

# SYNTACTIC CHANGE

## A MINIMALIST APPROACH TO GRAMMATICALIZATION

IAN ROBERTS

*Downing College, University of Cambridge*

and

ANNA ROUSSOU

*University of Patras*



**CAMBRIDGE**  
UNIVERSITY PRESS

published by the press syndicate of the university of cambridge  
The Pitt Building, Trumpington Street, Cambridge CB2 1RP, United Kingdom

cambridge university press  
The Edinburgh Building, Cambridge, CB2 2RU, UK  
40 West 20th Street, New York, NY 10011–4211, USA  
477 Williamstown Road, Port Melbourne, VIC 3207, Australia  
Ruiz de Alarcón 13, 28014 Madrid, Spain  
Dock House, The Waterfront, Cape Town 8001, South Africa  
<http://www.cambridge.org>

© Ian Roberts, Anna Roussou 2003

This book is in copyright. Subject to statutory exception  
and to the provisions of relevant collective licensing agreements,  
no reproduction of any part may take place without  
the written permission of Cambridge University Press.

First published 2003

Printed in the United Kingdom at the University Press, Cambridge

*Typeface* Times 10/13 pt.      *System* L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub> [tb]

*A catalogue record for this book is available from the British Library*

ISBN 0 521 79056 5 hardback

# Contents

---

<i>Acknowledgements</i>	<i>page xi</i>
<b>Introduction</b>	<b>1</b>
<b>1 Parameters, functional heads and language change</b>	<b>9</b>
1.1 Introduction: the logical problem of language change	9
1.2 Functional categories	17
1.3 The nature of parameters: interface interpretation of functional categories	27
1.4 Conclusion	33
<b>2 T elements</b>	<b>35</b>
2.0 Introduction	35
2.1 From verb to auxiliary: the development of English modals	36
2.2 Romance futures	48
2.3 The Greek future	58
2.4 Conclusion	71
<b>3 C elements</b>	<b>73</b>
3.0 Introduction	73
3.1 From complementizer to particle: the case of Greek <i>na</i>	74
3.2 From adverb to particle: Southern Italian <i>mu</i>	88
3.3 The infinitival marker <i>to</i> in English	97
3.4 The English complementizer <i>that</i>	110
3.5 From verb to complementizer: serial verb constructions	121
3.6 Conclusion	129
<b>4 D elements</b>	<b>131</b>
4.0 Introduction	131
4.1 Determiners out of demonstratives	131
4.2 Negative words and related expressions	136
4.3 Wh-elements	161

x *Contents*

4.4	Universal quantifiers	167
4.5	Clitic agreement: Northern Italian dialects	175
4.6	Affixal agreement: Welsh and Indo-European	186
4.7	Conclusion	192
<b>5</b>	<b>Theoretical consequences</b>	<b>194</b>
5.0	Introduction	194
5.1	A general characterization of grammaticalization	195
5.2	Grammaticalization and the theory of language change	202
5.3	On the nature of functional categories	218
5.4	Conclusion	234
	<i>References</i>	237
	<i>Index of languages</i>	257
	<i>Index of names</i>	260
	<i>Index of subjects</i>	264

# 1 *Parameters, functional heads and language change*

---

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

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

- (1) a. Parla italiano.  
      ‘He/she speaks Italian.’  
      b. \*Speaks Italian

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

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

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

- (2) a. Max<sub>i</sub> bewondert zichzelf<sub>i</sub>/\*zich<sub>i</sub>.  
       ‘Max<sub>i</sub> admires himself<sub>i</sub>.’  
       b. Jan<sub>i</sub> liet mij voor zich<sub>i</sub>/zichzelf<sub>i</sub> werken.  
           John made me for him work  
           ‘John<sub>i</sub> made me work for him<sub>i</sub>/\*himself<sub>i</sub>.’

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

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

(1b) (Rizzi 1986a). Roughly speaking, children have to determine, on the basis of experience, whether their language has the English-type or the Italian-type AgrS. Similar considerations extend to (2). Questions that remain open in current research include the characterization of the class of functional heads and the definition of the ways in which functional categories can be parameterized. The latter point is discussed in section 3 of the present chapter. We consider the former question in more detail in Chapter 5, where we will make some suggestions based on the evidence provided by grammaticalization.

According to what we have said so far, the acquisition of syntax is viewed as the process of parameter setting. Within this framework of assumptions, syntactic change can be viewed as change in the parametric values specified for a given language. In other words, parameter values can change as a function of time. We can in fact observe this very easily by comparing the Modern Romance languages with Latin in respect of word order. Latin word order was rather free, but object–verb order clearly predominated; on the other hand, the Modern Romance languages are all verb–object. The contrast is illustrated in (3), with Italian representing Modern Romance:

- (3) a. Ego . . . apros tres et quidem pulcherrimos cepi. (Pliny the Younger)  
(Object) (Verb)  
I boars three and indeed very-beautiful have-taken.  
b. Io . . . ho preso dei cinghiali, tre e anche bellissimi.  
(Verb) (Object)  
'I have taken three and indeed very beautiful boars.'

Thus, if there is a parameter determining the relative order of verb and direct object, its value has changed in the development of Latin into Romance. The central issue for diachronic syntax in the context of Principles and Parameters theory is accounting for how and why this can happen.

