

3

Acquisition, learnability, and syntactic change

Introduction	207	3.4. Markedness and complexity	251
3.1. First-language acquisition from a principles-and- parameters perspective	209	3.5. Parameter setting and change	266
3.2. The logical problem of language change	226	3.6. Conclusion to Chapter 3	282
3.3. The changing trigger	236	Further reading	284

Introduction

Having by now established that parameter change can describe many instances and types of syntactic change, here we look at the deeper questions this conclusion raises. The purpose of this chapter is to explore the idea, introduced in §2.1, that parameter change is driven by the first-language acquisition process (this idea is pursued in Lightfoot (1979; 1991; 1998); see Croft (2000: 47–9, 119) for critical discussion), and thereby to illustrate how the study of syntactic change may be relevant for our understanding of the processes involved in first-language acquisition. One way of construing the idea that parameter change is driven by the first-language acquisition process is to think that a parameter

value changes because an innovative alternative is more 'accessible' to acquirers, thereby rendering the conservative value in effect 'inaccessible', or unlearnable. This view has two important consequences. First, to the extent that abductive reanalysis of the sort discussed in §2.1 is a symptom of an underlying parameter change, it can explain the pervasiveness of this kind of reanalysis in syntactic change. Second, it entails that language **learnability** is intimately connected to change – in fact learnability becomes the key to understanding syntactic change. This further tightens the connection between L1 acquisition of syntax and syntactic change. In fact, Niyogi (2004: 462) points out that 'every theory of language acquisition also makes predictions about the nature of language change.'

We begin by looking in §3.1 at the current state of knowledge regarding first-language acquisition of syntax, basing our presentation fairly closely on the discussion in Guasti (2002). In §3.2 we consider what we call, following Clark and Roberts (1993), 'the **logical problem of language change**'; we will see that this is closely related to, and maybe subsumes, the Regress Problem discussed in §2.1. In this context, we introduce a central idea for much of the later discussion: the idea that the language learner (or, in more technical terms, the parameter-setting device) is computationally conservative, obeying a kind of 'least-effort' constraint, i.e. a general preference for simplicity of representations, which we formulate along the lines of Roberts and Roussou (2003: 201). In §3.3 we try to get a picture of what kinds of external circumstances could cause a parameter change; this relates closely to the discussion of the logical problem of language change. Under this heading, we discuss in a preliminary way the possible role of language contact (although this will be the focus of Chapter 5), as well as the notion of cue introduced by Dresher (1999) and discussed at length in relation to syntactic change in Lightfoot (1999). We also discuss the role of morphological change in triggering syntactic change. §3.4 introduces the notion of markedness and relates it to the characterization of complexity/simplicity given in §3.2. Finally, in §3.5, we try to bring the strands of the discussion together in a general proposal for the form of parameters, how they are set and how they may change. This concludes the general discussion of parametric change as the mechanism of syntactic change, the remaining chapters being concerned with the wider implications of this view. But let us now begin at the beginning, i.e. with first-language acquisition.

3.1. First-language acquisition from a principles-and-parameters perspective

In this section I will try to rather sketchily summarize some aspects of the burgeoning recent literature on first-language (L1) acquisition of syntax. My focus will be on the major empirical observations and their implications for the thesis that parametric change is driven by the acquisition process. These are, first, that many important parameters appear to be set rather early in the acquisition process (see (2) below), and, second, that there are two phenomena of interest in children's early production: the so-called **root** or **optional infinitives** (see Radford (1990; 1996); Platzack (1992); Pierce (1992); Wexler (1992; 1994; 1999); Poeppel and Wexler (1993); Rizzi (1994); Haegeman (1995b); Hoekstra and Hyams (1998); Hyams (1996); Schütze (1997); Hamann and Plunkett (1998); see also Guasti (2002: 128ff.) and the references given there) and 'early null subjects' (see *inter alia* Hyams (1986; 1992); Bloom (1990); Valian (1990); Gerken (1991); Weissenborn (1992); Hyams and Wexler (1993); Rizzi (1994; 2000); Clahsen, Kursawe, and Penke (1995); Haegeman (1995a); Guasti (1996; 2000); and the papers in Friedemann and Rizzi (2000), as well as the references given in Guasti (2002, Chapter 5)). The purpose of our discussion is to indicate to what extent our understanding of the parameter-setting process has been furthered by this work, and to see if in principle any connection with a parameter-changing approach to syntactic change can be discerned.

Before looking at the phenomena which have been observed, however, we need to be clear about our general conception of first-language acquisition. In the introduction to Chapter 1, I presented and tried to justify Chomsky's claim that the human language faculty is a facet of human cognition, physically instantiated in the brain and, most importantly, genetically inherited as an aspect of the human genome. Under this conception of the language faculty, first-language acquisition can be characterized in the following terms:¹

¹ As we mentioned in note 1 of Chapter 1, modularity – the idea that the language faculty is a distinct system of the mind/brain – may not play a role in the minimalist conception of the language faculty. But the crucial point for the purpose of the discussion of first-language acquisition here is that even if the language faculty (in either the broad or the narrow sense as defined by Hauser,

The language faculty is a distinct system of the mind/brain, with an initial state S_0 common to the species ... and apparently unique to it in certain respects [footnote omitted]. Given appropriate experience, this faculty passes from the state S_0 to some relatively stable state S_S , which then undergoes only peripheral modification. (Chomsky 1986: 25)

The initial state of language acquisition is nothing other than Universal Grammar, while the stable state is adult competence in a given language, which remains unaltered in essential respects from childhood on. The process of first-language acquisition is the process by which the language faculty 'passes from the state S_0 to some relatively stable state S_S ' in Chomsky's formulation. We thus have the following schematic notions:

- (1) a. $UG = S_0$ (initial state);
 b. Adult competence = S_S (stable state);
 c. Stages of acquisition = $\langle S_0, \dots, S_{i>0}, \dots, S_{n>i}, S_{n+1<S}, \dots, S_S \rangle$.

Here, (1c) indicates the stages of first-language acquisition. These can be thought of as an ordered n -tuple of states of indeterminate, but certainly finite, number, occurring later than S_0 and earlier than S_S . They correspond neither to UG nor to the adult competence, but rather to what we can think of as an immature competence. (I will say more about this notion of immaturity below.)

How are the various states defined in relation to one another? To put it crudely, what does S_{n+1} have that S_n lacks in (1c)? As we have seen, the 'innateness hypothesis' claims that S_0 is determined wholly by the genome, independently of any experience. On the other hand, S_S differs from S_0 in that it is at least partly determined by experience of the linguistic environment (the primary linguistic data, or PLD): the PLD, among other things such as providing the vocabulary of the first language, in some way causes parameters to be set to determinate values. We can therefore assume that the various intermediate states are distinguished by having differing values of various parameters. Each S_n differs from S_{n+1} in one of two ways: either insufficient experience has been accumulated at S_n for setting certain

Chomsky, and Fitch (2002)) is in some sense 'emergent' and may certainly lack a single neurophysiological locus in brain architecture and a single phylogenetic source, we can nevertheless meaningfully distinguish the initial state of the system (or of the relevant subparts) in the newborn child from the modified state which is the stable, adult state. This process appears to be subject to a critical period; see §5.4 for some recent evidence for this.

parameters, or the overall system has not matured sufficiently at S_n to permit certain parameters or parameter values to be attained. This in turn could be due to different parameters coming 'on line' for acquisition at different times, to the system gradually maturing with respect to the kinds of PLD it is able to accommodate, or to interactions among already-set parameters, perhaps in accordance with a general notion of markedness (see §3.4 and §3.5). In any case, we can consider that the various intermediate states differ from one another in representing successively closer approximations to the adult system (S_S) in terms of the values of the parameters. To put it another way: if m parameters are set to the adult value at stage S_n then at least $m + 1$ parameters are set to the adult value at stage S_{n+1} .

The nature of the intermediate grammars has been studied fairly intensively in the past twenty years or so, beginning with the pioneering study in Hyams (1986). Guasti (2002) points out that between the ages of two and three years old, i.e. some time before linguistic maturity if this is characterized as the stable state, children know at least the following about the parameter values of the language they are in the process of acquiring (the references are given in Guasti (2002: 148, 185, 242)):

- (2) a. the value of the head direction parameter in their native language;
- b. the value of the V-to-T parameter in their native language;
- c. the value of the topic-drop and null-subject parameters;
- d. the value of the parameters governing question formation, the one governing overt movement or in-situ placement of the *wh*-element and the one regulating T-to-C movement (inversion).

To this we can add that Hamann, Rizzi, and Frauenfelder (1996) show that as soon as a French-acquiring child produces clitics, they are placed in the correct clitic position for French, even though there is much parametric variation in clitic-placement across languages, and, following Wexler (1998), verb second. Wexler (25) describes this general phenomenon as follows: '[b]asic parameters are set correctly at the earliest observable states, that is, from the time that the child enters the two-word stage around 18 months of age.' He continues: '[q]uite possibly ... children have set basic parameters ... *before* the entry into the two-word stage'. This observation has become known as Very Early Parameter Setting, or VEPS.

Guasti also provides evidence that, also between the ages of two and three, children know the properties of unaccusative verbs (see §2.3.1).

Furthermore, by the age of four, children comprehend and produce passives based on actional verbs, although they have difficulty with passives of non-actional verbs (2002: 269); they have acquired most, but not all, of the principles governing the distribution and interpretation of anaphoric and other pronouns (the **binding theory**: see the textbooks mentioned in the Introduction for discussion of this aspect of syntactic theory) (2002: 310); they have also acquired the principles concerning the distinction between referential and **quantified expressions** (for example, *John vs. every boy*) and many aspects of the interpretation of quantified expressions (2002: 344); and, finally, that they have acquired many but not all aspects of the nature of the 'control' relation between a DP in a superordinate clause and the understood subject of a non-finite subordinate clause briefly alluded to in §2.4 (2002: 372).

Most of the parameters listed in (2) are familiar from Chapter 1. (2a) refers to word-order parameters, of which we identified several subtypes in the discussion in §1.6.1 (parameters F1–F6). The L1-acquisition literature has shown that children are sensitive to these parameters and that 'from the onset of multiword utterances (or even earlier)' (Guasti 2002: 103) they have correctly identified the relevant values for the ambient language. (2b) clearly refers to parameter B of §1.3.1. (2c) partly refers to parameter A of §1.2.1, although the notion of 'topic-drop' was not discussed there; I will return to this in the discussion of 'early null subjects' in L1 acquisition below. The second part of (2d) concerns parameter C, T-to-C movement, while the first part refers to parameter E of §1.5.

All of the parameters listed in (2) are important and salient for synchronic description, and Guasti's summary of the L1-acquisition evidence shows that they are salient for language acquirers; these parameters are acquired early and correctly, or so it appears. As it stands this observation supports the poverty-of-the-stimulus argument, as Guasti implies (2002: 147); children are able to glean the values of these parameters from the PLD almost before they are able to produce multiword utterances. This strongly suggests a predisposition to the task, given the very young age of the children (multiword utterances normally begin between twenty and twenty-four months old; see Guasti (2002: 98)), the complexity of PLD, and the rather abstract nature of these parameters. However, we saw in Chapter 1, several of these parameters can be shown to have changed their values in the recorded history of various languages. This brings us face to face with what Clark and Roberts (1993: 299–300) termed the logical problem of

language change, which I will discuss in more detail in the next section. For the moment, it suffices to note both the tension between Guasti's conclusions and what we saw in Chapter 1, and that the obvious resolution of this tension must involve some intergenerational change in the nature of the PLD; some of the issues surrounding this conclusion were discussed in §2.1, and will be discussed further later in this chapter.

Returning to the discussion of the intermediate stages of acquisition as defined in (1c), it may seem that we have little evidence regarding the setting of the parameters in (2) beyond the fact that they are typically set early and accurately. Far from shedding light on a parameter-resetting approach to syntactic change, this appears to pose a problem, as just noted. Of course, it is entirely possible that parameter-resetting can take place between twelve and twenty-four months, but this will be very difficult to document on the basis of children's utterances prior to the two-word stage (which normally starts at around eighteen months); it is possible that some experimental methodology could be developed in order to ascertain this, but I am aware of none that has been developed at present. (Wexler (1998: 25, note 1) makes the same point.) But the grammar of the earliest stages of multiword production is still an immature one in the sense that it is subject to modification through further stages of acquisition. Guasti's conclusions as listed in and immediately below (2) appear to show that there is an intermediate, fairly early, state of the language faculty $S_{i>0}$ at which a number of important parameters have been set, but that there are nevertheless further stages of acquisition, and presumably therefore of parameter-setting, $S_{n>1}$, $S_{n+1<S}$, etc., remaining. (See note 5 below for a conjecture as to the difference between the two sets of parameters.) Since these are stages during which children produce multiword utterances, we have in principle better access to the parameter-setting process here than in the case of the parameters whose values are set earlier. Can anything relevant for our conception of syntactic change as parameter-resetting be gleaned from these later stages?

What would be the 'ideal scenario' for relating parameter setting in first-language acquisition to parameter change? The kind of case which could link the evidence of production based on immature grammars of the type discussed in the L1-acquisition literature to the questions relevant for syntactic change would have to have four properties. First, we would want to compare the acquisition of two closely related languages L and L' where it is known that L' is syntactically innovative in relation to L

(in that it is known that the common parent language of L and L' set some parameter P to value v_i , and that L has value v_i for P while L' has value $v_j \neq i$ for P). Second, we would observe that early production in L showed a tendency for strings which appear to express value v_j for P, in the sense of parameter expression introduced in §2.1, (8). Third, we would observe that this apparently aberrant production in L ceases when some further feature F of L is acquired, i.e. manifested in production, and, fourth, we would observe that F was lost from L' when P changed value. So we link the change in P, *both in acquisition and in diachrony*, to the presence/absence of F.

The two main phenomena in early production which have been looked at, root/optional infinitives and early null subjects, come close to instantiating the 'ideal scenario' as just described in relation to some of the parameters listed in (2). It is probable, however, that neither of them truly instantiates this scenario. Nevertheless, it is worth looking at them.

The essence of the root (or optional) infinitive phenomenon is that it 'consists of producing main clauses containing an infinitive verb, rather than a finite one' (Guasti 2002: 128). It is 'peculiar to the earliest multiword productions and lasts until about 3 years' (*ibid*). Some examples from various languages are given in (3) (taken from Guasti (2002: 128–9); sources for the examples are given there):

- (3) a. hun sove (Swedish)
she sleep-infin
- b. eerst kleine boekje leze. (Dutch)
first little book read-infin
- c. pas manger la poupée (French)
not eat-infin the doll
- d. s[çh]okolade holen (German)
chocolate fetch-infin
- e. Papa have it. (English)

In each example, the verb has the form of an infinitive, despite the fact that these are all main clauses. For English, this implies that the verb has the bare-stem form: cf. *have* rather than *has* in (3e). In the other languages, the ending is recognizably that of the infinitive (for example, *-en* in Dutch and German, *-er* in French). Moreover, in some of the examples, the verb has the syntax of an infinitive: it follows the negative element *pas* in French in (3c), while finite verbs must precede *pas* (see §1.3.1), and it follows the

direct object in Dutch – (3b) – and German, (3d). Here the Dutch example is more telling, since this could not be a verb-second clause, given the presence of the adverb *earst* (*erst?*) ‘first’ in addition to the direct object.

The following constraints on root infinitives have been observed in the L1-acquisition literature:

- (4)
- a. Root infinitives do not occur in null-subject languages.
 - b. Root infinitives are not introduced by nonsubject XPs in V2 languages.
 - c. Root infinitives are incompatible with clitic and weak-pronoun subjects.
 - d. Root infinitives occur in declaratives, but not in wh-questions.
 - e. Root infinitives are incompatible with auxiliaries.

In connection with (4a), note that all the languages exemplified in (3) are standardly analysed as non-null-subject languages (see §1.2.1). (4b) strongly suggests that clauses containing root infinitives are not verb-second clauses, even in verb-second languages, despite being main clauses – see below. (4c) is fairly self-explanatory (recall the notions of clitic and weak pronoun from the discussion of the history of French in §1.2.2), and (4d, e) are straightforward.

Two principal types of analysis have been proposed for this phenomenon: the Tense-omission approach of Wexler (1994; 1999), and the clausal-truncation approach of Rizzi (1994). (Avrutin (1998) pursues a third option, observing that root infinitives occur in adult Russian under certain discourse conditions; it is therefore possible that the cases illustrated in (3) are straightforward instances of initial parameter-missetting; it is not clear, however, to what extent Russian and some related language might instantiate the ‘ideal scenario’ as described above). Both approaches rely, in different ways, on the idea that the T-position in main clauses optionally lacks some crucial property in the child grammars which produce root infinitives: for Wexler, the tense feature is not specified, while for Rizzi, all projections above VP may be ‘truncated’, i.e. simply not present, at this stage of grammar development; Rizzi (2005: 94–5) extends this possibility to the categories making up the ‘split CP’ he assumes. Both analyses can account for the absence of root infinitives in null-subject languages: for Wexler, this depends on specific assumptions about the relationship between the T-position and the agreement features characteristic of null-subject languages (see Guasti (2002: 137–9) for a summary); for Rizzi, this is because in null-subject languages infinitives must raise out of VP (see

Kayne (1991) and note 10 of Chapter 1 on movement of non-finite verbs), and this is of course impossible if the structure above VP is not there. Similarly, both analyses can account for the fact that clauses containing root infinitives cannot be V2 clauses (see (4b) above): in both cases this is because the C-position is not available as a target for movement; for Wexler, because non-finite verbs cannot raise there, and root infinitives are not specified as finite; for Rizzi, because CP is simply absent. The incompatibility of root infinitives with clitic and weak-pronoun subjects is again straightforward for Rizzi: it is well-known that these elements are attracted to the position bearing agreement features (which we take to be T), and so if this position is simply absent, such elements cannot be licensed. This restriction is problematic for Wexler's analysis, however. Property (4d), that root infinitives are not found in wh-questions, is once more straightforward for Rizzi: such questions clearly depend on the nature of the C-position, a position which is absent in a truncated clause. Wexler cannot handle this fact straightforwardly, since wh-questions may be either infinitival or finite, although of course infinitival wh-questions must be indirect questions in adult language: cf. *I don't know what to do/what I should do*, and in fact Wexler takes issue with this generalization about root infinitives, citing examples like *Where train go?* from child English. Intriguingly, though, such examples seem to be restricted to English.² Finally, both analyses can account for (4e) as long as it is assumed that auxiliaries require a fully-specified T-position, something lacking in the immature grammar on both analyses.

Root infinitives are typically no longer found after age three (Guasti 2002: 146). Both Wexler and Rizzi propose that this is due to the maturation of the grammatical system. They thus take the intermediate stages of language acquisition, or at least some of them, to represent literally immature grammars: grammars of a type that do not underlie any form of (non-pathological) adult linguistic behaviour. The idea that linguistic competence matures during the intermediate stages of language acquisition is, in the context of the assumption of a genetically-determined language faculty, quite reasonable. As Guasti says (2002: 146):

² There has been a debate about the analysis of such productions in child English. Roeper and Rohrbacher (2000) give numerous examples of the type *Where __ go?* Guasti (2002: 139, 202–8) provides a summary and references.

Maturation is likely to control some aspects of language development – for example, the fact that infants start to babble orally or manually around 6–8 months. According to the maturational view, a genetic program also controls the development of syntax ... and determines the timing by which components of UG become available to the child. Under this view, R[oot] I[nfinitive]s occur because principles of UG have not matured.

