Principles and Parameters of Universal Grammar

Citation

Published Version

Permanent link
https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37367394

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
CHAPTER 14

PRINCIPLES AND PARAMETERS OF UNIVERSAL GRAMMAR

C.-T. JAMES HUANG AND IAN ROBERTS

14.1 Introduction

The Principles and Parameters Theory (P&P), which took shape in the early 1980s, marked an important step forward in the history of generative grammatical studies. It offered a plausible framework in which to capture both the similarities and differences among languages within a rigorous formal theory. It led to the discovery of important patterns of variation across languages. Most important of all, it offered an explanatory model for the empirical analyses which opened a way to meet the challenge of ‘Plato’s Problem’ posed by children’s effortless—yet completely successful—acquisition of their grammars under the conditions of the poverty of the stimulus (see chapters 5, 10, 11, and 12).

Specifically, the P&P model led linguists to expand their scope of inquiry and enabled them to look at an unprecedented number of languages from the perspective of the formal theory of syntax, not only in familiar traditional domains of investigation; it also opened up some new frontiers, at the same time raising new questions about the nature of language which could not even have been formulated earlier. Another consequence was that it became possible to discover properties of one language (say English) by studying aspects of a distinct, genetically unrelated language (say, Chinese or Gungbe), and vice versa.

Most of the original proposals for parameters in the early days of P&P were of the form that we would now, with the benefit of hindsight, think of as macroparameters. They have the characteristic property of capturing the fact that parametric variations occur in clusters. As the theory developed, it became clear that such a model is

---

1 This work was partly supported by the ERC Advanced Grant 269752 Rethinking Comparative Syntax (ReCoS), Principal Investigator: I. Roberts.
inadequate for the description of micro-scale parametric variation across languages. In addition, certain correlations that were predicted by proposed well-known macroparameters turned out not to hold as more languages were brought into consideration. In the meantime, considerations of theoretical parsimony led to widespread adoption of the lexical parameterization hypothesis (known now as the Borer–Chomsky conjecture), ruling out much of the theoretical vocabulary used in earlier macroparametric proposals. These developments led to some doubts about the existence of macroparameters, and even the feasibility of the P&P program; see in particular Newmeyer (2005).

In this chapter, developing recent work, we will support the position that both macroparameters and microparameters exist (and indeed other levels of parametric variation—see section 14.5), and that there is really no adequate alternative to a parameter-setting model of language acquisition. Consistent with the ‘three-factors’ conception of language design (Chomsky 2005 and chapter 6), parametric variation can be seen as an emergent property of the three factors of language design (Biberauer 2011, Roberts and Holmberg 2010, Roberts 2012, and references given there): the first factor is a radically underspecified UG, the second the Primary Linguistic Data (PLD) for language acquisition (see chapters 5, 10, 11, and 12) and the third general learning strategies based on computational conservatism (see chapter 6). Using the facts of Chinese as a paradigm case (i.e., the macroparametric contrasts with English, macroparametric changes since Old/Archaic Chinese, and microvariation among dialects), we show that both macroparameters and microparameters exist, and the tension between descriptive and explanatory adequacy is resolved by the view that macroparameters are aggregates of microparameters acting in concert, with correlating values as driven by the third-factor learning strategies (Roberts and Holmberg 2010, Roberts 2012).

**14.2 The Principles and Parameters Theory**

Principles-and-Parameters theory emerged as a way of tackling what Chomsky (1986b) referred to as ‘Plato’s Problem.’ This is the basic observation that children acquire the intricacies of their native language in early life with little apparent effort and confronted with the impoverished stimulus (see chapters 5, 10, 11, and 12 for more details). As an illustration of the complexity of the task of language acquisition, consider the following sentences (this exposition is largely based on Roberts 2007:15–14):

(1) a. The clowns expect (everyone) to amuse them.
   b. The clowns expected (everyone) to amuse themselves.

If *everyone* is omitted in (1a), the pronoun *them* cannot correspond to *the clowns*, while if *everyone* is included, this is possible. If we simply change *them* to the reflexive pronoun
themselves, as in (1b), exactly the reverse results. In (1b), if everyone is included, the pronoun themselves must correspond to it. If everyone is left out, themselves must correspond to the clowns. The point here is not how these facts are to be analyzed, but rather the precision and the subtlety of the grammatical knowledge at the native speaker’s disposal. It is legitimate to ask where such knowledge comes from.

Another striking case involves the interpretation of missing material, as in (2):

(2) John will go to the party, and Bill will—too.

Here there is a notional gap following will, which we interpret as go to the party; this is the phenomenon known as VP-ellipsis. In (3), we have another example of VP-ellipsis:

(3) John said he would come to the party, and Bill said he would—too.

Here there is a further complication, as the pronoun he can, out of context, correspond to either John or Bill (or an unspecified third party). Now consider (4):

(4) John loves his mother, and Bill does—too.

Here the gap is interpreted as loves his mother. What is interesting is that the missing pronoun (the occurrence of his that isn’t there following does) has exactly the three-way ambiguity of he in (3): it may correspond to John, to Bill or to a third party. Example (4) shows we have the capacity to apprehend the ambiguity of a pronoun which we cannot hear. Again, a legitimate and, it seems, profound question is where this knowledge comes from.

The cases just discussed are examples of native grammatical knowledge. The basic point in each case is that native speakers of a language constantly hear and produce novel sentences in that language, and yet are able to distinguish well-formed sentences from ill-formed ones and make subtle interpretative distinctions of the kind illustrated in (4). The existence of this kind of knowledge is readily demonstrated and not in doubt. But it raises the question of the origin: where does this come from? How does it develop in the growing person? This is Plato’s problem, as Chomsky called it, otherwise known as the logical problem of language acquisition (Hornstein and Lightfoot 1981). It is seen as a logical problem because there appears to be a profound mismatch between the richness and intricacy of adult linguistic competence, illustrated by the examples given in (1), and the rather short time taken by language acquisition coupled with small children’s seemingly limited cognitive capacities.

This latter point brings us to the argument from the poverty of the stimulus. Here we briefly summarize this argument (for a more detailed presentation, see chapter 10, as well as Smith [1999:40–41], Jackendoff [2003:82–87], and, in particular, Guasti [2002:5–18]). As its name implies, the poverty-of-the-stimulus argument is based on the observation that there is a significant gap between what seems to be the experience facilitating first-language acquisition (the input or ‘stimulus’) and the nature of the linguistic knowledge
which results from first-language acquisition, i.e., one’s knowledge of one’s native language. The following quotation summarizes the essence of the argument:

The astronomical variety of sentences any natural language user can produce and understand has an important implication for language acquisition … A child is exposed to only a small proportion of the possible sentences in its language, thus limiting its database for constructing a more general version of that language in its own mind/brain. This point has logical implications for any system that attempts to acquire a natural language on the basis of limited data. It is immediately obvious that given a finite array of data, there are infinitely many theories consistent with it but inconsistent with one another. In the present case, there are in principle infinitely many target systems … consistent with the data of experience, and unless the search space and acquisition mechanisms are constrained, selection among them is impossible…. No known ‘general learning mechanism’ can acquire a natural language solely on the basis of positive or negative evidence, and the prospects for finding any such domain-independent device seem rather dim. The difficulty of this problem leads to the hypothesis that whatever system is responsible must be biased or constrained in certain ways. Such constraints have historically been termed ‘innate dispositions,’ with those underlying language referred to as ‘universal grammar’ (Hauser, Chomsky, and Fitch 2002:1576–1577).

Hence we are led to a biological model of grammar. The argument from the poverty of the stimulus leads us to the view that there are innate constraints on the possible form a grammar of a human language can take; the theory of these constraints is Universal Grammar (UG). But of course it is clear that experience plays a role; no one is suggesting that English or Chinese are innate. So UG provides some kind of bias, limit, or schema for possible grammars, and exposure to people speaking provides the experience causing this latent capacity to be realized as competence in a given actual human language. Adult competence, as illustrated for English speakers by the data such as that in (1–4), is the result of nature (UG) and nurture (exposure to people speaking).

The P&P model is a specific instantiation of this general approach to Plato’s problem. The view of first-language acquisition is that the child, armed with innate constraints on possible grammars furnished by UG, is exposed to Primary Linguistic Data (PLD, i.e., people speaking) and develops its particular grammar, which will be recognized in a given cultural context (e.g., in London, Boston, or Beijing) as the grammar of a particular language, English or Chinese. But it should be immediately apparent that London and Boston, or Beijing and Taipei, are not linguistically identical. Concepts such as ‘English’ and ‘Chinese’ are highly culture-bound and essentially prescientific. An individual’s mature competence, the end product of the process of first-language acquisition just sketched, is not really ‘English’ or ‘Chinese,’ but rather an individual, internal grammar, technically an I-grammar. We use the terms ‘English’ or ‘Chinese’ to designate different variants of I-grammar, but these terms are really only approximations (as are more narrowly defined terms such as ‘Standard Southern British English’ or ‘Standard Northern Mandarin Chinese,’ neither exactly corresponds to the I-grammar of Smith, Roberts, Li, or Huang).
What makes the different I-grammars of, to revert to prescientific terms for convenience, English or Chinese? This is where the notion of parameters of UG comes in. Since UG is an innate capacity, it must be invariant across the species: Smith, Roberts, Li, and Huang (as well as Saito, Rizzi, and Sportiche) are all the same in this regard. But these individuals were exposed to different forms of speech when they were small and hence reached the different final states of adult competence that we designate as English, Chinese, Japanese, Italian, or French. These cognitive states are all instantiations of UG, but they differ in parameter-settings, abstract patterns of variation in the restricted set of grammars allowed by UG. So, on the Principles and Parameters conception of UG, differing PLD was sufficient to cause Roberts to set his UG parameters one way, so as to become an English speaker, while Huang set his another way and became a Chinese speaker, Saito another way, Rizzi still another, and so forth. On this view language acquisition is seen as the process of fixing the parameter values left open by UG, on the basis of experience determined by the PLD.

The P&P model is a very powerful model of both linguistic diversity and language universals. More specifically, it provides a solution to Plato’s problem, the logical problem of language acquisition, in that the otherwise formidable task of language acquisition is reduced to a matter of parameter-setting. Moreover, it makes predictions about language typology: parameters make predictions about (possible) language types, as we will see in more detail in section 14.3 (see also chapter 15). Furthermore, it sets the agenda for research on language change, in that syntactic change can be seen as parameter change (see chapter 18). Finally, it draws research on different languages together as part of a general enterprise of discovering the precise nature of UG: we can discover properties of the English grammatical system (a particular set of parameter values) by investigating Chinese (or any other language), without knowing a word of English at all (and vice versa, of course). Let us now begin to look at the progress that has been made in this endeavor in more detail.

14.3 Principles and Parameters in GB

In this section, we will briefly review some notable examples of parameters that were put forward in the first phase of research in the P&P model in the 1980s, using the general framework of Government–Binding (GB) theory.

14.3.1 The Head Parameter

First proposed in Stowell (1981), and developed in Huang (1982), Koopman (1984), and Travis (1984), the head parameter can be stated as follows:

(5) In X′, X {precedes/follows} its complement YP.
This parameter regulates one of the most pervasive and well-studied instances of cross-linguistic variation: the variation in the linear order of heads and complements. Stated as (5), it predicts that all languages will be either rigidly head-initial (like English, the Bantu languages, the Romance languages, and the Celtic languages, among many others) or rigidly head-final (like Japanese, Korean, the Turkic languages, and the Dravidian languages). Of course many languages, including notably Chinese, show mixed, or dis-harmonic word order, suggesting that (5) needs to be relativized to categories, a matter we return to in section 14.4.1.

The simplest statement of this parameter is along the lines of (5), which assumes, as was standard in GB theory, that linear precedence and hierarchical relations (defined in terms of X′-theory) are entirely separate. In fact, X′-theory was held to be invariant, a matter of UG principles (or deriving from UG principles), while linear order was subject to parametric variation. Since Kayne (1994), other approaches to linearization have been put forward (starting with Kayne’s Linear Correspondence Axiom), some of which, like Kayne’s, connect precedence and hierarchy directly. The head parameter must then be reformulated accordingly. In Kayne’s (1994) approach, for example, complement–head order cannot be directly generated, but must be derived by leftward movement (in the simplest case, of complements). The parameter in (5) must therefore be restated so as to regulate this leftward movement. Takano (1996), Fukui and Takano (1998), and Haider (2012:5), on the other hand, propose that complement–head order is the more basic option, with surface head–complement order being derived by head movement. In that case, (5) may be connected to head movement (and the availability of landing sites for such movement, according to Haider).