Following a view that has been developed in terms of recent linguistic theory, primarily by Lightfoot (1979, 1991, 1998), we assume that parameter change is an aspect of the process of parameter setting. A change is initiated when (a population of) learners converge on a grammatical system which differs in at least one parameter value from the system internalized by the speakers whose linguistic behaviour provides the input to the learners. As the younger generation replaces the older one, the change is carried through the speech community. Of course, many social, historical and cultural factors influence speech communities, and hence the transmission of changes (see Labov 1972, 1994). From the perspective of linguistic theory, though, we abstract away from these factors and attempt, as far the historical record permits, to focus on change purely as a relation between grammatical systems.

The assumption that parameter change is an aspect of the process of parameter fixation raises an important issue for language acquisition. The issue is summed up in the following quotation from Niyogi and Berwick (1995):

it is generally assumed that children acquire their...target...grammars without error. However, if this were always true,...grammatical changes within a population would seemingly never occur, since generation after generation children would have successfully acquired the grammar of their parents. (Niyogi & Berwick 1995:1)

As the above quotation shows, the standard paradigm for language acquisition is not immediately compatible with the observation that grammatical systems change over time. To be more precise, it is generally assumed that language acquisition is a deterministic process: its final state converges with the target grammar that acquirers are exposed to. However, if convergence is always guaranteed, then the crucial question is how changes can ever take place. Clark and Roberts (1993, 1994) refer to this issue as the logical problem of language change, and sum it up as follows:

if the trigger experience of one generation, say  $g_l$ , permits members of  $g_l$  to set parameter  $p_k$  to value  $v_i$ , why is the trigger experience produced by  $g_l$  insufficient to cause the next generation to set  $p_k$  to  $v_i$ ? (Clark & Roberts 1994:12)

The simple answer to this question, which again goes back to Lightfoot (1979), is that  $v_i$  is unlearnable. In this case language acquirers have to revert to some other parametric option, thus triggering a change in the system. This way, the new setting for parameter  $p_k$  amounts to parameter resetting in comparison with the target grammar. If this is correct, we have to weaken and refine the notion of determinism, along the following lines: language acquisition is deterministic to the extent that all parameters have to be set. This allows for  $p_k$  to receive a different value from that found in the input, therefore making space for language change. This of course does not imply that changes have to take place; indeed, most of the time convergence is 'successful' in that children arrive at the same parameter values as their parents – this is reflected by Keenan's (1996) principle of inertia (see also Longobardi 2001a). A change occurs when the trigger experience for a parameter setting provided by the input has become obscure or ambiguous. This can happen in a variety of ways, for example through language contact, morphophonological erosion, etc. Fleshing this idea out requires us to develop an account of the relation between the learner and the trigger; it also requires us to be very precise about the nature and format of parameters. We



will discuss parameterization in section 1.3; here we will focus on the relation between the learner and the trigger.

The logical problem of language change interacts with the logical problem of language acquisition. For the latter, the question is how children succeed in setting the parameters correctly on the basis of the input they receive, given that this input may be insufficient and degenerate (see the ‘poverty of stimulus’ argument of Chomsky 1986a). If by ‘correctly’ we mean complete matching with the adult setting, then the logical problems of language acquisition and language change become contradictory. If, however, by ‘correctly’ we mean simply fixing a value consistent with the trigger experience, as suggested above, then the contradiction does not arise. Let us call this the weakly deterministic view of language acquisition: the goal of acquisition is to fix parameter values on the basis of experience – all parameter values must be fixed, but there is no requirement for convergence with the adult grammar (although this happens most of the time).

The relationship between the learner and the trigger can be thought of as mediated by a device which takes experience as input and produces parameter values as output. The trigger experience is naturally thought of as consisting of sets of sentences (cf. Clark & Roberts 1993, Gibson & Wexler 1994, among others). Lightfoot (1998) and Dresher (1999) argue that learners use input forms as ‘cues’ for setting parameters. The trigger in this case is not sets of sentences but fragments of utterances (partial structures) (cf. also Fodor 1998). For Dresher (1999) each parameter has a marked and a default setting, and comes with its cue, as part of the UG specification of parameters. Lightfoot (1998:149), however, takes a much stronger view and argues that ‘there are no independent “parameters”’; rather, some cues are found in all grammars, and some are found only in certain grammars, the latter constituting the points of variation’. Let us illustrate this with the loss of the verb-second (V2) phenomenon in Middle English. The presence of exactly one constituent other than the subject in immediately preverbal position is a cue for the learner that a given language is V2. According to Lightfoot (who follows Kroch & Taylor 1997), the Northern dialects of Middle English had a V2 grammar, which at some point ceased to exist – Modern English is not V2, as the grammaticality of sequences like *Yesterday John left* shows. Lightfoot proposes that the change was triggered by the following: (a) interaction with speakers of Southern dialects which didn’t have obligatory V2 and also didn’t treat subject pronouns as clitics, so the *XP–subject pronoun–V* sequence in the input was evidence against a positive setting for the ‘V2 parameter’; (b) the independent loss of all verb movement operations,

making verb movement to the second position impossible, pre-empting many V2 orders. In this way, the occurrence of the V2 cue was considerably reduced, leading to the consequent loss of V2. This approach, however, seems to involve circularity. It appears that V2 was lost because it was not cued, and that the cue was lost because V2 was undermined (owing to the factors given). It is not clear what the notion of cue is really explaining here; if we omitted it from our account, we would nevertheless have at least a plausible description of how V2 was lost. Also, Lightfoot's approach seems to involve a category mistake: cues are fragments of the trigger experience, sequences such as  $XP-V$  in the case of V2. But parameters are abstract properties of grammars, features of part of an individual's mental representation (his/her I-language). Although the notion of cue is useful, it must be kept distinct from the notion of parameter. Finally, Lightfoot's approach is too unconstrained: if there is no independent definition of cues, then we have no way of specifying the class of possible parameters, and hence the range along which languages may differ (synchronically or diachronically).