The principles of UG in question clearly concern the fact that a main-clause Tense has to be fully specified for finiteness and other features, something both Rizzi and Wexler articulate in differing ways.³ So the account of root infinitives relies in one way or another on the idea that the requirement for specification of these features matures, typically, at around three years old.

In terms of these analyses, then, the root-infinitive phenomenon does not come close to our 'ideal scenario' for relating syntactic change and the acquisition of syntax. It simply involves the transition from one stage of acquisition to the next owing to the genetically-determined maturation of UG principles determining the well-formedness of main-clause Tense. However, as mentioned above, another way of moving from one stage of acquisition to the next must be the incorporation of further data: the accumulation of experience through greater exposure to the PLD. This idea and the maturation approach are not incompatible: in fact, it is natural to think that there is positive feedback between continued exposure to PLD and maturation of aspects of UG, in that greater exposure to data may in fact cause the system to mature as long as a certain age threshold has been passed. This view is supported by the fact that it is known that there is a **critical period** for language acquisition in general (see §5.4 for discussion of recent evidence for this), and that environmental stimulus is required in order for the system to come into operation at all.

In these terms, we might conceivably relate the root-infinitive phenomenon to our ideal scenario by adopting a two-stage approach to the acquisition of verbal agreement. Suppose that the 'first pass' acquirers make to the acquisition of agreement, at a rather early stage of acquisition, involves setting the null-subject parameter. Thus, if the relevant kind of agreement

³ If the minimalist conception of UG is taken on board, there may be rather few components of UG which are in principle available to come 'on line' at different stages of language acquisition, something which would impose inherent restrictions on how far maturation could be invoked to explain properties of child production. It is not clear how far this affects the analyses of root infinitives under consideration here, though.

and other properties (see §1.2.1) are expressed in the PLD, the positive value of the null-subject parameter is expressed and thus set by the acquirer. At this point, let us suppose that the immature system treats all non-null-subject systems alike: as having no agreement. Now, in §2.1, (16), I suggested, following Vikner (1997), that a certain pattern of verbal agreement, less robust than that required for a positive value of the null-subject parameter and yet greater than zero, expresses a positive value for the V-to-T parameter. At this intermediate stage, where all non-null-subject languages are treated as entirely lacking in agreement, there can therefore be no verb-movement. This can provide an account of the root-infinitive stage. Recall the basic properties of root infinitives, given in (4) above, repeated here:

- (4) a. Root infinitives do not occur in null-subject languages.
 b. Root infinitives are not introduced by nonsubject XPs in V2 languages.
 c. Root infinitives are incompatible with clitic and weak-pronoun subjects.
 d. Root infinitives occur in declaratives, but not in wh-questions.
 e. Root infinitives are incompatible with auxiliaries.

Clearly, on this view (4a) is accounted for, as a central assumption is that root infinitives result from an immature negative setting of the null-subject parameter. To the extent that clitics and weak pronouns depend on the presence of strong agreement, then (4c) is accounted for. Properties (4b) and (4d) depend on verb-movement to C (recall that root wh-questions involve verb- or auxiliary-movement to C in all the languages in question), and this will not be available if V-to-T movement is not available (and if auxiliaries are not merged directly in T, see below). Finally, we can account for (4e) if we assume that auxiliaries are elements which must always either move to T or be merged there (see §2.1), independently of the parameter determining V-movement to T. So the root-infinitive phenomenon could conceivably arise if, in acquiring systems with *little* verbal inflection, children assume there is *none at all* at first. This idea is similar in some respects to Phillips' (1995) idea that root infinitives arise from the failure of the features of V and T to combine either through Move or Agree. The phenomenon disappears when children make a later, 'second pass' at the acquisition of agreement, and at this stage they are sensitive to the expression of the agreement present in some non-null-subject languages. (This second pass may arise either through exposure or maturation, or, most likely, a combination of the two as described above.) This second pass is

related to the acquisition of the parameters involving V-movement, particularly V-to-T movement. The idea that verbal agreement is acquired in two stages is supported by the fact that there is good evidence that children acquire the agreement marking in null-subject languages early (Guasti (2002: 120–2) and the references given there), while it is well known that the much-impoorished agreement marking in English is acquired much later (Cazden 1968, Brown 1970); indeed, there is some reason to think that English agreement morphology is acquired like irregular tense marking (Maria-Teresa Guasti (p.c.)). The situation in German, on the other hand, appears equivocal, with Wexler and Poeppel (1993) arguing for early acquisition of agreement marking and Clahsen and Smolka (1985) and Clahsen and Penke (1992) arguing that there are many errors in the early use of German agreement marking (Guasti *loc cit.*).

The interest of the account of root infinitives just sketched is that it can get us close to our ideal scenario for linking acquisition and change. In terms of the scenario, we can take L to be German and L' to be English. English is syntactically innovative in relation to German, in that it is probable that the common parent language, Proto-West Germanic, set the V-to-T parameter to the movement value, as we have assumed for German (see §1.3.1, notes 18 and 30). The reason for this is that Proto-West Germanic is usually reconstructed as having a very rich verbal agreement system, and hence enough agreement was present to express a positive value for the V-to-T parameter. Hogg (1992: 147ff.) presents a reconstructed stage of pre-OE; and the paradigms of the Gothic verb given in Jasanoff (2004: 900) nearly all have forms of both strong and weak verbs which distinguish all person-number combinations. Modern English, of course, sets this parameter to the negative value. Second, we observe that early production in German shows a tendency for strings which appear to express value v_j for P: these are the root infinitives.

So far, we are close to making a connection between the two areas of acquisition and change. However, the third step of our scenario involves the observation that the 'apparently aberrant production in L', i.e. the root infinitives, ceases when some further feature F of L is acquired. If we could establish that German agreement marking is acquired relatively late, and coincides with the loss of root infinitives, this would be just what we need. However, as mentioned above, there is a debate regarding the timing of the acquisition of German agreement (possibly because the researchers in question made use of differing experimental methodologies; Maria-Teresa

Guasti (p.c.)). And so here we cannot be sure of attaining the ideal scenario.

The fourth part of the scenario is straightforward, on the other hand. We have already observed the correlation between the loss of agreement marking (F) from English (L') when the V-to-T parameter changed value.

We see that it may be possible to link root infinitives to some aspects of the loss of V-to-T movement in English, although making the connection is not without problems.⁴ Let us now turn to the other L1-acquisition phenomenon of interest: early null subjects.

Early null subjects are illustrated in (5) (examples from Guasti (2002: 151); sources are given there):

- (5) a. Se, blomster har. (Swedish)
 see, flowers have/has
 'Look, (I/you/she/we) have/has flowers.'
 b. Tickles me.
 c. Mange du pain. (French)
 eat-3sg some bread

While similar to root infinitives, early null subjects differ from them in that the verb is clearly finite, as can be seen from the forms in (5), and the fact that they are compatible with the presence of auxiliaries (compare (4c) above).

At first sight, it might seem that early null subjects provide evidence that the null-subject parameter is not set as early as we have been supposing up to now, essentially following Guasti. In fact, in her pioneering work on this

⁴ Arguably the biggest problem is that Guasti (2002: 166) shows that there is an implicational relation between root infinitives and early null subjects: if a child's grammar allows root infinitives, it allows early null subjects. But if root infinitives result from an early, correct, negative setting of the null-subject parameter, then this seems paradoxical. One way to handle this would be to claim that the correct acquisition of the strong agreement morphology associated with a positive value of the null-subject parameter implies that the null subject can be syntactically represented in such systems (either as *pro* or in the verbal inflection itself – see §1.1.1), while the immature negative setting associated with complete lack of agreement implies that overt subjects are simply not subject to a licensing condition, while covert subjects must be inferred from argument structure of the verb (i.e. as a kind of 'implicit argument'). It seems that any analysis of root infinitives has to allow for the fact that subjects, when structurally manifested, cannot be licensed in the way they are in the adult system (for example, by Agree with T's φ -features as discussed in §2.3). In different ways, this is true for Wexler's and Rizzi's analysis of root infinitives, as well as the one sketched above.

phenomenon, Hyams (1986) proposed that early null subjects were an indication of 'parameter-missetting' in relation to the null-subject parameter, in that children acquiring non-null-subject languages initially set the parameter to the positive value. This led to the suggestion that the null-subject parameter may have a default positive value, a matter we return to briefly in §3.4 below.

Hyams' account of early null subjects comes close to fulfilling our ideal scenario for the connection between language acquisition and language change. In terms of that scenario, we could suppose that L is Italian and L' is English; P is the null-subject parameter; v_i is the positive value of that parameter; and v_j is the negative value. Since it is likely that Proto-Indo-European, the common parent of English and Italian, was a null-subject language (see §4.4.4), then English is syntactically innovative with respect to Italian as far as this parameter is concerned. The crucial factor F, causing children to fix the parameter correctly after a period of 'aberrant' production, could be either the presence of modal auxiliaries (originally identified by Hyams as inherently incompatible with null subjects) or overt expletive pronouns (often thought to be incompatible with a positive value for the null-subject parameter – see Rizzi (1986a), and, for a different view, Holmberg (2005)). The difficulty with this is that feature F, which, following Hyams, we take to be expletive pronouns or modal auxiliaries, is predicted to have arisen when English, or the relevant ancestor of English, ceased to be a null-subject language. But, first, we have little clear idea as to when that was; as just mentioned, it is likely that Proto-Indo-European was a null-subject language, and it is possible that Proto-Germanic was (see the above comment on archaic Germanic verbal agreement). As for the likely status of the null-subject parameter in the runic inscriptions (which may represent an early form of either North or Northwest Germanic, see Faarlund (2004a: 908)), Faarlund (2004a: 920) argues that the data from the surviving runic inscriptions is too sparse for any conclusions to be drawn. Second, both modal auxiliaries and overt expletive pronouns are innovations in the recorded history of English. It is usually thought that modal auxiliaries of the Modern English type arose in the sixteenth century (as was briefly discussed in §2.1; see note 4 of that section on the chronology of this change), while overt expletives appear during ME (A. Williams 2000; Biberauer 2003; Biberauer and Roberts 2005a). Thus there is a clear chronological mismatch regarding this feature, and so this vitiates this particular application of the scenario.

An alternative might be to take L' to be French. Here we are on much firmer ground regarding the parent language: it is clear that Latin was a null-subject language, and so French can be defined as syntactically innovative in relation to Italian in this respect. It is also clear that French shows early null subjects (see (5c)), and so we see the relevant kind of 'aberrant' production in child language. However, we run into difficulty with the third part of the scenario: neither French nor Italian has modal auxiliaries, and so the relevant factor must be overt expletive pronouns, which of course Modern French has but Italian lacks. But the problem is that OF, which, as we saw in §1.2.2, was a null-subject language, had overt expletives, as the following examples show:

- (6) a. Il est juget que nus les ocirum.
 it is judged that we them will-kill
 (*Roland*, 884; Roberts 1993a: 150)
- b. Il ne me chaut.
 it not to-me matters
 (Einhorn 1974: 123)

One possibility for saving this approach might be to claim that the expletives illustrated in (6) are in SpecCP (which they almost certainly are, given the V2 nature of OF; see §1.3.2), and that the relevant property for changing the null-subject parameter, in both acquisition and change, involves expletives in SpecTP.

However, there are examples of expletives in SpecTP in OF, such as the following:

- (7) car ainsin estoit il ordonne
 for thus was it ordained
 'for thus it was ordained'
 (Vance 1988 (26b), 159; Roberts 1993a (102b): 147)

We must therefore conclude that OF had expletive pronouns. We therefore do not know what caused the null-subject parameter to change its value in the history of French, and so we are unable to relate this change to the acquisition of a given value of the null-subject parameter.

The fundamental problem with the Italian–French comparison is that, since Hyams' early work, evidence has emerged that early null subjects are not the result of a 'missetting' of the null-subject parameter to the Italian value. The main reason for this is that early null subjects do not occur in the following environments:

- (8) a. questions with a fronted *wh*-element;
 b. subordinate clauses;
 c. matrix clauses with a fronted non-subject.

On the other hand, null subjects readily occur in these environments in adult null-subject languages, as the following Italian examples (from Guasti (2002: 159)) show:

- (9) a. Cosa __ hai detto? (wh-question)
 what have-2sg said
 'What did you say?'
 b. Gianni ha detto che __ verrà. (subordinate clause)
 John has said that (he) will come.
 c. Ieri __ ho parlato a Carlo. (root clause with fronted adverb)
 Yesterday (I) have spoken to Carlo.

Because of data like this, it has been widely concluded that, despite initial appearances, early null subjects are not a case of missetting of the null-subject parameter to the 'Italian' value. Hence our comparison of Italian and French above in relation to the 'ideal scenario' was to no avail.

Another option, pursued by Hyams (1992), was to claim that early null subjects result not from a 'subject-drop' option of the familiar Italian kind, but from a 'topic-drop' option of the kind seen in languages such as Chinese and Japanese (see in particular Huang (1984; 1989) on this). The advantage of this idea is that it reconciles the occurrence of early null subjects in languages with impoverished agreement systems with the known facts of adult languages: while null-subject languages like Italian appear to require 'rich' verbal agreement for the recovery of the content of null subjects (see §1.1.1), topic-drop languages like Chinese and Japanese have no agreement at all and yet allow null arguments of various kinds, as the following Chinese examples illustrate:

- (10) a. __ kanjian ta le.
 (he) see he Asp
 b. Ta kanjian __ le.
 he see (him) Asp
 'He saw him.'

The disadvantage of this approach is also apparent from (10b): topic-drop languages allow null objects fairly freely, in addition to null subjects. On the other hand, Hyams and Wexler (1993) show that early null objects are rather rare in child English, and Wang *et al.* (1992) show that child Chinese

allows null objects significantly more freely than child English. (For further discussion and statistical evidence, see Guasti (2002: 157–8).) So the idea that the putative parameter-missetting is in the ‘Chinese’ direction rather than in the ‘Italian’ one does not appear to hold up either.

So it appears that there is no obvious ‘parameter-missetting’ going on with early null subjects. Other possibilities which have been explored to account for this phenomenon include relating it to the **diary drop** we briefly saw in §1.1.1. The relevant examples are repeated here:

- (11) a. ... cried yesterday morning.
(Plath 1983: 288)
- b. Elle est alsacienne. ... Paraît intelligente.
She is Alsatian. Seems intelligent.
(Léautaud 1989: 48)

These examples have been argued to involve clausal truncation (but at a higher level of clausal structure than that involved in root infinitives as described above) by Haegeman (2000) and Rizzi (2000). Indeed, a truncation analysis seems to account for this phenomenon quite well; see Guasti (2002: 166ff.) for summary and discussion. The other accounts which have been put forward involve extra-syntactic factors, such as processing difficulties (Bloom 1990) and metrical difficulties (dropping of weakly stressed syllables) (Gerken 1991). As these do not involve factors which may be relatable in any direct way to parametric change, I will leave them aside here. (Once again, see Guasti (2002: 179–83) for discussion.)

In conclusion, for all their intrinsic interest and the light they shed on L1 acquisition, it seems that neither root infinitives nor early null subjects can clearly be related to the kinds of phenomena known in parametric syntactic change. Hence no clear connection can be made between studies of immature competence and the acquisition-driven conception of parametric change.⁵ Although perhaps disappointing, this is not surprising, and does

⁵ Rizzi (2005: 97–100) conjectures that root infinitives and early null subjects (as well as ‘determiner drop’ and ‘copular drop’, two other features of the production of two- and three-year-olds not discussed here) may arise from the fact that ‘the child initially assumes all the parametric values which facilitate the task of the immature production by reducing computational load’ (97). This general strategy is constrained by the values of the parameters which are fixed early, hence the observation that the null-subject parameter is correctly fixed in the acquisition of null-subject languages, but early null subjects and root infinitives may appear in the acquisition of non-null-subject languages owing to the adoption of the conjectured strategy.

not, at least in my view, prove the acquisition-driven conception of parametric change wrong (although very interesting potential evidence that it is right is thereby sadly lacking).

There are various explanations for this state of affairs. First, no-one has really looked carefully for a connection between acquisition and change: there is something of a sociological divide between linguists working on L1 acquisition and those working on diachronic syntax.⁶ This is a regrettable, but entirely contingent state of affairs, and something which can in principle easily be remedied. Second, good empirical coverage of early production is limited to a few languages: English, French, Dutch, German, and Italian most prominent among them to judge by Guasti (2002); in diachronic syntax, selected topics have been studied in the histories of a range of languages, but a good overall picture of the syntactic history of very few languages other than English and French is hardly available. Thus our database of languages is at present extremely small, and so our chances of finding the ideal case correspondingly restricted. Third, and most importantly, the nature of the data in both cases may make the ideal scenario described above hard to identify. The immature competence of small children goes hand in hand with a general cognitive immaturity, notably for example a smaller short-term memory capacity, which means that comparing children's grammars with adult grammars may really be like comparing chalk with cheese. The diachronic data we have is the output of adult competence, but of course the surviving texts have been subject to many vicissitudes of history; one of the principal goals of traditional philology is simply to unravel the sometimes tortuous histories of extant texts. And so in their different ways both the acquisition data and the diachronic data are corrupt, and this makes comparing data from the two sources in any reliable way all the more difficult. Of course, what we would ideally like is an acquisition study of an earlier stage of a language. Since

⁶ As mentioned several times already, Lightfoot has consistently made the connection between syntactic change and the acquisition of syntax. In particular, Lightfoot (1991) develops the 'degree-0 learnability' theory with a view to explaining aspects of both. Lightfoot's application of degree-0 learnability to word-order change in the history of English was briefly discussed in §2.5; he also applies the same notion to aspects of language acquisition. This, Clark and Roberts (1993), and Roberts (1999) are the only cases in the literature where an explicit connection is attempted between change and acquisition (although see also DeGraff (1999)), although the connection is mentioned in Hyams (1986: 23, n. 1).

acquisition studies only began in the mid-twentieth century (see Grégoire (1937–47); Jakobson (1941); Leopold (1939–49)), this is not possible.⁷

Although the conclusion of this section may seem pessimistic, three points are worth bearing in mind. First, as stated above, what is lacking is empirical confirmation through acquisition studies that parameter change is driven by acquisition; the lack of evidence does not disconfirm this approach, especially since we can quite easily see why such evidence may be lacking, and the conceptual arguments (see §2.1) are unaffected. Second, the ‘ideal scenario’ may well be realized by some new data; in fact, the root-infinitive evidence comes close to this, as we saw above, and an enhanced understanding either of this data or of parameter changes involving agreement may yet yield that scenario. Third, as mentioned above, Guasti’s conclusions regarding early and accurate setting of several important parameters, listed in (2), lead us to pose the intriguing logical problem of language change. It is to this last issue that we now turn.

