COPiL CAMBRIDGE OCCASIONAL PAPERS *in* LINGUISTICS
Volume **10**, Article 3: 38–65, 2017 | ISSN 2050-5949

# Factors 2 and 3: A principled approach*

THERESA BIBERAUER
*UNIVERSITY OF CAMBRIDGE & STELLENBOSCH UNIVERSITY*

ABSTRACT This paper concerns two components of Chomsky's (2005) Three Factors model that have not, to date, received the serious and systematic attention that they deserve: the data that a language-acquiring child picks up on during the process of language acquisition (Factor 2), and the non-language-specific, general cognitive considerations (Factor 3) that interact with Factor 2 and a minimal UG (Factor 1) to determine the form of I-languages. In relation to Factor 2, I introduce and motivate a principled approach, which builds on both classic structuralist and more recent Chomskyan ideas, and allows us to formulate a suitably precise hypothesis about which aspects of the input will qualify as 'intake' and, hence, serve as the basis for grammar construction. In relation to Factor 3, I highlight a specific cognitive bias that appears well motivated outside of language, while also having wide-ranging consequences for our understanding of how I-language grammars are constructed, and why they should have the crosslinguistically comparable form that generativists have always argued human languages have. This is Maximise Minimal Means (MMM). I demonstrate how its incorporation into our model of grammar acquisition allows us to understand diverse facts about natural language typology, acquisition, both in "stable" and "unstable" contexts, and also the ways in which it may change.

## 1 INTRODUCTION

The "traditional" generative perspective on the question of how adult speakers come to have the native-language knowledge that they do famously highlights the two ingredients given in (1):

(1) Universal Grammar (UG) + Primary Linguistic Data (PLD) → Adult Grammar (=I-language)

Here, UG is thought to be "rich in structure" (Chomsky 1981: 3), with the key consequence that the nurture component (the PLD) can be more restricted. In fact,

* This paper represents a slightly fleshed out version of a short position paper written for the 'Generative Syntax 2017: Questions, Crossroads, and Challenges' meeting organized by Ángel Gallego and Dennis Ott. I thank the audience at this meeting for their questions and comments, and the editors of the present volume for their editorial generosity. The research reported here was funded by the European Research Council Advanced Grant No. 269752 "Rethinking Comparative Syntax" (ReCoS), and a Cambridge Humanities Research Grant 'Learning from questions and commands: probing the nature and origins of native-speaker knowledge' (Biberauer 2015).

in the context of the classic Principles & Parameters era of the 1980s and 1990s, all the PLD has to provide is:

> limited evidence, just sufficient to fix the parameters of UG [which could — TB] …determine a grammar that may be very intricate and … in general lack grounding in experience in the sense of an inductive bias. (ibid.).

In view of the inescapability of Plato's Problem, the minimal grounding point has always been of particular significance: acquirers demonstrably go beyond the input in a range of, for the most part, surprisingly consistent ways; similarly, the nature and content of individual exposure also varies greatly, once again seemingly mostly not to the detriment of the uniformity of adult grammars. During the Minimalist era, the rich UG assumption has, however, been drawn into question, the objective in this context being to populate UG with only the grammar-shaping content that cannot be ascribed to more general cognitive principles. More specifically, Chomsky (2005) proposes the so-called Three Factors Model, represented in (2):

(2) UG + PLD + general cognitive factors → Adult Grammar (=I-language)

Here, the additional factor may, for example, include language acquisition biases ('principles of data analysis …used in language acquisition and other domains'; Chomsky 2005: 6), and constraints on the make-up and workings of the computational system underpinning human language ('principles of structural architecture' and 'principles of efficient computation'; ibid.).

To my mind, this Three Factors model has not received the serious and systematic attention that it deserves. In part, this follows from the vastness of the questions about its individual components — the Three Factors — on which there is currently no consensus. Consider, for example, the question of what a minimal UG should contain: here, researchers who would today describe themselves as "generative"/"Chomskyan" range from those, on the one hand, who would identify only (feature-blind) Merge, the basic combinatorial operation which produces Recursion (cf. Hauser, Chomsky & Fitch 2002 and many subsequent researchers[1]) to those, on the other, who assume richly specified cartographic structures (see i.a. Shlonsky 1999, Cinque 2013, Rizzi & Cinque 2016). An informal survey of generative colleagues of all ages also suggests that a great many remain committed to the necessary correctness of Chomsky' (2001: 10) proposal that UG 'specifies the features F that are available to fix each particular language L'. To the extent that parameters are still assumed to be a useful way of thinking about (the limits on) crosslinguistic

[1] This basic, feature-blind combinatorial operation is known by many names, including Core Merge (Fujita 2009), Set-Merge or Simplest Merge (Epstein, Kitahara & Seely 2012, 2008, Chomsky, Gallego & Ott 2017), Bare Merge (Boeckx 2015), and Concatenate (Hornstein & Nunes 2008, Hornstein & Pietroski 2009). See i.a. Mobbs (2015), and Freeman (2016) for discussion of the nature of syntactic Merge, and of the extent to which Merge as employed in syntactic derivations can be equated with the combination operation seen outside language.

variation,[2] both synchronically and diachronically, we also see significant unclarity regarding the nature and origins of minimalist parameters; some researchers assume a high number of innately specified choice-points (cf. i.a. Westergaard 2009, and the work of Richie Kayne more generally), while others assume these to be (in part) emergent in different ways (cf. i.a. Dresher 2009, 2014 in the domain of phonology; Gianollo, Guardiano & Longobardi 2008, and Guardiano & Longobardi 2017 for the proposal that specific parameters in fact reflect a limited number of innately specified parameter schema, and Rizzi 2014 for a proposal in the same spirit; see also Zeijlstra 2008, Biberauer 2011 et seq., Roberts 2012, Wiltschko 2014, Ramchand & Svenonius 2014, Biberauer & Roberts 2017 on specifically emergent parameters), and perhaps the majority leave aside explicit consideration of this "bigger picture" question. In relation to third factors, the picture is more rather than less opaque; see Mobbs (2015) for overview discussion. Finally, systematic consideration of the form that the 'triggering' input takes has barely advanced beyond the recognition that 'PLD' cannot be taken to mean "everything the child hears". Thus discussions like Evers & van Kampen (2008) highlight the difference between 'input' and 'intake', and Fodor & Sakas (2017) provide a useful overview of work to date on so-called 'triggering input.

Agreement — even in quite general terms — on what our conception of Factors 1, 2 and 3 should be thus remains to be reached. A positive perspective on this state of affairs would interpret it as following from the fact that a fleshed-out version of the Three Factors model and its components is precisely what current generative theory is — quite rightly, given what has been learned from earlier stages of the endeavor — striving for. Granting this positive interpretation, however, one would want to see explicit discussion of how progress towards this goal might be made; and it is my sense that we are not engaging in discussion of this kind — or at least, not systematically so. More specifically, we are not taking seriously enough the possibility of making new progress on the Big Question regarding the likely contents of UG — and on many other matters of generative concern, long-standing and otherwise.

What I would like to suggest here is that such progress can rather readily be made by probing the second and third factors in ways that generative and more general linguistic research to date puts 21st century generativists in an excellent position to exploit. Accordingly, this paper will seek to outline a (relatively) new model within which I believe productive investigation of all three factors might proceed (section 2). As my purpose here is to attempt a demonstration of how systematic investigation of Factors 2 and 3, and their interaction with Factor 1 might be undertaken, most of the discussion will focus on the former Factors (sections 2.2 and 2.3 respectively). Section 3 then considers some of the novel predictions the model makes, i.a. also considering its implications for our understanding of UG. Section 4 concludes.

[2] See i.a. Newmeyer (2004, 2005), Biberauer (2008, 2011, 2016, 2017b,c,d), Gallego (2011), many of the contributions in Picallo (2014), Eguren, Fernandez-Soriano & Mendikoetxea (2016), and also Biberauer & Roberts (2017).

## 2 A neo-emergentist approach to linguistic variation: the Maximise Minimal Means (MMM) model

The neo-emergentist model to be outlined here can be schematized as follows (Biberauer 2011 et seq.):

(3) UG + PLD + Maximise Minimal Means (MMM) → Adult Grammar
F1 F2 F3

The nature and assumed role of each factor will be discussed in the following sub-sections, but first a word on the "new" factor: Maximise Minimal Means. On the sense in which this model is 'neo-emergentist' see section 2.2 below.

As already noted, I am assuming MMM to be a general cognitive bias. Importantly, it is conceived as both (i) a generally applicable learning bias harnessed by the acquirer during acquisition, and (ii) a principle of structure building, facilitating the kind of efficient computation and also, crucially, the self-diversifying property that allows human language to be the powerful tool that it is. On this latter point, I follow Abler (1989), who highlights particulate (i.e. discrete combinatorial as opposed to blending) structure as the basis of self-diversification, on account of the way it facilitates the creation of so-called Humboldt systems, namely those:

(4) a. which 'make[ ] infinite use of finite means' (Humboldt 1836: 70), and, no less importantly,

b. whose 'synthesis creates something that is not present per se in any of the associated constituents' (Humboldt 1836: 67)

This, of course, calls to mind Hockett's (2016) 'duality of patterning' to which we return below; see section 2.2.

### *2.1 Factor 1: Universal Grammar*

On the present model, UG is thought to contribute the following to the I-language creation process:

(5) a. the basic operations: (i) feature-sensitive — as opposed to 'blind' or Simplest[3] — Merge, and (ii) likewise feature-sensitive Agree, plus

b. a formal feature template of some kind (e.g. [iF]/[uF]) or possibly just the notion 'formal feature, distinct from phonological and semantic feature' (i.e. [F]) to be fleshed out in ways appropriate to the substantive content of the formal features in the system.[4]

There may, additionally, be a very small set of universally specified formal features (=[F]s) not derivable from the input (see section 2.2); but not the full inventory from which acquirers make a one-time selection postulated in Chomsky (2001: 10): one of this model's objectives is precisely to try to make progress on the question of

[3] See note 1, and also i.a. Chomsky et al. 2017, Richards 2017, and Preminger 2017 for discussion of Simplest Merge. See section 3.1.1 for the suggestion that Simplest Merge might not in fact be the obvious default in the context of a system that makes maximal use of minimal means.