It is, however, possible to maintain that parameters can be independently defined and that learners also make use of cues provided by the input (this is closer to Dresher's view). Recall that according to current assumptions in the Principles and Parameters framework, parameters are lexical; it is also generally accepted that the lexicon has to be learnt, as it is language specific. There must be some learning device that enables acquirers to learn words (their syntactic, morphological, phonological and semantic properties). If parameters are linked to a subclass of lexical items, that is, functional elements, which also have to be learnt, then it follows that the same device is also responsible for setting parameters. This device may be part of UG, or it may be a separate device which interfaces with UG (we will tentatively assume the latter, mainly for clarity of exposition). Any part of the input that can provide the acquirers with information about the lexicon is a cue. This approach, unlike Lightfoot's, allows us to maintain both the notions of cues and parameters: cues are provided by the input, parameters are specified by UG and are set by the learning device on the basis of the interaction of cues and UG. The relation between the cues and the parameter values is indirect and is mediated by the learner.

We can make the notion of cue clearer if we consider the notion of parameter expression introduced by Clark and Roberts (1993:317):

- (4) Parameter expression:  
A sentence  $S$  expresses a parameter  $p_i$  just in case a grammar must have  $p_i$  set to a definite value in order to assign a well-formed representation to  $S$ .

As Clark and Roberts (*ibid.*) say: ‘When a given datum expresses some parameter value, the learner will be under pressure to set that parameter to the value expressed by the datum.’ This given datum is the trigger and is defined as in (5):

- (5) Trigger:  
A sentence  $S$  is a trigger for parameter  $p_j$  if  $S$  expresses  $p_j$ .

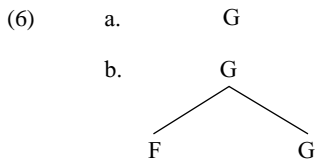
We can relate this notion of trigger to the notion of a cue by replacing ‘sentence’ in (4) and (5) by ‘substring of the input text’, as follows:

- (4’) Parameter expression:  
A substring of the input text  $S$  expresses a parameter  $p_i$  just in case a grammar must have  $p_i$  set to a definite value in order to assign a well-formed representation to  $S$ .
- (5’) Trigger:  
A substring of the input text  $S$  is a trigger for parameter  $p_j$  if  $S$  expresses  $p_j$ .

It is an empirical question what a substring may be. Arbitrarily, we will suppose that a substring can be no smaller than a morpheme (we are thus proposing that a morpheme is the minimal unit of grammatical analysis for language acquirers as well as for linguists) and no larger than a sentence (cf. Fodor 1998:17 for a similar proposal). If the parameter expression is robust enough, it will lead to the correct parameter setting. If, however, the parameter expression is ambiguous, then there must be some ‘safety mechanism’ in the learning device which leads to the assignment of a value – weak determinism requires this. This value will still be compatible with the input, but – again due to weak determinism – may differ from that of the target grammar, thus yielding a change.

The questions now are: (a) what is ‘robust’ parameter expression? (b) what is the ‘safety mechanism’ referred to in the above paragraph? We know of no good general answer to question (a), beyond observing that many parameters seem to be morphologically expressed, and when independent morphological or phonological changes conspire to remove or obscure this expression, a parameter change may take place (see Roberts 1999 on this). This answer is undoubtedly insufficiently precise and insufficiently general. Concerning question (b), we suppose, following Clark and Roberts (1993), that the learning device is computationally conservative in that it has a built-in preference for relatively simple representations. In other words, if the trigger is ambiguous, the learner will choose the option that yields the simpler representation. We will consider the question of how to define simplicity in detail in Chapter 5, but here we will provide a preliminary illustration of what we have in mind. Let

us assume that movement operations are adjunctions, as proposed by Kayne (1994); then movement always creates relatively complex representations, in the obvious sense that (6b) with F adjoined to G is a more complex structure than (6a), where no movement, and thus no adjunction, has taken place:



(Here G and F may have any amount of internal structure; in particular they may be either heads or XPs.) Loss of movement will lead to a reduction in complexity, that is, to a simpler representation. More precisely, if the learner postulates non-movement the simplicity preference will be satisfied. So movement must be robustly triggered (note that we are implicitly taking movement to be a parameter here – we develop this in section 3). If (6b) is not properly triggered, then (6a) will be preferred. Where (6b) changes to (6a) a movement operation is lost. However, there is another possible outcome where (6b) changes to (6a). The learner may analyse some instances of the moved category F as part of the inflectional system instantiated by G (this idea depends on the assumption that movement is always and only to a functional position – see section 3). This kind of ‘misanalysis’ results in recategorising a class of lexical elements as inflectional items; in (6b) F is reanalysed as G, essentially giving the structure in (6a). In other words, ‘misanalysis’, in the sense described here, can create new functional material. We will argue extensively in Chapters 2 to 4 that this kind of structural simplification is precisely the one that occurs in cases of grammaticalization. Another kind of structural simplification involves reanalysis of an XP, a category with a certain amount of internal syntactic structure, as a simple head X, a category with no internal syntactic structure. The same considerations relating to language acquisition apply to this kind of reanalysis as to the loss of movement, and we will see in Chapters 2 to 4 that this kind of reanalysis, among others, is also prevalent in grammaticalization.