14.3.2 The Null Subject Parameter

The basic observation motivating the postulation of this parameter is that some languages allow a definite pronominal subject of a finite clause to remain unexpressed, while others always require it to be expressed as a nominal bearing the subject function. Traditional grammars of languages such as Latin and Greek relate this to the fact that personal endings on the verb distinguish the person and number of the subject, thereby making a subject pronoun redundant. Languages that 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 2010a), the Celtic languages, with certain restrictions in the case of Modern Irish (see McCloskey and 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 (these appear to be ‘partial’ null subject languages in the sense of Holmberg, Nayudu & Sheehan 2009, Holmberg 2010b; see Duguine and Madariaga 2015 on Russian). Indeed, it seems that languages that allow null subjects are significantly more widespread than those which do not (Gilligan 1987, cited in Newmeyer 2005:85).
Since Rizzi (1986), it has been widely assumed that the null subject parameter involves the ability of Infl (or T, or AgrS) to license a null pronoun, pro, and so can be stated as in (6):

(6) T [licenses/does not license] pro in its Specifier.

Perlmutter (1971) observed that languages that allow null subjects also allow wh-movement of the subject from a finite embedded clause across a complementizer (this observation has since become known as ‘Perlmutter’s generalization’). Rizzi (1982) linked this to the possibility of so-called ‘free inversion,’ leading to the following parametric cluster:

(7) a. The possibility of a silent, referential, definite subject of finite clauses.
   b. ‘Free subject inversion.’
   c. The absence of complementizer–trace effects.

Rizzi showed that Italian has all of these properties while English lacks all of them. As with the head parameter, though, this cluster has empirical problems; see Gilligan (1987), Newmeyer (2005), and section 14.4.1.

14.3.3 The Null Topic Parameter

Huang (1984) observed that certain languages allow arguments to drop if they are construed as topics. In Chinese, a question about the whereabouts of Lisi or whether anyone has seen him, may be answered by either of the sentences in (8):

(8) a. Zhangsan kanjian-le.
    Zhangsan see-PERF
    ‘Zhangsan saw [him].’
   b. Zhangsan shuo ta mei kanjian.
    Zhangsan say he not see
    ‘Zhangsan said that he didn’t see [him].’

Huang argued that the understood object in each case is first topicalized before it drops. This conception is supported by parallel facts in German. Thus, a similar question about Lisi can be answered by either of (9) (see Ross 1982 and Huang 1984 for more examples):

(9) a. [e] hab’ ich schon gesehen. ‘I saw [him].’
    Have I already seen.
    ‘I have already seen him.’
b. [e] hab' ich in der Bibliothek gestern gesehen.
   'I saw him at the library yesterday.'

Note that the missing pronoun *ihn* ‘him’ (referring to *Lisi*) is licensed by virtue of being in the first (hence topic) position, as witnessed by the ill-formedness of *ich hab' [e] schon gesehen* where the topic position is filled by *ich*. The missing argument is thus not licensed by any formal feature of T (as it is in the case of null subjects). The Null Subject Parameter and the Null Topic Parameter thus jointly distinguish four language types.

(10) a. [null subject, –null topic]: Italian, Spanish, etc.
   b. [null subject, +null topic]: Chinese, Japanese, European Portuguese, etc.
   c. [–null subject, –null topic]: English, Modern French, etc.
   d. [–null subject, +null topic]: German, Swedish, etc.

(See Raposo 1985 on European Portuguese, and Sigurðsson 2011b on Swedish and Icelandic).

14.3.4 The Wh-Movement Parameter

This parameter, first proposed in Huang (1982), regulates the option of preposing a wh-constituent or leaving it in place (‘in situ’) in wh-questions. English is a language which requires movement of such constituents, as shown in (11a), while Chinese and Japanese are standard examples of ‘wh-in-situ’ languages, illustrated by (11b,c):

(11) a. What did John eat t_{what}?
   
   b. Hufei chi-le sheme (ne)
      Hufei eat-asp what Qwh
      ‘What did Hufei eat?’ (Cheng 1991:112–113)
   
   c. John-ga dare-o butta-ka?
      John-NOM what-ACC hit-Q
      ‘Who did John hit?’ (Baker 2001:184)

(11c) shows the standard, neutral SOV order of Japanese (cf. *John-ga Bill-o butta* ‘John hit Bill’ [Baker 2001], while (11a) illustrates that in English the object wh-constituent what is obligatorily fronted to the SpecCP position and (11b) illustrates wh-in-situ in SVO Chinese. To be more precise, English requires that exactly one wh-phrase be fronted in wh-questions. In multiple wh-questions, all wh-phrases except one stay in situ (and there are intricate constraints on which ones can or must be moved, as well as how they are interpreted in relation to one another). Some languages require all wh-expressions
to move in multiple questions. This is typical of the Slavonic languages, as (12), from Bulgarian (Rudin 1988), shows:

(12) Koj kogo e vidjal?
who whom aux saw-3s
‘Who saw whom?’

There appears to be a further dimension of variation here (but see Bošković 2002 for a different view).

14.3.5 The Nonconfigurality Parameter

This parameter was put forward by Hale (1983) to account for a range of facts in languages that show highly unconstrained (sometimes rather misleadingly referred to as ‘free’) word order, such as Warlbiri and other Australian languages, as well as Latin and other conservative Indo-European languages. Hale’s proposal was that the phrase structure of such languages was ‘flat’ (i.e., it did not show the ‘configurational’ pattern familiar from languages such as English). This accounts directly for the ‘free’ word order of these languages, as well as the existence of ‘discontinuous constituents,’ i.e., cases where a nominal modifier may be separated from the noun it modifies by intervening material which is clearly extraneous to the NP (or DP). Hale (1983) connected two further properties, the extensive use of null anaphora and the availability of A-movement operations such as passive, to this parameter. The precise formulation of the parameter was as follows:

(13) a. In configurational languages, the projection principle holds of the pair (LS, PS).
    b. In nonconfigurational languages, the projection principle holds of LS alone.

Here ‘LS’ refers to Lexical Structure, a level of representation at which the lexical requirements of predicates are represented, and ‘PS’ refers to standard phrase structure. The projection principle requires lexical selection (c-selection and/or s-selection) properties of predicates to be structurally represented. Hence in nonconfigurational languages, according to this approach, phrase structure does not have to directly instantiate argument structure, with the consequence that arguments can be freely omitted, there are no structural asymmetries among arguments and no syntactic operations ‘converting’ one grammatical function into another. Hale (1983) argued for a number of other consequences of this parameter, focusing in particular on Warlbiri.

14.3.6 The Polysynthesis Parameter

This parameter was argued for at length in Baker (1996). In fact, it can be broken up into two distinct parts. One aspect of it has to do with whether a language requires all arguments to show overt agreement with the main predicate (usually a verb); Baker (1996:17)
formulates this in terms of whether a language requires its arguments to be morpho-
logically or syntactically visible for θ-role assignment. A further option is whether a lan-
guage allows (robust) noun incorporation. English and Chinese allow neither of these
options, while Mohawk allows both. Navajo has the former but not the latter property.
Noun incorporation is restricted to languages that satisfy the visibility requirement
morphologically (Baker 1996:18), and so there are predicted to be no languages which
have noun incorporation without fully generalized agreement. Baker connects the fol-
lowing cluster of properties to the polysynthesis parameter (as well as a further six, see

\[(14)\]
\[\begin{align*}
  a. & \text{ Syntactic Noun Incorporation;} \\
  b. & \text{ obligatory object agreement;} \\
  c. & \text{ free pro-drop;} \\
  d. & \text{ free word order;} \\
  e. & \text{ no NP reflexives;} \\
  f. & \text{ no true quantifiers;} \\
  g. & \text{ no true determiners.}
\end{align*}\]

Baker’s parameter gives an elegant account of the major typological differences between
languages of the Mohawk type (known as head-marking nonconfigurational languages)
and those of the English/Chinese type.

14.3.7 The Nominal Mapping Parameter

This parameter was put forward by Chierchia (1998a,b), and concerns, as its name
implies, an aspect of the mapping from syntax to semantics. Chierchia observes that two
features characterize the general semantic properties of nominals across languages: they
can be argumental or predicative, or \([±\text{arg(ument)}], [±\text{pred(icate)}]\), reflecting the general
fact that nominals can function as arguments or predicates, as in \(\text{John}_{\text{arg}}^{} \text{is } [\text{ a doctor}_{\text{Pred}}^{}].\)
The parametric variation lies in which of the three possible combinations of values of
these features a given language allows (the fourth logical possibility, negative values for
both features, is ruled out as nominals of this sort would have no denotation at all).

In a \([+\text{arg}, \text{–pred}]\) language, every nominal is of type \(<\text{e}>\), i.e., nominals denote indi-
viduals rather than predicates. Languages with this parameter setting have the following
properties (Chierchia 1998b:354):

\[(15)\]
\[\begin{align*}
  i. & \text{ Generalized bare arguments;} \\
  ii. & \text{ the extension of all nouns is mass;} \\
  iii. & \text{ no plural marking;} \\
  iv. & \text{ generalized classifier system.}
\end{align*}\]
Languages with this value for the nominal-mapping parameter include Chinese and Japanese. Nominals appear as bare arguments in these languages as a direct consequence of being of type \(e\); hence count nouns can function directly as arguments with no article or quantifier (giving the equivalent of I saw cat, meaning ‘I saw a/the cat(s)’). Chierchia argues that all nouns have fundamentally mass denotations, so unadorned nouns will have property (ii); more generally, there is no mass-count distinction in these languages. Further, since mass nouns cannot pluralize, there is no plural marking and, finally, special devices have to be used in order to individuate noun denotations for counting; this is what underlies the classifier system.

In a \([-\text{arg},+\text{pred}]\) language, on the other hand, all nominals are predicates (type \(<e.t>\)). It follows that bare nouns can never be arguments. This, as is well known, is the situation in French and, with certain complications, the other Romance languages (see Longobardi 1994). Such languages can have plural marking and lack classifiers.

Finally, \([+\text{arg},+\text{pred}]\) languages allow mass nouns and plurals as bare arguments, but not singular count nouns, have plural marking, and lack classifiers. (Singular bare count nouns can function as predicates as in We elected John president). This is the English, and, more broadly, Germanic, setting for this parameter.

14.3.8 The Relativized X-Bar Parameter

Fukui (1986) presents a general theory of functional categories, arguing that only these categories project above X’, and hence only these categories have Specifiers. He further proposes that functional categories can be absent as a parametric option. He analyzes Japanese as lacking D and C, and having a very defective I (or T), in particular in lacking agreement features. It follows that Japanese has no landing site for wh-movement (and hence is a wh-in-situ language, as we saw in (11c)), no dedicated subject position of the English type and the concomitant possibility of multiple nominative (-ga) marked arguments in a single clause, no position in nominals for articles and the possibility of multiple genitive (-no) marked nominals inside a single complex nominal and of stacked appositive relatives.

14.3.9 Parametric Typology

The parameters we have briefly reviewed can be put together to give a characterization of the major grammatical properties of different languages, as shown in Table 14.1.

Table 14.1 illustrates, albeit in a rather approximate and (in certain cases, e.g., Chinese is head-final in DP) debatable form, how a reasonable number of parameters can give us a synoptic and highly informative characterization of the salient grammatical features of a system. Note that our three languages here all differ for their values of each parameter discussed, except polysynthesis (which of course has a distinct value in languages such as Mohawk). An approach of this general kind, known as the Parametric Comparison
Method, has been developed in detail by Giuseppe Longobardi and his associates; see in particular Gianollo, Guardiano, and Longobardi (2008), Longobardi (2003, 2005), Colonna et al. (2010) and chapter 16, especially Figure 16.1.

