The papers collected here represent some of the work carried out in the period 2002-2007 by the group working on the project “Null Subjects and the Structure of Parametric Theory”, funded by the Arts and Humanities Research Council, Great Britain (Grant No. APN14458). The group consisted of Theresa Biberauer, Anders Holmberg, Chris Johns, Ian Roberts, Michelle Sheehan and David Willis. The central goal of that project was to investigate and, if possible, refine the notion of parameter of Universal Grammar, as it has been understood in generative theory since roughly 1980, by looking carefully at the phenomena associated with one of the best-known and most widely discussed examples of a parameter: the null-subject parameter (NSP). This volume brings together a number of articles focusing on the nature of null subjects in a range of languages; Biberauer (2008a) is a sister volume arising from that project.

---

1 We’d like to thank the other members of the null-subject project group for their comments on various earlier versions of this work. We’d also like to thank Hans Broekhuis, Norbert Corver, Riny Huybregts, Ursula Kleinhenz and Jan Koster, the editors of Organising Grammar: Linguistic Studies in Honor of Henk van Riemsdijk (De Gruyter 2005) for their comments on the Roberts & Holmberg contribution to that collection, parts of which are taken up again in Sections 2.3 and 2.4. Our thanks also to Fritz Newmeyer, for criticising us (see Newmeyer 2006), and to Pino Longobardi, Cristina Guardiano and Claudia Gianollo for being such stimulating interlocutors. Last but not least, Roberts would like to thank Alastair Appleton, Bob Davies, Luke Donnan and Iain Mobbs, all students on the Cambridge Linguistics MPhil 2007-8, for listening to and reacting to some of the ideas discussed in Sections 2 and 3 so assiduously and so intelligently.
from the same project, which focuses more on parameter theory than on null subjects, while Holmberg (to appear) focuses on partial null-subject languages.²

In this Introduction, we would like to set the papers in context. Accordingly, we first discuss the phenomena from English and various Romance languages which originally motivated the postulation of the NSP. Next, we summarise the main kinds of null-subject system that have been identified in the comparative-syntax literature. We complete Section One by summarising the two principal approaches to the analysis of null subjects, whose classical exponents are Rizzi (1986) and Borer (1986).

Section Two focuses on the debates surrounding the classical formulations of the NSP, and in particular the “typological” predictions that were initially made, beginning with Rizzi (1982). Here we consider the systematic cross-linguistic investigation whose results apparently indicate that certain predicted correlations do not hold (Gilligan 1987), and the far-reaching and negative conclusions for parameter theory drawn partly on the basis of this by Newmeyer (2004, 2005). We once again attempt to defend a version of P&P theory against Newmeyer’s attacks (see also Roberts & Holmberg (2005), Newmeyer (2006), Biberauer (2008b)).

Section 3 takes up the wider question of the nature of parameters; here we observe certain problems with the original view, which associated parametric variation closely with a rich, domain-specific array of UG principles. This view can no longer be

² Other material, published and unpublished, generated by the project is detailed on the project website (http://people.pwf.cam.ac.uk/mtb23/NSP/Nullsubjectsprojecthome.html).
maintained in full in the context of the minimalist programme, which undertakes to eliminate as many UG principles as possible, and which calls into question the domain-specificity of what principles we are required to postulate. Ongoing comparative work over the past twenty years or so has tended by and large to favour the postulation of a large number of microparameters (see Kayne (2005) for discussion, and Baker (1996, 2008) for a different view); this view of parameters is readily compatible with minimalist assumptions, arguably more so than a “macroparametric” approach. We suggest that the proliferation of microparameters is an instance of the familiar tension between descriptive and explanatory adequacy, which at earlier stages in the development of the theory motivated the simplification of rule systems and led to the development of the principles-and-parameters approach itself,\(^3\) as before, what seems to be required is a radical increase in theoretical abstraction. We make some tentative suggestions in this direction in Sections 3.4 and 3.5, developing ideas in Gianollo, Guardiano & Longobardi (2008), Roberts (2007a:443, 2008:17, chapter 1), Roberts & Roussou (2003, Chapter 5) and Mobbs (2008). This leads us to briefly propose a hierarchical model of parameter schemata which combines the notion of micro- and macroparameters, and to speculate on the shape of comparative syntax beyond explanatory adequacy.

---

\(^3\) A similar point is made in Roberts (2001:90) in relation to parametric accounts of syntactic change.
1. **Null Subjects: The Basics**

1.1 **The observation**

Traditional grammars of many languages, for example Latin, observe that a pronominal subject is marked “in the verb,” i.e. by the person-number agreement inflection on a finite verb, and as such is not in need of expression by an independent pronoun. The following comment from a well-known traditional grammar of Latin is representative: “Here [in the finite verb, AH/IGR] the form contains in itself all the necessary elements .., the persons being indicated by the endings” (Gildersleeve & Lodge (1895/1997:144)). Jespersen takes up this idea in the following remark:

> In many languages the distinction between the three persons is found not only in pronouns, but in verbs as well .. in Latin .. Italian, Hebrew, Finnish, etc. In such languages many sentences have no explicit indication of the subject, and *ego amo, tu amas* is at first said only when it is necessary or desirable to lay special stress on the idea “I, thou.” (Jespersen (1924:213))

This idea has an ancient pedigree, as indicated by the following remark by Apollonius Dyscolus on Ancient Greek:

> The nominative [subject] is implicitly present in [finite] verbs, and it is definite (i.e. has definite reference) in the first and second persons, but indefinite in the third because of the unlimited number of possible referents.  

*(On Syntax, Book 1, §17; Householder (1981:25))"
What is being alluded to here is that, since a pronominal subject can be expressed “in the verb” in languages such as Greek and Latin, there is no general requirement to pronounce the subject separately as a nominative pronoun. The initial observation behind the NSP, in all its formulations, has to do with this second point: an overt pronominal subject is not required in finite clauses, and, when such a pronoun does not appear, there is no nominal element which overtly realises the subject function in the clause. This fact may reflect a trivial feature of morphology, person-marking on the verb, but the possibility of not expressing the nominal bearing the subject function is of obvious importance for syntax.

The first generative study of these matters was Perlmutter (1971). Perlmutter (pp. 100ff.) distinguished languages with the surface filter in (1), which he called Type A languages, from those lacking it, Type B languages:

(1) “Any sentence other than an Imperative in which there is an S that does not contain a subject in surface structure is ungrammatical.”

(Perlmutter’s (9), p. 100)

Perlmutter relates the presence of the surface filter in (1) to the possibility of null subjects and of wh-movement of the subject from a finite embedded clause across a complementiser (this observation has since become known as “Perlmutter’s generalisation”), to the presence of obligatory expletives in the relevant kind of impersonal constructions, and to the existence of an arbitrary subject pronoun as a true subject (as opposed, for example, to an arbitrary subject clitic pronoun which
surfaces as part of the object-clitic cluster). French and English are examples of Type A languages, while “Spanish, Italian, Serbo-Croatian, Arabic, Hebrew, Warlbiri and Basque” (Perlmutter (1971:115)), as well as numerous others are Type B languages. This “typological distinction,” as Perlmutter referred to it, is not, however, connected to the nature of agreement inflection in the Type B languages in Perlmutter’s treatment. With this exception, and with the important omission of any discussion of “free inversion” (see below), Perlmutter’s discussion identifies the NSP in all but name.4

As Perlmutter pointed out, the basic fact motivating the postulation of this parameter is that certain languages require finite clauses to overtly express a definite, referential, pronominal subject, while others do not. The contrast is illustrated by the following Italian and English examples:

(2) a. Parla italiano.

b. *Speaks English.

Spanish and Greek, among many other languages, pattern like Italian, while, as Perlmutter pointed out, French appears to pattern like English ((3c) is ungrammatical as a declarative, although it would be a well-formed imperative):

4 One might also observe, with hindsight, that the surface filter in (1) is very close to the original formulation of the Extended Projection Principle put forward in Chomsky (1982:10), i.e. the requirement that every clause must have a subject. For Perlmutter, (1) is parametrised (although, again, this terminology postdates his insights); see Section 1.3 on proposals that the NSP involves parametrising the requirement for a structural subject.
(3)  
   a.  Habla español.
   b.  Mila ellinika.
   c.  *Parle français.

Thus Italian, Spanish and Greek are null-subject languages, while English and French are non-null-subject languages.

The NSP relates, as stated above, to finite, discourse-neutral clauses, and canonically involves the interpretation of the null subject as a definite, referential pronoun. Many non-null-subject languages, including English, allow null subjects under other conditions. For example, both English and French extensively allow or require the subject of non-finite clauses to be null:5

(4)  
   a.  [ (Him) smoking ] bothers me.
   b.  John expects [ (Mary) to leave soon ].
   c.  Jean a essayé [ de – partir ].
      John has tried [ -- to leave ].

Such subjects have somewhat different properties from the null subjects of (2) and (3), in that in (4b,c) the empty subject of the infinitive must be coreferent with the subject of the main clause (this is subject control) and that in (4a) must be arbitrary.

5 Examples of this type do not fall under Perlmutter’s surface filter in (1) since it was generally assumed that such infinitives were VPs (resulting from Raising in (4b) where Mary is present, and Equi-NP Deletion where it is not, and in (4c), combined with tree-pruning in the sense of Ross (1967, Chapter 3)). In cases like (4a), the S-node of the gerund is pruned when the subject is not present.
Accordingly, they have generally been analysed in a different way from those of (2) and (3).\(^6\)

The initial observation, then, is that some languages allow a definite pronominal subject of a finite clause to remain unexpressed as a nominal bearing the subject function, while others do not. Traditional grammars of languages such as Latin and Greek relate this to the fact that personal endings on the verb distinguish person and number, thereby making a subject pronoun redundant. Languages which allow null subjects are very common: most of the older Indo-European languages fall into this category, as do most of the Modern Romance languages (with the exception of some varieties of French and some varieties of Rhaeto-Romansch; see Roberts (chapter 8), the Celtic languages, with certain restrictions in the case of Modern Irish (see McCloskey & Hale (1984), and, for arguments that Colloquial Welsh is not a null-subject language, Tallerman (1987)), West and South Slavic, but probably not East Slavic.

---

6 Both English and French also allow null subjects in special discourse environments or registers. Haegeman (2000:130) gives the following examples from what she calls “written abbreviated registers” (“written registers in which pressures of economy seem to overrule the ‘core’ grammar” (132), including diaries, short notes and some kinds of colloquial speech):

(i) a. --- cried yesterday morning.  
    (Plath (1983:288))

b. Elle est alsacienne. --- paraît intelligente.  
    (Léautaud (1989:48))

She is Alsatian. Seems intelligent.

Such null subjects, in addition to being restricted to certain types of discourse and/or register, have special properties which distinguish them from the canonical null subjects of (2) and (3) (see Haegeman (2000:138-141) for details). See Holmberg (chapter 2, section 2). We will leave these cases aside here.
Slavic (these appear to be “partial” null-subject languages in the sense of §1.2.4 below and Holmberg & Sheehan (chapter 3). Indeed, it seems that languages which allow null subjects are significantly more widespread than those which do not (Gilligan (1987), cited in Newmeyer (2005:85)). According to the Haspelmath, Dryer, Gil & Comrie (2005)’s *World Atlas of Language Structures* (WALS), of 674 languages for which data is available, subject pronouns can be omitted in 409, and cannot be omitted in 77 (the remainder form various kinds of mixed categories involving clitics, displaced pronouns, etc.; see Map 101 “Expression of Pronominal Subjects”). So null-subject languages, of one kind or another, are considerably more common than non-null-subject languages.

1.2 *Types of null-subject systems*

Since Rizzi’s early work on null subjects, it has been observed that there are different types of null-subject language. Rizzi (1982:143) proposed that the NSP be divided into two subcases, one applying to languages in which the unexpressed pronoun can only be an expletive, and one applying to languages where it is able to be referential. Huang (1984) observed that many East Asian languages show a much more liberal option of non-expression of pronominal elements, and that this could not be related to person agreement, since that kind of inflection is generally absent in these languages. More recently, the existence of “partial” null-subject languages has been observed: languages in which the pronominal subject may remain unexpressed under restricted conditions determined by both the morphological and the syntactic context. We now briefly describe each of these types of null-subject language one by one.
1.2.1 Consistent null-subject languages. These languages have been the most discussed and analysed among the various types of null-subject languages and have, mainly for historical reasons, often been taken to be the only kind of null-subject language. In consistent null-subject languages, all persons in all tenses can feature an unexpressed pronoun. These languages characteristically show “rich” agreement inflection, i.e. distinct personal endings on the verb, generally in all tenses. The Italian, Greek and Turkish forms in (5) illustrate:

(5) a. Italian
bevo  “I drink” (etc)
bevi
beve
beviamo
bevete
bevono

b. Greek
pino  “I drink” (etc)
pinis
pini
pinume

There are sometimes limited exceptions to this generalisation. In Italian, for example, the 2sg pronoun tu must appear when the verb is in the subjunctive. In the present subjunctive, the singular forms of the verb are not distinct: che (io) parli, che tu parli, che (lui/lei) parli (“that I/you/he/she speak”).
The Romance null-subject languages and Modern Greek are the paradigm examples of this kind of language, and have been much discussed and exemplified in the literature. These languages also illustrate the properties originally proposed to form a cluster determined by the positive value of the NSP, which we will turn to in Section Two below.

One further property that we can mention here is that, as pointed out by Jespersen in the quotation above, overt subject pronouns are generally allowed in finite clauses in null-subject languages, although they tend to have what we may loosely call an emphatic interpretation (this is indicated by putting the English pronoun in capitals in the translations below). Thus, alongside (2a) and (3a,b) we have:

---

But see Öztürk (2001, 2008) for the view that Turkish is a “discourse pro-drop” language in the sense of §1.2.3 below. This paradigm is from Csató & Johansen (1998:214).
(6)  a.  Lui parla italiano.  (Italian)
    HE speaks Italian.

    b.  Él habla español.  (Spanish)
    HE speaks Spanish.

    c.  Aftos mila ellinika.  (Greek)
    HE speaks Greek.

This aspect of the interpretation of overt pronominal subjects in null-subject languages emerges slightly more clearly in (7). Here the overt pronoun in the adverbial clause does not allow the interpretation in which it corresponds to the subject of the main clause (see Vanelli, Renzi & Benincà (1986), Samek-Lodovici (1996), and Frascarelli (2007) for discussion):

(7)  a.  Il professore ha parlato dopo che (lui) è arrivato.  (Italian)
    The professor has spoken after that (he) is arrived
    “The professor spoke after he arrived.”

    b.  I Maria jelase afou (afti) idhe ton Yianni.  (Greek)
    The Mary laughed after (she) saw the Yiannis.
    “Mary laughed after she saw Yiannis.”

In other words, the overt pronoun of the adjunct does not show the same ambiguity as its English and French counterparts in (8). Instead, it strongly prefers the interpretation which is disjoint from “the professor”, while the English and French
pronouns are, out of context, ambiguous between this interpretation and the one where they correspond to “the professor”:

(8) a. The professor spoke after he arrived.
    b. Le professeur a parlé après qu’il est arrivé. (=8a))

These interpretative differences involving the use of an overt pronoun appear to be related to the fact that subject pronouns may be unexpressed, i.e. to the positive value of the NSP.

For the moment, we take the two diagnostic features of a consistent null-subject language to be (i) the possibility of leaving the definite subject pronoun unexpressed in any person-number combination in any tense, and (ii) the rich agreement inflection on the verb. We will suggest other properties as we proceed.