3.2. The logical problem of language change

In this section, we will look in more detail at the apparent tension we noted in the previous section between the evidence that many important parameters are set at an early stage of language acquisition and the idea that syntactic change is driven by abductive reanalysis, associated with parametric change, of parts of the PLD. This will lead us to formulate the logical problem of language change, as a kind of paradox for learnability theory. (At the same time, some basic concepts of learnability theory are introduced.) In considering a possible approach to solving this problem, we

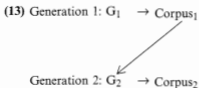
⁷ I am aware of one striking exception to this generalization: the journal kept by Jean Héroard in the period 1601–28 of the speech of the young dauphin Louis XIII (Ayres-Bennett 1996: 216ff.) Héroard, who was the dauphin’s personal physician, transcribed samples of the dauphin’s speech between the ages of 3;3 and 9;3 (i.e. in the years 1605–10). This immediately yields some interesting observations, notably that *ne* was already dropped, i.e. the dauphin’s production represented a very early instance of Stage III of Jespersen’s Cycle (Ayres-Bennett 1996: 221; see §2.2, note 8 and §1.4.2). Ernst (1985) is an edition of Héroard’s journal. Ayres-Bennett (2004: 185ff.) provides further examples of the sporadic omission of *ne* in seventeenth-century French.

introduce the **Inertia Principle** first put forward in early version of Keenan (2002), as developed in Longobardi (2001). This in turn leads to an explicit characterization of the conditions under which abductive change can take place, in terms of ambiguity and complexity. The discussion will sharpen some of the notions connected to parameter-setting that we are concerned with, and so our overall account of syntactic change will be further elucidated, and a number of questions raised for detailed consideration in later sections of this chapter.

Let us begin by recapitulating some of the ideas we have put forward up to now:

- (12) a. The central mechanism of syntactic change is parameter change.
 b. Parameter change is manifested as (clusters of) reanalyses.
 c. Reanalysis is due to the abductive nature of acquisition.

Concept (12a) was argued for at length in Chapters 1 and 2, and we noted in §2.3 that parameter change may not be the only mechanism of syntactic change, although it does appear to be the principal one. Both (12b) and (12c) were introduced in the discussion of reanalysis in §2.1. There we also saw Andersen's (1973) familiar schematization of abductive change, which I repeat here:



Owing to the abductive nature of language acquisition, G_2 may in principle not be identical to G_1 . If identity between grammars is defined in terms of identity of parameter-settings, this implies that G_2 may differ from G_1 in at least one parameter value, and, by (12b), this will give rise to a cluster of reanalyses. Following Roberts and Roussou (2003: 11), we can give the following characterization of abductive change (cf. also Lightfoot (1979; 1991; 1999)):

- (14) (A population of) language acquirers 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 those acquirers.

This essentially states what the schema in (13) illustrates. So where G_2 may differ from G_1 in at least one parameter value in (13), an abductive change takes place.

Let us consider (13) and (14) in the light of the main concepts of learnability theory, the abstract, formal theory which deals 'with idealized "learning procedures" for acquiring grammars on the basis of exposure to evidence about languages' (Pullum 2003: 434). In terms of this theory, any learning situation can be characterized in terms of the answers to the following questions (this presentation is from Bertolo (2001: 2ff.)):

- (15) a. 'What is being learned ...?'
 b. What kind of hypotheses is the learner capable of entertaining?
 c. How are the data from the target language presented to the learner?
 d. What are the conditions that govern how the learner updates her responses to the data?
 e. Under what conditions, exactly, do we say that a learner has been successful in the language learning task?

The answers to some of these questions are obvious, given the assumptions we are making here regarding UG and parametric variation: for example, the answer to (15b) is that the learner can only consider distinct values of parameters in the acquisition of syntax. Similarly, the answer to the question of how the data are presented to the learner – (15c) – appears to be simply in the form of spontaneous linguistic behaviour which makes up the PLD. As we saw in the Introduction to Chapter 1, no negative evidence is available to the learner. One could put the answer to this question more abstractly, and say that the data is presented in the form of P-expression – as defined in (8) of §2.1 – in the strings in the PLD.

Some of the other questions in (15) are trickier. For instance, our entire discussion of the possibility of evidence of 'parameter-missetting' in the previous section can be construed as addressing (15d). On the basis of that discussion, we have to conclude that no updating of parameter values takes place, which might be relevant. But it is (15a) and (15e) which are most important to our concern with the relation between learnability/acquisition and change. One possible answer to (15a) is that acquirers are learning the parameter values of the grammar that produces the PLD. But in that case the learning task would be seen as unsuccessful when abductive parametric change of the sort schematized in (13) takes place. This seems to be the wrong conclusion, since the learners in (13) have acquired *a* grammar, just not the parental one (since $G_1 \neq G_2$). So the answer to (15a) would be simply that a parametric system is learned. This point relates to (15e), too: the criterion for successful learning cannot be replication of the parental grammar, but approximation to it, in such a way as abductive change of the

sort shown in (15) is possible. For this last reason, it is often thought that **Probably Approximately Correct (PAC) algorithms are learning algorithms** which can provide useful simulations of language acquisition. (See Bertolo (2001: 8–10); Clark and Roberts (1993); Pullum (2003: 433); and in particular Niyogi (2004: 75ff.) and the references given there.)

Now, most work on L1 acquisition assumes that the stable state of acquisition, S_g , corresponds exactly to the target grammar. In other words, it assumes that no ‘mismatch’ arises between G_1 and G_2 in (13). In considering the stages of L1 acquisition in the previous section, we implicitly took this view in saying ‘we can consider that the various intermediate states differ from one another in representing successively closer approximations to the adult system (S_g) in terms of the values of the parameters. To put it another way: if m parameters are set to the adult value at stage S_n then at least $m + 1$ parameters are set to the adult value at stage S_{n+1} .’ But, as we have seen, abductive change requires a slightly looser formulation than this, in order to allow for the possibility of ‘mismatch’ between G_1 and G_2 in (13).

The following remark by Niyogi and Berwick (1995: 1) summarizes the difference between the standard assumptions in work on language acquisition and what seems to be required for an acquisition-driven account of change: ‘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’.

Thus language acquisition is usually taken to be deterministic in that its final state converges with the target grammar that acquirers are exposed to. The postulation of abductive change challenges exactly this assumption.

In the previous section, we saw good support for the deterministic assumption in L1 acquisition. Recall Guasti’s list of parameter values which appear to be correctly fixed from roughly the time of the earliest multiword utterances:

- (2) a. the value of the head direction parameter;
 b. the value of the V-to-T parameter;
 c. the value of the topic-drop and null-subject parameters;
 d. the value of the parameters governing question formation, the one governing overt movement or *in-situ* placement of the wh-element and the one regulating T-to-C movement (inversion).

Moreover, we saw in the previous section that two of the best studied phenomena of child production, root infinitives and early null subjects, are not easily or obviously analysed in terms of ‘parameter-missetting’, i.e. in terms of mismatches between G_1 and G_2 in (13). As Roberts and Roussou (2003: 12) put it: ‘the standard paradigm for language acquisition is not immediately compatible with the observation that grammatical systems change over time.’ This ‘standard paradigm for language acquisition’ is more than just a methodological simplification on the part of linguists working on L1 acquisition: (2) gives the evidence of accuracy and earliness in the setting of a number of important parameters. And yet we saw in Chapter 1 that all the parameters referred to in (2), with the exception of topic drop, have changed their values in the recorded histories of various languages. So we see a tension between the results of L1 acquisition research and what we can observe about syntactic change.

The obvious explanation for the fact that children are able to set many parameters very early lies in the highly restricted range of analyses of the PLD that UG allows them to entertain and the limited exposure to PLD needed for parameter fixation. (In this respect, the facts reported in (2) support the argument from the poverty of the stimulus, as noted in the previous section.) But then, how are we to explain the fact that these parameters are subject to change over time? Notice that our answers to the learnability questions in (15) do not answer this question; they simply allow for abductive parametric change without explaining how it happens.

This leads us to the logical problem of language change, which we can formulate as follows (this formulation is based on unpublished work with Robin Clark (Clark and Roberts 1994: 12)):

- (16) If the trigger experience of one generation permits members of that generation to set parameter p_k to value v_i , why is the trigger experience produced by that generation insufficient to cause the next generation to set p_k to v_i ?

The logical problem of language change as formulated here is close to the Regress Problem for reanalytical approaches to change which we introduced in §2.1 in the following terms: ‘an innovation in Corpus_2 [in (13)] may be ascribable to a mismatch in G_2 (compared to G_1), but it must have been triggered by something in Corpus_1 – otherwise where did it come from? But if Corpus_1 could trigger this, then how could G_1 produce this property without itself having the innovative property?’ Essentially, this

formulation puts the problem the other way around as compared to (16), in addition to not being directly phrased in terms of parameter change. I take it, though, that (16) subsumes what I called the Regress Problem in the earlier discussion.

The first thing that is required in order to find our way out of the dilemma stated in (16) is a slightly weaker notion of the deterministic nature of L1 acquisition than that which is usually assumed in the L1-acquisition literature. So let us propose, following Roberts and Roussou (2003: 13), that 'the goal of language 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'. More precisely, as mentioned in the discussion of (15e) above, let us suppose that the goal of acquisition is to *approximate* the parental grammar, not to *replicate* it. Making this move allows for p_k in (15) to receive a different value from that found in the input, therefore making space for language change. The stable state S_S of language acquisition now amounts to the situation where all parameters are fixed to a given value (cf. the remark in relation to parameter-setting in §1.1 that 'not deciding is not an option'). Let us call this view of the endpoint of language acquisition 'weak determinism'.

The 'approximation' approach may seem too weak, in particular in that it does not appear to account for the results of L1-acquisition research as summarized in (2), since it in principle allows parameters to vary freely and randomly from generation to generation. However, Roberts and Roussou (2003: 13) add an important proviso to the above quotation to the effect that convergence with the adult grammar 'happens most of the time'; that is, approximation usually amounts to replication. This brings us to an important principle of syntactic change, first put forward in Keenan (2002): the Inertia Principle. Keenan formulates this as follows:

- (17) Things stay as they are unless acted on by an outside force or decay.
(Keenan (2002:2))

This principle is very general; in fact it holds of the physical world in general, taking decay to include entropy, i.e. the second law of thermodynamics. For our purposes, we can take (17) to mean that, although L1 acquisition is not inherently deterministic but rather weakly deterministic in Roberts and Roussou's sense, the target system is successfully converged on, i.e. the stable state S_S of acquisition has the same parameter values as that of the parent system; G_1 and G_2 in (13) do not differ. This is no doubt

due to the highly restricted range of analyses of the PLD that UG allows and the limited exposure to PLD needed for parameter fixation, i.e. standard poverty-of-stimulus considerations.

Longobardi (2001: 278) adopts Keenan's principle, and puts forward the following very interesting version of it:

- (18) 'syntactic change should not arise, unless it can be shown to be *caused*'
(emphasis his)

In other words, as Longobardi says, '*syntax*, by itself, is diachronically completely inert' (277–8). It is clear that this view is compatible with the results of L1-acquisition research, as reported in the previous section and in (2). If we combine (18) with Roberts and Roussou's weak determinism, we arrive at the following:

- (19) If a definite value v_i is expressed for a parameter p_i in the PLD, then (a population of) acquirers will converge on v_i .

In other words, given adequate P-expression, inertia will hold. So inertia implies that most of the time abductive change does not happen. P-expression was introduced and defined in §2.1; the definition, and the corollary definition of trigger, is repeated here:

- (20) a. 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.
- b. Trigger:
A substring of the input text S is a trigger for parameter p_i if S expresses p_i .

As long as there is a trigger for a given parameter value, then, inertia will hold and abductive change will not take place.

Under what circumstances does abductive change happen, then? This must be when no definite value v_i is expressed for a parameter p_i in the PLD. According to Longobardi's version of Inertia in (18), this lack of robust P-expression must be 'a well-motivated consequence of other types of change (phonological changes and semantic changes, including the appearance/disappearance of whole lexical items) or, recursively, of other syntactic changes' (278).

More precisely, we propose, following our discussion of reanalysis in §2.1, that both ambiguity and opacity of the P-expression are required in

order for abductive change to take place. Ambiguity is defined in relation to parametric systems as follows:

(21) a. P-ambiguity:

A substring of the input text S is strongly P-ambiguous with respect to a parameter p_i just in case a grammar can have p_i set to either value and assign a well-formed representation to S .

b. A strongly P-ambiguous string may express either value of p_i and therefore trigger either value of p_i .

c. A weakly P-ambiguous string expresses neither value of p_i and therefore triggers neither value of p_i .

In fact, as we saw in §2.1, strong P-ambiguity is what is required for reanalysis. These definitions are repeated from that discussion, where they were illustrated in relation to certain reanalytical changes. Weak P-ambiguity arises where some parameter value is undetermined, also possibly leading to change in a parameter value, although not necessarily through reanalysis.

We can define opacity in terms of complexity (see Lightfoot (1979)). Following an idea developed in Clark and Roberts (1993), let us assume that learners are conservative in that they have a built-in preference for relatively simple representations (the precise characterization of simplicity will be discussed in §3.4 and §3.5). If a given piece of PLD is P-ambiguous, there will be at least two representations for it, each corresponding to a different grammar, i.e. representing systems with distinct parameter values. Assuming that any two representations differ in complexity and therefore opacity, the learner will choose the option that yields the simpler representation. The more complex representation will be both opaque (in virtue of being more complex than the other available representation(s)) and ambiguous (by definition). Therefore it is inaccessible to the learner, i.e. it is effectively unlearnable. The Inertia Principle tells us that the strong P-ambiguity of the trigger (and therefore the relative opacity, assuming that any two alternative representations differ in overall complexity) arises through either extra-syntactic factors or as the consequence of an independent syntactic change.

Actually, closer reflection reveals that the circumstances just described do not guarantee a change; they merely suspend inertia, since we can take it that P-expression (and therefore strong P-ambiguity) and the preference for relative simplicity are forces acting on the learner, in the sense relevant for the Inertia Principle as stated in (17). Hence, it is possible that things

will not stay as they are. Whether a parameter value of G_1 actually changes will depend on the relative complexity of the representations of aspects of the PLD entailed by the parameter-settings in G_1 compared to those of G_2 : a parameter will change if it corresponds to the single option expressed by the PLD for G_1 and the more opaque of two options expressed by the strongly ambiguous PLD for G_2 . Here again weak determinism is relevant, in that it implies that under these conditions a definite value v_i for p_i will be assigned. This value will still be compatible with the input, but – again thanks to weak determinism – may differ from that of the target grammar, in which case an abductive change takes place. So we see that the simplicity metric is the ‘safety mechanism’ alluded to earlier.

So, our tentative answer to the question posed in (16) is that between the two generations in question there is a change in the PLD. In other words, some extra-syntactic factor, or at least a factor independent of the change in question, introduces P-ambiguity into the expression of at least one parameter in the PLD of G_2 . Still assuming that any two representations can be distinguished in terms of complexity (and still leaving complexity undefined, for the time being), complexity/opacity will then choose between the two values, possibly leading to a parametric change.⁸ The crucial question becomes that of locating the factors that may introduce P-ambiguity into the PLD. This is what we turn to in the next section.

Clearly, all of the above discussion turns on the notion of complexity, which we must therefore now define. The commonest way of determining this is by simply counting some aspect of a derivation or representation. (See in particular Chomsky and Halle (1968, Chapter 9) on complexity and markedness in phonological systems; we return to the discussion of markedness in §3.4.) As Roberts and Roussou (2003: 200) point out, syntactic representations offer several possibilities:

In principle, there are several formal options available in syntactic representations or derivations: one could count nodes, branching nodes, traces [i.e. copies – IGR], chain links, symbols or features.

⁸ Kroch (2000: 700) points out that it is also possible that ‘extrasyntactic’ change may be attributable to some property of the learner, for example, age at the time of acquisition. This is relevant ‘in the case of change induced through second-language acquisition by adults in situations of language contact’ (*ibid.*). I will discuss this case briefly in the next section, returning to it in detail in Chapter 5.

After considering the various options, Roberts and Roussou opt for a feature-counting approach; this is in fact very much in the original spirit of Chomsky and Halle (1968). Here I give a slightly simplified version of Roberts and Roussou's proposal (for the original, see Roberts and Roussou (2003: 201)):

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

The notion of 'formal feature' here is the standard one in current versions of syntactic theory, as introduced in §1.4.1: it includes features such as Person, Number, Gender, Case, and Negation. And, as we saw in §2.5, movement takes place where the Probe of an Agree relation has uninterpretable formal features and 'an extra property' triggering movement. Chomsky (2000; 2001) proposes that that 'extra property' is in fact a further formal feature known as the EPP feature. This means that Probes, in terms of formal features, are more complex than non-Probes, and Probes that cause movement are more complex than those which do not. We will see the effects of this definition of complexity in more detail in the next section.

In this section, we resumed certain aspects of the discussion of reanalysis in §2.1, notably the question of abductive change. Applying this idea strictly to parameter change, we arrived at the logical problem of language change as stated in (16), partly on the basis of some of the observations regarding language acquisition made in the previous section. We suggested a way of solving this problem on the basis of the Inertia Principle of (17) and the corollary stated by Longobardi in (18). We are led to the conclusion that abductive parametric change only occurs when the trigger for a given parameter value, as defined in (20b), is both ambiguous and opaque. P-ambiguity is defined in (21) and opacity/complexity in (22). P-ambiguity can only be introduced through extrasyntactic factors, for example, through language contact (see note 8), morphophonological erosion, or through an independent syntactic change.

In the remaining sections of this chapter, I will address the two main issues that this account of change raises: the nature of changes to the trigger, and the nature of complexity as the principal force which acts on parameter-setting, preventing things from staying the same (see (17)); this will be linked to the concept of markedness of parameter values. Finally I will draw these threads together and attempt a formal characterization of parameters.

3.3. The changing trigger

Following on from our rather abstract discussion of the logical problem of language change, this section looks at the idea that changes in the trigger experience – in particular the introduction of strong P-ambiguity – are responsible for language acquirers resetting parameter values. In other words, we investigate what both Kroch (2000: 700) and Longobardi (2001: 277–8) put forward implicitly as the solution to the problem: the idea that intergenerational changes in the PLD render earlier parameter-settings prone to abductive change. We saw that this is due to the expression of these parameters becoming ambiguous and opaque.

I will discuss three ways in which the PLD may be rendered ambiguous and/or opaque: contact-driven parameter-resetting, as suggested in King (2000), Kroch and A. Taylor (1997) and Kroch, Taylor, and Ringe (2000); cue-based resetting of the type advocated by Drescher (1999) and Lightfoot (1999) (the cue-based model was originally proposed by Drescher and Kaye (1990)); and morphology-driven parameter-resetting, as suggested in Roberts (1985; 1999); Roberts and Roussou (2003). These ways of rendering the PLD ambiguous and opaque are not mutually exclusive; it is very likely that all three possibilities exist.

3.3.1. *Contact-driven parameter-resetting*

The contact-driven view of parameter-resetting can be construed, in our terms, as the case where PLD is affected by an alien grammatical system. What this means is that Generation 2 in the schema for abductive change in (13) is subjected to a different kind of PLD from Generation 1 in that Generation 2 receives PLD that either directly or indirectly reflects a distinct grammatical system (i.e. set of parameter values) from that which underlay the PLD for Generation 1.

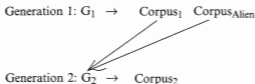
The direct case of contact is that where the PLD simply contains a significant quantity of tokens from a distinct system (where ‘distinct’ means that the grammar in question generates strings that cannot express the original grammar); this would naturally arise where Generation 2 is brought up in an environment which contains a language or dialect absent from the early experience of Generation 1. Such situations can and do arise through

many different types of historical contingency: emigration, invasion, and intermarriage being the most obvious but certainly not the only ones.