### 14.4 Macroparameters, Microparameters, and Parametric Clusters

#### 14.4.1 Macroparameters and Clustering

Most of the parameters proposed in the GB era, of which those discussed in the previous section are a representative sample, have the character of being *macroparameters*, in that their effects are readily observed across the board in almost any sentence in any language. This can be easily observed in the case of the head parameter and the non-configurationality parameter, but any finite clause with a definite pronominal subject can express the null subject parameter, any *wh*-interrogative expresses the *wh*-parameter, any realization of arguments expresses the polysynthesis parameter, any nominal containing a singular count noun the nominal mapping parameter and any nominal or clause Fukui’s functional-category parameter. The effects of these parameters are thus pervasive. This also means that their settings are salient in the PLD, making them, presumably, easy for acquirers to observe and thereby fix (see chapters 11 and 12). This also means that observed variations are predicted to typically cluster together. For example, the Head parameter (all else being equal) predicts that V-final, N-final, P-final, and A-final orders will all co-occur (in addition to, depending on what one assumes about functional categories, T-final, C-final, and D-orders). As noted in section 14.3.2, the classical null subject parameter predicts the clustering of null subjects, free inversion, and apparent long subject-extraction as in (7) (as well as, possibly, differences between French and Italian long clitic-climbing and infinitival V-movement [Kayne 1989, 1991]). Chierchia’s nominal mapping parameter predicts the cluster of surface properties given in (15) and Baker’s

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Chinese</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Japanese</td>
<td>Yes</td>
<td>Yes</td>
<td>??</td>
<td>No</td>
<td>??</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 14.1 Summary of values of parameters discussed in this section for English, Chinese, and Japanese
Polysynthesis Parameter in (14). Similarly, the DP/NP parameter, more recently proposed by Bošković (2008), predicts that left branch extraction (as in *Whose did you read book?*), adjunct extraction from NP, scrambling, adnominal double genitives, superlatives with a ‘more than half’ reading, and other properties cluster together. This, as was first pointed out in Chomsky (1981a), gives macroparameters their potential explanatory value: an acquirer need only observe one of the clustering properties which express the parameter to get all the others ‘for free,’ as an automatic consequence of their UG-mandated clustering. For example, merely recognizing that finite clauses allow definite pronominal subjects not to appear overtly automatically guarantees that the much more recondite property of long *wh*-extraction of subjects over complementizers is thereby acquired. In this way, the principles and parameters approach brought biolinguistics and language typology together; this point emerges particularly clearly when groups of parameters are presented together as in Table 14.1 and, much more strikingly, Figure 16.1.

However, since the mid-1980s it has gradually emerged that there are problems with the conception of macroparameters. These are of both a theoretical and an empirical nature. On the empirical side, it has emerged that many of the typological predictions made by macroparameters are not borne out. This is particularly clear in the case of the Head Parameter which, formulated as in (5), predicts that all languages will be either rigidly, harmonically head-initial or rigidly, harmonically head-final. It is, of course, well known that this is not true: German, Mandarin, and Latin are all clear examples of very well-studied languages which show disharmonic orders (and see Cinque 2013 for the suggestion that fully harmonic systems may be very rare). However, at the same time it is not true that just anything goes: on the one hand, languages tend toward cross-categorial harmony (as first shown in detail in Hawkins 1983; see also Dryer 1992 and chapter 15); second, there appear to be general constraints on possible combinations of head-initial and head-final structures (see for example Biberauer, Holmberg, and Roberts 2014). Concerning the predictions made by the putative cluster associated with the classical null subject parameter, see the extensive critique in Newmeyer (2005), and the response in Roberts and Holmberg (2010). Similar comments could be made about the other parameters listed in section 14.3.

From a theoretical point of view, there are two basic problems with macroparameters. First, they put an extra burden on linguistic theory, in that they have to be stated somewhere in the model. The original conception of parameters as variable properties associated with invariant UG principles dealt elegantly with this question, but most of the parameters listed in section 14.3 do not seem to be straightforwardly formulable in this way. Second, it is not clear why just these parameters are what they are; there is, in other words, a certain arbitrariness in where variation may or may not occur which is not explained by any aspect of the theory.

In short, macroparameters, while having great potential merit from the perspective of explanatory adequacy, have often fallen short in descriptive terms by making excessively strong empirical predictions. Moreover, there has been no natural intensional characterization of the notion of what a possible macroparameter can be, rendering their theoretical status somewhat questionable.
14.4.2 \textit{Microparameters}

Here the key theoretical proposal is the lexical parameterization hypothesis (Borer 1984; Chomsky 1995b). This can be thought of, following Baker (2008a:3, 2008b:155–156), as the ‘Borer–Chomsky conjecture,’ or BCC:

\begin{equation}
(16) \text{All parameters of variation are attributable to differences in the features of particular items (e.g., the functional heads) in the Lexicon.}
\end{equation}

More precisely, we can restrict parameters of variation to a particular class of features, namely formal features in the sense of Chomsky (1995b) (Case, $\phi$, 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. First, it is clearly a highly restrictive theory: reducing all parametric variations to features of (functional) lexical items means that many possible parameters simply cannot be stated. An example might be a putative parameter concerning the ‘arity’ of Merge, i.e., how many elements can be combined by a single operation of Merge. Such a parameter might restrict some languages to binary Merge but allow others to have ternary or $n$-ary Merge, perhaps giving rise to the effects of nonconfigurationality along the lines of Hale (1983) (see section 14.3.5).

The second advantage of a microparametric approach has to do with language acquisition. As originally pointed out by Borer, ‘associating parameter values with lexical entries reduces them to the one part of a language which clearly must be learned anyway: the lexicon’ (Borer 1984:29).

Third, the microparametric approach implies a restriction on the form of parameters, along roughly the lines of (17):

\begin{equation}
(17) \text{For some formal feature } F, \text{ } P = \pm F.
\end{equation}

Here are some concrete, rather plausible, examples instantiating the schema in (17):

\begin{enumerate}
\item (18a) T is $\pm \phi$;
\item (18b) N is $\pm \text{Num}$;
\item (18c) T is $\pm \text{EPP}$.
\end{enumerate}

(18a) captures the difference between a language in which verbs inflect for person and number, such as English (in a limited way) and most other European languages, on the one hand, and languages like Chinese and Japanese, on the other, in which they do not. This may have many consequences for the syntactic properties of verbs and subjects (cf. the discussion of Japanese in Fukui 1986 mentioned in section 14.3.8). (18b) captures the difference between a language in which number does not have to be marked on (count) nouns, such as Mandarin Chinese, and one in which it does, as in English; this difference may be connected to the nominal mapping parameter (see section 14.3.7). (18c)
determines the position of the overt subject; in conjunction with V-to-T movement, a negative value of this parameter gives VSO word order, providing a minimal difference between, for example, Welsh and French (see McCloskey 1996; Roberts 2005).

This simplicity of formulation of microparameters, along with the general conception of the BCC, should be compared to the theoretical objections to macroparameters discussed at the end of the previous section. It seems clear that microparameters represent a theoretically preferable approach to the macroparametric one illustrated in section 14.3.

Fourth, the microparametric view allows us to put an upper bound on the set of grammars. Suppose we have two potential parameter values per formal feature (i.e., each feature offers a binary parametric choice as stated in (17)), then we can define the quantity $n$ as follows:

\begin{equation}
\tag{19}
n = |F|, \text{ the cardinality of the set of formal features.}
\end{equation}

It then follows that the cardinality of the set of parameter values $|P|$ is $2n$ and the cardinality of the set of grammatical systems $|G|$ is $2^n$. So, if $|F| = 30$, then $|P| = 60$ and $|G| = 2^{30}$, or $1,073,741,824$. Or if, following Kayne (2005a:14), $|F| = 100$, then $|G| = 1,267,650,600,228,229,401,496,703,205,376$. Kayne states that ‘[t]here is no problem here (except, perhaps, for those who think that linguists must study every possible language)’ (2005a:14). However, one consequence is clear: the learning device must be able to search this huge space very efficiently, otherwise selection among such a large range of options would be impossible for acquirers (see chapter 11, section 5, for the problems that this kind of space poses for ‘search-based’ parameter-setting).

It may be, though, that the observation of this extremely large space brings to light a fatal weakness of the microparametric approach. To see this, consider a thought experiment (variants of this have been presented in Roberts 2001 and Roberts 2014). Suppose that at present approximately 5,000 languages are spoken and that this figure has been constant throughout human history (back to the emergence of language faculty in modern *homo sapiens*; see the brief discussion of the evolution of language in chapter 1). Suppose further that every language changes in at least one parameter value with every generation. Then, if we have a new generation every 25 years, we have 20,000 languages per century. Finally, suppose that modern humans with modern UG have existed for 100,000 years, i.e., 1,000 centuries. It then follows that 20,000,000 languages have been spoken in the whole of human history, i.e., $10^7 \times 2$. This number is 27 orders of magnitude smaller than the number of possible grammatical systems arising from the postulation of 100 independent binary parameters.

While there are many problems with the detailed assumptions just presented (several of them related to the Uniformitarian Principle, the idea that linguistic prehistory must have been essentially similar to recorded linguistic history; see Roberts [forthcoming] for discussion and a more refined statement of the argument), the conclusion is that, if the parameter space is as large as Kayne suggests, there simply has not been enough time since the emergence of the species (and therefore, we assume, of UG) for anything other than a tiny fraction of the total range of possibilities offered by UG to be realized. This implies that we could never know whether a language of the past corresponded to
the UG of the present or not, since the overwhelming likelihood is that these languages could be typologically different from any language that existed before or since, perhaps radically so. More generally, even with a UG containing just 100 independent parameters we should expect that languages appear to ‘differ from each other without limit and in unpredictable ways’ in the famous words of Joos (1957:96). But of course, we can observe language types, and note diachronic drift from one type to another.

We conclude that, despite the clear merits of the microparametric approach, it appears that a way must be found to lower the upper bound on the number of parameters, on a principled basis.

14.5 Principles and Parameters in Minimalism

The exploratory program for linguistic theory known as the Minimalist Program (MP henceforth) has as its principal goal to go ‘beyond explanatory adequacy,’ that is, beyond explaining the ‘poverty of stimulus’ problem (see in particular Chomsky 2004a and chapters 5, 6, and 10). This goal has both theoretical and empirical aspects. On the theoretical side, the goal is to fulfill the Galilean ideal of maximally simple explanation (see also Chen-Ning Yang 1982). On the empirical side, the goal is to explain the ‘brevity of evolution’ problem. Estimates regarding the date of the origin of language vary widely, with anything between 200,000 and 50,000 years ago being proposed (Tallerman and Gibson 2012:239–245, and chapter 1). It is not necessary to take a precise view on the date of the origin of language here, because anywhere within this range is a very short period for the development of such a seemingly complex cognitive capacity. It seems that there has been little time for the processes of random mutation and natural selection to operate so as to give rise to this capacity, unless we view the origin of the language faculty as due to a relatively small set of mutations which spread through a small, genetically homogeneous population in a very short time (in evolutionary terms). Hence, from the biological or neurological perspective, the core properties of the language faculty must be rather few. Combining this with the Galilean desideratum just mentioned, we then expect UG, at least the domain-specific aspects of cognition which are essential to language, to be few and simple.

In trying to approach these goals, then, there has been an endeavor to reduce the 'size,’ complexity, and the overall contents of UG; see Mobbs (2015) for an excellent discussion and overview. One important conceptual shift in this direction was Chomsky’s (2005) articulation of the three factors of language design. These are as follows:

(20) a. Genetic endowment: UG.

b. Experience: PLD.

c. Other independent, non-domain-specific cognitive systems, such as other cognitive abilities (logical reasoning, memory), computational efficiency, minimality, and general laws of nature.
From this perspective, many things which were previously attributed directly to UG as principles of grammar can be ascribed to the third factor (see in particular chapter 6 for discussion). Regarding the question of parametric variation, since there are few or no UG principles to be parametrized along the earlier, GB-style lines, all parameters must be stated as microparameters, and indeed in general the BCC has been the dominant view of where parameters fit into a minimalist approach (see in particular Baker 2008b).

More generally, the nature of the rather speculative and, at least in principle, restrictive and programmatic proposals of the MP has meant in practice that there are numerous empirical problems that have been known since the GB era or before which have been largely left untouched. For example, many of the results of the intensive technical work on phenomena associated with Empty Category Principle in the GB era, particularly those developing the proposals in Chomsky (1986b), have not been carried forward, in part because some of the mechanisms and notions introduced earlier have been made unavailable (notably the various concepts of government, proper government, head/lexical government, and antecedent government; see Huang 1982, Lasnik and Saito 1984, 1992; Cinque 1991; and references given there).

To a degree, the GB notion of parameter, as summarized and illustrated in section 14.3, has suffered a similar fate. Traditional macroparameters cannot be stated within Minimalist vocabulary, and so all parametric variation must be seen as microparametric variation, stated as variations in the nature of formal features of individual functional categories. So the ‘traditional’ macroparameters are completely excluded as such. This, combined with the empirical problems associated with clusters discussed in section 14.4.1, has led many to conclude that the entire P&P enterprise should have been abandoned (see especially Boeckx 2014), although no clear alternative proposals for how to deal with synchronic and diachronic linguistic diversity have emerged.

