (like I said)</td>
<td>Negative objects to negative markers</td>
</tr>
<tr>
<td>Modals: v &gt; ASP &gt; T</td>
<td>Negative verbs to auxiliaries</td>
</tr>
<tr>
<td>To: P &gt; ASP &gt; M &gt; C</td>
<td>PP &gt; C (for him to do that ...)</td>
</tr>
</tbody>
</table>

Table 1.5: Examples of the LMP

Chomsky (2001) uses the terms ‘external merge' for the initial merge (the traditional Merge) and ‘internal merge' when an element is merges for a second time (the traditional Move). Since Move is seen as a special case of Merge, i.e. Remerge, it is not less economical than Merge (see Chomsky 2001; 2005: 12). One could argue that (17) is still valid since the special Merge, i.e. internal Merge, requires additional steps. Traces are no
longer allowed; they would introduce new material into the derivation after the initial selection and therefore copies are included in the derivation, as in (10). Move/internal merge is not Move but ‘Copy, Merge, and Delete.’ Since the derivation contains more copies of the lexical item to be internally merged, and since those copies have to be deleted, (13) could still hold as an Economy Principle. In addition, Chomsky (2005: 14) suggests an important difference: external merge is relevant to the argument structure, whereas internal merge is relevant for scope and discourse phenomena. This establishes a crucial difference between the two operations.

It is also possible to think of syntax as inert and reformulate Late Merge in terms of feature change and loss. This is currently the thinking about linguistic variation: parametric variation is due to different features connected to lexical items. Starting with Chomsky (1995), the features relevant for and accessible during the derivation are formal. Formal features can be interpretable (relevant to the semantic interface) or uninterpretable (only relevant to move elements to certain positions). Interpretable features are acquired before uninterpretable ones, as argued in Radford (2000), but are later reinterpreted as uninterpretable, triggering the functional/grammatical system. The same happens in language change. Changes in negatives can be explained by arguing that their (initially) semantic features are reanalyzed as interpretable and then as uninterpretable, as in (23). Phrases like never have interpretable negative and phi-features that are probed by a probe in a functional category. Once the phrase is reanalyzed as a head (e.g. Old English ne ‘not’), another element is required. One could therefore argue there is a principle at work, as formulated in van Gelderen (2007; 2008ab; 2009c), namely (23).

(23) Feature Economy
Minimize the semantic and interpretable features in the derivation, e.g:

<table>
<thead>
<tr>
<th>Adjunct</th>
<th>Specifier</th>
<th>Head</th>
<th>affix</th>
</tr>
</thead>
<tbody>
<tr>
<td>semantic</td>
<td>[iF]</td>
<td>[uF]</td>
<td>[uF]</td>
</tr>
</tbody>
</table>
The change represented in (23) occurs in the case of the negative cycle as well as the subject cycle: the interpretable person (and gender) features of a full pronoun are reanalyzed as uninterpretable when they become agreement.

(24) **Subject Agreement Cycle**

\[
\text{emphatic} \quad \rightarrow \quad \text{full pronoun} \quad \rightarrow \quad \text{head pronoun} \quad \rightarrow \quad \text{agreement}
\]

\[
[i\text{-phi}] \quad [i\text{-phi}] \quad [u\text{-1/2}] \quad [i\text{-3}] \quad [u\text{-phi}]
\]

This is compatible with Chomsky's (1995: 230; 381) views on features: "formal features have semantic correlates and reflect semantic properties (accusative Case and transitivity, for example)." This makes sense if a language learner uses the semantic features in the derivation, these features turning into interpretable ones so to speak. Chomsky assumes that uninterpretable features need to be valued and I follow that but see Pesetsky & Torrego (2007) for the alternative that +/- valued is independent from +/- interpretable.

In (23) and (24), I connect heads with having uninterpretable features. I think that is correct and could be made to work. Due to the number of features a head can have, this is sometimes hard to be consistent about, however. A major problem in the approach I develop regarding Feature Economy is that the status, number, and use of features is still very fluid in Minimalism. Proliferation of features needs to be avoided. In the concluding chapter, I will briefly discuss this.

Schütze (1997; 2009) uses an Accord Maximization Principle which is very similar to (23) and Hicks’ Principle in (25) comes close to Feature Economy. He justifies Feature Economy as a component of merge (Hicks 2008: 220).

(25) **Maximize Featural Economy**

Establish dependencies via syntactic operations where possible

Unvalued features are more economical since they allow the relevant dependencies to be established ‘for free’ by syntactic means. This principle provides the rationale for the change from semantic to uninterpretable features. It also explains why there may be a
difference in Feature Economy where uninterpretable features of the probe and those of the goal are concerned. Those of the probe keep the derivation going.

