1. Background

The theory of principles and parameters (P&P) as first proposed in detail in Chomsky (1981) and developed in versions of the minimalist programme for linguistic theory offers a promising approach to the classical question of explanatory adequacy. It is much less clear, however, whether this approach offers a way of approaching the question of complexity. This paper aims to address this question: what (if anything) can modern parametric theory tell us about the formal complexity of grammatical systems?

In order to approach this question, we have to be clear about the nature of P&P. Its central idea can be summarised in (1), which essentially paraphrases Chomsky (1995):

(1) An I-language is an instantiation of the innate language faculty with options specified.
Here ‘I-language’ is taken in the sense of Chomsky (1986): the internal, individual language faculty characterised in intension by a generative grammar. The innate language faculty is that aspect of the human genome, apparently unique to humans, which makes the possession of an I-language possible, given appropriate environmental stimulus in early life. The theory of this faculty is Universal Grammar (UG). The “options” of (1) are the parameters of UG, whose nature is the focus of much of the discussion below; suffice it to say for the moment that the parametric options create the space of variation occupied, at least in part, by the typological diversity attested in actually occurring languages.

The P&P approach represents a major advance on earlier conceptions of language acquisition (see in particular the discussion in the Introduction to Chomsky 1981). Chomsky (1964) identified the goal of achieving explanatory adequacy in linguistic theory as accounting for the acquisition of a grammar in relation to Universal Grammar (UG). Earlier approaches, prior to 1981, had defined UG as a grammatical metatheory specifying a broad format for rules and some general conditions on rule application; a particular grammar as a system of language-specific, construction-specific rules; and language acquisition as rule induction, aided by an evaluation metric. This theory offered little hope for insights into either language typology or language acquisition. P&P stood in stark contrast to this from its inception. The leading idea was that UG contains an invariant set of principles associated with parameters which define the space of possible variation among the grammars of actual, individual I-languages. In these terms, language acquisition could be seen as setting the parameters of the native language on the combined basis of the innate UG and the triggering aspects of the primary linguistic data (PLD).

In short, P&P appeared to significantly simplify the learning task, while at the same time
providing typological insights in the form of “parametric clusters”. Thus, it provided a way of connecting biolinguistics with language typology.

Despite its conceptual advantages over earlier approaches and its initial empirical promise in facilitating a new approach to typological questions, the P&P approach nonetheless has drawbacks, and these have gradually come to the fore in recent years.

The main difficulty encountered by P&P theory in recent years reflects an empirical issue. As research in comparative syntax has advanced, many more parameters than originally envisaged have been proposed to account for observed cross-linguistic variation. Much of this work has been quite successful, and there can be little doubt that our knowledge of the syntax of many of the world’s languages and also of crosslinguistically recurring patterns has increased enormously since 1981. P&P theory has been an excellent heuristic. But Chomsky’s criterion of explanatory adequacy requires more than this. Arguably, the direction that P&P theory has taken reflects the familiar tension between the exigencies of empirical description, which lead us to postulate ever more entities, and the need for explanation, which requires us to eliminate as many entities as possible. In other words, parametric descriptions as they have emerged in much recent work tend to sacrifice the explanatory power of parameters of Universal Grammar in order to achieve a high level of descriptive adequacy. The result is that the learning task remains mysterious, and the utility of P&P in solving this problem, which at the outset seemed so clear, is questionable.

Newmeyer (2004, 2005) was the first to construct a detailed critique of P&P theory, concluding that it was not living up to its promise. He advanced a number of criticisms of the approach, not all of which we agree with, and we do not endorse his conclusion that the approach should be abandoned. But Newmeyer (2005:83) makes one extremely telling point: “... we are
not yet at the point of being able to ‘prove’ that the child is not equipped with 7,846 /.../ parameters, each of whose settings is fixed by some relevant triggering experience. I would put my money, however, on the fact that evolution has not endowed human beings in such an exuberant fashion.”. In other words, P&P theory places too much content in the innate endowment, and aside from general plausibility questions, this places an almost intolerable burden on any account of the evolution of language.

Finally, one of the most difficult problems for acquisition/learnability theory remains. This is often referred to as the Linking Problem (cf. Pinker 1984, Dresher 1999 and Fasanella & Fortuny 2013 for recent discussion). Parameters are defined over abstract linguistic entities, with the result that the language-acquiring child has to link these mental representations to actual physical entities in the speech signal. It is every bit as unclear in P&P theory as in almost any other approach to language acquisition how this happens.

Newmeyer’s point holds in full force if learners must link and set large numbers of innately specified parameters, each independently of all the others. On these assumptions, the learnability problem takes its starkest form. Moreover, in this case, we might expect all grammatical systems to be equally complex. It would therefore appear on this view that P&P theory has little or nothing to say about the relative complexity of grammatical systems. If, however, parameters are interconnected in various ways, then this may simplify the learning task substantially. A concomitant of this is that the possibility then arises that certain parametric “routes” to steady-state grammars are shorter – and hence in an obvious intuitive sense simpler – than others. This is the central idea that we try to develop in what follows.
Let us begin with a very simple example. It has been known since at least Kayne (1981) that languages vary as to whether they allow “Exceptional Case-marking” infinitives or not. English allows this construction and French does not (in the canonical context involving believe-type verbs; see Kayne 1981):

(2)  
  a. John believes Paul to write the best songs.
  b. *Jean croit Paul écrire les meilleures chansons.

We could therefore posit a parametric difference between English and French. It is also known that some languages have no infinitives at all, e.g. Modern Greek. In such languages, constructions corresponding to raising, control and ECM typically involve finite clauses in the indicative (complement to believe) or subjunctive (complement to expect) (D. Michelioudakis, p.c.).

So there are at least three options made available by UG: (i) both ECM and non-ECM infinitives (English); (ii) non-ECM, but not ECM infinitives (under believe-type verbs) (French); and (iii) no infinitives at all (Modern Greek). It is clear that the “no infinitives” option obviates the need to choose between ECM and no ECM. In that straightforward sense, the third option is simpler than either of the other two: there is less for the learner to do. Moreover, we see that there are advantages from the learnability perspective in linking parametric options, and of course the more “intrinsic” those links can be, as in this example, the better.

