Factors 2 and 3: Towards a principled approach*

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

Abstract

This paper seeks to make progress in our understanding of the non-UG components of Chomsky's (2005) Three Factors model. In relation to the input (Factor 2), I argue for the need 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 facilitates understanding of diverse facts about natural language typology, acquisition, both in "stable" and "unstable" contexts, and also the ways in which linguistic systems may change over time.

Keywords: three factors; Universal Grammar; acquisition; crosslinguistic variation; poverty of the stimulus

Resum. *Factors 2 i 3: cap a un enfocament fonamentat*

Aquest treball pretén fer progressos en la comprensió dels components que no són UG del model de tres factors de Chomsky (2005). En relació amb l'entrada (factor 2), argumento la necessitat de formular una hipòtesi adequada i precisa sobre quins aspectes de l'entrada es qualificaran com a

«ingesta» i, per tant, seran la base de la construcció gramatical. En relació amb el factor 3, destaco un biaix cognitiu específic que apareix força motivat fora del llenguatge, alhora que té àmplies conseqüències per a la nostra comprensió de com es construeixen les gramàtiques del llenguatge I, i per què haurien de tenir la forma interlingüísticament comparable als generativistes. Es tracta de maximitzar els mitjans mínims (MMM). Demostro que la seva incorporació al nostre model d'adquisició gramatical facilita la comprensió de fets diversos sobre tipologia de llenguatge natural, adquisició, tant en contextos «estables» com «inestables», i també de les maneres de canviar els sistemes lingüístics amb el pas del temps.

Paraules clau: tres factors; gramàtica universal; adquisició; variació interlingüística; pobresa de l'estímul

Table of Contents

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 (= an I-language)

Here, the nature component – UG – is thought to be "rich in structure" (Chomsky 1981: 3), with the key consequence that the nurture component – the PLD – does not need to be so elaborate. The connection between UG and the PLD in the context of the classic Principles & Parameters era of the 1980s and 1990s was in fact assumed to be much closer than is often appreciated in current discussion, with UG fulfilling a "steering" function in relation to the PLD. Chomsky (1981: 10), for example, characterises the UG specification as entailing

> concepts that can plausibly be assumed to provide a preliminary, prelinguistic analysis of a reasonable selection of presented data, that is, to provide the primary linguistic data that are mapped by the language faculty to a grammar…

In other words, the PLD, as initially conceived, was not assumed to be "everything the acquirer hears", but, instead, that part of the input that UG facilitated initial access to. On this model, all the PLD had to provide was:

> limited evidence, just sufficient to fix the parameters of UG [which – TB] … determine a grammar that may be very intricate and … in general lack grounding in experience in the sense of an inductive bias. (Chomsky 1981: 3)

The PLD, then, was expected to be readily accessible and quite simple in structure, with the "rich deductive structure" of parameters[1] accounting for the fact that our linguistic knowledge ultimately seems to vastly outstrip the input.

In view of the inescapability of Plato's Problem[2], the minimal grounding point raised above has always been of particular significance: acquirers demonstrably go beyond the finite input to which they are exposed 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 essential uniformity of adult grammars. During the Minimalist era, the rich UG assumption and, thus, its potential as a solution to Plato's Problem has, however, been drawn into question: the objective in this context is 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 (= an I-language)

Here, the additional factor – the "general cognitive factors" in (2) – 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 very little, if any, real consensus. Consider, for example, the question of what a minimal UG should contain. 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[3]) to those, on the other, who assume richly specified cartographic or even nanosyntactic structures (see i.a. Shlonsky 2010, Cinque 2013, Rizzi & Cinque 2016 on the former, and i.a. Caha 2009, Starke 2009, 2014, and Baunaz, De Clercq, Haegeman & Lander 2018 on the latter). An informal survey of generative colleagues of all

