

1. Background

The theory of principles and parameters as first proposed in detail in Chomsky (1981) and developed in versions of the minimalist programme for linguistic theory (see in particular Chomsky 1995, 2005) offers a promising approach to the classical question of explanatory adequacy (Chomsky 1964). It is much less clear, however, whether this approach offers a way of approaching the question of complexity. This paper aims to address exactly 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 principles and parameters theory. The central idea of this theory 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 faculty of language 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 the

* This work is funded by the European Research Council Advanced Grant No. 269752 “Rethinking Comparative Syntax”. Many thanks to the other members of the project at the time of the University of Washington Workshop on Complexity, Jenneke van der Wal, Tim Bazalgette and Alison Biggs, for comments and critique. Thanks also to Fritz Newmeyer, and the other participants at that workshop, especially Edith Aldridge, Ray Jackendoff and Andrea Moro.

parametric options create the space of variation which is occupied, at least in part, by the typological diversity attested in actually occurring (E-)languages.

The principles and parameters (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 (island constraints, etc.); a particular grammar as a system of language-specific, construction-specific rules; and language acquisition as rule induction, aided by an evaluation metric (see i.a. Peters 1972, Baker 1979, McCarthy 1981, and Lasnik 1981 for discussion of the importance of this latter component). This theory offered little hope for insights into either language typology or language acquisition. The P&P approach 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, the P&P approach 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 (the study of the innate language faculty) 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.

A first point is that, contrary to what was first thought (and explicitly stated in Chomsky 1981), an evaluation metric is still needed. This was first pointed out, to our knowledge, by Clark (1992). Clark observed that a parameter-setting approach to acquisition did not on its own guarantee convergence on a single grammar, which would therefore by definition be the most highly valued since it is the only possible one given the interaction of UG and the PLD. Instead, he pointed out that it is possible to construct fairly plausible scenarios in which UG (even one reflecting markedness considerations in the form of default parametric specifications) and the PLD alone lead to indeterminacy, and so some further factor must be at work in order to

guarantee convergence (his example involves what he called “shifting” grammars in relation to the Subset Principle, but the details of his illustration need not detain us here; see also Clark & Roberts 1993 on this and the related notion of P(arametric)-ambiguity, and Gibson & Wexler 1994). This further factor chooses among possible UG-sanctioned grammars consistent with the PLD on the basis of relative complexity, and as such corresponds very closely to the earlier notion of an evaluation metric. Hence what was thought to be a conceptual advantage in adopting P&P theory – the elimination of the evaluation metric – turned out not to hold (note that this holds however we construe the evaluation metric; the point is a conceptual one, somewhat distant from the detail of language acquisition).

The second difficulty encountered by P&P theory in recent years reflects a more directly empirical issue. As research in comparative syntax has advanced, naturally enough many more parameters than originally envisaged have been proposed in order to account for observed cross-linguistic variation. Descriptively, 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 (superficially non-evident) crosslinguistically recurring patterns has increased enormously since 1981. At the absolute minimum, then, 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, in this case ever more parametric variation, and the need for explanation, which requires us to eliminate as many entities as possible, including parameters. In other words, parametric descriptions as they have emerged in much recent work (see in particular Manzini & Savoia 2005, which is a work of monumental empirical breadth, but which poses serious conceptual problems given the richness of the set of parametric options presupposed) tend to sacrifice the explanatory power of the notion of ‘parameter 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 the P&P approach in solving this problem, which at the outset seemed so clear, is in question.

Newmeyer (2004, 2005) was the first to construct a detailed critique of P&P theory, concluding that it was not living up to its initial promise. He put forward a number of criticisms of the approach, not all of which we agree with (see in particular the discussion in Roberts & Holmberg (2010) and Biberauer (2008)), and we certainly do not endorse his conclusion that the

approach should be abandoned. But Newmeyer (2005: 83) makes one extremely telling point, as follows: “... 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, as standardly formulated, 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, 1989, Fodor 2001, Tomasello 2005, and Gervain & Mehler 2010; Dresher 1999 refers to the same problem as the *Epistemological Problem*). 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 (see the above-mentioned authors and also i.a. Berwick 1985, Gibson & Wexler 1994, Evers & van Kampen 2008, Gervain & Mehler 2010, and Fasanella & Fortuny 2013 for discussion). It is every bit as unclear in P&P theory as in almost any other approach to language acquisition (certainly the pre-P&P generative accounts) how this happens.

Newmeyer’s point as summarised above holds in full force if learners must link and set all of the many hundreds, perhaps thousands, of putatively 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, as has often been proposed, then this may simplify the learning task substantially (and, with a rich enough theory of the nature of the interconnections¹, it may even permit us to circumvent the objection entirely). 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 will try to develop in what follows.

¹ Cf. Baker (2008) for a discussion of micro- vs. macro-parameters, and also the discussion to follow.

Let us begin with a very simple – and, indeed, as far as the syntactic details are concerned, greatly over-simplified – example. It has been known since at least Kayne (1981) that languages vary as to whether they allow “Exceptional Case-marking” infinitives or not.² English allows this construction and French does not (in the canonical context involving *believe*-type verbs; see Kayne 1981 for further discussion of French):

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

So we are led to posit a parametric difference between English and French which tells us why English has this possibility and French does not. It is also known that some languages have no infinitives at all, e.g. Modern Greek. In a language of this kind, the constructions corresponding to raising, control and (in English) ECM typically involve finite clauses (in the indicative or subjunctive; the complement to *believe* is in the indicative, while the complement to *expect* is in the subjunctive, D. Michelioudakis, p.c.), as shown in (3):

- (3) O John pistevi oti o Paul ghrafi ta kalitera traghudhja.³
 the John believes that the Paul.NOM writes the best songs
 ‘John believes that Paul writes the best songs’

So there are at least three options made available by UG: some languages have both ECM and non-ECM infinitives (English), some languages have non-ECM, but no ECM infinitives (under *believe*-type verbs), e.g. French, and, third, some languages have no infinitives at all (Modern Greek). It is clear that if the learner figures out that it is dealing with the third option then it does not need to set the ECM vs non-ECM option. Choosing 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 simply less for the learner to do. At the same time as appreciating the role of simplicity here, we see that there are advantages from the learnability

² Here we take no view on the question of whether this kind of infinitive involves “raising to object” or not (see Postal 1974, Lasnik & Saito 1992, etc.). For us, the cross-linguistic observations are what matters here, not the details of analysis, although of course these are ultimately extremely important.

³ Thanks to Dimitris Michelioudakis for this example.

perspective in linking parametric options, and of course the more “intrinsic” those links can be, as in this little example, the better.

Moreover, the minimalist programme (Chomsky 1995 *et seq.*), while continuing to pursue the basic P&P idea, offers the possibility of seeing the nature of parameters from a very different perspective from the earlier one discussed above, and in a way which quite clearly offers a solution to the very real problem identified by Newmeyer. To see this, consider the three factors of language design put forward in Chomsky (2005):

- (4) 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 (although we concur with Chomsky’s conceptual argument that there is great naturalness in specifying a biological capacity in terms of what is innate, what is triggered by the environment, and the restrictions imposed by general laws of nature and logic). Below, we will attempt to do this in relation to parametric variation and language acquisition. The general view that we take, then, is that parametric variation is an emergent property of the interaction of the three factors listed in (4), 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 potentially very 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 proposal

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_{HL} :

- (5) a. certain formal features;
- b. recursive, binary Merge;
- c. a labelling algorithm;
- d. Agree (feature-valuation, relating elements of syntactic structures).

Obviously much more needs to be said about all of (5a-d).⁴ For present purposes, we take the class of formal features to include categorial features ($\pm N$, $\pm V$, etc), structural Case features (or equivalent), person, number and gender features (collectively ϕ -features), other features such as $[\pm wh]$, $[\pm neg]$, $[\pm tense]$, etc, as well as purely diacritic features which simply trigger operations (different kinds of Merge, usually). Sigurðsson (2011) suggests that a single very abstract feature-schema might be all that is required in UG, with the precise content and values of formal features also an emergent, three-factor-driven property, while Biberauer (2011, 2013) allows for the innate specification of a very small number of formal features (e.g. a linearization diacritic, [Person], [Case]), arguing that crosslinguistically variant formal features that can plausibly be acquired on the basis of the interaction between a UG-given $[uF/iF]/[valued/unvalued F]$ distinction and the PLD are emergent. For present purposes, however, we keep to the rather more “conservative” position which attributes some such content directly to UG.

Following Chomsky (1995: 243ff.), we take Merge to recursively combine two syntactic objects α and β to form a set $\{\alpha, \beta\}$, the objects may be drawn from the Lexicon (technically from the Numeration – this is External Merge), or, if the members of an existing set $\{\alpha, \beta\}$ have internal structure, from within α or β (giving Internal Merge). The set formed by Merge requires

⁴ Another potential candidate is 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.

a label K (in other words, Merge creates the object $\{K, \{\alpha, \beta\}\}$ (see Chomsky 1995, Sheehan to appear b for discussion). K is determined by either α or β , giving the effect of “projection” of a syntactic category label, and hence endocentric structures.⁵ 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:___], something 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 in a fairly standard structural sense).⁶ All of this is a fairly mainstream set of technical assumptions; for more details, see Chomsky (2001). This is what we take to be the invariant core of UG.

Aside from the proposals in Biberauer (2011, 2013) just alluded to, our most important 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 head 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.⁷ This underspecification can take three forms, as follows:

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

⁵ Chomsky (2013) proposes a different approach to labelling, which gives endocentricity a much lesser role in syntax, at least potentially. We will leave this alternative proposal aside here, however.

⁶ Following Chomsky (2000, 2001), a Goal must also bear the relevant feature in order to be visible to a given Probe. There are empirical wrinkles facing such a view, however (notably the apparent existence of defective intervention – cf. i.a. McGinnis 1998, Holmberg & Hróarsdóttir 2003, Hartmann 2012 and Bruening 2012 for dicussion), so we leave the matter to one side here.