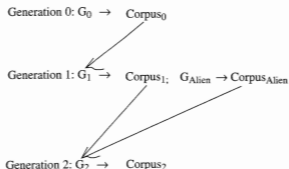
The indirect case of contact arises where Generation 1 uses a second language in interaction with Generation 2. Here the PLD for the two generations is very obviously distinct. In the case where this second language is a pidgin, Generation 2 may form a creole; I will leave this particular situation aside here and return to it in §5.3. If Generation 1 has learnt the second language after the critical period for language acquisition, the PLD will consist of **interlanguage** and many parameter-settings may be radically underdetermined by the PLD. (I will come back to this in the discussion of second-language acquisition and the nature of interlanguage in §5.1.) This situation gives rise to weak P-ambiguity and hence potentially to change, which may have the properties of the creation of new grammatical features *ex nihilo*.

The direct and indirect cases of language contact influencing PLD are diagrammed in (23):

(23) a. Direct contact:



b. Indirect contact:



As already mentioned, issues to do with the nature of language contact, second-language acquisition and possibly creolization arise here, but we will leave them aside and return to them in Chapter 5.

An example of what looks like direct contact, that is, contact-induced borrowing affecting aspects of syntax, comes from King's (2000) study of the French spoken on Prince Edward Island in Canada. Prince Edward Island (PEI) French is a variety of Acadian French, also spoken in the other Canadian Maritime Provinces (New Brunswick and Nova Scotia) and in parts of Newfoundland. French has been spoken in Acadia since the early seventeenth century (King 2000: 7), but according to the 1991 census, only 4.2 per cent of the population of PEI were native speakers of French, and only 2.3 per cent spoke the language at home. 86 per cent of these people lived in a single area, Prince County (King (2000: 19)). Contact with English is clearly very extensive, and King (2000, Chapter 6) documents much lexical borrowing and code-switching.

What is of interest in the present context is that King (Chapter 7) documents cases of Preposition-stranding in PEI French. As we saw in Box 1.5 of Chapter 1, Preposition-stranding is the cross-linguistically rather rare option of moving the complement of a preposition, while leaving the preposition 'stranded'. English and Mainland Scandinavian languages allow this, both in *wh*-questions, illustrated in (24a), and in passives, shown in (24b) (Icelandic in fact allows the equivalent of (24a), but not (24b) with a nominative subject: Kayne (1984: 117)):

- (24) a. Who did you speak to __?
b. John was spoken to __.

It is not clear what exactly permits this in English and the Scandinavian languages. In most other languages which have overt *wh*-movement, the preposition must move with the *wh*-phrase, i.e. it must be pied-piped (see §1.5 on the *wh*-movement parameter). This is the case of Standard French, for example (both examples in (25) are equivalent to (24a)):

- (25) a. *Qui as-tu parlé à __?
b. A qui as-tu parlé __?

Of course, English also allows the equivalent of (25b) (*To whom did you speak __?*). This is probably a case of formal optionality in English; a [+*wh*] C with an EPP feature can cause either the DP complement of the preposition or the whole PP to move, although Agree_{wh} holds just between

C and the DP. (We will look in more detail at the concept of formal optionality in §4.1.) Optionality has social value, in that the pied-piping option is characteristic of formal registers. French has some property which requires pied-piping of the DP in all cases. Kayne (1984) made an interesting and influential proposal concerning this, but in the context of technical assumptions which have not been carried over into most versions of minimalism. The central idea in Kayne's analysis was that English prepositions resemble verbs in that their complements are always accusative and able to undergo movement. French prepositions, on the other hand, have inherently Case-marked complements (in the sense defined in §2.3.2), which cannot undergo movement independently of the preposition. So the parameter distinguishing French from English is connected to the differential lexical properties of prepositions in the two languages. The existence of inherent Case in French is morphologically signalled by the contrast between dative and accusative 3rd-person clitic pronouns, for example, *le* (acc) vs. *lui* (dat.); English has no comparable contrast. (Icelandic distinguishes accusative and dative complements, and has Preposition-stranding of some types, as mentioned above; see Kayne (1984: 117) for discussion).

PEI French behaves like English in allowing Preposition-stranding. (26) illustrates this in a wh-question, a relative clause and a passive:⁹

- (26) a. Où ce-qu'elle vient de __?
 where that she comes from
 'Where does she come from?'
 (King 2000: 136, (5))
- b. Ça, c'est le weekend que je me souviens de __
 That it is the weekend that I me remember of
 'That's the weekend that I remember.'
 (King 2000: 136, (6))
- c. Robert a été beaucoup parlé de __ au meeting.
 Robert has been much talked of at-the meeting
 'Robert was talked about a lot at the meeting.'
 (King 2000: 141, (32))

⁹ King points out (138) that Preposition-stranding has been observed in Montreal French (by Vinet (1984: 239)), but the phenomenon is much more restricted in that variety, according to King's account of Vinet's observations. Roberge (1998; 1999) surveys a range of Canadian varieties of French and observes that Alberta French is intermediate between Montreal French and PEI French.

Data like this indicate that PEI French has what appears to be English-style Preposition-stranding. Thus, PEI French, thanks to the very extensive contact with English, has developed a parametric option which is lacking in Standard French, and not fully instantiated in Montreal French; see note 9.

King insightfully and convincingly relates PEI French Preposition-stranding to the fact that PEI French has borrowed a number of English prepositions. Some of these are illustrated in (27):

- (27) a. Ils avont layé off du monde à la factorie.
 They have laid off some people at the factory
 'They have laid off people at the factory.'
 (King 2000: 142, (39))
- b. Il a parlé about le lien fixe.
 he has talked about the link fixed
 'He talked about the fixed link.'
 (King 2000: 143, (41))

These prepositions also allow stranding:

- (28) a. Qui ce-qu'a été layé off __?
 who that has been laid off
 'Who has been laid off?'
 (King 2000: 142, (40))
- b. Quoi ce-qu'il a parlé about __?
 what that he has talked about
 'What did he talk about?'
 (King 2000: 143, (42))

King argues that 'the direct borrowing of English-origin prepositions has resulted in the extension of a property of English prepositions, the ability to be stranded, to the whole set of Prince Edward Island prepositions' (147). If the option of stranding is genuinely a lexical property of prepositions, as roughly sketched in our remarks on Kayne (1984) above, then we might expect that option to be borrowed with the English prepositions, although PEI French does retain an accusative-dative distinction in pronouns, as King (2000: 64, Table 5.2) shows. So here we have a fairly clear case of direct contact: at some point in the history of PEI French, elements from an alien grammatical system – English prepositions – were borrowed and this affected the parameter governing Preposition-stranding. (Acquirers seem to have generalized the input based on the English prepositions; we will return to this notion of 'generalization of the input' in §3.5.) This is

borrowing, rather than in any sense imperfect learning of French by native speakers of French; nor does it reflect imperfect learning of English by French speakers.

In fact, PEI French appears to allow Preposition-stranding in contexts where English does not. Hornstein and Weinberg (1981) observed that examples like the following are unacceptable for most English speakers:¹⁰

(29) *Who did Pugsley give a book yesterday to __?

PEI French appears to allow examples equivalent to (29):

- (30) a. Quoi ce-que tu as parlé hier à Jean de __?
 what that that you have spoken yesterday to John of
 'What did you speak yesterday to John about?'
 (King 2000: 146, (53))
- b. Quoi ce-que tu as parlé hier de __ à Jean?
 what that that you have spoken yesterday of to John
 'What did you speak yesterday about to John?'
 (King 2000: 146, (57))

King relates this to the independent fact that 'French does not have the strong adjacency requirements found in English' (147).¹¹

The Prince Edward Island case is, as we said, a clear case of direct contact affecting the trigger experience. Reanalysis is relevant to the extent that the structure [_{PP} P DP] changes its properties as DP becomes extractable from PP. It is hard to evaluate the role of strong P-ambiguity here, as the pied-piping option is apparently only available with certain prepositions; with *de*, it is not found: **De quel enfant as-tu parlé?* ('About which child did you speak?') is not good, but in other cases it is possible: *Pour quelle raison qu'il a*

¹⁰ Hornstein and Weinberg suggest that this is because the verb must c-command the stranded preposition, but that in (29) the PP is 'extraposed' outside of VP and to its right, as its position relative to the adverb *yesterday* shows, and therefore the PP is not c-commanded by the verb. The relevant parts of the structure would thus be approximately as in (i):'

(i) [... [_{VP} V] yesterday] PP

The definition of c-command was given in §1.4, (90).

¹¹ It could in fact be connected to the possibility of raising a participle to a slightly higher structural position than that occupied by English participles, allowing the verb a wider range of c-command possibilities, roughly in line with what is suggested in note 10. On raising of French participles, see Pollock (1989: 417). We mentioned movement of non-finite verbs in French in §1.3.1, note 10.

parti?/Quelle raison qu'il a parti pour? ('For what reason did he leave?/What reason did he leave for?') (Ruth King, p.c.). It is clear, though, that strings with Preposition-stranding crucially affected the PLD at the time of contact, leading PEI French to change the value of the Preposition-stranding parameter, thus creating a difference with Standard French.

3.3.2. Cue-driven parameter-resetting

Let us now turn to the second way of making the trigger experience ambiguous and opaque: the cue-based approach. Lightfoot (1999) and Dresher (1999) argue that learners use input forms, i.e. pieces of PLD, 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) on the potential importance of fragments of sentences for parameter-setting). 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. For example, Dresher (30ff.) proposes that there is a parameter determining whether a given language's stress system is quantity-sensitive (QS) or not (Q(uality)I(n)sensitive)). English, for example, is QS, in that the basic stress rule states that the penultimate syllable is stressed if heavy; otherwise the antepenult is stressed (cf. *Cánada*, with a non-heavy CV penult, as opposed to *Vancóu:ver*, with a heavy CV: penult). Thus the heaviness (or quantity) of a syllable plays a role in determining stress-assignment. Not all languages have quantity-sensitive stress-assignment; QS thus represents a value of a particular parameter. The parameter in question is formulated as follows:

(31) *Quantity (in)sensitivity*

- a. *Parameter*: The language {does not/does} distinguish between light and heavy syllables ...
- b. *Default*: Assume all syllables have the same status (QI).
- c. *Cue*: Words of *n* syllables, conflicting stress contours (QS).
(Dresher's (7): 31)

Dresher (1999: 31) points out that:

In QI systems all words with *n* syllables should have the same stress contour, since they are all effectively equivalent. Taking quantity insensitivity to be the default case, a learner will continue to assume that stress is QI until it encounters evidence that words of equal length can have different stress contours.

Dresher goes on to point out that there is abundant evidence that English is not QI, and so the learner quickly sets this parameter to the QS value. Thus we see how the unmarked value requires no evidence, and the marked value is associated with the cue. We will look more closely at the question of the marked and default settings of syntactic parameters in the next section.

Lightfoot (1999: 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'. He illustrates this approach with the loss of V-to-T movement in ENE. He assumes that the NE situation whereby tense and agreement morphology is realized on V is the default (Lightfoot says that this is a morphological rule, but we can continue to think of it as an instance of Agree; see §1.4.1), and so the V-to-T grammar needs the cue $[_T V]$ ($[_I V]$ in his notation). Lightfoot suggests that the main expression of the cue (where his notion of 'expression' is like our notion of P-expression in being a structure which requires the cue in order to be grammatical) to be subject-verb inversion, as in (75) of Chapter 1, repeated here as (32):

(32) What *menythe this pryste?*

What does this priest mean?

(1466-7: Anon., from J. Gairdner (ed.), 1876, *The Historical Collections of a London Citizen*; Gray 1985: 11; Roberts 1993a: 247)

This cue was perturbed by three factors: (i) the reanalysis of modals as T-elements (see §2.1); (ii) the development of dummy *do* in the sixteenth century, also a T-element (we briefly mentioned this in our discussion of the loss of V-to-T in §2.1; it may well have been the same change as that affecting modals); (iii) the loss of V2, which clearly took away many environments in which verb-subject order had formerly been found. He concludes:

with the reanalysis of the modal auxiliaries, the increasing frequency of periphrastic *do*, and the loss of the verb-second system, the expression of $_I[V]$ in English became less and less robust in the PLD. That is, there was no longer anything very robust in children's experience which had to be analysed as $_I[V]$, which *required* V to I, given that the morphological I-lowering operation was always available as the default. (Lightfoot 1999: 164)

This account is very similar to the one we proposed in §2.1, with the notion of cue playing the role of our notion of P-expression. Lightfoot also points

out that 'weak' verbal agreement inflection is a precondition for the change ('the *possibility* of V to I not being triggered first arose in the history of English with the loss of rich verbal inflection' (164)), although he does not explicitly say that this inflection is the expression of a cue, which would be the equivalent in his terminology of our claim in §2.1 that the relevant agreement morphology expresses the parameter.

The similarities between Lightfoot's cue-based account of the loss of V-to-T and the one we gave in §2.1, are, as we see, very close. In fact, the definition of 'trigger' in (20b), in making reference to 'a substring of the input text S', is equivalent to the Lightfoot/Dresher notion of cue. However, in this sense, cues cannot be identified with parameters: parameters are abstract properties of grammars, features of part of an individual's mental representation. Although the notion of cue is useful, it must be kept distinct from the notion of parameter.

A further point is that Lightfoot's cue-based 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); if we do this we do not run into this difficulty. So, it seems reasonable to take the view that cues, i.e. triggers as defined in (20b), are provided by the input; parameters are specified by UG and are set by the learner on the basis of the interaction of cues/triggers and UG (and internal properties of the learner – see §3.4 and §3.5). Construed this way, the Lightfoot/Dresher view is essentially the one I have been presenting here, as the close similarities in the analysis of the loss of V-to-T show.

However, there is a difference: Lightfoot has no account for the shift in the cue. He says:

this model ... has nothing to say about why the distribution of cues should change. This may be explained by claims about language contact or socially defined speech fashions, but it is not a function of theories of grammar, acquisition or change – except under one set of circumstances, where the new distribution of cues results from an earlier grammatical shift; in that circumstance, one has a 'chain' of grammatical changes. One example would be the recategorization of the modal auxiliaries ..., which resulted in the loss of V to I.
(Lightfoot 1999: 166)

A morphologically-based approach like that sketched earlier can, on the other hand, explain the change in the cue/P-expression in terms of morphological loss. Let us now return to that account.

3.3.3. Morphologically-driven parameter-resetting

It is worth making two points here. First, we observed in §2.1 that there was a strong P-ambiguity in the analysis of very simple positive declarative sentences like *John walks* in sixteenth-century English, as schematized in (33) (repeated from (13) of Chapter 2):

- (33) a. John [_T walks] [_{VP} ... (walks) ...]
 b. John T [_{VP} walks]

We saw that Lightfoot also makes this observation. (33a) represents the conservative structure with V-to-T movement, and (33b) the innovative structure without V-to-T movement. The proposal was that the conservative system was preferred as long as there was a morphological expression of V-to-T movement through the agreement system. This was stated in terms of the following postulate, linking agreement marking on the verb to V-to-T movement (repeated from (18) of Chapter 2):

- (34) If (finite) V is marked with person agreement in all simple tenses, this expresses a positive value for the V-to-T parameter.

We then proposed that the loss of much verbal agreement, particularly plural endings in both simple tenses, led to the loss of the morphological expression of the V-to-T parameter and thus to a reanalysis of (33a) as (33b) with the concomitant change in the value of the V-to-T parameter. Since (33a) contains an occurrence of movement, V-to-T movement, missing from (33b), it must have at least one more formal feature than (33b). Hence, by (22), it is more complex than (33b). So here the crucial factor creating ambiguity and opacity in the PLD is the erosion and loss of certain endings, something I take to be a morphological property. (In fact, it is more than likely that it is ultimately phonological; see Lass (1992: 134ff.)) This solves the Regress Problem and gives an account of what changed in the P-expression/cue-expression, unlike Lightfoot's analysis.

The second point concerns weak P-ambiguity, as defined in (21c). As a comparison of the definition of strong P-ambiguity in (21b) and that of weak P-ambiguity in (21c) shows, the essential difference lies in the fact that

a weakly P-ambiguous string triggers neither value of a parameter. This notion can be relevant to understanding certain aspects of change in that an independent change can render a former trigger weakly P-ambiguous, i.e. render it irrelevant for triggering some value of a parameter that it triggered prior to the independent change. The loss of V-to-T movement in ENE exemplifies this. The reanalysis of the modals and *do* as functional elements merged in T had the consequence that examples like (35) (repeated from (14) of Chapter 2) no longer triggered anything, i.e. they were weakly P-ambiguous:

- (35) a. I may not speak.
b. I do not speak.

Prior to reanalysis of modals and *do* as T-elements, such examples provided an unambiguous trigger for a positive setting for the V-to-T parameter, in that modals and *do* were verbs (with plural agreement marking) which moved to T and expressed the morphological trigger for V-to-T movement. Once modals and *do* are merged in T, such sentences become weakly P-ambiguous in relation to the V-to-T parameter in that they are compatible with either value of the parameter. As such, an important, and frequently occurring, kind of trigger for the positive setting of the V-to-T parameter is lost. Weak P-ambiguity may be relevant to understanding certain changes in this way. (We will discuss weak P-ambiguity more in §5.3 and §5.4.)

Another change discussed in Chapter 2 illustrates further how morphological change may affect the PLD in such a way as to create ambiguity and opacity in triggers and hence abductive change. This concerns the loss of so-called recipient passives in thirteenth-century English, as discussed in §2.3.2. There we schematized the crucial reanalysis as in (36) ((43) of Chapter 2):

- (36) [_{CP} Him+DAT [_{TP}[_T was] [_{VP} v [_{VP} helped (him+DAT)]]]]>
[_{TP} He+NOM [_{TI_{OP}} was] [_{VP} v [_{VP} helped (he+NOM)]]]

We treated this reanalysis as directly caused by the loss of dative-case morphology, i.e. by the loss of any morphological distinction between morphological accusative and morphological dative case. This led, we proposed, to a parametric change in the nature of *v*, in that it henceforth had a new uninterpretable Case feature which was available in all transitive clauses, with the consequence that all inherently Case marked non-subject arguments were lost.

We tied the reanalysis in (36), and the associated parametric change in the Case-features of *v*, to the loss of case-marking distinctions among complements. This is directly supported by Allen's (1995) detailed account of the breakdown of case morphology in English, as we saw. Let us now consider the effect of the loss of dative-case morphology on the relevant PLD in more detail. Consider a variant of (36) with a non-pronominal argument, which therefore after the loss of dative-case morphology is fully ambiguous as to which Case-feature it bears. We also make the argument singular, so that verbal agreement cannot tell us whether it is Nominative. The string in question is thus *the man was helped*. One could assume that the situation is straightforward here: if there is no dative morphology there is no abstract Dative Case. However, two points militate against this simple assumption. First, Allen (1995: 351ff.) shows that direct passives (i.e. those with Nominative subjects) appear in dialects where the dative/accusative distinction still remains. Second, the relation between abstract Case and concrete case is not usually one-to-one; instead, it normally has the form of a one-way implication, viz. (see Kayne (1984: 116–17) on this):

- (37) If a DP has morphological dative case, then the grammar has abstract Dative Case.

This is in fact the simplest statement of the relation between morphological case and abstract Case that one can postulate, assuming the existence of any kind of abstract Case at all. What this implies is that as long as there was morphological dative case there could be no ambiguity at all regarding these constructions. However, it says nothing about the situation once the case morphology has been lost. One might conclude that the string in question is weakly P-ambiguous, since it provides no unambiguous information regarding the parametric property of *v*. But the very fact that, thanks to the one-way implication in (37), there could be an abstract DAT, shows that the structure must be strongly P-ambiguous (since the presence of DAT implies one feature make-up for *v*, while the absence of DAT implies that active *v** has uninterpretable *φ*-features; see §2.3.2).