[4] Thanks to Jeroen van Craenenbroeck for discussion.

what kinds of [F]s are required to characterize natural-language syntax, and also to what extent those [F]s need to derive from UG. The working hypothesis is that [F]s which cannot be acquired on the basis of (i) cues of the type outlined in section 2.2 below and/or (ii) the manner in which these cues are interpreted as a consequence of the interaction of Factors 1 and 3 (see section 2.3) must constitute a 'UG residue'.

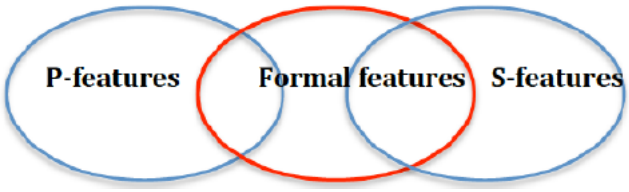
Importantly, the perspective on formal features here elaborates in a particular way on Chomsky's (1995) distinction between phonological ([P]), semantic ([S]), and formal features ([F]). In particular, we take [P]-[S]-based mappings to produce the Saussurean arbitrariness familiar from the literature (see (6a) below). Human language, however, (uniquely?) goes beyond this level of arbitrariness; it additionally involves a "higher" level of arbitrariness defined by Formal ([F]-) features. These map onto [P]- and [S]-features in systematic ways; see (6b) and (7) below, and also section 2.2 for more detailed discussion. The proposal, then, is that there are degrees of arbitrariness in human language:

(6) a. lexically stored, idiosyncratic conventionalized sound-meaning mappings involving just [P]- and [S]-features, and

b. grammatically regulated and thus more systematically conventionalised sound-meaning mappings, involving [P]-, [S]- and [F]-features.

(7) gives a rough schematization of the proposed interaction between the universally uncontroversial ('virtually conceptually necessary'; Chomsky 1993 et seq.) form ([P]) and meaning ([S]) components of language, and Chomsky's (1995) formal features ([F]). As this diagram indicates, the [F]s are assumed to piggy-back on [P] and/or [S]-features, a point to which we return in more detail below:

(7)




In the absence of a UG-given inventory of [F]s, and, further, no innately given parametric specifications, the question is, of course, where the putatively recurring systematic patterns in natural-language syntax come from. In this model, the answer is from the interaction of (i) the minimal UG outlined in this section with (ii) specific aspects of the input to be introduced in the following section and (iii) MMM, which is the focus of section 2.3.

### *2.2 Factor 2: PLD (the intake)*

As is clear from (1), PLD has been part of the generative model of language acquisition from the outset. There has, however, never been a systematic attempt to specify precisely what this entails, or why it should be credible that the child is able to draw on it. The 'limited evidence' orientation of the classic P&P era (see p.1 above) is

partly to blame here as the 'deductive richness' expectation of classic parameters was precisely concerned with alleviating the need for acquirers to notice every regularity in their target systems. This, it is important to note, remains a goal that needs to be pursued in the current context, given the seeming existence of regularities for which the input is either rare or non-existent, i.e. where acquisition would require negative evidence of a type not assumed to be available to the acquirer. Insofar as UG-PLD match-up is concerned, however, there is also a challenge that was quite widely acknowledged during the classic P&P era, namely the so-called Linking Problem (cf. i.a. Pinker 1984, Gervain & Mehler 2010, Ambridge, Pine & Lieven 2013, Fasanella 2014, and Fasanella & Fortuny 2016 for discussion). This revolves around the question of how the contents of UG, rich or otherwise, is to be linked up to the actual linguistic input that acquirers are exposed to. From the classic P&P perspective, how do acquirers 'recognize' the empirical facts that will allow them to set pre-specified parameters in the appropriate way? (see Fodor & Sakas 2017 for overview discussion, and i.a. the work of Lightfoot, Fodor, and Westergaard for some phenomenon-specific attempts to pinpoint the nature of the input strings that would "cue" parametric settings/I-language specifications of different kinds.) The same question, of course, arises in the context of an impoverished UG model of the kind under consideration here. Regardless of one's assumptions about UG, then, better understanding of the notion 'acquisitionally significant input' (= 'PLD' = 'intake') is required.

In the absence of an overarching theory of why certain data matters, while other data does not (as much), generativists have left themselves open to (not entirely unjustified) accusations about the seriousness with which they approach the empirical side of their linguistic theorizing. What I would like to do in this connection is briefly introduce and motivate what I believe to be a principled approach, which builds on both classic structuralist and more recent Chomskyan ideas and allows us to formulate a suitably precise hypothesis about which aspects of the input seem likely to qualify as credible 'intake' and, hence, to serve as the basis for grammar construction. What follows is a highly simplified version of an approach I have been developing since 2011 in the context of the research projects listed in note *.

Against the backdrop of the model introduced above, the proposal is that the child is specifically looking for what I will call systematic departures from Saussurean arbitrariness, i.e. for systematic departures from idiosyncratic one-to-one form-meaning mappings of the kind characterized in (6a) above.[5] More specifically, these (6b)-type, [F]-entailing mappings include:

(8) a. Doubling/Agreement and expletives/dummy elements, i.e. cases where we have two/multiple forms and one meaning (cf. also Zeijlstra 2008), or one form with no meaning.[6] Instead of just postulating the relevant

[5] Cf. also Fasanella (2014) and Fasanella & Fortuny (2016) on the so-called Chunking Procedure.

[6] The idea that expletives add "no meaning" to structures of which they are part and are, consequently, LF-replaceable Chomsky (cf. 1995). is widespread in Chomskyan syntax (see i.a. Vikner 1995, Svenonius 2002 for discussion). That even the most familiar English-type "pure" expletives Lasnik 1995 have interpretive consequences is, however, also clear: English *there*, for example, consistently blocks wide-

semantic ([S]) feature in cases like these, an appropriate [F] also needs to be postulated (see (8) below for an illustration relating to the postulation of formal [negation] features).

b. Systematic silence, e.g. null exponence, null arguments, null complementisers, ellipsis, etc., i.e. cases where there appears to be meaning which arises systematically despite the absence of form. In these cases, regulating the distribution of the nullness triggers the postulation of [F]s. To capture the distribution of null subjects in Finnish, for example, we need to reference particular phi-features (the 1st and 2nd or [participant] features), plus the features distinguishing finite from non-finite clauses (see Holmberg 2005).

c. Multifunctionality, i.e. cases where there appears to be what we might think of as system-defining homophony, i.e. a pattern in terms of which single forms can contribute multiple meanings, depending on their placement/distribution (cf. also Wiltschko 2014). Importantly, for the acquirer to diagnose a systematic departure from Saussurean arbitrariness, the grammar being acquired must feature multiple apparently homophonous forms whose distribution is key to their interpretation. This is a striking property of many East Asian languages, for example (see Duffield 2013, 2017, Biberauer 2017a/in press). In cases of this sort, acquirers postulate underspecified 'homophones' lacking the [F]s that determine the distribution of the element in question; these [Fs] the acquirer instead assigns to phonologically null functional heads, which serve as Merge-sites for the relevant underspecified forms.

d. Movement, i.e. assuming Chomsky's (2000) notion of 'duality of semantics' — roughly, that human language expresses both thematic and discourse/scopal meaning — we can see that movement will often result in "extra" meaning. This would, for example, be true in topicalization- and focus-fronting cases. Also relevant here, however, is what we might think of as 'higher-level duality of patterning', deriving from the contrast between "neutral/basic" and "marked" orders. Just like Hockettian duality of patterning assumes two levels of structuring — meaningless phonemes which combine to create meaningful phoneme combinations — we might think of syntax as involving "meaningless" structuring that contrasts with meaningful structuring. More specifically, consider on the one hand meaningless "basic" word-order choices like OV vs VO — which are, significantly, known to be acquired early (cf. Tsimpli 2014 for overview discussion) — and meaningless obligatory filling choices like V's spellout position or the need to fill Spec-TP or Spec-CP; on the other hand, we

scope readings (Milsark 1974, Bobaljik 2002). To the extent that they are primarily grammatical rather than content elements which contribute to interpretation by blocking otherwise available meanings, expletives may thus better be classified as instantiations of (8d)-type departures from Saussurean arbitrariness. If one considers expletives beyond English — e.g. Icelandic topic expletives, Basque, Korean and Sardinian verbal expletives — this latter classification in fact seems more appropriate.

would have meaningful optional movements like T-to-C in English, or the nature of the XP that raises to Spec-CP. Here, the meaningless conventions require fixing — just like the contents of the phoneme inventory does — whereafter they can serve as the basis for further, potentially meaningful ordering patterns, which contrast with the "basic" one.[7]

e. Recursion (cf. much work by Tom Roeper and William Snyder, i.a. Roeper 2011, Roeper & Snyder 2004, 2005.Recursion here involves repetition patterns that cannot be ascribed to [P]- or [S]-properties. It differentiates "exceptional" domains from truly productive grammar, and [F]s are required to capture the nature of that productivity (Roeper & Snyder 2005: 158; cf. also Yang 2016).[8]

A word on high-frequency recurring collocation, i.e. unduly frequent forms with a consistent, relatively minimal meaning, and a consistent position relative to contentful lexical items, is in order here. This case boils down to the distinction between content/lexical and function words, which we know acquirers to be sensitive to from the very earliest stages of acquisition.[9] Importantly, the difference between the two is signalled both prosodically — function words are shorter and more reduced than content words (lower number of syllables, less complex syllables, less diphthongization, shorter vowel duration, diminished amplitude, etc.) — and in frequency, and, crucially, distribution terms — function words are more frequent, and occupy the edges of syntactic domains (see also below). Again, [F]s are ascribed directly to these 'edge-elements' where they exhibit regular, non-homophony-type departures from Saussurean arbitrariness, e.g. where they trigger agreement, or movement, or ellipsis or nullness of some other kind, or recursion. In other words, functional elements per se are not necessarily ascribed [F]s, leaving open the existence of (largely) [F]-less auxiliaries, determiners, etc., in some languages, i.e. of less grammaticalised functional elements. This seems useful when we compare "particle"-type auxiliaries and determiners with "full" counterparts, either crosslinguistically or within a single

[7] Having both levels of duality of patterning allows the system to maximise the contribution of both the Lexical Items — i.e. the elements (containing the features) that are manipulated by the computational system — and that system's structure-building operations, (External and Internal) Merge, as MMM would lead us to expect.