⁷ It could be noted that what we might call the textbook view of parameters as prespecified options is not actually expressed, or even necessarily assumed in the best known classical works on parameters in the literature. The parameters formulated there can typically be interpreted as defining points of underspecification without any change of formulation; see Roberts and Holmberg (2010: 55-56) and Holmberg (2010).

Certain heads are intrinsically potential bearers of formal features; this set may well be limited to the class of functional heads (though see Pesetsky & Torrego 2006 on the featural make-up of lexical heads, which are also assumed to be (externally) merged on the basis of their *formal* feature specifications). So, 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 (usually finite) verb and the subject is rare in Indo-European,⁸ but found in many Semitic languages including Classical Arabic, for example. Furthermore, as (6b) 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 – consider, for example, Miyagawa’s (2010) proposals in this regard. The options in (6c) concern the distribution of the ‘EPP’ movement-triggering feature, which, following Biberauer, Holmberg & Roberts (to appear, BHR henceforth) we write as “^”. This sub-feature can be associated with any kind of syntactic dependency triggering Internal Merge, with languages differing in relation to which dependencies are associated with ^.

There is clearly a close relation between (5) and (6). In fact, (6) really says that a subset of the core properties of UG is optional in a given instantiation of UG (an 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 a very apt formulation put forward by Biberauer & Richards (2006), parametric variation emerges where UG “doesn’t mind”.

To elaborate slightly and give some more concrete examples of (6), (6a) includes such options as the mapping of features to heads, i.e. feature-scattering vs. feature-syncretism (see Giorgi & Pianesi 1997 on the latter, and see Biberauer & Roberts (2013) for a third-factor driven elaboration of this approach); presence vs. absence of features (e.g. [Gender], [Tense]) on heads, as already mentioned; differing distribution and internal make-up of properties such as finiteness in clauses – this may underlie the variation between English, French and Modern Greek

⁸ The Italo-Romance variety of Ripatransone is a rare case of this. See Ledgeway (2012: 299-310) for discussion and illustration.

discussed above. Variation in Agree, and in its overt 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 for discussion), gives rise to the differing properties of subject-agreement in English vs Italian vs Japanese, etc., for example, and also to “doubling” effects, e.g. Negative Concord, ‘forked’ modality in many South-East Asian languages (see Cheng & Sybesma 2003), “bracketed” relative clauses (see Peng 2011 on Jambi Teochew and Bradshaw 2009 on Oceanic languages spoken around the Huon Gulf, and Hendery 2012), and many other phenomena. Finally, (6c) 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 (6b), to different kinds of case systems including arguably the distinction between (the various kinds of) ergative and accusative systems (cf. Sheehan 2013); and possibly variation in “basic” head-complement order (see BHR and Sheehan to appear for opposing views on this last point).

In essence, (6) reduces to the statement in (7):

- (7) 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:

- (8) $Qh_{h \in P} [F(h)]$

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

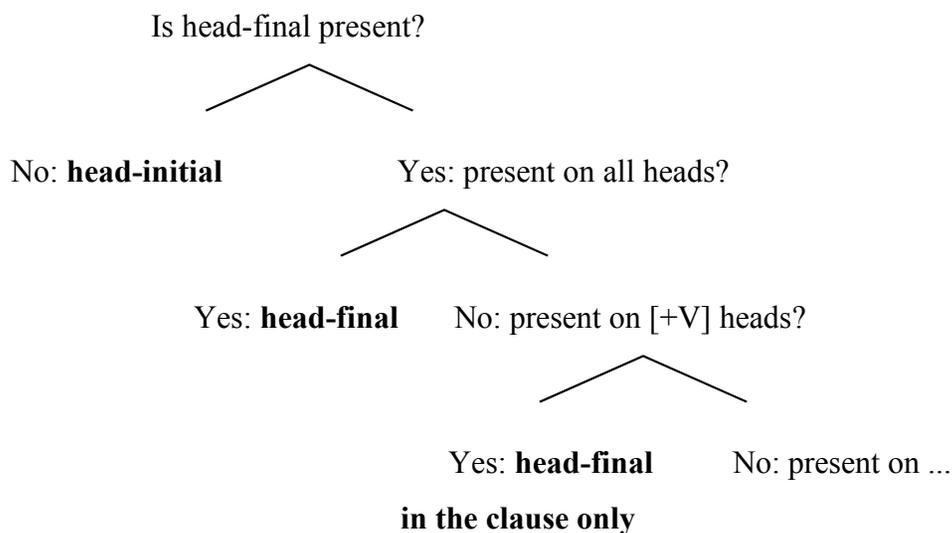
This approach gives rise to the following informal taxonomy of parameters (Biberauer 2011 Biberauer & Roberts 2012a,b, 2013):

- (9) For a given value v_i of a parametrically variant feature F :

- a. **Macroparameters:** all functional heads of the relevant type 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 clear that the different kinds of parameters listed in (9) are hierarchically related to one another. So we are led to postulate different kinds of parameter hierarchies. Roberts (2012) suggests the following (among others, see below):

(10) ***Hierarchy 1: Word order:***



Here we use the neutral term “head-final”. This can be seen 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 (10) 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 (9). The third option approximates to the typical Continental West Germanic situation (in which all clausal heads except C and probably T follow their complement); by the definition in (9), this is 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). These can be stated as follows:

(11) 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 .

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, with the result that so-called *superset traps* are circumvented as a consequence of the fact that 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).⁹ We take the

⁹ 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). Worth noting here is that approaches assuming the set of formal features employed within a particular language only to be partially specified by UG are radically maturational in the sense that there are no or, at best, very few pre-given categories to map the input to. In this case, the acquirer cannot in any meaningful sense be said to find himself in a “superset trap” as a result of “erroneous” application of IG: this will simply be the consequence of his not yet having acquired a distinction present in the target system; as soon as the relevant distinction is acquired, the question whether an existing pattern is to be extended to the new (sub)category created by this

conditions in (11) (perhaps along with the Subset Principle (Berwick 1985)) to arise from general cognitive optimisation strategies, *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 (11) and the PLD. It is in this sense that parametric variation emerges from the three factors of language design given in (4). 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 (and diachronic) 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 (8): given a head h , the set P of heads bearing the relevant formal properties (ϕ -features, movement-triggering features, etc.), and the set F of features, the general form of hierarchies, as determined by FE and IG, will be as follows:

- (12) a. Hypothesis I (ahead of any experience/analysis of PLD):
No head in P has F ($\forall h_{h \in P} \neg [F(h)]$); 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 ($\forall h_{h \in P} [F(h)]$); 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 ($\neg \forall h_{h \in P} [F(h)]$); both FE and IG overridden by PLD

The left branches of (10) reflect this ordering of (progressively weaker) hypotheses; (12a-c) reflect the **macroparametric** options; at the next level, generalisation ranges over $P' \subset P$, where P' is defined as a linguistically natural class, 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,

distinction arises anew, with IG only continuing if the newly analysed input is compatible with further generalisation (see Biberauer 2011, 2013 for further discussion).

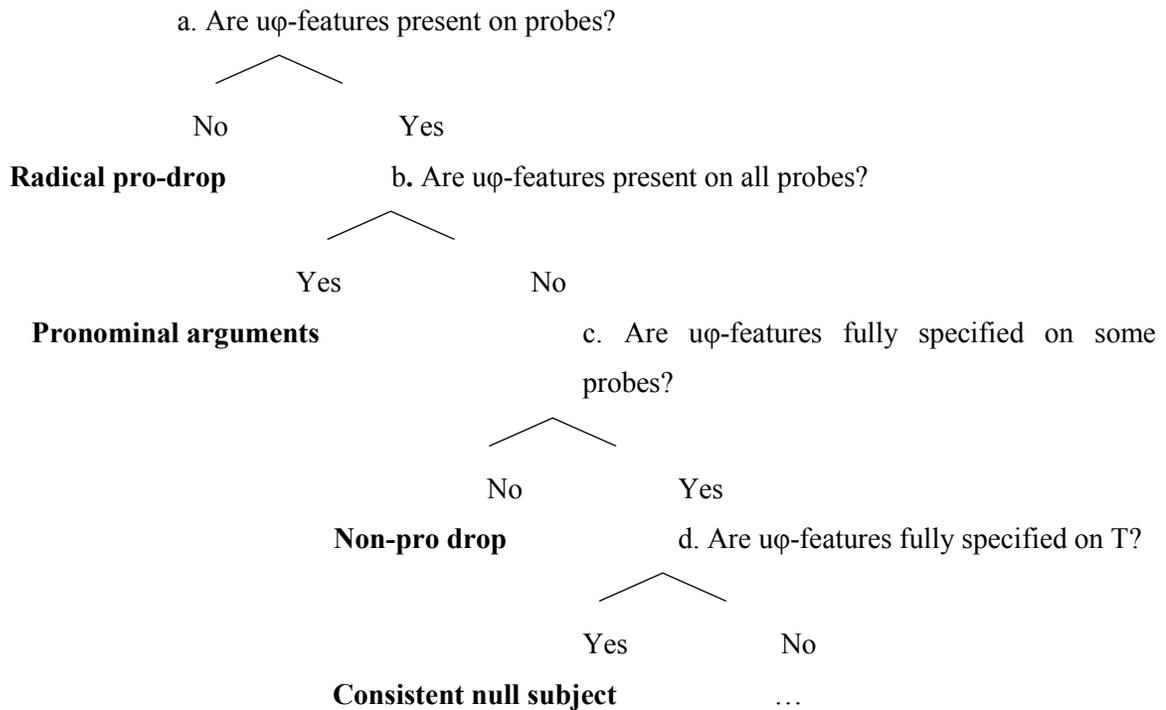
generalising over P gives no clear outcome. FE and IG conspire to make each step refer to the minimal (FE) and the maximal (IG) proper subset of categories, hence the next level is the **mesoparametric** one. The **microparametric** level operates on still smaller subsets $P'' \subset P' \subset P$. The **nanoparametric** level operates on the smallest feasible subset (individual lexical items).