So the question that arises is whether macroparameters really exist, and if so, how they can be accommodated in a minimalist UG. Furthermore, as our brief discussion of microparameters at the end of the previous section shows, given the large number of microparameters based on individual formal features, a question we have to ask is whether Plato’s Problem arises again. How can the acquirer search a space containing $1,267,651$ trillion trillion possible grammars in the few years of first-language acquisition (see chapter 11 on the question of searching the grammatical space, and chapter 12 on the time-course of first-language acquisition)? Do we not risk sacrificing the earlier notion of explanatory adequacy in our attempt to go beyond it?

Perhaps surprisingly, these questions have not been at the forefront of theoretical discussion in the context of the MP. Nonetheless, some interesting views have been articulated recently. Here we will briefly discuss those of Kayne (2005a, 2013, i.a.), Baker (2008b), Gianollo, Guardiano and Longobardi (2008), Holmberg (2010b), Roberts and Holmberg (2010) and Biberauer and Roberts (2012, 2015a,b, forthcoming).

Kayne (2005a, 2013, i.a.) emphasizes the fact that there is no doubt as to the existence of microparameters. The particular value of this approach lies in the idea that, in looking very carefully at very closely related languages or dialects (e.g., the Italo-Romance varieties), we detect many useful generalizations that would not have been visible on a microparametric
approach. Microparametric research has two methodological advantages. On the one hand, it gives us a restrictive theory of parametric variation, as the BCC clearly illustrates (see the discussion in section 14.4.2). Second, it permits something close to a ‘controlled experiment’: by looking at very closely related varieties, we control for as many potential variable factors (which may obscure the facet of variation we are interested in) as possible, making it possible to focus on the single variant property, or at least relatively few properties of interest. To give a very simple example (for which Kayne is not responsible), if we observe differences in verb/clitic orders across two or more Romance varieties (which is very easy to do; see Kayne 1991, Roberts in press), we are unlikely to treat these as reflexes of more general (‘macro’) differences in word order parameters, since all the Romance languages share the same very strong tendency to harmonic head-initial order.

Furthermore, many or most (perhaps all) macroparameters can be broken up into microparameters. As Kayne (2013:137n23) points out:

> It might also be that all ‘large’ language differences, e.g. polysynthetic vs. non- (cf. Baker (1996)) or analytic vs. non- (cf. Huang 2010 [=2013]), are understandable as particular arrays built up of small differences of the sort that might distinguish one language from another very similar one, in other words that all parameters are microparameters [emphasis added].

This last idea was developed by Roberts (2012); see also the discussion of Biberauer and Roberts (2015a,b, forthcoming) later in this section.

Baker (2008a,b) argues for the need for macroparameters in addition to microparameters. He argues that certain macroparameters go a long way towards reducing the range of actual occurring variation:

> The strict microparametric view predicts that there will be many more languages that look like roughly equal mixtures of two properties than there are pure languages, whereas the macroparametric-plus-microparametric approach predicts that there will be more languages that look like pure or almost pure instances of the extreme types, and fewer that are roughly equal mixtures (Baker 2008b:361).

On the other hand, the macroparametric view predicts, falsely, rigid division of all languages into clear types (head-initial vs. head-final, etc.). Regarding this possibility, Baker comments (2008b:359) that ‘[w]e now know beyond any reasonable doubt that this is not the true situation.’

Baker further observes that, combining 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. And, as he points out, this often appears to be the case, for example regarding the correlation originally proposed by Greenberg (1963/2007) between verb–object order and preposition–object order. The figures from the most recent version of The World Atlas of Language Structures (WALS) are as follows (these figures leave aside a range of minority patterns such as ‘inpositions,’ languages lacking adpositions, and the cases Dryer classifies as ‘no dominant order’ in either category):
It is very clear that here we see the kind of normal distribution predicted by a combination of macro- and microparameters. Baker therefore concludes that the theory of comparative syntax needs some notion of macroparameter alongside microparameters. He also makes the important point that many macroparameters could probably never have been discovered simply by comparing dialects of Indo-European languages.

Gianollo, Guardiano, and Longobardi (2008) propose a distinction between parameters themselves, construed along the lines of the BCC, and hence microparameters, and parameter schemata (see also chapter 16, section 9). On this view, 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. They suggest the following schemata, where in each case F is a formal feature of a functional head, lexically encoded as such in line with the BCC:

\[
\begin{align*}
&\text{a. Grammaticalization: Is F grammaticalized?} \\
&\text{b. Checking: Is F, a grammaticalized feature, checked by X, X a category?} \\
&\text{c. Spread: Is F, a grammaticalized feature, spread on Y, Y a category?} \\
&\text{d. Strength: Is F a grammaticalized feature checked by X, strong? (i.e., does it overtly attract X?)} \\
&\text{e. Size: Is F a grammaticalized feature checked by a head X (or something bigger?)}
\end{align*}
\]

Gianollo, Guardiano, and Longobardi (2008:121–122) illustrate the workings of these schemata for the [definiteness] feature in relation to 47 parameters concerning internal structure of DP (e.g., is there a null article? is there an enclitic article? are demonstratives in SpecDP? do demonstratives combine with articles? What is the position of adnominal adjectives? etc.) across 24 languages (this is an example of Modularized Parametric Comparison; see chapter 16). A very important aspect of Gianollo, Guardiano, and Longobardi’s position, taken up by Roberts and Holmberg (2010) and Biberauer and Roberts (2012, 2015a,b, forthcoming) is the idea that parameters are not primitives in a minimalist system, but derive from other aspects of the system.

Holmberg (2010b:8) makes a further important observation: it is possible to consider parameters as underspecifications in UG, entirely in line with minimalist considerations. He says:

A parameter is what we get when a principle of UG is underdetermined with respect to some property. It is a principle minus something, namely a specification of a feature value, or a movement, or a linear order, etc.
(In fact, Kayne argued for this view in the 1980s; see Uriagereka 1998:539.) Roberts and Holmberg (2010:53) combine these last two ideas and suggest that the existence of parameter variation, and in fact the parameters themselves, are emergent properties, resulting from the three factors of language design given in (20). They propose that, formally, parameters involve generalized quantification over formal features, as in (23):

\[(23) \quad Q(f \in C) [P(f)] \]

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.). In these terms, one of the standard formulations of the null subject parameter as involving ‘pronominal’ \(T/\text{Infl}\), instantiates the general schema as follows:

\[(24) \quad \exists_{f \in D} [S(D, T_{\text{Fin}})] \]

(24) reads ‘For some feature \(D\), \(D\) is a sublabel of finite \(T\)’ (where ‘sublabel’ is understood as in Chomsky 1995b:268).

On this view, UG does not even provide the parameter schemata. As Roberts and Holmberg put it:

In essence, parameters reduce to the quantificational schema in [(20)], 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 generalized 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 (Roberts and Holmberg 2010:60).

The role of the second and third factors is developed and clarified in Roberts (2012) and, in particular, in Biberauer and Roberts (2012, 2015a,b, forthcoming), summarizing and developing earlier work (see the references given). The third factor principles are seen as principles manifesting optimal use of cognitive resources, i.e., general computational conservativity. In particular, the following two acquisition strategies are proposed:

\[(25) \quad \begin{align*}
&(i) \text{ Feature Economy (FE) (see Roberts and Roussou 2003:201):} \\
&\quad \text{Postulate as few formal features as possible.}
&
&(ii) \text{ Input Generalization (IG) (see Roberts 2007:275):} \\
&\quad \text{Maximize available features.}
\end{align*} \]

Biberauer and Roberts (2014:7) say:

From an acquirer’s perspective, FE requires the postulation of the minimum number of formal features consistent with the input. IG embodies the logically invalid, but
heuristically useful learning mechanism of moving from an existential to a universal
generalisation. Like FE, it is stated as a preference, since it is always defeasible by the
PLD. More precisely, we do not see the PLD as an undifferentiated mass, but we take
the acquirer to be sensitive to particular aspects of PLD such as movement, agreement,
etc., readily encountered in simple declaratives, questions and imperatives. So
we see that the interaction of the second (PLD) and third (FE, IG) factors is crucial.

The effect of parametric variation arises from this interaction of PLD and FE/IG with
the underspecification of the formal features of functional heads in UG. In further work,
Biberauer (2011) in fact suggests that the formal features themselves may represent
emergent properties, with UG contributing merely the general notion of ‘(un)interpretable
formal feature’ rather than an inventory of features to be selected from; see also
Biberauer and Roberts (2015a). This clearly represents a further step towards general
minimalist desiderata of overall simplicity, as well as arguably going beyond explana-
tory adequacy.

This emergentist approach has two interesting consequences. One is that it leads to
the postulation of a learning path along the following lines: acquirers will always by
default postulate that no heads bear a given feature F; this maximally satisfies FE and
IG. Once F is detected in the PLD, IG requires that that feature is generalized to all rel-
evant heads (of course this violates FE, but PLD will defeat the third-factor strategies).
As a third step, if a head which does not bear F is detected, the learner retreats from the
maximal generalization and postulates that some heads bear F. This creates a distinction
between the set of heads bearing F and its complement set, and the procedure is iter-
ated for the subset (this procedure is very similar to Dresher’s [2009, 2013] Successive
Division Algorithm, as well as learning procedures observed in other domains, as

Related to the NO>ALL>SOme procedure is a finer-grained distinction among
classes of parameters (originating in Biberauer and Roberts 2012), as follows:

(26) For a given value \( v_i \) of a parametrically variant feature F:
  a. Macroparameters: all heads of the relevant type, e.g., all probes, all phase
     heads, etc., share \( v_i \);
  b. Mesoparameters: all heads of a given natural class, e.g., [+V] or a core
     functional category, share \( v_i \);
  c. Microparameters: a small, lexically definable subclass of functional heads
     (e.g., modal auxiliaries, subject clitics) shows \( v_i \);
  d. Nanoparameters: one or more individual lexical items is/are specified for \( v_i \).

Biberauer & Roberts (2015b) illustrate and support these distinctions in relation to para-
metric changes in the history of English.

It is clear that the kinds of parameters defined in (26) fall into a hierarchy. Beginning with Roberts and Holmberg (2010) and developing through Roberts (2012), Biberauer and Roberts (2012, 2014, 2015a,b, forthcoming) and numerous
references given there (notably, but not only, Biberauer, Holmberg, Roberts and
Sheehan [2014] and Sheehan [2014, to appear]; see also the references at http:// recos-dtal.mml.cam.ac.uk/papers). One advantage of parameter hierarchies is that they reduce the space of possible grammars created by parameters by making certain parameter values interdependent; see Biberauer, Holmberg, Roberts, and Sheehan (2014) for more discussion. We will return to some further implications of parameter hierarchies in section 14.9.

So we see that the change in theoretical perspective brought about by the MP does not, in itself, invalidate the aims, methods, or the results achieved in the GB era, nor is it inconsistent with P&P theory, once parameters are seen as points of underspecification in UG, with other aspects of parametrization resulting from the interaction of UG so conceived with the second and third factors.

In what follows, we give a case study of parametric variation both within varieties of Chinese (synchronically and diachronically), and between (mostly Mandarin) Chinese and English. This case study is intended to provide empirical support for the following claims and proposals:

A: Both macroparameters and microparameters are needed in linguistic theory.
B: Macroparameters are simply aggregates of microparameters acting in concert on the basis of a conservative learning strategy (see the discussion of work by Biberauer and Roberts in the preceding paragraphs).
C: The (micro)parameters are themselves hierarchically organized (again see the discussion in the foregoing); we will also tentatively identify a candidate meso-parametric cluster, supporting the idea that there is hierarchy ‘all the way down.’

In the next three sections, we will develop and support each of Points A-C in turn.

## 14.6 Evidence for Macroparameters and Microparameters

### 14.6.1 Synchronic Variation: Macroparametric Contrasts between Modern Chinese and English

Modern Chinese shows a number of properties that Huang (2015) characterizes as indicating a general property of ‘high analyticity’:

(i) Chinese has light-verb constructions where English has (typically denominal) unergative intransitives:


b. English: *to fish*, *to phone*, *to sneeze* …
(ii) Chinese has ‘pseudo-incorporation’ (Massam 2001a), otherwise known as phrasal compound verbs, where English has simple transitives or intransitives:

    b. English: to fish, to feed, to skin …

(iii) Chinese typically has compound and phrasal accomplishment verbs, where English has simple verbs:

    b. English: break, etc.