---

1. This "rich deductive structure" refers to the assumption that parameters, by hypothesis, determined not just the phenomenon associated with their triggering input, but additionally also a cluster of at first sight unrelated, and, in part, very complex properties.
2. '[T]he problem of explaining how we can know so much given that we have such limited evidence' (Chomsky 1986: xxv).
3. 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, 2013; Chomsky, Gallego & Ott this volume), *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.

ages also suggests that a great many remain committed to the necessary correctness of Chomsky's (2001: 10) proposal that UG 'specifies the features F that are available to fix each particular language L'. This would, however, entail a much richer UG than the Merge-only entity assumed in Hauser, Chomsky & Fitch (2002), and work following that line of thinking. To the extent that parameters are still assumed to be a useful way of thinking about (the limits on) crosslinguistic variation[4] both synchronically and diachronically, we also see significant unclarity regarding the nature and origins of minimalist parameters, with some researchers assuming a high number of innately specified choice-points (cf. i.a. Westergaard 2009, and the work of Richie Kayne more generally), and others assuming these to be (in part) emergent in different ways (cf. i.a. Dresher 2009, 2014 in the domain of phonology; Gianollo, Guardiano & Longobardi 2008; Guardiano & Longobardi 2017, and Longobardi 2018 for the proposal that specific parameters in fact reflect a limited number of innately specified parameter schema, and Rizzi 2014, 2015 for a proposal in the same spirit; see also i.a. Zeijlstra 2008; Biberauer 2018 *et seq.*; Roberts 2012, 2019; Wiltschko 2014; Ramchand & Svenonius 2014; and Biberauer & Roberts 2015, 2017 on different types of specifically emergent parameters), and perhaps the majority leaving 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 by now long-standing recognition that 'PLD' cannot be taken to mean "everything the child hears". Thus discussions like Evers & van Kampen (2008), Gagliardi (2012), and Lidz & Gagliardi (2015) highlight the difference between 'input' and 'intake',[5] while 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 more explicitly articulated version of the Three Factors model and its components is precisely what current generative theory is, at this point, in the process of 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 via routes that generative and more general linguistic research to date puts 21st century generativists – and researchers

---

4. 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).
5. See also Gass (1997) on this distinction in the L2 context.

more generally – in an excellent position to exploit. Accordingly, this paper will seek to outline a 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. 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" ingredient: 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), an early proponent of the idea that at first sight very different-seeming complex systems – like those underlying chemical interactions, biological inheritance, and human language – may be constructed on the basis of common principles. More particularly, Abler argued that chemistry, genetics, and human language all share a hierarchical organisation centred on "particulate" – i.e. discrete – units, which combine in such a way that the systems in question are able to self-diversify. In other words, they are able to behave in the manner of what Abler designates *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)

The component in (4a) is much-cited in generative work, with (4b) typically going unmentioned. Here, I would, however, like to suggest that the novel "more-than-the-sum-of-the-parts" (henceforth: *more-than*) products emerging from the synthesis of simpler elements are no less fundamental to our understanding of the make-up of language structure – and, in fact, also that structure's use – than

the oft-mentioned infinity-generating finite means: that one would get more than just the sum of the (finite) parts is precisely what MMM would lead us to expect, as the following discussion will show.

2.1. Factor 1: Universal Grammar

Our starting hypothesis in respect of UG is that it will contribute the following to the I-language creation process:

(5)   a.   the basic operations: (i) feature-sensitive – as opposed to 'blind' or Simplest[6] – Merge, and (ii) likewise feature-sensitive Agree,

      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.[7]

There may, additionally, be a very small set of universally specified formal features (=[F]s) not derivable from the inpuy (see section 2.2), and/or a set of universal spine-defining categorisers of the kind assumed in the work of Wiltschko (2014), Ramchand & Svenonius (2014), and Song (2019); 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 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 that can credibly be ascribed to the input (see section 2.2 below for discussion) and/or (ii) the manner in which these input cues are interpreted as a consequence of the interaction of Factors 1 and 3 (see section 2.3) must constitute part of the 'UG residue' in the sense of Chomsky (2007: 19).[8]

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 give the essence of the Saussurean arbitrariness that is 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. As we will see, these [F]s map onto [P]- and [S]-features in systematic ways (see (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.   See note 3, and also i.a. Chomsky, Gallego & Ott (this volume), 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.