A further hierarchy, first put forward by Roberts & Holmberg (2010:49), concerns null arguments. In terms of (12), we expect it to take the following form:

- (13) 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 $P' \subset P$) has uninterpretable φ -features.

The system of hypotheses in (13) can be graphically illustrated by the diagram in (14), to which we have added a further mesoparametric option at the lowest level shown here:

(14) **Hierarchy 2: Null arguments:**



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).¹⁰ “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 (9): all relevant heads act in concert, showing the same property. Question (c) corresponds to Hypothesis III in (12). 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.¹¹ 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)) or in Mainland Scandinavian; if this is true, then English and Mainland Scandinavian lack fully specified T-probes. Why only T and not at least *v* becomes relevant at this point in this hierarchy is not clear to us at this point, 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 -

¹⁰ Saito’s proposals are not uncontroversial; notably Miyagawa (2010) has given evidence for the existence of φ -features in Japanese. But if Miyagawa is correct, the nature of the hierarchy does not change, only the place of Japanese (and perhaps other radical-pro-drop languages) in it. For the remainder of this paper, we will tentatively assume that Saito’s analysis of Japanese is correct; if it is not, this would have some bearing on the discussion of the relative grammatical complexity of Japanese in §3.

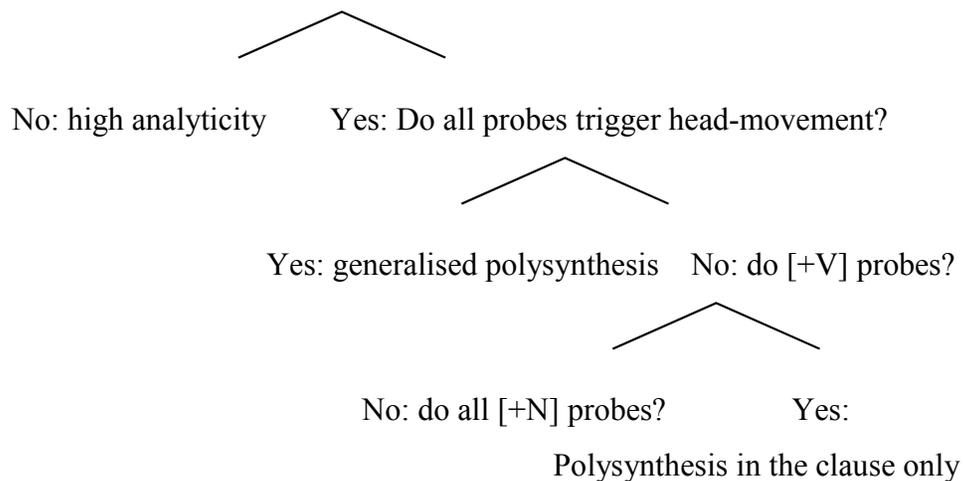
¹¹ It may seem odd to refer to finite T as a *class* of heads, but recall that it is necessarily present in all finite clauses and that it is also morphologically instantiated on all finite verbs and auxiliaries in a given language.

features are only partially specified on T and only in certain contexts. Ultimately, highly specific systems such as Literary French as described in Roberts (2010a) may fall into this class.

A further parameter hierarchy concerns word structure. Assume complex words to be formed by head-movement (cf Baker 1988, 1996, Julien 2000), Hypothesis I here would assume no head-movement, i.e. highly analytic morphosyntax, Hypothesis II highly prevalent head-movement, i.e. polysynthesis, and Hypothesis III more limited head-movement, leading to the mesoparametric options, roughly as shown in (15):

(15) ***Hierarchy 3: Word structure***

Do some probes trigger head-movement?



We take head-movement to be the formal operation which determines the internal structure of complex words. More specifically, following Roberts (2010b), 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” in the sense defined by Roberts 2010b). 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). In addition to its very well-known property of extreme morphological analyticity, Huang identifies a range of syntactic effects of this property, which we construe as showing lack of head-movement in the system as a whole. The next option is polysynthesis, as discussed and analysed in detail 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 and they are high in the relevant hierarchy. 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 mesoparametric 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, lower options in a given hierarchy 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 (see Biberauer & Roberts 2013 for further discussion). We return to this point in section 3 below.

Biberauer & Roberts (2012b) and Branigan (2012) further propose that “lower” parametric options are progressively more diachronically unstable. To put it another way, the “higher” settings are more diachronically conserved. The general picture is as follows:

- (16) a. Macroparameters are stable over millennia.
- b. Mesoparameters are somewhat stable, often characterising genera in the sense of Dryer (1992).¹²
- c. Microparameters are somewhat unstable.
- d. Nanoparameters are highly unstable.

The diachronic patterns follow from the idea, common but not unique to most of the generative literature on language change, that grammatical systems change diachronically owing to abductive reanalysis of PLD by acquirers (the classic statement of this position for syntax is Lightfoot 1979; see Walkden 2011, 2012a for critical discussion). In terms of the kinds of hierarchies seen in (10-12), the “higher” options are so massively attested (or “expressed” to use the terminology of Clark & Roberts 1993) in the PLD that they are readily set, and so relatively immune to the effects of abductive reanalysis. To put it rather simplistically, acquirers, and therefore I-languages, favour higher positions in the hierarchy.

We therefore expect to find examples of highly stable macroparameters, i.e. syntactic properties which characterise whole language families and which we can infer have remained stable for millennia. One example, pointed out to us by Phil Branigan (see Branigan 2012) concerns multiple incorporation in the Algonquian languages. The Algonquian family is spread across the North American continent, and is thought to be about 3,000 years old (Campbell 1997:104; Campbell & Poser 2008: 342). Unsurprisingly then, the family shows a fair amount of lexical and phonological diversity. But one feature is constant: all of these languages show a great deal of incorporation of various kinds. In other words, all of them choose a “high” option in the word-structure hierarchy in (15).

¹² A group of languages related at approximately the level of the main subgroups of Indo-European (Germanic, Italic, Celtic, etc), i.e. to a time depth of about two millennia.

Wichmann and Holman (2009) have carried out an interesting investigation of the relative diachronic stability of the linguistic features in the *World Atlas of Language Structures* (*WALS*; cf. Dryer & Haspelmath 2011). The metric they implement is estimating the stability of a feature by “assessing the extent to which phylogenetically related languages are more similar with respect to the feature than are unrelated languages”. Applied to the features in *WALS*, this metric assigns a numerical stability value (in percentage) to each feature. With the range divided four ways (very stable, stable, unstable, and very unstable), the word-order features (VO vs. OV, P NP vs. NP P, NAdj vs. AdjN, etc.) come out as very stable. They note that this result is not obvious from a European perspective, since many of the languages in Europe have recently (in the last 1000 years) undergone change from OV to VO. However, in a global perspective, this is apparently not a common phenomenon, as indeed predicted by our Hierarchy 1. A word-order system which we know has remained stable for a very long time is harmonic head-final order in Dravidian. According to Steever (1998: 31) this is a feature of almost all extant languages in this family (one or two northerly Dravidian languages, e.g. Maho and Brahui, have some head-initial patterns, but this is almost certainly due to contact with Indo-Aryan, as Steever points out). Proto-Dravidian is standardly reconstructed as rigidly and harmonically head-final, and is dated as being about 4,000 years old (Steever 1998:31).

We expect mesoparameters to be somewhat less strongly conserved, and to perhaps be more open to contact-induced change. Possible examples of mesoparameters might include the null-subject parameter in (most) Romance languages. The positive setting for this parameter has been inherited from Latin (and possibly from Indo-European; see Roberts 1998), and is common to all Romance varieties except certain varieties of French, some Northern Italian and Franco-Provençal dialects and most varieties of Rhaeto-Romansch. These latter varieties are just the varieties of Romance which have had the most contact with Germanic, which may explain why null subjects were wholly or partially lost there. General V2 (at least in all root clauses) in Germanic is another possible case; this property is common to all of Germanic except for English, which lost it in 15th century (see Biberauer & Roberts 2008; Fischer et al. 2000) under conditions which remain unclear. Here there is a difficulty, though, in establishing whether the

property was common to Proto-Germanic or a parallel innovation; for recent discussion and analysis see Walkden (2012b).¹³

Examples of both microparametric and nanoparametric change taken from the Modern English auxiliary system are discussed in Biberauer & Roberts (2012a). Auxiliary movement to T in Modern English was innovated post-1600, emerging from the loss of general V-to-T movement and the grammaticalisation of various lexical verbs as functional elements, primarily the modals and *do* (see Roberts 1993, Warner 1997, Biberauer & Roberts 2008). It is well known that this system is idiosyncratic both typologically and genetically (nothing comparable is found anywhere else in Germanic). Moreover, as argued in particular detail by Denison (1998), the auxiliary system is in a state of ongoing collapse in contemporary varieties of English, giving rise to a great deal of regional variation, reflecting idiosyncratic properties of individual lexical items. Denison illustrates certain striking recent changes in this connection, notably the loss of generalised “conditional inversion” (as in *Should he leave, let me know*), applying to all modals and *do*, in the past century and a half.

We are now in a position to turn to the question of complexity in relation to this conception of parametric variation.

3. Complexity

We begin this section by recapitulating the most important points of the conception of parametric variation given above for any attempt to determine the relative complexity of grammatical systems. First, the lower positions in the hierarchies correspond to more microparametric options; in fact, 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

¹³ The feature in *WALS* which corresponds most closely to the null subject parameter is ‘Expression of pronominal subjects’. In Wichmann and Holman’s (2009) four-way classification (very stable-stable-unstable-very unstable), it comes out as ‘unstable’.

favours the higher options in the absence of PLD regarding more specified options (see Biberauer 2011, 2013 and Biberauer & Roberts (2013) for a detailed illustration of this point in the context of an even “smaller” UG than that schematised in (5)). Finally, where hierarchically lower (i.e. “smaller”) 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 “analagised out of the system” over time (note in this connection the close similarity between Input Generalisation and the neo-grammarians’ notion of analogy).