[8] Significantly, recursion also guarantees Distinctness in the sense of Richards (2010), i.e. the requirement that appears to characterize all components of language structure, and in terms of which formally identical elements, which compete for the same positions, cannot surface adjacent to each other within the same domain (cf. the diverse OCP effects that have been identified in phonology and morphosyntax). Cf. also D'Alessandro & van Oostendorp (2016) on so-called Gravitational Grammar. That we would see the kinds of repulsion and attraction effects highlighted in this work — and also properties like Relativized Minimality — follows quite directly from the approach outlined here: in systems that maximize minimal means, we expect the number of features and the composite objects constructed from them to be limited in such a way that attraction, repulsion, and intervention effects would be expected to emerge. In a system with too many distinct [F]s, the observed interactions could not be modelled as falling out from simple similarity and difference "calculations".

[9] Shi, Werker & Morgan (1999), for example, show that newborns can distinguish the prosodic cues associated with content and function words, respectively, while Shi & Werker (2001) demonstrate that a content-word preference already emerges at 6 months.

language (see Biberauer 2017a/in press for discussion), and also when we think about the process via which functional elements become grammaticalised.

Taking (9a-e) together, then, the driving intuition is that [F]s are postulated if they can be seen to regulate some form of systematic contrast, which cannot be explained by appealing only to semantic or phonological considerations. Consider the case of negation. (10-12) illustrates three types of systematic departure from Saussurean arbitrariness that the approach outlined here predicts to cue the presence of a formal feature ([F]); here [negation]:

(10) *Ons is nie laat nie.* [Afrikaans]
us is not late NEG
'We are not late.'

(11) a. [*With no job*] *would she be happy.* [English]
(neutral order: *She would be happy with no job.*)

b. [*Never in my life*] *did I expect that to happen!*
(neutral order: *I never in my life expected that to happen.*)

(12) a. *a gva at̚.* [Mbili, Grassfields Bantu, Niger-Congo; Cameroon]
3SG fell tree
'He fells a tree.' (affirmative: VO)

b. *a ka at̚gva.*
3SG not tree
'He does not fell a tree.' (Ayuninjam 1998: 339, via Dryer 2009)

In (10), two negative markers are required to express a single negation, a regular pattern in Afrikaans, which acquirers are thus expected to pick up on;[10] since the doubling is specifically keyed to negation, the formal feature [negation] is postulated. Property-type (8a) thus cues the presence of [negation] here. (11), in turn, presents two structures in which a negative phrase has been fronted, triggering Verb Second, a further non-neutral word-order pattern in modern English. The contrast between the neutral SVO-structures and these V2-fronting structures requires reference to the formal feature [negation] (and possibly also [focus], given the more general nature of modern English's V2 profile, a point we leave aside here). Optional movement — one instantiation of property-type (8d) — thus cues the presence of [negation] in this case. Finally, (12) demonstrates the consistent word-order difference between affirmative and negative clauses in Mbili, a case of "basic" word order facts pointing

[10] Since this negative doubling is necessarily expressed in every negative imperative structure (see (i)), the child will receive considerable amounts of input signalling the formal (i.e. grammaticalised) nature of negation.

(i) *Moenie jou tas vergeet nie!*
must.not your case forget NEG
'Don't forget your suitcase!'

The formal features cued in imperatives seem to us good candidates for 'early' acquisition in the sense of Wexler (1998) and Tsimpli (2014); see also main text.

to the grammatical relevance of negation, i.e. the other instantiation of property-type (8d) signalling the need to postulate [negation].

Strikingly, it appears to be the case that [P]-features alone — notably prosody — serve as the initial stepping-stone into grammar: much research during the past 20 years has demonstrated acquirers' sensitivity from birth to the prosodic profile (e.g. strong-weak vs weak-strong) of their target language, and it has similarly been shown that children are able to pick up on the 'edge-marking' nature of function words during the pre-linguistic stage (cf. also note 9), a capacity which may, in turn, give them access to core properties like syntactic headedness (see Biberauer 2017b,d for further discussion). With basic, purely P-mediated regularities in place, the child can then proceed to draw on the cues provided by (8a-e)-type phenomena. Worth noting in the latter connection is the seeming significance of the cues provided by certain high-frequency, relatively simple, but strikingly syntax-rich structures, notably questions and imperatives (Biberauer 2015, 2017c, Biberauer, Bockmühl, Herrmann & Shah 2017). The current hypothesis is that [F]s cued in these structures will play a key role in structuring the earliest child grammars. As we will see in section 3.1.1 below, this also leads to the prediction that these [F]s will be the target of different kinds of 'recycling'. For present purposes, the key point is that the approach outlined here does suggest both a 'way in' for the postulation of [F]s — the P(honological)-route — and also a potential basis on which purely P-mediated [F]s can then be combined with substantive features like [negation], [tense], etc.

Evidently, the systematic morphosyntactic and morphosemantic contrasts that an acquirer encounters will vary by language; hence the language-specific 'content' of what it means to "be" categories of different types, and also what features are grammaticalised (i.e. [F]s) is, on the account proposed here, expected to vary (cf. also i.a. Haspelmath 2010, Ritter & Wiltschko 2009, 2014, Wiltschko 2014, and Chung 2012 on this). That grammars will always be characterized in terms of the distribution of formal features (cf. Baker's so-called Borer-Chomsky Conjecture) and the way in which these regulate the operations of Merge and Agree, however, crucially distinguishes the present approach from "standard" emergentist approaches, e.g. those in the Construction Grammar tradition. We therefore designate the current approach neo-emergenist.

Since both the [F]s and the categories they define will be emergent, we do need to understand how it is that the current proposal does not just predict rampant and unconstrained variation. Having considered the respective contributions of Factors 1 and 2, it is time to turn to Factor 3: Maximise Minimal Means (MMM).

### *2.3 Factor 3: MMM*

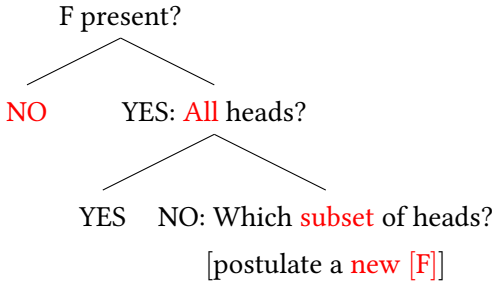
MMM is, as noted at the outset, a general cognitive bias, which I assume to play a key role in steering acquisition. In the linguistic context, I assume it to have — possibly among others — the language-specific manifestations in (13-14):

(13) Feature Economy (FE): postulate as few formal features as possible to account for the input (=intake) [generalised from Roberts & Roussou 2003]

(14) Input Generalisation (IG): maximise already-postulated features [generalised from Roberts 2007]

Together, FE and IG result in a learning pattern/path (hierarchy) with the following general "shape" (cf. also Biberauer & Roberts 2016, 2017):

(15) The NONE>ALL>SOME learning path




Here, the idea is that (15) models the interaction between the three factors in (3) as follows: an acquirer who does not pick up on a systematic departure from Saussurean arbitrariness in the input will not pose the 'F present?' question, with the result that the initial NO is a default which the comparatively oriented linguist can juxtapose with the initial YES, the answer that necessarily results when triggering data like that in (8) leads to this question being posed. The initial NO (or the NONE-system), then, respects both FE and IG. The initial YES (or the ALL-system) necessarily violates FE — as all [F]-postulation and thus, (further) grammar construction, will — but respects IG as the newly identified [F] is assumed to be present on all heads in the relevant domain (all heads in the case of headedness; all argument-licensing heads in the case of null-argument phenomena; all verbal heads in the case of finiteness marking, etc.). Should it emerge that the postulated [F] is not sufficient to delineate the domain over which the property in question is distributed, a further [F] will be postulated, thus producing a SOME-system (at later acquisition stages, this [F] may already be part of the system). If the relevant regularity is still not suitably demarcated, a further [F] is postulated, as before, producing another SOME-system. And so on until the relevant regularity has been appropriately characterized.[11]

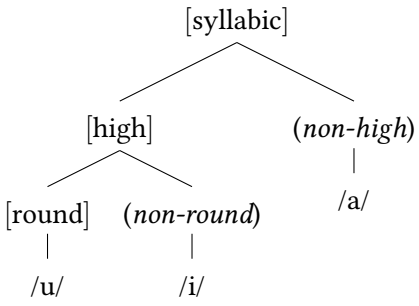
Importantly, there appears to be non-syntactic evidence in favour of the validity of postulating MMM and, more specifically, the NONE>ALL>SOME learning path it gives rise to. Dresher's (2009) Successive Division Algorithm approaches the acquisition of phonology, and thus, by extension, phonological typology in the same way (see (16) below), while the work of Dany Jaspers (cf. i.a. Jaspers 2013, Seuren & Jaspers 2014) independently postulates a NONE>ALL>SOME algorithm in the domain of logico-cognitive concept formation (see (17) below)) and also to

[11] The proposed learning path thus progresses from super- to subset, which might at first sight suggest a 'superset trap' problem. Since the supersets in play here plausibly follow from the acquirer's initial 'ignorance', however, with subsets being postulated precisely because it is clear that the existing superset grammar is deficient, the classic Subset Principle reasoning does not apply here (see also Branigan 2012 on this). A superset 'grammar' is always defeasible by the input. Additionally, see i.a. Fodor & Sakas (2005, 2017) and Biberauer & Roberts (2009) for critical discussion of the extent to which 'grammar size' can meaningfully be translated into super- and subset relations.