7.   Thanks to Jeroen van Craenenbroeck for discussion of this point.

8.   That is, 'UG is the residue when third factor effects are abstracted. The richer the residue, the harder it will be to account for the evolution of UG, evidently.'

(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 the in part more directly accessible [P]- and/or [S]-features, a point to which we return in more detail below:

(7)
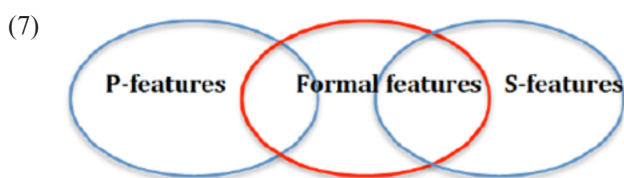
In the absence of a UG-given inventory of [F]s, and, further, no innately given parametric specifications, the question is, of course, where the seemingly recurring systematic patterns in natural-language syntax come from. In this model, the answer is from the interaction of (i) the minimal UG outlined 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. That is, natural language syntax is the more-than outcome of the interaction of Chomsky's three factors (see again (4) above).

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: without exposure to specific linguistic input, no grammar will develop (cf. i.a. Crain & Pietroski 2001; Lidz & Gagliardi 2015). There has, however, never been a **systematic** attempt to specify precisely what the PLD actually entails in concrete terms, 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 alleviation, it is important to note, remains a goal that needs to be pursued in the current context, given the clear existence in both "stable" and developing grammars of regularities for which the input is either rare or non-existent (see section 3.1.3 for discussion of a specific case).

Insofar as the relationship between UG and the PLD is concerned, there was also, during the classic P&P era, a challenge that was quite widely acknowledged, namely the so-called *Linking Problem* (cf. i.a. Pinker 1984; Gervain & Mehler 2010; Ambridge, Pine & Lieven 2013; Fasanella 2014; Fasanella & Fortuny 2016;

and Pearl & Sprouse in press for discussion). This revolves around the question of how the contents of UG, rich or otherwise, are to be linked up to the actual linguistic input that acquirers are exposed to. From the classic P&P perspective, the question is how acquirers actually '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 naturally 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 matter, while other data do 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 introduce and motivate what I believe to be a principled approach, which builds, on the one hand, on what we have learned about acquisition in the last four decades or so, and, on the other, on both classic structuralist and more recent Chomskyan ideas, thereby allowing 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 and subsequent research listed in the first note.

In the absence of a rich UG for an appropriately articulated learning theory to link to the input acquirers receive, we clearly have to let Factors 2 and 3 work harder than was previously the case. And key insights from the past 30 years' language acquisition research suggest that this may indeed be feasible. Consider, for example, the research demonstrating *in utero* and very early post-birth sensitivity to aspects of prosody (see Gervain & Werker 2008 for an overview). In brief, it is known that the fetal auditory system is functional from around 6 months' gestation (Mehler & Dupoux 1994; Moore 2002). While fine details of speech are filtered out, less fine-grained prosodic properties, like intonational contours (e.g. a language's characteristic "tune", which is closely tied to its basic headedness properties; see below) and rhythmic properties are detectable *in utero*. This fact appears to underlie newborn infants' repeatedly demonstrated ability to distinguish the maternal language from a prosodically distinct – and oppositely headed – language-type, e.g. English vs Japanese (cf. i.a. Mehler et al. 1988; Nazzi et al. 1998; Gervain et al. 2007), and also, subsequently, their strikingly early ability to establish the "basic" (i.e. lexical/bottom-of-extended-projection) head-directionality of the system they are acquiring: simplifying greatly, OV has a basic 'strong-weak' prosodic contour, while VO has a basic 'weak-strong' contour (cf. i.a. Wexler 1998 and Tsimpli 2014 on basic word order as a very acquired property, a Very Early Parameter or VEP).