Moreover, the minimalist programme (Chomsky 1995 et seq.) offers the possibility of seeing the nature of parameters in a new way, one which clearly offers a solution to the problem.
identified by Newmeyer. To see this, consider the three factors of language design put forward in Chomsky (2005):

(3) a. Factor 1: innate endowment (UG)
    b. Factor 2: experience (PLD)
    c. Factor 3: non-language-specific innate capacities

The first and second factors do not require much comment here and we note only that Factor 1, from a minimalist perspective probably contains far less than was assumed in former stages of the P&P approach. The “third factors”, according to Chomsky, include “(a) principles of data analysis that might be used in language acquisition and other domains; (b) principles of structural architecture and developmental constraints ... including principles of efficient computation” (Chomsky 2005:6). These factors clearly require further elucidation before the overall approach can be evaluated. Below, we will attempt to do this in relation to parametric variation and language acquisition. The general view that we take is that parametric variation is an emergent property of the interaction of the three factors listed in (3), and that parameters emerge as a consequence of the learning process. All that is prespecified is (a) a small number of invariant properties of UG (first factor) and (b) general computational conservatism of the learning device (third factor).

As we hope to show, this view allows us to flesh out in interesting ways the question of whether languages differ in complexity. We address that question in §3 below. In §2, we set out in more detail our approach to parametric variation.
2. The Nature of Parameters: A Proposal

2.1 On the Nature of Parameters

In this section, we introduce and illustrate the “emergentist” approach to parameters just described. We must first state what does not vary, i.e. what is part of UG. UG determines the following properties of the linguistic computational system C\textsubscript{HL}:

\begin{enumerate}
\item certain formal features
\item recursive, binary Merge;
\item a labelling algorithm;
\item Agree (feature-valuation, relating elements of syntactic structures).
\end{enumerate}

Obviously much more needs to be said about all of (5a-d).\textsuperscript{1} For present purposes, we take the class of formal features to include categorial features (±N, ±V, etc), structural Case features (or equivalent), person, number and gender features (collectively \(\phi\)-features), other features such as

\textsuperscript{1} This exposition leaves two important issues open. First, the status of thematic roles, which, if some version of Baker (1988) is right, are structurally determined. It remains unclear whether the correlation between thematic role and relative syntactic position is determined by UG or emerges from some connection between event participation and structural prominence. Second, locality conditions, notably Relativised Minimality and the Phase Impenetrability Condition. We take Relativised Minimality to be a case of the general third-factor strategy of Minimal Search. Phase Impenetrability might be too. On the (im)possibility of connecting these two principles, see Rizzi (2013).
[±wh], [±neg], [±tense], etc, as well as purely diacritic features which simply trigger operations (different kinds of Merge, usually).²

Following Chomsky (1995:243ff.), we take Merge to recursively combine two syntactic objects α and β to form a set \{α, β\}. The objects may be drawn from the Lexicon/Numeration (External Merge), or, if the members of an existing set \{α, β\} have internal structure, from within α or β (Internal Merge). The set formed by Merge requires a label K, i.e. Merge creates the object \{K, \{α, β\}\}. K is determined by either α or β, giving the effect of “projection” of a syntactic category label, and hence endocentric structures. Finally, Agree involves valuing of formal features, which we take to be attribute-value pairs of the form [Person:3], [Number:plural], i.e. [Att(ribute):Val(ue)]. Features may enter the syntax without a value, i.e. as [Att:__], which the interpretative devices of the interfaces cannot read. Agree takes a pair of syntactic feature-bearing elements γ and δ such that for some feature F, one of γ and δ has the form [Att:__] and the other has the form [Att:Val]; the former is the Probe and the latter the Goal. The Probe must asymmetrically c-command the Goal and there must be no Goal' bearing an unvalued F such that the Probe asymmetrically c-commands Goal' and Goal' asymmetrically c-commands the Goal, i.e. the Goal must be the “closest” possible Goal to the Probe. All of this is a fairly mainstream set of technical assumptions; for more details, see Chomsky (2001). The mechanisms described above are what we take to be the invariant core of UG.

² See Sigurðsson (2011) and Biberauer (2011, 2013) for the claim that all or many formal features may also acquired on the basis of the interaction between a basic UG-given schema and the PLD. For present purposes, however, we keep to the rather more “conservative” position which attributes some such content directly to UG.
Our principal departure from standard P&P thinking concerns the nature of parameters. Rather than taking them to be prespecified options of the kind “A head X \{precedes/follows\} its complement YP”, “A head H, drawn from a set of heads L of licensing heads, formally licenses some element E in configuration C”, etc., which are somehow genetically encoded, we take them to arise from **underspecification** of formal features in UG. This underspecification can take three forms, as follows:

(5)  
   a. association of formal features with (functional) heads;  
   b. values of formal features, triggering Agree;  
   c. purely diacritic features triggering movement.

Certain heads are intrinsically potential bearers of formal features; this set may well be limited to the class of functional heads. For example, T bears ϕ-features of various kinds in many languages. In most Indo-European languages, T has Person and Number features and so we see agreement between the verb and the subject. Gender agreement between the verb and the subject is rare in Indo-European, but found in many Semitic languages including Classical Arabic. Furthermore, as (5b) states, formal features may have their value specified or not; if they do not, then Agree is triggered, and it does seem to be the case that languages can vary as to the specific Agree operations they require (see Miyagawa 2010). The options in (5c) concern the distribution of the movement-triggering feature, which, following Biberauer, Holmberg & Roberts (to appear, BHR henceforth) we write “^”. This sub-feature can be associated with any kind of syntactic

---

3 The Italo-Romance variety of Ripatransone is a rare case of this in Europe. See Ledgeway (2012:299-310) for discussion and illustration.
dependency triggering Internal Merge, with languages differing in relation to which dependencies are associated with ^.

There is clearly a close relation between (4) and (5). In fact, (5) really says that a subset of the core properties of UG is optional in a given I-language; this is the content, for us, of Chomsky’s statement in (1) that a given I-language is an instantiation of UG “with options specified.” The “theory of parameters” is nothing more than this: some subset of the universally available set of features is optional. In other words, to paraphrase Biberauer & Richards (2006), parametric variation emerges where UG “doesn’t mind”.