(iv) Chinese requires overt classifiers for count nouns:

    b. English: three books

(v) Chinese needs overt localizers to express locations:

(31) a. Chinese: zou dao zhuozi pangbian ‘walk to table’s side’  
    b. English: walked to the table

(vi) Chinese has the canonical ‘Kaynean word order’: Subject–Adjunct–Verb–Complement:

(32) Zhangsan zuijin changchang bu neng hui jia chi fan.  
    Zhangsan recently often not can return home eat rice
    ‘Recently Zhangsan often cannot come home for dinner.’

(vii) Chinese has wh-in-situ (instead of overt wh-movement), cf. (11a,b), repeated here:

(11) a. What did John eat twhat ?  
    b. Hufei chi-le sheme (ne)  
    Hufei eat-asp what Qwh
    ‘What did Hufei eat?’

(viii) Chinese has no forms equivalent to nobody or each other:

(33) Negative quantifiers:
    a. John did not see anybody.  
    b. John saw nobody.  
    c. Zhangsan mei you kanjian renhe ren.  
       Zhangsan not have see any person  
    d. *Zhangsan kanjian-le meiyou ren.  
       Zhangsan see-PERF no person
Reciprocals:

a. They each criticized the other(s).
b. They criticized each other.
c. Tamen ge piping-le duifang. They each criticize-PERF other
d. *Tamen piping-le bici. They criticize-PERF each-other

(ix) Chinese is restricted to ‘analytic’ adverbial and adjectival modification, unlike English. Regarding adverbial modification, examples such as (35a,b) are essentially synonymous in English:

(35) a. Jennifer types fast, Dorothy drives fast, etc.
   b. Jennifer is a fast typist, Dorothy is a fast driver, etc.

On the other hand, in Chinese adverbs equivalent to English *fast* can only modify the verb, not the derived noun (see Lin and Liu 2005):

(36) a. Zhangsan shi yi-ge (da zi) da-de hen kuai de daziyuan. Zhangsan be one-CL (type) type very fast DE typist
    ‘Zhangsan is a typist who types very fast.’
b. *Zhangsan shi yi-ge hen kuai de daziyuan. Zhangsan be one-CL very fast DE typist.

Regarding adjectival modification, in English (37) is ambiguous (see Cinque 2010 for extensive discussion):

(37) Jennifer is a beautiful singer.

This example is ambiguous between the reading ‘Jennifer is beautiful and a singer,’ and ‘Jennifer sings beautifully.’ In Chinese, on the other hand, these two readings must be expressed by quite different structures, in the one case with *hen piaolang* (‘very beautiful’) modifying ‘singer,’ in the other case with it modifying ‘sing’:

(38) a. Amei shi yi-ge hen piaolang de geshou. Amei be one-CL very beautiful DE singer
    ‘Amei is a singer who is beautiful.’
b. Amei shi yi-ge chang-de hen piaolang de geshou. Amei be one-CL sing very beautifully DE singer
    ‘Amei is a singer who sings beautifully.'
(x) Chinese has no equivalents of English articles (although it has the equivalents of numeral *one* and demonstrative *this, that*).

(xi) Chinese lacks ‘coercion’ in the sense of Pustejovsky (1995). In English, a sentence like (39a) can be understood, depending on the context and what we know about John, as any of (39b–d):

(39)  
   d. John began editing a book.

On the other hand, in Chinese the equivalent of (39a) is ungrammatical; the implicit subordinate verb must be overtly expressed (see Lin and Liu 2005):

(40)  
   a. *Zhangsan kaishi yi-ben shu.*  
   Zhangsan begin one-CL book  
   ‘Zhangsan began to read a book.’
   b. Zhangsan kaishi kan yi-ben shu.  
   Zhangsan begin read one-CL book  
   ‘Zhangsan began to read a book.’
   c. Zhangsan kaishi xie yi-ben shu.  
   Zhangsan begin write one-CL book  
   ‘Zhangsan began to write a book.’
   d. Zhangsan kaishi bian yi-ben shu.  
   Zhangsan begin edit one-CL book  
   ‘Zhangsan began to edit a book.’

(xii) Chinese lacks (canonical) gapping:

(41)  
   a. John eats rice, and Bill spaghetti.  
   b. *Zhangsan chi fan, Lisi mian.*  
   Zhangsan eat rice, Lisi noodles

(xiii) Chinese has no ‘ga–no conversion,’ i.e., nominative–genitive alternation, as often found in languages with prenominal relatives. Thus in Japanese object relatives, the subject of the relative clause may be case-marked with nominative -*ga* or genitive -*no*, indicating the influence of the nominal phrase that dominates it.

(42)  
   John-ga/-no katta sakana  
   John-Nom/Gen bought fish  
   ‘the fish that John bought’
This phenomenon is commonly attributed to the ‘strong’ nominal nature of the relative-clause CP and TP (see Ochi 2001 and references there, among many others). In Chinese the subject cannot bear genitive case:

(43) Zhangsan (*de) mai de yu
    Zhangsan (*’s) bought REL fish
    the fish that Zhangsan (*’s) bought’

(xiv) Chinese lacks gerundive nominalization with a genitive subject:

(44) a. John’s buying that car was a stupid decision.
    b. Zhangsan (*de) mai na-bu che shi ge yuchunde jueding
    Zhangsan (*’s) buy that-CL car be CL stupid decision

(xv) Chinese shows a series of syntax–semantics mismatches (see Huang 1997 et seq). One famous case is when a pseudo-noun incorporation construction is separated by a low adverbial after the verb is raised:

(45) ta chi-le yi-ge zhongtou (de) fan, hai mei chi-bao.
    he eat-PERF one-CL hour (*’s) rice, still not finish
    ‘He ate for a whole hour, and is still not done.’
    (Literally: He ate a whole hour’s rice, and is still not done with eating.)

(xvi) Chinese has analytic passivization, with the so-called ‘bei passive’ being somewhat akin to the English get-passive. Instead of employing passive morphology that intransitivizes an active transitive verb, Chinese forms a passive by superimposing a semi-lexical verb bei (whose meaning approximates ‘undergo’) on the main predicate without passivizing the latter:

(46) Zhangsan bei [Lisi qipian-le liang ci]
    Zhangsan bei Lisi deceived two time
    ‘Zhangsan got twice deceived by Lisi.’

The important thing to observe here is the clustering of these sixteen properties in Chinese to the exclusion of them in English. (Other properties could be added to this list, including those related to argument structure, as argued in Huang 2006 for Mandarin resultatives, and in Lin [2001 et seq.] on noncanonical subjects and objects; see also Barrie and Li 2015 for related discussion.) Some of these properties have previously been attributed to macroparameters (e.g., the Wh-Movement Parameter and Nominal Mapping Parameters mentioned in section 14.3), but the degree of clustering shown here had not been observed prior to Huang (2005, 2015) and indicates a macro-parameter of high analyticity; following Huang (2005, 2015) this macroparameter can be opposed to Baker’s Polysynthesis Parameter (in fact, in terms of the Biberauer and Roberts-style NO>ALL>SOME learning path/parameter hierarchy, they can be seen as representing the two extreme NO vs. ALL options for some UG-underspecified
property; we develop this idea below in section 14.11). So this is a clear case of macro-parametric clustering.

14.6.2 Macroparametric Properties of Old Chinese (vs. Modern Chinese)

Rather like the contrasts between Modern Chinese and English, we can observe a number of syntactic properties which distinguish Old (or Archaic) Chinese (OC: 500 BC to AD 200) from Modern Chinese (MnC). These are as follows:

(i) OC lacks light verbs, but instead has denominalized unergative intransitives: *yu* ‘to fish’ (instead of *da yu*);
(ii) OC lacks pseudo-incorporation: *fan* ‘have rice’ (instead of *chi fan* ‘eat rice’);
(iii) OC has simplex accomplishments: *po* ‘break’ (instead of *da-po* ‘make break’);
(iv) OC does not have overt classifiers for count nouns: *san ren* ‘3 persons,’ *er yang* ‘two sheep’ (see Peyraube 1996 among others);
(v) OC does not have overt localizers, as illustrated in the famous line from the Confucian Analects (Peyraube 2003, Huang 2009 for other examples):

(47) 八侑舞於庭, 是可忍也, 聽不可忍也? (論語: 八侑)

bayu wu yu ting, shi ke ren ye, shu bu ke ren ye?
‘To hold the 8x8 court dance in his own court, if this can be tolerated, what else cannot be tolerated?’

Note *yu ting* ‘in the court,’ instead of *yu ting-zhong* ‘at court’s inside.’

(vi) OC has passive-like sentences that are arguably derived by NP-movement:

(48) 劳心者治人, 劳力者治于人。（孟子·滕文公）

laoxinzhe zhi ren, laolizhe zhi yu ren
‘Mental workers govern people, physical works govern by others; Mental workers govern people; physical works are governed by people.’

(vii) OC has overt *wh*-movement (although to an apparently clause-medial rather than a left-peripheral position):

(49) 吾谁欺? 欺天乎? (論語: 子)

wu shei qi, qi tian hu? (Analects: Zihan)
‘Who do I deceive? Do I deceive the Heavens?’
(viii) OC relatives involve operator movement of a relative pronoun, the particle *suo*:

(50)  魚,我所欲也; 熊掌,亦我所欲也。 (孟子: 告子)

> yu, wo suō yu ye; xiongzhāng, yì wǒ suō yu ye.

′Fish is what I want; bear-paw is also what I would like to have.′

(ix) OC had focus movement: an object focused by *wei* ′only′ is preposed to a Spec,FocusP position:

(51) 唯命是从 (左傳: 邑公12年)

> wei mìng shì cong ( Zuozhuan: Zhaogong 12)

′only the order have I followed′

(x) OC allowed postverbal adjuncts:

(52) 易之以羊 (孟子: 梁惠王上)

> yì zhī yī yáng (Mencius: Lianghuiwang I)

′replace it with a sheep′

(xi) OC shows canonical gapping, as shown by Wu (2002):

(53) 為客治飯而自藜藿。 (淮南子: 話林)

> wei ke zì fàn ěr zi_ Ø_ lihuo (Huainanzi: Shuolin)

′For guests cook rice, but for oneself [cook] grass.′

(xii) OC exhibits nominative–genitive alternation in prenominal relatives. In (54) we have two (free) relative clauses whose subjects are Genitive-marked by *zhi*, indicating that the relative CPs are highly nominal:

(54) 是聰耳之所不能聽也, 明目之所不能見也, … (荀子: 儒效) (Xunzi: Rixiao)

> shì chōng ěr zhi suō bu neng ting ye, míng-mu zhi suō this sharp-ear Gen what not can hear Prt, bright-eye Gen what bu neng jian ye.

not can see Prt

′This is what a sharp ear cannot hear, and what a bright eye cannot see, …′
(xiii) OC allows extensive use of gerundive constructions with genitive subjects, again revealing the nominal nature of the embedded CP:

(55) 寡人之有五子，猶心之有四支。（晏子·內篇諫上）

‘My having five sons is like the heart’s having four supports.’

We observe the same clustering of properties in OC that distinguish this system en bloc from MnC. In fact, OC seems to pattern consistently like English regarding these properties, and against MnC. Again, this clustering is macroparametric.

We conclude, on the basis of the evidence presented in this and the preceding section that macroparametric variation exists. Therefore our theory of variation must capture these kinds of clusterings of properties.

14.7 Evidence for Microparameters: Synchronic Microvariation among Chinese Dialects

There is considerable microvariation among the various ‘dialects’ of Chinese. Here we list a few striking examples of syntactic microvariation which have been discussed in the recent literature, mainly regarding differences among Mandarin, Cantonese, and Taiwanese Southern Min (TSM).

A first set of differences involves classifier stranding (see Cheng and Sybesma 2005). This operation allows for deletion of the numeral associated with a classifier under the relevant conditions. It is schematically illustrated in (56):

\[
\text{yi ben shu ‘one classifier book’ } \rightarrow \text{ yi ben shu ‘classifier book’}
\]

The dialects of Chinese vary as to the syntactic positions which allow for this kind of deletion. In Mandarin, it is allowed in object position but not subject position:

(57) Mandarin: ok Object, *Subject

a. wo yao mai ge roubaozi lai chi.

‘I want buy CL meat- bun to eat

‘I want to buy a meat bun to eat.’
b. *ge roubaozi tai xian le.
   CL meat-bun too salty Prt.
   ‘A/the meat bun is too salty.’