Before proceeding to a more detailed discussion of complexity, though, we need to introduce two further parameter hierarchies. Hierarchy 4, notably, 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. Hierarchy 4, concerns what, following Kiss (1995), we can loosely refer to as “discourse configurability”. 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]).¹⁴ We assumed 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 (cf. Aboh 2010 for the strong view that information structure is always reflected in the featural make-up of the lexical items making up a Numeration, regardless of the language in question). Crosslinguistic investigation has shown that this is often to left-

¹⁴ Our approach, then, departs from the view that “A-movement is IM (Internal Merge) contingent on probe by uninterpretable inflectional features [i.e. uninterpretable [ϕ]-features – TB, AH, IR, MS], while A’-movement is IM driven by EF” (Chomsky 2007: 24). This seems necessary given that some A’-movements exhibit sensitivity to Relativized Minimality, an empirical scenario that is readily understood if one postulates an appropriate range of A’-features (see Rizzi 2013 for recent overview discussion and references); furthermore, fronting options vary from language to language depending on factors such as whether topics are quantificational or not, implicating the relevance of substantive formal features (see again Rizzi 2013 for recent discussion). These effects clearly cannot be accounted for on the “blind” EF-driven approach to A’-movement suggested by Chomsky.

peripheral positions within CP, vP and DP (cf. the wealth of research in the cartographic framework in particular for discussion, and also Neeleman, Titov, van de Koot & Vermeulen (2009) for discussion from a very different theoretical standpoint, which, however, nevertheless points to the peripherality consideration). 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).¹⁵ 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-cyclicality of *wh*-movement; but see Sheehan & Hinzen 2011 for a different approach). 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 this cannot be targeted by either of the operations Agree or Move¹⁶) 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)¹⁷). Precisely

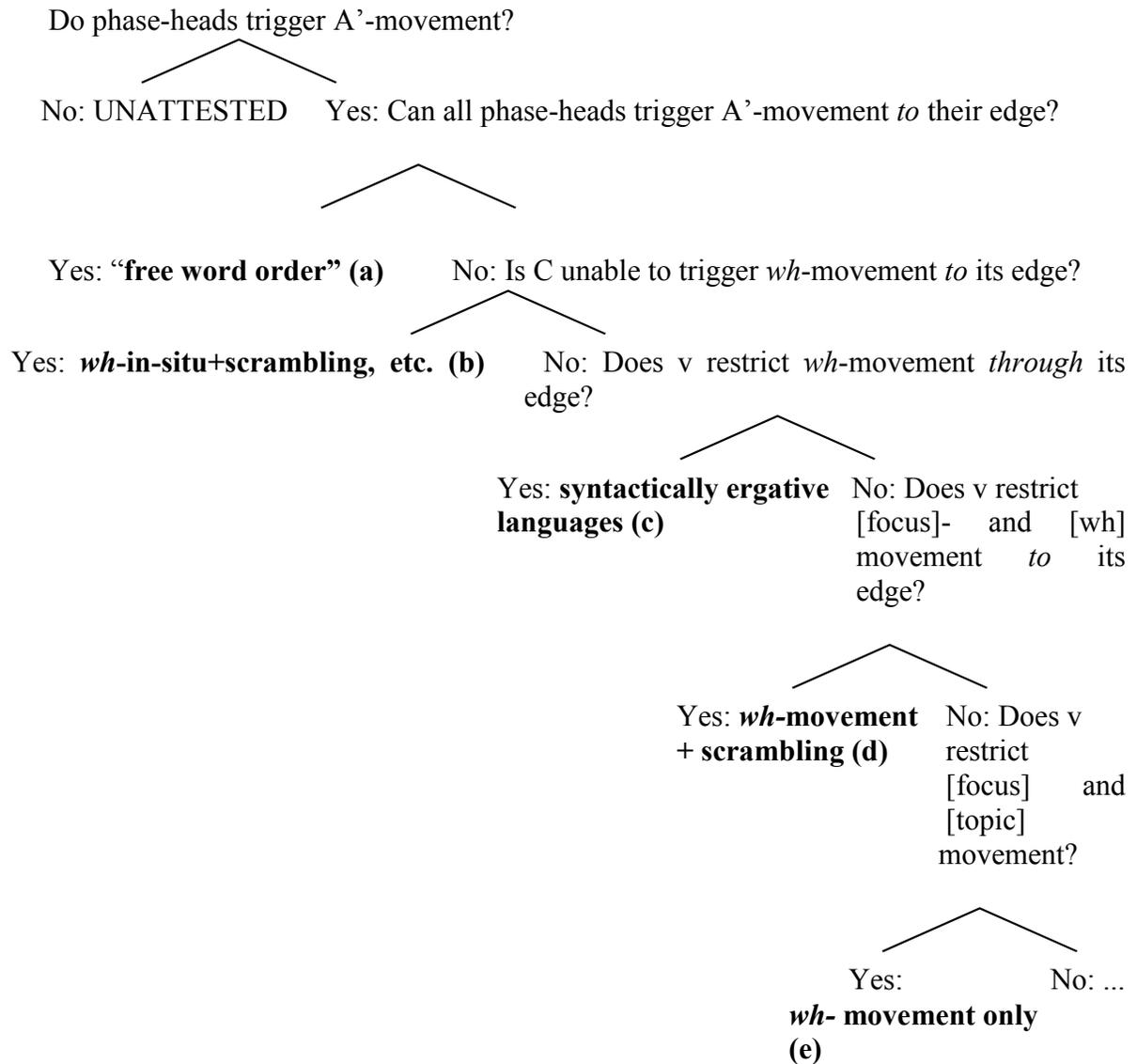
¹⁵ We assume movement *through* a phase edge to be purely Edge Feature-/EPP-feature-/movement diacritic-driven movement, while movement *to* a phase edge involves both a movement diacritic and a substantive A’-feature. We return to this point in the main text; see also note 18 below.

¹⁶ Clearly, this is only convergently possible where the featural requirements of individual elements within the spelled-out domain have been satisfied; the only features that may plausibly remain unvalued in the context of a (non-root CP) spelled-out constituent are peripheral features like [Case] and [wh], i.e. features which may plausibly percolate to the phase-edge, from where they are accessible to external probes. Spelled-out constituents of this type may be thought of as islands by virtue of having been atomised, i.e. collapsed into an element which the computational system sees as a head (cf. i.a. Uriagereka 1999, Nunes & Uriagereka 2000, Hsieh & Sybesma 2007, Takita 2010, and Fowlie 2013).

¹⁷ Importantly, this nominal head is distinct from the basic nominaliser (*n*), which may be a universally attested functional head, and also from the clausal nominalisation head in languages which systematically nominalise (some or all) clausal complements, and from the nominaliser that has been argued to facilitate certain types of finite clause complementation (see Biberauer & Sheehan in press for recent discussion and references). All of these heads are, crucially, nominal heads which are plausibly associated with non-left-peripheral components of the nominal phase, i.e. they do not encode [definiteness], [specificity] or other “discourse”-related features of a kind which would interfere with the extraction of an element likewise marked with an A’-feature of this kind (the assumption here

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. More specifically, consider (17):

(17) **Hierarchy 4: A'-movement**



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 will return to the matter of such no-choice parameters

again being that elements will only be visible for extraction to a phase edge if they bear a left-peripheral/A'-feature of some kind).

below. Type (a) languages include Warlpiri 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 adjectival 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 Edge Features/EPP-features (=movement diacritics; henceforth $\hat{}$), and in which the highly specified island-creating nominal head mentioned above (and see also note 17) 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 $\hat{}$, not associated with a specific substantive formal feature¹⁸) and topic-, *wh*- and focus-movement *to* their edge (by virtue of their ability to bear [topic], [wh] and [focus] features respectively alongside $\hat{}$). In other words, 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: specifically, phase-heads bearing [topic] and [focus] features are also associated with $\hat{}$, i.e. probing for [topic] and [focus] results in both Agreement and movement; phase-heads bearing [wh]-features, by contrast, are not associated with $\hat{}$, with the result that *wh*-elements remain in situ (unless they are moved for non-[wh] reasons, e.g. because the element in question is

¹⁸ The idea is that phase-heads endowed with only a movement diacritic will “blindly” target the element located at the outermost phase-edge of the next phase down, i.e. C will target the element located in the highest spec-vP. Bottommost phase-heads within a given clausal or nominal structure will target the element in v’s complement domain (i.e. VP) that bears a [topic], [wh] and/or [focus] feature, i.e. an element specifically associated with a left-peripheral (i.e. A’-) feature, which will consequently be “active”. Importantly, this movement will, as with other successive-cyclic movements, not be Agree-driven as the phase-head in question is not an A’-probe.

topicalised or focalised).¹⁹ Given the clear existence of island effects, the restrictions on extraction from nominals and also on the more general availability of discontinuous constituents (Takita 2010), it also seems correct to postulate the presence of the extraction-barring nominal head.²⁰ 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 (2012) 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.²¹ The properties of Type (c) are clearly also in part determined by aspects of Hierarchy 5, pointing to the fact that the hierarchies may interact with one another. The details evidently require further research. What seems clear, however, is that languages of this type do not necessarily entail the postulation of more distinct types of C, v and D, but that the feature specification of the phase heads is not unified in the way it is in Type (a) and even Type (b) languages: whereas C, v and D in Type (a) languages may all bear [focus], [wh] and [topic] features associated with $\hat{}$ (where they constitute movement-*to* heads) and also independent $\hat{}$ (where they constitute movement-*through* heads), only C and (non-island-inducing) D in Type (b) languages have this profile, while v may only bear $\hat{}$ -associated [focus] and [topic]-features in addition to permitting independent $\hat{}$; in Type (c) languages, in turn, C and D have the same properties as those of Type (b) languages, but v differs from C and D in permitting only a single $\hat{}$ where [wh] is present. Unlike for Type (a) and (b) languages, then, the Type (c)-acquiring child must postulate an idiosyncratic specification for one of the phase heads (s)he needs to acquire.²²

¹⁹ As the hierarchies under discussion here are intended as learning paths and questions are known to be salient in the speech addressed to children (cf. Pine 1994), it is worth noting that it is plausible that A'-related differences relating to the instantiation of [wh]-features should feature high up in the A'-hierarchy.