Further, various 'edge'-oriented cues allow acquirers to begin to "chunk" the input-strings in accordance with the grammar of their input-language(s) long before

they have any lexical knowledge. Function items consistently differ from content items in respect of their phonological properties (they are, in general, shorter, with individual syllables being less complex with less diphthongisation, shorter vowel duration, and diminished amplitude), their frequency (individual function words are much more frequent than individual content items), and, particularly crucially in the current context, their distribution (functional items tend to occupy the edges of syntactic domains). These properties appear to alert pre-lexical infants to the distinction between content and functional items, leaving 6-month-olds with a preference for the former (see Shi, Werker & Morgan 1999; Shi & Werker 2001; and the overview in Gervain & Werker 2008). Thereafter, more fine-grained details become available, with, for example, the distribution of consonants and vowels within already-identified linguistic chunks contributing specifically to the articulation of acquirers' knowledge of, respectively, vocabulary and associated inflectional morphology (Nespor, Peña & Mehler 2003).

Importantly, then, the picture that emerges is of acquirers making the most of the cues that are accessible to them at every stage of the acquisition process, as one would expect on an MMM view (see section 2.3 below). More specifically, we see that acquirers seem initially to focus just on the linguistic systematicities that do not require any mapping between form and meaning: salient and typically recurring (and thus high-frequency) phrase-level prosodic regularities. Prosody, in other words, seems to be the minimal means which serves as the stepping-stone into grammar. Once accustomed to the initially registered patterns, acquirers appear to be become "bored" by them, and we see a shift in interest to more fine-grained, high-frequency aspects of prosodic encoding – such as those underlying the difference between content and functional items – which the now "boring" initial prosodic regularity has rendered accessible to the acquirer. And so the process continues, with the acquirer's attention to linguistic properties becoming successively more finely tuned as their linguistic knowledge at each stage of the acquisition process facilitates ever more detailed access to the regularities in the input.

On the MMM view, then, the acquirer's attention to the input is at least partly "steered" by what the grammatical specification of their grammar makes accessible to them. Initially having access to only a limited component of what is in the input – i.e. to a highly restricted intake – appears to allow acquirers to make efficient headway in fleshing out the complex formal system to which they are exposed on a "Less is More" basis (see Newport 1990; Elman 1993; much recent work by Charles Yang, and the discussion to follow in section 2.3).

Crucially, acquirer's hypothesised initial "sound-side" focus provides them with various kinds of *distributional* knowledge, which can then be fleshed out on the basis of input requiring sensitivity to both sound and meaning.[9] In this connection, the distinction between the fully arbitrary form-meaning mappings that define classic Saussurean arbitrariness ((6a) above), and the still arbitrary, but more

---

9. In the case of sign languages, this initial sensitivity would, of course, be expected to centre on relevant aspects of sign-language prosody, which has been said to involve body posture and various manual cues (timing, size; see Sandler 2010, 2012 for an overview).

systematic form-meaning mappings that constitute grammar ((6b) above) is argued to be particularly important. More specifically, Biberauer (2017e) highlights the key relevance of so-called *systematic departures from Saussurean arbitrariness* – i.e. consistent departures from the arbitrary one-to-one form-meaning ([P]-[S]-based) mappings that underlie the core content lexicon – in alerting the acquirer to a domain in which the postulation of (grammatical) formal features ([F]s) would facilitate more economical – in our terms, MMM-driven – learning and knowledge representation (see also Schuler et al. 2016, and Pearl in press; and Fasanella 2014 and Fasanella & Fortuny 2016 on the so-called *Chunking Procedure*). The proposed [F]-signalling mappings include:

(8) a. Doubling/Agreement and expletives/dummy elements, i.e. cases where there is, in a relevant sense "too much form". In the doubling/agreement case, for example, we have two/multiple forms, the prosodically weaker one of which "echoes" (part of) the meaning of the other (cf. also Zeijlstra 2008).[10] In the expletive/dummy case, we have a form with no (non-relational) meaning.[11] Instead of just postulating the relevant semantic ([S])

---

10. The fact that agreement/doubling "echoes" part of the meaning of its controller does not rule out the possibility that it may, in the context of particular structures, serve to signal meaning that might not otherwise be (so) evident. In German (i-ii) suggested by an anonymous reviewer, for example, the verbal agreement serves (potentially alongside intonation in speech) to distinguish two quite different meanings:

(i) Peter hat   Frauen einen Brot gebacken.
    Peter have.SG women a.ACC bread buy.PART
    'Peter baked a loaf of bread for women.'

(ii) Peter haben   Frauen einen Brot gebacken.
     Peter have.PL women a. ACC bread bake. PART
     'For Peter women baked a loaf of bread.'

   The claim about agreement/doubling as an [F]-cue is simply that its systematically dependent, "echoing" nature will be salient to child acquirers in a context where they are trying to establish generalisations for systematically recurring patterns. As units, agreement/doubling markers carry "derivative" meanings, but this "derivative" meaning may serve disambiguating, emphasizing, or other interpretively significant functions in certain structural configurations. That individual elements will be able to serve both neutral/unmarked and non-neutral/marked functions, depending on their structural environment, is, in fact, precisely the kind of more-than effect that the MMM model would predict.

11. The idea that expletives add "no meaning" to structures of which they are part and are, consequently, LF-replaceable (cf. Chomsky 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-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, movement-derived 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, all of which play a role in information-structurally marked structures – this latter classification in fact seems more appropriate. It is also worth noting that expletive elements generally seem to contribute to meaning principally as a consequence of the relations they enter into with the contentful components of the structures they feature in, i.e. their interpretive contributions depend less on the

feature in cases like these, an appropriate [F] also needs to be postulated (see (9) below for an illustration relating to the postulation of a formal [negation] feature).

b. Systematic silence, e.g. null exponence, null arguments, null complementisers, ellipsis, etc. These are cases where there appears to be meaning which arises systematically despite the absence of form. If acquirers, as a result of their encounters with the content lexicon in particular, operate on the default assumption that meaning is paired with overtly realised form, we might expect them to "notice" circumstances where they systematically interpret meanings that don't correlate with overt form. The evidence from child acquisition suggests that certain types of nullness – notably, null arguments – are correctly produced and understood very early, by the age of 2 (Tsimpli 2014). Other types, like VP ellipsis are likewise produced and understood surprisingly early, by the age of 3-4 (see i.a. Foley et al. 2003, and Santos 2009), although this may not be full acquisition of all aspects of the relevant phenomena (cf. i.a. Göksun et al. 2011 for discussion). The fact that null elements alternate with overt counterparts undoubtedly plays a key role in the identification of nullness-related [F]s, with cases where the overt form is necessarily emphatic – null subjects are a case in point – being acquired particularly readily.

c. Multifunctionality, or 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; isolated homophonies (as in that centring on English *bank*) are not predicted to trigger [F]-postulation. Systematic homophony is a striking property of many East Asian languages, for example (see Duffield 2013, 2017, Biberauer 2017a). Thus the Vietnamese modal system discussed in Duffield (*op. cit.*) comprises three distinct lexical items – or, more accurately, units of language in Wiltschko's (2014) sense – whose immediately preverbal, postverbal or clause-final placement determines their modal force (deontic, abilitative, epistemic, respectively). In cases of this sort, acquirers postulate an underspecified 'homophone' (or unit of language) lacking the [F](s) that determine the distribution of the element in question; these [F](s), instead, the acquirer assigns to phonologically null functional heads, which serve as Merge-sites for the relevant underspecified forms. Distributional cues, then, are key to capitalizing on this [F]-cue.