To summarize, in this section we considered the general assumption that parameters are a property of lexical items. We discussed the general approach to language acquisition in the Principles and Parameters model, according to which the process is viewed as parameter setting. Syntactic changes, on the other hand, are the result of changing parametric settings. Learnability issues

connect to both language acquisition and language change, as there has to be some mechanism that allows the learner to set or reset parameters on the basis of the trigger experience. The latter happens when the trigger (or cue) is obscure. In this case, we propose that the learner will opt for the default option as part of the built-in preference of the learning device for simpler representations. The logical problem of language change is addressed in terms of the idea that the learning device is computationally conservative; a value  $v_i$  of parameter  $p_k$  can be changed where the trigger experience (or cue) for  $v_i$  is not sufficient to prevent a simpler option being chosen. This ‘insufficiency’ of the trigger experience can arise through the effects of other syntactic changes, phonological changes, language or dialect contact, etc.

One question that this approach gives rise to is: why are grammars not tending towards some maximally simple state, which, at the very least, would be free of movement operations? The answer is that the simplifications effected by changes are always local, and may increase complexity elsewhere in the system. In fact, grammaticalization is a case in point: as already noted by Meillet (1912), grammaticalization may increase the notional expressive power of the grammatical system. Von Stechow (1995:184) notes that under grammaticalization ‘the meaning of a lexical category is composed with a functional meaning to yield a new, more complex functional meaning’. In our terms, grammaticalization may provide a functional category with new exponents – this will become clearer in the next two sections. But, as just sketched, grammaticalization nevertheless arises from the learning device’s bias towards simpler representations.

In the remainder of this chapter we will make more precise what it means to say that parameters are lexically associated with functional categories, beginning with a general discussion of functional heads themselves.

## **1.2 Functional categories**

In the previous section, we presented the recent Principles and Parameters approach to cross-linguistic variation, according to which parameters are associated with functional categories. Parameterization as such then is restricted to the lexicon. Syntax connects the Phonological and Logical Forms (PF and LF respectively), that is sound and meaning. This is achieved with the help of the two basic mechanisms: Merge and Attract/Agree (Chomsky 1995, 2000). Merge is a binary operation that recursively combines elements, thereby building phrase structure. Agree is the operation that manipulates combinations, by establishing a relation between lexical items within a syntactic space. A simple

example is the agreement that we see between the subject and the verb in a sentence like:

- (7) John likes/\*like apples.

The agreement *-s* on the verb is the morphological expression of the relation that holds between the subject and the verb.

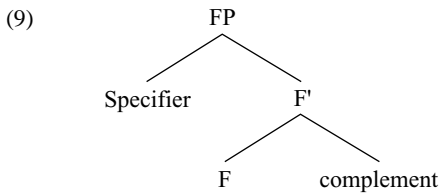
Lexical items belong to various categories, and this information is relevant for the syntactic operations of Merge and Agree. Categories then are primitive symbols associated with lexical items. The distinction between lexical and functional categories has its antecedents in traditional grammar. While Nouns, Verbs, Adjectives, and (at least some) Prepositions are lexical categories, elements such as Tense, Complementizers, Determiners, Negation, to name but a few, belong to the set of functional categories. The distinction between two kinds of item is an old one (cf. the Aristotelian distinction between *substance* and *accidence*), and it comes under various names, such as open versus closed class or lexical versus grammatical categories. In general, the basic distinguishing property is that lexical categories have descriptive content while functional categories do not; instead they carry grammatical meaning (cf. Radford 1997, Chapter 2 for a recent introductory discussion).

This distinction is widely accepted as one which holds in the lexicon. The question though is whether functional categories also have a syntactic representation. In other words, are functional categories also subject to syntactic operations, such as Merge and Agree? Some grammatical material seems to be purely morphological, and to have no role to play in syntax; for example, this seems to be the case of noun declensions or verb conjugations in languages like Latin or Classical Greek. Other material appears simply to duplicate other elements in the clause. For example, negation in French is realized by means of two elements, that is, *ne...pas* as in (8a), although only one of them (*pas*) is the ‘true’ negation. Similarly, expletives such as *there* in English double the postverbal subject in a construction like (8b) (the same can be argued for the subject agreement that we see on the verb in (7)):

- (8) a. Marie *n’aime pas* Jean.  
       Mary not loves not John  
       ‘Mary doesn’t love John.’  
       b. *There* arrived *three students*.

We can also see from (7) and (8) that grammatical material is lexically specified for morphological properties: *pas* in (8a) is a free morpheme, while *-s* in (7) is a bound one. Most importantly, grammatical properties such as those in (7)

and (8) turn out to be relevant in syntax as well: subject–verb agreement as in (7) is sensitive to the syntactic notion ‘subject’, while both *ne* and *there* in (8) have been argued to be syntactic markers of different kinds of scope (see Kayne 1984 on *ne*, Williams 1984 on *there*), also a notion standardly defined over syntactic structures. If this is correct, then we have to ensure that they are somehow syntactically present. In other words, we need syntax to be able to make reference to features associated with functional categories. Now, since Chomsky (1970), categories have been analysed as feature matrices. This means that it is possible – and, given the considerations just raised in connection with (7) and (8), desirable – to analyse grammatical features like agreement, negation, tense, etc., as syntactic categories. Given both the standard view of phrase structure (X'-theory), and the more recent Bare Phrase Structure of Chomsky (1995), that means grammatical features can function as heads which project a phrasal category containing a specifier and a complement, as follows (cf. (2) of the Introduction; here F is any feature):