Based on van Gelderen (2008e, 2009a), I briefly review how the grammaticalization of prepositions discussed earlier as (19) to (22) can be seen in terms of Feature Economy and Late Merge. In chapter 7, a slightly more elaborate version appears. A preposition such as *after* has semantic features (e.g. [time, order, past]) and phonological ones (two syllables, etc.). These are not accessible during the derivation, though [time] may be interpretable, as in Pesetsky & Torrego (2004). In addition, there are formal features, which are accessible during the computation and include categorial, Case, and phi-features, at least in Chomsky (1995: 230-2). Assuming that prepositions have unvalued phi-features, they value the Case of the DP in their domain. So, the Case of the DP is valued after agreement with an appropriate probe (I use ACC to show the Case features but nothing hinges on this choice).

(26)  PP

```
  P     DP
   after [uACC] 
   [u-phi]  [3S]
   [ACC]
```

Thus, there is a formal uninterpretable and unvalued feature [u-phi] that makes prepositions into probes (see also Baker 2008a: 112-4). This is the feature relevant for the derivation; other features are a burden on the computational system since they do not keep the derivation going. Language learners thus use (23) to eliminate [ACC] from the lexical item. With the interpretable [ACC] feature removed, the features of *after* are as listed in (27a), making it a C looking for interpretable phi-features from a Goal that doesn’t need to value its case, i.e. as in (27b).

---

10 In later work, following Marantz (1997), lexical items are seen not as specified for category but as roots that are nominalized or verbalized through Merge with an n or v.

11 This accounts for the difference between a preposition and an adverb: the former need an NP in its complement, the latter does not.
Thus, we have gone from (19) to (21). The same changes occur with *like* and *for* and a number of other prepositions. For instance, Higashiizumi (2006: 69-74) provides a good overview of how temporal prepositions are reanalyzed as causal. The uninterpretable, unvalued features of C will probe into the clause it c-commands and find a goal in the lower TP to value its phi-features. It is well-known that CPs (as subjects) trigger third person singular agreement on the verb. This is expected if the complementizer has phi-features (that are overt in many languages).

Chomsky (2004; 2007: 3-4) argues that we need to attribute as little as possible to UG and rely instead on principles not specific to the language faculty. Many Economy Principles, including (15), (17) and (23), fall into the latter category in that they reduce the computational burden.

Like the Head Preference Principle, Late Merge is argued to be a motivating force for linguistic change, accounting for the change from specifier to higher specifier and head to higher head. These Principles help language learners reanalyze their linguistic input. I have reformulated the LMP as a Feature Economy Principle; the same can be done for the Head Preference Principle. When phrases are reanalyzed as heads, they lose semantic and formal features, as we will see.

### 2.4 Cyclical Change

The two principles just mentioned, and their reformulations in terms of Feature Economy, account for grammaticalization and Feature Economy is responsible for cyclical change. Let's see in more detail what happens when we combine the effects of the HPP and the
LMP. Figure 1.1 shows how a Spec(ifier) can be reanalyzed as a head (through the HPP) and how the Specifier position can be filled by a phrase from a lower layer (LMP).

Figure 1.1: The Linguistic Cycle

This scenario works perfectly for changes where a negative object such as Old English *na wiht* 'no creature' becomes a specifier (LMP) and subsequently a head *not* of a NegP (HPP).

A stage not yet accounted for is the shift of the head to zero, as in the case of negative heads: Old English and Modern French *ne* and *n't* are null in many varieties of Modern English. The main reason for head deletion is head movement of other heads: the auxiliary moving via the Neg(ative) head to C may lead to Feature Syncretism (where one word has two functions). Words such as *won't* and Old English *nis* 'not-is' tend to be reanalyzed as expressing only one feature. See Faarlund (2008: 231) who argues for the principle in (28).

(28) **Null hypothesis of language acquisition**

A string is a word with lexical content.

He explains that "[i]n terms of acquisition and reanalysis, this means that the child misses some of the boundary cues, and interprets the input string as having a weaker boundary (fewer slashes, stronger coherence) at a certain point" (2008: 236), as in (29).

(29) ///>||>\
Roberts & Roussou (2003: 16; 200; 210) similarly rely on "simplicity" and "simpler representations" to help the acquisition process. These are fleshed out further (2003: 200ff; 210ff) and rely on feature syncretisms like Faarlund: if one lexical item "spell[s] out the features of two (or perhaps more) heads" (2001: 200), a reanalysis can take place since syncretism is not economical (2003: 201).

However, it is not clear that avoiding syncretism is most economical. One could argue that there are two possibilities, as stated in (30) and (31).

(30) Movement links two positions and is thereby economical; i.e. assume uninterpretable features and movement.

(31) Avoid syncretism; Iconicity is economical; i.e. assume interpretable features.

Feature Economy favors (30) since it keeps the derivation going. If you select two words from the lexicon with only interpretable features, they will not interact or merge.

As a result of (23), one needs ‘new’ features. Once the functional element has lost its semantic and interpretable features, it becomes a probe looking for an element to value its uninterpretable features. I assume it can value these with elements that are themselves inactive; otherwise renewal would not take place. Some elements are straightforward renewers: demonstratives have phi-features and adverbs temporal or spatial ones. A few are given in Table 1.6 and will be discussed in later chapters.