1.2.2. Expletive null subjects. Some languages apparently allow expletive null subjects, but not referential ones. German is one such language, as are some varieties of Dutch and Afrikaans, and a range of creoles (Nicolis (2005, 2008) mentions Cape Verdean, Berbice Dutch, Kriyol, Mauritian, Papiamentu and Saramaccan; Roberts (2007a: 413) adds Haitian and Jamaican). In (9a) the expletive pronoun ès cannot be expressed, while in (9b) the same pronoun in the same syntactic position, only now with a referential interpretation, must be expressed (examples from Cardinaletti (1990:5-6)):
Owing to this restriction on their null subjects, languages of this type are not regarded as “full” null-subject languages. Rizzi (1982:143) identifies what he called two “related but autonomous parameters”: one concerns whether an unexpressed pronoun is allowed at all, and the other whether referential pronouns are allowed to be unexpressed. In languages like English, both parameters are negative, while in Italian both are positive. In German and the creoles just mentioned, the first is positive and the second negative. Hence German allows non-referential null subjects, as in (9a), but not referential ones. According to Rizzi, the fourth logical option is impossible (“if an inflection cannot be pronominal, it cannot be referential either” (Rizzi (1982:143)). There is thus an implicational relation between the presence of referential null subjects and the presence of expletive null subjects (see Holmberg (chapter 2; this point is discussed in some detail in Roberts (2007b: 31-38)). For an alternative analysis of the German facts illustrated in (9), and related phenomena elsewhere in Germanic, see Biberauer (Chapter 4), and the references given there.
Expletive null-subject languages (sometimes called ‘semi-pro-drop languages’), then, are distinguished from consistent null-subject languages in that non-dummy pronouns cannot be left unexpressed.

1.2.3 “Discourse pro-drop” (also called “Radical pro-drop”). A good number of languages which are otherwise typologically and genetically distinct (Chinese, Japanese, Korean, Thai, Vietnamese and others) allow null subjects quite freely, but seem to be entirely without agreement marking of any kind. The case of Chinese is discussed in Huang (1984). Chinese allows both subjects and objects to remain unexpressed and have a definite pronominal interpretation, as illustrated in (10):

\[(10) \quad \begin{align*}
  a. & \quad \text{-- kanjian ta le} \\
     & \quad \text{(he) see he Asp} \\
  b. & \quad \text{Ta kanjian – le.} \\
     & \quad \text{He see (him) Asp} \\
     & \quad \text{“he saw him.”}
\end{align*}\]

Both pronouns can be dropped under the appropriate discourse conditions. It has been suggested since the earliest studies (Huang (1984), Rizzi (1986)) that the total absence of agreement marking may play a role in facilitating the very liberal availability of null subjects in these languages. Recently, three specific hypotheses have been put forward in this connection. First, Tomioka (2003:336) proposed the “Discourse Pro-Drop Generalization” (see Jayaseelan (1999) for a similar idea):
“All languages which allow discourse pro-drop allow (robust) bare NP arguments”; “Null pronouns in Discourse Pro-Drop languages are simply the result of N’-Deletion/NP-Ellipsis without determiner stranding”.

This idea expresses a relation between discourse pro-drop and the availability of bare NP arguments (i.e. the grammaticality of a sentence such as (I) saw cat, thereby relating discourse pro-drop in an interesting way to Chierchia’s (1998) Nominal Mapping Parameter). It also relates discourse pro-drop to ellipsis, as does Saito (2007) (see below and the discussion in Roberts (chapter 1)).

A different proposal is made by Neeleman & Szendröi (2007, 2008). These authors treat fully specified nominals as KPs (since they inherently contain a syntactic position for Case) and posit an operation of context-free KP-deletion. In languages with fusional pronoun morphology, this context-free operation is blocked by the principle of disjunctive ordering (the Elsewhere Condition of Kiparsky (1973)), which states that a more specific operation blocks a more general one in the case where both structural descriptions are met. They further adopt a “realisational” approach to the insertion of pronouns into positions created by the syntax; for example, the English pronoun him is the realisation, or “spell out,” of the feature complex [KP +pronoun, -anaphor, 3rd person, Singular, Masculine, Accusative]. The general “radical pro-drop” rule is the context-free zero-realisation rule (11):

9 Chierchia (1998) formulated the Nominal Mapping Parameter, which distinguishes languages in which bare nouns are able to function as arguments from those in which they are not (more technically, can NP map directly into type <e>?). If yes, then the language has generalized bare arguments (allows bare singulars and plurals), has a generalized classifier system and lacks plural morphology. Chierchia proposes that Chinese has the positive value for this parameter, while English has the negative value. See Ramchand & Svenonius (2008) for a different view.
The Elsewhere Condition will always block this realisation of pronouns in English, since, given their fusional nature, English pronouns always have more complex spell-out rules whose structural descriptions properly include that of (11). But this is not true in every language: in some languages, e.g. Japanese, regular, agglutinating case-markers are added to the pronominal root (*watasi-ga* “I”; *watasi-o* “me”, etc.). Japanese thus has separate spell-out rules for the Case (K) morpheme and for the pronoun, which is a category distinct from KP (probably NP). And here is the central idea of their analysis: because of the non-fusional make-up of pronominal KPs, neither the radical pro-drop realisation of KP nor the specific rules for NP and K are in an “elsewhere” relation. Hence Japanese pronominal KPs are optionally allowed a zero realisation. The analysis leads to the following generalisation: fusional pronouns block radical pro-drop. Neeleman & Szendrői’s analysis entails a clear implicational relation between non-fusional pronoun morphology and discourse pro-drop, which they show holds across a very wide range of languages.

The third analysis of discourse pro-drop put forward recently is due to Saito (2007). Saito (2007) suggests that a single covert grammatical mechanism allows for radical pro-drop and argumentellipsis. This mechanism involves covert copying of elements into argument positions from a set of discourse-given entities. Understood pronouns may be included in this set, along with antecedents for the recovery of elided arguments. The precondition for this covert operation is, effectively, the lack of surface agreement triggers. Hence arguments are not required to be overtly present in
order to trigger surface agreement on verbs and other categories. Thus the relation
between radical pro-drop and absence of agreement-marking is established (see also
Kuroda (1988)). This analysis of discourse pro-drop is discussed in more detail in
Roberts (chapter 1); see also Section 3.4 below.

Whatever the correct analysis, the characteristics of discourse pro-drop languages
which distinguish them from consistent null-subject languages are (i) general
possibility of non-expression/ellipsis of nominal arguments in various functions in
addition to the subject; (ii) lack of person-agreement marking on verbs.

1.2.4 Partial null-subject languages. The existence of partial null-subject languages
as a separate type of null-subject language has been more difficult to establish.
However, Holmberg (2005:548-550), Holmberg and Sheehan (chapter 3) and
Holmberg, Nayudu & Sheehan (to appear) and Holmberg (chapters 2 and 5) identify a
number of characteristics which can serve to distinguish languages of this type from
languages of the Italian type described in Section 1.2.1 (see also the papers in
Holmberg (2008)). These include Finnish, Hebrew, Russian, Icelandic, Marathi and
probably several other Indic languages, and Brazilian Portuguese.  

Here we take Finnish as our example of a partial null-subject language. We observe
three things which distinguish the null subjects of Finnish from those of a language

---

Brazillian Portuguese is an interesting case, as it seems to differ from European Portuguese in
being a partial null-subject language, while European Portuguese is a consistent null-subject language.
Duarte (1995) traces this development in 19th- and 20th-century Brazilian Portuguese. For some
speculation regarding the diachronic development of Brazilian Portuguese in this respect, see Roberts
(forthcoming).
like Italian. First, only 1st and 2nd person pronouns are freely able to be left unexpressed in any finite context:

(12) (Minä) puhun englantia  “I speak English, etc.”
    I      speak-1sg English

    (Sinä) puhut englantia
    You    speak-2sg English

    *(Hän) puhuu englantia
    He/she speak-3sg English

    (Me) puhumme englantia
    We      speak-1pl English

    (Te) puhutte englantia
    You speak-2pl English

    *(He) puhuvat englantia
    They speak-3pl English     (Holmberg’s (2005: 539))

However, it is not the case that 3rd person pronouns can never be unexpressed. As Holmberg (2005: 539) says: “A 3rd person definite subject pronoun can be null when
it is bound by a higher argument, under conditions that are rather poorly understood.”

This possibility is illustrated by examples like the following:

(13) Pekka väittää [ että hän/jØ/puhuu englantia hyvin ].

Pekka claims that he speaks English well

(Holmberg’s (2005: 539))

This is characteristic of partial null-subject languages. Holmberg & Sheehan (chapter 3) is an investigation of the conditions under which these null subjects are licensed.

Finally, in partial null-subject languages “generic pronouns can, and must, be null” (Holmberg’s (2005: 540)):

(14) Täällä ei saa polttaa

Here not may smoke

“One can’t smoke here”

This contrasts with languages like Italian and Greek, where a special clitic (Italian) or verb form (Greek) is required:

(15) a. Qui non si può fumare. [Italian]

Here not SI can smoke

b. Apogherevete to kapnisma. [Greek]
In Italian, omission of the *si*-clitic gives rise to a string comparable to that in the Finnish example in (14), but the unexpressed pronoun must be understood as definite: “Here s/he can’t smoke”. The Greek example features the mediopassive form of the verb “to prohibit”, which gives rise to the same kind of impersonal interpretation as Italian *si*.

The above differences suffice to establish that partial null-subject languages have a range of properties distinguishing them from consistent null-subject languages. It is very likely that many languages that have been seen as consistent null-subject languages are in fact partial null-subject languages; in this connection, the simple descriptive facts need to be clarified. One should regard categorical statements in descriptive grammars to the effect that a given language is a null-subject language with some scepticism (with the possible exception of the very best studied languages). The fact that subjects are sometimes unexpressed does not make a language a null-subject language in the technical sense. As is well known, even (spoken) English can drop its subjects in certain contexts (see Note 6). Distinguishing a consistent from a partial null-subject language requires consideration of contexts such as (7) and the syntax of impersonal constructions at a level of detail which is, probably, rarely attained in the research behind descriptive grammars.

1.2.5 Conclusion. We see that there are four identifiable types of null-subject language. We can range them along a scale of “liberality” as follows:
(16) expletive null subjects ⊃ partial null subjects ⊃ consistent null subjects ⊃ discourse pro-drop

Placing discourse pro-drop languages at the right edge of the scale is motivated particularly if other arguments (direct and indirect objects) are taken into account; discourse pro-drop languages allow (referential) null objects as well as subjects, which is not the case for consistent or partial null-subject languages, across the board. And of course, we could add non-null-subject languages such as English at the left-hand edge of the scale in (16). All other things being equal, for each system $S_i$ in (16), the set of positions in which a pronoun can remain unexpressed in $S_i$ is a proper subset of the set of positions in which a pronoun can remain unexpressed in all systems $S_j$, where $S_i$ is to the left of $S_j$ in (16).\footnote{As it stands, this is an idealisation. There is, in fact, considerable variation among the discourse pro-drop languages as regards the use of pro-drop. Chinese is apparently more restricted in this respect, making more use of overt pronouns, than, for example, Japanese, and possibly more than many consistent null-subject languages: see Cole (forthcoming).} We will return to the possibility of arranging different kinds of systems into some kind of hierarchy in Sections 3.3 and 3.4.

1.3 Two analyses of null subjects

There have been two main views on the nature of null subjects and the NSP in the literature for some time. One view, most influentially put forward in Rizzi (1986), is that null subjects are occurrences of a phonologically unrealised, or empty, pronoun $pro$ in the subject position (see Section 1 of Roberts (chapter 1) and Section 9 of
Holmberg (chapter 2) for more details of Rizzi’s account). The other view, which has its origins in Borer (1986), is that there is no overarching requirement for a subject position as such (i.e. that the Extended Projection Principle (EPP) of Chomsky (1982:10) does not hold, or at least does not hold universally). The null subject may then be directly expressed by the rich verbal agreement inflection, and there is therefore no need for a distinct empty pronoun to realise the subject function. This view implicitly accepts that agreement inflection can function like a pronoun, in that it can bear a grammatical function, and a thematic relation, in the way that nominal expressions typically do. Since agreement is located in the Infl, or I, position, this view is known as the “I-subject” view. This view articulates the intuition behind comments in traditional grammars of the kind exemplified in Section 1.1 above.

The two views just sketched survive in current work. Developing Borer’s (1986) I-subjects idea, it has been suggested by various authors that, since person-number specification of the subject can be exhaustively computed from the verbal inflection, the preverbal subject is effectively optional and when it appears it acts as a left-dislocated (or more precisely “clitic left-dislocated”) element occupying a position peripheral to the clause (i.e. one not associated with a grammatical function) and with the verbal inflection functioning analogously to a clitic pronoun. In differing ways, this view is put forward by Alexiadou & Anagnostopoulou (1998), Barbosa (1995, 2006, to appear), Fassi Fehri (1993), Manzini & Savoia (2005), Nash & Rouveret (1997), Ordoñez (1997), Platzack (2004) and Pollock (1997). Holmberg (chapter 2) articulates what might be viewed as a version of this hypothesis.

On the other hand, Cardinaletti (1997, 2004), Holmberg (2005) and Sheehan (2006) have argued that the subject position is present, at least in some null-subject
languages, and hence it is filled by the null pronoun \textit{pro} (see for example Sheehan (chapter 6 on this). Holmberg (2005) and Roberts (chapter 1) follow Cardinaletti & Starke (1999) in taking \textit{pro} to be a weak pronoun, that is a ‘deficient’ pronoun whose distribution is restricted to certain designated positions. Furthermore, following a long line of work going back at least to Rizzi (1982), they take \textit{pro} to be licensed by a special pronominal feature (usually termed a D-feature) associated with the head bearing the features realised as person-agreement on the verb.

The two views can be taken to converge on the idea that the inflectional head must be pronoun-like in a null-subject language (they may converge in other respects too; see Roberts (chapter 1, §2.5)). This appears to be the core of the null-subject parameter, whatever the further details. Let us encode this property formally as the presence or absence of a D-feature associated with T. Following the yes/no-question format for parameters adopted in Roberts (2007a), let us then state the NSP as follows:

\begin{align*}
(17) \quad \textit{The Null Subject Parameter:} \\
\text{Does } T \text{ bear a D-feature?}
\end{align*}

T clearly does not bear a D-feature in non-null-subject languages, while it clearly does in consistent null-subject languages. Partial null-subject languages and languages only allowing null expletives do not have a D-feature in T, which is why they allow referential null subjects only under very restricted circumstances (partial null-subject languages) or not at all; see Holmberg (chapter 2), Holmberg & Sheehan (chapter 3) and Roberts (chapter 1) for refinements. We leave open for now the status of discourse pro-drop languages, although it is worth noting that if, following the
Nominal Mapping Parameter, non-D elements can function as arguments, perhaps the
requirement for a D-feature on T is waived in such languages. We will explore this a
little more in Section 3.4 below (and see Roberts (chapter 1, §2.5)).

2. The NSP in the context of P&P theory

2.1 Rizzi (1982): clustering properties in Romance and English

The NSP has played a prominent role in the theoretical study of comparative syntax in
recent years, not just because of the characterisation it gives us of languages like
Italian, and how they differ from English, but primarily because it has been seen as a
good example of the way in which rather abstract grammatical properties (such as that
given in (18)) may have proliferating effects, unifying apparently unrelated surface
phenomena.

To see the full importance of this idea, we need to consider Chomsky’s (1964:28f.)
definitions of levels of adequacy for linguistic theory. These were observational,
descriptive and explanatory adequacy. An observationally adequate grammar presents
the data correctly, while a descriptively adequate grammar “specifies the observed
data .. in terms of significant generalizations that express underlying regularities in the
language” (Chomsky (1964:28)). Explanatory adequacy “can be interpreted as
asserting that data of the observed kind will enable a speaker whose intrinsic
capacities are as represented in th[e] general theory to construct for himself a
grammar that characterizes exactly this intuition”; in other words, attaining
explanatory adequacy involves showing how a given empirical phenomenon can be deduced from UG.