To elaborate slightly and give some more concrete examples of (5), (5a) includes such options as the mapping of features to heads, i.e. feature-scattering vs. feature-syncretism (see Giorgi & Pianesi 1997); presence vs. absence of features on heads; differing distribution and internal make-up of properties such as finiteness in clauses – this may underlie the variation between English, French and Modern Greek discussed above. Variation in Agree, and in its expression through inflectional morphology (which we take to be fairly closely associated to the presence of the features for learnability reasons; see Holmberg & Roberts 2012), gives rise to the differing properties of subject-agreement in English vs Italian vs Japanese, etc., and also to “doubling” effects, e.g. Negative Concord, ‘forked’ modality in many South-East Asian languages (see Cheng & Sybesma 2003), “bracketed” relative clauses (Peng 2011, Bradshaw 2009, Hendery 2012), etc. (5c) gives a range of movement options: V-movement in English vs. French (Pollock 1989) vs. Germanic verb-second languages (Holmberg & Platzack 1995); wh-movement in English vs Chinese (Huang 1982); and, in combination with (5b), to different kinds of case systems including the distinction between ergative and accusative systems (Sheehan 2013); and variation in “basic” head-complement order (BHR, Sheehan to appear, a).
Essentially, (5) reduces to the statement in (6):

(6) A given formal feature F may associate with a different set of heads (including the empty set) in different languages.

Here, for attribute-value features, “F” ranges over [Att:val] and [Att:__]). To put things a little more formally, we can say that parameters involve generalised quantification over formal features, as follows:

(7) \( Qh \in P [F(h)] \)

Here Q is a quantifier, h is a head, P is the set of heads bearing the relevant formal properties (\( \varphi \)-features, movement-triggering features, etc.), and F is the set of formal features. Both F and P may be null in a given system, i.e. a given option may fail to apply.

This approach gives rise to the following informal taxonomy of parameters (Biberauer 2011, Biberauer & Roberts 2012a,b, 2013):

(8) For a given value \( v_i \) of a parametrically variant feature F:
   a. **Macroparameters**: all (functional) heads share \( v_i \);
   b. **Mesoparameters**: all functional heads of a given naturally definable class, e.g. [+V], share \( v_i \);
   c. **Microparameters**: a small subclass of functional heads (e.g. modal auxiliaries, pronouns) shows \( v_i \);
d. **Nanoparameters**: one or more individual lexical items is/are specified for $v_i$.

It is clear that the different kinds of parameters listed in (8) are hierarchically related to one another. So we are led to postulate different kinds of parameter hierarchies, whose relevance for acquisition we discuss below.

2.2 Word Order

Roberts (2012) suggests the following parameter hierarchy for word order:

\[(9) \quad \textit{Hierarchy 1: Word order}\]

\begin{itemize}
  \item a. Is head-final present?
  \begin{itemize}
    \item No: **head-initial**
    \item b. Yes: present on all heads?
      \begin{itemize}
        \item Yes: **head-final**
        \item c. No: present on [+V] heads?
          \begin{itemize}
            \item Yes: **head-final**
            \item d. No: present on ...
              \begin{itemize}
                \item in the clause only
              \end{itemize}
          \end{itemize}
      \end{itemize}
  \end{itemize}
\end{itemize}

Here we use the neutral term “head-final”. More technically, head-finality may be instantiated as a complement-movement feature, following the general approach in Kayne (1994), or perhaps as a PF head parameter of the kind proposed by Richards (2004) and Sheehan (to appear b); for
present purposes we do not need to choose among these options. The higher nodes in this hierarchy define, first, rigidly head-initial systems and, next, rigidly head-final systems; in these systems all heads capable of varying in linear order in relation to their complements show a single, consistent order (we return below to the question of how a hierarchy structured as in (9) can be viewed as defining a learning path. These are macroparametric options both in the intuitive sense that they have massive effects in the grammars they determine, and in the sense defined in (8). The third output in (9) approximates to the typical Continental West Germanic situation (in which all clausal heads except C follow their complement); by the definition in (8), this represents a mesoparameter. Further “down the hierarchy” on the unspecified lowest right branch, we define micro- and nanoparameters, ultimately specifying, for example, that in English the single lexical item _enough_ follows rather precedes the adjective it degree-modifies, unlike all other degree modifiers in English (i.e. _tall enough/*enough tall vs very tall/*tall very_).

Roberts (2012) proposes that the parameter hierarchies arise from two interacting markedness conditions, Feature Economy (FE) (Roberts & Roussou 2003:201) and Input Generalisation (IG) (Roberts 2007:273-5). These can be stated as follows:

(10)  

a. **Feature Economy (FE):**  
Given two structural representations R and R’ for a substring of input text S, R is less marked than R’ iff R contains fewer formal features than R’;

b. **Input Generalisation (IG):**  
If a functional head F sets parameter $P_j$ to value $v_i$ then there is a preference for similar functional heads to set $P_j$ to value $v_i$. 

13
Input Generalisation plausibly follows from the acquirer’s initial “ignorance”: not initially knowing what the categories in the target language are, the acquirer assumes an identified property/pattern to apply maximally generally; recognition that a new (sub)category needs to be distinguished, however, leads to re-evaluation of the initial input generalisation. So-called *superset traps* are therefore circumvented because the child is assumed to be establishing the relevant inventory of syntactic categories incrementally (see Biberauer (2011, 2013), Branigan (2011, 2012) for more detailed discussion).\(^4\) We take the conditions in (10) (perhaps along with the Subset Principle (Berwick 1985)) to arise from general cognitive optimisation strategies applied to the task of language acquisition, *not* from UG. So the hierarchies are not part of UG, but determined by the underspecified parts of UG, interacting with conditions like those in (10) and the PLD. It is in this sense that parametric variation emerges from the three factors of language design given in (3). Since they do not form part of UG, the hierarchies cannot directly determine explanatory adequacy in Chomsky’s (1964) sense. In fact, the hierarchies are descriptive taxonomies of the emergent system, i.e. epiphenomena. Since that system emerges from the interaction of the three factors in language design, and explicitly relates typological generalisations to language acquisition, and since the hierarchies aim to provide an explicit characterisation of the way in which syntactic variation is structured, they obviously have explanatory value, though.

We can state things more precisely following the notation introduced in (7): given a head \(h\), the set \(P\) of heads bearing the relevant formal properties (\(\phi\)-features, movement-...