<table>
<thead>
<tr>
<th>Agreement:</th>
<th>Emphatic pronouns/nouns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case:</td>
<td>Demonstratives</td>
</tr>
<tr>
<td>Future:</td>
<td>Verbs of motion</td>
</tr>
<tr>
<td>Negative:</td>
<td>Minimizers/Negative DPs/Negative APs</td>
</tr>
</tbody>
</table>

Table 1.6: Examples of Renewal

---

12 For independent reasons, the form hallen is not used and ungrammatical, but Faarlund uses (29) for the morphology.
Representative cycles can therefore be seen as (32), rather than as (1) or (23). This indicates that there are two types of cycles for certain elements. In chapters 2, 3, 4, and 6, the change represented in (32a) is exemplified; in chapter 7, (32b) is relevant, and in chapters 5 and 8, both are.

(32) a. Adjunct > Specifier > Head > 0
   semantic    [iF]                        [uF]      --

b. Head > (higher) Head > 0
   [uF]                            [uF]
   [iF]

Cyclicity is mentioned only three times in Hopper & Traugott (2003). They talk about grammaticalization as a “continuously occurring phenomenon” (p. 124) and point out that the cyclical model is “extremely problematic because it suggests that a stage of a language can exist when it is difficult or even impossible to express some concept” (p. 124). Having a probe with uninterpretable features automatically look for another element avoids such a stage, I think.

Before ending this section on Economy, I’ll add a brief note on analogy which has been a major explanatory principle for hundreds of years and shows similarities to economy. Meillet (1912: 13) is famous for emphasizing both phonetic regularities and analogy in morphological paradigms in linguistic change. “La recherche des lois générales, tant morphologiques que phonétiques, doit être désormais l’un des principaux objets de la linguistique” 13. Within the morphological laws, he recognizes two principles, analogy and grammaticalization (1912: 130-1) where analogy is mainly restricted to paradigmatic regularity. After him many others have used this concept, e.g. Kurtyłowicz and Antilla. Kiparsky has claimed that the “direction of analogical change is optimization of the evaluation measure” (1974: 203-4) provided by UG. Thus, the child choosing between grammars that fit the data it encounters will choose the morphological

13 “Research of general laws, both phonetic and morphological, must from now on be one of the main goals of linguistics.” [EvG translation]
representation that is more regular. Most examples in Kiparsky (1974; 2000) involve lexical/paradigmatic relationships and are therefore different in range from the Economy Principles mentioned above. Lightfoot (1979: 347-8) notes that “analogy plays an extensive role in syntactic change” because UG puts constraints on the levels of opacity. Fischer (2007: 135-145) also invokes analogy as playing a very important role in language change. Her examples involve cases of grammaticalization such as going to reanalyzing as a future on analogy of the Auxiliary + bare infinitive pattern; “it in fact joins another token-set” (2007: 145). Analogical reasons have been criticized for their looseness and their arbitrariness but in Fischer’s Construction Grammar framework are less loose.

In this section, I provided background on Minimalism and Economy and showed how Feature Economy is responsible for the linguistic cycle. In the next section, we will see how these principles are at work in language acquisition.

3. Economy and Language Acquisition

Acquisition data are difficult to interpret. If the child uses the pronoun me, is that a full phrase (as in adult English), an N head, or a D head? In this section, with data taken from the CHILDES corpus, I first show that children use what look like heads rather than phrases (in accordance with the HPP), though this may be for more general reasons. I then argue that children also create grammatical/functional elements from lexical ones (the LMP), and use what looks like Feature Economy when they develop beyond the lexical word stage.

3.1 Head Preference

Children are well-known for using single lexical words in the early stages, as in (33), from the two-word and telegraphic stages.

(33) a. all gone (Allison 1:8, Bloom 1973)
b. walk school (Allison 1:8, Bloom 1973)
c. baby eat cookie (Allison 1:10, Bloom 1973)
d. sit down right here next truck (Allison 1:10, Bloom 1973)

There is no evidence of overt coordination in Allison's speech at age 1:10, as (34) shows. This is expected if only heads are used.

(34) horse cow
    ‘horse and cow’ (Allison 1:10, Bloom 1973)

The earliest pronoun produced is given in (35), a dislocated structure, the only one in the file for age 1.8.

(35) Momy you wiping (Allison 1:8, Bloom 1973)

No files for Allison were collected between 1:10 and 2:4. At age 1:10, there are no pronouns; at age 2:4, pronouns are used correctly, as (36) shows.

(36) a. she (i)s not eating snack right there (Allison 2:4, Bloom 1973)
    b. I'm gonna sit on here (Allison 2:4, Bloom 1973)
    c. oh I don't want drink it (Allison 2:4, Bloom 1973)

The first use of demonstratives with nouns,\(^\text{14}\) clearly making a head into a phrase, is eight months later than (35), namely (37).