The postulation of parametric variation in UG principles was a very large step in the direction of explanatory adequacy, since, one could assume, if we can say that this syntactic feature of this language is due to setting that parameter to that value, we have provided an explanatorily adequate account of the syntactic feature in question in that we have related it directly to UG, as well as a descriptively adequate account to the extent that the analysis of the relevant property of the language is correct. Moreover, the parametric account has immediate cross-linguistic implications, since it implies that another language lacking the property in question will set the parameter in question to a different value. Now, if each parameter value determines a cluster of disparate syntactic features, then explanatory adequacy is enhanced, especially if certain features are readily accessible to acquirers on the basis of impoverished evidence while others are hardly likely to be easily accessible. In this case, arriving at a parameter value determining both the accessible and relatively inaccessible feature gives us a simple account of how the inaccessible feature can be acquired, thus accounting for an aspect of the poverty of the stimulus to language acquisition and thereby, again, reaching explanatory adequacy. At the same time, other things being equal, a “typological” prediction is made: the inaccessible feature will be acquired whenever the acquired one is, since both reflect the same abstract property of Universal Grammar, the setting of a given parameter to a given value.
Let us state the following conjecture in relation the “clustering effect” associated with parameters:\(^\text{12}\)

\( (18) \) A substring of the input text \( S \) expresses a parameter \( P \) just in case a grammar must have \( P \) set to a definite value in order to assign a well-formed representation to \( S \). \( \text{(Roberts (2007a:133))} \)

\( (19) \) For any UG parameter \( P \), fixing \( P \) at value \( v_i \) entails a cluster \( C \) of grammatical expressions of \( P(v_i) \), while setting \( P \) to value \( v_j \neq i \) entails a disjoint cluster \( C' \) of expressions of \( P(v_j) \).

\( (20) \) A cluster of \( P \)-expressions is a set of properties of a surface morphosyntactic form of cardinality equal to or greater than 2, which are reflexes of \( P \)’s setting to a determinate value \( v_i \).

\( (21) \) A substring of the input text \( S \) is a trigger for parameter \( P \) if \( S \) expresses a determinate value \( v_i \) of \( P \). \( \text{(Roberts (2007a:133))} \)

\(^{12}\) (19) is simplified in various ways. First, we are assuming that parameters always have binary values, although of course one could in principle add to (19) a clause specifying how values \( v_i, i+1 \ldots n \) each entail distinct clusters of \( P \)-expressions \( C_1, \ldots C_n \), each of which may be partially disjoint from the others. For simplicity, and following general practice in discussions of parameters, we assume that all parameters are in fact, or can be formulated as if they were, binary. Second, taking \( C' \) to be the negative correlate of \( C \) is clearly the simplest assumption we can make, although not required. We will nevertheless make it.
It follows from these definitions that a trigger for a given parameter value is included in the cluster of expressions of that parameter value, and indeed properly included in that set if triggers must be unique (if there is to be a one-to-one mapping from triggers to parameter values). A trigger must be accessible in the primary linguistic data (PLD), while the other expressions in a given cluster C may be relatively inaccessible, along the lines described above.

The properties connected to the NSP by Perlmutter (1971) and, in particular, Rizzi (1982), can be seen as a cluster. We take the NSP to be a UG parameter, and fixing the NSP at value + entails the cluster C of grammatical expressions of NSP(+) in (22):

(22) a. The possibility of a silent, referential, definite subject of finite clauses.
    b. “Free subject inversion”
    c. The apparent absence of complementiser-trace effects.
    d. Rich agreement inflection on finite verbs.

(22b) refers to the general possibility of expressing an overt subject, usually with a focus interpretation, in postverbal position:

(23) a. Hanno telefonato molti studenti.
    b. *Ont téléphoné beaucoup d’étudiants.

Have telephoned many students.

“Many students have telephoned.”
“Free inversion” is in fact subject to slightly differing constraints in different languages, being more freely available in Spanish and Greek than in Italian, for example; see Sheehan (2006, chapter 6) and the references given there.

(22c) relates to Perlmutter’s generalisation, since it originates in Perlmutter’s (1971) pioneering work. Perlmutter’s generalisation expresses the fact that in non-null-subject languages the subject of a finite clause cannot undergo wh-movement if the complementiser introducing the clause is present. This constraint holds of English and French, as the following examples show:

(24) a. *Who did you say that – wrote this book?  
     b. *Qui as-tu dit qu’ – a écrit ce livre?  

Here the questioned constituent (who/qui) corresponds to the subject of the subordinate clause, so there is a “gap” in that position. The ungrammaticality of (24a) is known as a “complementiser-trace effect”, since in many versions of the theory of movement it is held that the empty subject position at the movement site in the complement clause contains a trace of the moved wh-element. The idea that the presence of the complementiser determines the ungrammaticality of such examples is supported by the fact that (25a) becomes grammatical if that is omitted. In French, (25b) can be rendered grammatical by altering the form of the complementiser from que to qui. These points are illustrated in (26):

(26) a. Who did you say – wrote this book?  
     b. Qui as-tu dit qui – a écrit ce livre?  

(=(24a))
In null-subject languages, as Perlmutter observed, it appears that complementiser-trace effects are not found (Rizzi (1982) argued that in fact this is not true if certain structures covertly derived at the level of Logical Form are taken into consideration). The subject of a finite clause introduced by a complementiser can readily be questioned in these languages:

(27)  
a. Chi hai detto che -- ha scritto questo libro? (Italian)  
Who have-2sg said that -- has written this book  
b. Pjos ipes oti – egrapse afto to vivlio? (Greek)  
who said-2sg that -- wrote this the book  
“Who did you say wrote this book?”

This feature of the null-subject cluster can be reasonably thought of as relatively inaccessible in the PLD, while rich agreement inflection is presumably very accessible (especially given the known sensitivity of acquirers to inflections; see Wexler (1998)), and the other two properties may be somewhat accessible.

In terms of the definitions above, setting the NSP to the value - entails the disjoint cluster C’ of expressions of NSP(-): no null subjects, no free subject inversion, complementiser-trace effects and “poor” agreement inflection. English, French and the Mainland Scandinavian languages have the cluster C’, as do many creoles.

The above discussion relates primarily to language acquisition, and shows how parametric clusters take us towards explanatory adequacy. But it is also clear that this
approach defines language types. In this way, typology, in the sense of the establishment of cross-linguistic relations and of a structure to cross-linguistic variation, and acquisition become intrinsically related. This is a very positive development as it clearly opens the way to a two-pronged empirical approach to understanding the nature of UG.

It is also worth noting that the definitions given in (19-22) provide very strong definitions of types, far stronger than what is usually found in the Greenbergian typological tradition. According to these definitions, all the properties in C are biconditionally related to one another. In other words, all things being equal, we can derive the following set of implicational statements:

(27)  
a. A language has null subjects iff it has rich agreement.  
b. A language has null subjects iff it has free subject inversion.  
c. A language has null subjects iff it does not show complementiser-trace effects.  
d. A language has rich agreement iff it has free subject inversion.  
e. A language has rich agreement iff it does not show complementiser-trace effects.  
f. A language has free subject inversion iff it does not show complementiser-trace effects.

In Rizzi’s (1982) terms, all of these properties are connected by the presence of the silent pronoun pro in the subject position. This element is licensed by rich agreement inflection, and can satisfy the general requirement for a subject position (the Extended
Projection Principle of Chomsky (1982:10)), allowing an overt subject to appear in the “freely inverted” position (see Sheehan chapter 6 for more on this), and indeed to be \(wh\)-moved from this position (as argued by Rizzi (1982)). Thus the formal property which underlies the NSP, on this analysis, is the availability of \(pro\) subjects. Once acquirers deduce this (on the basis of the universal principles determining the availability of null pronouns), they will immediately deduce the existence of the other properties in the cluster, and the very strong implicational links among the properties in (28) follow. Hence we expect strong typological correlations to support parametric clusters, and thereby to motivate analyses of the general type instantiated by Rizzi’s account of the cluster associated with the NSP. One might also, at least initially, think that all languages must fall on one side of the other of a parametric divide that distinguishes cluster NSP+ from cluster NSP-, and that perhaps this is true of all clusters defined by all parameters.

2.2 Gilligan (1987): universal clusters

Building on Rizzi (1982), then, it was possible to take the NSP to define two disjoint language types, in which the four properties in (22) are biconditionally related to one another. Putting aside the partial null-subject language Brazilian Portuguese, the validity of this typology across the Romance looks reasonable: all the Romance languages except French (and some varieties of Rhaeto-Romansch) appear to show the properties in (22).\(^{13}\)

\(^{13}\) See Holmberg & Sheehan (chapter 3), and especially Sheehan (2006: ch. 6) on Brazilian Portuguese.
But of course if this is a valid linguistic typology, it should extend further. Consider, then, the Celtic languages Welsh and Irish. (Literary) Welsh and Irish are both null-subject languages (on Irish, see McCloskey & Hale (1984); on Welsh, see Roberts (2005, Chapter 2), Borsley, Tallerman & Willis (2007:34)). In both systems, the presence of null subjects appears to be fairly closely tied to rich agreement inflection on the finite verb. However, it is very hard to evaluate the status of the correlation with the absence of complementiser-trace effects, since subjects never appear adjacent to finite complementisers in these languages owing to the fact that the basic word order in finite clauses in VSO. This highlights a major difficulty in trying to generalise the correlations in (22): in general, such tightly connected clusters can only hold among languages where a range of further conditions are held constant, and this in practice often means rather closely related languages. Even in the case of closely related languages there may be other, independent differences which disguise the effect of a parameter. Among the Germanic languages English and the Mainland Scandinavian languages show none of the properties of (22) (except for the dialects which allow that-trace; see Sobin (19??), Lohndahl (2007)). But the status of Afrikaans, Dutch, and German is less clear, since their SOV order makes it difficult to determine whether free inversion occurs. As for Icelandic, it shares with Brazilian Portuguese the property of being a partial null-subject language, which therefore does not conform to the classical NSP (see Holmberg, chapter 2).

Simple, surface evaluation of correlations of the kind in (22) is not possible, without making further assumptions. In this respect, such correlations differ from Greenbergian correlations such as “If a language is OV, then it has postpositions” in that more assumptions are required in order to test them. This, of course, is connected
to the fact that the correlations arose from a deeper syntactic analysis of the languages in question than that underlying the simpler Greenberg-style correlations.

Our intention is not to criticise either Rizzi’s or Greenberg’s approach to establishing typological correlations; both may be able to reveal previously unsuspected cross-linguistic correlations. The point is simply that correlations of the type in (22) may be very difficult to establish and maintain as more and more extra variables are brought into play. For example, what do we conclude from the fact that Welsh and Irish are null-subject languages in which the presence of complementiser-trace effects cannot be determined in finite clauses owing to the existence of VSO order? It seems reasonable to conclude that VSO order neutralises this property, and so here we do not have a true counterexample to the proposed cluster, but at the same time we do not have the strong biconditional relation of the type in (27). Such biconditional statements must be subject to a general qualification along the lines of “all other things being equal”.

Another potential counterexample is posed by those dialects of Mainland Scandinavian and English which do not show the that-trace effect in finite clauses (cf. Sobin (19??), Lohndahl (2007)).

(28) Vem tror du att _ är intelligentast? (Fenno-Swedish)

who think you that is most intelligent

‘Who do you think is most intelligent?’

Again, such facts indicate that the correlation between the that-trace effect and the other properties of (28) is not biconditional. Null subjects and free inversion entail the
absence of *that*-trace effects, but not vice versa. This is not necessarily an argument against Rizzi’s theory. Note that the theory has an explanation of the correlation of absence of *that*-trace effect with the other properties, which is that the presence of *pro* in SpecTP makes it possible for A’-movement to by-pass this position, moving directly from a lower subject position to Spec,CP. In recent work, Rizzi argues that spec,IP is a ‘criterial position’ in his terms, which means that movement out of this position is universally impossible; cf. Rizzi & Shlonsky (2005). This creates a problem for subject extraction which every language has to deal with somehow. In null-subject languages the problem is solved quite simply by by-passing Spec,TP. English and French solve it by operating on C, which in English is morphologically realised as deletion of *that*, in French as substitution of *qui* for *que*, thus modifying the criterial property of spec,IP (see Rizzi & Shlonsky for details). Fenno-Swedish, Ozark English, etc. have to solve this problem, too. The most likely hypothesis is that these varieties do not solve it in the manner of the null-subject languages, but perhaps in a manner akin to that in other varieties of Mainland Scandinavian and English, by operating on C, but without a morphological effect on the complementiser. A further alternative may be that the EPP does not apply at SpecTP here, allowing that position to be bypassed by “long” wh-movement. The point here is that complementiser-trace violations are not predicted not to occur in non-null-subject languages. Again, we see that the correlations are not accurately formulated as biconditionals.

Returning to the Celtic languages, it is difficult to be sure of the status of free inversion in these languages. One might claim that the position occupied by the subject in VSO order is in fact that of a freely-inverted subject, although the fact that the subject position is interpolated between an auxiliary and non-finite verb in compound tenses in VSO languages, while in cases of free inversion the subject
follows both the auxiliary and the non-finite verb (see (24)), argues against this. Nonetheless, we could tentatively conclude for Celtic that complementiser-trace effects cannot be tested owing to effect of VSO order. In both cases, the very strict nature of the correlations is part of the problem. Again, if the correlations were stated as one-way implications, or were hedged by a statement such as “all else, particularly basic word order, being equal”, the situation would be clearer in that the status of these languages as true counterexamples would be easier to determine.

The Celtic languages, although typologically quite different in a number of superficial respects from Germanic and Romance (see Haspelmath (2001)), are nonetheless Indo-European languages and have been in contact with some Germanic and Romance languages for millennia. If we want to really establish the universal coverage of a cluster like (22) and the typology it implies, we should of course move far beyond merely comparing Germanic, Romance and Celtic. However, we can only expect that the kind of difficulty involved in interpreting the data, sketched above in relation to Celtic, will multiply, perhaps out of control. Again, this stems from the degree of syntactic analysis needed to establish the relevant properties combined with the very strict nature of the clustering we have postulated.

Gilligan (1987) tested the correlations put forward by Rizzi (1982) against a 100-language sample, which he attempted to correct for areal and genetic bias. As reported in Newmeyer (2004:202-6; 2005:88-92) and Croft (2003:80-84), Gilligan found just four robust cross-linguistic correlations, taking the properties discussed by Rizzi pairwise:

\[(29) \quad \text{a. Free Inversion} \rightarrow \text{expletive null subjects}\]
b. Free Inversion $\rightarrow$ allow complementiser-trace violations

c. Referential null subjects $\rightarrow$ expletive null subjects

d. Allow complementiser-trace violations $\rightarrow$ expletive null subjects

Newmeyer (2005:90-1) concludes that “[t]hese results are not very heartening, .. In three of the four correlations, null non-thematic subjects are entailed, but that is obviously a simple consequence of the virtual non-existence of languages that manifest overt non-thematic subjects.” In a sense, this situation is exactly what we would expect, given the above considerations: expanding the database from roughly ten to roughly one hundred languages simply multiplies the number of uncontrolled variables to a point where, without further detailed analysis of a wide range of constructions in a wide range of languages, the correlations can no longer be discerned. In a word, then, these results are inconclusive (this is not at all the conclusion Newmeyer draws, as we will see in the next section).

However, Gilligan’s results are not quite as inconclusive as they first appear. Consider the only implication which does not involve null expletive pronouns: (30b). This relates to one of the most important conjectures in Rizzi’s (1982) analysis: that true complementiser-trace violations are in fact universally ruled out (due to the ‘criterial’ nature of spec,IP, according to Rizzi & Shlonsky (2005)), and null-subject languages can circumvent them owing to the availability of a different position from which movement can take place in finite clauses introduced by C, the “freely inverted” subject position, as discussed above.\(^{14}\) This claim has very clear explanatory force in

\(^{14}\) On this basis, we might have expected (29b) to go the other way, stating a one-way implicational relation between complementiser-trace violations and free inversion. But as discussed
relation to the poverty of the stimulus: the acquirer encountering the relatively accessible phenomenon of free inversion in the PLD will thereby “acquire” the possibility of the complementiser-trace violations, an otherwise fairly inaccessible aspect of the PLD.