²⁰ Strictly speaking, the island-creating nominal is an independent lexical item whose presence in a system is not regulated by Hierarchy 4. As such, we should expect to find free word-order languages lacking or exhibiting more restricted options in relation to the discontinuous nominals found in languages like Warlpiri, Latin and the Slavonic languages.

²¹ 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. (2013) 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.

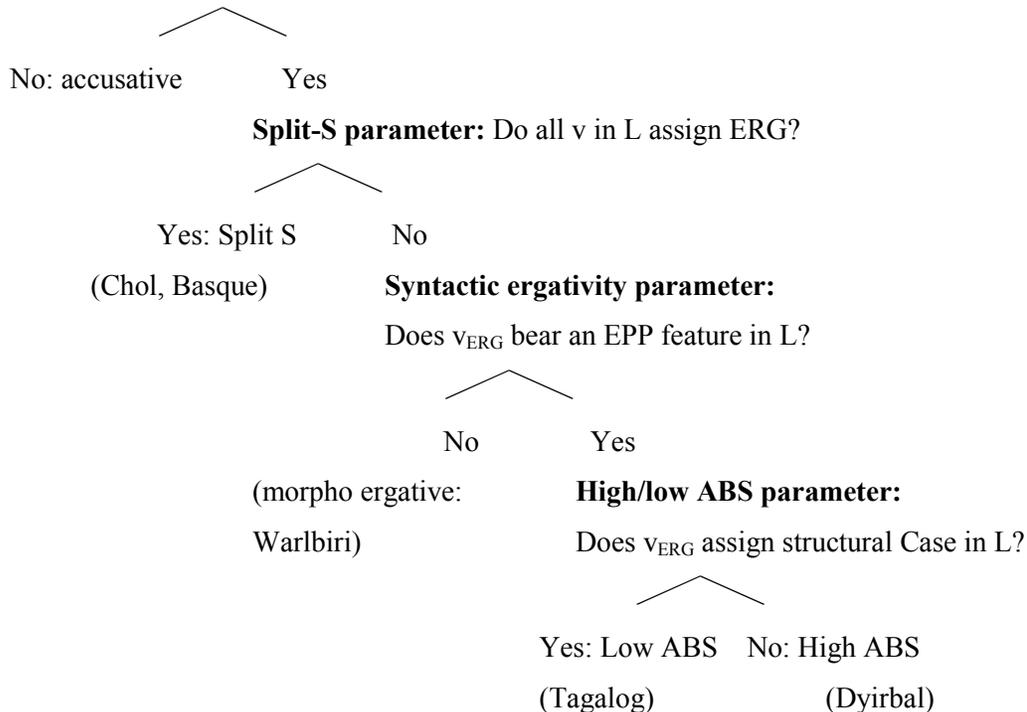
²² Given that Type (d) and (e) systems – which are lower down the learning path in (15) – involve non-idiosyncratic departures from the macro option, in the same way that Type (b) does, one might question whether syntactically

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-inducing) D may once again be specified as for Type (a) and (b) languages, but where v cannot bear $\hat{\text{}}$ -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.²³ In formal terms, these languages involve C and (non-island-inducing) D may once again be specified as for Type (a), (b) and (c) languages, but v in this case may not be associated with substantive features of any kind; it may only bear $\hat{\text{}}$, serving as an escape hatch for movement on to C.

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):

(18) **Basic alignment parameter:** Does transitive v assign θ -related case (ERG) to its specifier in L?



ergative languages should not be located lower on Hierarchy 4. The relevant consideration here would seem to be what role input frequency/salience plays, a matter we must necessarily leave aside here.

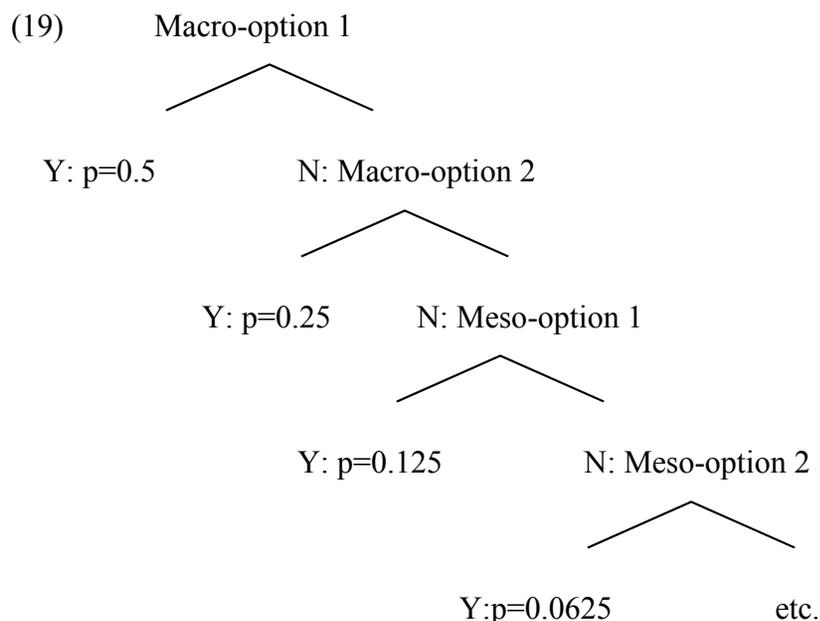
²³ Though Spanish may have scrambling in marked VOS orders (see Ordóñez, 2000).

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-bar extraction of ergative-marked DPs (a property which has come to be known as ‘syntactic ergativity’, cf. Coon et al 2012 and also the discussion above). 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).

These last two hierarchies depart a little from the form of the earlier three hierarchies. This is most clearly the case in relation to lower positions. However, it is also true, as Hierarchy 5 stands, of the highest position; although the first three options follow the “all not > all > not all” pattern of the higher positions in the other hierarchies. The hierarchies in (10-12) and (17) 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 and 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 for discussion). As the parameterisation of structural Case remains somewhat controversial, though, we leave this option open here, subject to further research. These minor differences aside, the five parameter hierarchies provide a fairly rich characterisation of the grammar of natural languages and open up new possibilities regarding the calculation of grammatical complexity.

There are at least two distinct ways to calculate a notion of complexity using the hierarchies. In the first case, we can equate complexity with probability. We reason as follows: all else being equal, there should be a roughly 50/50 chance of a given choice 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. Here is a dummy hierarchy with dummy relative probabilities:



Several interesting points immediately arise concerning this way of reasoning, in conjunction with the fact that we have posited five interacting, but (in principle) independent hierarchies. The first point is that it seems unlikely that any system will be maximally unmarked. 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:

- (20)
- 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 parameter 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 and Phimsawat 2011 on Thai, and Badan & Del Gobbo 2011 for recent discussion and references relating to Chinese). 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 for further discussion of cases of this type). 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 (11)), 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. also Fortuny 2010, Biberauer 2011). Since UG, as a formal system, is entirely indifferent to questions of expressivity, the formal options exist, but, essentially for reasons to do with thought, conceptualisation and communication – i.e. the cognitive systems UG interfaces with directly or indirectly – they are never instantiated. Parameters which offer this non-choice are referred to as “no-choice” parameters. An important additional 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? 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 the 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. (19)). We can then calculate the probability of a given language by multiplying these five probabilities together.

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 look at each 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 each language (i.e. the grammar of that language), equivalent to the probability of this grammar. The smaller this value is, the more complex (and thus less probable) the grammar of the language is. These values are intended to be indicative only of the general proposal, as many questions concerning the precise shape of the hierarchies and also regarding the significance of what we have referred to above as *nanoparametric* properties remain. In relation to the latter, it is clear that the degree of complexity added by elements with idiosyncratic formal specifications does not seem amenable to the kind of regular, probability-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 complex questions aside for the present, focusing exclusively on parametric options that seem more readily 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),²⁴ 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.

²⁴ 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 (15), then the English option would be at the 5th level of embedding. For expository purposes, that is what we assume here.

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,²⁵ 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, 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.²⁶ 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, 2007, Huang, Li & Li 2008) 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), and no *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).²⁷

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 & Szendrői 2007), *wh*-in-situ + scrambling (0.125) and accusative (0.5). This gives an overall complexity index (probability) of 0.391%, making Japanese even less grammatically complex than Mohawk.