The structural ambiguity is partially represented in (38), bearing in mind that English was a V2 language at this time, and supposing that our example is a main clause:¹²

¹² See note 17 of Chapter 2 on the reason for assuming that these clauses are V2-clauses, i.e. CPs.

- (38) [_{CP} The man+DAT/NOM [_C was] [_{TP} [_T_[u_{TP}] (was)] [_{VP} v [_{VP} helped (the man+DAT/NOM)]]]]

Example (38) shows that the clause is a CP, and that *the man*, which we are taking to be ambiguously Nominative or Dative, occupies SpecCP. What is left unclear in (38) is the nature of SpecTP and the way in which T's uninterpretable φ -features are eliminated where *the man* is DAT. It is usually assumed that SpecTP must be filled, and whatever fills this position must be able to Agree with some feature in T, i.e. that T has an EPP feature (see §2.5). Where *the man* is NOM, we can assume that it moves through the SpecTP position on its way to SpecCP, thereby satisfying the EPP. And of course, *the man*'s NOM-feature Agrees with T's uninterpretable φ -features and so these uninterpretable features are able to be eliminated from the representation and the NOM-feature is valued. On the other hand, where *the man* is DAT, it has no feature which can Agree with T. (This is clear where we have a plural argument, as there is no agreement in number between a dative argument and the verb, i.e. between a DAT DP and T's φ -features; see Allen (1995: 70ff., 142ff.) for discussion.) So, if *the man* is DAT, it is unable to move through SpecTP on its way to SpecCP as it cannot Agree with any feature of T.¹³ Therefore the EPP must be satisfied in some other way in this situation. There are two options as to what can fill SpecTP. On the one hand, we can postulate an expletive null subject (i.e. *pro*), an element which can also bear NOM and thus Agree with T (see §1.2.1 on expletive, i.e. non-referential, *pro*). The other option is to assume 'massive movement', in the sense introduced in §2.5, of either vP or VP into SpecTP (see examples (83–8) in Chapter 2). In order for massive movement of this kind to satisfy the requirement that the element in SpecTP Agree with some feature of T, we have to assume that the moved category may contain the element which Agrees with T. In the present case, we may assume, following Baker, Johnson, and Roberts (1989), that the passive marker itself is a nominal element capable of bearing a Case feature which can Agree with T. (This idea is updated in the context of massive

¹³ Chomsky (2000: 128) suggests that the very similar dative subjects in Icelandic might Agree for a Person feature with T, but, as mentioned in note 17 of Chapter 2, there is no evidence that preposed datives were subjects in passives in OE. It thus seems correct to take the DAT DP to occupy SpecCP and to have not moved through SpecTP. So, at least in OE and ME, there is no Agree for person features in this construction.

movement by Richards and Biberauer (2005.) So in fact our example in (38) really manifests a strong P-ambiguity between (39a) and either (39b) or (39b') (this is a more elaborate version of the reanalysis given in (36) above and in (43) of Chapter 2):

- (39) a. [_{CP} The man+NOM [_C was] [_{TP} (the man+NOM) [_{T[_uo}] (was)] [_{VP} v [_{VP} helped (the man+NOM)]]]]]
 b. [_{CP} The man+DAT [_C was] [_{TP} *pro*+NOM [_{T[_uo}] (was)] [_{VP} v [_{VP} helped (the man+DAT)]]]]]
 b'. [_{CP} The man+DAT [_C was] [_{TP} [_{VP} v [_{VP} helped+NOM (the man+DAT)]]_{T[_uo}] (was)] (vP)]]

Example (39a) is quite unproblematic: *the man* Agrees for Nominative with T, and moves through SpecTP on its way to SpecCP, satisfying the requirement for an element in SpecTP, which Agrees with T. In (39b), *the man* bears the interpretable DAT feature which does not need to Agree with anything. It moves in one step to SpecCP. SpecTP is filled by expletive *pro*, which is NOM and so Agrees with T. In (39b'), *the man* behaves exactly as in (39b). However, vP raises to SpecTP and the passive marker on *helped*, which is contained in vP, Agrees with T.

The ambiguity in (39) is created by the loss of dative morphology. Given (37), where there is dative morphology, one of the options (39b) or (39b') is the only one. Once dative case is lost, (39a) becomes available. (39a) involves two movements of *the man+NOM*: first to SpecTP and then to SpecCP. (39b) involves just one movement of *the man+DAT* to SpecCP, and insertion of expletive *pro* in SpecTP. (39b) therefore appears to be less complex than (39a). We could attempt to introduce some further complexity cost associated with the postulation of expletive *pro*, but probably the best course of action is to assume that expletive *pro* is not relevant here and that instead the correct representation of the Dative option involves a structure with massive movement like (39b'). Since this structure involves copying of all the vP-internal material, along with all its formal features, this structure will be more complex than (39a). In that case, the definition of complexity in (22) gives the right results. As a consequence of this reanalysis and the associated parametric change, v's feature make-up was changed as described in §2.3.2 with the consequences outlined there. Without the loss of dative-case morphology, Inertia ensures that the structure remained as (39b') (not (39b) if we rule out expletive *pro*).

A final illustration of the role of morphological change causing the PLD to change comes from our discussion of complementation in Latin and Romance in §2.4. Among various other changes discussed there, we suggested that Latin infinitives were associated with a T-position which was able to Agree for Accusative with an argument in the subordinate clause. This gives rise to the accusative + infinitive construction in Latin (or, more precisely, one variant of it – see the discussion in §2.4 and below), as in (repeated from (56a) of Chapter 2):

- (40) Dico te venisse.
 I-say you-Acc come-perf-infin
 'I say that you have come.'

It was tentatively proposed that Latin non-finite T lost this capacity when the tense/aspect forms of the infinitive such as perfect *venisse* were lost. In line with (34) and (37) above, let us state this as a one-way implicational statement, as follows:

- (41) If T[-finite] has an Accusative feature, then it shows inflectional distinctions marking tense/aspect.

This means that after loss of forms such as *venisse*, the unmarked (formerly present) form of the infinitive was no longer associated with an Accusative-bearing T. Nevertheless, examples like (42), with an Accusative subject of the complement clause and the unmarked form of the infinitive, would have been possible:

- (42) Dico te venire.
 I-say you-Acc to-come

We noted in §2.4 that this construction was ambiguous in Classical Latin in that the Accusative feature of the subject of the infinitive *te* could originate either in the infinitival T, or in the main-clause v^* . (The evidence for this comes from the two attested passive constructions in (53a) and (54) of Chapter 2.) After the loss of the tense/aspect forms of the infinitives, examples like (42) became less ambiguous than previously, in that they became unambiguously English-style accusative + infinitives, with the Accusative subject agreeing with v^* in the superordinate clause. This possibility was, however, then ruled out by a change in the value of Parameter G, itself connected to the development of Romance-style prepositional complementizers, and so Accusative subjects of infinitives were eliminated in general. So the loss of morphology played a role in eliminating

the accusative + infinitive construction in the development from Latin to Romance, although in this instance by eliminating an ambiguity rather than creating one.

3.3.4. *Conclusion*

Here we have tried to apply the solution to the logical problem of language change proposed in the previous section to actual cases of change discussed in the earlier chapters. If contact, cues, or morphology cause changes in PLD, then P-ambiguity and consequent opacity of one representation result, leading to abductive reanalysis and associated parametric change. We also saw that there is often an implicational relation between a morphological trigger and a parameter value. We will return to this last point briefly in the next section.

The main question that has been begged throughout the discussion is that of the definition of opacity. This, along with the closely-related notion of markedness, is the subject of the next section.

3.4. **Markedness and complexity**

The purpose of this section is to connect complexity as defined in (22) above to markedness, and thereby arrive at a basis for defining the marked and unmarked values of parameters with a view to formulating parameters along the lines of Dresher's proposal illustrated in (31) above. The concept of complexity is closely related to that of markedness. Here I will discuss an approach to determining in general the marked and unmarked values of parameters which correlates marked parameter values to the relative opacity or complexity of representations or derivations. The idea is that marked settings are associated with opaque, that is relatively complex, constructions.

3.4.1. *The concept of markedness*

The concept of markedness originated in Prague School phonology, apparently with Trubetzkoy, and was taken up by Jakobson (1941). (The history

of the concept is described in detail in Battistella (1996: 19ff.)) The basic idea can be stated as follows: given a binary opposition, the two terms of the opposition may stand in a symmetric relation or in an asymmetric one. In the former case, we say that the terms are equipollent; in the latter, we refer to one term as the marked one and the other as the unmarked one. The asymmetry lies in how the absence of specification of the terms is interpreted: the absence of the marked term implies the unmarked term, but the *absence of the unmarked term does not on its own imply the marked term*. In other words, all other things being equal we assume the presence of the unmarked term; the presence of the marked term requires something special, however, i.e. some kind of 'mark'. This asymmetric formulation can be maintained independently of the nature of the terms involved, the nature of the asymmetry, or the correlates of the asymmetry in some other domain. It is often supposed that the marked term is associated with relatively greater complexity; this is arguably inherent in the idea of it requiring an extra 'mark'. For example, Cinque (1999) proposes a series of markedness conventions for the features associated with various functional heads in his analysis of clause structure. He states that marked features are 'more restricted [in] application, less frequent, conceptually more complex, expressed by overt morphology' (128), while unmarked features are in each case the opposite.

We can illustrate the essential asymmetry that characterizes markedness relations with phonological distinctive features. A phonological opposition, for example that of voicing, can be thought of as an equipollent opposition between [+Voice] and [-Voice] or an asymmetric opposition between [mVoice] and [uVoice]. (Here and below, 'u' before the name of a feature means 'unmarked', not 'uninterpretable' as in the specifications of formal syntactic features in earlier sections; 'm' means 'marked'.) Chomsky and Halle (1968, Chapter 9) discuss markedness in relation to the phonological distinctive-feature system they proposed; the approach to markedness and complexity adopted here is largely inspired by their discussion. Where the opposition is equipollent, if a segment is not [+Voice] then it is [-Voice] and vice-versa. But where the opposition involves a markedness asymmetry, markedness conventions and perhaps other statements are required to determine the value of the coefficient of a feature (Chomsky and Halle 1968: 403ff.). Moreover, there need not be a straightforward relation between the *u/m* values and the *+/-* values; for example, Chomsky and Halle (1968: 406) proposed that [uVoice] is [-Voice] if the segment is

[–sonorant], but [+Voice] if it is [+sonorant]. Marking conventions of the type first put forward by Chomsky and Halle imply that the underspecified feature [Voice] can ‘default’ to a given value under various circumstances, something impossible in the case of the equipollent +/– opposition.¹⁴ In turn, this leads to the possibility of an ‘elsewhere convention’, a notion going back to the Sanskrit grammarians of Indian antiquity (see Kiparsky (1973)): in the absence of specification, the unmarked feature value is assumed, while a marked value requires positive specification, and therefore a longer description. For Chomsky and Halle, the markedness asymmetry relates to the evaluation metric they propose for determining the relative simplicity of rule systems: ‘the unmarked value of a feature was cost-free with respect to the evaluation metric, while the marked values were counted by the metric’ (Battistella 1996: 75). The correlates of the asymmetry were stated by the marking conventions (Chomsky and Halle (1968: 404–7) propose thirty-nine of these), which are intended to capture aspects of the intrinsic content of distinctive features. The correlates of markedness in the distinctive-feature system include: cross-linguistic frequency of unmarked terms (all languages have voiceless obstruents, but not all have voiced ones: note how in implicational universals, the marked value of an opposition entails the unmarked one (see Croft (2003) for discussion)); unmarked terms appear earlier than marked ones in language acquisition and are lost later in language deficits (this was first proposed by Jakobson (1941)); and the fact that unmarked values emerge under neutralization in certain positions, for example, the coda of a syllable or the end of a word (for example, final-obstruent devoicing is cross-linguistically very common, while obstruent voicing in this context is relatively rare). Kenstowicz (1991: 61–4) discusses these points in more detail in relation to phonology.

3.4.2. *Markedness and parameters*

Since we take parameters to have binary values (see §1.1 for general discussion, and note that all the examples of parameters we have discussed have been formulated in a strictly binary fashion), we can in principle apply

¹⁴ The idea that there is no single unmarked value for a feature, but that this may depend on other features, represents an important difference between Chomsky and Halle’s approach and the Prague School approach to markedness.

markedness logic to the opposition between these values; in other words, we can treat the binary opposition between the two values of a parameter as an asymmetric one in the sense described above. This has in fact been suggested in various places ever since the earliest formulations of principles-and-parameters theory (see Chomsky (1981), and the discussion in Battistella (1996: 82ff.)).¹⁵ If we do this, we have to answer three questions: (i) what is the nature of the features involved in the asymmetric relation? (ii) what is the nature of the asymmetry? (iii) what are the correlates of the asymmetry in other domains?

Regarding question (i), it suffices for now to simply treat the features in question as the values of a parameter; giving a fuller answer requires a proper statement of the form of parameters, something we have yet to do. In the next section, I will attempt a general characterization of parameters which will provide a more substantive answer to this question, and thereby facilitate the statement of parametric marking conventions similar to those introduced into phonological theory in Chomsky and Halle (1968, Chapter 9).

We could answer question (ii) in terms of the definition of complexity given in (22), which we repeat for convenience:

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

The nature of the asymmetry between the parameter values lies in the complexity of the structures generated by the grammars determined by the different values. The unmarked value of a parameter determines a grammar which generates simpler structures than those generated by the marked value. In the next section, we will suggest that parameter interactions give rise to a slightly more complex and interesting situation than this.

¹⁵ Chomsky's discussions of markedness here and in *Knowledge of Language* (1986), make use of the distinction between the 'core grammar' and the 'periphery' in various ways. I am not assuming this distinction, as seems to be more in line with Chomsky's assumptions in his more recent work on the Minimalist Program, where this distinction should no longer play a role. (This is what I take to be the implication of Chomsky's remark that it 'should be regarded as an expository device, reflecting a level of understanding that should be superseded as clarification of the nature of linguistic inquiry advances' (Chomsky 1995: 163, note 3)). As the text discussion will make clear, I follow Chomsky's thinking in taking markedness to impose a preference structure on the parameters of (core) grammar for the language acquirer.

Question (iii) brings us back to the issue of most concern here. At least some of the correlates of the markedness asymmetry between parameter values lie in syntactic change: since abductive reanalysis and parametric change arise through P-ambiguity and opacity/complexity of the trigger, with the less complex structure being preferred, then – all other things being equal – parametric change will be in the direction of unmarked values.¹⁶ We expect to find correlates in language acquisition (for example, marked values being harder to acquire and hence acquired later) and typology (marked values being less cross-linguistically frequent). I will not comment further on language acquisition, given the conclusion of §3.1 that it is hard to observe the postulated connection between acquisition and change. I will come back to the relationship between markedness and typology in the next section.

The approach to complexity in (22) comes remarkably close to the notion of ‘value’ put forward by Chomsky and Halle (1968: 334): ‘The “value” of a sequence of rules is the reciprocal of the number of symbols in its minimal representation’. Taking the relevant symbols to be formal features, which are the important symbols in the syntactic representation in the theory of syntax being assumed here, the definition in (22) would make it possible to value syntactic derivations just as Chomsky and Halle propose valuing phonological derivations. We did essentially this in our discussion of the role of complexity/opacity in abductive reanalysis in the previous section.

A further point arises from this. Roberts and Roussou (2003: 210–13) derive a series of markedness hierarchies from the definition of complexity in (22). Here I give a simplified version of their hierarchy:

(43) Move > Agree > neither

Here ‘>’ means ‘is more marked than’. So a category set to a parameter value which requires movement is more marked than one which merely causes an Agree relation, which is in turn more complex than one which has neither property. This follows straightforwardly from the feature-counting idea: in order to give rise to movement, a category must have both

¹⁶ Of course, we do not want only this kind of change to be possible: change from unmarked to marked must be allowed somehow. This point will be dealt with in the next section.

uninterpretable formal features and the movement-triggering (EPP) feature. In order to give rise to an Agree relation, a category need only have an uninterpretable formal feature. Finally, in order to trigger neither operation, a category should lack both EPP and uninterpretable formal features.¹⁷ This approach is developed in some detail in Roberts and Roussou (2003, Chapter 5), and we will consider some of its implications below.

3.4.3. *The Subset Principle*

Other approaches to determining the marked and unmarked values of parameters have been put forward. One important and influential proposal was the Subset Principle of Berwick (1985). This states that 'the learner selects the grammar that generates the smallest possible language that is compatible with the data' (Manzini and Wexler 1987: 425). The interest of the Subset Principle is that it relates to an important and fairly well-established aspect of language acquisition: the fact that language acquirers do not have access to negative evidence. In other words, language acquirers are not presented with ungrammatical sentences which are marked as such. As Guasti (2002: 4) puts it, 'negative evidence is not provided to all children on all occasions, is generally noisy, and is not sufficient ... Children have the best chance to succeed in acquiring language by relying on positive evidence [emphasis omitted – IGR], the utterances they hear around them – a

¹⁷ One might wonder whether a category with three uninterpretable features is more marked than one with one uninterpretable feature and an EPP feature. This is predicted by (22), but is not consistent with the hierarchy in (43). (43) should be understood as holding in relation to a given feature: in that case, Move-F will always be more complex, and therefore more marked, than Agree-F, for any F, as Move requires the EPP feature in addition to F. Chomsky (2005c) introduces the possibility of Move occurring independently of Agree, possibly triggered by a further kind of feature known as an Edge Feature (EF). EF-triggered movement is characteristic of wh-movement, topicalization and focalization, movements which typically target the 'left periphery' of the clause. It is possible, and would follow from (22), that this type of movement is less marked than that triggered by EPP where Agree is involved. See the discussion of how grammars may innovate marked properties in §3.5 below.

resource that is abundantly available'. (We touched on the absence of negative evidence in our discussion of the poverty of the stimulus in Chapter 1).

Because they only have access to positive evidence, acquirers are, other things being equal, at risk of falling into a 'superset trap.' This can happen if acquirers posit a grammar which generates a language which is a superset of the language generated by the actual grammar in their environment, in the sense that it contains no examples that are incompatible with the PLD to which the children are exposed, but it generates examples that are incompatible with the target grammar. (This situation would correspond to the schema for abductive change in (13) of §3.2 above, with G_1 a subset of G_2 .) If children only have access to positive evidence, they will never hear any example which causes them to 'retreat' from the superset grammar. Thus they may posit a grammar which is incompatible with the target, and recall that it is a standard assumption in work on language acquisition that this does not happen; this is what underlies the empirical force of the Inertia Principle – see the discussion in §3.2 above.

In order to rule out the risk of superset traps, the Subset Principle is proposed as a condition on language acquisition. The Subset Principle, as just given in the quotation from Manzini and Wexler (1987), forces children to hypothesize the grammar which generates the smallest language compatible with the trigger experience. In this way, it is argued, they do not run the risk of falling into superset traps.

The notion of markedness which derives from this then is that marked parameter values will generate bigger languages. The null-subject parameter may serve as a (slightly artificial) example. As we saw in §1.1.1, null-subject languages allow a definite, referential pronoun subject to be dropped in finite clauses, while non-null-subject languages do not:

- (44) a. Parla italiano.
b. *Speaks Italian.

On the other hand, null-subject languages typically allow the pronominal subject to be expressed, just like non-null-subject languages:

- (45) a. Lui parla italiano.
b. He speaks Italian.