(37) a. want eat my snack (Allison 2:4, Bloom 1973)
    b. where the chair mommy (Allison 2:4, Bloom 1973)
    c. it a puddle from juice I spilled (Allison 2:4, Bloom 1973)

\(^{14}\) There are earlier instance of modification, as in (i) and (ii), but these are compounds:
    (i) there baby cup (Allison 1:10, Bloom 1973)
    (ii) more apple juice (Allison 1:10, Bloom 1973)
This suggests that there is an initial preference for heads. The same is true for other children. Peter (Bloom 1970) uses both *I* and *me* at age 2:1.

(38)  
   a. I writing  
   b. I'm writing

(39)  
   a. me found it # I find it  
   b. me take it off # no me take it off  
   c. me too # go home

(all from Peter 2:11, Bloom 1970)

Even though *I* and *me* could be phrases in (38) and (39), there are no instances of coordination of (pro)nouns.

If we consider a more complex phenomenon, i.e. the choice between a *wh*-relative (in the specifier position in adult language) and no relative or *that* (a head), the same preference for heads holds. Using data from four children, Diessel (2004: 137) shows that when the children start to produce relative pronouns, out of a total of 297 relative clauses, 165 have *that*, 6 have *who* (all by one child), and 126 have no marker. This shows that children avoid phrases completely. The 6 instances of *who* can be argued to be heads since *whom/to who* do not occur. The percentages are: 56% *that*, 42% zero, and 2% *who*.

In the CHILDES- Kuczaj corpus (Kuczaj 1976), Abe produces 82% *that* relatives, as in (40) and 18% *wh*-relatives. There is no evidence that the *wh*, as in (41), is not a head, however, since they are never full phrases.

(40)   a dragon *that* was this little (Abe, 4;0.16)

(41)   You know the round part *where* they dig (Abe 4; 1.5)

Abe's use of *who* is sometimes in direct imitation of a caregiver, as in (42).

(42)   FAT: no the person *who* eats em.  
   CHI: the person *who* eats em? (Abe, 3: 1.8)
So, one could argue that adults may already be avoiding phrases. Researchers estimate that in adult speech *that* (or zero) relatives appear 80-90% of the time, as opposed to 10-20% in written genres.

### 3.2 Late Merge and Feature Economy

Lexical categories are acquired before grammatical ones. For instance, Clark (1971) and Friederici (1983), have shown that lexical prepositions are acquired before grammatical ones, and more recently Littlefield (2006: 148-9) has done the same. Among the lexical ones, spatial prepositions are acquired before temporal ones (Grimm 1975; Tomasello 1987). I demonstrate in this section that children seem to turn lexical categories, such as prepositions, into grammatical ones, such as complementizers, seemingly without explicit input. This means they are employing Feature Economy, reanalyzing semantic features as grammatical ones.

To argue decisively that language change is determined by principles at work in acquisition, it would be great if children completely abandoned using the lexical item with semantic features in favor of using the grammatical item. However, most languages have words that are ambiguous between lexical and grammatical status, e.g. prepositions and complementizers. Pronouns may be an instance of a clear change from lexical use to grammatical use. In what follows, I provide instances of words ending up doing double duty, so to speak: first as lexical words and then as both lexical and grammatical words.

Josefsson & Håkansson (2000: 398) argue that Swedish "children first acquire the PP and then, directly after that the subordinate clause:" (43) first and then (44).

(43) *precis som en kan* / *som en kanin*
   just like a rab/ like a rabbit

(44) *grisen, den som heter Ola*
   pig that who is-called Ola
   (Embla, 27 months, both from Josefsson & Håkansson 2000: 410)
According to Josefsson & Håkansson, the children do not start using complementizers at all until they have reached a 90% use of prepositions in obligatory contexts.

In the CHILDES-Kuczaj corpus, English *like* and *for* are used similarly as in Swedish. Initially, the child uses *like* as a lexical category only, as in (45) to (48), and later adds the complementizer, as in (49) and (50).

(45)  *like* a cookie (Abe, 3.7.5)
(46)  no the monster crashed the planes down *like* this *like* that (Abe, 3.7.5)
(47)  I wan(t) (t)a show you something # I mean *like* this thin ? (Abe, 3.7.5)
(48)  I feel *like* having a pet do you? (Abe, 4.8.20)
(49)  watch it walks *like* a person walks. (Abe, 4.9.19)
(50)  Daddy # do you teach *like* you do [//] *like* how they do in your school? (Abe, 4.10.1)

This is all the more interesting in that the caregivers in the transcript only use *like* as a C after *sound* or *look*, as in (51). Although the child may have heard the use of *like* as a C in another context, it could also be a spontaneous invention by the child guided by Late Merge.