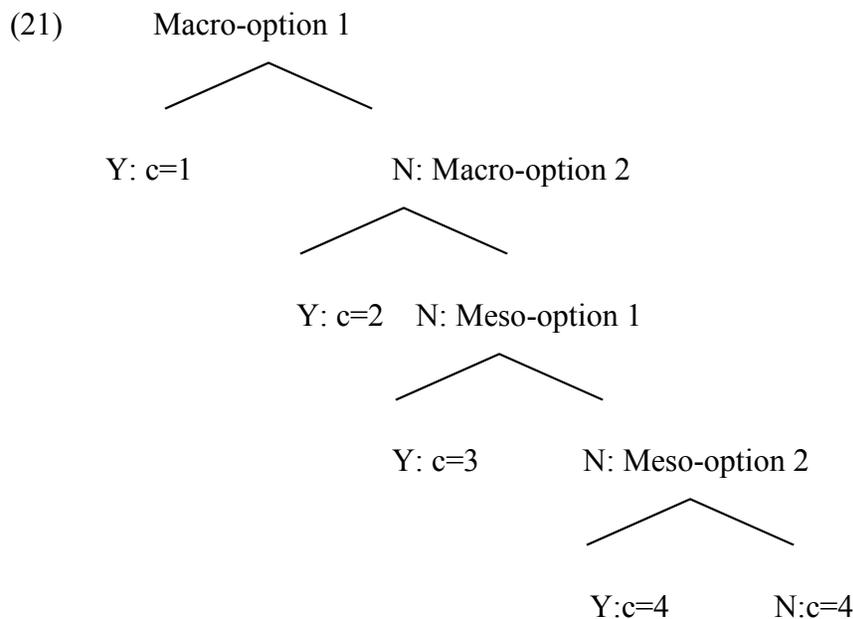
²⁵ One might question this conclusion on the basis of the fact that Mohawk features complement clauses that seem rather similar to English *that*-clauses, both in respect of the fact that they systematically surface postverbally and respect of their being introduced by an optional complementiser-like element, *tsi* (cf. Ikeda 1991 for further discussion).

²⁶ If one accepts the strong empirical arguments in favour of ascribing some type of “marked” status to head-final orders (cf. i.a. Biberauer, Holmberg & Roberts to appear and Sheehan to appear b), the emergentist approach, then, arguably facilitates a non-stipulative take on the notoriously problematic notion of ‘default parameter setting’ in the domain of word order.

²⁷ There is a question whether a language which lacks ϕ -features can be considered to be accusative. It is possible that Hierarchy 5 is not even activated in a system which lacks ϕ -features and Case. We leave this matter to one side here. See Diercks (2012) and the brief discussion above for the proposal that Case is parameterized.

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

The second method for calculating the grammatical complexity of a language using the hierarchies involves simply assigning each output a complexity index directly, based on the number of parametric choices it entails (i.e. the number of parameters that have to be postulated and set) and taking the average across all five hierarchies. This distinguishes the notion of complexity from probabilities, 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 (c = parametric choice):



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 1), 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 head-initial word order ($c=1$), pronominal arguments ($c=2$), polysynthesis ($c=2$), “free” word order ($c=2^{28}$), split-S alignment ($c=2$), and. This gives an average complexity index of 1.8, again somewhat less complex than English.

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

- (22) Japanese: 1.6
- Mohawk: 1.8
- Mandarin: 2
- Basque: 2
- English: 3

Interestingly, while the indices themselves are different, the relative complexities calculated in this way, are identical to the relative complexities calculated in terms of probabilities, as discussed above and summarised in (23):

- (23) Japanese: 0.391%
- Mohawk: 0.195%
- Mandarin: 0.098%
- Basque: 0.098%
- English: 0.003%

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 (early-stage) 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), which is an encouraging

²⁸ This index is based on the assumption that the “non-choice” instantiated by the first parametric option (i.e. NO to “Do phase heads trigger A’-movement?”) counts in the same way as a regular option yielding two “real” choices. This may not be correct, but the issues involved go beyond the scope of this paper.

outcome. Second, English emerges, in both instances, as the most complex case, 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.²⁹ 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 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 at least possible that the regional prominence of these properties may reflect the fact that they are, as suggested by our hierarchies, 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

What we have tried to illustrate in the foregoing is a new approach to parametric variation, which, we believe, (i) overcomes the earlier objections of Newmeyer (2005), (ii) can be reconciled with certain types of functionalist approaches (cf. the notion of “no-choice” parameter introduced in §2), (iii) makes new diachronic, typological and acquisitional predictions and (iv) 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

²⁹ Of course, this 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 much, the result is doubly encouraging.

markedness, of grammars. For all of these reasons, we believe our approach to be worth considering and developing further.

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) (cf. Boeckx 2010 for a critique of this kind). 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

- Aboh, Enoch (2010). Information structuring begins with the Numeration. *Iberia* 2(1): 12-42.
- Aldridge, Edith (2004). *Ergativity and word order in Austronesian languages*. Ph.D. dissertation: Cornell University.
- Anderson, Stephen (1976). On the notion of subject in ergative languages, in C. N. Li (ed.). *Subject and Topic*. New York: Academic Press, 1-24.
- Assmann, Anke, Doreen Georgi, Fabian Heck, Gereon Müller and Phillip Weisser. 2013. Ergatives move too early: on an instance of opacity in syntax, in F. Heck & A. Assmann (eds). *Rule Interaction in Grammar (Linguistische Arbeitsberichte 90)*: 363-412.
- Badan, Linda and Francesca Del Gobbo (2011). On the syntax of Topic and Focus in Chinese, in *Mapping the Left Periphery. The Cartography of Syntactic Structure* (volume 5). Oxford: Oxford University Press, 63-90.
- Bakker, Peter (1997). *A Language of Our Own: The Genesis of Michif, the Mixed Cree-French Language of the Canadian Metis*. New York: Oxford University Press.
- Baker, Carl Lee (1979). Syntactic theory and the Projection Problem. *Linguistic Inquiry* 10(1): 533-81.
- Baker, Mark (1988). *Incorporation: a theory of grammatical function changing*. Chicago: University of Chicago Press.
- Baker, Mark (1996) *The Polysynthesis Parameter*. New York: Oxford University Press.

- Baker, Mark (2008). The macroparameter in a microparametric world, in T. Biberauer (ed.). *The Limits of Syntactic Variation*. Amsterdam: Benjamins, 351-74.
- Bernstein, Judy (1991). DPs in French and Walloon: Evidence for parametric variation in nominal head movement. *Probus* 3(2): 101-26.
- Berwick, Robert (1985). *The Acquisition of Syntactic Knowledge*. Cambridge, MA: MIT Press.
- Biberauer, Theresa (2008). Introduction, in T. Biberauer (ed.). *The Limits of Syntactic Variation*. Amsterdam: Benjamins, 1-72.
- Biberauer, T. (2011). In defence of lexico-centric parametric variation: two 3rd factor-constrained case studies. Paper presented at the Workshop on Formal Grammar and Syntactic Variation: Rethinking Parameters (Madrid).
- Biberauer, T. (2013). Offers that cannot be refused: mafioso parameters and the construction of minimalist parameters. Paper presented at the 39th Incontro di grammatica generativa (Modena and Reggio Emilia).
- Biberauer, Theresa, Anders Holmberg and Ian Roberts (to appear). A syntactic universal and its consequences. To appear in *Linguistic Inquiry*.
- Biberauer, Theresa, Anders Holmberg, Ian Roberts and Michelle Sheehan (2010). Reconciling formalism and: a minimalist perspective. Presented at the LAGB Annual Meeting (Leeds).
- Biberauer, Theresa and Marc Richards. (2006). True optionality: when the grammar doesn't mind, in C. Boeckx (ed.). *Minimalist Essays*. Amsterdam: John Benjamins, 35-67.
- Biberauer, Theresa and Ian Roberts (2008.) Cascading parameter changes: internally-driven change in Middle and Early Modern English, in T. Eythórsson (ed.). *Grammatical Change and Linguistic Theory. The Rosendal Papers*. Amsterdam: Benjamins, 79-114.
- Biberauer, Theresa and Ian Roberts (2012a). Towards a parameter hierarchy for auxiliaries: diachronic considerations, in J. Chancharu, X. Hu & M. Mitrović (eds). *Cambridge Occasional Papers in Linguistics* 6: 209-36.
- Biberauer, Theresa and Ian Roberts (2012b). On the significance of what doesn't happen. Paper presented at the 14th Diachronic Generative Syntax (DiGS 14) conference (Lisbon).
- Biberauer, Theresa and Ian Roberts (2013). Clausal hierarchies. Unpublished ms: University of Cambridge.
- Biberauer, Theresa, Ian Roberts and Michelle Sheehan (2013). No-choice Parameters and the Limits of Syntactic Variation. To appear in: N. Danton, D. Kostadinovska & R. Santana-

- LaBarge (eds). *Proceedings of the 31st West Coast Conference on Formal Linguistics (WCCFL 31)*. Somerville, MA: Cascadilla.
- Biberauer, Theresa and Michelle Sheehan (in press). Introduction: Theoretical Approaches to Disharmonic Word Orders, in T. Biberauer & M. Sheehan (eds), *Theoretical Approaches to Disharmonic Word Orders*. Oxford: OUP
- Boeckx, Cedric (2010). *What Principles and Parameters Got Wrong*. Unpublished ms: UAB/ICREA. (<http://ling.auf.net/lingbuzz/001118>; accessed 27 March 2012)
- Bradshaw, Joel (2009). Relative-clause bracketing in Oceanic languages around the Huon Gulf of New Guinea, in A. Adelaar & A. Pawley (eds). *Austronesian Historical Linguistics and Culture History. A Festschrift for Robert Blust*. Canberra: Pacific Linguistics, 143-162.
- Branigan, Phil (2011). In defense of macroparameters: Algonquian evidence. Paper presented at the Workshop on Formal Grammar and Syntactic Variation: Rethinking Parameters (Madrid).
- Branigan, Phil (2012). Macroparameter learnability: an Algonquian Case Study. Unpublished ms: Memorial University of Newfoundland.
- Bruening, Benjamin (2012). No such thing as “Defective Intervention”. Unpublished ms: University of Delaware. [<http://udel.edu/~bruening/Downloads/DefectiveIntervention1.pdf>; accessed 3 March 2013]
- Campbell, Lyle (1997). *American Indian Languages. The Historical Linguistics of Native America*. Oxford: Oxford University Press.
- Campbell, Lyle and William Poser (2008) *Language Classification. History and Method*. Cambridge: Cambridge University Press.
- Cheng, Lisa and Rint Sybesma (2003). Forked modality. *Linguistics in the Netherlands* 13-23.
- Chomsky, Noam (1964). *Current Issues in Linguistic Theory*. The Hague: Mouton.
- Chomsky, Noam (1973). Conditions on transformations, in S. Anderson & P. Kiparsky (eds). *A Festschrift for Morris Halle*. New York: Holt, Reinhart and Winston, 232-86.
- Chomsky, Noam (1981). *Lectures on Government and Binding. The Pisa Lectures*. Dordrecht: Foris.
- Chomsky, Noam (1986). *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger.
- Chomsky, Noam (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