\(^4\) This approach, then, can be characterised as falling into the class of maturational (rather than continuity) approaches to syntactic development (cf. Rizzi 1994 for discussion of the differences between these types of approach).
triggers, etc.), and the set $F$ of features, the general form of hierarchies and the steps in the learning path, as determined by $FE$ and $IG$, will be as follows:

(11) a. Hypothesis I (ahead of any experience/analysis of PLD):

No head in $P$ has $F$ ($\forall h \in P \neg[F(h)]$); this hypothesis maximally satisfies $FE$ and $IG$

b. Hypothesis II (at least one occurrence of $F$ is detected in the PLD):

All heads in $P$ have $F$ ($\forall h \in P [F(h)]$); $FE$ is overridden by PLD, $IG$ is still satisfied

c. Hypothesis III (at least one non-occurrence of $F$ is detected):

Some heads in $P$ have $F$ ($\neg\forall h \in P [F(h)]$); both $FE$ and $IG$ overridden by PLD

The left branches of (9) reflect this ordering of (progressively weaker) hypotheses; (9/11a-b) reflect the macroparametric options; at the next level, generalisation ranges over $P' \subseteq P$, where $P'$ is defined as a linguistically natural class (in (9), the class of [+V] heads), and Hypotheses I-III are iterated over these classes; the shift from generalising over $P$ to generalising over $P'$ takes place since, at Hypothesis III, generalising over $P$ gives no clear outcome. $FE$ and $IG$ conspire to make each step refer to the minimal ($FE$) and the maximal ($IG$) proper subset of categories, hence the next level is the mesoparametric one. The microparametric level operates on still smaller subsets $P'' \subseteq P' \subseteq P$. The nanoparametric level operates on the smallest feasible subset (individual lexical items).

2.3 Null arguments
A further hierarchy, first put forward by Roberts & Holmberg (2010:49), concerns null arguments. In terms of (11), we expect it to take the following form:

(12)  
  a. Hypothesis I: no head in P (the set of probes) has uninterpretable φ-features.  
  b. Hypothesis II: all heads in P have uninterpretable φ-features.  
  c. Hypothesis III: some subset of P (the largest natural class P’⊂ P) has uninterpretable φ-features.

The system of hypotheses in (12) can be graphically illustrated by the diagram in (13), to which we have added a further mesoparametric option at the lowest level shown here:

(13)  
  **Hierarchy 2: Null arguments**  
  a. Are uφ-features present on probes?  
     No | Yes  
     |      
     | Radical pro-drop | Pronominal arguments  
     | Yes | No  
     |     |      
     | No | Yes  
     |     |      
     |     |     | c. Are uφ-features fully specified on some probes?
Here “radical pro-drop” refers to languages of the Chinese-Japanese type, which allow any pronominal argument to be “dropped”, and lack agreement inflections which could “track” such arguments (see Huang 1984, Tomioka 2003, Saito 2007, Neeleman & Szendrői 2007 for discussion and differing analyses of this phenomenon). “Pronominal argument” is intended in the sense put forward by Jelinek (1984): languages of this kind typically have very rich agreement marking for many, if not all, grammatical functions and a high degree of word-order freedom. Jelinek proposes that the agreement markers are the true arguments, incorporated into the verb from argument positions, with the optional realised nominal “doubles” of these arguments being adjuncts, hence their somewhat free order (see also Speas 1990, Baker 1996). Again, these options are macroparametric, both in the clear sense that they have massively proliferating effects in the grammars they determine and in the sense defined in (8), with all heads sharing the same property. Question (c) corresponds to Hypothesis III in (11). At this point, then, mesoparametric options become relevant. This is arguably where the classical null-subject parameter comes in. This parameter refers to a sub-class of heads (finite T), as is usual for
mesoparameters. Also, we have changed “present on probes” to “fully specified on probes”, since it is possible, for example, that [Person] is not present on T in English (its only putative instantiation is 3sg present –s which may be a default (Roberts to appear) or a [Number] morpheme (Kayne 1989)); if this is true, then English lacks fully specified T-probes. Why only T and not at least v becomes relevant at this point in this hierarchy is not yet clear to us, but it seems empirically correct that (a) subject agreement is cross-linguistically more common than object agreement and (b) (definite) null subjects are more common than null objects. Note, though, that the pattern of options is the same (features nowhere > features everywhere > features somewhere). The lower reaches of this hierarchy, not shown here, probably specify various kinds of partial null-subject systems, in the sense of Holmberg, Nayudu & Sheehan (2009), Holmberg (2010), in which φ-features are only partially specified on T.

2.4 Word Structure

A further parameter hierarchy concerns word structure. We assume that complex words are formed by head-movement (cf. Baker 1988, 1996, Julien 2002). Hypothesis I then assumes no head-movement, i.e. highly analytic morphosyntax, while Hypothesis II assumes exceptionlessly instantiated head-movement, i.e. polysynthesis. Finally, Hypothesis III assumes more limited head-movement, giving the mesoparametric options. This hierarchy is shown in (14):

5 It may seem odd to refer to finite T as a class of heads, but recall that it is necessarily present in all finite clauses and that it is also morphologically instantiated on all finite verbs and auxiliaries in a given language.
We take head-movement to be the formal operation which determines the internal structure of complex words. More specifically, following Roberts (2010), we see head-movement as a subcase of Agree, where the moved head has a proper subset of the features of the Probe (it is a “defective Goal”). The notion of high analyticity referred to here is taken from Huang (2013), who sees exactly this property as a macroparameter characteristic of Chinese (of various kinds, including Modern Mandarin). Huang identifies a range of syntactic effects of this property, alongside extreme morphological analyticity, which are construed as showing lack of head-movement in the system as a whole. The next option is polysynthesis, as discussed and analysed by Baker (1996). Following Baker, we take “the polysynthesis parameter” to involve head-movement as a highly prevalent property of the system, affecting all lexical categories. Both analyticity and polysynthesis are again macroparameters: they have massive effects on the systems they determine. Fully polysynthetic languages, i.e. those which show polysynthesis both
in the clausal and the nominal domains fall under Hypothesis II at the macro-level. A positive value for [+V] at the mesoparametric level would give rise to a language which is polysynthetic only in the clausal domain (e.g. Michif; Bakker 1997). Lower in the hierarchy we encounter parameters determining familiar cases of head-movement; Biberauer & Roberts (2012a) propose a sub-hierarchy for verb-movement which encompasses a meso-parametric option for V-to-T movement (as in Pollock 1989), microparametric options for auxiliary-movement (encompassing Modern English “have/be-raising” (Emonds 1976)), and nanoparameters affecting individual auxiliaries in certain varieties of Modern English. N-movement of the kind discussed Bernstein (1991) and Longobardi (1994) defines a further mesoparametric sub-hierarchy.