(51)  Abe's father:  it looks *like* some birds have eaten some of the bread. (Kuczaj file 206)

The situation with *for* is similar to that of *like* although *for* is used by the caregivers both as complementizer and preposition. Initially, the preposition is used to express benefactor thematic role, as in (52) and (53). Later, this meaning is extended to time, as in (54), and more abstract use in (55) and (56).

(52)  Mom # this white one *for* me? (Abe 2.7.18)
(53)  this picture is mine *for* myself (Abe 2.7.18)
(54)  how long you grow up *for* a minute (Abe 2.9.27)
(55)  Mom # I'm glad you are making a rug *for* out in the hall. (Abe 2.8.14)
The first target-like C is in (57). Another month and a half later, Abe produces his second one (at least on the transcripts we have), namely (58), and three months later the third one (59).

(57) yeah and I said I was waiting and waiting for you to come and I [/] (Abe, 3.2.1)
(58) this crocodile was standing around waiting for someone to drop around and what did he see when he saw it? (Abe, 3.3.18)
(58) yeah maybe it's time for it to rain we'll have a storm. (Abe, 3.6.26)

Abe continues to use for as a preposition, as in (59), as well as a complementizer in (60).

(59) a. two Dad # how come some people have cookies for lunch sometimes? (Abe 3.7.5)
   b. ok then we could go way # way # way down from the stairs and dig for that rock I saw (Abe 3.7.5)
   c. because people don't use hatchets for hunting butterflies (Abe 3.7.5)
(60) it's not too high up # but I'm waiting for Silver to get ready. (Abe 5.0, file 210)

The total number of for complementizers before infinitives, as in (61), by adults in this corpus is 35. This makes the situation different from that of like where no such input exists.

(61) Abe's father: ok # go ahead and call me when you're ready for me to play are you still playing by yourself?

The total number of sentences with a complementizer for, such as (60), that are uttered by Abe is 21. There are a few non-target like sentences like the one in (62).
(62) CHI: ropes.
MOT: what for?
CHI: for cows to catch with.
FAT: for what?
CHI: for cows to catch with. (Abe 3.0.7)

The HPP predicts that if there is evidence for a pronoun to be both a phrase and a head, the child/adult will analyze it initially as a head unless there is also evidence in the grammar (e.g. from coordination) that pronouns function as full DPs. Feature Economy helps the child reanalyze a lexical element as a grammatical one. For instance, when for and like are used as complementizers, they ‘exchange' interpretable for uninterpretable features. All cross-linguistic variation is therefore in the lexicon. I will now turn to some other evidence for Feature Economy from acquisition data.

Radford (2000) argues that interpretable features are acquired before uninterpretable ones, but are later reinterpreted as uninterpretable, triggering the functional/grammatical system. His data show that pronouns appear late; nouns carrying interpretable Case and phi-features are used instead. See (33), repeated here as (63).

(63) baby eat cookie (Allison 1:10, Bloom 1973)

The Case features are then reanalyzed as uninterpretable, according to Radford.

Radford also argues that verbs with –ing, as in (38a), initially carry interpretable aspect features, later reanalyzed as uninterpretable. One can debate the correct analysis (e.g. it can be argued that the features of -ing remain interpretable in the adult grammar and that the auxiliary carries the uninterpretable features), all evidence seems to suggest that children start out with interpretable features. It would be good to investigate all instances where currently uninterpretable features are postulated and see how the child represents these.
If the acquisition picture that has been sketched is accurate, acquisition (and maturation) of features mirrors the grammaticalization process and provides evidence that Feature Economy, as in (23), is a cognitive process driving language change.

4. External factors to language change

Many historical linguists see language change as determined by two kinds of factors. There are internal reasons for change, such as those instigated by the Economy Principles or by ‘Ease’, as in Jespersen (1922: chapter 14, §6), which deal mainly with the articulatory ease of pronunciation. Children acquiring a language use these principles to analyze their input. This probably also happens in the case of creoles. As Mufwene (2001) claims, there is a feature pool of constructions that learners choose from for various reasons (internal as well as external ones). There are also external factors for language change such as a need on the part of speakers to be innovative, polite, creative, or conservative. I will discuss the latter briefly in this section since the remainder of the book emphasizes internal factors.

External factors include pragmatic ones. The urge of speakers to be innovative may introduce new, loosely adjoined elements into the structure. Hagège (1993: 153) uses the term Expressive Renewal. Speakers may want to be explicit and therefore chose full phrases rather than heads. One source of new specifiers and words is borrowing. Heine & Kuteva (2005: 3) give examples of wh-interrogatives being expanded to relatives in Tariana under the influence of Portuguese, where interrogatives and relatives share the same form. The same pattern occurred in the history of English: wh-pronouns were used in questions but were later extended to relative contexts under the influence of French (see van Gelderen 2004: 88). Heine & Kuteva (2005: 73) give many other examples, e.g. Tariana speakers renewing their evidentials by using Portuguese expressions such as eu vi ‘I saw.’ Lehmann (2002: 20) provides other examples of renewal: Latin antē ‘before' gets an ab prefix to become French avant; Latin ille, which had become reduced, was reinforced by *eccu illu to Italian quello, as we will see in Chapter 6.