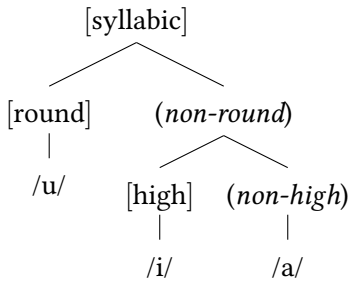
account for human colour perception (Jaspers 2012). More generally, there is evidence from (developmental) cognitive psychology showing that object classification also seems to develop on the basis of 'hierarchical inclusiveness', with superordinate/more inclusive/less specified categories being acquired before subordinate/less inclusive/more specified categories (cf. i.a. Bornstein & Arterberry 2010).

(16) NONE>ALL>SOME in phonology: the basis for the successive divisions is not dictated by UG, and may therefore target different features, producing systems with different natural classes (diagram from Dresher 2014)

a. [*high*]>[*round*]




b. [*round*]>[*high*]




[F] signifies a marked feature, and (*non-F*) and unmarked one

(17) NONE>ALL>SOME in the domain of the propositional calculus operators (following Jaspers 2013)

a. *Domain*

| | |
|---|---|
| 1 | 1 |
| 1 | 0 |
| 0 | 1 |
| 0 | 0 |

All truth-value pairs

b. *Step 1*

| | | |
|---|---|---|
| 1 | 1 | |
| 1 | 0 | *or* |
| 0 | 1 | |
| 0 | 0 | *nor* |

**Contradiction**
Something is true
vs.
Nothing is true

c. *Step 2*

| | | |
|---|---|---|
| 1 | 1 | *and* |
| 1 | 0 | |
| 0 | 1 | incl. *or* |
| 0 | 0 | |

**Implication**
Something is true
vs.
Everything is true

d. *Step 2'*

| | | |
|---|---|---|
| 1 | 1 | *and* |
| 1 | 0 | |
| 0 | 1 | excl. *or* |
| 0 | 0 | |

Something but not everything is true
vs.
Everything is true

Various child language acquisition phenomena also point in this direction — e.g. the "shadow" noun-class markers that have been said to precede fully specified noun-class markers in the acquisition of Bantu languages (Demuth 1994, 2003), the way in which free anaphors develop in French (van Kampen 2004; cf. also Lleó 1998, 2001, and Lleó & Demuth 1999 for Spanish), and the 'root infinitive' phenomenon more generally. We will discuss further domains in which NONE>ALL>SOME seems to emerge in section 3 below.

With the main components of the model in place, we are now in a position to consider some of its predictions.

## 3 Novel predictions of the model

We will consider predictions of two types here: those relating to the general formal properties that we expect to find in natural-language systems, on the one hand, and those relating to predicted patterns in what I will call 'Going beyond the input' scenarios on the other (see i.a. Biberauer 2016, 2017b for more detailed discussion of a wider range of predictions).

### *3.1 General formal properties*

#### *3.1.1 Recycling*

Given MMM, we expect what we might generally think of as 'recycling' effects to be a distinctive property of natural-language systems. This does indeed appear to be correct. Consider, for example:

(18) a. the pervasiveness of grammaticalisation phenomena in natural language, and the way in which 'pragmaticalisation' (broadly, speaker-hearer-oriented grammaticalization) also draws on existing elements and features in the system;

b. the way in which certain features serve multiple functions in the same grammar (e.g. case stacking, where case-marking marks not just thematic and/or grammatical relations, but also discourse prominence; or the numerous uses to which agreement can be put, sometimes within the same language, Archi seemingly being the extreme case here; see Bond, Corbett, Chumakina & Brown 2016);

c. the "specialised" use of C(onsonant) and V(owel), stress, and basic linearization in acquiring the lexicon and morphosyntatic regularities (see i.a. Nespor, Peña & Mehler 2003, and Gervain & Mehler (2010) for overview discussion); and

d. the various ways in which the earliest-acquired categories (V and N) are put to "extended" use in grammar structuring: V can signal notions that can be lexically expressed too, e.g. declarative vs interrogative, main vs subordinate, or realis vs irrealis; V often acts as a reference point for focus (see recent work by Kriszta Szendrői & Fatima Hamlaoui, and Vieri Samek-Lodovici), or for the A'-domain (as in V2 systems, and Hungarian — cf. Kiss 2008, who distinguishes a "nonconfigurational" post-V zone from a configurational pre-V zone; a similar, apparently "configurationality"-distinguishing pre- and post-V zone is found in Kiowa — Adger, Harbour & Watkins 2009); the existence of extended projections (Grimshaw 1991 et seq.), typically thought to be defined by lexical categorial features (e.g. V, N, P, etc.) — these impose structural constraints of different kinds, as we will see in section 3.1.2; verbalization and nominalization, where the latter also seems to serve both a general "subordinating" function, e.g. in relation to subordination and embedding (cf. Franco 2012 for discussion and references; and Huddleston 1984: 379–380 for the distinction between

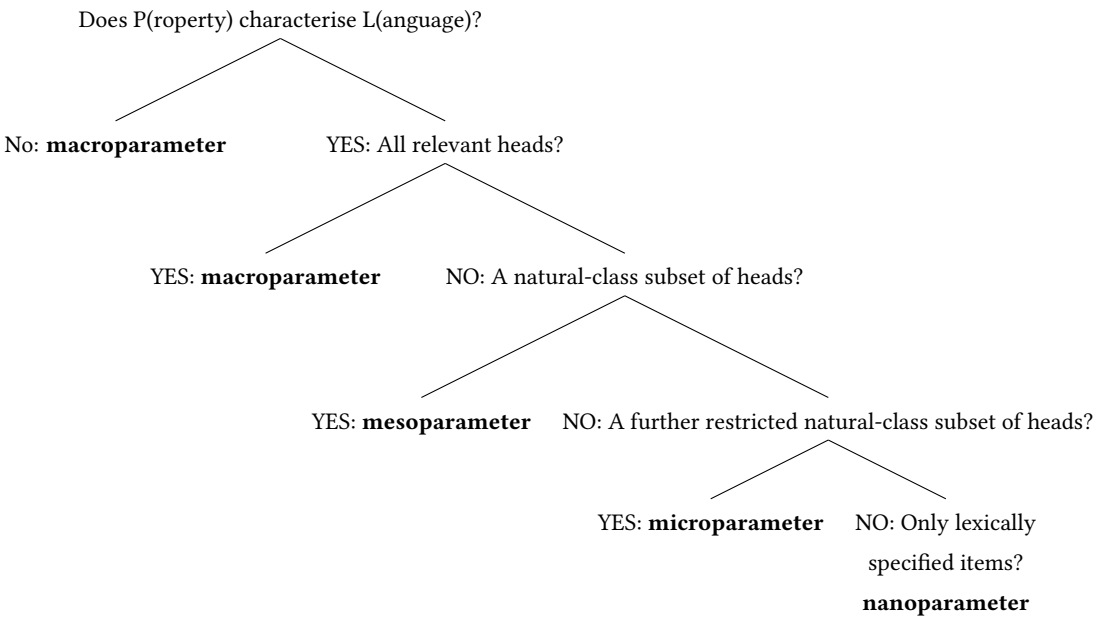
these two), but also for the reverse foregrounding purpose (as in VP topicalization/focus).

Importantly, the MMM logic also suggests a perspective in terms of which Simplest Merge, conceived of as an [F]-blind operation, may not in fact be the simplest or 'most minimal' option (see note 1). In a system which maximizes minimal means, in which [F]s already serve as the basis on which Agree operates, one might expect [F]s also to regulate Merge: if the computational system can "see" these entities for the purposes of one operation, it requires a stipulation to render them "invisible" for the purposes of the other putatively universally given computational operation. If that is correct, the problems associated with 'free generation' can be eliminated (cf. also Preminger 2017 on this).

### *3.1.2 The shape of grammatical (parametric) variation and its connection to the course of acquisition*

The NONE>ALL>SOME learning path also leads us to expect "the same" phenomenon to surface across languages in different sized versions. (19) schematises one way of thinking about this, with (20) attempting a rough characterization of what is at stake (cf. also Biberauer & Roberts 2016, 2017):[12]

(19)




(20) 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 $v_i$;

[12] Importantly, the proposed parameter types must be thought of in relative rather than absolute terms, i.e. a different approach to that assumed during the classic P&P era, where the Head Parameter, for example, constituted a macroparameter; the null-subject parameter a mesoparameter, and so on.

c. Microparameters: a small subclass of functional heads (e.g. modal auxiliaries) shows $v_i$;

d. Nanoparameters: one or more individual lexical items is/are specified for $v_i$.

That the types of head-final systems that can be identified crosslinguistically can be (partially) distinguished along the lines in (21) thus fits with the expectations of the model (see i.a. Cinque 2005, 2017, Biberauer 2008, Biberauer & Sheehan 2013, Biberauer 2017d for discussion):

(21) a. "rigid" head-finality: Japanese, Malayalam, etc.

b. clausal head-finality, nominal head-initiality, and vice versa: Chinese, Thai, Gungbe, etc.

c. "leaking" OV of different kinds, e.g. West Germanic

d. OVX, where O is the direct object (Hawkins 2009)

e. $O_{[F]}VX$, where $O_{[F]}$ is a restricted object-type (e.g. Neg, Focused, Specific, etc.)