(As we saw in connection with examples (14) and (15) in Chapter 1, there are interpretative differences between null-subject and non-null-subject

languages where subject pronouns are expressed in the former; I will gloss over these for the purposes of illustration of the Subset Principle, however). Thus the grammar of Italian generates a larger set of strings than that of English. In other words, non-null-subject languages are a subset of null-subject languages. This can be illustrated as follows:

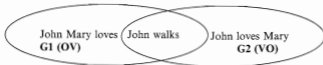
(46)



The Subset Principle therefore implies that the positive setting of the null-subject parameter is more marked than the negative setting.

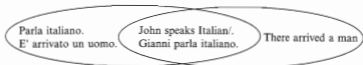
One empirical problem that one could raise here is the evidence of early null subjects in L1 acquisition of non-null-subject languages. However, as we saw in §3.1, this phenomenon probably is not related to a 'missetting' of the null-subject parameter at an early stage of language acquisition, and so the objection does not hold. A much more serious problem with the above line of reasoning emerges if we consider the various parameters we have put forward in our discussion: verb-movement parameters (both V-to-T and V2), the negative-concord parameter, the wh-parameter, and word-order parameters all define intersecting grammars. That is, in each case, one setting of the parameter allows one type of structure S and disallows another type S', while the other setting allows S' and disallows S. This is clearest in the case of word-order parameters: one setting of parameter F1, for example, allows VO and at the same time disallows OV, while the other setting has just the opposite effect. The intersection relation can be illustrated as follows:

(47)



The material in the intersection is weakly P-ambiguous in relation to the parameter P in question in terms of the definition in (21c), while the material in the complement of the intersection expresses the value of P.) The same exercise could be repeated for the verb-movement parameters, the negative-concord parameter, and the wh-movement parameter. In fact, it can also be repeated for the null-subject parameter, to the extent that it is true that null-subject languages do not have overt expletives:

(48)



As Battistella (1996: 113) points out: '[i]f markedness relations obtain between parameters that are not in a subset relation, they must be accounted for in some other way'. The above considerations seem to indicate that the Subset Principle is not a useful way of predicting markedness relations in general among parameters, since most – if not all – parameters define intersection relations of the kind seen in (47) and (48).

A further issue arises if we look at the Subset Principle in the diachronic domain. If non-null-subject languages are subsets of null-subject languages, then we expect a diachronic preference for change from the positive to the negative value of this parameter. We know that the null-subject parameter must have changed from positive to negative at some point in the history of Germanic (see the remarks on this in §3.1 above), and this change has certainly happened in the history of French and certain Northern Italian dialects, as we saw in §1.1.2. So this much is consistent with what the Subset Principle predicts. However, we also saw there that this parameter may have changed its value in the opposite direction in exactly these Romance varieties, with some question as to what may be the best analysis of contemporary French.

So we conclude, rather reluctantly, that the Subset Principle is not useful in providing the basis for determining the markedness of parameter values in cases like the above. The reluctance is due to the fact that the Subset Principle has the great conceptual merit of being firmly grounded in an important fact about acquisition: that children do not have access to negative evidence.

One area where the Subset Principle may be useful is in distinguishing between a grammar which allows genuine formal optionality and one which does not. Abstractly, a case of this type would be where G_1 allows an alternation between two constructions C_1 and C_2 while G_2 , thanks to a

different parameter-setting, does not. An example might be the difference between English and (Standard) French regarding Preposition-stranding or pied-piping. (This was discussed in Chapter 1, Box 1.5 and in the previous section.) English has the option of Preposition-stranding or pied-piping, while French only allows the latter. The French situation is illustrated by (25) above, which we repeat here:

- (49) a. Who did you speak to ___?
 b. To whom did you speak ___?
 (25) a. *Qui as-tu parlé à ___?
 b. A qui as-tu parlé ___?

The fact that English allows both options, while French only allows one of them means that the French parameter-settings generate a language which is a subset of the English one. We could, therefore, regard English as marked in relation to French in this respect. Of course, (49b) is characteristic of a relatively 'high' register as compared to the more colloquial (49a), but for the purpose of this illustration of the logic of the Subset Principle I abstract away from this; we will come back to the concept of the 'social value' of variants in §4.2.

The Subset Principle might also lie behind the phenomenon of 'restriction of function', whereby in one system a given operation applies more freely than in another. The more restricted grammar then produces a subset of the grammatical strings of the more liberal one. An example of this might be the restriction on OV orders to negative and quantified objects in fifteenth-century English which we mentioned briefly in §2.5.3. Given that OV order was an option with non-negative, non-quantified DPs in the earlier stage (i.e. ME from roughly 1200 to 1400), we have a situation where the fifteenth-century grammar only allowed OV for a particular class of objects and required VO elsewhere, while the earlier grammar allowed OV with any kind of object. Thus object-movement, assuming that is the correct analysis of this construction (see §2.5.4), was restricted in function. Here there may be a tension between concepts of markedness based on the Subset Principle and those based on feature-counting, since the more restricted variant requires more features.

We see then that the Subset Principle has a major conceptual advantage, being based on what appears to be an important fact about language acquisition: namely, that language acquirers do not make use of negative evidence. Its actual application to parametric systems may be somewhat restricted, since so many parameters appear to define languages in

intersection, rather than inclusion, relations. However, it may play a role in relation to true formal optionality, in predicting that such systems would be marked, and it may play a role in accounting for the diachronic phenomenon of 'restriction of function.'

3.4.4. *Markedness and core grammar*

Another proposal, which was not intended to form the basis of a general account of the markedness of parameter values, was made by Hyams (1986: 156ff.). She took the view that markedness was a feature only of the 'periphery' of the grammar (in the sense briefly discussed in note 15 above). The null-subject parameter, however, is a property of core grammar, and so the question of the markedness of its settings does not arise. Nevertheless, on the basis of her observation of early null subjects in the production of children acquiring non-null-subject languages, she argues that the null-subject value is the more accessible value (Hyams 1986: 162-3) since null subjects do not require the costly process of lexicalization of pronouns (163). Hence children acquiring English begin with the assumption that it is a null-subject language, and the parameter is reset during the course of language acquisition to the negative value (on the basis of the evidence from modals and overt expletive subjects, as mentioned in §4.1). One could imagine that this would favour a tendency in language change in the direction of null subjects, but the general view now held amongst researchers on L1 acquisition is that early null subjects do not reflect a 'missetting' of the null-subject parameter, but rather some property of immature competence. For this reason we leave this proposal aside.

3.4.5. *Markedness and inflectional morphology*

At this point it is justifiable to ask what the advantages of an analysis of parameter values into marked and unmarked values might be. Aside from connecting syntactic change to the form of parameters, as we have done, one independent point has to do with the nature of language acquisition. Lasnik (1983) observes that there is an intrinsic connection between markedness in L1 acquisition and the question of indirect negative evidence. The notion of indirect negative evidence is discussed by Chomsky (1981: 8-9): although, as

we have suggested and as is the standard view amongst L1-acquisition researchers, children do not have direct negative evidence in the sense of not having access to the information that a given structure or string is ungrammatical, Chomsky suggests that indirect negative evidence may nevertheless be available in the case where some feature is expected by acquirers but is not actually found in the PLD. In such a situation, the lack of the 'expected' feature may be a kind of evidence: indirect negative evidence.

A sufficiently clear and robust characterization of the markedness of parameter-settings may provide a form of indirect negative evidence. If the marked value of a parameter associated with a given feature is that associated with movement, then if there is no evidence for movement in the PLD, the acquirer has indirect evidence that the marked value of the parameter in question does not hold. In other words, evidence for marked features requires direct positive evidence, and indirect negative evidence that the positive setting does not hold arises simply when the direct positive evidence is not available. That this is the case follows from our basic characterization of the asymmetric nature of markedness relations: the presence of the marked feature must be in some way signalled. So if there is no evidence for a marked parameter value, it is not assumed. This in itself is a form of indirect negative evidence.

The feature-counting notion of markedness of parameter values that was introduced in §3.4.2 above is a purely formal one. As such, it differs from other approaches which have tried to relate markedness to **substantive universals**, either directly or indirectly. Chomsky and Halle's (1968) marking conventions relate the purely formal, feature-counting evaluation metric they propose to substantive phonetic and phonological universals. We also mentioned that Cinque (1999: 128) proposes a series of markedness conventions for the features associated with various functional heads in his analysis of clause structure. His postulations of marked and unmarked values are based on familiar Jakobsonian criteria, as we saw. For example, Cinque assumes that the unmarked value of his postulated Mood_{Speech Act} category is 'declarative', while the marked value is '-declarative'; the unmarked value of Mod_{epistemic} is 'direct evidence' and the marked value is '-direct evidence', the idea being that in each case the unmarked value is inherently simpler than the marked one.

How do Cinque's proposals regarding markedness relate to the proposal made above regarding the relation between complexity and markedness? The two notions are quite distinct, in several important respects. The

fundamental difference between the two is that Cinque's proposals regarding markedness relate to the substantive content of features of functional heads, ultimately their notional semantic properties, while what was sketched in §3.4.2 is a purely formal, feature-counting notion associated with a complexity metric. We therefore might want to keep the two kinds of markedness distinct. We could call the complexity-based notion of markedness discussed above formal markedness and Cinque's notion substantive markedness. (This distinction is proposed in Roberts and Roussou (2003: 214), although on slightly different grounds.)

We saw earlier that Chomsky and Halle (1968, Chapter 9) link formal markedness (their feature-counting evaluation metric) to substantive markedness by means of markedness conventions. We might want to contemplate a similar move in the present context. One reason for this is that, as we saw above, Cinque proposes as one criterion of markedness a greater likelihood of morphological expression. This connects to the postulates introduced in the previous section regarding the morphological expression of certain parameter values.

Let us repeat those statements here:

- (34) If (finite) V is marked with person agreement in all simple tenses, this expresses a positive value for the V-to-T parameter.
- (37) If a DP has morphological dative case, then the grammar has abstract Dative Case.
- (41) If T[-finite] has an Accusative feature, then it shows inflectional distinctions marking tense/aspect.

We can note that, directly in the case of (34) and indirectly (by means of the marked way of realizing SpecTP in the absence of a Nominative DP in the case of (37)), the realization of a morphological feature implies the marked value of the parameter ((41) seems to go the other way, though). (34) and (37) suggest that the following general template might hold for the relationship between morphological expression of a parameter and the markedness of that parameter:¹⁸

¹⁸ (50) is deliberately vague in formulation. The expression 'C is associated with a marked parameter value' is formulated so as to allow for (37), where the connection between morphological dative case and markedness, in terms of the complexity of the structure in (39b') above where the relevant DP is Dative, is somewhat indirect (although it is in fact the consequence of the fact that Dative Case is interpretable and therefore unable to check a feature of T – see the discussion in §2.3.2). If it is anywhere near correct, (50) no doubt requires a great deal of refinement.

- (50) If a formal feature of a category C is inflectionally expressed, then C is associated with a marked parameter value.

Although rather vague as it stands, something like (50) could serve as a marking convention linking overt inflectional morphology with the marked values of syntactic parameters, as well as providing a clear general statement of the kind of morphological triggers (or cues, in Lightfoot's (1999) terminology) that are relevant in acquisition and change. It also predicts that the loss of inflectional morphology, at least for certain types of inflection, may perturb the PLD in such a way as to lead to abductive change along the lines we saw in the previous section. Taking (41) into consideration suggests that the implication might go either way, but we nevertheless observe a relation between morphology and the changing and setting of parameters. In the next section we will propose a further marking convention related to cross-categorical harmony in word-order patterns and word-order change.

3.4.6. *Markedness, directionality, and uniformitarianism*

One final very general point regarding markedness concerns the concept of uniformitarianism. We briefly mentioned this concept in §2.4 in our discussion of diachronic aspects of subordination. This idea is formulated by Croft (2003: 233) as follows: '[t]he languages of the past ... are not different in nature from those of the present'. In terms of the principles-and-parameters approach to syntax, we can take this to mean that all languages at all times (in the history of our species) reflect the same basic UG and therefore the same set of parametric options, and that those parametric options have the same markedness properties.

Stated as above, the uniformitarian hypothesis seems very plausible. In fact, one can argue that it is a precondition for applying the principles-and-parameters approach to diachronic questions (see Roberts (2001: 89)).¹⁹

¹⁹ Or indeed any kind of historical linguistics. Interestingly, for most of the history of linguistic thought in the West, uniformitarianism was not assumed, in that it was thought that the three languages of the Holy Scriptures, Latin, Greek, and Hebrew, were not subject to change or decay. See the discussion of Dante's *De vulgari eloquentia* in Law (2003: 190, 230). Clearly, the assumption that Latin and Greek could change was necessary for comparative Indo-European philology to be possible, although the Renaissance recognition of the changeability of Latin did not give rise to the postulation of the Indo-European family (see Law (2003: 260ff.) and Simone (1998: 215) on this).

However, one question we can raise has to do with the transition from unmarked to marked parameter values, an issue we have postponed until the next section (see note 16). We clearly want to allow for the transition to marked parameter values, although we have not yet seen how this may be possible in terms of the general approach outlined above. If we do not allow for the innovation of marked values, then two highly problematic issues arise. First, we predict that all languages are tending towards a steady state, from which they will not be able to escape, where all parameters are fixed to unmarked values. Second, we have to explain where the marked parameter values currently observable in the world's languages came from. So it is highly desirable to have a mechanism for the innovation of marked parameter values.

The question of uniformitarianism arises here, in that if every language were in the maximally steady state we would have a violation of a strong version of this thesis. However, at least a weaker interpretation, of the kind just given in terms of principles and parameters, would allow for the idea that change from marked to unmarked is more regular and frequent than change from unmarked to marked. This would entail that the set of languages in the world would gradually change towards ever less marked systems. On this view, UG and the available parameters do not change, and so uniformitarianism is not violated, but at the same time the range of different sets of options actually instantiated in the world's languages steadily diminishes. In other words, if we think of the set of parameters as defining an abstract space (perhaps a 'state-space' in the terminology of dynamical systems – see §4.3.3) within which grammars can exist, a general move towards more and more unmarked systems implies that ever smaller pockets of the available space are occupied by actually existing systems. Something like this is certainly possible in principle; whether it is actually happening is an empirical question, albeit a rather difficult one to answer with any certainty. At first sight, there appears to be some evidence for something like this from typological studies: Nichols (1992: 250–1), for example, observes that the overall level of structural diversity in (some aspects of) grammatical systems is lower in the Old World than in the New World and the Pacific. She points out that '[t]he high diversity there [in the New World and the Pacific – IGR] can be regarded as a peripheral conservatism in dialect-geographical terms; these areas, secondarily settled, are far enough from the Old World centers of early spread to have escaped the developments that have lowered genetic density and structural diversity in the Old

World' (250). However, determining whether there is a global tendency for reduction in diversity requires knowledge of change at very great time depths, greater than the maximum of 8,000–10,000 years which the traditional method of comparative reconstruction seems to allow. Nichols (1992) addresses this very question, and in fact concludes that 'today's linguistic universals are the linguistic universals of the early prehistory of language' (Nichols 1992: 278). This conclusion strongly favours the uniformitarian view, and the concomitant view that the world's languages are no less evenly spread among the options made available by UG than they were in prehistory. As Nichols states '[t]he only thing that has demonstrably changed is the geographical distribution of diversity' (277). We thus clearly need a mechanism for introducing marked parameter values, as there seems to be an overall equilibrium over time in the grammatical systems attested, as far as we can tell; I will return to this point in the next section.

3.4.7. Conclusion

In this section we introduced the concept of markedness and applied it to parameter values, in terms of the definition of complexity given in (22). We also looked at other approaches to the markedness of parameter values, notably the Subset Principle. Further, we briefly considered the relationship between markedness and indirect negative evidence, as defined by Chomsky (1981), and Cinque's (1999) proposals for substantive markedness values associated with functional heads. We considered the relationship between inflectional morphology and syntactic markedness, tentatively suggesting the correlation in (50). Last, the possibility that the world's languages are tending towards ever more unmarked systems was considered and rejected, following Nichols' (1992) conclusions.

In the next section I will try to conclude the general discussion of parameter-setting which has been the theme of this chapter by making a proposal for the form of parameters and considering some of its consequences.

3.5. Parameter setting and change

In this section I attempt to synthesize the discussion in the preceding sections, by proposing a general format for parameters and suggesting an

account of how they are set in language acquisition and change. The goal of this exercise is to give a clear view of the issues involved, and to bring together the strands of the discussion in the rest of this chapter, rather than to make new theoretical proposals. The conclusions we reach here will also form the basis of the discussion in much of the remaining two chapters. Accordingly, we first present a general statement of the format for parameters, basing ourselves fairly closely on Dresher's (1999) formulation, as illustrated in (31) above. We then flesh out further the discussion of markedness from the previous section, presenting a further markedness convention (in addition to that presented in (50)), and showing how the concept of **markedness reversal** may play a role in certain types of syntactic change, primarily word-order change. This leads naturally to a discussion of networks of parameters; here we summarize the very interesting proposals in Baker (2001). The final question we look at, although rather briefly since it will be taken up in more detail in §4.3.4, has to do with 'cascades' of parametric change: the extent to which one parameter change may lead to another and how, once again, markedness considerations may play a role.

3.5.1. *A format for parameters*

The first issue concerns a general statement of the exact form of parameters. This is something that we have not broached until now, having contented ourselves with rather informal statements when we introduced the various parameters we have been considering in Chapters 1 and 2. Let us first recapitulate those statements:

- (51) A. Does every finite clause require an overt subject?
 YES: non-null-subject languages (French, English ...).
 NO: null-subject languages (Italian, Spanish, Greek, Japanese ...)
- B. Does V move to T in finite clauses?
 YES: French, Welsh, Italian, Icelandic ...
 NO: English, Swedish, Danish ...
- C. Does the finite verb move to C in finite main clauses?
 YES: German, Dutch, Swedish, Icelandic, Danish, Kashmiri, Romansch ...
 NO: English, French, Italian, Welsh ...
- D. Are (non-inverse) Negative Agree relations found?
 YES: French, Italian, Welsh ...
 NO: English

- E. Does a *wh*-phrase move to the Specifier of an interrogative CP?
 YES: English, Italian, Spanish, German, Welsh ...
 NO: Chinese, Japanese, Thai, Korean, Turkish, Armenian ...
- F6. For all heads H, does the structural complement of a head H precede or follow H in overt order?
 PRECEDE: Malayalam, Turkish, Japanese, Basque ...
 FOLLOW: Romance, Celtic ...
- G. Does L allow accusative subjects in SpecTP of a non-finite clause?
 YES: English, Latin, Classical Greek, Irish ...
 NO: French, Italian ...

Each of these parameters is formulated as a yes/no question, or, in the case of F6, as a disjunctive question (precede vs. follow). As we pointed out in §1.1, one could imagine that the two-year-old mind/brain has some means of interrogating the PLD along these lines. It is now time to try to flesh this rather crude notion out in a more precise fashion.

We can immediately observe that (51B, C, E) have to do with triggering movement, and can thus conclude that the variation is due to the presence or absence of the movement-triggering feature on the head in question (finite T in (51B), finite C in (51C), and interrogative C in (51E)). Furthermore, if word-order variation is to be accounted in terms of movement relations, as suggested in §2.5.4, we may be able to see parameter F6 (or the group of parameters which determine head-complement order for a range of heads) as a case of the presence or absence of a movement-triggering feature. (51A) and (51G) are slightly different in that they concern the type of subject which can appear in SpecTP: whether there can be a null subject in the Specifier of a finite T or an overt Accusative subject in the Specifier of a non-finite T. Assuming that the possibility of a null subject of finite T is connected to 'rich' agreement, a property we can associate with T, then both of these parameters have to do with the nature of T's features and therefore what kinds of elements T may Agree with. Finally, (51D) concerns the possibility of a particular feature entering a particular type of Agree relation.