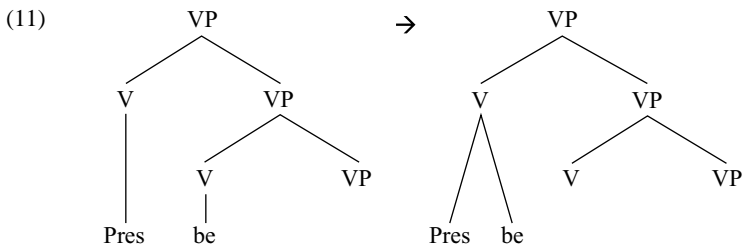


We will now provide some arguments in favour of having functional elements syntactically present in the sense just described. Let us begin by looking at the English auxiliary system:

- (10)
- a. Do you like fish?/ \*Like you fish?
  - b. I don't like fish/ \*I like not fish.
  - c. Fred likes fish and Bill does/ \*likes too.
  - d. I should go/ \*I should do go/ \*I do to go.

As the examples in (10a)–(10b) show, main verbs do not invert and cannot support negation; in both cases the auxiliary *do* must be present. Similarly, *do* can occur in elliptical contexts, while main verbs cannot – see (10c). The examples in (10d) show that the auxiliary *do* is in complementary distribution with modals, such as *should*, and with the infinitival marker *to*. The empirical evidence in (10) is the standard way of distinguishing between auxiliaries and main verbs. In the late 1960s and early 1970s, there was some debate as to whether data like that in (10) justified the postulation of a distinct category Aux, or whether it simply meant that certain verbs (e.g. *do*, *shall*, *have*, *be*, etc.)

were to be assigned various kinds of exception features. McCawley (1971) took the latter view. As he points out: ‘Auxiliaries are exceptional by virtue of undergoing a transformation “tense-attraction”, which combines them with the immediately preceding tense morpheme. All other transformations that might appear to treat auxiliaries in a special way (for example, subject verb inversion) are simply transformations that follow “tense-attraction” and have a structural description calling for the first verb.’ He proposes the following structure:



In (11) the auxiliary *be* is attracted by the higher V which carries tense information (i.e. Present in this case). A similar approach is taken by Emonds (1970, 1976) who argues that verb raising attaches *have* and *be* (which, unlike modals, are treated as members of V) to the Aux node. Verb raising next feeds subject inversion. *Do*-insertion inserts *do* under V to the left of the main V, while it is verb raising again that places *do* under Aux. Finally, *do*-deletion deletes *do* where Aux appears immediately adjacent to VP. What Emonds and McCawley have in common is that they isolate a given position – the highest V for McCawley, Aux for Emonds – as the structural position associated with tense-marking, and that auxiliary verbs can move into that position. This is the position Chomsky (1981) called I(nflection), and which more recently has become known as Tense.

In addition to noting the common points between McCawley and Emonds, we can make two further observations about the English auxiliaries, both of which are relevant for understanding the notion of functional category. First, even if we categorize auxiliaries *do*, *be* and *have*, the modals, tense, even the infinitival marker *to* (see Pullum & Wilson 1977) as verbs, we have to accept that they are morphologically irregular, have special syntactic properties and form a closed class of items. It is also important to observe that they lack a central lexico-semantic property of verbs, namely argument structure (with the possible exception of dynamic modals; see 2.1). Second, tense, modals and auxiliaries project like other categories. In current terms, this means that Tense heads the phrasal category TP. Its Specifier is arguably the subject position



(this is proposed in Chomsky (1995, 4.10)) and its complement may be VP. Consider (10a–b) again: subject–Aux inversion indicates that *do* must have a syntactic position, as it can invert with the subject; the same holds for (10b) as *do* can support the negative element *not*. Furthermore, as Ross (1967) argued, deletion processes show that auxiliary elements are part of syntactic phrases:

- (12) Fred could have been killed and Bill (could (have (been (killed)))) too.

As (12) indicates, any of the bracketed material can be deleted. Ross argues that the simplest account of these facts is to treat each bracketed constituent as a separate VP, headed by the respective verb or auxiliary. If we treat the highest auxiliary (*could* in (12)) as Tense, then we have a reason for thinking that it forms a constituent with the following bracketed material (this constituent may be T', if the Specifier of TP is the subject, as just mentioned; this point does not alter the fact that deletion processes show that functional heads project phrasal categories).

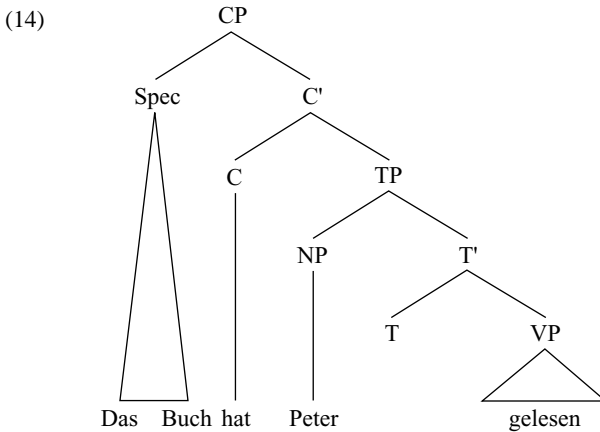
Having seen some evidence in favour of TP, let us now turn to another functional element, the Complementizer. Bresnan (1972) argued for the syntactic presence of a C(omplementizer) position as part of the extended structure of the sentence notated as S' (as distinct from the 'core' sentence S, so subordinate clauses were assigned a structure like [<sub>S'</sub> *that* [<sub>S</sub> *John left*]]). Clearly, Complementizers like *that* and *if/whether* differ in that the former appears with declaratives, while the latter introduce embedded interrogatives. The C position thus carries information about clause-type, and as such it is natural to think of it as the head of the subordinate clause. Given that C can also bear the +wh specification (as in *whether*-clauses), its Specifier can be identified as the landing site of wh-fronting (see Chomsky 1986b). The structural complement of C is TP.