Moreover, we can in fact combine the implications in (30) to give a modest implicational scale along the following lines:

(30) Free Inversion $\rightarrow$ (allow that-trace violations $\rightarrow$ expletive null subjects).

This scale defines three types of language: Type I has all three properties; Type II allows complementiser-trace violations and hence allows expletive null subjects, and Type III only allows expletive null subjects (here “allowing complementiser-trace violations” is intended to mean that no manipulation of C is required for long wh-movement of the lower subject to be possible, hence the varieties of English and Mainland Scandinavian discussed above are not relevant here). It is worth noting that we have weakened the original clustering claim to a series of one-way implicational statements; in other words, the claims in (30) are weaker and therefore easier to support empirically than those in (28). As we have now observed several times, this seems desirable for establishing parameter-based typologies. This seems to be due to the intricate nature of parameter interactions.

above, languages have different strategies for avoiding complementiser-trace violations, including an operation on C, as in English and French. On the other hand, (29b) makes the highly substantive claim that any language with free inversion will allow complementiser-trace violations.
In his study of the null-subject parameter in creoles, Nicolis (2005, 2008) observes that Cape Verdean, Berbice Dutch, Kriyol, Mauritian, Papiamentu and Saramaccan all have expletive null subjects and tolerate that-trace violations, but do not allow referential null subjects (Nicolis (2008:x)); while Haitian and basilectal Jamaican have expletive null subjects but do not tolerate that-trace violations. No creole has free inversion. So we have the following picture:\textsuperscript{15}

(31) Type I: Italian, Spanish, Greek, etc.
Type II: Cape Verdean, Berbice Dutch, Kriyol, Mauritian, Papiamentu, Saramaccan
Type III: Haitian, basilectal Jamaican

So we see that a very strict, theory-driven typology such as that which emerges from Rizzi (1982) makes very strong predictions which almost immediately become very difficult to evaluate as soon as the cross-linguistic database is extended only modestly. Unsurprisingly, it becomes impossible to evaluate once further languages are taken to consideration, as shown by Gilligan’s survey. When a very large number of genetically and typologically highly diverse languages are compared for the “same” properties, with no control as to the other typological features of these languages, the original correlations were shown not to hold in their original form, although four implicational statements could still be gleaned. To us, this does not seem like a bad or shocking result for parametric theory, but rather a fairly promising result from the admixture of a very large amount of essentially random data into an

\textsuperscript{15} See Roberts (2007b:31ff.) on the possibility of including referential null subjects in the scale in (30).
originally carefully controlled database. The fact that any coherent patterns survived is telling, and a sign that Rizzi’s observations were clearly on the right track.

We conclude that the claim that results such as Gilligan’s invalidate the particular parametric cluster proposed is not warranted since, first, we simply do not know enough about dozens of problematic languages in order to be sure whether they are genuinely counterexamples; second, the most intriguing, non-obvious and explanatory conclusion of Rizzi (1982) remains (namely, the implication in (29b)); third, a more modest implicational hierarchy can be constructed on the basis of Gilligan’s results, one which shows that Rizzi’s cluster has some significance for language typology. In the next section, we review the consequences of drawing the opposite conclusion from this one.


We have seen that the kind of clustering of properties predicted by a parameter like the NSP can be construed as making very strong typological predictions, but the exact evaluation of these predictions may in practice be rather difficult. At every step, possible counterexamples and difficult cases proliferate, and, although in our judgement Rizzi’s original proposals certainly have a valid crosslinguistic core, it is of course always possible that a given proposal is wrong. This alone would not, of course, invalidate the approach.

In earlier work (Roberts & Holmberg (2005)), aside from defending parametric theory in general against a range of specific charges made against it in Newmeyer (2004), we
suggested a different example of parametric clustering. This concerns an abstract feature of Agr, or T, which accounts for a range of differences between the Insular Scandinavian languages (ISc) and the mainland ones (MSc) involving the possibility of null nonreferential subjects, non-nominative subjects, stylistic fronting, V-to-T in embedded clauses and relatively rich subject verb agreement; see Holmberg (chapter 2). Newmeyer’s (2006) response to this is that the parametric cluster in question lacks sufficient cross-linguistic justification:

For the facts in (1a-e)[i.e. the correlating syntactic properties just listed, AH/IR] to support parametric theory, it would be necessary to demonstrate that in language after language the constellation of properties particular to ISC and MSC reappear. But aside from some brief remarks about Middle English, Yiddish, and Old French, two of which, like Scandinavian, are Germanic, and the third of which was heavily influenced by Germanic, they ignore the typological dimension entirely. (Newmeyer (2006:3))

Here it becomes clear that Newmeyer sees it as a requirement that, in order for any observed, or predicted, cluster of properties “to support parametric theory”, what he refers to as “the typological dimension” must be taken into account. As becomes clear from the following discussion of each of the surface properties we proposed to be connected to a single formal grammatical property, what he calls the “typological dimension” means bringing in as much unanalysed data from as many unrelated languages as possible. For the reasons given in the previous section, the likelihood is that this approach will only obscure any patterns that may be observable in more controlled data samples (although, as we have observed, any partial or weakened
generalisations will be all the more striking, as is the case for (28b)). But for Newmeyer only the strongest possible justification can satisfy the “typological dimension”: every language must fall on one side or the other of the proposed parametric divide (i.e. every parameter must be shown to define a cluster corresponding to the value \( P^+ \), and the complementary cluster \( C' = P^- \) in terms of (18-21) above), and this must be ascertainable on the basis of a superficial survey of the relevant properties. If not, the parameter lacks empirical support. But almost no known distinction among languages will meet this desideratum for empirical support. For example, the correlation between VO vs OV order and pre- vs. postpositional order has 141 languages defined as not falling into one of the four types, and 48 straight counterexamples, according to Dryer (2005:386). Taking Newmeyer’s “typological” stricture seriously would presumably cast doubt on this correlation, too. More generally, it will simply result in the abandonment of any attempt to establish cross-linguistic generalisations.

Newmeyer (2004, 2005, 2006) concludes from his discussion of the NSP, and equally superficial discussion of a range of other proposed parameters, that parameter-based approaches to cross-linguistic variation “have failed to live up to their promise” (2005:181) and that “the hopeful vision of UG as providing a small number of principles each admitting of a small number of parameter settings is simply not workable” (2005:185). Instead, he suggests that “language-particular differences are captured by differences in language-particular rules” (2005:183). The nature of these “rules” is never really clarified, and so the last claim is hard to evaluate. One case he discusses, though, is that of the head-complement parameter, which, according to Newmeyer (2004, 2005: 74) should be replaced by two rules: (a) Complements are to
the right of the head, and (b) Complements are to the left of the head. Some languages apply (a), other languages (b). In Roberts & Holmberg (2005) we pointed out that this is equivalent to a parameter in the sense of P&P theory: UG leaves two options open, and acquisition of head-complement order is a matter of choosing between the two options on the basis of experience. Newmeyer (2006) agrees, but insists that if a set of grammatical facts can be characterized as a rule, then it should be, rather than being characterized as the value of a parameter. 16

However, if rules are, at least in certain cases, empirically equivalent to parameters, we can conclude that, in his view, they will fare just as badly, by his criteria, in accounting for cross-linguistic variation. At the very best, the proposed rules are as inadequate as he proposes parameters are. At worst, they may be much worse: the only construal of “rule” in the literature on generative syntax refers to phrase-structure rules and transformational rules. Such rules were thought of as largely language-particular, language acquisition consisting of the selection among the infinite class of such rules compatible with the PLD on the basis of an evaluation metric. If this is what Newmeyer is advocating, it is clearly a retrograde step, and undoubtedly represents a retreat from explanatory adequacy (the same observation is made in Dryer’s (2008) review of Newmeyer (2005): cf. Dryer’s (245-6) remark that “Newmeyer argues for .. a retreat to a version of C[homskyan]G[enerative]T[heory] from the period of Chomsky (1973) up to, but not including Chomsky (1981)”). If it is

16 A possible reason behind Newmeyer’s opposition to characterizing the choice between head-precedes-complement and complement-precedes-head as two values of a parameter is that, in this particular case, the values are not given by properties of UG, but rather follow from the ultimately physical fact that words must be linearly ordered, allowing exactly those two options. We return briefly to this point in §3.6.
not what he is advocating, he owes us an account of his notion of rule. Even then, if he retains the view that rules are largely equivalent to parameters, there is no reason to adopt his approach over a parameter-based one, since, as he has argued, the parameter-based approach is inadequate. What must be shown is that the notion of rule, as opposed to parameter, is both empirically superior to the notion of parameter (i.e. more descriptively adequate) and can take us closer to understanding UG, i.e. to explanatory adequacy, than the notion of parameter. But as far as we can tell, Newmeyer merely gives up the pursuit of explanatory adequacy along with the notion of parameter. For this reason, we cannot accept his conclusions.

Implicit in the above discussion has been a distinction between two approaches to universals, or to cross-linguistic correlations of any kind. These are, roughly speaking, the Chomskyan approach and the Greenbergian approach (see also Rizzi (1994)). The Chomskyan approach is based on the pursuit of explanatory adequacy. Hence the central idea is that a given piece of grammatical knowledge hardly accessible through experience given normal assumptions about the nature of the PLD can be acquired “for free” given the existence of a correlating, more accessible piece of evidence. Rizzi’s discussion of the NSP meets this desideratum, and the survival of implication (29b) under Gilligan’s empirical scrutiny is highly significant in this connection. The other approach involves the observation of properties which covary on the surface, without prejudice as to any deeper correlations or poverty-of-the-stimulus considerations. The empirical scope of the generalisations unearthed in this tradition is impressive (see in particular WALS, Haspelmath, Dryer, Gil & Comrie (2005)), and seems to set an important research agenda, and a series of empirical challenges, to
generative approaches committed to more abstract analyses and, in particular, to achieving explanatory adequacy.

As Gianollo, Guardiano & Longobardi (henceforth GGL, 2008:x) point out, both approaches to typology have flaws, particularly if they are seen as heuristics for the accumulation of information regarding which properties (co-)vary and which do not. As they point out, the Chomsky-Rizzi approach, in which poverty-of-stimulus considerations remain paramount, “provides deep and often correct grammatical insights, but, alone, it may lead little further than the study of single parameters in a few languages” (ibid). We have seen above the reasons why this may be so. The Greenbergian approach, they say, is subject to two objections. First, “it is practically hardly usable for the relevant purposes, because of the depth of grammatical investigation required by sound parametric analysis” (again, we have made this point in the foregoing). Second, and perhaps more challengingly, “it is anyway likely to be insufficient if investigation is not guided by strong abductive and theory-oriented considerations, even if hypothetically extended to all languages, because the cardinality of the universal parameter set is so wide as to generate a number of possible languages of which the actual existing or known ones represent an infinitesimal sample” (ibid). In other words, typological observations, however well-supported by data from the currently available set of languages, may nevertheless represent accidental correlations coexisting contingently at this historical moment, and not reflect the true nature of a UG which generates a far larger set of languages than those currently extant. This second point, in fact, undermines Newmeyer’s objection to parameters such as the one put forward by Holmberg & Roberts (2005), whose effects are most visible in Germanic languages, on the grounds that “it leaves
open the possibility of historical accident (language contact, drift due to some non-parametric cause)” (Newmeyer (2006:5)); so does the Greenbergian approach, only on a wider historical and geographical scale. GGL go on to advocate a very interesting kind of “halfway” approach, based on their notion of Modularised Global Parametrisation, which they argue is largely immune to both types of objection.

Whatever the merits of GGL’s approach, we suggest that the fundamental error in Newmeyer’s critique of parameters is that he conflates the two original approaches. He therefore judges parameters proposed in the Chomskyan tradition, and which have something to offer from the point of view of explanatory adequacy, by the extensional, taxonomic, surface-oriented criteria of the Greenbergian approach. This inevitably leads to his highly negative assessment of work such as Rizzi’s on the NSP. Essentially, any parameter would have to show clustering effects which exhaustively partition all languages into two classes, with this partitioning immediately accessible to superficial scrutiny in all cases, i.e. with no masking effects due to extraneous properties or extensional overlaps in coverage. Ideally, some of the clustering properties would be relatively inaccessible in the PLD, giving the parameter explanatory depth. Of course, no proposed parameter (or Greenbergian universal) meets such stringent requirements. In the case of the NSP, for example, any and every kind of phonologically non-realised subject ought to fall under its purview, and this mistake underlies much of the criticism of the nature of the parameters and the alleged failure of the correlations. But it has been known since the earliest work that there are at least two, if not three, different kinds of null-subject systems: the Italian-style “consistent” system, those allowing only non-referential null subjects and the East Asian type of “discourse pro-drop” (more recently, “partial” systems have been added
to this listy, as we have seen). Newmeyer arrives at his position because his critique of the parametric approach conflates and confuses the aims and methods of two distinct traditions. Given this, it is no surprise that he advocates a retreat from the goal of explanatory adequacy. And it is fundamentally for this reason that we cannot accept his conclusions.

2.4 A performance-efficiency based alternative

So, after setting the bar for descriptive adequacy for any proposed parameter impossibly high with his allusion to the “typological dimension,” Newmeyer proposes a retreat from explanatory adequacy in the form of a return to the rule-based systems of the 1960s and 1970s, thereby dissolving the link between acquisition and typology discussed in §2.1 above. Nonetheless, he does believe that a number of typological observations (mostly made in the Greenbergian tradition) reflect genuine cross-linguistic generalisations. He suggests, however, that these are amenable to “performance explanations” of the type advocated by Hawkins (1994, 2004).

Hawkins’ central idea is the Performance-Grammar Correspondence Hypothesis (PGCH), which we state as follows:

(32) Grammars have conventionalized syntactic structures in proportion to their degree of preference in performance, as evidenced by patterns of selection in corpora and ease of processing in psycholinguistic experiments.

Preferences are the reflection of efficiency: “speakers attempt to increase efficiency by reducing structural complexity” (Newmeyer (2005:122-3)). Hawkins proposes three efficiency principles, one of which is Minimize Domains (MiD). This efficiency principle is said to be able to account for the fact that OV and postpositional orders tend to cooccur and VO and prepositional orders tend to cooccur, but OV tends not to cooccur with PO and VO tends not to cooccur with OP. The idea is that in the consistently “head-initial” (VO and PO) and “head-final” (OV and OP) orders, the “distance” from V to P or P to V is shorter than in the inconsistent orders. The structures in (34) and (35), adapted from Newmeyer (2005:124), illustrate:

\[(33)\]

a. \[
[VP \ V \ NP \ [PP \ P \ NP ]]
\]

\[\ldots\]

\[\text{VO and PO} \]

\[\text{(not far from V to P)}\]

b. \[
[VP \ [PP \ NP \ P ] \ NP \ V ]
\]

\[\ldots\]

\[\text{OV and OP} \]

\[\text{(not far from P to V)}\]

\[(34)\]

a. \[
[VP \ V \ NP \ [PP \ NP \ P ]]
\]

\[\ldots\]

\[\text{VO and OP} \]

\[\text{(too far from V to P)}\]

b. \[
[VP \ [PP \ P \ NP] \ NP \ V ]
\]

\[\ldots\]

\[\text{OV and PO} \]

\[\text{(too far from P to V)}\]

According to these schematised structures, in the frequently occurring orders of (34), only one NP intervenes between the two heads V and P, favouring these structures on the grounds that they are easy to process. The rarer orders in (35) are disfavoured.
because two NPs intervene. Newmeyer concludes that MiD and the similar efficiency
principles can do the work that parameters have been proposed to do, only better.