In Cantonese, it is allowed in both subject and object position:

(58) Cantonese: ok Object, ok Subject
   a. ngo yiu maai go zyuyuk-baau lai sik.
      I want buy CL meat-bun to eat
   b. go zuuyuk-baau taai ham la.
      CL meat-bun too salty SFP

In TSM, it is not allowed in either position:

(59) TSM: *Object, *Subject
   a. *gua be boe liap bapao-a lai tsia.
      I want buy CL meat-bun to eat
   b. *liap bapao-a ukau kiam
      CL meat-bun very salty

This looks rather similar to the distribution of bare nominals in European languages: Italian allows them in object but not subject position (Longobardi 1994): *Latte è buono/Quì si beve latte (‘Milk is good/Here one drinks milk’); Germanic allows them in both positions: Milk is good/I drink milk; French doesn’t allow them in either position: *Lait est bon/*Je bois lait (equivalent to the English examples just given). There may thus be a parallel between the incidence of bare nominals in European languages and the incidence of classifier stranding in Chinese varieties. Clearly this observation merits further explanation.

Second, dialects differ in the extent to which they make use of postverbal suffixes. Mandarin has some aspectual suffixes (e.g., the progressive zhe, the perfective le, and the experiential guo). Cantonese has a considerably more elaborate system, employing additional postverbal suffixes like saai, dak, and ngaang for expressions of exhaustivity, exclusivity, and obligation (see Tang 2006:14–15):

(60) a. keoi tai-saai bun syu.
       he read-up CL book
       ‘He finished reading the entire book.’

b. keoi tai-dak jat-bun syu.
       he read-only one-CL book
       ‘He only read one book.’

c. keoi tai-ngaang nei-bun syu.
       he read-should this-CL book
       ‘He should read this book.’
Some of the suffixes may stack, indicating the considerable height of the verb, for example with the exhaustive on the experiential:

(61) keoi tai-go-saai nei-di syu.
    he read-Exp-Exhaust these book
    ‘He has read up all these books.’

On the other hand, TSM is much more restricted. While the experiential kuei may arguably be a suffix in TSM as it is in Mandarin, the cognates of Mandarin progressive zhe and perfective le are not. Instead, the progressive and the perfective are rendered with preverbal auxiliaries, an analytic strategy:

(62) a. gua ti khoann tiansi.
    I Prog watch TV
    ‘I am watching the TV.’

   b. li u chia-pa bou?
    you have eat-full not
    ‘Have you finished eating?’

Third, the dialects vary regarding their verb–object order preferences (see Liu 2002; Tang 2006). Mandarin allows both OV or VO orders, while Cantonese is strongly VO and TSM strongly OV. The following patterns of preference are typical:

(63) a. Cantonese: ngo tai-zo (bun) syu. ??ngo (bun) syu tai-zo.
    I read-Perf CL book I CL book read-Perf
    ‘I have read the book.’

   b. Mandarin: wo kan-le shu le. wo shu kan-le.
    I read-Perf book SFP I book read-Perf-SFP

   c. TSM: ??gua khoann-kuei tshe a. gua tshe khoann-kuei a.
    I read-Exp book SFP I book read-Exp SFP

Fourth, there is variation regarding the position of the motion verb qu ‘go’ (see Lamarre 2008). Corresponding to the English sentence ‘Zhangsan went to Beijing,’ Mandarin allows both the ‘analytic’ strategy (64a) and the ‘synthetic’ strategy (64b):

(64) a. Zhangsan dao Beijing qu le.
    Zhangsan to Beijing go Perf
    ‘Zhangsan to Beijing went.’

   b. Zhangsan qu-le Beijing.
    Zhangsan go-Perf Beijing
    ‘Zhangsan went Beijing.’
Cantonese allows only the synthetic strategy, whereas Pre-Modern Chinese (as illustrated in textbooks used during Ming–Qing dynasties) allows only the analytic strategy. Assuming that (64b) is derived by V-movement to a null light verb position otherwise occupied by dao in (64a), this pattern shows that V–v movement is obligatory in Cantonese, optional in Mandarin, but did not take place in Pre-Modern Chinese.

We conclude then that there is clear empirical evidence from varieties of Chinese that, alongside macroparameters of the kind illustrated in the previous section, microparameters also exist, with varying (but lesser) degrees of clustering. We will see more examples of microparameters in section 14.10.5.

### 14.8 Macroparameters as Aggregates of Microparameters

The idea that macroparameters are not primitive aspects of UG, but rather derive from more primitive elements, was first suggested in Kayne (2005a:10). It is also mentioned by Baker (2008b:354n2). However, it has been developed in various ways in recent work, starting from Roberts and Holmberg (2010) and Roberts (2012), by Biberauer and Roberts (2012, 2014, 2015a,b, forthcoming), Biberauer, Holmberg, Roberts, and Sheehan (2014), Sheehan (2014, to appear); see again the references at http://recos-dtal.mml.cam.ac.uk/papers.

On this view, macroparameters are seen as aggregates of microparameters with correlating values: a macroparametric effect arises when a group of microparameters act together (clearly, meso-parameters, as in (26), can be defined in a parallel fashion). Hence macroparameters are in a sense epiphenomenal; each microparameter that makes up a macroparameter falls under the BCC, limiting variation to formal features of functional heads.

The microparameters act in concert for reasons of markedness, related to the general conservatism of the learner, and therefore arguably to the third factor (see chapter 6). The two principal markedness constraints are Feature Economy and Input Generalization, as given in (25), repeated here:

1. **Feature Economy (FE)** (see Roberts and Roussou 2003:201):
   - Postulate as few formal features as possible.
2. **Input Generalization (IG)** (see Roberts 2007:275):
   - Maximize available features.

Together these constitute a minimax search and optimization strategy: assume as little as possible and use it as much as possible. As Biberauer and Roberts (2014) show, there are analogs to this strategy in phonology (Dresher 2009, 2013) and in other cognitive domains (see in particular Jaspers 2012). Note also that IG generalizes the known to the unknown, and so can be seen as a form of bootstrapping. The interaction of FE and IG
give rise to the NO>ALL>SOME learning path described in section 14.5. We can now present that idea in a more precise fashion as follows (see also Biberauer, Holmberg, Roberts, and Sheehan 2014:111):

(65) (i) default assumption: ¬∃h [ F(h)]
(ii) if F(h) is detected, generalize F to all relevant cases
(∃h [ F(h)] → ∀h [ F(h)]);
(iii) if ∃h ¬[ F(h)] is detected, restrict h and go back to (i);
(iv) if no further F(h) is detected, stop.

Here h designates functional heads, and F is the predicate ‘feature-of,’ so F(h) means ‘formal feature of a head H.’ As we have said, the procedure in (65) says that acquirers first postulate NO heads bearing feature F. This maximally satisfies FE and IG. Then, once F is detected in the PLD, that feature is generalized to ALL relevant heads, satisfying IG but not FE. This step, in other words the operation of the third-factor strategy IG, gives rise to clustering effects, i.e., aggregates of microparameters acting in concert as macroparameters. The existence of macroparameters and clustering, and therefore many large-scale typological generalizations such as the tendency towards harmonic word order, or high analyticity as in MnC, follows from the interaction of the three factors in language design in a way which is entirely compatible with both the letter and the spirit of minimalism. This establishes Point B in section 14.5.

14.9 The Hierarchical Organization of Parameters

The idea of a hierarchy of parameters was first put forward in Baker (2001:170). Baker suggested a single hierarchy, and, while his specific proposal had some empirical problems, the proposal had two principal merits, both of which are intrinsic to the concept of a hierarchy. First, it forces us to think about the relations among parameter settings, both conceptually in terms of how they interact in relation to the architecture of the grammar (do we want to connect parameters of stress to parameters of word order, for example? See chapter 12 for relevant discussion in relation to first language acquisition), how they interact logically (it is impossible to have inflected infinitives in a system which lacks infinitives, for example), and empirically on the basis of typological observations (e.g., to account for the lack of SVO ergative languages, as observed by Mahajan 1994, among others). Second, parameter hierarchies can restrict the space of possible grammars, and hence reduce the predicted amount of typological variation and simplify the task for a search-based learner (see chapter 11). Given a hierarchical approach, 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. So, if, for example, there are just 5 hierarchies with 20 parameters each. Then |G| is 21^5, or 4,084,101 for 5 × 20 = 100 possible choice
points. Compared to $2^{100}$, this is a very small number, entailing the concomitant simplification of the task of a search-based learner (see again chapter 11, section 6).

Roberts and Roussou (2003:210–213) suggested organizing the following set of options relating to a given formal feature $F$ on the basis of their proposal that grammaticalization is a diachronic operation affecting functional categories:

(66) $F$? (formal feature?)

| No | yes |
| STOP | Does $F$ Agree? |
| No | yes |
| STOP | Does $F$ have an EPP feature? |
| No | Yes |

(head-init) (head-final)

Does $|F$ trigger head-movement? Is $F$ realized by external Merge?

| No | Yes |
| STOP | Does every $F$ trigger movement |
| No | Yes |

High analyticity Agglutinating

Synthesis Polysynthesis

Notice how this hierarchy derives the four traditionally recognized morphological types (Sapir 1921). It also connects analyticity and head-initiality on the one hand, and agglutination and head-finality on the other (see also Julien 2002 on the latter).

Gianollo, Guardiano, and Longobardi (2008, see chapter 16) developed the Roberts and Roussou approach, and, as we have seen, introduced the very important idea that the parameters are not primitives of UG, but created by the hierarchies (‘schemata’ in their terminology). Roberts and Holmberg (2010) proposed two distinct hierarchies for word order and null argument phenomena, and Roberts (2012) and Biberauer, Holmberg, Roberts, and Sheehan proposed three more, dealing with word structure
(polysynthesis/analyticity and microparametric options in between giving various kinds of fusional systems), A’-movement (wh-movement, scrambling, topicalization, and focalization), and alignment. In connection with the last of these, Sheehan (2014, to appear) has developed several hierarchies relating to ergativity, causatives, and ditransitives, and Sheehan and Roberts (2015) have developed a hierarchy for passives. Each of this last class of hierarchies has the general form in (67):

(67) \[ \text{Does H a head have F (i.e. is F in the system at all?)} \]

\begin{align*}
\text{NO} & \quad \text{YES: is F generalised to all H?} \\
\text{YES} & \quad \text{NO: is F limited to a subset of transitive H?} \\
& \quad \text{YES} \quad \text{NO: is F extended to a further subset?} \\
& \quad \text{YES} \quad \text{NO: does H have an EPP feature?} \\
& \quad \text{NO} \quad \text{YES: does H have an “extra” Case/phi feature?} \\
\end{align*}

Sheehan (2014, to appear) shows that a hierarchy of this kind applies to F an inherent Case feature of v (for ergativity), F a feature of Appl (causatives/ditransitives) and F a feature of Voice (passives; see Sheehan and Roberts 2015). Other hierarchies have been proposed for Person, Tense, and Negation (on the latter, see Biberauer 2011).

These hierarchies are empirically successful in capturing wide typological variation of both the macro- and microparametric kind (for example, Sheehan and Roberts’ passive hierarchy covers Yoruba, Thai, Yidin, Turkish, Dutch, German, Latin, Danish, Norwegian, Hebrew, Spanish, French, English, Swedish, Jamaican Creole, and Sami). As already mentioned, this hierarchical organization of the elements of parametrization reduces the potential number of options that a child has, thereby easing the learning procedure. Hence, Plato’s problem is solved.

It is important to emphasize that the macroparameters, and the parameter hierarchies, are not primitives: they are created by the interaction of FE and IG. UG’s role is reducible to a bare minimum: it simply leaves certain options open. In this way, we approach the minimalist desideratum of moving beyond explanatory adequacy (see chapters 5 and 6). Note also that if Biberauer’s (2011, 2015) proposal that the formal features themselves are emergent properties resulting from the interaction of the three factors is adopted, then a still further step is taken in this direction.

We now illustrate these ideas concretely, taking the variation discussed in section 14.6 in Modern Chinese, Old Chinese, and Modern Chinese dialects as case studies.
14.10 Back to Chinese

14.10.1 Summary and a First Attempt to Characterize Chinese in Parametric Terms

To summarize so far, we have made the following proposals. First, there are macroparameters, which capture important, sometimes sweeping, clusters of typological properties. Second, there are also microparameters with fewer or no clustering effects. The macroparameters give rise to general patterns, while the microparameters give rise to the mixed, exceptional cases, plus details of variation or change. Third, macroparameters are aggregates of microparameters acting in concert driven by FE and IG. Fourth, parameters conceived as cases of underspecification do not add to the burdens of UG and are consistent with minimalist theorizing. Fifth, parameters are hierarchically organized so the number of occurring options to choose from is greatly reduced, as is the burden on the learner. Sixth, a small number of parameter hierarchies (which appear to be fairly isomorphic in general, if Sheehan is right) is enough to account for a large amount of cross-linguistic variation (languages, dialects, idiolects).