C plays an important role in the analysis of other phenomena. For example, den Besten (1983) showed that many main-clause inversion processes target this position, so C must also be the head of main (or root) clauses:

- (13) a. Peter hat das Buch gelesen.  
Peter has the book read  
b. Das Buch hat Peter gelesen.  
\*Das Buch Peter hat gelesen.  
'Peter has read the book.'

This is the well-known case of the verb-second (V2) phenomenon found in root declaratives in nearly all Germanic languages (and already mentioned in the previous section). The obligatory subject–Aux inversion in (13b) suggests

that the auxiliary has moved to a higher position, namely C. Notice that even if we treat auxiliaries as a class of defective verbs, we cannot do the same with complementizers. Thus, C at least is a grammatical element that has syntactic reality. The availability of inversion in matrix declaratives in V2 clauses, as in (13), brought to attention the intimate relation that holds between complementizers and auxiliaries, or to be more precise the relation between the C and T heads (cf. Platzack 1987 on Germanic, tensed complementizers in Irish discussed by Cottell 1995, etc.). This relation of course further supports the claim that T elements must be syntactically represented. Thus there are at least two functional elements, C and T, that project syntactically. This gives rise to a structure like the following for an example such as the well-formed sentence in (13b):



Here we see how both CP and TP follow the  $X'$ -schema (see (9)). We also see that TP is the complement of C and VP is the complement of T. The properties that characterize the C-T system have been extended to the nominal system, leading to the postulation of a D(eterminer) category that takes the NP as its complement (cf. Abney 1987, Horrocks & Stavrou 1987, Szabolsci 1983/1984, for some early approaches).

Having provided some evidence for the syntactic presence of categories such as C and T, mainly based on English, let us now turn to their morphological properties. Consider C, for example, which in English can be realized by means of free morphemes, such as *that*, *if*, *whether*. At the same time, question formation in (10a) and Germanic V2 in (13) show another kind of morphological realization of C, namely by means of moving a verbal element to the C position, giving rise to inversion. In other languages, complementizers are realized by

means of affixes, as is the case in Korean (a rigidly head-final agglutinating language – see Cinque 1999:53–54 and the references given there):

- (15) cap-hi- si- ess- ess- keyss- sup- ti- kka  
 V- Passive-Agr Ant Past Epistemic Agr Evid Q  
 ‘Did you feel that (unspecified argument) had been caught?’

Here (interrogative) C is realized by the suffix *-kka*. It is a matter for debate whether languages like Korean are C-final, or whether TP moves to the Specifier of CP (see Kayne 1994 on the latter possibility). It is also possible to find languages which have no realization (alternatively use a zero morpheme) for C; this is in fact an option in English under certain conditions (cf. *I said (that) John left*). The same pattern can be found in the T domain: elements indicating tense, mood, etc., can be realized as free morphemes, like English modals; as bound ones, like the passive, epistemic and evidential morphemes in (15); or receive no realisation at all, as in the English simple present. Similarly, D can be free (as in English), bound (as in Rumanian) or zero (as in Latin). Thus, to summarize the discussion so far, we see that functional categories are subject to cross-linguistic variation in their realization, are like normal lexical categories in that they can project their properties, but differ from normal lexical categories in being closed-class and (as is clear when we compare auxiliaries and main verbs in English) in being inherently ‘defective’ in various ways.

In the recent theoretical literature, it is possible to identify two main views on functional categories: one is to deny their existence, the other to accept them. We have already provided evidence against the first view, as grammatical information is relevant for syntax and appears to have configurational instantiations. Further evidence comes from the areas of typology, diachrony, and language acquisition. In particular, typological studies have shown that languages undeniably differ in word order and morphology. As we mentioned in the Introduction, positing functional categories, and making them the locus of cross-linguistic variation allows us to reduce these two axes of variation to one. Regarding diachrony, it is one of the aims of this book to provide evidence from grammaticalization for the syntactic presence of functional elements (see also the first section of this chapter). Finally, recent work on language acquisition, starting with Hyams (1986) and Radford (1990), has shown the relevance of functional categories in the development of grammars, as early grammars differ from the adult ones in the way functional categories are realized.

Having argued then that functional categories must be syntactically present in some way, various options are open for how this idea may be implemented – particularly in accounting for cross-linguistic variation in word order and

inflection. There are in fact various ways ranging from accepting a very small to a quite large number of functional heads. The first of these approaches could be called the ‘What you see is what you get’ (WYSIWYG) analysis. As its name implies, the only functional categories postulated as present in a given language, or even a given sentence, are the ones for which we see some kind of realization. On this view, it is a matter of parametric variation as to which functional heads are present in which languages. For example, Grimshaw (1997) proposes that structures can ‘stretch’, or be compressed, even in the same language, depending on the number of lexical items available. To illustrate this point, consider (16):

(16) I think [<sub>VP</sub> it rained] (Grimshaw 1997:410)

According to Grimshaw, the embedded clause here is just a VP, as no auxiliaries or complementizers are present to instantiate T or C. However, as the examples in (17a–c) show, this can’t be right: if the subject was inside the VP, then substitution, fronting or deletion of the VP should also include the subject, contrary to fact:

- (17) a. *do so*: I thought it rained and \*(it) did so.  
 b. *VP Fronting*: I thought it rained and rain it did/\*it rain did.  
 c. *VP Deletion*: I thought it rained and \*(it) did.