True macroparameters sit close to the top of each hierarchy, as here all heads parametrised for the feature(s) in question behave as one. Moving down the hierarchy, parameters become more “micro”, behaving in a non-uniform, differentiated fashion. Crucially for present purposes, microparameters are also inherently more complex than the systems defined higher in the tree. This is simply because a smaller number of parameter settings are needed to give rise to the higher options, as we have seen, while systems defined lower down in the hierarchy require more parameters and apply first to natural subsets of the entire set (mesoparameters), to smaller classes of functional categories F (microparameters), and ultimately to single lexical items (nanoparameters). Moving “down a hierarchy”, then, systems become more marked, having a longer and more complex description than the higher options. The possibility also arises that lower options are further along a given learning path. Here it is worth noting that many, perhaps all, nanoparametric options fall outside the core system defined by the hierarchies under discussion here. To the extent that nanoparametric options involve high-frequency elements, they appear to be acquired as independent lexical items, independently of
the more general properties of the system to which they belong; hence the much-discussed U-shaped acquisition pattern associated with the acquisition of high-frequency irregulars (cf. Marcus et al. 1992 for detailed discussion). In our terms, forms of this type would therefore not be acquired as a result of progressing down a given hierarchy, although their connection to specific hierarchies – in the sense that they appear to represent isolated instantiations within a given system of a pattern that can be seen to hold more systematically in other systems – is clear. We return to this point in section 3 below.

2.4 A’-Movement

The fourth hierarchy is more tentative than the others, and we introduce it largely to illustrate how our approach can shed light on the general question of grammatical complexity. It concerns what, following Kiss (1995), we can loosely refer to as “discourse configurationality”. More technically, it concerns options of A’-movement. As such, two ingredients are crucial: the concept of phase, as introduced in Chomsky (2000) and developed in Chomsky (2001), and A’-related formal features (for simplicity, we will refer to [focus], [wh] and [topic]). We assume, in part following Chomsky (2007), that phase-heads define local domains, license movement to and/or through their left periphery, and trigger A’-movement. Suppose that C, D and v are phase-heads (there may of course be others). Suppose further that there is universal functional pressure for systems to encode focalisation/topicalisation, these being a component of the “second” type of semantics Chomsky highlights in referring to “duality of semantics” (we return to this point below). Formally, let us assume that elements which are to undergo focalisation/topicalisation and A’-movement more generally will be “inflected” to reflect this fact, i.e. they will differ from elements which can remain in situ in virtue of bearing one or more A’-features of the relevant kind. Crosslinguistic investigation has shown that A’-movement is often to left-peripheral
positions within CP, vP and DP. At the same time, syntactic locality (subjacency/island conditions) severely restricts movement to the left periphery, forcing all (long-distance) movement to be successive-cyclic. Phase heads can function as escape hatches (licensing cyclic movement *through* their left periphery, without interpretive effect) or as targets (licensing movement *to* their left periphery, giving an appropriate discourse interpretation). Under certain circumstances, this phasal “escape hatch” is not available, however. Let us suppose that all phase-heads can, in principle, allow successive-cyclic movement *to* their edge and, where they do not represent the last-merged phase-head in the clausal domain, also *through* their edge (this is, then, an extension of Chomsky’s 1973 proposals regarding the successive-cyclicity of *wh*-movement). The *through* options available to non-last-merged phase-heads are clearly restricted where island effects are observed. Two considerations which appear to be relevant in determining the possibility of escaping from phasal domains are (i) the relevant domain having been spelled out (which we take to mean that its internal structure has become invisible to the computational system, with the result that it cannot be targeted by either of the operations Agree or Move) and (ii) the relevant domain having been “sealed off” by a highly specified nominal head whose rich featural specification precludes the possibility of other elements being extracted across it (i.e. Relativized Minimality considerations of the type discussed i.a. in Starke (2001) and Rizzi (2001, 2013)). Precisely which *to* and *through* options are permitted and whether a given system includes a nominal head of the relevant kind we assumed to be a matter of parametric specification (the distinction between preposition-stranding and obligatory pied-piping would presumably be determined by this kind of option applying to PP). More specifically, consider (15):
(15) \textbf{Hierarchy 4: A'-movement}

<table>
<thead>
<tr>
<th>a. Do phase-heads trigger A'-movement?</th>
</tr>
</thead>
<tbody>
<tr>
<td>No: UNATTESTED</td>
</tr>
<tr>
<td>Yes: “free word order” (a)</td>
</tr>
<tr>
<td>Yes: \textit{wh-in-situ+scrambling} (b)</td>
</tr>
<tr>
<td>Ye: syntactically ergative languages (c)</td>
</tr>
<tr>
<td>Yes: \textit{wh-movement} + scrambling (d)</td>
</tr>
</tbody>
</table>

\footnote{\textbf{6} This negative formulation aims to highlight the fact that C only attracts a subset of A’-elements to its edge, [focus]- and [topic]-bearing ones. Given the prominence of \textit{wh}-interrogatives in the input, it is plausible to assume that acquirers will register \textit{wh}-in-situ structures, with the result that they will conclude that a specific phase head, C, does not trigger movement of \textit{wh}-elements and that, therefore, A’-movement, which had initially been assumed to be possible across-the-board, is (potentially) restricted to topics and foci. The proposal is there \textit{not} that the child is “interrogating” the input on the basis of the c.-question.}
Here, we see that one of the options given by the broadest question, namely that of foregoing A’-movement, is in fact a non-choice. We return to the matter of such no-choice parameters below.

Type (a) languages include Warlbiri and many other Australian languages, Latin, the Slavonic languages and others. These languages have very liberal scrambling, both to the Mittelfeld and to the left-periphery, and also subextraction from nominals, creating the possibility that adjetival and other adnominal modifiers can appear somewhat distant from the noun they modify, one characteristic often thought to characterise “free word order”. This type of language we assume to be the reflex of a formal system in which all phase-heads (and relevant clause-internal elements; see below) have the possibility of being specified for one or more of [topic], [wh] and [focus] with one or more associated ^-features, triggering movement, and in which the highly specified island-creating nominal head mentioned above is absent. More specifically, C, v and D in languages of this type will all be able to trigger both movement through their edge (by virtue of their being able to bear an independent ^, not associated with a specific substantive formal feature) and topic-, wh- and focus-movement to their edge (by virtue of their ability to bear [topic], [wh] and [focus] features respectively alongside ^). Thus systems of this kind can be thought of as instantiating a macro option in relation to phase-heads in that they treat all the moving to and through options associated with these elements identically. This means that fewer
types of C, v and D (i.e. fewer sub-categories) need to be acquired, as one would expect for an option located high on a hierarchy defining a learning path.

Type (b) languages include Japanese and Korean; these languages have quite liberal scrambling, but no clausal-level overt wh-movement in interrogatives. These phenomena we view as indicative of the fact that C, v and D cannot be treated identically, as was the case of Type (a) languages. Instead, it seems to be necessary to distinguish between the way in which C, v and D bear [topic]- and [focus]-features on the one hand and [wh]-features on the other. Consequently, languages of this type require the postulation of a larger number of distinctly specified heads than Type (a) languages.