---

independent content they contribute to the wholes of which they become a part, and more on the interpretive contrasts they facilitate with otherwise required, but, in the structures in which they occur, unrealised derivational operations – obligatory substantive subject-, topic-, or verb-raising in the cases mentioned here.

d. Movement, i.e. assuming Chomsky's (2000) notion of 'duality of seman-tics' – roughly, that human language expresses both thematic and dis-course/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.[12] 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 (Hockett 1958) assumes two levels of structuring – meaning-less phonemes which combine to create meaningful phoneme combina-tions[13] – we might think of syntax as involving "meaningless" structuring that contrasts with meaningful structuring (see also Fortuny 2010). 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 obliga-tory filling choices like V's spellout position or the need to fill Spec-TP or Spec-CP; on the other hand, we would have meaningful optional move-ments 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 content of the phoneme inventory does – whereafter they can serve as the basis or reference point for further, potentially meaningful ordering patterns, which contrast with the "basic" one.[14]

e. A particular kind of recursion, namely that which produces the structured repetition patterns that underlie productive compounding(-like) patterns (e.g. noun-noun or verb-verb compounding, verb-serializing, and verb clustering; cf. much work by Tom Roeper, William Snyder and Ana Pérez-Leroux, i.a. Roeper 2011; Roeper & Snyder 2004, 2005; Pérez-Leroux, Castilla-Earls, Bejar & Massam 2012; Pérez-Leroux, Peterson, Bejar, Castilla-Earls, Massam & Roberge 2018). Acquirers can be expected to "notice" this kind of recursion – thus rendering it a credible [F]-trigger – on account of their keenness to postulate memory-saving generalizations, i.e. formal rules (see Roeper & Snyder 2005: 158; and also

12. Importantly, though, the fact that a particular movement operation is interpreted as topicalisation or focalisation does not automatically result in the postulation of a [topic]- or [focus]-feature. The third-factor bias to maximise minimal means, to be discussed in the next section, will, in the first instance, drive the acquirer to seek to recycle an already-postulated [F]. In this connection, the growing number of analyses of topicalisation and focalisation phenomena that diagnose [F]s like [person] (see i.a. Richards 2008; Leffel, Simik & Wierzba 2013) and [case] (see i.a. Pesetsky 2014; Levin 2016, and (18b) in the main text) as the syntactically active [F]s being manipulated by Merge and Agree is precisely what the MMM approach would predict (see also section 3.1.1.).

13. Duality of patterning rather clearly seems to instantiate the second aspect of language's Humboldtian character (cf. (4b) above); and the same is true for the "higher-level" variety pro-posed here.

14. 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 compu-tational system – and that system's structure-building operations, (External and Internal) Merge, as MMM would lead us to expect.

Yang 2016; Schuler et al. 2016). More fundamentally, the requirement that van Riemsdijk (2008) and Leivada (2017) label *Identity Avoidance* and Richards (2010) *Distinctness* drives acquirers in the direction of [F]-postulation wherever apparently identical elements surface adjacent to each other within the same domain. This drive plausibly reflects a very basic heuristic that children are more famously known to employ in word learning, namely the Principle of Contrast (Clark 1993). From the current perspective, the recursion at stake here is just MMM driving the acquirer to make use of the Principle of Contrast not just in the core lexical domain (i.e. in relation to (6a) above), but also in grammar structuring (i.e. in relation to (6b) above too).[15,16]

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 also 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 (see again the discussion of Shi, Werker & Morgan 1999; Shi & Werker 2001 above). As noted above, function words are edge-elements, located at the left or right boundary of their XP. [F]s are assigned directly to these elements in cases 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, or where their presence is obligatory wherever a substantive element of some kind is present (French determiners would be a case in point). As discussed above, [F]s are **not** assigned directly to (8c)-type homophonous elements: doing so would create an unwieldly, homophone-rich lexicon which fails to register many systematic generalisations. Importantly, then, functional elements *per se* are not necessarily ascribed [F]s, leaving open the possibility of (largely) [F]-less auxiliaries, determiners, etc., in some languages, i.e. of less grammaticalised functional elements.