Although there are a number of obvious objections that can be raised against this
rather simplified account,\textsuperscript{17} we will concentrate our discussion here on the conceptual

\textsuperscript{17}These include: (i) why are all verbs with prepositional complements assumed to also have a
direct object? English and most other languages have verbs like \textit{rely}, \textit{hope}, \textit{depend}, etc, which can take
PP but not NP complements. Do we therefore expect a different typological generalisation for this type
of verb? If not, why not? (ii) why is V’s PP complement assumed to pattern with its NP complement
such that both always systematically precede/follow the verb? In other words, why are orders such as
the following not considered?

\begin{itemize}
\item[(i)] $[\text{VP NP} \ V \ [\text{PP P NP}]]$ (OV and PO)
\item[(ii)] $[\text{VP} \ [\text{PP NP P} ] \text{V NP}]$ (VO and OP)
\end{itemize}

Here the “distance” – however that is computed – between V and P is smaller than in the cross-
linguistically common cases in (34). Therefore, all other things being equal, we expect to find these
more frequently than (34), which is not the case. “Mixed” orderings of complements like these do
exist: depending on one’s analysis of elements such as Mandarin and other Chinese varieties may have
$[\text{VP} \ [\text{PP P NP} ] \text{V NP}]$ (see Li’s (1990) analysis of the \textit{ba}-construction), which is predicted to be as
common as (34) by MiD, but is cross-linguistically rare. German and Dutch allow the order in (i) in
subordinate clauses with “PP-extraposition”; again, this pattern is predicted to be more frequent than
(34) and (35), which is very probably not the case. These objections concern the presence of an “extra”
NP in the representations given by Newmeyer and schematised in (34). Hawkins (1990:238-9) leaves
the extraneous NP out, and makes essentially the same argument. Here, too, some objections can be
made. First, what about adjunct PPs, which presumably occur in a different configuration from those in
(34) and (35); do we predict a different typological generalisation for these? Second, Dryer’s
(2005:386) figures show that (35b) is four times as rare as (35a). Why? Third, Dryer states that in 141
of the 1,033 languages he surveys the word order cannot be determined. It may be that some of these
languages have “free” word order on the surface. For Newmeyer/Hawkins, such orders, if truly free,
would presumably be the hardest of all to process, since there is no consistent “distance” in the
issues. The key point, we submit, is the notion of “efficiency”. This is left implicit in Newmeyer’s account, although some general notion of “ease” of processing is clearly intended. Since we are dealing with a cognitive capacity, processing as an aspect of performance, it is reasonable to think of efficiency in terms of reduction in computational effort. If that is so, then, following Mobbs (2008), we suspect that there may indeed be a case to made for viewing typological skewings as the consequence of general computational principles. Rather than applying at the level of performance and processing, we suspect, like Mobbs, that they may be more deeply implicated in the organisation of the language faculty and the learning device. We will return to these points in §3.5 below, and look a little more at the other efficiency principles proposed by Hawkins, in addition to MiD.

2.5 Conclusion

Here we have described and illustrated the idea of parametric clustering in relation to Rizzi’s classic (1982) work on the NSP. We have observed the difficulties in

“domain” (or perhaps the easiest, for the same reason). But actually they are neither the rarest type, being much commoner than the inconsistent types in (35), nor particularly common, being much rarer than the consistent types. Of course, it would be possible to assume an underlying fixed order as in generative work, but such a move will make a processing account hard to maintain and raise the possibility of movement operations perturbing underlying order in the other cases. Whichever way things are construed, then, more than 10% of the languages in Dryer’s sample pose an insuperable problem for Hawkins’ approach. Newmeyer’s strictures regarding the “typological dimension”, so ruthlessly applied to generative work, mean that he cannot simply appeal to analyses that have not yet been carried out. Jumping the gun somewhat, then, we find Newmeyer’s preferred alternative as unconvincing as his arguments against the principles and parameters approach.
straightforwardly generalising the predictions without further analysis and/or taking into account interfering factors (such as VSO order in the Celtic languages). Newmeyer’s critique of the NSP and of parameter-based approaches to cross-linguistic work completely overlooks this point and conflates the two distinct approaches to comparative syntax that have emerged in recent years. This leads to him setting the bar for even descriptive adequacy for the postulation of parameters impossibly high, with the logical but regrettable result that he retreats entirely from the goal of explanatory adequacy in comparative syntax. We briefly illustrated his preferred approach, based on Hawkins’ Performance-Grammar Correspondence Hypothesis (PGCH) and associated efficiency principles, which we return to in §3.5. By and large, though, we reject Newmeyer’s conclusions regarding the principles and parameters approach to comparative syntax. At the same time, we recognise that the kind of clustering effect discussed in §2.1 and 2.2 is of limited typological interest, and that extending the typological purview of parameters like the traditional NSP may be problematic (although the brief discussion in §2.2 indicates that weaker clusterings and correlations can be envisaged). What is needed, as also pointed out by Biberauer (2008b), is a more systematic basis for extending the typological domain in which theoretically inspired parameters may be more readily evaluated. We now turn to this point.

3. **Parametric Theory**

3.1 **Problems with the “classical” formulation**
Although we believe that aspects of Newmeyer’s critique of parameter-based comparative syntax are seriously misguided, and, for the reasons given above, we cannot accept his conclusions, it is nonetheless true that there are problems with the original conception of parameters. In our view, this warrants a refinement of the idea, rather than its abandonment (see also Biberauer (2008b)).

One valid criticism is raised by Newmeyer (2005:83). A consequence of the large amount of comparative work that has been done since the 1980s is that there has been a proliferation of parameters as descriptive devices. This has been particularly apparent in the “microparametric” work on closely related languages and dialects, typified by the papers collected in Kayne (2000, 2005). Given this proliferation of parameters, a natural question which arises is, quite simply, how many parameters there are. It is very likely that the number of parameters is in the hundreds, and at least possible that it is the thousands. For example, Roberts (2007) discusses five well-known parameters (null subjects, V-to-T, T-to-C, negative concord, wh-movement). The “head parameter” is arguably non-unitary, and breaks up into several sub-parameters, probably at least ten. There is also a parameter governing subject-raising to SpecTP (whose negative value, along with a positive value for V-to-T raising, gives rise to VSO orders), at least four parameters governing auxiliary selection (and this does not take into account the impressive microvariation found in Central-Southern Italo-Romance), several parameters governing ergativity (to account for varieties of split ergativity), presumably pertaining to the feature content of v, a parameter concerning the ability of C to Agree for Case with the subject of the TP it introduces (related to “Exceptional Case-marking”), a parameter determining the availability of the option of preposition-stranding and a parameter concerning whether
wh-expressions are DPs. This brings the total to 24. Adding Polysynthesis, Subject Side and Serial Verbs from Baker (2001), the total comes to 27. Baker (2008b) proposes two macroparameters governing agreement (Direction of Agreement and Case-Dependency of Agreement). Furthermore, the null-subject parameter may break up into at least three parameters (see Holmberg chapter 2). Finally, Longobardi & Guardiano (2008) propose 51 parameters which affect DP-internal syntax only. This brings the number of parameters up to 80, and, there is little doubt many more than this would be needed just to reach descriptive adequacy.

Newmeyer concludes that “we are not yet at the point of being able to ‘prove’ that the child is not equipped with 7,846 .. parameters, each of whose settings is fixed by some relevant triggering experience. I would put my money, however, on the fact that evolution has not endowed human beings in such an exuberant fashion” (2005: 83). Although, as Newmeyer implicitly admits, this is only a plausibility argument, we agree with him. It seems highly implausible that UG should specify detailed microparameters governing the nature of clitic systems or agreement systems (or classifier systems or tone systems) when so many languages lack such systems entirely. Clearly, what is needed is some structure to parameter systems, at the very least along the lines of specifying “if L has a clitic/agreement/tone/classifier system, then what particular kind of system does L have?”, where the consequent may break down into a further series of implicational choices.

A related point arises from the discussion above of the correlations connected to the NSP in (22) and (26). All other things being equal, this type of clustering, deriving from a parameter closely linked to a UG principle (such as, in this case, licensing of a
particular empty category), does yield a series of biconditional relations determining a range of properties that are predicted to be either all present or all absent in any language. We saw above, very briefly, the kinds of analytic difficulty that approach can entail. What is preferable, in fact, is weaker, one-way implications of the kind in (29), or the kind typically put forward in the Greenbergian tradition. But these, too, require a rather different approach to the nature of parameters than the classical one illustrated by Rizzi (1982).

Arguably the real issue underlying both of these points is the familiar tension between descriptive and explanatory adequacy. Parameters have in recent years to an extent shared the fate of 1960s-style transformations. They are very powerful formal devices that make possible, for the first time ever, the precise, theory-internal description of cross-linguistic relations (and, correspondingly, descriptions of what children must be able to acquire). However, if over-exploited, and especially in the absence of any general restrictions on their form and functioning, these devices become mere facilitators of taxonomies. Newmeyer quite correctly observes that the very large number of parameters that we seem to need, just to approach cross-linguistic descriptive adequacy, reduces their efficacy as explanatory devices. What is required, as took place with the theory of transformations from late 1960s onwards, is a theory of parameters which places substantive restrictions on their form and function while maintaining their descriptive power.18

18 A similar point is made in connection with diachronic linguistics by Roberts (2001). GGL (2008: 6) observe a tension between explanatory and what they term “evolutionary” adequacy: “once parameters are included by a theory of UG, the minimization of the genetic endowment (produced by the supposed economy constraint on stored innate knowledge) should probably amount to minimizing
3.2 The minimalist view

In the context of the minimalist program for linguistic theory, as it has been pursued in relation to syntax for almost two decades, a slightly different view of parameters has been widely accepted. This can be thought of, following Baker (2008a:3, 2008b:155f.), as the “Borer-Chomsky conjecture,” or BCC:

(35) All parameters of variation are attributable to differences in the features of particular items (e.g. the functional heads) in the Lexicon.

More precisely, we can restrict parameters of variation to a particular class of features, namely formal features in the sense of Chomsky (1995) (Case, φ and categorial features) or, perhaps still more strongly, to attraction/repulsion features (EPP features, Edge Features, etc.). This view has a number of advantages, especially as compared with the earlier view that parameters were points of variation associated with UG principles, and arguably takes us a little way towards resolving the tension between descriptive and explanatory adequacy at the parametric level as described in the previous section. Let us see why this is so, particularly in relation to the latter point.

First, the BCC imposes a strong limit on what can vary. In most versions of syntactic theory informed by minimalist ideas there are a small number of extremely general

the number of parameters as well as resulting, if anything, in a reduction, rather than in the observable extension, of the space of variation.”
principles: Merge, Agree, Select, etc. It is highly unlikely that any of these are subject to parametric variation (pace Baker (2008a)). For example, it would be strange to propose a parameter determining whether or not Merge has to be binary, or whether or not internal Merge (i.e. Move) is possible (Kayne (2008b:1) suggests that internal merge may be a defining feature of human language, i.e. the “narrow language faculty” in the sense of Hauser, Chomsky & Fitch (2002)). Similarly, although languages may vary a fair amount as to which features may be subject to Agree relations, the nature of the operation itself (the locality restrictions operative, the nature of Match, the definition of an active Probe, etc.) seems unlikely to vary. Of course, we have no proof that these kinds of variation do not exist, but they seem unlikely, have not – to our knowledge – been postulated, and would be excluded by the BCC.

Second, as has often been pointed out (initially by Borer (1984:29)), an advantage of the BCC is that associating parameter values with lexical entries reduces them to the one part of a language which clearly must be learned anyway. Ultimately, on this view, parametric variation reduces to the fact that different languages have different lexica, in that sound-meaning pairs vary arbitrarily: the most fundamental and inescapable dimension of cross-linguistic variation. The child acquires the values of the parameters valid for its native language as it acquires the vocabulary (more precisely, as it acquires the formal features associated with the functional categories of its native language).

A third advantage of the BCC is that it imposes a restriction on the form of parameters. The attraction property of a functional head, which may be responsible for fairly major aspects of word-order variation, can be formulated in terms of a
simple diacritic associated with a given functional head, or a given feature of a functional head. For example, the matrix C of a V2 language may be associated with a movement-triggering diacritic, causing an XP to move into its specifier position; in a “residual V2” language like English or French, the diacritic would be associated with a [wh]-feature of C, giving rise to movement of a wh-XP into SpecCP exactly where C has the [wh]-feature. In a language like Chinese, where there is no movement to SpecCP at all, there would be no such feature associated with C. We will give a more general formulation of possible parametric variation, consistent with the BCC, in §3.4 below. Clearly, if parametric variation can be restricted to formulations of this type, this would be a step in the direction of explanation.

This simplicity of formulation in turn makes possible a statement of parametric variation at the UG level which relies on the logic of underspecification. It may, for example, be possible to state that parameter P has value \( v_i \), perhaps a movement-triggering/attraction property, when this is stated as such, and \( v_j \), the absence of this property, elsewhere. This in turn raises the prospect of applying disjunctive ordering to parameter setting, and thereby the setting up of markedness relations. For the moment, we merely note that the simplicity of the formulation of parameters, given the BCC and the typical minimalist conception of parameter setting, allows this. In subsequent sections we will explore the possible implications of this idea further.

A further point, also made by Roberts (2001) and Roberts & Roussou (2003), is that the BCC, combined with a simple restriction of parameters to formal features of functional heads, allows us in principle to calculate the upper bound to the set of grammars. Suppose we have two potential parameter values per formal feature (i.e,
each feature offers a binary parametric choice), then for \( n = |F| \), the cardinality of the set of formal features, the cardinality of the set of parameters \( |P| \) is \( 2n \) and the cardinality of the set of grammatical systems \( |G| \) is \( 2^{2n} \). For (rather implausible) illustration, assume 15 formal features, then we have \( n = 15 \), then \( |P| = 30 \) and \( |G| = 2^{30} \), or 1,073,741,824. One further parametrisable formal feature raises \( |G| \) to \( 2^{32} \), and so on. Since it is likely that UG makes available more than 15 formal parametrised features, the upper bound to the number of possible grammatical systems is likely to be considerably greater than just over a billion, as in the example just given. Of course these are upper bounds; what is not taken account of is interaction among parameters so as to rule out possible combinations of values; as GGL (2008:x) show, such interactions are likely to be pervasive. The lower bound is therefore very likely to be a great deal lower than the kinds of numbers resulting from the above calculation; at this stage it is hard to determine a general characterisation for this (for a general speculation, derived from Kauffman’s (1995) work on dynamical systems, see Roberts (2001:91-93)). Although not in itself a step towards greater explanatory adequacy, this feature of the BCC has the heuristic value of allowing, in principle, an exact quantification of the cost of the postulation of a new parametrised formal feature.

What has the above to do with the NSP? Rizzi (1982:143) identified the core property allowing null subjects as a “pronominal Agr” in Infl in terms of the system he was assuming. In more contemporary terms, this can easily be restated as the presence of a D-feature on T (see §1.3, (19)), or, following Barbosa (1995) and Alexiadou & Anagnostopoulou (1998), on the verbal inflection itself. These two alternatives correspond to the two basic hypotheses for accounting for (non-discourse-driven) null
subjects in Holmberg (2005:536-7) and discussed in Roberts (chapter 1). If the requirement for a subject in SpecTP (Chomsky’s (1982:10) Extended Projection Principle) can be reduced to the realisation of a D-feature on T or its Specifier, then there is no need for the subject to raise to SpecTP in a language with D in T, giving rise to the possibility of free subject inversion and the associated possibility of extraction from the inverted position, evading the ban on extraction from SpecTP when C is finite and overtly realised (the complementiser-trace effect). Finally, if the D-feature is associated with rich agreement (see Roberts chapter 1 for a proposal for this), then we arrive at an account which ties together all four clustering properties associated with the NSP. Hence it is very easy to formulate the NSP in terms of a formal feature of a functional head.19

Finally, although the idea of reducing parameters to formal features of functional heads has largely been associated with “microparametric” approaches, it is not necessarily restricted to that case. In fact, we have just sketched a way to replicate the clustering predictions of Rizzi’s original NSP in these terms. Such clustering is usually seen as the hallmark of a “macroparameter.” Baker (2008a:4) points out that it “is perfectly possible that a lexical parameter consistent with [the BCC, AH/IGR] could have a substantial impact on the language generated, particularly if it is concerned with some very prominent item (such as the finite Tense node)”. This is, in fact, exactly what we have just seen in relation to the NSP.