Now let us consider the macroproperties of Chinese. Viewed synchronically, Modern Chinese exhibits high analyticity in a macroparametric way, in systematic contrast to many other languages (including English, which compared to many other Indo-European languages is often described as somewhat analytic in a pre-theoretical sense). Moreover, Chinese is analytic at all levels: lexical, functional, and at the level of argument structure.

Viewed diachronically: Old Chinese underwent macroparametric change from a substantially synthetic language (typologically closer to English, as we observed in section 14.6.2) to a highly analytic language, at all levels, with analyticity peaking at the end of the Six-Dynasties period and the Tang–Song period, followed by small-scale new changes that result in the major dialects of Modern Chinese, with some varying degrees of small-scale synthesis. We thus observe a partial diachronic cycle: synthetic to analytic to synthetic. (For the shift from synthetic to analytic, see also Mei 2003, Xu 2006, and Peyraube 2014).

The question now is how to characterize the macroparametric properties. One possibility would be a simple ‘analytic–synthetic’ parameter, with the features [±analytic], [±synthetic], so that Chinese is [+analytic, −synthetic], Old Chinese (and say English) are [+analytic, +synthetic], and some other languages (say Romance) are [−analytic, +synthetic].

This is not a good approach, we believe, for two main reasons. First, such a view is purely descriptive and does not reveal the real nature of linguistic variation. For one thing, there are exceptions that must be accounted for, and such exceptions must resort to microparametric descriptions. A binary-value parameter cannot reveal the nature of the gradation that characterizes cross-linguistic variation and diachronic changes. This is the basic problem with many macroparameters that formed the basis of Newmeyer's
(2005) critique. Second, such a view makes use of concepts unavailable in the theoretical vocabulary of a minimalist grammar: what are the features [±analytic], [±synthetic]? While it may have been possible to countenance such features in GB, it is against the spirit, and arguably the letter, of minimalist theorizing.

14.10.2 A Second Approach: Lexical and Phrasal Domains

Suppose, following Hale and Keyser (1993), Chomsky (1995b), and much subsequent work (see in particular Borer 2005, Ramchand 2008) that transitive and unergative predicates involve a form of ‘VP shell.’ In English, unergatives like telephone, transitives like peel and verbs which can freely alternate between the two like fish are associated with a basic structure like that in (68):

(68)

```
|   vP  |
| |     |
| vP   |
| |     |
| DP_EA |
| |     |
| v    |
| |     |
| v'   |
| |     |
| v    |
| |     |
| NP   |
| |     |
| DO   |
| |     |
| telephone |
| |     |
| DO   |
| |     |
| fish |
| |     |
| DO   |
| |     |
| peel |
```

Here DO is an abstract predicate assigning an Agent θ-role to the external argument (EA) in its Specifier (Dowty 1979, 1991; Borer 2005; Folli and Harley 2007; Ramchand 2008). The head of the complement of v incorporates into v, giving rise to a derived verb. Head movement is the operation which gives rise to synthetic structures (and, of course, maximally generalized head movement gives rise to polysynthesis according to Baker 1996). Hence we understand the pretheoretical terms ‘synthetic’ and ‘analytic’ to mean, respectively, having/lacking head movement. This applies across various domains, as we will see.

With verbs showing the anti-causative alternation in languages like English, we posit a CAUSE head above VP, as in (69) (see again Folli and Harley 2007):

(69)

```
|   vP  |
| |     |
| vP   |
| |     |
| DP_EA |
| |     |
| v    |
| |     |
| v'   |
| |     |
| v    |
| |     |
| CAUSE |
| |     |
| VP   |
| |     |
| ...  |
| break |
```


And for the transitive version of denominal verbs, we have a further CAUSE head above vP as in (70), for the transitive feed, i.e., ‘EA causes IA to do food/eat’:

(70)

\[
\begin{array}{c}
\text{NP} \\
\text{food} \\
\vP \\
\text{DPEA} \\
\text{v'} \\
\text{v} \\
\text{VP} \\
\text{CAUSE} \\
\text{DP}_{IA} \\
\text{V'} \\
\text{DO} \\
\text{NP} \\
\text{food}
\end{array}
\]

English v may be in the form of a phonetically null light verb DO or CAUSE, which are assumed to have the following properties: they both have formal features which need to Agree, do not contain EPP, and do trigger head movement (these properties may all be connected in terms of the general approach to head movement developed in Roberts 2010d). Head movement equates to synthesis, and English abounds in simplex denominal verbs like telephone, fish, peel and simplex causatives like break or feed.

In Modern Chinese, on the other hand, v is occupied by an overt light verb such as da for an unergative or a ‘cognate verb’ for pseudo-incorporation:

(71)

\[
\begin{array}{c}
\text{vP} \\
\text{v'} \\
\text{v} \\
\text{NP} \\
\text{da} \\
\text{dianhua} \ ‘\text{telephone}' \\
\text{da} \\
\text{yu} \ ‘\text{fish}' \\
\text{bo} \ ‘\text{peel}' \\
\text{pi} \ ‘\text{skin}' \\
\text{nian} \ ‘\text{read}' \\
\text{shu} \ ‘\text{book}'
\end{array}
\]
For causatives, either an inchoative verb combines with a light/cognate verb to form a compound (rather than moving into a null v forming a simplex causative):

(72)  
\[
\begin{array}{c}
\text{DP}_{EA} \\
\text{v'} \\
\text{v} \\
\text{da/nong} \\
\text{‘do/make’} \\
\text{VP} \\
\end{array}
\]

Or we have a periphrastic causative, with heavy verbs like shi ‘cause,’ rang ‘let,’ and so forth.

(73)  
\[
\begin{array}{c}
\text{DP}_{EA} \\
\text{v'} \\
\text{v} \\
\text{rang} \\
\text{let} \\
\text{DP}_{IA} \\
\text{V'} \\
\text{NP} \\
\text{chi} \\
\text{eat} \\
\text{fan} \\
\text{rice} \\
\end{array}
\]

‘let someone eat rice’

Unlike English, Chinese does not have the phonetically null CAUSE and DO. Instead, it resorts to lexical (light or heavy) verbs which do not trigger head movement (though they may trigger compounding), leading to high analyticity. Instead of simplex denominalized action verbs or simplex causatives, Chinese resorts to more complex expressions, and abounds in light verb constructions, pseudo-incorporation, resultative compounds or phrases, and periphrastic causatives. The high analyticity of
Chinese derives from the absence of incorporation into the abstract DO and CAUSE. These labels are really shorthand for certain event- and \( \theta \)-role-related features of \( \nu \), whose exact nature need not detain us here; these features are lexically instantiated in Chinese by verbs such as \( da \) and \( rang \) which, as lexical roots in this language, repel head movement.

Let us now look at how IG can give rise to macroparametric clustering. By IG, if \( \nu \) can attract a head, then, all other things being equal, \( n, a \), and \( p \) also have that property (this represents the unmarked option as it conforms to IG). Chinese has lexical classifiers, nominal localizers, an adjectival degree marker, and (discontinuous) prepositions, while English generally has such categories in null or affixal form. So high analyticity generalizes across all the principal lexical categories in Chinese.

Looking at the specific cases, Chinese count nouns are formed by an overt 'light noun' (i.e., a classifier):

\[ [\text{CL} \text{ben} [\text{NP shu}]] = \text{count noun} \]

By IG, the light noun does not trigger head movement, so \( \text{ben shu} \) is the Chinese 'count noun'; i.e., an analytic 'count noun phrase.' On the other hand, English count nouns are formed by incorporating the noun root into an empty CL-head (see Borer 2005):

\[ [\text{CL} \text{CL [NP book]}] \rightarrow [\text{CL} \text{book+CL [NP t]}] = \text{count noun} \]

By IG, CL has a formal feature that Agrees, has no EPP and triggers head movement, so the count noun is synthetic.

As we saw in section 14.6, Chinese forms locational NPs with overt localizers (see also Biggs 2014):

\[ [\text{zhuozi} [\text{nali}]] \]

'the table's location'

The word \( \text{nali} \) means 'place.' Here too there is no head movement and so the locative expression is analytic in the sense we have defined. English forms such NPs by incorporating silent PLACE (see Kayne 2005b):

\[ [\text{table} [\text{PLACE}]] \]

Hence \( \text{table} \) here is synthetic.

Chinese adjectives have lexical \( \text{hen} \) ('very'), which marks absolute degree: \( \text{hen hao} \) ('very good'). Kennedy (2005, 2007) proposes treating a gradable adjective as being headed by a \( \text{Deg}^0 \) in the form of covert \( \text{pos} \), e.g., \( [\text{DegP pos [AP happy]}] \), which we may think of as HEN, the covert counterpart of \( \text{hen} \). English adjectives incorporate into null HEN and are synthetic, but Chinese adjectives do not incorporate but remain analytic.
The Deg⁰ head hen or HEN turns a state adjective into a degree word, which is then able to combine with comparatives and superlatives, much as a classifier turns a mass or kind into a count noun so it can be combined with a number word. (See Dong 2005 and Liu 2010 for relevant discussions.)

Chinese complex PPs take a 'discontinuous' form:

(78) zai zhuozi pangbian
     at table side
     'by the table side'

Again, this is an analytic construction. English complex PPs are formed by incorporation, as can be fairly transparently seen in some cases, e.g., beside:

(79) be-the table-side → beside the table

Here side incorporates to be (presumably a morphologically conditioned variant of by), an example of a synthetic preposition (see Svenonius 2010 and the other papers in Cinque and Rizzi 2010 for cartographic analyses of the extended PP). Similarly, one could analyze English in the box along with its Chinese counterpart as AT the box's in(side), thus taking all locative prepositions to be underlingly headed by the light preposition AT.

14.10.3 The Clausal, Inflectional Domain

Mandarin Chinese has aspectual suffixes that are functional heads (they are grammaticalized verbs) and instantiate formal features of those heads. As such, they enter into Agree with appropriate verb stems. However, they do not trigger (overt) head movement:

(80) Zhangsan [ASP PERF] [VP zuotian qu-le Kaohsiung] (PERF Agrees with le)
    Zhangsan yesterday go-le Kaohsiung
    'Zhangsan went to Kaohsiung yesterday.'

English T and Asp heads are similar to Mandarin in this respect. These clausal heads are functional, they enter into Agree with the inflected verb and they do not attract the inflected verbs:

(81) John [T TNS] [VP often kisses Mary in the kitchen] (TNS Agrees with kisses)

In Romance languages, as has been well known since Pollock (1989), T and Asp attract lexical verbs (see Schifano 2015 for an extensive analysis of verb movement across a range of Romance languages, which effectively supports this conclusion, with some important provisos). Thus, while English is synthetic in the v-domain, it is not synthetic in the T-domain: only some, but not all, Fs trigger head movement (in this respect
English may be more marked than either Romance or Mandarin; Biberauer, Holmberg, Roberts, and Sheehan [2014:126] arrive at the same conclusion comparing English to other languages). Chinese is more consistently analytic than English is synthetic; hence it is less marked in this regard than English.

14.10.4 Old Chinese

Let us turn now to Old Chinese, looking first at the lexical domain. In this domain, Old Chinese is similar to English (as we observed in section 14.6.2). Like English, Old Chinese possessed null DO and null CAUSE as higher lexical heads (both reconstructed as ‘s- by Tsu-Lin Mei [1989, 2012] and references given there) which trigger head movement (see also Feng [2005, 2015] for extensive other examples of head movement in OC). This gives rise to the following properties:

(82) a. No light verb, but denominalization:
   \[\text{yu ‘to fish’: } [\text{VP } \ast s [\text{NP yu}]] \rightarrow \text{yu}_\text{verb}\]
   b. No pseudo-incorporation:
   \[\text{fan ‘have rice’: } [\text{VP } \ast s [\text{NP fan}]] \rightarrow \text{fan}_\text{verb}\]
   c. No compounds: synthetic accomplishments.
   \[\text{po ‘break’: } [\text{VP } \ast s [\text{VP-inchoative po}]] \rightarrow \text{po}_\text{verb-causative}\]

And by IG, the properties in (83):

(83) a. No overt classifiers for count nouns (no need for ‘light noun’);
   b. No need for overt localizers (no need for ‘light noun’).