Thus at least TP must be present.

Grimshaw’s (1997) analysis has to account for cross-linguistic variation in a different way by imposing a different ranking of constraints across languages. The result of this is the reduction of functional heads at the expense of a proliferation of constraints. However, it is a conceptual consequence of the distinction between syntax and phonology that certain elements may be present at one level and absent at the other. We therefore expect that syntactic categories can be silent. So there is no conceptual advantage in a position like Grimshaw’s. The next question then is whether this kind of view has empirical advantages. The answer seems to be negative for a number of reasons. Clearly, Grimshaw’s analysis of (16) cannot be right, as we have just seen. More generally, though, WYSIWYG approaches complicate the statement of cross-linguistic variation, as we have to assume that variation lies in differing selections from a universally given pool of categories such that simple sentences may have quite different structures in different languages; but if functional categories have semantic content, then we might expect that simple declarative sentences all have the same category across languages – this assumption is natural from the perspective of

the mapping from syntax to semantics, and simplifies the task of the language acquirer. However, this is explicitly denied by the WYSIWYG approach, which may say, for example, that German main clauses are CPs while English ones are VPs. Also, grammaticalization is harder to understand on the WYSIWYG view. If grammaticalization involves the development of new functional material, it must be analysed as a structural change rather than a simple category change. Given the assumptions about learning and change articulated in the previous section, structural change is hard to account for (in fact, the innovation of whole phrases is impossible on this view – surely a desirable result in the light of the logical problem of language change). Categorical reanalysis is a natural change, though, as we sketched there (see the discussion of (6)).

Another possibility is to assume that functional categories are always present, but in a very restricted fashion. For example, Chomsky (1995, 2000) argues that categories like C, T and D are present as they carry clause-typing, temporal and referential information respectively. Other functional categories that were postulated in earlier versions of the theory, such as subject agreement (AgrS) and object agreement (AgrO) should be dispensed with, given that they are not conceptually necessary (see Chomsky 1995, 4.10 and section 4.5 of the present work for discussion). This looks like a viable option, but the problem is that this kind of reduced structure brings along other complications. In particular, in order to accommodate lexical material structurally we need to assume that there can be multiple Specifiers, subject to parametric variation. Although it is desirable to keep the number of functional heads to a minimum, this kind of approach has the consequence of becoming less restrictive.

There are at least two further possible approaches. One is to say that we accept a relatively large number of functional heads, provided we find empirical support for their existence. This is the line of reasoning followed by many recent studies in different ways. For example, Kayne (1994, 1998) accounts for certain ambiguities (e.g. in *I will force you to marry no one*), which in earlier frameworks are assumed to involve covert movement in order to determine scope (of the quantifier *no one* in the example just given), by means of overt movement to a number of functional projections. The existence of functional positions is justified on this basis, but there is no further attempt to justify their presence conceptually by means of identifying the properties that trigger movement in the first place.

Cinque (1999) argues for a number of functional heads based on the distribution of adverbs. Each position carries the property identified with the interpretation of the adverb, resulting in the following set (we give only the labels of

the categories; from left to right, each takes the maximal projection of the next as its immediate structural complement in the sense of the  $X'$ -schema in (9)):

(18)

Mood <sub>Speech Act</sub>	Mood <sub>Evaluative</sub>	Mood <sub>Evidential</sub>	Mod <sub>Epistemic</sub>	T(Past)T(Future)
Mood <sub>Irrealis</sub>	Mod <sub>Necessity</sub>	Mod <sub>Possibility</sub>	Asp <sub>Habitual</sub>	Asp <sub>Repetitive(I)</sub>
Asp <sub>Frequentative(I)</sub>	Asp <sub>Continuative</sub>	Mod <sub>Volitional</sub>	Asp <sub>Celerative(I)</sub>	T(Anterior)
Asp <sub>Terminative</sub>	Asp <sub>Generic/progressive</sub>	Asp <sub>Perfect(?)</sub>	Asp <sub>Retrospective</sub>	Asp <sub>Proximative</sub>
Asp <sub>Durative</sub>	Voice	Asp <sub>Prospective</sub>	Asp <sub>SgCompletive(I)</sub>	
Asp <sub>PlCompletive</sub>	Asp <sub>Celerative(II)</sub>	Asp <sub>SgCompletive(II)</sub>	Asp <sub>Repetitive(II)</sub>	
Asp <sub>Frequentative(II)</sub>	Asp <sub>SgCompletive(II)</sub>			

This 32-head structure, as Cinque stresses, is a conservative estimate of the number of functional heads in ‘TP’. No account is taken here of Negation Phrases or Agreement Phrases, for example. A similar approach is also taken by Manzini and Savoia (forthcoming) and Poletto (2000) who postulate a number of functional heads in the C and T domain, based primarily on the possible clitic strings found in Italian dialects and their interaction with verb movement, negation, particles, etc. Giorgi and Pianesi (1997) adopt another variant. They assume a universal set of functional features, all of which are in principle able to project – in this respect their approach is like those just mentioned. However, they also assume that – while there is a universal hierarchy of functional projections – features can ‘scatter’ over a structure in different ways in different languages. To put it another way, features can syncretize onto heads as long as the universal ordering (which is thus not a total ordering, in the technical sense) is not violated. Cinque (1999:133) criticizes this approach on the grounds that it is excessively complex (a special convention is needed to interpret syncretic heads); we will return to this issue in our discussion of markedness in Chapter 5.