So we can draw two conclusions. First, the parameters all concern formal operations of the syntactic system: Agree and Move. They do not seem to relate directly to morphological, phonological, or semantic properties of language. Second, we can see that all of the parameters relate to the features associated with heads; in fact they all involve the features of functional heads, except perhaps for some cases of F6. Again, the features

in question are all formal features, i.e. they are features which play a role in determining the application of formal operations such as Move and Agree. So, all the parameters we have looked at involve the formal-feature specification of heads, principally functional heads; this corresponds exactly to what is proposed in Chomsky (1995: 6).

These observations make possible a general statement of the form of parameters. Following Dresher's (1999) approach, as illustrated in (31) above, we present each parameter along with a statement of its default value and its cue. We will continue to present parameters as binary in nature. On the basis of the discussion of markedness and complexity in the preceding section, we assume that the default value must involve a smaller number of features than the marked value (see (22)). Also on the basis of the discussion in the preceding section, we take it that the cue may be a morphological property. The other obvious cue is word order itself.

The general format for parameters will thus look like this:

- (52) a. *Parameter*: A (functional) head H (has/does not have) feature F (in a given formal relation).²⁰
 b. *Default*: F is absent.
 c. *Cue/expression*: properties of inflectional morphology and linear order of elements.²¹

²⁰ Given the nature of the syntactic operations postulated in recent minimalism, we really only have four options for (52a), and these are implicationally related, as follows:

- i. does H have a feature triggering Agree?
- ii. if so, does H have an EPP feature?

If we distinguish head-movement and XP-movement, then we have two further options:

- iii. if (ii), does H require pied-piping of the Goal?
- iv. if so, how large a category is pied-piped?

This last option was implicit in our discussion of 'massive movement' in §2.5. For more technical and empirical details, see Richards and Biberauer (2005); Biberauer and Richards (2006); Biberauer and Roberts (2005a). I will not pursue the options in (i-iv) systematically here, although the statement of the parameters in (54) is not incompatible with them. We will encounter pied-piping again in §4.1.4, when we look more closely at the nature of formal optionality.

²¹ See §3.2 above for a discussion of the similarities and differences between Clark and Roberts' (1993) notion of P-expression and the Lightfoot/Dresher notion of cue.

Introducing parameters in §1.1, I pointed out that they have four important properties. These can be summarized as follows:

- (53) a. Parameter values must be able to be set on the basis of rather salient elements of the PLD.
 b. Parameter values must be set: not deciding is not an option.
 c. Parameters may be determined by 'gaps' in UG principles.
 d. Parameters are binary.

How does the schema in (52) capture the properties of parameters as listed in (53)? Let us consider (53a–d) one by one. (53a) clearly relates to the cuing or expression of parameters. The linear order of constituents and inflectional morphology are both salient features of the PLD, and are both things that acquirers appear to be sensitive to, given that they are able to set word-order parameters very early (as we saw in §3.1) and they acquire the morphological properties of verbs, including agreement and finiteness marking, equally early. (This is shown in detail by Guasti (2002: 120ff.)) So the schema in (52) can clearly capture this property of parameters.

Example (53b) relates partly to the default clause in (52), in that we can assume that in the absence of a clear expression of the value of a given parameter (i.e. if all the relevant PLD is weakly P-ambiguous in the sense defined in (21c)), the default option is always taken. A further point which comes up here concerns the relations among parameters. We saw in §1.5.1 that there is a further parameter distinguishing among languages with the positive value for parameter E (i.e. those with overt wh-movement) determining whether just one wh-phrase is moved to an interrogative C or whether all available wh-phrases must be moved. Naturally, this further parameter is not relevant in systems where parameter E has the negative value. We take it that this parameter must take on the default value in this kind of case. What is at issue here is the question of the implicational relations amongst features, a point I will return to below.

Example (53c) can be reconciled with (52) if we make the obvious inference from (52) that, to some extent, the feature make-up of functional (and perhaps some other) heads is underspecified by UG. It may be that UG only requires a very minimal feature specification for functional heads: just enough to distinguish what Chomsky (2000; 2001) refers to as the 'core functional categories': C, T and v. Further specification may be entirely a matter of parametric variation. (Giorgi and Pianesi (1997, §1.4) make a proposal similar to this.) Indeed, to the extent that the formal features of functional categories are primarily relevant for the internal workings of

syntax, and that, in the context of the Minimalist Program, these internal workings of syntax are as elementary as possible, it seems very reasonable to think that UG imposes no particular further requirements on the feature make-up of functional categories. So we see that (53c) can be captured by the format in (52). This gives us a way to understand why parameters exist at all, which we hinted at in Chapter 1: they simply force a consistent choice where UG leaves things open, i.e. an individual system cannot have gaps and does not tolerate randomness. Every underspecified point must be 'filled in' in a consistent way. (The consistency might in fact be created by the learning device; I will briefly take this point up again in §5.1.) Finally, (53d) is built in to the statement in (52a).

Since (52) seems to capture the important properties of parameters as listed in (53), we will take it to be a general format for the statement of parameters. More specifically, the parameters in (51) can now be reformulated as follows:

- (54) A. *Null subjects*
- a. *Parameter*: Finite T {has/does not have} sufficient specification of agreement features φ to bear the subject thematic role/Agree with *pro* in SpecTP.²²
 - b. *Default*: φ is absent.
 - c. *Cue/expression*: 'rich' agreement morphology on T- and/or V-elements.
- B. *V-to-T movement*
- a. *Parameter*: Finite T {has/does not have} an EPP feature which attracts V.
 - b. *Default*: EPP is absent.
 - c. *Cue/expression*: (finite) V is marked with person agreement in all simple tenses.
- C. *Verb second*
- a. *Parameter*: Finite, root C {has/does not have} an EPP feature which attracts T.
 - b. *Default*: EPP is absent.
 - c. *Cue/expression*: consistent XP V order in the left periphery of CP (see Lightfoot (1999: 153)).
- D. *Negative concord*
- a. *Parameter*: non-inverse Agree_{Neg} relations.
 - b. *Default*: such relations are absent.
 - c. *Cue/expression*: clausal negation which either can or must be uninterpretable.

²² Recall that in our discussion of the null-subject parameter in §1.1, we did not decide between these two analyses of null subjects.

E. *Wh-movement*

- a. *Parameter*: [+wh] C {has/does not have} an EPP feature triggering movement of a wh-phrase to its Specifier.
- b. *Default*: EPP is absent.
- c. *Cue/expression*: 'displaced' wh-phrases, wh-marking on D.

F. *The head parameter(s)*

- a. *Parameter*: a head H {has/does not have} an EPP feature triggering movement of its complement to its specifier.
- b. *Default*: EPP is absent.
- c. *Cue/expression*: overt complement > head orders.

G. *Accusative + infinitive*

- a. *Parameter*: non-finite T {has/does not have} features which Agree with a DP in its specifier.
- b. *Default*: such features are absent.
- c. *Cue/expression*: overt Accusative subjects of infinitives.

So we have a general format for parameters, which seems to have the right kinds of properties, and we are able to reformulate the parameters which we have been interested in by using this format. At the very least, this is a useful exercise, but the combination of the requirement to state each parameter in terms of formal features and to state both the default value and the cue has clear implications for both language acquisition and language change. In essence, our expectation is that, if the cue is not sufficiently robustly attested in the PLD, the parameter will revert to its default value. Given the discussion in the preceding sections, we can see that this has clear implications for both acquisition and change. Thus the exercise of formulating parameters along these lines is one which amounts to making empirical predictions in these two domains.

3.5.2. *A markedness convention for syntax*

In terms of (54), it is easy to see how a parameter changes from a marked to a default value, and it is easy to see how the default values are related to the general simplicity metric in (22), since in each case some feature (or, in the case of (54D), a relation) which is present in the marked state is absent in the default state. But, as we discussed at the end of previous section, we must allow for change in the opposite direction too. One way to do this is by considering how markedness considerations may relate to *systems* of parameters, or perhaps subsystems of related parameters, rather than to

individual parameters. Hence, rather as in the case of distinctive features as discussed by Chomsky and Halle (1968, Chapter 9), it may be that the markedness of a particular parameter will depend on the values assumed by other parameters in a given system. Let us explore this idea further, and consider an illustrative possibility.

We saw in §1.5.1 and §2.5 that the head parameter (54F) is rather problematic as stated. If it were a single parameter, it would predict a spectacular clustering of properties, which is not actually attested in the majority of languages. As we mentioned in the discussion of word-order correlations in §1.6.1, Dryer (1992) shows that a minority of the languages in his sample actually conform to the predictions of this putative parameter across the whole range of head-complement relations. The majority of languages diverge at least in some respects. We suggested in Chapter 2 that (54F) should in fact be broken down into a series of related parameters relating to each head-complement pair. However, without some further statement, all predictions regarding word-order correlations are thereby lost. The preference for 'harmonic' ordering seems to derive from an overriding tendency for independent parameters to conspire to produce a certain type of grammar. To capture this, we tentatively suggested that a restatement of J. Hawkins' (1983) generalization regarding cross-categorial harmony is needed, along the following lines (repeated from (84) of §2.5):

- (55) There is a preference for the EPP feature of a functional head F to generalize to other functional heads G, H ...

Now it is time to relate (55) to the ideas about markedness we have been developing. We can think of (55) as an approximation to a markedness convention of the type proposed for phonology by Chomsky and Halle (1968).

To take a specific example, suppose, following Kayne (1994) and the discussion in §2.5.4, that VO is the universal underlying order and that OV orders derive from the combination of V-to-v raising and remnant VP-fronting to SpecvP, as illustrated in (56):

- (56) [_{IP} [_{VP} O (V)] v+V (VP)]

In terms of (54F), v has a marked property here. Following Chomsky and Halle's notation, let's write this as the *mEPP* value for v.²³ In rigidly head-final languages like Malayalam (see §1.5.1), many, perhaps all, functional

²³ Presumably v actually has two EPP features, since it attracts both V and VP. Here I am only concerned with the one which attracts VP.

heads will have at least one EPP feature in this way. Such systems will therefore emerge as very marked indeed, in terms of what we have said so far, and yet they are more common than 'mixed' types like Latin, German, etc., which would be less marked on this approach.

It is here that markedness conventions and the concept of the markedness of a whole system, or subsystem, of parameters comes in. Let us postulate, for concreteness, the following convention:

$$(57) \text{ For a class of heads } H, \text{ uEPP for } H_{\text{uF}} \neq v \rightarrow \left\{ \begin{array}{l} [+EPP] / \eta_{\{+EPP\}} \\ [-EPP] \end{array} \right\}$$

What (57) says is that the unmarked value of the EPP feature for some head of a particular type with an uninterpretable feature (i.e. a Probe, capable in principle of triggering movement) is [+EPP], i.e. the presence of an EPP feature, just where v has an EPP feature, i.e. in an OV system. (Here the EPP feature is understood to refer to attraction of VP rather than V; see note 23.) This convention would replace the default statement associated with the head parameter in (54F, b). This has the effect that, for all head-complement pairs which are subject to word-order variation, head-final is the unmarked order in an OV system, and head-initial in a VO system. In these terms, rigidly head-final languages are relatively unmarked, as of course are rigidly head-initial languages, while 'mixed' languages are relatively marked (and one can in principle quantify exactly how marked different types of mixed systems would be). Furthermore, Dryer's observation that VO vs. OV order is the basic determinant of ordering among other head-complement pairs is directly captured by (57). (See again the discussion of Dryer's results in §1.5.1.) What remains unclear, however, is how to specify the class of heads (57) refers to.

A possible disadvantage of (57) is that it appears to disconnect markedness from the simplicity metric in (22), in that we are now claiming that systems where the EPP feature is present on all possible heads are relatively unmarked. Hence a simple feature-counting approach to simplicity and thence to markedness no longer suffices. However, we can think that the simplicity metric itself derives from a more general notion of the conservatism of the learner, in that the learner will strive to assign the simplest representation or derivation possible to the PLD it is exposed to. In these terms, we can understand a markedness convention like (57) in terms of the conservatism of the learner, assuming that another conservative aspect of

the learner would be to exploit pieces of, perhaps marked, input to the full. So we could entertain something like the following:

(58) Generalization of the input:

If acquirers assign a marked value to H, they will assign the same value to all comparable heads.

Example (57) can naturally be understood in terms of (58), and both (57) and (22) can be seen as different aspects of the overall conservatism of the learner, which is essentially trying to set parameters in the most efficient way possible. The Subset Principle, as discussed in §3.4.3 above, can also be seen in this light: one aspect of the learner's conservatism is to avoid superset traps.

3.5.3. *From unmarked to marked*

A markedness convention like (57) also gives us a way of seeing how individual parameters may change from an unmarked to a marked value. All heads which are capable of bearing EPP features have the inherently unmarked property of not bearing this feature, but, where *v* has an EPP feature the opposite is true. Thus, if *v* acquires an EPP feature, a markedness reversal takes place for all the other heads in the system, and this creates pressure, ultimately due to (58) as a property of the learner, in the direction of acquiring the [+EPP] value for all other heads. Of course, this does not answer the question of how *v* might acquire an EPP feature. One might object to this approach along the same lines as Song's (2001: 304) objection to the Lehmann–Vennemann approach to word-order change discussed in §2.5.2. The preference for markedness-induced harmony must be weak enough to permit 'incongruous' word orders to arise in otherwise consistent grammars but strong enough to cause 'endogenous optimization' in Kiparsky's (1996: 150ff.) sense, i.e. harmonization of the relevant attraction properties of other functional heads. Kiparsky (1996: 153) defends his position as follows:

We can therefore legitimately posit a universal preference which is not universally instantiated, provided that we specify the other factors that allow (or force) it to be subverted. In principle, they might be either intersecting structural or functional constraints, or historical processes. Motivating the latter would, in the case at hand, amount to demonstrating a natural origin for OV syntax.

Kiparsky goes on to suggest that OV syntax could arise from a system in which objects (and perhaps other arguments) are in apposition to pronouns,

and hence frequently left-dislocated for reasons to do with information structure (topicalization, focalization, etc.). If the pronouns become agreement markers and/or disappear, OV syntax may emerge. Although speculative, this scenario gives an indication of how a dispreferred, possibly incongruous, order might arise through a separate, natural kind of change.

We can readily rephrase Kiparsky's speculation using the formal notions adopted here. In these terms, the question of the origin of OV orders becomes the question of how *v* could acquire an EPP feature. The most likely scenario for *v* acquiring an EPP feature is that whereby optional, discourse-driven object-movement becomes obligatory through the loss of the discourse effect, as Kiparsky suggests. This can be fairly naturally stated in terms of some of Chomsky's recent assumptions, since he allows for optional movement triggers (i.e. EPP features) as long as their presence has an effect on output, i.e. creates some kind of discourse effect (Chomsky 2001: 34). At an earlier stage *v*'s optional EPP feature would be dissociated from its uninterpretable φ -features (which might, following Kiparsky's suggestion, Agree with a resumptive pronoun in object position), but later the two sets of features would coalesce and the EPP feature would thereby be obligatorily associated with Agree in φ -features with the object, giving rise to object-movement (or possibly VP-pied-piping, as described in §2.5.4).

That the discourse effect is associated with a complication of structure through the imposition of an extra EPP feature gives a formal expression of the traditional intuition that the drive for expressivity is a factor in language change alongside the drive for simplicity (see Martinet (1955) for similar ideas in the context of sound change), and that, in the long run, these two forces create an overall equilibrium. This may be what prevents languages from developing the maximally unmarked steady state. Nichols' (1992) evidence that the overall degree of diversity in the world's languages has not changed since prehistory supports the idea that marked structures must be able to be innovated; in the terms just described, the tension between expressivity and simplicity balances out over the very long term, and there is thus no net increase or decrease in the markedness of the systems that are attested at any historical moment.

Expressivity may cause EPP features to be introduced, while simplicity causes them to be eliminated. Presumably, a constraint like (58) causing generalization of the input causes them to become obligatory. Furthermore, it is very likely that the 'coalescence' of EPP and φ -features alluded to above is driven by the preference for simplicity. We can thus envisage a

sequence of changes, starting from an optional EPP feature, to obligatory EPP combined with Agree, to simply Agree. For *v*, this would give rise to a sequence of changes from optional OV associated with a discourse effect, to obligatory OV associated with case marking and/or agreement, to VO. As already mentioned, this is fully consistent with the proposals in Kiparsky (1996).²⁴ Clearly, these speculations require a great deal more work before they can really be considered as established hypotheses, but we can see in principle at least how relatively marked systems might be innovated, and hence avoid the problems with positing marked and default parameter values which we noted in §3.4.6 above.

3.5.4. Networks of parameters

Looking at the ways in which parameters may interact in change and in determining the markedness of a whole system leads naturally to the question of networks of parameters. We already saw a very simple example of how one parameter may determine the value of another one in our brief discussion of the relation between the parameter determining multiple wh-movement and parameter (54E). We can think of this relation as a kind of intrinsic ordering, in that the multiple wh-movement parameter depends on the value of the superordinate parameter (54E): if (54E) has the negative value, then the multiple wh-movement parameter can only have the default value; it is effectively 'switched off' as an independent parameter.

A natural question to ask is how far this kind of intrinsic ordering among parameters can be taken. It is clearly a desirable feature of a parametric system, as it automatically ensures certain empirical predictions. If P_1 is superordinate to P_2 in the sense just defined, then P_1 must be acquired before P_2 and a diachronic change in the value of P_1 will potentially affect the value of P_2 , but not vice versa.

Baker (2001: 163) formulates exactly this notion of intrinsic ordering among parameters as follows:

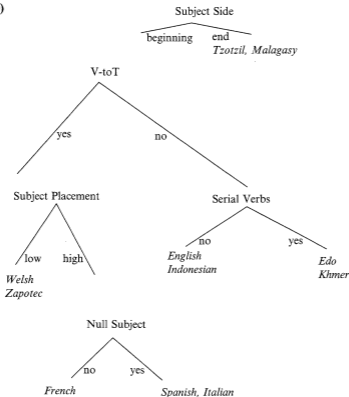
²⁴ In terms of Chomsky's (2005c) proposal that there may be a further variety of movement which is entirely separate from Agree, triggered by the Edge Feature EF (see note 17), we might replace the optional EPP feature at the first stage of the cycle with an EF feature. In that case, *v* changes through all the formal options the current theory makes available.

- (59) Parameter X ranks higher than parameter Y if Y produces a difference in one type of language defined by X, but not in the other.

Keeping to our rather simple example involving *wh*-movement for the purpose of illustration, (54E) would be parameter X in Baker's formulation, and the multiple *wh*-movement parameter would be parameter Y, since the latter produces a difference only in those languages with overt *wh*-movement and not in those without.