Another external factor is the need of society to be conservative and prescriptive. This may stop change altogether. In the chapters that follow, we will examine some
examples. For now, I will mention stranding and negatives, where prescriptive rules are very strong. Considering Economy, a principle such as (64) is expected.

(64) **Stranding Principle**  
Move as little as possible.

This principle has been used to explain why speakers in English typically front the DP, as in (65) to (68), rather than the full PP in (69) and (70) or the full Quantifier Phrase in (71) and (72). I have given examples from Modern and Middle English.

(65) **Who** did you talk to **who**?
(66) **Quilc men mai get wundren on**  
‘which men may yet wonder about’  
(Genesis & Exodus 3715, Morris 1865, from Denison 1993: 132)
(67) **The children** might have been all the **children** reading happily.
(68) **The roote of ri3twis men** shal not ben al the roote of ri3twis men moued.  
(1382 Wyclif *Prov.* xii. 3 from the OED)
(69) **To whom** did you talk to **whom**?
(70) **fro hwat** he scal his sunne uor-saken (HC ME1)
(71) **All the children** might have been all the **children** reading happily.
(72) The sterres also and all the **fyrmamente** she maketh to retorne abacke. (from the OED, 1490 Caxton *Eneyd* xxiii. 87)

Preposition stranding in English, as in (65) and (66), started in the 13th century (Denison 1993: 125 ff.). It is preferred under (64) and it is estimated that in speech 86% of prepositions are stranded while in writing only 7% are. This difference between spoken and written data points towards strong prescriptive pressure. According to Diessel (2004) and Snyder (2007), young English speakers produce only stranded constructions, as in (73) and (74), but as adults they are taught to take the preposition along, as in (75).

(73) **where's** the bolt go **in** (Peter 2:1, Bloom 1970)
those little things that you play with (Adam 4:10, from Diessel 2004: 137).

things with which you play.

The data on Quantifier-stranding, or Quantifier-float, are not as straightforward. Unexpectedly, pied piped instances of the quantifier all, as in (71), are quite popular in English (see Wenger 2005), so something else must be going on. Floating quantifiers are very infrequent in child data.

Bullokar's grammar from 1586 contains stranded prepositions, but one century later, most grammarians prescribe against its use: Poole (1646: 38), Dryden (1691 letter 17), Lowth (1762), and Coar (1796). Yáñez-Bouza (2004; 2007) finds that these prescriptivists indeed had an influence on the language. On the basis of an analysis of 285 different works on grammar written between 1700 and 1800, she argues that the proscription against stranding goes back to the 17th century and has a real influence on usage, especially in the 18th century. Other languages may have prescriptive pressure as well (see Oppenrieder 1991). As chronicled in great depth in Fleischer (2002), in many varieties of German preposition stranding is frequent though some describe it as "älter oder umgangssprachlich" (137). It is also common in North-American varieties of French; see e.g. Roberge & Rosen (1999).

Some languages do not strand prepositions even though they have no obvious prescriptive tradition. There is something in their grammars that disallows preposition stranding. Van Riemsdijk (1978) and Hornstein & Weinberg (1981) discuss this phenomenon early on. For preposition stranding to be allowed, the stranded preposition and the verb need to be adjacent: talk and to in (65). This is also the case in (76), from Jacaltec, a Mayan language, and in many languages of the Kru family, as in (77), from Gbadi.

(76) mac chach to munil yin
    Who you go work for

    ‘Who are you going to work for?’ (Craig 1977: 15)

(77) tablE yI wa kE-IO lilE kIU jiIE
    table WH they FUT-FOC food on put
‘It is the table they will put the food on' (Koopman 1984: 54, but tones left off).

Law (1998) offers another explanation: if a language has D-to-P incorporation, preposition stranding is not allowed. This holds in many of the Romance languages as well as in Dutch and German (see chapter 3, section 7). Formulating the exact grammatical constraint against preposition stranding is less important here. I just wanted to show that English speakers allow it even though prescriptive grammar does not.

As for negatives, at least since the 18th century, there has been such a prohibition against multiple negatives to express sentential negation that, even though an overt negative object in (78) with a negative n't would be expected, this will not happen in most standard varieties of English. Thus, the negative cycle is not continuing in the way one expects it.

(78) I can't do nothing for you either, Billy.
    (Ken Kesey, One flew ... p. 118)

To renew the weakened negative –n’t, never is used instead, as in (79). Even in (79), prescriptive rules say to only use never when you mean ‘at no time, not ever.’

(79) a. I never saw the outline of a plane, just this incredible ball of flame.
    (BNC - CH2 12700)

b. No, I never see him these days (BNC - A9H 350)

The adverb never seems in the same position as not/n't because it typically follows the auxiliary and precedes the main verb. It must, however, be in a Specifier position since the auxiliary moves across it and since it can be modified as in ‘almost never.’