Type (c) languages include Tagalog and many other Polynesian languages. Strikingly, these are ergative languages, which restrict wh-extraction to absolutive-marked arguments. In terms of analyses like Aldridge (2004), Coon, Mateo Pedro & Preminger (2011) and Sheehan (2013), this restriction entails that only arguments that can be targeted for movement through the edge of vP by virtue of the fact that they are not first-merged within that edge can in fact be extracted.7 The properties of Type (c) are clearly also in part determined by aspects of Hierarchy 5, which we discuss below, pointing to the fact that the hierarchies may interact with one another (a topic which we must leave aside here).

Type (d) includes German and Dutch, i.e. systems which feature Mittelfeld scrambling, and overt wh-movement. In formal terms, these are systems in which C and (non-island-

7 These analyses propose that movement of the absolutive to the outer specifier of vP serves to trap the transitive (ergative) subject inside that phase. As Assmann et al. (2012) show, this restriction affects only transitive ergative subjects and not other arguments inside vP, suggesting that what blocks extraction of the transitive subject is its base-generation in the phase edge.
inducing) D may be specified as for Type (a) and (b) languages, but where v cannot bear ^-associated [focus] or [wh]; thus only [topic]-elements can remain within the vP-edge (scrambling), while [focus] and wh-elements may move through this edge to CP.

Type (e) includes English, North Germanic and the Romance languages, which permit little or no scrambling, but do feature overt wh-movement. In these languages, C and (non-island-inducing) D are specified as for Type (a), (b) and (c) languages, but v is not associated with substantive features of any kind; it only bears ^, serving as an escape hatch for movement on to C.

2.6 Alignment

The fifth and final hierarchy concerns alignment, in the general sense of how the core grammatical functions are marked in the case/agreement system. Here we present a version of this hierarchy proposed in Sheehan (2013):

(16) **Hierarchy 5: Case and alignment**

a. Does transitive v assign θ-related case (ERG) to its specifier in L?

   No: accusative  

   b. Yes: Do all v in L assign ERG?

      Yes: Split S  

      (Chol, Basque)

   c. No: Does v_{ERG} bear an EPP feature in L?

   8 Though Spanish may have scrambling in marked VOS orders (see Ordóñez 2000, Gallego 2013).
The first option distinguishes the familiar accusative alignment, found covertly in English, overtly in Latin, Russian, Japanese, etc., from all non-accusative systems. The second parameter separates split-S languages, also known as *stative-active languages*, which show ergative alignment only with the single argument of an unaccusative verb (cf. Mithun 1991, Laka 2006).

The third distinguishes languages in which ergative alignment is purely a matter of case and/or agreement marking (cf. Anderson 1976) from those which disallow the A’-extraction of ergative-marked DPs (a property which has come to be known as ‘syntactic ergativity’, cf. Coon et al 2011). The final parameter concerns the source of absolutive case and hence the extent to which the absolutive argument shows “subject properties” of various kinds (ability to be controlled in non-finite clauses, absence in non-finite contexts). In transitive clauses, then, the internal argument can display these properties in High-ABS languages, because Absolutive is uniformly assigned by T (cf. Legate 2008, 2011).

Hierarchy 5 departs a little from the form of the earlier hierarchies. The hierarchies in (12-15) all have at the highest node the question of whether the relevant property is instantiated in the system at all (the “head-final” feature, φ-features, head-movement and A’-movement respectively). In this way, the highest option maximally satisfies both EF (no feature) and IG (generalisation of the absence of the feature). It would be possible, obviously, to add a macro-parametric option to the top of hierarchy 5, determining whether structural Case – and therefore
A-movement – is present in a given language (see Diercks 2012). As the parameterisation of structural Case remains somewhat controversial, though, we leave this option open here. This minor difference aside, the five parameter hierarchies provide a fairly rich characterisation of the grammar of natural languages and, we argue, create new possibilities regarding the calculation of grammatical complexity.

3. Complexity and Emergent Parameters

Consider now how the parameter hierarchies presented above might contribute to an understanding of grammatical complexity. First, the lower positions in the hierarchies correspond to more microparametric options: going down a given hierarchy, we move from macro- to meso- to micro-variation (as noted above, nanovariation is lexically idiosyncratic and thus in a sense outside the hierarchies). Second, the lower options behave in a non-uniform, differentiated fashion which is inherently more complex than the systems defined higher up. Third, each parameter hierarchy can be thought to define a learning path, much in the sense of Dresher (1999), with the higher options inherently preferred by the acquirer, because Input Generalisation favours the higher options in the absence of PLD regarding more specified options. Finally, where hierarchically lower options rely on low-frequency components of the input, we predict Input Generalisation to lead to overgeneralisation, which may, in turn, lead to the loss of such options, resulting in a less complex system. Essentially, highly irregular “low” options will either be lost or “analogised out of the system” over time (note in this connection the close similarity between Input Generalisation and the neo-grammian notion of analogy).
There are at least two distinct ways to calculate the notion of complexity using hierarchies and we will now consider these in turn.

3.1 Complexity and Probability

One way of calculating parametric complexity involves equating complexity with probability. We reason as follows: all else being equal, there should be a roughly 50/50 chance of a taking a given option at each independent choice point, making lower positions in the hierarchy cumulatively less probable. We can quantify the probability associated with a given output of the hierarchy as $0.5^n$, where $n$ is the level of embedding in that hierarchy. (17) provides a dummy hierarchy with dummy relative probabilities:

\[
(17) \quad \begin{array}{c}
\text{Macro-option 1} \\
\text{Y: } p=0.5 & \text{N: Macro-option 2} \\
\text{Y: } p=0.25 & \text{N: Meso-option 1} \\
\text{Y: } p=0.125 & \text{N: Meso-option 2} \\
\text{Y: } p=0.0625 & \text{etc.}
\end{array}
\]

We look at a language’s position in each of the five hierarchies (to the extent that this is feasible), assign a value for $p = 0.5^n$ in each case, and then give the product of the five
independent probabilities. This gives a complexity index for (the syntax of) each language, equivalent to the probability of this grammar. The smaller this value, the more complex the grammar of the language.