19 Newmeyer (2005:208) asserts that the minimalist approach to parameters, which involves seeing them as inherently connected to features of functional heads, “makes it all but impossible to predict any significant degree of clustering.” This assertion is false, as we have just seen.
In this section we have described the BCC, the idea that parametric variation is associated with formal features of functional heads, and we have indicated what we see as some of its advantages. We have pointed out that, by postulating a restriction on what can be parametrised, on the form of parameters, and, possibly, interactions among parameters determined by disjunctive ordering (and therefore markedness relations), the BCC can take us some way to resolving the tension between descriptive and explanatory adequacy we observed in §3.1. However, if we take seriously the idea that parameters are lexically determined, even as properties of functional heads, we run the risk of seeing them as “less universal” in various ways. For example, could there be a language-particular parameter? This presumably depends on whether we are willing to countenance language-particular features of functional heads. A second question is whether a single language, i.e. lexicon, can tolerate contrary specifications of the same parameter on distinct lexical items. Third, can a parameter simply fail to have a value in a given language, by no having lexical item which realises this feature? (Thanks to David Willis for raising these issues). We will return to this last question in §3.4 below; GGL, for example, allow for this possibility. But the most important objection to a purely microparametric approach remains the question of the highly specific innate endowment in formal features that the BCC appears to demand; this is clearly a stumbling block to true explanation. We will look at this question more closely in the next two sections.

3.3 Micro- and macroparameters
In introducing the BCC in the previous section, we have already introduced one way of thinking about the distinction between microparameters and macroparameters. Although it does not require it, as we saw, the BCC favours a microparametric approach. According to this view, cross-linguistic variation consists of variant features of (a subclass of) lexical items which determine a small range of variation, and larger-scale differences among languages represent the accumulation of numerous microvariants of this kind. On the other hand, macroparameters such as the Polysynthesis Parameter of Baker (1996) and, possibly, the Head Parameter determine in one fell swoop a huge range of possibly variant properties. As Baker (2008a:5) puts it “there are at most a few simple (not composite) parameters that define typologically distinct sorts of languages.”

Baker gives interesting arguments for the existence of macroparameters alongside microparameters. In addition to two arguments based on the cross-linguistic distribution of different kinds of agreement marking (these are developed at greater length in Baker (2008b), but we will not go into them here), he gives an interesting statistical argument. Essentially, his point is that if all variation were microvariation, we would not expect to find coarse-grained types of the “head-initial”, “head-final” kind. If each category were able to vary freely, independently of all others, for its linear order in relation to its complement (this can of course be phrased in terms of triggering movement of its complement or not), then we would expect there to be a normal distribution of word-order variants across languages. As he says (Baker (2008a:10)), “there should be many mixed languages of different kinds, and relatively few pure languages of one kind or the other”. On the other hand, if there were only macroparameters, we predict, falsely, the kind of situation described above in relation
to the NSP: every category in every language should pattern in one way or the other. But if we admit both macroparameters and microparameters, we expect to find a bimodal distribution: languages should tend to cluster around one type or another, with a certain amount of noise and a few outliers from either one of the principal patterns. This, Baker points out (again drawing on the statistics for OV/OP vs. VO/PO order in Dryer (2005:386)), is essentially what we find. He suggests (pp. 11-12, citing his earlier 1996 work) that the same is true regarding polysynthesis.

We find Baker’s argument fairly convincing. One might add a similar diachronic argument. A canonical example of microparametric variation comes from the Italo-Romance dialects. Although the variation among these varieties is highly impressive (as the 2,500 pages of Manzini & Savoia (2005) amply attest), a large number of features remain constant: all Italian dialects are SVO, all are prepositional, none show a systematic ergative case/agreement pattern (although some “split-ergativity” is attested), none is fully polysynthetic, none shows the Chinese value of Chierchia’s (1998) Nominal Mapping Parameter (i.e. in allowing a singular count noun to stand alone as an argument, giving *I saw cat*), all have definite and indefinite articles, all have moderately rich agreement systems, all (except a small number of Rhaeto-Romansch varieties; see Benincà & Poletto (2005)) have complement clitics, none has a full morphological case system, etc. On the other hand, the microparametric variation involving the existence and behaviour of subject clitics, the expression of negation, the position of both finite and non-finite verbs in relation to subject and complement clitics and various classes of adverbs, the nature of object- and subject-agreement on past-participles in compound tenses, the nature and choice of aspectual auxiliaries, the expression of various forms of finite and non-finite complementation,
and a range of other properties, is extremely intricate. These are exactly the conditions which favour productive microparametric work, as Kayne has convincingly argued (see in particular Kayne (2005)).

But, one could ask, why are certain properties variable in Italo-Romance and others not? The microparametric answer, as it were, is that no theoretical significance should be attached to what varies and what does not in this particular synchronic geographically defined domain; this is attributable to a historical accident, in that the common features are due to a shared inheritance. But if we consider Latin, we find OV order, a full morphological case system, the complete absence of pronominal clitics, no (active) compound tenses, and a system of complementation in which finite clausal subordination was a minority pattern. As has often been observed, the Modern Romance languages (or the Italo-Romance subgroup) are more similar to one another than any of them are to their common ancestor Latin. The microparametric explanation for this observation would presumably appeal to the accumulation of microparametric changes in the common ancestor language before it broke up into the dialects, i.e. in Late or Vulgar Latin. The question here, though, is to what extent Vulgar Latin can be reasonably regarded as a single system; the term is generally used a cover term for the varieties of non-literary Latin spoken in Italy and elsewhere in the Roman Empire, whose written records are somewhat uniform but have been argued to form a koiné (Palmer (1961:223)). In this connection, Clackson (2004:790) says: “the construction of a uniform ‘Vulgar Latin’ probably oversimplifies a very complex linguistic situation. Different communities of speakers used different varieties”. If there ever was a single “Proto(-Italo)-Romance” variety, it would probably have to be dated rather early, as Hall (1950) suggests on phonological grounds (proposing 250-
200BC, exactly the period in which Roman rule was extended to the whole Italian peninsula). Although the Latin of this period is known to differ somewhat from Classical Latin, and to have certain “Vulgar” features, it is highly unlikely that it had the syntactic characteristics of Romance rather than Classical Latin (OV rather than VO order, etc).

It seems then that the current microparametric variation either derives historically from an archaic, typologically distinct, single ancestor variety of Latin, or there is no ancestor variety common to all the dialects. Either way, the major typological differences between Latin and (Italo)-Romance cannot be traced to a single microparametric change or series of microparametric changes in a single variety; there must have been typological drift across the varieties of Vulgar Latin. This poses a problem for a purely microparametric approach: other things being equal, we might have expected some dialects to have retained a case system, or OV order, or synthetic passive forms and not to have developed clitics, etc., others to have developed in the way we observe, and still others to have developed in a mixed fashion, preserving certain archaic features and innovating others. But what we observe, instead, is typological drift: from OV to VO, and in the general direction of greater analyticity (as elsewhere in Indo-European), allowing for a considerable amount of truly microparametric variation of the kind that we observe to develop. The simplest account of this kind of parallel development involves distinguishing macroparametric from microparametric change: certain macroparameters (OV vs VO, for example) changed in the transition from Latin to Romance, while much of the synchronically
observable variation among the Romance languages, and certainly among Italian
dialects, involves microparameters.\(^{20}\)

So let us conclude, with Baker, that macroparameters exist alongside
microparameters. Then two related questions arise: (i) what are their properties? (ii)
how are they distinguished from microparameters? Two rather unsatisfactory and
partial answers to these questions are that macroparameters ought to be rather few in
number, and they ought to be extremely pervasive in their influence on the
grammatical system. The first point holds because, as Baker (2008a:7) says “[i]f there
were many macroparameters and they interacted with one another in complex ways,
then languages could differ crazily in ways that would be hard to pull apart.” But this
does not seem to be the case in practice: universal properties and microparameters
account for much that is shared and much that varies. The second point holds because
macroparameters, perhaps by definition, can affect large-scale aspects of the grammar
such as all headed phrases, or all instances of Agree.

\(^{20}\) One could perhaps attempt a contact-based explanation for the parallel developments. The
Southern dialects were in contact with Greek and the other Italic varieties of Indo-European: Oscan,
Umbrian and related varieties, collectively known as Sabellian. These are broadly similar to Latin in
typological terms, being predominantly SOV (Wallace (2004:832)), and so are unlikely to be
responsible for the common development of the dialects. The Northern varieties were in contact with
forms of Celtic and Venetic. In Tuscany and indeed in the early days of Rome itself, there was contact
with Etruscan, a non-Indo-European SOV language (Rix (2004:961)). A very thorough study of contact
between Latin and all of these languages, at various times and places, is Adams (2003). One
possibility, which could have had far-reaching consequences, is that Latin demonstratives developed
into articles partly due to contact with Greek, which had an article system. Adams (2003:518) points to
the use of demonstratives as articles in a passage of Plautus. This is significant because it can be traced
to Greek influence and because of its early date (see Adams’ discussion for details).
In the preceding section we argued that the BCC, which Baker quite reasonably takes to underlie microparametric variation, has a number of desirable consequences, and may even take us some way towards constraining the form of parameters in the way that is required in order to resolve the tension between explanatory and descriptive adequacy which has arguably arisen in this domain. But here we have suggested that Baker is right in suggesting that a small number of macroparameters may also exist. Baker (2008a:3) explicitly proposes that macroparameters are to be formulated in a manner incompatible with the BCC. So we appear to be in a quandary.

We tentatively suggest a way out of this quandary which, we believe, points the way to truly resolving the tension between explanatory and descriptive adequacy in the parametric domain. This involves retaining a formally “microparametric” view of macroparameters, i.e. seeing macroparameters as aggregates of microparametric settings, but as proposing that these aggregate settings are favoured by markedness considerations. This proposal was made in Roberts (2007a:274) for the Head Parameter (and is suggested as an “intermediate” approach to the question of macro- vs. microparametric variation by Baker (2008a, Note 2)).

It has often been noted that the Head Parameter is rather problematic. If it were a single (macro)parameter, determining the order of head and complement across all categories once and for all, it would predict a spectacular clustering of properties, which is not actually attested in the majority of languages. If it is broken down into a series of related microparameters relating to each head-complement pair then, without some further statement, all predictions regarding word-order correlations are 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, Roberts (2007a:194) suggested that a restatement of Hawkins’ (1983) generalisation regarding cross-categorial harmony is needed, along the following lines:

(36) There is a preference for the EPP feature of a functional head F to generalise to other functional heads G, H …

We can think of (36) as an approximation to a markedness convention of the type proposed for phonology by Chomsky & Halle (1968, Chapter 9).

To take a specific example, suppose, following Kayne (1994), that VO is the universal underlying order and, following Biberauer (2003), that OV orders derive from the combination of V-to-v raising and remnant VP-fronting to SpecvP, as illustrated in (37):

(37) \[ vP \ [vP O (V) \ ] v+V (VP) \ ] ]

If movement represents a marked option, as suggested by Roberts & Roussou (2003), then v is set to a marked parameter value here. Following Chomsky & Halle’s notation, let’s write this as the mEPP value for v.\(^{21}\) In rigidly head-final languages like

\(^{21}\) Here we are only concerned with the feature which attracts VP; let us leave aside whatever it is that attracts V. In Biberauer’s system, VP-movement is triggered by an EPP feature associated with
Malayalam or Japanese, many, perhaps all, functional heads will have at least one EPP-feature of this kind. 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:

\[(38) \quad \text{For a class of heads } H, \ uEPP \text{ for } H_{[F,-]} \neq v \Rightarrow \{ [+EPP] / v_{+[EPP]}; \} \]

\{ [-EPP] elsewhere \}

What (38) says is that the unmarked value of the EPP feature for some head of a particular type with an unvalued 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. This has the effect that, for all head-complement pairs, 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 (1992) observation that VO vs. OV order is the basic determinant of ordering among other head-complement pairs is the v’s property of probing for D, VP-movement representing the “pied-piping” option as compared to object shift (movement of the object DP).
directly captured. One might speculate that v is the crucial category determining the markedness of the system with respect to word order because it is the head of the phase which determines argument structure and therefore the category whose features are most important for determining the positioning and licensing of arguments.

We can understand a markedness convention like (38) in terms of conservatism of the learner, assuming that the learner exploits pieces of input, perhaps marked input, to the full. So we could entertain something like the following:

(39) Generalisation of the input:
If acquirers assign a marked value to H, they will assign the same value to all comparable heads.

(38) can naturally be understood in terms of (39). Moreover, both (38) and a simple feature-counting simplicity metric like that put forward by Roberts & Roussou (2003:210) can be seen as different aspects of the overall conservatism of the learner, which essentially tries to set parameters in the most efficient way possible. The Subset Principle can also be seen in this light: one aspect of the learner’s conservatism is to avoid superset traps (see Berwick (1985), Clark & Roberts (1993), Biberauer & Roberts (forthcoming)). We will return to the question of why something like (39) should hold in §3.5. For the moment we may observe that macroparametric effects arise from aggregations of microparameters acting in concert for markedness reasons.

One could perhaps extend this to the Polysynthesis Parameter. According to Baker (1996:14, 17, 496), the central property of polysynthetic languages is that all
argumental DPs must be correspond to a morphological expression in the head that θ-marks them. The reflexes of this very general condition (which Baker calls the Morphological Visibility Condition) are syntactic noun-incorporation, rich object- and subject-agreement marking, “free pro-drop” of all arguments, free word order and a range of other major morphosyntactic properties (see Table 11.1, Baker (1996:498-9)). Kayne (2005:7) observes that there is a similarity between these features of polysynthetic languages such as Mohawk and what is found in clitic doubling/dislocation constructions in Romance. He speculates that “it could alternatively be the case that the systematic obligatoriness of pronominal agreement morphemes in Mohawk is just an extreme example of what is found to a lesser extent in (some) Romance” (ibid). Suppose, then, that Mohawk and other polysynthetic languages have generalised clitic left-dislocation (CLLD), while Romance languages have partial CLLD and languages such as English lack it altogether. There is a link to the NSP here, in that rich subject-agreement on the verb may be thought of as comparable to a clitic and the subject then seen as clitic left-dislocated (this is one variant of Hypothesis A in §1.3, (17), pursued by Barbosa (1995, to appear) and Alexiadou & Anagnostopoulou (1998)). Clitics can be thought of as the overt realisation of both the probe’s and the goal’s matching φ-features on the probe (see Roberts chapter 1, in preparation); this prevents the goal from being realised in its grammatical-function position (this applies to non-c-clitic-doubling; the difference between clitic-doubling and non-clitic-doubling may have to do with the interpretability of the features, but we will leave that aside here). We can thus envisage a markedness statement based on (38) which would specify that realising all features on all possible probes is the more marked option than never realising them, but less marked than realising them sporadically. Again, the Polysynthesis Parameter
may reduce to an aggregation of microparameters concerning clitics/agreement, governed by a markedness constraint.

Are there other macroparameters? One intriguing recent suggestion is due to Huang (2007). He observes that a range of properties, up to 20 or more, appear to cluster together in Chinese, as compared to English and other familiar European languages, or to Japanese, for example. Space prevents full discussion and illustration of these, but among them are the following:

(40) a. Generalized bare N (denoting kinds)
    b. A generalized classifier system
    c. No plural morphology

(41) a. Action verbs are atelic
    b. No simplex accomplishment verbs.
    c. Resultative compounds or phrases
    d. Periphrastic causatives
    e. Extensive use of light verbs.