Here it is worth highlighting the SOME-options reflected in (21), i.e. the systems for which the original head-initial/-final decision did not go all in one or other direction (see Biberauer & Roberts 2017 for simplified discussion, and Biberauer 2017b for more detailed consideration). That uniformly head-initial/-final clausal or nominal structures should occur once again reflects the expectation that early-acquired V and N will play a key structuring role in natural-language grammars (cf. (18d) above). Importantly, we can, from a typological perspective, think of V and N as fulfilling parallel roles in structuring different grammars (just as [high] and [round] did in (16) above; cf. also Wiltschko 2014 on the distinct, but formally parallel choice of one of [tense], [person] and [location] as the substantive content for INFL). More specialised SOME-systems will require the postulation of more [F]s in order to constrain the domain of head-finality. Here again, different [F]s may serve parallel structuring roles, with [aspect] potentially defining a domain of head-finality in one system, and [tense] in another, for example. As [F]-postulation is assumed to be driven by regularities in the input (section 2.2), and as there is no innately specified learning path, there is no expectation that these [F]s will be "tested" in a sequence of any kind (pace the earliest parameter hierarchies proposed within the ReCoS project; cf. Biberauer, Holmberg & Roberts 2014 for some exemplars). Instead, a linguists' (typologically oriented) amalgamated representation of the potential learning paths would indicate that these SOME-options are typologically equivalent, i.e. choices made at the same stage of the learning path. The possibility of thinking about typological equivalence in this in part acquisition-oriented way is a new one, which arises directly from the way the present model is constructed.

As also pointed out by Biberauer & Roberts (2012, 2016, 2017), the "size"-based parametric approach set out in (19-20) also leads to novel diachronic predictions. The expectation would, for example, be that "larger" (more macro) choices which require fewer [F]s exhibit greater stability over time. And this seems to be true: rigid

head-finality, for example, seems very stable, whereas West-Germanic-style OV is far less so. Furthermore, we predict that change in the direction of "smaller" (more micro) choices will exhibit a particular character, namely one which references [F]s that are already present in the system. Again, this seems to be correct. If we consider the case of OV-loss/restriction, it seems that what we observe is a process along the lines of (22) (Biberauer & Roberts 2008 show that OV-loss in the history of English appears to have followed the kind of "cascading" pathway sketched out in (22b,c):

(22) (simplified) schema of potential changes in the nature of the preverbal position in an initially "rigidly" head-final OV system:

a. all Os preverbal > all non-clausal complements (DP, PP, etc.)[13]

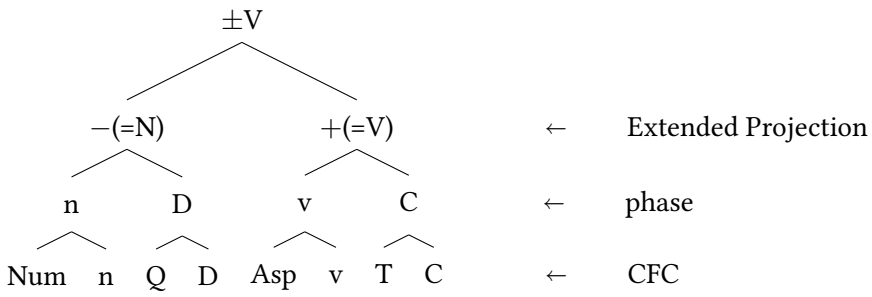
b. all non-clausal complements (DP, PP, etc.) > all DPs (nominal objects only)

c. all DPs (nominal objects only) > specific sub-type of DP (e.g. $DP_{[negative]}$, $DP_{[focus]}$, $DP_{[topic]}$) > pronominal object > clitic pronominal object, etc.

Alternatively, it could also be that the OV-constraining factor is not nominal-oriented, as in (22), but clause-oriented, with the restriction referencing [tense], [aspect], [finiteness], etc.

A key feature of the NONE>ALL>SOME learning paths is that they lead us to expect natural classes constructed on the basis of "nested" featural specifications. Thinking of syntactic category formation, for example, we might expect something like (23) rather than the kind of bottom-up approach to the acquisition of syntactic structure that was popular in the classic P&P era (cf. i.a. Radford's 1990 Small Clause Hypothesis, Rizzi's 1993/4 Truncation model, the ATOM model of Schütze & Wexler 1996; see Biberauer & Roberts 2015 for discussion of (23)):

(23)




In terms of (23), we expect acquirers to want to utilize the (in part prosodically mediated) [F] facilitating the initial V vs N distinction (here: [±V]) as the basis for further category distinctions. Taking seriously the significance of interrogative and imperative structures in the input (see again Biberauer 2015, 2017c, and also the observed fact that children are confident about "basic" interrogative properties like

[13] Intensive contact seems to be necessary to trigger a change from a rigidly head-final system to something less head-final; and it also seems necessary to introduce a head-initial nominal/D so that CPs can begin to undergo extraposition (see Biberauer & Sheehan 2012 on this).

wh-movement before they have grasped the workings of the auxiliary system or, indeed, all the specifics of the C-system (cf. i.a. Thornton 1995 for discussion and references relating to English), there seems to be good motivation for proposing that the (clause-typing-related) category C may define the second [+V] category-type acquired by children. In phase-based systems (Chomsky 2001 et seq.), this head instantiates a phase-head, whose properties further determine the properties of T (cf. again Chomsky 2001); in the present approach, T's properties are expected to build on and further elaborate — by means of newly postulated/harnessed [F]s — those already present on C. In other words, the connection between C and T is entirely expected. Similar reasoning can be applied in relation to v and one or more associated non-phase heads, and, likewise, to the corresponding heads in the nominal domain.

What is important for our purposes here is that the NONE>ALL>SOME learning path in (15) assumes an acquirer keen to generalize over as large a domain as possible to create formally defined domains sharing a particular property. This works against the kind of incremental upwards learning often assumed, suggesting instead that acquirers will successively postulate initially underspecified elements which can then be fleshed out to create sub-types of different kinds, each building upon the [F]s of the initially underspecified category, which, in turn, builds on that of earlier underspecified categories. This leads to the creation of monotonic natural classes, meaning that we expect to find considerable evidence of monotonicity in crosslinguistic variation. And this expectation does appear to be borne out. Consider, for example, the Final-over-Final Condition[14] (FOFC; see i.a. Biberauer et al. 2014, Sheehan 2013, Sheehan, Biberauer, Holmberg & Roberts 2017). FOFC is stated in (24):

(24) **The Final-over-Final Condition (FOFC)**
A head-final phrase αP cannot dominate a head-initial phrase βP where α and β are heads in the same Extended Projection.
(cf. Biberauer, Holmberg & Roberts 2008 et seq., notably BHR 2014)

What (24) requires is that head-finality start at the bottom of an Extended Projection, i.e. with V or N (see Grimshaw 1991 et seq.), and that once a head-final sequence has "stopped", it cannot restart within the same EP. Contrast the structures in (25) and (26) in this respect ($^{\wedge}$ signifies head-finality in each case):

(25) Three very basic FOFC-respecting patterns:
a. $[_{CP}\ C^{\wedge}\ [_{TP}\ T^{\wedge}\ [_{VP}\ V^{\wedge}]]]$
b. $[_{CP}\ C\ [_{TP}\ T^{\wedge}\ [_{VP}\ V^{\wedge}]]]$
c. $[_{CP}\ C\ [_{TP}\ T\ [_{VP}\ V^{\wedge}]]]$
> **monotonicity**: structurally adjacent heads consistently bear $^{\wedge}$

[14] Note that the *C* in *FOFC* stands for *Condition* as of January 2017. *Final-over-Final Condition* is still not as transparent a name for the word-order constraint as we would like, but the revised form at least does not misstate the nature of the constraint in play: final-over-final is precisely what is required, and not what is ruled out, as the constraint-oriented acronym seemingly suggested. Final-over-Initial is what is barred.

(26) Three basic FOFC-violating patterns:

a. *[$_{CP}$ C$^{\wedge}$ [$_{TP}$ T [$_{VP}$ V$^{\wedge}$]]]

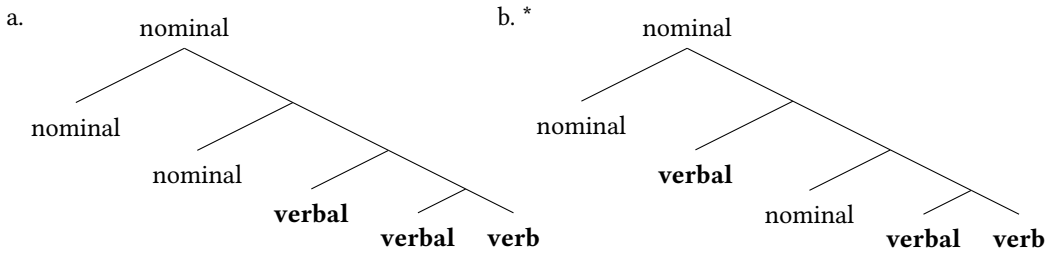
b. *[$_{CP}$ C$^{\wedge}$ [$_{TP}$ T [$_{VP}$ V]]]

c. *[$_{CP}$ C$^{\wedge}$ [$_{TP}$ T$^{\wedge}$ [$_{VP}$ V]]]

> **non-monotonicity**: structurally adjacent heads vary in their $^{\wedge}$-specification; an "on-off" pattern

As noted elsewhere (Biberauer et al. 2008, Biberauer, Newton & Sheehan 2009, Biberauer, Sheehan & Newton 2010, Biberauer et al. 2014, Sheehan et al. 2017, this requirement has diachronic implications: OV>VO changes must proceed top-down, and VO>OV changes bottom-up, which seems to be correct. Very significantly for our current purposes, however, FOFC-style monotonicity effects are not restricted to the domain of word order. Something strikingly similar emerges in relation to categorization: see Panagiotidis (2014) and references therein on so-called *Phrasal Coherence*, which is illustrated in (27)

(27) **Phrasal Coherence**: an initially verbal structure may subsequently be nominalized (see (a)); once it has been nominalized, there can be no return to verbalization. Further initially nominal structures cannot be verbalized (i.e. verbal = the equivalent of head-final in the word-order domain).[15]




And similarly, in the domain of Agreement, we see (non)-agreement "cut-off" effects exhibiting the same profile (see Biberauer 2017b for discussion). Additionally, the various hierarchies proposed by typologists and others, and the recently much-discussed *ABA syncretism constraint (cf. i.a. Caha 2009, Bobaljik & Sauerland 2017 for discussion and references) instantiate further examples of monotonicity effects in grammar — precisely what we would expect if grammars are structured on the basis of the kind of featurally regulated acquisition pathways outlined above. The same is true for the "extended FOFC effects" discussed in Biberauer (2017b, in progress).