Several interesting points immediately arise concerning this way of reasoning, in conjunction with the fact that we have posited five interacting, but independent hierarchies. The first point is that it seems unlikely that any system will be maximally unmarked (in featural terms). To be maximally unmarked would entail being, as it were, “at the top” of all 5 hierarchies. Although this would be the preferred option in terms of maximal satisfaction of both EF and IG, it may be ruled out for independent reasons. Consider what the properties of the least-marked possible system would have to be:

(18) a. Harmonically head-initial;
    b. Radical pro-drop;
    c. High analyticity;
    d. No A’-movement (i.e. no mechanism of focussing, topicalisation, \(wh\)-movement, scrambling)
    e. Accusative alignment (or no Case if hierarchy 5 is expanded upwards)

We conjecture that no language has a system of this kind. Thai, Vietnamese and possibly other South-East Asian languages come close, but all of them, to our knowledge, show some word-order disharmony (final modals, some head-final orders within the nominal; see i.a. Duffield 2001, Enfield 2003 and Simpson 2005) and they also permit information structure-related A-bar movement (cf. Hinds 1989, Phimsawat 2011, Badan & Del Gobbo 2011). Many creoles also
come close, but all feature $wh$-movement. If the maximally unmarked system were found, the prediction is that it would presumably represent a “basin of attraction” in that it would be impossible – or at least extremely difficult – for such a system to change, for the reasons we discussed above in relation to the diachronic conservativity of macroparametric settings. An important question, then, is why such systems do not seem to exist.

At this point, functional considerations come into the picture. We propose that certain options, which are left open by UG in principle, are impossible in practice for functional reasons (cf. Biberauer, Holmberg, Roberts & Sheehan 2010, Biberauer, Roberts & Sheehan 2013, Sheehan 2013, and Biberauer 2011, 2013). This is perhaps clearest in the case of the least-marked options in Hierarchy 4. As we saw, to be consistent with our general markedness conditions and with the first three hierarchies, the most unmarked system here, as dictated by FE and IG (see (10)), is that in which there are no A’-movement triggers at all. We take it that UG in principle allows such an option, but that functional considerations rule it out of the parametric “gene pool”: no system entirely lacking a formal means to focalise/topicalise constituents is likely to survive as it falls short of basic expressivity needs – it arguably undermines one of the two types of semantics in Chomsky’s “duality” (cf. Fortuny 2010, Biberauer 2011). Since UG, as a formal system, is entirely indifferent to questions of expressivity, the formal options exist, but, for reasons to do with thought, conceptualisation and communication – i.e. the cognitive systems UG interfaces with – they are never instantiated. Parameters which offer this non-choice are referred to as “no-choice” parameters. Another important factor is contingent pressure on languages from language contact. While we take contact-induced change to be constrained by UG and its interaction with more general non-language-specific considerations, such change may follow a different path from internally-triggered change. As such, marked systems can come into
existence out of less-marked systems because of (i) functional pressures and/or (ii) language contact.

We are now in a position to address the central question of this paper: do grammars differ in complexity and, specifically, do the differ in terms of complexity defined in terms of parametric probability? To answer this, we propose a thought experiment (in advance of the real experiments, which are the object of ongoing work). Applying the formula discussed above, as we go down a given hierarchy, the probability of being assigned a given parameter value decreases as a function of depth: \( p = 0.5^n \) (where \( n \) = level of embedding; cf. (18)). We can then calculate the probability of a given language by multiplying together all these independent probabilities (assuming, for illustrative purposes, that these probabilities are indeed independent).

So let us see how this works for the grammars of some fairly well-known and well-studied languages across a reasonable typological, genetic and areal range. We leave nanoparameters aside, since it is clear that the degree of complexity added by elements with idiosyncratic formal specifications does not seem amenable to the kind of regular, level-of-embedding-based quantification we are proposing for macro-, meso- and microparametric properties. In this respect, nanoparameters may be just “noise” from the point of view of computing overall complexity. Moreover, considerations such as frequency need to be taken into account in some way when calculating the complexity added by irregulars of different types. We leave these questions aside for the present, focusing on parametric options that seem more amenable to quantification on the basis of the hierarchies that we have been discussing.
First, **English** is: (basically) harmonically head-initial (0.5 on Hierarchy 1), non-pro-drop (0.125 on Hierarchy 2), shows Aux but not V-movement (0.03125 on Hierarchy 3), has *wh*-movement but no scrambling (0.03125 on Hierarchy 4), and is accusative (0.5 on Hierarchy 5). The product of these probabilities is 0.003%, making English a relatively complex language.

Consider next **Mohawk** (here our information comes from Baker 1996 and the references given there). Leaving aside head-initiality/finality for a moment for a reason that will immediately become clear, this language has pronominal arguments (0.25), polysynthesis (0.25), free word order (0.25), and split-S alignment (0.25). Baker (1996) argues extensively that it is impossible to ascertain the nature of head-complement order owing to the language’s pronominal-argument, polysynthetic nature, which has the consequence that all nominal arguments, both in the clause and inside the nominal (e.g. possessors) are adjuncts which can appear either left- or right-adjoined to the clause/nominal. If this is true, then we can infer that the basic word-order parameter is never set. This scenario is arguably problematic in the context of traditional parametric approaches; in the context of the emergentist approach argued for here,

---

9 Biberauer & Roberts (2012a) tentatively place the Modern English Aux-movement option at the 3rd level of embedding in the verb-movement hierarchy they propose. But this hierarchy must be further embedded in Hierarchy 3. If it is embedded at the deepest right branch in (14), then the English option would be at the 5th level of embedding. For expository purposes, that is what we assume here.

10 One might question this conclusion on the basis of the fact that Mohawk features complement clauses that seem rather similar to English *that*-clauses, both in respect of the fact that they systematically surface postverbally and in respect of their being introduced by an optional complementiser-like element, *tsi* (cf. Ikeda 1991 for further discussion).
however, it simply entails that no question ever arises regarding the presence of head-finality in the system, with the consequence that the word-order parameter is set to head-initial, giving a complexity value of 0.5 in this domain. The product of these probabilities is 0.195%. Hence we see that Mohawk is, perhaps surprisingly, somewhat less complex than English.