15. In fact, in emergentist approaches to phonology such as that of the so-called *Toronto School* (see i.a. Hall 2007; Dresher 2009, 2014), the Principle of Contrast is also assumed to be operative in the structuring of phonological systems: in accordance with Hall's *Contrastivist Hypothesis,* phonological features are only postulated if they account for a phonological contrast in the system being acquired. To the extent that all Identity Avoidance phenomena can be ascribed to the workings of the basic Principle of Contrast heuristic, the diverse Obligatory Contour Principle (OCP)/haplology phenomena that have been identified in phonology and morphosyntax can all be understood as a reflex of this same heuristic. Formally identical elements may not Merge with each other and thus surface adjacent to each other in the same domain any more than identical phonological units may do so.

16. Cf. also D'Alessandro & van Oostendorp (2018) on so-called *Magnetic 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 complete or partial similarity- and difference-based relations like attraction, repulsion, and intervention effects would be expected to become calculable and, thus, to play a role in regulating language structure. 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".

This seems useful when we compare "particle"-type auxiliaries and determiners with "full" counterparts, either crosslinguistically or within a single language (see Biberauer 2017a for extensive discussion), and also when we think about the process via which functional elements become grammaticalised (an [F] not previously associated with a content item needs to be ascribed to it).

Taking (8a-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. (9-11) 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]:

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

(10) 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.)

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

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

In (9), two negative markers are required to express a single negation, a regular pattern in Afrikaans, which acquirers are thus expected to pick up on;[17] since the doubling is specifically keyed to negation, the formal feature [negation] is postulated. Property-type (8a) thus cues the presence of [negation] here. (10), in turn, presents two structures in which a negative phrase has been fronted, triggering Verb Second, a non-neutral word-order pattern in modern English. The contrast between the neutral SVO-structures and these V2-fronting structures requires reference to

---

17. 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) Moe**nie** jou   tas   vergeet **nie**!
        must.not your case forget NEG
        'Don't forget your suitcase!'

    More generally, 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.

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. Interpretively significant optional movement – one instantiation of property-type (8d) – thus cues the presence of [negation] in this case. Finally, (11) demonstrates the consistent word-order difference between affirmative and negative clauses in Mbili, a case of "basic" word order facts pointing to the grammatical relevance of negation, i.e. the other instantiation of property-type (8d)) signalling the need to postulate [negation].

   As already noted, it appears to be the case that [P]-features alone – notably prosodic properties – serve as the initial stepping-stone into grammar. 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 an initial 'way in' for the postulation of [F]s – the P(honological)-route – and also a potential basis on which the acquirers may initially move beyond purely [P]-mediated [F]s to those cued by systematic departures of the kind in (8).

   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-emergentist*.

   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 (12-13):
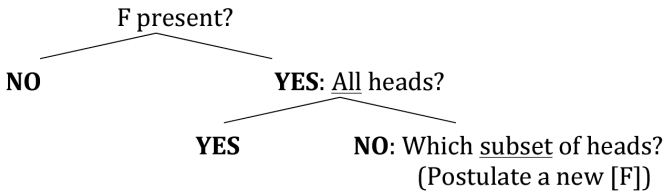
(12) **Feature Economy** (FE): postulate as few formal features as possible to account for the input (=intake) [generalised from Roberts & Roussou 2003]

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

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

(14) The **NONE>ALL>SOME** learning path

                    F present?

     NO   **YES**: <u>All</u> heads?