What seems to be at stake here, then, are higher-level generalizations about recurring patterns of grammar structuring that could not readily have been ascribed to parameters — or even readily identified, to begin with — during the classic P&P era. These, we contend, are precisely the kinds of newly discovered patterns that generativists can now investigate seriously. From our perspective, they also appear

[15] Derivational forms like *anti-disestablishmentarianism* and *recategorisability* famously do not exhibit this coherence, of course. Thanks to Jeroen van Craenenbroeck for reminding me of this matter, which has been on my 'Future research' list for rather too long already, but necessarily remains there at this point.

to be the kinds of generalizations that are best understood as the product of the kind of three-way interaction between UG, the input and MMM proposed here.

### *3.1.3 Going Beyond the Input scenarios*

For Chomskyans, there is a sense in which all acquisition requires the acquirer to go beyond the input. Here, though, we will briefly consider just two scenarios that rather uncontroversially involve going beyond the input. One relates to artificial language learning, and the other to real-life learning.

Experimental work by i.a. Hudson Kam & Newport (2005) has revealed that 'children learn unpredictable variation differently than adults. They have a *stronger tendency to impose systematicity on inconsistent input* …(my emphasis; TB)' (Hudson Kam & Newport 2005: 184; see Mobbs 2015 for overview discussion). In particular, while adults demonstrate frequency-matching, approximately replicating the variability in the original input, child acquirers employ regularization strategies. The nature of these strategies is of particular interest here. Consider (28) in this connection:

(28) The types of regularization that children impose on the input:

a. **minimization**: use the variable form none of the time **(NONE)**

b. **maximization**: use the variable form all the time[6] **(ALL)**

c. **linguistically governed selection**: use the variable form in a grammatically defined subset of contexts, e.g. only with transitive Vs **(SOME)**

It is worth noting that (28c) was the most rarely used strategy; nevertheless, the picture that emerges from this (and other studies) is that child acquirers appear to appeal to MMM-driven regularization strategies of precisely the kind assumed in this model.

Our real-life example comes from English, and, specifically, the domain of number-marking in modern British English vernaculars (see Willis 2016 for more detailed discussion of this data). Let us first consider the present tense. Here standard English number-marking is restricted to first and third person on *BE* (i.e. *am/are*, *is/are*), and 3rd person singular on lexical verbs and (non-modal) auxiliaries. In vernacular varieties, the following patterns emerge:

(29) a. generalization throughout the paradigm, either
(i) to s-forms throughout (*she sings, they sings*) **(ALL)**, or
(ii) to s-less forms (*she sing, they sing*) throughout **(NONE)**.

b. use with specific sub-types of subjects, as in the Northern Subject Rule, which takes a number of different forms, picking up on the form of the subject (e.g. full DP, pronoun) and potentially the position of the subject (pre-/post-auxiliary), and so on **(SOME)**.

As indicated, then, NONE>ALL>SOME patterns once again emerge.

In the past tense, number marking is even more restricted, surfacing only on *BE* (i.e. *was/were*) in standard English. In the vernacular varieties, we once again see different patterns emerging, namely:

(30) a. generalization throughout the paradigm, either to all *was* or all *were* **(ALL/NONE)**

b. specialization relative to polarity: *were* (i.e. *weren't*) in negative clauses, regardless of person and number, with *was* occurring in affirmative clauses, regardless of number. **(SOME)**

(31) a. *They* ***was*** *writing a lot of tests that time.*

b. *He* ***weren't*** *doing much else.*

The grammatically defined SOME-choices that emerge in the past tense thus centre on [polarity]. The question is why? A highly plausible conditioning factor here would be the evidence that acquirers get from interrogative structures that auxiliaries are fundamentally concerned with polarity. Consider (32) in this regard:

(32) a. *They* ***were*** *all picnicking in the sunshine.*

*b.* ***Were*** *they all picnicking in the sunshine?*

*c. They* ***ate*** *a lot of cake.*

*d.* ***Did*** *they all eat a lot of cake?*

Here we see a very fundamental declarative-interrogative contrast in respect of auxiliary positioning (cf. (8d) above) and realization (cf. (8b) above). That English-acquiring children initially relate auxiliaries to interrogativity — i.e. open polarity — and, more generally, non-neutral affirmative polarity rather than tense-marking is strongly suggested by child data (see again Thornton 1995, and notably also Roeper 2016 for recent discussion and references).[16] [Polarity] then seems to be an early-acquired [F], which, in the context of our model, would therefore be expected to serve as the basis for input structuring in cases where the input is in some way compromised.

## 4 Conclusion

Our objective here has been to try to show why it is both productive and important for generativists to take the Three Factors model seriously, and also to flesh out how we might want to approach its empirical and general cognitive components, and their interaction with each other, and with whatever is left in UG. I introduce a neo-emergentist model of language acquisition, variation, and change that, like its classic P&P predecessor, seeks to understand language variation (and change) as a reflex of the way in which language is acquired. Where the explanatory burden previously rested largely on UG and its hypothetically rich parametric content, we have instead considered how parametrically shaped adult grammars might arise in the absence of a UG-given parametric endowment. Each of the three factors in Chomsky's (2005) model were ascribed a role in the context of the model presented here, with the general cognitive factor, Maximise Minimal Means, being argued to

[16] The strong connection to non-affirmative polarity is also evident in the history of the rise of *do*-support (see Kroch 1989, and Wallage 2017 for discussion and references).

be particularly significant in facilitating new understanding of crosslinguistically recurring patterns that would not — had they been noticed during the classic P&P era — have received a satisfactory "two-factors" explanation. At the same time, we have emphasised the importance of engaging seriously with the input, and, more specifically, those aspects of it which serve as the basis for UG-mediated, MMM-driven generalisation. The current minimalist perspective on crosslinguistic variation and language typology, then, would seem to be both more complex and more interesting than that expressed in Chomsky (1995: 6):

> Within the P&P approach the problem of typology and language variation arises in a somewhat different form than before. Language differences and typology should be reducible to choice of values of parameters.

In fact, it may be that we are, finally, starting to reach the point where we can make progress on matters like those initially highlighted in Chomsky's review of Skinner (emphasis mine, TB):[17]

As far as acquisition of language is concerned, it seems clear that reinforcement, casual observation, and natural inquisitiveness (coupled with a strong tendency to imitate) are important factors, as is **the remarkable capacity of the child to generalize, hypothesize, and "process information" in a variety of very special and apparently highly complex ways which we cannot yet describe or begin to understand, and which may be largely innate, or may develop through some sort of learning or through maturation of the nervous system.** The manner in which such factors operate and interact in language acquisition is completely unknown. (Chomsky 1959: 43).

## References

Abler, W. 1989. On the particulate principle of self-diversifying systems. *Journal of Social Biological Structure* 12. 1–13.

Adger, D, D. Harbour & L. Watkins. 2009. *Mirrors and Microparameters: Phrase Structure beyond Free Word Order*. Cambridge: Cambridge University Press.

Ambridge, B., J. Pine & E. Lieven. 2013. Child language acquisition: Why Universal Grammar doesn't help. *Language* 90(3). e53–e90.

Biberauer, T. 2008. Introduction. In T. Biberauer (ed.), *The Limits of Syntactic Variation*, 1–72. Amsterdam: Benjamins.

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. 2015. Learning from questions and commands: probing the nature and origins of native-speaker knowledge. Cambridge Humanities Research Grant Proposal.

[17] Thanks to Itziar Laka for drawing attention to this important extract during a recent generative linguistics event in Reading.

Biberauer, T. 2016. Going beyond the input (and UG): an emergentist generative perspective on syntactic variation, stability and change. Invited talk at the Language Contact, Continuity and Change in the Emergence of Modern Hebrew conference (Jerusalem).

Biberauer, T. 2017a/in press. Particles and the Final-over-Final Condition. In M. Sheehan, T. Biberauer, A. Holmberg & I. Roberts (eds.), *The Final-over-Final Condition*, 1–72. Cambridge, MA: MIT Press.

Biberauer, T. 2017b. Maximising Minimal Means: typological, acquisitional and diachronic perspectives. Three-day invited Lecture Series at the Center for Research in Syntax, Semantics and Phonology (CRiSSP) (Brussels).

Biberauer, T. 2017c. Learning from commands in contact situations: some Southern African case studies. Talk given at the Stellenbosch Linguistic Research Seminar (Stellenbosch, 5 April).

Biberauer, T. 2017d. Word-order variation and change in systems that maximize minimal means. Paper presented at the Variation and Change in the Verb Phrase Workshop (Oslo, 15 May).

Biberauer, T. in progress. Beyond word order (and syntax): FOFC revisited. Unpublished ms: University of Cambridge/Stellenbosch University.

Biberauer, T., J. Bockmühl, E. Herrmann & S. Shah. 2017. Imperative variation: the case of Afrikaans, Namibian German and Kroondal German. Paper presented at the 4th Formal Ways of Analyzing Variation (FWAV4) conference (York, 29 June).