- Chomsky, Noam (2000). Minimalist inquiries: the framework, in R. Martin, D. Michaels and J. Uriagereka (eds). *Step by step: essays on minimalist syntax in honor of Howard Lasnik*. Cambridge, MA: MIT Press, 89-156.
- Chomsky, Noam (2001) Derivation by phase, in M. Kenstowicz (ed.). *Ken Hale: a life in language*. Cambridge, MA: MIT Press, 1-53.
- Chomsky, Noam (2005). Three factors in language design. *Linguistic Inquiry* 36: 1-22.
- Chomsky, Noam (2007) Approaching UG from Below. In U. Sauerland & H.-M. Gärtner (eds). *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax-Semantics*. Berlin: Mouton de Gruyter.
- Chomsky, Noam (2013). Problems of Projection. *Lingua* 130: 33-49.
- Clark, Robin (1992). The selection of syntactic knowledge. *Language Acquisition* 2: 85-149.
- Clark, Robin, and Roberts, Ian (1993). A Computational Model of Language Learnability and Language Change. *Linguistic Inquiry* 24: 299-345.
- Coon, Jessica, Pedro Mateo Pedro and Omer Preminger (2011). The Role of Case in A-Bar Extraction Asymmetries: evidence from Mayan. Unpublished ms: McGill, Harvard and Syracuse Universities.
[\[http://people.linguistics.mcgill.ca/~jessica/Papers_and_handouts_files/CoonMateoPreminger-AF.pdf\]](http://people.linguistics.mcgill.ca/~jessica/Papers_and_handouts_files/CoonMateoPreminger-AF.pdf); last accessed 20 March 2013]
- Coon, Jessica and Omer Preminger (2012). Taking 'ergativity' out of split ergativity: a structural account of aspect and person splits. Unpublished ms: McGill, Harvard and Syracuse. (lingbuzz/001556; accessed 20 March 2013).
- Denison, David (1998). Syntax, in S. Romaine (ed.). *The Cambridge History of the English Language*. Cambridge: Cambridge University Press, 92-329.
- Diercks, Michael (2012). Parameterizing Case: evidence from Bantu. *Syntax* 15: 253-86.
- Dresher, Elan (1999). Charting the Learning Path: Cues to Parameter Setting. *Linguistic Inquiry* 30: 27-67.
- Dryer, Matthew (1992). The Greenbergian word order correlations. *Language* 68: 81-138.
- Dryer, Matthew (1998). Aspects of word order in the languages of Europe, in A. Siewierska (ed.) *Constituent order in the languages of Europe*. Berlin: Walter de Gruyter, 283-320.

- Dryer, Matthew and Martin Haspelmath (eds). 2011. *The World Atlas of Language Structures On-line*. Munich: Max Planck Digital Library. [<http://wals.info/>; last accessed 14 March 2013]
- Duffield, Nigel (2001). On certain head-final effects in Vietnamese, in K. Megerdooian & L.A. Bar-el (eds). *Proceedings of WCCFL 20*. Somerville, MA: Cascadilla Press, 150-63.
- Emonds, Joseph (1976). *A transformational approach to English syntax: root, structure-preserving, and local transformations*. New York: Academic Press.
- Enfield, Neil (2003). *Linguistic Epidemiology. Semantics and Grammar of Language Contact in Mainland Southeast Asia*. London: Routledge.
- Evers, Arnold and Jacqueline van Kampen (2008). Parameter setting and input reduction, in T. Biberauer (ed.). *The Limits of Syntactic Variation*. Amsterdam: Benjamins, 483-14.
- Fasanella, Adriana and Jordi Fortuny (2013). Deriving linguistic variation from learnability conditions: the Chunking Procedure. Unpublished ms: Universitat Autònoma de Barcelona.
- Fischer, Olga, Ans van Kemenade, Willem Koopman and Wim van der Wurff (2000). *The Syntax of Early English*. Cambridge: Cambridge University Press.
- Fodor, Janet Dean (2001). Setting syntactic parameters, in M. Baltin and C. Collins (eds). *The Handbook of Contemporary Syntactic Theory*. Oxford: Blackwell, 730-67.
- Fortuny, Jordi (2010). On the duality of patterning, in M. de Vries & J-W. Zwart (eds) *Structure preserved. Studies in syntax for Jan Koster*. Amsterdam: Benjamins, 131-40.
- Fowlie, Meaghan (2013). Multiple multiple spellout, in T. Biberauer & Ian Roberts (eds). *Challenges to Linearization*. Berlin: Mouton de Gruyter, 129-69.
- Gervain, Jacques and Jacques Mehler (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology* 61: 191-218.
- Gibson, Edward and Kenneth Wexler (1994). Triggers. *Linguistic Inquiry* 25: 355-407.
- Giorgi, Alessandra and Fabio Pianesi (1997). *Tense and Aspect. From Semantics to Morphosyntax*. Oxford: Oxford University Press.
- Hartman, Jeremy (2012). (Non-)Intervention in A-Movement: some cross-constructional and cross-linguistic considerations. *Linguistic Variation* 11: 121-48.
- Hendery, Rachel (2012). *Relative Clauses in Time and Space. A case study in the methods of diachronic typology*. Amsterdam: Benjamins.
- Hinds, John (1989). Left-Dislocation in Thai. *Language and Linguistics* 2: 61-83.

- Holmberg, Anders (2010). Null subject parameters, in T. Biberauer, A. Holmberg, I. Roberts and M. Sheehan *Parametric Variation*. Cambridge: Cambridge University Press, 88-124.
- Holmberg, Anders and Thorbjörg Hróarsdóttir (2003). Agreement and Movement in Icelandic Raising Constructions. *Lingua* 113: 997–1019.
- Holmberg, Anders, Aarti Nayudu and Michelle Sheehan (2009). Three Partial null-subject languages: a comparison of Brazilian Portuguese, Finnish and Marathi. *Studia Linguistica* 63: 59-97.
- Holmberg, Anders and Christer Platzack (1995). *The Role of Inflection in Scandinavian Syntax*. New York; Oxford: Oxford University Press.
- Holmberg, Anders and Ian Roberts (2012). The Syntax-Morphology relation. To appear in *Lingua*.
- Hsieh, Feng-fan and Rint Sybesma (2007). On the linearization of Chinese Sentence-Final Particles: Max Spell Out and Why CP Moves. Unpublished ms: MIT and Leiden Universiteit.
- Huang, C-T. James (1982). Move-wh in a language without wh-movement. *The Linguistic Review* 1: 369-416.
- Huang, C-T. James (1984). On the distribution and reference of empty pronouns. *Linguistic Inquiry* 15: 531-74.
- Huang, C-T. James (2013). On syntactic analyticity and parametric theory. To appear in: C-T. James Huang, Andrew Simpson & Audrey Li (eds). *Handbook of Chinese Linguistics*. Oxford: Wiley-Blackwell.
- Huang, C-T. James, Y-H. Audrey Li and Yafei Li (2009). *The Syntax of Chinese*. Cambridge: Cambridge University Press.
- Ikeda, Edward (1991). *Sentential Complementation in Mohawk*. MA dissertation: McGill. (<http://www.collectionscanada.gc.ca/obj/thesescanada/vol2/QMM/TC-QMM-61093.pdf>; accessed 27 March 2013)
- Jelinek, Eloise (1984). Empty categories, case, and configurationality. *Natural Language and Linguistic Theory* 2: 39-76.
- Julien, Marit (2002). *Syntactic Heads and Word Formation*. New York: Oxford University Press.
- Kayne, Richard (1981). On certain differences between French and English. *Linguistic Inquiry* 12: 349-71.

- Kayne, Richard (1994). *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.
- Kayne, Richard (1989). Notes on English Agreement. *CIEFL Bulletin, Hyderabad Working Papers in Linguistics* 1: 41-67.
- Kiss, Katalin E. (1995). *Discourse Configurational Languages*. New York: Oxford University Press.
- Laka, Itziar (2006). On the nature of case in Basque: structural or inherent? in H. Broekhuis, N. Corver, J. Koster, R. Huybregts & U. Kleinhenz (eds). *Organizing Grammar: Linguistic Studies in Honor of Henk van Riemsdijk*. Berlin/New York: Mouton de Gruyter, 374-82.
- Lasnik, Howard (1981). Learnability, restrictiveness, and the evaluation metric, in C. L. Baker and J. McCarthy (eds). *The Logical Problem of Language Acquisition*. Cambridge, MA: MIT Press, 1-21.
- Lasnik, Howard and Mamoru Saito (1992). *Move [α]: conditions on its applications and output*. Cambridge, MA: MIT Press.
- Ledgeway, Adam (2012). *From Latin to Romance: morphosyntactic typology and change*. Oxford: Oxford University Press
- Legate, Judy (2008). Morphological and abstract case. *Linguistic Inquiry* 39: 55-101.
- Legate, Judy (2011). Types of ergativity. *Lingua* 122: 181-91.
- Lightfoot, David (1979). *Principles of Diachronic Syntax*. Cambridge: Cambridge University Press.
- Longobardi, Giuseppe (1994). Reference and proper names. *Linguistic Inquiry* 25: 609-65.
- Manzini, Rita and Leonardo Savoia (2005). *I dialetti italiani e romanci. Morfosintassi generativa*. (3 volumes) Alessandria: Edizioni dell' Orso.
- Marcus, Gary, Steven Pinker, Michael Ullman, Michelle Hollander, John Rosen and Fei Xu (1992). *Overregularization in Language Acquisition*. Chicago: University of Chicago Press.
- McCarthy, John (1981). The role of the evaluation metric in the acquisition of phonology, in C.L. Baker and J. McCarthy (eds). *The Logical Problem of Language Acquisition*. Cambridge, MA: MIT Press, 218-48
- McGinnis, Martha (1998). *Locality in A-Movement*. Ph.D. dissertation: MIT.
- Menges, Karl. H. (1975). *Altajische Studien II. Japanisch und Altajisch*. Wiesbaden: Franz Steiner Verlag.