(42) a. No agreement, tense, case morphology
    b. No wh-movement.
    c. Word order: “V2” counting backwards (no Kaynean VP-movement over adjuncts).\textsuperscript{22}

\textsuperscript{22} This refers to the fact that Chinese canonically shows the order Accunct V Complement rather than V Complement Adjunct (e.g. John often visits Mary vs. John visits Mary often). Following Kayne
d. No V-to-T movement

(43) Radical Pro drop

Huang proposes that these properties seem to cluster together in one language-type to the exclusion of another, because they are manifestations of the same generalization. He proposes the macro-parameter High Analyticity, which states that Chinese lexical items are highly analytic at three levels: the level of lexical categories, functional categories and argument structure. In particular, in a detailed study of the diachronic development of Chinese, Huang shows that the following properties have been lost:

(44) a. wh-movement
    b. VP movement
    c. NP movement
    d. EPP movement
    e. Plural morphology
    f. Denominal suffix
    g. Causativizing suffix

These losses have taken place in conjunction with the development of numerous particles, particularly for marking tense, mood, aspect and illocutionary force. It (1994), Cinque (1999), the English order may result from VP-raising into the functional field; Chinese lacks this operation. Hence, in general, the only material following V is the complement (and, given verb serialization, V usually has exactly one argument). This gives the appearance of “reverse V2”.

72
seems tempting, then, to try to maintain that “analyticisation” is loss of movement, associated to some degree with loss of morphology. In particular, it seems that Chinese lacks movement at the lowest structural level, inside the lexical phase. It is plausible to think of the properties in (40) as being related to the lack of N-to-n movement (assuming classifiers to be ns and plural marking to be determined or fed by N-to-n movement). Similarly, at least (41a-c) could be connected to the lack of V-to-v movement, assuming that a complex event structure such as that involved in accomplishments requires some form of incorporation in the vP phase (see Ramchand (2008)). The properties in (42) are more familiar, but perhaps depend on the prior formation of verbs by V-to-v movement (in the case of V-to-T movement) and the prior existence of lexical elements in the n/v position that can be probed by uninterpretable φ-features merged outside the first phase, in the case of (42a) for example. Furthermore, there appear to be no EPP-type movement triggers at the higher phase level (see (42b, c)), although, since Chinese is topic-prominent, Edge Features must be present on C. We will return to the relation of radical pro-drop to the absence of φ-features below; see also Saito (2007), Roberts chapter 1). Putting all of this together with the very well-known fact that in Chinese it is impossible to distinguish nominal and verbal roots by their morphological shape, and that very many basic roots are entirely ambiguous between nominal and verbal interpretation, it is tempting to suggest that head-movement, in particular L-to-l movement (i.e. movement of the lexical root to the local phase head) is systematically absent in Chinese. More generally, phase heads appear to lack Agree-related movement triggers (EPP features). In the non-lexical phases (CP and DP), this has the consequence that probes are largely absent, as their putative goals are too deeply buried in the lexical
phase to be accessible (assuming the version of the Phase Impenetrability Condition in Chomsky (2000), rather than the less restrictive version in Chomsky (2001)).

So we might, very tentatively, conclude that Huang’s High Analyticity macroparameter results from an aggregate of head-movement parameters acting together, in this case in not triggering movement. In more familiar languages, we have the inherently more marked situation where heads sporadically trigger movement (although fairly systematically at first-phase level; but see Biberauer & Roberts (chapter 7) for the proposal that (Modern) English lacks V-to-v movement).

A final, very speculative suggestion comes to mind at this point. If there are macroparameters determining polysynthesis and high analyticity, are there parameters determining other morphological types? In fact, Julien (2002, Chapter 3), following a proposal in Kayne (1994), proposed that very many OV agglutinating languages typically showed the following structure, for all (or most) X:

\[
\text{(45)} \quad [xP \ YP \ [x \text{affix }] \ (YP)]
\]

This is in fact the variant of the general OV parameter, generalised by (38), where the host head contains a bound morpheme. So (38) may fall into two subcases, depending on precisely this, the former giving OV order, the latter agglutinating morphology and OV order. It has been observed many times that rigidly OV languages tend to be agglutinating.
Finally, we are led to think that perhaps the very first kind of typology ever proposed, the morphological typology put forward by Schleicher (1862) and Sapir (1921) among others, was on the right track after all, but it was really a pre-theoretical observation about syntactic macroparameters. The values of the macroparameters are so salient that one could not fail to notice their effects in the data, but, in the absence of a theory of syntax, it was not possible to discern their true nature. And so the generalisations were mistakenly thought to be morphological. Current parametric theory, especially given the distinction between macroparameters and microparameters as construed here, enables us to tentatively begin to do this.

Here we have argued first that Baker (2008a) is correct in distinguishing macroparameters from microparameters, but that it is inadvisable to abandon the BCC. Instead, we have suggested that macroparameters are the result of aggregates of microparameters acting in concert, guided by the acquisition-based markedness constraint in (38). In the next section, we will relate all this to the NSP (and to the somewhat incidental question of whether this is a macro- or microparameter), to GGL’s proposal for “parameter schemata” and to a proposal for learning paths, in the sense of Dresher (1990), which relate macro- and microparameters.

---

23 The earliest version of this typology was put forward by Schlegel (1817); see Morpurgo-Davies (1998:71-75) for discussion of this and its 17th and 18th-century antecedents.

24 At this point, it is natural to ask whether there is an Inflectional/Fusional Parameter. We suspect not; this parameter arises where none of Polysynthesis, Analyticity or Agglutination are set to their positive values. Inflection/fusion seems to involve non-uniform behaviour among functional heads, and hence might be seen as a marked system. Certainly, these systems are prone to change, as much of the history of Indo-European attests, and inflectional languages tend to be of mixed type (cf. German, Latin, Sanskrit and other conservative, highly inflecting Indo-European languages).
3.4 Epigenetic parameter-setting

A further critical question concerning the general nature of parameters that Newmeyer (2005: 44) very correctly raises is “whether all parameters are applicable to all languages.” In the principles-and-parameters literature, the answer to this question has generally been positive, although it does lead to the questions concerning the status of a complex range of microparameters related to clitics, or agreement, or classifiers, in languages where these properties are lacking, as discussed in §2.1.

GGL (2008) explicitly propose that the answer to this question should be negative. They propose that, instead of innate parameters, UG makes available a small set of parameter schemata, which, in conjunction with the PLD, create the parameters that determine the non-universal aspects of the grammatical system. In this way, parameters are created through interaction with the PLD in a fashion reminiscent of the Piagetian concept of epigenesis. They propose the following form for their schemata (pp. 7-8):

(46)  
   a. Grammaticalisation: is F, a functional feature, grammaticalised?
   b. Checking: is F, a grammaticalised feature, checked by X, X a category?

25 Boden (2006: 493) characterises epigenesis as follows: “a self-organising dialectic between biological maturation and experience.” A similar suggestion, but restricted to what he calls “core parameters” (approximately equivalent to Baker’s notion of macroparameter), is put forward by Uriagereka (2007:106ff). See also Piattelli-Palmarini & Vercelli (to appear).
c. Spread: is F, a grammaticalised feature, spread on Y, Y a category?
d. Strength: is F a grammaticalised feature checked by X, strong? (i.e. does it overtly attract X?)

They illustrate the functioning of such schemata in detail, in relation to the feature definiteness, and its effects on the internal syntax of DPs in some of the 24 languages for which they have obtained data regarding 46 parameters.

In a similar vein, Roberts & Roussou (2003:213) propose the following set of options relating to a given formal feature F on the basis of their extensive analysis of grammaticalisation as a diachronic operation affecting the realisation of functional categories:

(47) a. is F realised by (external) Merge (i.e. does it correspond to an overt grammatical formative?)
b. does F enter an Agree relation?
c. if so, does F attract?
d. if so, does F attract a head or an XP?
e. if (c), does F attract both a head and an XP?
f. does F combine realisation by external and internal Merge?
g. if so, does F attract a head or an XP?

(Roberts & Roussou do not assume GGL’s initial question: whether F is present at all, assuming instead that all languages use the same set of formal features (see Roberts & Roussou (2003:29))). What (47) and (47) share is specifying a range of formal
operations which can be associated with a given type of substantive feature (a formal feature of a functional head). They differ in detail, and this is certainly not the place to evaluate their relative merits. What we can note is that the sequence of statements involves a steady increase in specificity in each case. In fact, each statement is close to being disjunctively ordered in relation to the previous one, and it would certainly not be difficult to reformulate either (47) or (47) so as to make this more precise). Roberts & Roussou explicitly state that their system reflects a markedness hierarchy; GGL on the other hand make no such claim. Moreover, each set of statements has a kind of “branching” structure, which we can illustrate as follows for (47b-d), replacing (47a) with GGL’s option of the presence of F as a formal feature, and simplifying slightly:26

(48) F?
   No \_\_\_ yes

STOP does F Agree?
   No \_\_\_ yes

STOP does F have an EPP feature?
   No \_\_\_ yes

Does F trigger head-movement?
   No \_\_\_ yes

26 In the theory of incorporation put forward in Roberts (forthcoming), a probe which triggers incorporation cannot have an EPP feature for principled reasons, which justifies the way the options are presented in (49), as opposed to the slightly richer set of options put forward by Roberts & Roussou.
(This can also be done for (46)). Each “yes” option entails a further option. The “yes” options that do not dominate anything may entail further options regarding the type of head-movement, or, in the case of XP-movement, pied-piping options (see Richards & Biberauer (2005), Biberauer & Richards (2006) on the latter option); we leave further specifications aside here. Each more deeply embedded option is more marked than all less deeply embedded ones, since effectively the description of the parameter is the conjunction of all the dominating nodes, and so it increases in length as embedding deepens. “STOP” options on left branches are relatively unmarked options in each case. So here we see a parameter schema given as a network of options, each more embedded option representing a more specific, and therefore a more marked, option. Importantly, we can consider networks like (49) to define “learning paths” in the sense of Dresher (1990); again the conservatism of the learner is such that it prefers the path to be as short as possible, and so deeply embedded options are relatively marked owing to the fact that they have longer descriptions. Following GGL, we assume that the schema and the overall pool of possible features are given by UG; the network is created through epigenesis in acquisition, and markedness follows, on one standard construal, from increasing specificity (length of description relevant to F’s role in the grammar, and hence greater computational burden; we will say more about markedness in the next section).

Now, parameter schemata of the kind in (46-48) apply to individual formal features. As such, they are classic examples of microparameters (and have many of the advantages of this kind of formulation of parameters discussed in §3.2). But we are
now in a position to fruitfully combine these, or some of them at least, with markedness statements of the kind in (37), to derive some of the macroparameters discussed in the previous section. In a nutshell, macroparameters quantify over F in networks like (48).

To see how this works, consider the EPP option embedded two levels down in (49). The markedness statement in (36) essentially says that the unmarked option for the grammatical system (i.e. not necessarily for F itself) is “no F has this value”, and that the next least marked option is “all F have this value”, and that the trigger for choosing is the value taken by v. More mixed, and therefore more marked, systems may relate the possession of F to further categorial features, and the options may become progressively more specific (have longer descriptions) and more marked. In other words, we have a cross-cutting set of options of the form:

(49)  a. Are movement-triggering features absent from all probes?
    b. If not, are movement-triggering features obligatory on all probes?
    c. If neither (a) nor (b), are movement-triggering features present on \{T, v, \ldots\}?  

The positive value of (49a) gives a rigidly, harmonically head-initial language like Welsh. The positive value of (49b) gives a rigidly, harmonically head-final language such as Japanese or Turkish. Again, (49c) breaks up into a series of microparameters, with a range of other factors enter here (options regarding pied-piping, and general constraints on disharmonic orders of the kind explored in Biberauer, Holmberg & Roberts (2007)).
The existence of this set of cross-cutting options is determined by generalisation of the input. Applying (48) where F is universally quantified, we come very close to deriving the macroparameters discussed in the previous section (we have added one further option on the most embedded right branch, for further illustration):

(50) F?
No \(\Rightarrow\) yes
STOP does F Agree?
No \(\Rightarrow\) yes
STOP does F have an EPP feature?
No \(\Rightarrow\) head-initial \(\Rightarrow\) yes head-final
Does F trigger head-movement? Is F realised by external Merge?
No \(\Rightarrow\) yes
STOP polysynthesis
High analyticity
No \(\Rightarrow\) Yes:
STOP agglutinating

(This may wrongly predict that polysynthetic languages are head-initial, but actually they appear to have free word order, precisely owing to their polysynthetic nature (see Baker (1996:10ff.)). So we can finally arrive at a picture of the form of parameters as involving generalised quantification over formal features, as follows:

(51) \(Q(f \in C) [P(f)]\)
Here Q is a quantifier, f is a formal feature, C is a class of grammatical categories providing the restriction on the quantifier, and P is a set of predicates defining formal operations of the system (“Agrees”, “has an EPP feature”, “attracts a head”, etc.). The longer the characterisation of either C or P, the more deeply embedded in a network/schema the parameter will be, the more marked it will be, and the further along the learning path it will be. True macroparameters sit at the top of the learning path, and in fact involve unrestricted universal quantification or its negation, as we saw in (49) (Baker (2008a:9) also suggests that macroparameters might be a kind of default). This seems to us to be a maximally simple theory of parameters, since ultimately it involves relations between sets of features of categories and predicates defining grammatical operations.

Can we relate the NSP to these considerations? Roberts (chapter 1) speculates along just these lines. He suggests that putting together Saito’s (2007) proposals regarding the nature of radical pro-drop (which were briefly described in §1.2.3) with the particular account of the “rich agreement” that facilitates consistent null subjects based on Müller’s (2005) notion of impoverishment (again, see Roberts (chapter 1) for details) we arrive at the following generalisations:

(52) a. Radical pro-drop is possible iff φ-agreement is not obligatory.

b. Consistent null subjects are possible iff there is no impoverishment of T’s φ-features.
Where (52a) holds “discourse pro” is possible (i.e. subject pronouns can be merged at LF); where (52b) holds, deletion of subject pronouns is possible prior to LF (again, see Roberts’ paper for details). We see that the two systems are derivational mirror images of one another, and that this is the direct consequence of the different status of \( \phi \)-features on probes (fully optional vs. obligatorily present and unimpoverished), which in turn is typically reflected in the agreement morphology (totally absent vs. “richly” realised). \(^{27}\)

This further suggests a rethinking of the typological generalisations surrounding null arguments: perhaps the fundamental dimension of parametric variation is “radical” vs “consistent” null-subject (or null-argument) languages, with partial and non-null-subject languages being subcases of the “consistent” type featuring varying degrees of impoverishment of the goal. The basic form of the parameter would then be as in (54):

\(^{27}\) The Mainland Scandinavian languages do not show any subject agreement in finite clauses (see Holmberg & Platzack 1995: pp), yet they are not radical pro-drop languages, but instead non-null subject languages. We maintain that they have a generalised \( u \phi \)-feature in finite T which does not have a morphologically realised valued form, but is nevertheless visible in virtue of obligatory movement of a nominal subject to spec,TP; see Holmberg (chapter 2).
(54)  a. Are uφ-features obligatory on all probes?

No /\ yes

Radical b. Are uφ-features fully specified on all probes?

Pro-drop Yes /\ No

Polysynthesis c. Are uφ-features fully specified on some probes?

No: Yes

Non-null-subject d. Are the uφ-features of \{T, v, \ldots\} impoverished?

As indicated, the “No” value in (54a) gives radical prodrop. The negative value of (54b) may give rise to a polysynthetic system (or at least to consistent head-marking; Baker’s notion of polysynthesis seems to combine this with consistent head-movement of arguments; see (50)). A positive value for (54c) gives a non-null-subject language like English. (58d) is intended to simply indicate the ways in which the null-subject parameter starts to “break up” into microparameters as individual probes are evaluated in relation to it (cf. §1.2.5,(16)). Clearly, a “no” value for T and a “Yes”
value for \( v \) will give a consistent null-subject language like Italian.\(^{28}\) In terms of the general schema for parameters in (51), we can state the NSP as follows:

\[
\exists f \in D \; [ \; S(D, T_{Fin}) \; ]
\]

(54) reads “For some feature \( D \), \( D \) is a sublabel of finite \( T \)”, where “sublabel” is understood as in Chomsky (1995:268). This captures the force of the informal statement given in (18) above, and shows how the NSP fits with the general format for parameters, and how it is part of the parameter network in (54).