Biberauer, T., A. Holmberg & I. Roberts. 2008. Disharmonic word-order systems and the Final-over-Final-Constraint (FOFC). In A. Bisetto & F. Barbieri (eds.), *Proceedings of XXXIII Incontro di Grammatica Generativa*, MIT Press. Available on-line at: http://amsacta.cib.unibo.it/archive/00002397/01/PROCEEDINGS_IGG33.pdf.

Biberauer, T., A. Holmberg & I. Roberts. 2014. A syntactic universal and its consequences. *Linguistic Inquiry* 45(2). 169–225.

Biberauer, T., G. Newton & M. Sheehan. 2009. Limiting synchronic and diachronic variation and change: the Final-Over-Final Constraint. *Language and Linguistics* 10(4). 699–741.

Biberauer, T. & I. 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*, 79–113. Amsterdam: Benjamins.

Biberauer, T. & I. Roberts. 2009. The Return of the Subset Principle. In P. Crisma & G. Longobardi (eds.), *Historical Syntax and Linguistic Theory*, 58–74. Oxford: Oxford University Press.

Biberauer, T. & I. Roberts. 2012. On the significance of what hasn't happened. Paper presented at the DiGS 14 Conference (Lisbon).

Biberauer, T. & I. Roberts. 2015. Rethinking formal hierarchies: a proposed unification. In X. Hu J. Chancharu & M. Mitrović (eds.), *Cambridge Occasional Papers in Linguistics*, vol. 7, 1–31.

Biberauer, T. & I. Roberts. 2016. Parameter typology from a diachronic perspective: the case of Conditional Inversion. In E. Bidese, F. Cognola & M. Moroni (eds.),

*Theoretical Approaches to Linguistic Variation*, 259–291. Amsterdam: Benjamins.
Biberauer, T. & I. Roberts. 2017. Parameter setting. In A. Ledgeway & I. Roberts (eds.), *The Cambridge Handbook of Historical Syntax*, 134–162. Cambridge: Cambridge University Press.
Biberauer, T. & M. Sheehan. 2012. Disharmony, antisymmetry, and the Final-over-Final Constraint. In V. Elguea & M. Etxebarria (eds.), *Ways of Structure Building*, 206–244. Oxford: Oxford University Press.
Biberauer, T. & M. Sheehan. 2013. Introduction. In T. Biberauer & M. Sheehan (eds.), *Theoretical Approaches to Disharmonic Word Order*, 1–46. Oxford: Oxford University Press.
Biberauer, T., M. Sheehan & G. Newton. 2010. Impossible changes and impossible borrowings: the Final-over-Final Constraint. In A. Breitbarth, C. Lucas, S. Watts & D. Willis (eds.), *Continuity and Change in Grammar*, 35–60. Amsterdam: John Benjamins.
Bobaljik, J. 2002. A-Chains at the PF Interface: copies and covert movement. *Natural Language and Linguistic Theory* 20(2). 197–267.
Bobaljik, J. & U. Sauerland. 2017. *ABA and the combinatorics of morphological rules. Unpublished ms: U. Conn & Berlin. https://ling.auf.net/lingbuzz/003320.
Boeckx, C. 2015. *Elementary Syntactic Structures: Prospects of a Feature-Free Syntax*. Cambridge: Cambridge University Press.
Bond, O., G. Corbett, M. Chumakina & D. Brown. 2016. *Archi: Complexities of Agreement in Cross-theoretical Perspective*. Oxford: Oxford University Press.
Bornstein, M. & M. Arterberry. 2010. The development of object categorization in young children: Hierarchical inclusiveness, age, perceptual attribute and group versus individual analyses. *Developmental Psychology* 46. 350–365.
Branigan, P. 2012. Macroparameter learnability: an Algonquian Case Study. Unpublished ms: Memorial University of Newfoundland.
Caha, P. 2009. *The Nanosyntax of Case*: CASTL Tromsø dissertation.
Chomsky, N. 1959. Review of Verbal Behaviour by B.F. Skinner. *Language* 35. 26–58.
Chomsky, N. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
Chomsky, N. 1993. Essays in Linguistics in Honor of Sylvain Bromberger. In K. Hale & S. J. Keyser (eds.), *The View from Building 20*, 1–52. Cambridge, MA: MIT Press.
Chomsky, N. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
Chomsky, N. 2000. Minimalist inquiries: the framework. In R. Martin, D. Michaels & J. Uriagereka (eds.), *Step by Step: Essays on Minimalist Syntax in honor of Howard Lasnik*, 89–156. Cambridge, MA: MIT Press.
Chomsky, N. 2001. Ken Hale: A Life in Language. In M. Kenstowicz (ed.), *Derivation by Phase*, 1–50. Cambridge, MA: MIT Press.
Chomsky, N. 2005. Three factors in Language Design. *Linguistic Inquiry* 36. 1–22.
Chomsky, N., Á Gallego & D. Ott. 2017. Generative Grammar and the Faculty of Language: insights, questions, and challenges. To appear in *Catalan Journal of Linguistics*.

Chung, S. 2012. Are lexical categories universal? The view from Chamorro. *Theoretical Linguistics* 38. 1–56.

Cinque, G. 2005. A note on Verb/Object order and Head/Relative clause order. *University of Venice Working Papers in Linguistics* 15. 49–104.

Cinque, G. 2013. Cognition, universal grammar, and typological generalizations. *Lingua* 130. 50–65.

Cinque, G. 2017. A microparametric approach to the head-initial/head-final parameter. To appear in *Linguistic Analysis*.

D'Alessandro, R. & M. van Oostendorp. 2016. Gravitational Grammar. Talk given in Stuttgart.

Dresher, E. 2009. *The Contrastive Hierarchy in Phonology*. Cambridge: Cambridge University Press.

Dresher, E. 2014. The arch not the stones: universal feature theory without universal features. *Nordlyd* 41(2). 165–181.

Dryer, M. 2009. Verb-object-negative order in Central Africa. In N. Cyffer, E. Ebermann & G. Ziegelmeyer (eds.), *Negation Patterns in West African Languages and Beyond*, 307–362. Amsterdam: Benjamins.

Duffield, N. 2013. Minimalism and semantic syntax: interpreting multifunctionality in Vietnamese. Unpublished ms: Konan University (LingBuzz/001919; last accessed 21 September 2015).

Duffield, N. 2017. On what projects in Vietnamese. Unpublished ms: Konan University (http://lingbuzz.auf.net/lingbuzz?_s=FfHy5mpWKuBOjEn_k=7X0_B3wAj8C8do1a; last accessed 21 May 2017).

Eguren, L., O. Fernandez-Soriano & A. Mendikoetxea. 2016. *Rethinking Parameters*. Oxford: Oxford University Press.

Epstein, S. D., H. Kitahara & T. D. Seely. 2008. Simplest Merge generates set intersection: Implications for complementizer-trace explanation. In N. Goto, K. Otaki, A. Sato & K. Takita (eds.), *Proceedings of GLOW in Asia IX*, 77–92. Mie University: MIT Press. Reprinted in: Epstein, S. D., H. Kitahara & T. D. Seely (2015). *Explorations in Maximizing Syntactic Minimization*. London: Routledge, 175–94.

Epstein, S. D., H. Kitahara & T. D. Seely. 2012. Structure building that can't be! In M. Uribe-Etxebarria & V. Valmala (eds.), *Ways of Structure Building*, 253–270. Oxford: Oxford University Press. Reprinted in: Epstein, S. D., H. Kitahara & T. D. Seely (2015). *Explorations in Maximizing Syntactic Minimization*. London: Routledge, 156–174.

Evers, A. & J. van Kampen. 2008. Parameter setting and input reduction. In T. Biberauer (ed.), *The Limits of Syntactic Variation*, 483–515. Amsterdam: Benjamins.

Fasanella, A. 2014. *On how Learning Mechanisms shape Natural Languages*: UAB dissertation. filcat.uab.cat/clt/publicacions/tesis/pdf/fasanella_14.pdf.

Fasanella, A. & J. Fortuny. 2016. Deriving linguistic variation from learnability conditions: the Chunking Procedure. In L. Eguren, O. Fernandez-Soriano & A. Mendikoetxea (eds.), *Rethinking Parameters*, 105–132. Oxford: Oxford University Press.

Fodor, J. & W. Sakas. 2005. The subset principle in syntax: costs of compliance. *Journal of Linguistics* 41. 513–569.

Fodor, J. D. & W. G. Sakas. 2017. Learnability. In I. Roberts (ed.), *The Oxford Handbook of Universal Grammar*, 249–269. Oxford: Oxford University Press.

Franco, L. 2012. Against the identity of complementizers and (demonstrative) pronouns. In K. Dziubalska-Kolaczyk (ed.), *Poznan Studies in Contemporary Linguistics*, vol. 48, 565–569.

Freeman, T. 2016. Metaphysical syntax. M.Phil. dissertation.

Fujita, K. 2009. A prospect for evolutionary adequacy: Merge and the evolution and development of human language. *Biolinguistics* 3. 128–153.

Gallego, Á. 2011. Parameters. In C. Boeckx (ed.), *Oxford Handbook of Linguistic Minimalism*, vol. 48, 523–550. Oxford: Oxford University Press.

Gervain, J. & J. Mehler. 2010. Speech perception and language acquisition in the first year of life. *Annual Review of Psychology* 61. 191–218.

Gianollo, C., C. Guardiano & G. Longobardi. 2008. Three fundamental issues in parametric linguistics. In T. Biberauer (ed.), *The Limits of Syntactic Variation*, 109–412. Amsterdam: Benjamins.

Grimshaw, J. 1991. Extended Projection. Unpublished ms: Brandeis.

Guardiano, C. & G. Longobardi. 2017. Parameter theory and parametric comparison. In I. Roberts (ed.), *The Oxford Handbook of Universal Grammar*, 377–398. Oxford: Oxford University Press.