Turning now to the clausal functional heads, Old Chinese TP differs from Modern Chinese in the nature of at least one clausal functional head (probably more than one) in the TP region, immediately below the subject. Let us call this FP (possibly standing for focus phrase). F has an unvalued feature that requires it to Agree with an appropriate element and an EPP feature requiring XP movement. This gives rise to the following XP movements in OC:

(84) a. Wh-movement;
   b. suo-movement for relatives;
   c. focus-movement (of only-phrases);
   d. postverbal adjuncts.

Furthermore, it is possible that F also triggered head movement, giving rise to canonical gapping (Wu 2002, He 2010), assuming, following Johnson (1994) and Tang (2001), that gapping is across-the-board V-movement from a coordinated v/VP. The MnC–OC
contrast follows from the general lack of $v$-movement beyond $vP$ in MnC, and the availability of such movement (e.g., into FP) in OC.

### 14.10.5 Mesoparametric Variation in Modern Chinese Dialects

Here we observe microvariation with small degrees of clustering among Mandarin, Cantonese, and TSM, creating, at least regarding the contrasts between Mandarin and TSM, a mesoparametric effect. Relatively speaking, Cantonese has undergone more grammaticalization and is the least analytic (or most synthetic) of the three dialects. Mandarin (and Shanghai) has developed some suffixes but remains analytic in that these suffixes do not trigger (overt) head movement. TSM remains the most analytic, in having developed the least number of suffixes.

Clearly, here we have only scratched the surface of the dialectal variation to be found in 'Chinese.' Such microparametric differences are sure to increase when more dialects are examined, either contemporary dialects or dialects at any given historical stage. Hence although there is the appearance of macroparametric changes from OC to Modern Chinese, the truth must be that the actual changes took place on a microparametric level.

Let us look more closely at some of the microparametric differences between TSM and Mandarin. Together with those we have touched upon, we can identify 10 differences that distinguish them:

(i) **Classifier stranding.** As mentioned more generally in section 14.7, while Mandarin allows deletion of an unstressed *yi* 'one' in certain positions thereby stranding a classifier, TSM does not allow classifier stranding. Compare the following, repeated from (57a) and (59a):

\[
(85) \text{Mandarin: } \text{wo } \text{yao } \text{mai (yi) } \text{ge } \text{roubaozi } \text{la } \text{chi.}
\]

I want buy (one) CL meat-bun to eat

'I want to buy a meat bun to eat.'

\[
(85) \text{TSM: } \text{gua } \text{be } \text{boe *(tsit) liap bapao-a } \text{la } \text{tsia.}
\]

I want buy *(one) CL meat-bun to eat

(ii) **Aspectual suffix vs. auxiliary.** While the perfective aspect in Mandarin employs the suffix *le*, TSM resorts to a lexical auxiliary *u* 'have.' Compare:

\[
(86) \text{Mandarin: } \text{ni } \text{chi-bao-le ma?}
\]

you eat-full-Perf Q

'Have you finished eating?'

\[
(86) \text{TSM: } \text{li u } \text{tsia-pa bou?}
\]

you have eat-full Q?
That is, in Mandarin the Asp holds an Agree relation with the verb, in TSM a lexical auxiliary does away with the Agree relation. The use of *u ‘have* as an auxiliary is in fact generalized to all other categories, expressing existence of the main predicate’s denotation. Thus, as an auxiliary of a telic *vP*, it expresses perfectivity (as in (86)). It may also be used with an atelic *VP*, or with an *AP, PP, or AspP* predicate, expressing existence of the relevant eventuality:

(87)  a. li u tsia hun bou? (VP)
     you have eat tobacco Q?
     ‘Do you smoke?’

     b. gua bou ai tsit-nia sann (VP)
     I not-have like this-CL shirt
     ‘I don’t like this shirt.’

     c. in u te kong tsit-hang taisi. (AspP)
     they have at discuss this-CL thing
     ‘They have been discussing this thing.’

     d. i tsima bou ti tshu. (PP)
     he now not-have at home
     ‘He is presently not at home.’

     e. tsit-tiunn too u sui. (AP)
     this-CL picture have pretty
     ‘This picture is pretty.’

(iii) Aspectual suffix vs. resultative verb. While Mandarin perfective *le* is a suffix denoting a viewpoint aspect, the corresponding item in TSM *liau* is still a resultative verb meaning ‘finished.’

(88) Mandarin: ta chi-le fan le.
     he eat-Perf rice Prt
     ‘He has eaten / He ate.’

     TSM: i chia-liau peng a.
     he eat-finished rice SFP
     ‘He finished the rice.’

(iv) Null vs. lexical light verb. In Mandarin, there is an interesting ‘possessive agent’ construction, illustrated here:

(89)  a. ni tan nide tade xiaoshuo. (VP)
     you play your piano, he read his novels
     ‘You did your playing piano; he did his reading novels.’

     b. ta ku tade, ni shui nide.
     he cry his, you sleep your
     ‘He did his crying; you did your sleeping.’
In (89a), the possessives nide ‘your’ or tade ‘his/her’ do not denote the possessor of the NP they modify (a piano or a novel). And in (89b), the possessives are presented without a possessee head noun. In each case, the genitive pronoun is understood as the agent of an event, represented as a gerundive phrase in the translation. Huang (1997) argued that these sentences involve a null light verb DO taking a gerundive phrase as its complement. The surface form is obtained when the verb moves out of the gerund into the position of DO.

(90)  

a. ni     DO  nide  [GerundP [VP tan gangqin]]  
   you   DO  your                   play piano  

b. ta     DO  tade  [GerundP [VP ku]]  
   he    DO  his                          cry

These examples thus illustrate a limited kind of denominalization (whereby a verb moves out of a gerundive into DO). Interestingly, the corresponding expressions in TSM take the form of a lexical cho, literally ‘do,’ in place of the null DO, thus repelling head movement:

(91)  

a. li tso [li tuann kengkhim]; i tso [i khuann siosuat]  
   you do you play piano he do he read novel  
   ‘You do your piano-playing; he does his novel reading.’

b. i tso [i khao]; li tso [li khun]  
   he do he cry you do you sleep  
   ‘He went on crying, and you went on sleeping.’

(v) Position of (definite) bare objects. As indicated, Mandarin allows a definite object in postverbal position, while TSM prefers a preverbal object. This preference is particularly strong with bare nouns with definite reference:

(92) Mandarin: ta mei zhao-dao shu.  
   he not seek-get book  
   ‘He did not find the book.’

TSM: i tshe tshuei-bou.  
   he book seek-not-have.  
   ‘He did not find the book.’

(In the TSM example, placing the object tshe after the verb would render it non-referential, meaning ‘he didn’t find any book.’)
(vi) Objects of verb-resultative constructions. In Mandarin they may appear after the main verb, but in TSM they are strongly preferred in preverbal position with ka:

(93) Mandarin: wo ma-de ta ku-le qi-lai.
   I scold-to he cry-Perf begin
   ‘I scolded him to tears.’

   TSM: gua ka yi me-ka khau a.
   I ka he scold-to cry Prt
   ‘I scolded him to tears.’
   (*?gua me-ka yi khao.)

(vii) Complex causative constructions. Mandarin allows V-to-CAUSE raising in forming complex causative constructions, while in TSM such constructions are strictly periphrastic with a lexical causative verb. This is not just a strong preference:

(94) Mandarin: zhe-jian shi gaoxing-de ta liuchu-le
   this-CL thing happy-to he flow-Perf
   ‘This thing pleased him to tears.’

   TSM: tsit-tsan taitsi hoo i huannhi-ka lao-baksai.
   this-CL thing cause he happy-to tears
   ‘This thing caused him to be happy to tears.’
   (*tsit-tsan taitsi huannhhi-ka i lao-baksai.)
   this-CL thing pleased-to him tears

(viii) Outer objects and applicative arguments. In Mandarin the verb may raise above an outer or applicative object, but in TSM it must be licensed by the applicative head ka preverbally:

(95) Mandarin: wo da-le Zhangsan yi-ge erguang.
   I hit-PERF Zhangsan one-CL slap
   ‘I slapped Zhangsan once.’

   TSM: gua ka Abing sian tsit-e tshui-phuei.
   I KA Abing slap one slap
   ‘I slapped Abing once.’
   (*gua sian Abing tsit-e tshui-phuei.)
   I slap Abing one slap

(ix) Noncanonical double-object construction. Both Mandarin and TSM have double-object constructions in the form of V-DP1-DP2. In Mandarin, DP1 can denote a recipient (the canonical DOC) or an affectee (the ‘noncanonical DOC,’ after Tsai 2007). TSM, however, has only the canonical DOC. Thus, (96) in Mandarin has both the ‘lend’ and ‘borrow’ reading, but (97) in TSM has only the ‘lend’ reading:
(96) ta jie-le wo liang-ben shu. 
    he lend/borrow-PERF me two-CL book 
    a. 'He lent me two books.' 
    b. 'He borrowed two books from me.' 

(97) i tsio gua neng-pun tshe. 
    he lent me two-CL book
    'He lent me two books.' 

For the 'borrow' meaning, the affectee (or source) DP1 must be introduced by the applicative ka head:

(98) i ka gua tsio neng-pun tshe. 
    he ka me borrow two-CL book
    'He borrowed two books from me.' 

The contrast shows that the main verb may raise to a null applicative head position in Mandarin, but not in TSM.

(x) ka vs. ba. The above observations also lead us to the fact that, although the Mandarin ba (as used in the well-known ba-construction) is often equated with, and usually translates into TSM ka, the latter has a much wider semantic 'bandwidth' than the former. Generally the Mandarin ba-construction is used only with a preverbal low-level object (Theme or Patient), but the TSM ka-construction occurs with other, 'non-core' arguments, including affectees of varying heights—low and mid applicatives as illustrated above, and high applicatives—adversatives or (often sarcastically) benefactives, as illustrated here:

(99) i tshittsapetsa to ka gua tsao-teng-khi. 
    i 7-early-8-early already KA me go-back
    'He quit and went home on me at such an early time!'

(100) li to-ai ka gua kha kuai-le o. 
    you should KA me more obedient SFP
    'You should be more obedient for my sake, okay?'

We see then that while certain higher functional heads in the vP domain may be null in Mandarin, they seem to be consistently lexical in TSM.

Arguably, in all these cases of differences between Mandarin and TSM, we see some small-scale clustering. In fact, we may be dealing here with one or two mesoparameters as defined in (26). Again, we see the pervasive effects of IG. If we take each difference as indicative of one microparameter, then we have observed ten microparameters. Logically there could be $2^{10} = 1,024$ independent TSM dialects that differ from each
other by at least one parameter value. But it is unlikely that these parametric values are equally distributed. Rather, the likely norm is that they cluster together with respect to certain values. Hence here we have a mesoparameter, expressing special cases of TSM as consistently more analytic than Mandarin, i.e., a range of heads in TSM lacks the formal features giving rise to Agree or head movement in the corresponding cases in Mandarin.

Finally, not all speakers agree on the observations made in the preceding discussion, thus reflecting dialectal and idiolectal differences. This is not surprising, as micro-variations typically arise among individual speakers. Here we may also find cases of nanovariation.

### 14.11 Summary and Conclusion

We began by sketching and exemplifying the GB conception of parameter of Universal Grammar as a parametrized principle, where the principles, the parameters, and the possible settings of the parameters were all considered to be innate. We described how there was a gradual move away from this view, with the introduction of microparameters and also the lexical-parametrization hypothesis (the ‘Borer–Chomsky Conjecture’). We briefly discussed some of the conceptual difficulties with this view, particularly concerning the hyperastronomical number of grammars this predicts and the concomitant problems this poses for typology, diachrony and, in particular, acquisition, where the explanatory value of the whole approach may be called into question. We summarized some of the arguments, notably that put forward in Baker (2008b), for combining macro- and microparameters. We suggested that all (or at least the great majority of) parameters can be described in terms of lexical features (hence following the BCC, and reducing them effectively to microparameters), but we pointed out, following Baker, that we nonetheless do see large macroparametric patterns in the form of clusters. In this connection, we looked at Chinese, showing the remarkable extent of clustering. Here each property can be described by a microparameter, both in respect to synchronic variation and diachronic change, both in respect to typological differences and dialectal variations. The solution we proposed was to adopt the emergentist approach recently developed by Roberts and Holmberg (2010), Roberts (2012), and Biberauer and Roberts (2012, 2014, 2015a,b), and demonstrated how this approach can elegantly describe and explain the observed variation, achieving a high level of explanatory adequacy in the traditional sense (i.e., solving Plato’s Problem), while at the same time, in emptying UG of any statement regarding parameters beyond simple feature-underspecification, taking us in the desired direction, beyond explanatory adequacy.