Jespersen (1922) formulates this tension between internal and external reasons for change as a ‘tug-of-war' and says: "the correct inference can only be that the tendency towards ease may be at work in some cases, though not in all, because there are other forces which may at times neutralize it or prove stronger than it" (1922: 262). It may be that the new utterance is hard to understand, for instance, and is not economical for the hearer. Lightfoot (1979) distinguishes between "changes necessitated by various principles of
grammar" and those "provoked by extra-grammatical factors." Von der Gabelentz (1891/1901: 251/256) uses "Deutlichkeit" ('clarity') and "Bequemlichkeit" ('comfort') as important (competing) factors, as in the well-known passage in (80).

(80) Nun bewegt sich die Geschichte der Sprachen in der Diagonale zweier Kräfte: des Bequemlichkeitstriebes, der zur Abnutzung der Laute führt, und des Deutlichkeitstriebes, der jene Abnutzung nicht zur Zerstörung der Sprache ausarten lässt. Die Affixe verschleifen sich, verschwinden am Ende spurlos; ihre Funktionen aber oder ähnliche drängen wieder nach Ausdruck. Diesen Ausdruck erhalten sie, nach der Methode der isolierenden Sprachen, durch Wortstellung oder verdeutlichende Wörter. Letztere unterliegen wiederum mit der Zeit dem Agglutinationsprozesse, dem Verschleife und Schwunde, und derweile bereitet sich für das Verderbende neuer Ersatz vor ... ; immer gilt das Gleiche: die Entwicklungslinie krümmt sich zurück nach der Seite der Isolation, nicht in die alte Bahn, sondern in eine annähernnd parallele. Darum vergleiche ich sie der Spirale. (von der Gabelentz 1901: 256)\footnote{\textit{The history of language moves in the diagonal of two forces: the impulse toward comfort, which leads to the wearing down of sounds, and that toward clarity, which disallows the wearing down to destroy the language. The affixes grind themselves down, disappear without a trace; their functions or similar ones, however, require new expression. They acquire this expression, by the method of isolating languages, through word order or clarifying words. The latter, in the course of time, undergo agglutination, erosion, and in the mean time renewal is prepared: periphrastic expressions are preferred ... always the same: the development curves back towards isolation, not in the old way, but in a parallel fashion. That's why I compare them to spirals} (my translation, EvG).}

For many of the early 20\textsuperscript{th}-century (structuralist) approaches, the emphasis on Ease implies an emphasis on performance factors, or E-language, e.g. Hawkins (2004). In the approach used in this book, I explain some of the 'Ease' principles as part of the I-language.

Merge, I assume, is universally available since it is a UG principle. Morphology differs from language to language. If we believe that, at the Conceptual-Intentional (CI) interface, both thematic and discourse/scope information have to be marked, we need morphology or position to indicate this. If Merge was the crucial evolutionary jump, morphology and other options come later. These various options arise through grammaticalization, and I assume formal principles account for these changes. Differences between languages arise because they are in different stages of a particular cycle.
5. **Typology and parameters**

Three types of languages are often recognized: isolating (Chinese, creoles), dependent-marking (Korean, Malayalam), and head-marking (Navajo). They represent the different ways that semantic, grammatical, and pragmatic (definiteness/specificity) information may be coded in one language. In some languages, semantic roles are marked through dependent-marking (Case or adpositions), grammatical relations through agreement, and pragmatic information through word order or articles, but in most cases these overlap. The result is a varied morphology, as can be seen in Table 1.7, and which I work out in more detail in chapter 5.

<table>
<thead>
<tr>
<th></th>
<th>Semantic</th>
<th>Grammatical</th>
<th>Discourse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adpositions</td>
<td>yes</td>
<td>(some)</td>
<td>(some)</td>
</tr>
<tr>
<td>Case-inherent</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Case-structural</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Agreement</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Aspect</td>
<td>no</td>
<td>(some)</td>
<td>yes</td>
</tr>
<tr>
<td>D</td>
<td>no</td>
<td>(some)</td>
<td>yes</td>
</tr>
<tr>
<td>&quot;word order&quot;</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

Table 1.7: Morphological and Syntactic Markers

In this book, I will ascribe the difference between head-marking and non-head-marking languages to different stages of the agreement cycle. As I show, the setting changes fast in some languages without too many other characteristics changing. For instance, the Northern Athabascan languages lack object polysynthesis whereas the Southern ones have it; otherwise, the two are quite similar. This is a problem for approaches such as Baker’s (2001) hierarchy of macroparameters. Figure 1.2 provides a simplified version of Baker’s hierarchy.

| Polysynthesis | | |
The choice is between polysynthesis and no polysynthesis. Polysynthetic languages treat adjectives as nouns or verbs, and that is the second parameter to set. Non-polysynthetic languages have many parameters to set. One familiar parameter is headedness. It is clear that there are many problems with this, as pointed out in Baltin (2004: 551). The main objection is that the approach has a flavor of arbitrariness and is difficult to envision as part of UG. How would this have come about as an evolutionary shift?