Haspelmath, M. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687.

Hauser, M., N. Chomsky & T. Fitch. 2002. The faculty of language: what is it, who has it, and how did it evolve? *Science* 298(5598). 1569–1579.

Hawkins, J. 2009. An asymmetry between VO and OV languages: the ordering of obliques. In G. Corbett & M. Noonan (eds.), *Case and Grammatical Relations: Studies in Honor of Bernard Comrie*, 167–190. Amsterdam: Benjamins.

Hockett, C. 2016. *A Course in Modern Linguistics.* New York: Macmillan Company.

Holmberg, A. 2005. Is there a little pro? Evidence from Finnish. *Linguistic Inquiry* 36. 533–564.

Hornstein, N. & J. Nunes. 2008. Adjunction, labeling, and bare phrase structure. *Biolinguistics* 2. 57–86.

Hornstein, N. & P. Pietroski. 2009. Basic operations: Minimal syntax-semantics. *Catalan Journal of Linguistics* 8. 113–139.

Huddleston, G. 1984. *Introduction to the Grammar of English.* Cambridge: Cambridge University Press.

Hudson Kam, C. & E. Newport. 2005. Regularizing unpredictable variation: the roles of adult and child learners in language formation and change. *Language Learning and Development* 1. 151–195.

Humboldt, W. 1836. *Über die verschiedenheit des menschlichen sprachbaues.* Paderborn: Verlag Ferdinanc Schöningh.

Jaspers, D. 2012. Logic and colour. *Logica Universalis* 6. 227–248.

Jaspers, D. 2013. Constraints on Concept Formation. Poster presented at GLOW 36 (Lund).

van Kampen, J. 2004. Learnability order in the French pronominal system. In R. Bok-Bennema, B. Hollebrandse, B. Kampers-Manhe & P. Sleeman (eds.), *Romance Languages and Linguistic Theory 2002: Selected Papers from 'Going Romance', Groningen 28-30 November 2002*, 163–182. Amsterdam: Benjamins.

Kiss, K. É. 2008. Free word order, (non)configurationality, and phases. *Linguistic Inquiry* 39(3). 441–475.

Kroch, A. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1. 199–244.

Lasnik, H. 1995. Case and expletives revisited: on Greed and other human failings. *Linguistic Inquiry* 26. 615–633.

Lleó, C. 1998. Proto-articles in the acquisition of Spanish: Interface between Phonology and Morphology. In R. Fabri, A. Ortmann & T. Parodi (eds.), *Modelle der Flexion: 18. Jahresetagung der Deutschen Gesellschaft für Sprachwissenschaft*, Tübingen: Niemeyer.

Lleó, C. 2001. Early fillers: undoubtedly more than phonological stuffing. *Journal of Child Language* 28. 262–265.

Lleó, C. & K. Demuth. 1999. Prosodic constraints on the emergence of grammatical morphemes: crosslinguistic evidence from Germanic and Romance languages. In A. Greenhill, H. Littlefield & C. Tano (eds.), *Proceedings of the 23rd Annual Boston University Conference on Language Development (BUCLD 23)*, vol. 2, Somerville, MA: Cascadilla Press.

Milsark, G. 1974. *Existential sentences in English*: MIT dissertation.

Mobbs, I. 2015. *Minimalism and the Design of the Language Faculty*: University of Cambridge dissertation.

Nespor, M., M. Peña & J. Mehler. 2003. On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio* 2. 203–229.

Newmeyer, F. 2004. Against a parameter-setting approach to language variation. In P. Pica, J. Rooryck & J. van Craenenbroeck (eds.), *Language Variation Yearbook*, vol. 4, 181–234. Amsterdam: Benjamins.

Newmeyer, F. 2005. *Possible and Probable Languages.* Oxford: Oxford University Press.

Panagiotidis, P. 2014. *Categorial Features: A Generative Theory of Word Class Categories.* Cambridge: Cambridge University Press.

Picallo, C. 2014. *Linguistic Variation in the Minimalist Framework.* Oxford: Oxford University Press.

Pinker, S. 1984. *Language Learnability and Language Development.* Cambridge, MA: Harvard University Press.

Preminger, O. 2017. Back to the Future: Non-generation, filtration, and the heartbreak of interface-driven minimalism. Unpublished ms: (http://lingbuzz.auf.net/lingbuzz/003447).

Radford, A. 1990. *Syntactic Theory and the Acquisition of English Syntax: The Nature of Early Child Grammars of English.* Oxford: Blackwell.

Ramchand, G. & P. Svenonius. 2014. Deriving the functional hierarchy. *Language Sciences* 46. 152–174.

Richards, M. 2017. Problems of 'Problems of Projection': breaking a conceptual tie. Paper presented at the Generative Syntax 2016: Questions, Crossroads, and Challenges workshop (Barcelona 23 June 2017).

Richards, N. 2010. *Uttering Trees*. Cambridge: MIT Press.

Ritter, E. & M. Wiltschko. 2009. Varieties of INFL: TENSE, LOCATION and PERSON. In J. van Craenenbroeck (ed.), *Alternatives to Cartography*, 153–201. Berlin: Mouton de Gruyter.

Ritter, E. & M. Wiltschko. 2014. The composition of INFL. *Natural Language and Linguistic Theory* 32(4). 1331–1386.

Rizzi, L. 1993/4. Some notes on linguistic theory and language development: the case of Root Infinitives. *Language Acquisition* 3. 341–393.

Rizzi, L. 2014. Essays in Linguistics in Honor of Sylvain Bromberger. In C. Picallo (ed.), *Linguistic Variation in the Minimalist Framework*, 13–35. Oxford: Oxford University Press.

Rizzi, L. & G. Cinque. 2016. Functional categories and syntactic theory. *Annual Review of Linguistics* 2. 139–163.

Roberts, I. 2007. *Diachronic Syntax*. Oxford: Oxford University Press.

Roberts, I. 2012. On the nature of syntactic parameters: a programme for research. In C. Galves, S. Cyrino, R. Lopes, F. Sandalo & J. Avelar (eds.), *Parameter Theory and Linguistic Change*, 319–334. Oxford: Oxford University Press.

Roberts, I. & A. Roussou. 2003. *Syntactic Change. A Minimalist Approach to Grammaticalization*. Cambridge: Cambridge University Press.

Roeper, T. 2011. The acquisition of recursion: how formalism articulates the acquisition path. *Biolinguistics* 5. 57–86.

Roeper, T. & W. Snyder. 2004. Recursion as an analytic device in acquisition. In J. van Kampen & P. Coopmans (eds.), *Proceedings of GALA 2003*, 401–408. Utrecht: LOT.

Roeper, T. & W. Snyder. 2005. Language learnability and the forms of recursion. In A-M. Di Sciullo (ed.), *UG and External Systems: Language, Brain and Computation*, 155–169. Amsterdam: Benjamins.

Roeper, T. R. Woods. 2016. Separating Tense and Assertion: Evidence from embedded V2 and Child Language. Paper presented at the DGfS Workshop Verb Second in Grammar and Processing: its Causes and Consequences (DGfS, Konstanz).

Schütze, C. T. & K. Wexler. 1996. Subject case licensing and English root infinitives. In A. Stringfellow et al (ed.), *BUCLD 20*, 670–681. Cascadilla Press.

Seuren, P. & D. Jaspers. 2014. Logico-cognitive structure in the lexicon. *Language* 90. 607–643.

Sheehan, M. 2013. Explaining the Final-over-Final Constraint: formal and functional approaches. In T. Biberauer & M. Sheehan (eds.), *Theoretical Approaches to Disharmonic Word Order*, 407–444. Oxford: Oxford University Press.

Sheehan, M., T. Biberauer, A. Holmberg & I. Roberts. 2017. *The Final-over-Final Condition. A Syntactic Universal*. Cambridge, MA: MIT Press.

Shi, R. & J. Werker. 2001. Six-month-old infants' preference for lexical words. *Psychological Science* 12(1). 70–75.

Shi, R., J. Werker & J. Morgan. 1999. Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition* 72(2). B11–B21.
Shlonsky, U. 1999. The cartographic enterprise in syntax. *Language and Linguistics Compass* 4(6). 417–429.
Svenonius, P. 2002. Introduction. In P. Svenonius (ed.), *Subjects, Expletives and the EPP*, 3–28. Oxford: Oxford University Press.
Thornton, R. 1995. Referentiality and wh-movement in child English: Juvenile D-Linkuency. *Language Acquisition* 4. 139–175.
Tsimpli, I. 2014. Early, late or very late? Timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism* 4(3). 283–313.
Vikner, S. 1995. *Verb Movement and Expletive Subjects in the Germanic Languages*. Oxford: Oxford University Press.
Wallage, P. 2017. *Negation in Early English: Grammatical and Functional Change*. Cambridge: Cambridge University Press.
Westergaard, M. 2009. *The Acquisition of Word Order: Micro-cues, Information Structure and Economy*. Amsterdam: Benjamins.
Wexler, K. 1998. Very early parameter setting and the unique checking constraint: a new explanation of the optional infinitive stage. *Lingua* 106. 23–79.
Willis, D. 2016. Exaptation and degrammaticalization within an acquisition-based model of abductive reanalysis. In M. Norde & F. van de Velde (eds.), *Exaptation in Language Change*, 227–260. Amsterdam: Benjamins.
Wiltschko, M. 2014. *The Universal Structure of Categories. Towards a Formal Typology*. Cambridge: Cambridge University Press.
Yang, C. 2016. *Price of Productivity. How Children Learn and Break Rules of Language*. Cambridge, MA: MIT Press.
Zeijlstra, H. 2008. On the syntactic flexibility of formal features. In T. Biberauer (ed.), *The Limits of Syntactic Variation*, 143–174. Amsterdam: Benjamins.

Theresa Biberauer
University of Cambridge & Stellenbosch University
samtb23@gmail.com