**Mandarin Chinese** (Huang 1982, 2013, Huang, Li & Li 2009 is harmonically head-final in [+N] but not in [+V] (0.0625), radical pro-drop (0.5), highly analytic (0.5), has topicalisation to the left-periphery (Badan & Del Gobbo 2011), scrambling (Soh 1998), no \textit{wh}-movement (0.125), and accusative alignment (0.5), so the overall probability for this language is 0.098%, meaning that Mandarin falls somewhere between English and Mohawk in terms of complexity (calculated in terms of probability).\footnote{There is a question whether a language which lacks \( \phi \)-features can be considered to be accusative. It is possible that Hierarchy 5 is not even activated in a system which lacks \( \phi \)-features and Case. We leave this matter to one side here.}

**Japanese** is harmonically head-final (0.25), radical pro-drop (0.5), agglutinating in both verbal and nominal domains (0.5; see Julien 2002 and Neeleman & Szendrói 2007), \textit{wh}-in-situ + scrambling (0.125) and accusative (0.5). This gives an overall complexity index (probability) of 0.391%, making Japanese less grammatically complex than Mohawk.

Finally, **Basque** is harmonically head-final (0.25), has pronominal arguments (0.25), is agglutinating (0.5), has \textit{wh}-movement+scrambling (0.125) and split-S alignment (0.25), giving a complexity index of 0.098%, identical to that of Mandarin Chinese.

### 3.2 Complexity indices
A further method for calculating the grammatical complexity of a language using the hierarchies involves assigning each output a complexity index directly, based on the number of choices it entails and taking the average across all five hierarchies. This approach distinguishes the notion of complexity from that of probability, and interestingly gives a slightly different picture for the languages under discussion. The following diagram indicates the complexities associated with the various outputs of a binary branching hierarchy of the kind we have been discussing (the higher the index the higher the complexity):

(19) Macro-option 1

Y: c=1  N: Macro-option 2

Y: c=2  N: Meso-option 1

Y: c=3  N: Meso-option 2

Y: c=4  N: c=4

We can now use these complexity indices to calculate the average grammatical complexity of a given language, where this time, the higher the number, the more complex the language. Consider first English. It is (basically) harmonically head-initial (c=1 on Hierarchy
112), non-pro-drop \((c=3\) on Hierarchy 2), shows Aux, but not V-movement \((c=5\) on Hierarchy 3), has \(wh\)-movement but no scrambling \((c=5\) on Hierarchy 4), and is accusative \((c=1\) on Hierarchy 5). This gives an average complexity index of 3 across the five hierarchies.

Now consider **Mohawk**, which, as discussed above has pronominal arguments \((c=2)\), polysynthesis \((c=2)\), free word order \((c=2)\), split-S alignment \((c=2)\), and head-initial word order \((c=1)\). Mohawk thus has an average complexity index of 1.8, again slightly less complex than English.

Applying the same methodology to Mandarin, Japanese and Basque gives the following overall picture:

\[
\begin{align*}
\text{(20)} & \quad \text{Japanese: 1.6} \\
& \quad \text{Mohawk: 1.8} \\
& \quad \text{Mandarin: 2} \\
& \quad \text{Basque: 2} \\
& \quad \text{English: 3}
\end{align*}
\]

Compare the results from the probability-based calculations:

\[
\begin{align*}
\text{(21)} & \quad \text{Japanese: 0.391\%} \\
& \quad \text{Mohawk: 0.195\%}
\end{align*}
\]

12 If the highest option in a hierarchy does not in fact involve an explicit choice, as suggested above, it may not be correct to assign this option a choice-count of 1. We leave this matter aside here.
Mandarin: 0.098%
Basque: 0.098%
English: 0.003%

We observe, then, that, while the overall complexity indices are different, the relative complexities are identical whether we calculate complexity by probability or choice-count. We stress that these calculations are intended as purely illustrative, as a “proof-of-concept” of the idea that our parameter hierarchies can yield ways of quantifying the complexity of a given grammar. At this stage, we are comparing just the relatively “large-grained” properties that are the focus of our current research.

A couple of comments are worth making here. First, we can observe that there are no extreme outliers (except possibly English if probabilities are used). This is an encouraging outcome, although of course this conclusion is extremely tentative, being based on just five languages. Second, English emerges, in both instances, as the most complex language, which is not surprising, in particular given the contribution made by what is known to be a very unusual and marked property of this language, namely its auxiliary system.\(^{13}\) Third, Japanese seems remarkably simple if we compare it to English, Basque and Mandarin in particular. In connection with Japanese, it is worth pointing out that harmonically head-final word order, radical pro-drop, agglutinating morphology, \(wh\)-in-situ with scrambling and accusative alignment are properties of

\(^{13}\) Of course, the measured complexity of English could also be an effect of the fact that it is the best-studied language in generative grammar. Note crucially, though, that an English bias might lead us to expect it to be the least marked system, contrary to fact. In this way, the result is doubly encouraging.
a large number of languages spoken across the northern part of Asia (essentially the allegedly “Altaic” languages, comprising Japanese, Korean, and the Turkic, Mongolian and Tungusic families; Menges 1975). We do not wish to assert that this supports any version of the Altaic hypothesis, but it is possible that the regional prominence of these properties may reflect the fact that they are relatively unmarked properties which are therefore of the kind that we might expect to spread readily through contact. Nichols (1992) identifies North Asia as one of the world’s major “spread zones”, i.e. areas in which the topography permits extensive and innovative language contact (see also Dryer 1998).

A final point is that of course here we are only looking at (morpho-)syntactic complexity; it would be revealing to carry out a similar exercise in relation to phonological properties, whereafter the further questions naturally arise of whether and, potentially, how phonological and morphosyntactic complexity might be combined to give an overall picture of the formal complexity of language systems. But that would go well beyond the scope of this paper.

4. Conclusion

We have tried to illustrate in the foregoing a new approach to parametric variation, which, we believe, has a number of virtues. First, it overcomes the earlier objections of Newmeyer (2005) to classical parametric theory. Second, it can be reconciled with certain types of functionalist approaches (cf. the notion of “no-choice” parameter introduced in §2). Third, it makes new diachronic, typological and acquisitional predictions and fourth, it is fully compatible with minimalist assumptions on language design, arguably allowing certain aspects of Chomsky’s third factor to be made more precise. Finally, as we saw in §3, our approach may allow for an
overall quantification of the complexity, and hence perhaps the markedness, of grammars. For all of these reasons, we believe our approach to be worth considering and developing further.

We close with a final remark on terminology. It is possible that the term “parameter” may no longer really be appropriate, as the sense in which we understand it is fairly different to its original sense in Chomsky (1981). However, two considerations lead us to retain the term. First, introducing a new term would almost certainly create unwelcome terminological confusion. Second, we see this work as maintaining the spirit of the principles and parameters approach. It should not be forgotten that technical scientific terms change their denotations, both their extension and their intension, as knowledge progresses. In this respect, the ways in which the denotation of “syntactic parameter” has changed since 1981 should be seen as a sign of progress.

REFERENCES