The results of these approaches are no doubt enlightening and can complement an alternative view which attempts to identify functional heads on the basis of their interpretation. This is the view that we will pursue in the present book. In particular, we will argue that only those functional heads that have logico-semantic content can be present. This allows us to postulate a rather large number of functional heads, but at the same time the requirement for interpretability constrains what can be a functional head. For example, functional projections that play the sole role of being place-holders cannot exist. We sketch this approach in the next section, and return to it in detail in Chapter 5.

To summarize, in the present section we considered the reasoning behind accepting functional (grammatical) elements as syntactic entities. It is clear that

the presence of functional elements in syntax has considerable implications for typological studies, for the study of diachrony as well as acquisition. The next question of course is to identify what does and does not count as a syntactic functional head. We fully address this question in Chapter 5. In the next section, we will present in detail how we see functional heads providing the means to express cross-linguistic variation.

### **1.3    *The nature of parameters: interface interpretation of functional categories***

So far we have established a couple of main points. First, cross-linguistic variation is associated with functional elements and is restricted to the lexicon. Second, language acquisition is the process of setting parameters, while syntactic change is the result of changing (resetting) parametric values, in the sense discussed in 1.1; the parameter-setting device may, under certain conditions, fix a parameter  $p$  differently from the value assigned to  $p$  in the grammar that underlies the trigger experience. Finally, functional categories are syntactically present: they project their categorial features following the  $X'$ -schema in (9). On this basis, they are manipulated by syntactic operations such as Merge and Agree. What we need to do next is to clarify the nature of parameters, so that we can provide an account of grammaticalization. Since parameters are associated with functional heads, we need to specify the lines along which these heads may vary.

The approach we will outline here is based on Roberts and Roussou (1999), who aim at giving expression to the idea that movement, cross-linguistic variation and at least some morphophonological properties are reflexes of a single property of the computational system of human language ( $C_{HL}$ ). This property of  $C_{HL}$  is driven by the interfaces, and is referred to as **interface interpretability**. The analysis takes the standard view of the interfaces as PF and LF, that is, the interfaces with the Articulatory-Perceptual and the Conceptual-Intentional systems respectively. Interpretability is the property of mapping a syntactic feature onto a PF or LF expression. To take a very simple example, the noun *table* maps onto a PF representation (/teibl/) and an LF representation, that is, its denotation ([[table]]). We cannot go into detail here as to the nature of the PF or LF representations, but it suffices to state quite simply that, in principle, any syntactic symbol may or may not be mapped onto a PF or LF representation. The lexicon provides the information determining the mapping. For ease of exposition at this point we could designate a syntactic symbol which has a PF mapping as +p, and a syntactic symbol which has an LF mapping +l. So *table*

is both PF- and LF-interpretable. In fact, we can observe that the lexical entries of lexical categories, such as Nouns and Verbs, always contain a specification  $+p, +l$ .

Consider next a functional element such as C, which, as we mentioned in the previous section, provides information about clause-typing, among other things. This kind of information contributes to the interpretation of the sentence, so we take C (or more precisely the features associated with C) to be LF-interpretable, that is  $+l$ . We saw in the previous section that the realization of C is subject to parametric variation: a matrix declarative C in German is realized by means of a verbal element, which is partly responsible for the V2 construction, while its English counterpart receives no such realization. Instead the matrix declarative C in English is not spelled out, or alternatively is spelled out as zero. Another example of parametric variation was discussed in section 1.1 in relation to the 'pro-drop' parameter. Let us assume that AgrS is the position associated with the nominal features of the subject. As such it receives an interpretation at LF  $(+l)^1$ . Its realization, however, differs across languages. For example, English AgrS requires an overt subject, while its Italian counterpart allows for a null subject. We see then that functional elements are not necessarily  $+p$ . Variation in  $\pm p$  leads to cross-linguistic variation in which functional categories are overtly realized, as we will see in detail below. In general, then, we see that functional categories may be defined as that class of syntactic categories which is not obligatorily  $+p$ .

Among the functional features, Q, WH, Neg, T and D, at least, are LF-interpretable, that is,  $+l$ . These features clearly contribute to the interpretation of any phrase-marker they appear in. We assume that UG contains a vocabulary of substantive universals, which are realised as functional features in every language. These are the interpretable features. On the other hand, the  $\pm p$  property varies across languages. In fact, *pace* Chomsky (1995, 2000, 2001), we do not postulate uninterpretable functional categories or features. We believe that it is possible to maintain that all such features are LF-interpretable. If so, then there are no  $[-l]$  features.

The  $\pm p$  and  $\pm l$  properties are lexically determined, and as such are listed in the lexical entries of morphemes (cf. Cormack & Smith 1999 for a slightly similar approach). Assuming that lexical items are bundles of features, we can say that any category with N and V features is always  $+p, +l$ , while functional

1. Chomsky (1995, 2000) argues against the postulation of an AgrS category, on the grounds that phi-features are interpretable for nominals and not for verbs. However, if we take AgrS to correspond to a position that encodes the nominal features of the subject in the clause structure, then its presence becomes legitimate as these features are clearly interpretable. See section 4.5.