It is possible to rephrase Baker’s macroparameter hierarchy in terms of features and see differences between languages as different feature choices, as is common in Minimalism. An attempt is made in Figure 1.3, which will be added to in Chapter 9.

---

16 Baker (2008a: 39; 153) argues that T is not a probe in Japanese, which is similar to not having probing phi-features.
This means that languages could be head-marking, dependent-marking, both, or neither. It goes against the current thinking that both phi- and Case-features are relevant for all languages. Other attempts involving features are found in Biberauer & Richards (2006), Richards (2008a), and Roberts & Holmberg (to appear).

In approaches such as these, the choices are not deep and are feature-based. This means that the child gets hints to pay attention to features, which would have to be part of UG. The differences expressed in Figure 1.3 are also in accordance with the view that parameters are only relevant to lexical items.

6. Methodological and practical issues

To find evidence of cycles and cyclical change, I have read many grammars (and am grateful to the careful work of the authors), but haven’t done that in a typologically and geographically systematic way. Where relevant, I have consulted the online World Atlas of Language Structures (www.wals.info) to get an idea of where certain phenomena might be found. Since I use a variety of languages and stages of languages, the examples I found constitute only the tip of the proverbial iceberg and are heavy on Indo-European and languages of North America. For contemporary languages, I use data mainly from grammars, the internet, and corpora. This brings up the issue of competence vs. performance, I- vs. E-language. Finding a pattern in a spoken corpus shows that there is something systematic going on: repeatedly finding shouldof and shoulda in contemporary texts (as well as those from the 15th century) indicates that something interesting is happening with modals and perfect auxiliaries. Similarly, if first (and second) person pronouns are 'repeated' more than third person ones, this indicates an important difference between those persons in the language. This difference cannot be attributed to performance but must be determined by the I-language.

The corpora used include the 100-million word British National Corpus or BNC (thetis.bl.uk), the fairly small French Corpus d'entretiens spontanés or CdES (www.uwe.ac.uk/hlss/llas/iclru/corpus.pdf), the Helsinki Corpus of older English texts or
HC (see Kytö 1993 for a description of this corpus), and the 2-million word Corpus of Professional Spoken American English (www.athel.com). The last one is abbreviated here as CSE and contains transcriptions of committee meetings (COMM), faculty meetings (FACMT), and White House briefings (WH). With many of these texts, I have used the concordance program MonoConc. I have also made use of individual electronic texts, made available by the Oxford Text Archive and the Dictionary of Old English project (DOE) at the University of Toronto. The latter is wonderful in that it allows one to search all of Old English. For languages such as Norwegian, Swedish, Dutch, Urdu/Hindi, and French, I have used Google searches. The choice of which to use depends on the size of the query. For instance, if looking at all instances of me, it is more convenient to select a smaller corpus.

Mostly, I provide bibliographical references for the primary sources (e.g. King Alfred's works and the Poetic Edda), but in some cases I use only electronic resources (e.g. Jane Austen’s and Ken Kesey's novels). Therefore, if the reference can be found easily, e.g. on the internet, it will not be listed in the bibliography. I did not want to clutter up the bibliography with works I use only once or twice as example sentences.

I simplify glosses where appropriate but have kept the glosses as much as possible as they were in the original source. When discussing negation and the agreement on the verb or demonstrative is not relevant, I leave that information out. Once in a while, I have left special characters out, e.g. tones in Chinese sentences, since native speakers say they can get the meaning without them in a full sentence. Data from corpora and google-searches often incorporate unusual spellings or punctuation. I have left these as they were in the original.

It is often very hard to decide which name of a language to use. I have been pragmatic rather than always politically correct simply because it is impossible to be the latter. This results in inconsistency. For instance, Chipewyan is an older name that is well known to linguists but the speakers prefer Dene Súłiné (even though they are not in agreement about the spelling of it). I have therefore used Chipewyan/Déné Súliné. I have tried to use language names that were explicitly chosen by its speakers, e.g. Tohono O’odham, instead of Papago, and Athabascan (with b and c) rather than Athabaskan. Other problems arise using umbrella terms like Persian, French, Mandarin, Urdu/Hindi, and of course English.
7. Outline

This book consists of four parts. The first two examine how arguments are marked and how changes in this marking proceed in a cyclical fashion. Part I, which contains three chapters, deals with a cycle of head-marking, i.e. increase and loss of subject and object agreement. Part II discusses dependent-marking, sometimes called Case. It has two chapters, one on the origin of Case, the other on definiteness. Part III looks at how temporal, modal, aspectual and negative information is expressed and how this is cyclical as well. Part IV contains two chapters on how cycles are important for work in typology and language evolution.