Baker develops what he calls a 'periodic table' for parameters using the notion of ranking (what we have been calling intrinsic ordering) in (59). A subpart of this, which involves some of the parameters in (54), is given in (60) (Baker's Figure 6.4, 183, presents the full system he proposes):

- (60)



V-to-T is the familiar parameter (54B), and Null Subject is our (54A). We have seen the Subject Placement parameter, notably in §1.2.1; this parameter determines whether the subject raises to SpecTP. Baker's Subject Side parameter determines, roughly, whether the subject appears at the beginning of the clause, or at the end. Tzotzil and Malagasy are VOS languages, in which the canonical position of the subject is final, while all the other languages listed here are SVO or VSO, depending on the value of the Subject Placement parameter; the head parameter is superordinate to all the parameters in (60).²⁵ The Serial Verb parameter determines how many verbs a single VP (or perhaps vP) can contain (Baker 2001: 141). English doesn't allow more than one (main) verb per VP/vP, but many languages do, for example Edo (a Niger-Congo language spoken in Nigeria) does:

- (61) Ózò ghá lè èvbàré khiɛ'n.
 Ozo will cook food sell
 'Ozo will cook the food and sell it.'
 (Baker 2001: 140)

What (60) actually states is a series of intrinsic ordering relations among parameters: the Subject Side parameter is superordinate to the others given here in that only the 'beginning' value for this parameter allows a choice in the V-to-T parameter, since the clause-final position of the subject means that the position of V in T or lower would not affect word order. Similarly, only the positive value of V-to-T allows an option regarding Subject Placement, since the negative value will result in SVO order whether or not V moves. Only a negative value of V-to-T is compatible with the option of serial verbs, since if there are two verbs in vP/VP it is impossible for both to raise to a single T-position.

A hierarchy of parameters of the kind in (60) makes interesting predictions regarding typology, acquisition, and change. Regarding typology, it

²⁵ It has recently been proposed that VOS order should be derived by raising VP (excluding the subject, which is taken to be merged in SpecvP) to SpecTP. This was first proposed by Massam and Smallwood (1997); see also Massam (2000; 2005); Rackowski and Travis (2000); Chung (2005). This analysis would be consistent with Baker's proposals as given in (60), as VP-fronting to SpecTP would arguably 'bleed' both V-to-T movement (since the verb must remain in the fronted VP for the VOS word order to result from VP-fronting) and subject raising, if VP-movement satisfies T's EPP-feature, as suggested in the references just given.

predicts a series of implicational universals: if a language has serial verbs, it is SVO (VSO depends on the positive value of V-to-T, and VOS on the 'end' value for Subject Side; recall that OV languages are determined by the higher-order head parameter); if a language has null subjects, it has V-to-T; if a language is VSO, it does not have serial verbs. Clearly, all of these predictions are testable (and some of them are false: for example, (60) predicts that if a language is VSO it does not have null subjects, but Welsh, Irish, and Classical Arabic are all VSO null-subject languages; this point is also made by Newmeyer (2004: 201; 2005: 86)).

In the domain of language acquisition, (60) predicts what Dresher, following Lightfoot (1989), calls a 'learning path'. (Baker (2001: 192–6) also makes this point.) The setting of a superordinate parameter will determine whether or not there is an option to set a subordinate parameter. For example, choosing the 'end' value of Subject Side pre-empts the setting of any of the other parameters in (60). In terms of what we said earlier regarding parameter interactions and the schema for parameters in (52), we could take this to mean that all the subordinate parameters automatically take on the default value. (This would be the case because all the PLD would be weakly P-ambiguous in the sense of (21).) We thus predict that acquirers of Italian follow a learning path starting from Subject Side (beginning), and going on to V-to-T (yes), Subject Placement (high) and Null Subject (yes). Acquirers of English, on the other hand, set V-to-T to the negative value and then the Serial Verb parameter to the negative value. Again, the predictions for language acquisition are clear in principle. However, once again, the evidence for very early parameter-setting discussed in §3.1 makes it difficult to test these in practice (Baker's conclusion is slightly more optimistic than this, however).

Finally, (60) makes interesting predictions about relations among parametric changes. For example, if a language loses V-to-T movement, then it simultaneously loses the possibility of having VSO order or null subjects, but may go on to develop serial verbs. The history of English since the loss of V-to-T in the Early Modern period is consistent with this, but only in a rather unrevealing way, since serial verbs have not in fact developed. English- and Romance-based creoles, on the other hand, support this, in that they tend to lack V-to-T and (argumental) null subjects, and to have SVO order and serial verbs. (See Muysken (1988); the papers in DeGraff (1999); Nicholis (2004); and §5.3.2 on the syntactic properties of creoles.)

There are two obvious objections one can make to (60). First, it is empirically incorrect, in that certain pairs of parameters are set in the wrong relation with one another. We mentioned an example of this above: the incorrect prediction that there are no null-subject VSO languages. Another incorrect prediction is that there are no null-subject languages which have serial verbs: many East Asian languages, including Chinese, Thai, and Vietnamese, appear to show both of these properties. However, such difficulties can be dealt with quite easily; the hierarchy simply needs to be appropriately revised. The second difficulty is perhaps more serious: Baker allows for the possibility that a given pair of parameters may be logically independent, and in fact discusses (184ff.) several well-established parameters which do not appear to fit into his hierarchy, notably the *wh*-movement parameter (54E). But of course if too many pairs of parameters are independent from one another the hierarchy may start to lose its clear structure. The greatest difficulty would arise if one could show that, for a triad of parameters P_1, P_2, P_3 , P_1 is superordinate to P_2 , P_2 is superordinate to P_3 , but that P_3 is superordinate to P_1 ; this would create a kind of ordering paradox, since the relation ' P_n is superordinate/subordinate to P_m ' is, one assumes, logically transitive. It is not clear whether a case like this actually exists. If it does, then the concept of parameter hierarchy would have to be abandoned in favour of a looser notion of network, and presumably some of the wide-ranging predictions that a hierarchy like (60) makes would be lost. As things stand, though, Baker's proposals, or some variant of this hierarchy, are of great interest especially for establishing connections between acquisition and change, and have not been shown to be unworkable.²⁶

A final point regarding parameter interactions concerns the possibility of 'cascades' of changes: a situation where an initial parameter change perturbs a system in such a way that a whole series of changes follows, perhaps over many centuries, creating the appearance of typological drift.

²⁶ Newmeyer (2004; 2005) argues against Baker's parameter hierarchy. But his main critique is really the same as the first point just made: some of the parameters may be placed in the wrong relationship to one another. As pointed out by Roberts and Holmberg (2005), this is not really a criticism of the concept of a parameter hierarchy (still less of the concept of parameter itself), but rather of Baker's specific implementation of it. Roberts and Holmberg further take issue with a number of Newmeyer's criticisms of the principles-and-parameters approach to comparative syntax.

As we noted earlier, Longobardi (2001: 278) mentions this possibility in his discussion of Inertia, pointing out that syntactic change can be the consequence 'recursively, of other syntactic changes'. This possibility undoubtedly exists, and may be behind the observations of typological drift that have been made. I will defer detailed discussion of this to §4.3.4, where I explore this idea by looking at a series of changes which took place in the history of English between roughly 1400 and 1700. For the moment, the only relevant point is that the intermediate grammars during the sequence of changes must all be relatively highly marked, and therefore prone to change. Clearly, typologically 'mixed' systems will be of this kind, given the postulated markedness convention in (57).

3.5.5. Conclusion

In this section, we have attempted to consolidate the discussion in the earlier sections of the chapter, and to some degree that in the earlier chapters as well, by considering the format for parameters and the various ways in which parameters may interact, giving rise to networks or hierarchies, as well as the concept of markedness of an entire system. As we saw, markedness of a system may override the markedness specification of an individual parameter.

3.6. Conclusion to Chapter 3

This chapter has attempted to consolidate the ideas which were introduced in the first two chapters. There we first tried to demonstrate the utility of the notion of parameter of UG for analysing syntactic change (Chapter 1) and for giving a (near-) unified account of different types of change (Chapter 2). Here, we tried to show how parameter change can be seen as driven by language acquisition. The essential notion is that of the conservatism of the learning device, which always attempts to set parameters on the basis of the greatest computational efficiency. This has at least two consequences that we have seen: a strong tendency to favour relatively simple representations or derivations, which we stated as (22); and a tendency to generalize the input, which we formulated as (58), underlying the markedness

convention in (57). Both of these properties motivate a formulation of the default values of parameters (although (57) concerns the markedness of systems, and as such may create a markedness preference which overrides the markedness value of a single parameter). This contributed to our general statement of the format of parameters in (52), inspired by the proposals in Dresher (1999), whereby each parameter consists of a formal statement (itself highly restricted by the impoverished mechanisms of minimalist syntax – see note 20), a statement of the (defeasible) default, and a statement of the cue or expression of the parameter. This format seems to have the right properties and is useful for looking at both acquisition and change.

In seeking to relate parametric change to language acquisition, we undertook a survey of recent work on the acquisition of syntax in §3.1. Here we encountered the Very Early Parameter Setting observation, which to some extent hampers establishing a straightforward relationship between acquisition and change, although it does not preclude such a relationship. §3.2 discussed the logical problem of language change, which led us to the formulation of the simplicity metric in (22). The subject of §3.3 was the changing trigger. Given Inertia, i.e. the idea that syntactic change must be caused (to paraphrase Longobardi (2001: 278)), we considered how contact and morphological erosion may induce change. §3.4 dealt with markedness at some length. We suggested that (22), along with various markedness conventions in the pattern of Chomsky and Halle (1968, Chapter 9), should form the basis of markedness. The Subset Principle may also be relevant if parameter systems allow formal optionality. Finally, in §3.5 we arrived at our formulation of parameters in (52) and considered its implications for networks and hierarchies of parameters, paying particular attention to the proposals in Baker (2001). We also made a suggestion for how marked properties may be innovated, at least in the case of EPP features. This suggestion seems to capture the old idea that much of language change is caused by a tension between a drive for simplicity and a drive for expressivity.

In the next two chapters, we look at the consequences of the general view of syntactic change that we have outlined over the preceding chapters. We begin, in Chapter 4, by looking at the dynamic aspect of syntactic change – and considering how it might be handled in the terms described here. Chapter 5 focuses on questions connected to contact, substratum effects, and creoles.

Further reading

Principles-and-parameters theory

Baker (2001) is an excellent introduction to the principles-and-parameters conception of UG. Baker pursues a sustained analogy between contemporary comparative syntax and nineteenth-century chemistry, which culminates in a 'periodic table' of parameters, part of which is reproduced in (60). Baker observes that any analogue to the quantum-theoretic explanation of why the periodic table has the properties it has is far off. **Newmeyer (2004; 2005)** argues at length that the principles-and-parameters approach to comparative syntax has failed, and that variation across grammatical systems should be handled in terms of performance systems of various kinds. **Roberts and Holmberg (2005)** is a reply to Newmeyer, arguing that the principles-and-parameters approach is a valid and useful approach to comparative syntax.

Learnability and markedness

Lasnik (1983) is an early discussion of learnability in relation to principles-and-parameters theory. **Lightfoot (1989)** is the first statement of the degree-0 learnability idea, developed at much greater length in Lightfoot (1991). **Fodor (1998)** proposes a learnability theory for syntax. **Roberts (2001)** looks at the relation between syntactic change and learnability, proposing a version of the simplicity-based approach to markedness summarized in this chapter. **Dresher and Kaye (1990)** is the initial proposal for cue-based learning of phonological parameters, later developed in Dresher (1999). **Berwick (1985)** first put forward the Subset Principle as a natural learnability-driven constraint on the language-acquisition process. **Manzini and Wexler (1987)** offer an account of parametric variation involving long-distance reflexives, which makes explicit reference to the subset relations among the languages produced by grammatical systems defined by the different parameter-settings proposed. This represents a further case where the Subset Principle may be relevant for understanding the relations among parameter values, and perhaps as a basis for a theory of markedness. **Niyogi and Berwick (1995)** is a pioneering study of how syntactic

change can be mathematically modelled. **Battistella (1996)** is an introduction to and historical overview of the concept of markedness, with particular reference to syntax.

Language acquisition

Hyams (1986) is the ground-breaking study of the acquisition of syntax using the principles-and-parameters approach, in which the phenomenon of early null subjects is first described. Hyams analysed these as null subjects of the Italian type, an idea she has abandoned in subsequent work (see **Hyams (1992)**; **Hyams (1996)**; **Hyams and Wexler (1993)**). **Radford (1990)** was the first to generalize Hyams' (1986) account, and argue that English children, at least, go through a stage of acquisition in which no functional categories are available at all. This work led directly to the postulation of root infinitives and early null subjects. **Pierce (1992)** is a pioneering study of Early French, in which it is shown that at the root-infinitive stage, the infinitival form of the verb does not raise to T while the optional finite form does. **Poeppl and Wexler (1993)** is an important study of Early German, in which they argue for a root-infinitive stage in that language, and that infinitive verbs do not undergo the verb-second operation (i.e. they do not move to C). **Rizzi (1994)** is an influential study of root infinitives, in which it is argued that these derive from the possibility of 'clausal truncation', i.e. realising a clause as a VP only, at a stage of acquisition in which the language faculty is not fully mature. **Rizzi (2000)** proposes something similar for 'diary-drop'. **Clahsen, Kursawe, and Penke (1995)**; **Clahsen and Penke (1992)**; and **Clahsen and Smolka (1985)** are all studies of Early German, in which it is shown that the complex adult verb-movement system develops according to a series of well-defined stages. **Guasti (1996; 2000)**; **Haegeman (1995a)**; **Haegeman (1995b)**; **Hamann and Plunkett (1998)**; and **Hoekstra and Hyams (1998)** are all studies of the early stages of the acquisition of various Romance and Germanic languages from the perspective of principles-and-parameters theory. Hoekstra and Hyams' article is noteworthy for advocating that Early Null Subjects of the kind found in non-null-subject languages such as English and other Germanic languages are not to be equated with those found in null-subject languages such as Italian. **Wexler (1992; 1994; 1999)** provides overviews

and summaries of much of the work in this field, as well as developing more general ideas, notably the Very Early Parameter Setting observation discussed in §3.1. **Friedemann and Rizzi (2000)** is a collection of important articles on the acquisition of the syntax of a range of Germanic and Romance languages. **Brown (1970)** is an early and very influential study of the first-language acquisition of English, while **Jakobson (1941)** is, among other things, a ground-breaking study of language acquisition and language disorders, in which the concept of markedness plays an important role. **Ernst (1985)** is an extremely detailed study of Héroard's journal, in which the speech of the young dauphin was recorded over a period of several years. Héroard's journal is a unique document, of potentially great interest for language acquisition and language change, as well as providing a valuable record of the nature of spoken French in the early seventeenth century. **DeGraff (1999)** is a highly original collection of articles dealing with creolization, language acquisition, and language change. The Introduction and Epilogues are extremely useful and thought-provoking. This collection represents a unique attempt to bring together these areas, which have often been studied somewhat in isolation from one another.

The null-subject parameter

Huang (1984; 1989) develops a 'generalized-control' approach to null subjects in Chinese and Italian, covering also the distribution of the null subject of non-finite clauses in languages such as English (conventionally known as PRO in government-binding theory). **Nicholis (2004)** is a detailed study of the status of the cross-linguistic predictions made by the version of the null-subject parameter put forward in Rizzi (1982), and given in (28) of §1.2.1. He concludes that the correlations hold up fairly well across a wide range of languages, but that the distribution of expletive null subjects in creoles is problematic (see §5.3.2, on these). **Holmberg (2005)** is a recent and very original paper on null subjects, arguing, on the basis of the fact that Finnish has an overt expletive subject which appears to be in complementary distribution with a referential null subject, that null subjects are structurally pronouns.

Other works on syntactic theory

Giorgi and Pianesi (1997) propose a general theory of the syntax of temporal relations, and an analysis of the temporal systems of Italian, English, and Latin. A facet of their approach is the idea that functional heads are structurally present only when needed in order to bear certain features. This view differs notably from that put forward by Cinque (1999). **Massam (2005)** and **Massam and Smallwood (1997)** are analyses of Polynesian languages with VSO and VOS orders in which the central idea is that the V-initial orders derive from VP-fronting, possibly of a remnant VP. **Chomsky (2005c)** is, at the time of writing, the most recent statement of certain technical aspects of minimalism. This paper pays particular attention to the 'A'-system', i.e. wh-movement, topicalization, and focalization, all movements to the Specifier(s) (or 'edge') of CP. It is proposed that these movements are triggered by the E(dge) F(eature), a feature that has no connection with the Agree system. **Richards and Biberauer (2005)** develop an analysis of the distribution of what have often been seen as overt and null expletives in Germanic (see the discussion of null expletives in §1.2.1) which makes use of the twin notions of 'massive movement' of vP to SpecTP (as briefly described in §2.5.4) and the optionality of pied-piping operations. **Biberauer and Richards (2006)** make a very interesting and well-argued case for formal optionality in syntax, arguing in particular that this is a natural outcome of the kind of minimalist syntax proposed in Chomsky (2000; 2001). We will look at some of their proposals in more detail in §4.1.4.

Historical and typological syntax

King (2000) is a detailed and very interesting study of Prince Edward Island French. In addition to arguing convincingly that Preposition-stranding in this variety is the result of extensive borrowing of English prepositions, as summarized in §3.3 above, King looks at the syntactic consequences of the borrowing of the particle *back* and the wh-elements *whoever*, *whichever*, etc., into this variety of French. **Keenan (2002)** is a very detailed, original and interesting study of the development of English reflexives, arguing convincingly that they were originally emphatic forms. It is here that the

Inertia Principle is proposed for the first time. **Kroch et al. (1997)** is a detailed study of the loss of V2 in the history of English, arguing that this change was driven by contact between Northern and Southern dialects of ME. **Croft (2000)** puts forward a general account of syntactic change in functional-typological terms, hence differing in its basic assumptions for what is being put forward here. **Nichols (1992)** put forward a number of important innovations in language typology. The distinction between 'head-marking' and 'dependent-marking' was first made here. This is the distinction between a system in which a grammatical notion is marked on a head or on a dependent of that head, for example, marking grammatical functions through verb-agreement (head-marking) vs. case on nominals (dependent-marking). Nichols also made a number of proposals regarding the areal distribution of typologically variant properties. **Hogg (1992–2001)** is the invaluable *Cambridge History of the English Language*, a six-volume work which provides extremely detailed information about every aspect of the history of the language, from its Germanic and Indo-European origins to the present day. **Jasanoff (2004)** is a description of Gothic, a contribution to the *Cambridge Encyclopedia of Ancient Languages*, which gives descriptions of all languages known to have existed prior to 500AD for which a reasonable amount of data is available.

Phonological theory and phonological change

Chomsky and Halle (1968) is the classic exposition of generative phonology. It is notable for the system of distinctive features proposed, for the explication of the functioning of an ordered system of phonological rules, for the postulation of the levels of 'systematic phonetics' and 'systematic phonemics', for the analysis of the cyclic nature of stress-assignment in English, and for the markedness conventions connected to the evaluation metric discussed in §3.4. **Kiparsky (1973)** is a treatment of the nature of rule-ordering in the standard model of generative phonology as put forward by Chomsky and Halle (1968), in which the Elsewhere Condition is put forward as a condition determining one kind of rule-ordering. The origins of this condition in the works of Pāṇini's are explicitly acknowledged. **Kenstowicz (1991)** is a standard, comprehensive introduction to (pre-optimality-theory) generative phonology. **Martinet (1955)** is a classic

structuralist account of phonetic and phonological change, in which the fundamental idea is that sound changes arise from the interaction of economy (ease of articulation) and expressiveness (the need to make distinctions). **Lass (1992)** is a further contribution to the *Cambridge History of the English Language*, in which the phonology and morphology of Middle and Early Modern English are described in detail.