Partial null-subject languages, and intricate cases like certain registers of French (see Roberts (chapter 8)), require still further specification. Again, these represent progressively more marked options, located more deeply in the schema/network and further down the learning path. In these terms, we can immediately observe a connection between the null-subject parameter and the other parameter schemata/networks discussed here.

A further advantage of hierarchies of the type sketched above is that they restrict the upper bound of grammars that a given set of parameters can generate. The cardinality of \( G \), the set of grammars, is equivalent to the cardinality of \( P \), the set of parameters, plus 1, to the power of the number of hierarchies:

\[
|G| = (|P| + 1)^n, \text{ where } n = |H|
\]

\(^{28}\) The clustering properties might follow, given §2.2(30/32). On the connection of rich agreement to all of this, see Roberts (chapter 1, §2.5).
Suppose, arbitrarily, that there are 5 hierarchies (we have seen two; there must be one for word order, and it is very easy to imagine that there are at least two more), and suppose that there are 30 parameters. Then $|G| = 31^5$, or $28,629,151$. This is a large number, but recall that 30 parameters yielded over a billion grammars on our earlier calculation based on unhierarchised microparameters (see §3.2).

In this section and the previous one, we have clearly made some progress towards resolving the tension between descriptive and explanatory adequacy that we observed in §3.1. In particular, the question of the highly specific innate endowment in formal features that the BCC appears to demand has been eliminated, and we have also clarified the relation between micro- and macro-parameters. All parameters ultimately have the extremely simple form in (51), and they form schemata/networks which are related to markedness of the general kind in (48) and (50). The NSP for example is a case of the parameter schema in (53), specifically (53c) as it relates to T. We believe that these proposals go a long way towards to restoring the explanatory value of parameters (as well as giving wide empirical coverage, if our speculations about macroparameters are on the right track). In the next section, we will suggest that they may show the way beyond explanatory adequacy.

3.5 Why parameters? Comparative syntax beyond explanatory adequacy

In recent work, Chomsky (2004, 2005, 2007) has proposed that an important property of the minimalist programme is that it can take us beyond explanatory adequacy (see Chomsky (2007:19) for a lucid statement of this idea). Accordingly, we attribute the adult state of linguistic knowledge, adult competence, to three factors: (i) the genetic
endowment, UG, (ii) experience of the PLD, (iii) principles not specific to language. The last have become known as “third-factor principles” and have to do in particular with principles of optimal and efficient computation. In this section, we would like to show how the view of parameters arrived at in the preceding sections can begin to take us beyond explanatory adequacy in this domain.

The classical view of how principles and parameters interact to produce adult competence was based on the idea that the parametric options were specified as such as part of the genetic endowment, made manifest in the PLD (perhaps with inaccessible properties being triggered by accessible properties in a parametric cluster) and thereby fixed during language acquisition, to give an adult system which was an instantiation of UG with all parametric options fixed. So this view relied entirely in the interaction of factors (i) and (ii) above, characteristic of the classical notion of explanatory adequacy.

The view of principles and parameters that follows from the considerations in the previous sections is rather different. UG does not even provide the parameter schemata. In essence, parameters reduce to the quantificational schema in (49), in which UG contributes the elements quantified over (formal features), the restriction (grammatical categories) and the nuclear scope (predicates defining grammatical operations such as Agree, etc). The quantification relation itself is not given by UG, since we take it that generalised quantification – the ability to compute relations among sets -- is an aspect of general human computational abilities not restricted to language. So even the basic schema for parameters results from an interaction of UG elements and general computation.
The parameter schemata form networks defined by markedness relations. The markedness notions we have invoked include relative length of description, and generalisation of the input (relevant for statements such as (38), which ultimately define macroparameters). Both rely on a general notion of computational conservatism, which again we can think of as a facet of computational efficiency. Again, then, the schemata arise from third-factor properties. The different points in the schemata all instantiate the schema in (49); they differ in the specificity of the two arguments to the quantifier \( C \), the class of grammatical categories, and \( P \), the predicates defining (conjunctions of) grammatical operations. The more specific either of these arguments, the more marked, and indeed the more “micro”, the parameter.

Third-factor considerations may contribute to markedness in other ways, too. Mobbs (2008) critically reviews Hawkins’ (2004) three performance efficiency principles we mentioned in §2.5, and which were invoked by Newmeyer (2005) as an alternative to parametric accounts. We give them here:

(56) a. Minimize Domains (MiD)
   b. Minimize Forms (MiF)
   c. Maximize Online Processing (MaOP)

We saw in §2.5 how Newmeyer, following Hawkins, tries to invoke MiD in order to account for a well-known Greenbergian implicational universal. Whether or not that account is successful (and there are many reasons to think it is not, see Note 19), one can wonder as to the precise status of these efficiency principles. Mobbs convincingly
argues that Hawkins’ Performance-Grammar Correspondence Hypothesis (PGCH) reflects only a correlation, not causation, and that we might do better to reconsider the efficiency principles as third-factor principles that play a role in defining UG.

In these terms, MiD naturally relates to the “minimal search” considerations which underlie locality conditions: the non-intervention condition on Agree and the PIC. It also relates to the No Tampering Condition (existing relations should not be altered by later operations) and cyclicity generally (Mobbs (2008:8-9)). So a version of one of Hawkins’ efficiency principles, more abstractly construed as a general computational principle informing competence, rather than constituting performance, may lie behind certain fundamental properties of the language faculty.

What concerns us more directly here, however, are third-factor constraints related to markedness. Mobbs suggests very plausibly that Hawkins’ Minimise Forms (MiF) constraint may underlie the markedness preference articulated in Roberts & Roussou (2003) for relatively simple forms. More importantly, he proposes a further efficiency constraint, Generalise Features (GenF), which he states as follows:

(57) Human computation shares features over forms in the same domain (Mobbs (2008:11)).

This, he suggests, may underlie the “generalisation of the input” form of markedness put forward in (38) above. We concur with this suggestion.
It seems, then, that Hawkins-style efficiency principles may have a role to play in our theory of universals and typology, but not as performance constraints on processing. Instead, they may reflect deeper, rather general computational principles, which contribute to the third factor determining adult competence. They may contribute to typology to the extent that they inform the markedness principles which determine parameter schemata/networks. Further, we have suggested that the general form of parameters themselves results from the interaction of UG primitives with the general principles of quantification.

Finally, we can ask the most difficult question of all, but one which the minimalist programme requires us to ask: why do we have parameters at all? Our general format for parameters in (51), inasmuch as it allows Q to be a negative quantifier, basically states that formal features of functional heads are all in principle optional. UG says nothing more than this, which is about as little as could possibly be said (in particular, this is a more “minimal” statement than either forbidding or requiring the presence of such features). Moreover, the quantificational schema is maximally liberal: it states that the formal features may be in any set-theoretic relation with any predicate defined by the theory of grammar. So parametric variation arises because UG really doesn’t mind about the distribution of formal features in any given grammatical system. But we know that speakers fixate on given grammatical systems during language acquisition, and speech communities recognise given (aggregates of) grammatical systems as languages. Neither of these aspects of parametric variation directly concerns UG, however: “fixing” parameters may be a facet (actually, almost a definition) of learning. So the kind of stable parametric variation we observe in adults arises from the fixation on random values. These values take on cultural and social
value – a very different kind of value – as “languages” for the kinds of reasons that have been revealed by sociolinguistics.

But, even if UG doesn’t mind, how and why could variation in grammatical systems have emerged? Here, the work of Niyogi & Berwick (1995, 1997) and Niyogi (2006) on modelling the acquisition and change of grammatical systems in populations of learners in revealing. They show that given a learning algorithm A, a probability distribution of linguistic tokens across a population (random PLD), and a restricted class of grammars from which to select (UG), variability will result as long as the time allowed for the selection of hypotheses is restricted. This idea emerges most clearly in the following quotation from Niyogi (2006:14-15):

imagine a world in which there are just two languages, $L_{h1}$ and $L_{h2}$. Given a completely homogeneous community where all adults speak $L_{h1}$, and an infinite number of sentences in the Primary Linguistic Data, the child will always be able to apply a learning algorithm to converge on the language of the adults, and change will never take place … Now consider the possibility that the child is not exposed to an infinite number of sentences but only to a finite number $N$ after which it matures and its language crystallizes. Whatever grammatical hypothesis the child has after $N$ sentences, it retains for the rest of its life. Under such a setting, if $N$ is large enough, it might be the case that most children learn $L_{h1}$, but a small proportion $\epsilon$ end up acquiring $L_{h2}$. In one generation, a completely homogeneous community has lost its pure character.
In other words, if we combine the heterogeneity in any speech community, the random distribution of PLD (poverty of the stimulus) and the limited time for learning (i.e. the critical period for language acquisition), change in grammatical systems is inevitable. If change is inevitable in the diachronic dimension, then variation is inevitable in the synchronic dimension. But, again, none of this reflects any aspect of UG except an indifference as to the distribution of formal features, as captured by (51). Here we begin to see the shape of comparative syntax, beyond explanatory adequacy.

3.6 Where are parameters? The locus of parametric variation

We have already discussed the BCC, the idea that parameters are specified in the lexical entries of lexical items, at some length. However, there is also a trend in recent minimalist theory to locate all linguistic variation, including all syntactic variation, in the post-spell-out morphology and phonology components (see for example, Sigurðsson (2006a,b), Poole & Burton-Roberts (2006a,b), Boeckx (forthcoming). The following quote is representative:

/.../ it is not implausible to think of narrow syntax as completely uniform (meeting LF-demands), and not affected (design-wise) or adapted to cope with or code for variation in the guise of (syntactic) parameters. (Boeckx, forthcoming).

This hypothesis (which Boeckx dubs the the Strong Uniformity Theory, SUT) is still at a programmatic stage, and thus has not been seriously put to the test, yet. It is an interesting hypothesis, though, and very much in the spirit of the minimalist
programme. What we are proposing here can indeed be construed as coming close to the SUT. For example, whether to delete or pronounce a pronoun is clearly a matter of PF, and variation with regard to EPP-features discussed above in section 3.3. is presumably best construed as taking effect after spell-out (which is made technically possible by phase-theory, according to which spell-out applies at several stages in the derivation of a sentence; Chomsky (2000, 2001, 2008), Svenonius 2004)). We would maintain, however, that some variation is encoded in narrow syntax. In fact, so does Boeckx. He concedes that the following variation is encoded there:

\[(58)\]

\[\begin{align*}
\text{a.} & \quad \text{Features } F_1 \text{ and } F_2 \text{ may be expressed separately or as a bundle;} \\
\text{b.} & \quad F \text{ may or may not exhibit a uF variant;} \\
\text{c.} & \quad \text{A given phase head may or may be strong, i.e. uF-bearing, or weak (defective).}
\end{align*}\]

(The similarities with GGL’s parameter schema in (46) and Roberts & Roussou’s proposal in (47) are obvious; see also the other proposals mentioned by Boeckx). But, he says, “all other ‘parametric’ options arise in the post-syntactic morpho-phonological component, such as whether a head H allows its specifier to be filled by overt material, or whether the head or the tail of a chain can/must be pronounced, or whether a given head is affixal and requires its expression in the vicinity of another head, or whether a head H precedes or follows its complement.” (Boeckx, ibid.). This is not significantly different from what we are proposing. In fact, Boeckx’s parameter schemata (a,b,c) fall under our more general parameter schema (51) (if P
can be ‘is bundled together with $F_2$', an idea which we have not discussed, but which will be exploited in Holmberg (chapter 1)).

Interestingly, Boeckx presents these ideas in the context of a paper where he distances himself from GB-style parametric theory: “[T]he idea that a GB-style Principles-and-Parameters architecture provides the right format for a solution to Plato’s Problem is, I think, seriously mistaken, on both empirical and conceptual grounds” and “if minimalists are right, there cannot be any parametrized principles, and the notion of parametric variation must be rethought.” (Boeckx, ibid.). As should be obvious from the discussion above, this is not exactly how we see it. Instead, we see the theory of linguistic variation which is developed within the minimalist framework as a refinement of the “GB-style Principles-and-Parameters architecture” (see also Biberauer (2008b) for essentially the same proposal).

One reason why Boeckx dismisses GB-style parameters, thus concurring with Newmeyer (2004, 2005, 2006) (discussed above in section 2.3), Culicover (1999), and Jackendoff (2002), is that he thinks of the parameters as being, by definition, principles which come as an addition to the set of universal principles of UG, but with the difference that they specify a range of options for some grammatical property. In this view parameters make UG larger, with more specifications, and thereby also make it more specific to the language faculty, in apparent conflict with the minimalist programme of reducing UG, as far as possible, to third factor effects, as discussed in

---

29 The concept “F and G form a feature bundle” can also be expressed in the restriction, as follows (where “S” is again the predicate “is a sublabel of”):

(i) $\exists f \in F \exists g \in F \ [ S(f, C) \text{ and } S(g, C) ]$. 

---
the previous section. It also raises the question how such a rich UG could have evolved (as pointed out by Boeckx, ibid.).

But although the view of parameters as constituting an additional set of specifications to UG is often voiced in textbook presentations of parametric theory, it is not the only view, and it is not inherent to P&P theory. An alternative is that parameters are just those grammatical options which are not specified by UG, as we have tried to specify in the preceding sections. An obvious example is the traditional head-complement parameter. In the case of this parameter the values (head precedes complement or head follows complement) are not given by properties of UG, but instead follow from the ultimately physical fact that words must be linearly ordered, allowing exactly those two options (as Boeckx notes, in fact). However, this does not make it any less of a parameter in the P&P sense, as long as it remains true that there is a finite range of options which are left open by UG in that UG does not make the choice for the learner. Another classical parameter of GB theory which clearly has this character is the wh-parameter of Huang (1982):

(59) Wh-movement takes place before/after S-structure.

Here UG prescribes that wh-movement happens whenever a wh-phrase is selected from the lexicon, but does not specify when. Given the GB-model, and given the semantic properties of wh-expressions, there are logically two possibilities: before S-structure or after, on the LF-side (in Huang’s (1982) terms “in Syntax or in LF”). In this case the options are given by independently motivated architectural properties of
the system. Specifying the options as in (59) is done in the name of explicitness, presumably, but is in fact redundant, as should be obvious to a critical reader.

Disagreement between us and Boeckx as regards the historical relation between GB-theory and minimalism is of limited importance, though, in the context of this volume. What is more important is whether we agree on the substantial issues concerning the nature of UG and the explanation of variation. One substantial, empirical claim that Boeckx (ibid.) makes, which we do not agree with, is that all parameters (he refers to them as ‘nano-parameters’) are independent; hierarchies of parameters and macroparameters do not exist. But the hierarchies we have proposed are determined by third-factor principles, as we have seen, and so there is no cost to UG in proposing these. Unaltered, Boeckx’s view will either be descriptively inadequate (too few parameters to account for the attested variation) or predict astronomical numbers of unattested systems, for the reasons alluded to in §3.2. So we see that the opposite view should be taken seriously.

An even more substantial point on which we agree with Boeckx is that “the minimalist program offers us a different, more adequate way of exploring how principles and parameters may interact” (Boeckx’s Note 2, p. 2). We hope to have demonstrated this in the preceding sections. See also Biberauer (2008b).

4. Conclusion

The above discussions, combined with the papers to follow and those collected in Biberauer (2008a) and Holmberg (2008), attest to the continuing validity of the
principles-and-parameters approach to UG. Naturally, this approach has changed significantly over the thirty years since it was first suggested (in Rizzi (1978)). However, the essential idea can still provide a way to resolve the traditional tension between description and explanation in comparative syntax. Moreover, we have seen that, by attempting to eliminate the same tension at the parametric level (caused by a proliferation of highly specific microparameters), we can arrive at interesting characterisations of epigenetic parameter-setting, parameter schemata/networks, and the relation between macro- and microparameters. A central concept in all of this is markedness, which may be largely determined by third-factor considerations of computational efficiency. As we already said, we are beginning to see how comparative syntax might look, beyond explanatory adequacy.