- Mithun, Marianne (1991). Active/agentive Case Marking and Its Motivations. *Language* 67(3): 510-46.
- Miyagawa, Shigeru (2010). *Why Agree? Why Move? Unifying Agreement-based and Discourse-Configurational Languages*. Cambridge, MA: MIT Press.
- Neeleman, Ad, Elena Titov, Hans van de Koot and Reiko Vermeulen (2009). A syntactic typology of topic, focus and contrast, in J. van Craenenbroeck (ed.). *Alternatives to Cartography*. Berlin: Mouton de Gruyter, 15-52.
- Neeleman, Ad and Kriszta Szendrői (2007). Radical pro-drop and the morphology of pronouns. *Linguistic Inquiry* 38: 671-714.
- Newmeyer, Frederick (2004). Against a parameter-setting approach to language variation. *Linguistic Variation Yearbook* 4: 181-234.
- Newmeyer, Frederick (2005). *Possible and Probable Languages. A Generative Perspective on Linguistic Typology*. Oxford: Oxford University Press.
- Nichols, Johanna (1992). *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.
- Nunes, Jairo and Juan Uriagereka (2000). Cyclicity and Extraction Domains. *Syntax* 3: 20-43.
- Ordóñez, Francisco (2000). *The Clausal Structure of Spanish*. New York: Garland.
- Peng, Anne (2011). Head-final and head-initial relative clauses in Jambi Teochew, in K. Otaki, H. Takeyasu and S. Tanigawa (eds), *Online Proceedings of GLOW in Asia Workshop for Young Scholars 2011*. [http://faculty.human.mie-u.ac.jp/~glow_mie/Workshop_Proceedings/20Peng.pdf]; last accessed 14 March 2013]
- Pesetsky, David and Esther Torrego (2006). Probes, Goals and Syntactic Categories, in Y. Otsu (ed.). *Proceedings of the 7th Annual Tokyo Conference on Psycholinguistics*. Tokyo: Hituzi Syobo Publishing Company, 25-61. [http://web.mit.edu/linguistics/www/pesetsky/Probes_Goals_Pesetsky_Torrego.pdf]; accessed 3 March 2013]
- Peters, Stanley (1972). The Projection Problem: how is a grammar to be selection?, in S. Peters (ed.). *Goals of Linguistic Theory*. Englewood Cliffs, New Jersey: Prentice-Hall, 171-88.
- Phimsawat, On-Usa (2011). *The Syntax of pro-drop in Thai*. Ph.D. dissertation: Newcastle University.

- Pine, Julian (1994). The language of primary caregivers, in C. Gallaway & B. Richards (eds). *Input and Interaction in Language Acquisition*. Cambridge: Cambridge University Press, 15-37.
- Pinker, Stephen (1984). *Language Learnability and Language Development*. Boston: Harvard University Press.
- Pinker, Stephen (1989). *Learnability and Cognition. The Acquisition of Argument Structure*. Cambridge: The MIT Press
- Pollock, Jean-Yves. (1989). Verb Movement, Universal Grammar and the Structure of IP. *Linguistic Inquiry* 20: 365-424.
- Postal, Paul (1984). Contrasting extraction types. *Journal of Linguistics* 30: 159-86.
- Richards, Marc (2004). *Object Shift and Scrambling in North and West Germanic. A Case Study in Symmetrical Syntax*. Ph.D. dissertation: University of Cambridge.
- Richards, Marc (2008). Defective Agree, Case alternations, and the prominence of Person, in M. Richards & A. Malchukov (eds), *Scales (Linguistische Arbeitsberichte 86)*, 137-161.
- Rizzi, Luigi (1994). Some remarks on linguistic theory and language development: the case of root infinitives. *Language Acquisition* 4(3): 371-393.
- Rizzi, Luigi (2001). Relativized minimality effects, in M. Baltin & C. Collins (eds) *A Handbook of Syntactic Theory*. Oxford: Blackwell, 89-110.
- Rizzi, Luigi (2013). Locality. *Lingua* 130: 169-186.
- Roberts, Ian (1993). *Verbs and Diachronic Syntax*. Dordrecht: Kluwer.
- Roberts, Ian (1998). Review of A. Harris and L. Campbell *Historical Syntax in Cross-Linguistic Perspective*. *Romance Philology* 51: 363-370.
- Roberts, Ian (2007). *Diachronic Syntax*. Oxford: Oxford University Press.
- Roberts, Ian (2010a). Varieties of French and the Null Subject Parameter, in T. Biberauer, A. Holmberg, I. Roberts and M. Sheehan, *Parametric Variation*. Cambridge: Cambridge University Press, 303-27.
- Roberts, Ian (2010b). *Agreement and Head Movement. Clitics and Defective Goals*. Cambridge, MA: MIT Press.
- Roberts, Ian (2012). Macroparameters and Minimalism: A Programme for Comparative Research, in C. Galves, S. Cyrino, R. Lopes, F. Sândalo and J. Avelar (eds). *Parameter Theory and Linguistic Change*. Oxford: Oxford University Press, 319-34.

- Roberts, Ian (to appear) Some Comments on “The Structure of Syntactic Typologies”. To appear in *Mind and Language*.
- Roberts, Ian and Anders Holmberg (2010). Introduction: parameters in minimalist theory, in T. Biberauer, A. Holmberg, I. Roberts and M. Sheehan, *Parametric Variation. Null Subjects in Minimalist Theory*. Cambridge: Cambridge University Press, 1-57.
- Roberts, Ian and Anna Roussou (2003). *Syntactic Change. A Minimalist Approach to Grammaticalization*. Cambridge: Cambridge University Press.
- Saito, Mamoru (2007). Notes on East Asian argument ellipsis. *Language Research* 43: 203-227.
- Sheehan, Michelle (2013). Towards a general alignment parameter hierarchy. To appear in: N. Danton, D. Kostadinovska & R. Santana-LaBarge (eds). *Proceedings of the 31st West Coast Conference on Formal Linguistics (WCCFL 31)*. Somerville, MA: Cascadilla.
- Sheehan, Michelle (to appear a). Explaining the Final-over-Final Constraint: formal and functional approaches. To appear in T. Biberauer and M. Sheehan (eds). *Theoretical Approaches to Disharmonic Word Orders*. Oxford: Oxford University Press.
- Sheehan, Michelle (to appear b). Some implications of a copy theory of labeling. To appear in *Syntax*.
- Sheehan, Michelle and Wolfram Hinzen (2011). Moving towards the edge: the grammar of reference. *Linguistic Analysis* 37: 405-458.
- Sigurðsson, Halldor (2011). Uniformity and diversity: a minimalist perspective. *Linguistic Variation Yearbook* 11: 189-222.
- Simpson, Andrew (2005). Classifiers and DP Structure in Southeast Asia, in R. Kayne & G. Cinque (eds). *The Oxford Handbook of Comparative Syntax*. Oxford: Oxford University Press, 806-38.
- Soh, Hooi Ling (1998). *Object Scrambling in Chinese*. Ph.D. dissertation: MIT.
- Speas, Margaret (1990). *Phrase Structure in Natural Language*. Dordrecht: Kluwer.
- Starke, Michal (2001). *Move reduces to Merge: a theory of locality*. Ph.D. dissertation: University of Geneva.
- Steever, Sanford (1998) *The Dravidian Languages*. London: Routledge.

- Takita, Kensuke (2010). *Cyclic Linearization and Constraints on Movement and Ellipsis*. Ph.D. dissertation: Nanzan University.
- Tomasello, Michael (2005). Beyond formalities: the case of language acquisition. *The Linguistic Review* 22: 183-97.
- Tomioka, Satoshi (2003). The semantics of Japanese null pronouns and its cross-linguistic implications, K. Schwabe and S. Winkler (eds), *The Interfaces. Deriving and Interpreting Omitted Structures*. Amsterdam: Benjamins, 321-40.
- Uriagereka, Juan (1999). Multiple Spell-out, in S.D. Epstein & N. Hornstein (eds). *Working Minimalism*. Cambridge, MA: MIT Press, 251-82.
- Walkden, George (2011). Abduction or Inertia? The logic of syntactic change, in C. Cummins, C-H. Elder, T. Godard, M. Macleod, E. Schmidt and G. Walkden (eds), *Proceedings of the Sixth Cambridge Postgraduate Conference in Language Research*, 230-39.
- Walkden, George (2012a). Against inertia. *Lingua* 122: 891-901.
- Walkden, George (2012b). *Syntactic Reconstruction and Proto-Germanic*. Ph.D. dissertation: University of Cambridge.
- Warner, Anthony (1997). The structure of parametric change, and V-movement in the history of English, in A. van Kemenade and N. Vincent (eds). *Parameters of Morphosyntactic Change*. Cambridge: Cambridge University Press, 380-93.
- Wichmann, Søren and Eric Holman (2009). *Assessing Temporal Stability for Linguistic Typological Features*. München: LINCOM Europa.

Aspects of the Theory of Syntax (known in linguistic circles simply as Aspects) is a book on linguistics written by American linguist Noam Chomsky, first published in 1965. In Aspects, Chomsky presented a deeper, more extensive reformulation of transformational generative grammar (TGG), a new kind of syntactic theory that he had introduced in the 1950s with the publication of his first book, Syntactic Structures. Aspects is widely considered to be the foundational document and a proper book-length