              YES   **NO**: Which <u>subset</u> of heads?
                                     (Postulate a new [F])

Here, the idea is that (14) models the interaction between the three factors
in (3) as follows: the initial NO represents an acquirer who does not pick up
on a systematic departure from Saussurean arbitrariness in the input; they will
therefore not pose the 'F present?' question. The initial NO thus needs to be
interpreted as a default which the comparatively oriented linguist can juxtapose
with the initial YES, the answer that necessarily results when some form of trig-
gering data (see again (8) above) leads to this question being posed. The initial
NO (or the NONE-system), then, respects both FE and IG; it literally requires
the acquirer to do nothing. The initial YES (or the ALL-system), by contrast,
necessarily violates FE – as all [F]-postulation and thus, (further) grammar con-
struction, will – but it 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; see sec-
tion 3.1.3 for some discussion illustrating this case). 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 appro-
priately characterized.[18]

18. 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 (see
    also Branigan 2012 on this). The superset 'grammars' postulated on the basis of (14) are always
    defeasible by the input. Independently of this, see i.a. Fodor & Sakas (2005, 2017) and Biberauer
    & Roberts (2009) for critical discussion of the extent to which 'grammar size' can in fact be
    meaningfully translated into super- and subset relations: implementation of something like a Subset
    Principle in the acquisition context poses numerous non-trivial problems.
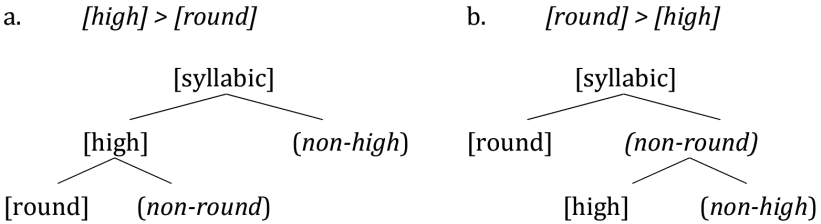
Very importantly, the assumption that [F]-postulation by acquirers is regulated by MMM means that [F]s already in the system will always, where possible, serve as the point of departure for further refinements of the existing grammar (see section 3.1.1 below on [F]-'recycling'). MMM will also tend to produce "nested" natural classes, with different (linguistic) phenomena being sensitive to more or less specific [F]-combinations (see the immediately following discussion, and also section 3.1.2 below). From an acquisition perspective, this also has the consequence that the ALL>SOME component of the NONE>ALL>SOME-defined learning path must be understood in relative terms. More specifically, as soon as an [F] is postulated to constrain the distribution of a grammatical regularity, it effectively becomes, for the acquirer, an ALL-option in relation to the class of heads under consideration at that point in the acquisition process; what "counts" as ALL vs SOME thus needs to be interpreted dynamically from the perspective of the language-acquiring child. From a typological perspective – i.e. the type of perspective a comparatively oriented linguist might hold – the existence of languages employing both more and less featurally constrained versions of "the same" phenomenon (head directionality, null subjects, verb-raising, etc.) – the NONE>ALL>SOME perspective remains useful in more fixed form (though see section 3.1.2. below for further discussion). And the same is true for the acquisitionist, who may find it useful to think of earlier and later stages of an acquirer's grammar in NONE>ALL>SOME terms.

If MMM and, more specifically, the NONE>ALL>SOME learning path it gives rise to are to be credibly conceived of as third-factor-related, there would need to be non-syntactic evidence favouring their postulation. Significantly, there does appear to be evidence of precisely this kind. Dresher (2009), for example, postulates the Successive Division Algorithm (SDA), which approaches the acquisition of phonology, and thus, by extension, phonological typology in NONE>ALL>SOME terms. The SDA is given in (15):

(15) a. Begin with *no* feature specifications: assume all sounds are allophones of a single undifferentiated phoneme.

    b. If the set is found to consist of more than one contrasting member, select a feature and divide the set into as many subsets as the feature allows for.

    c. Repeat step (b) in each subset: keep dividing up the inventory into sets, applying successive features in turn, until every set has only one member.
<div align="right">(Dresher 2009: 16)</div>

Importantly, the basis for the successive divisions is not dictated by UG; these divisions may therefore target different features in different systems, producing phonological systems with natural classes that are not structured in the same way. Consider (16) by way of example:

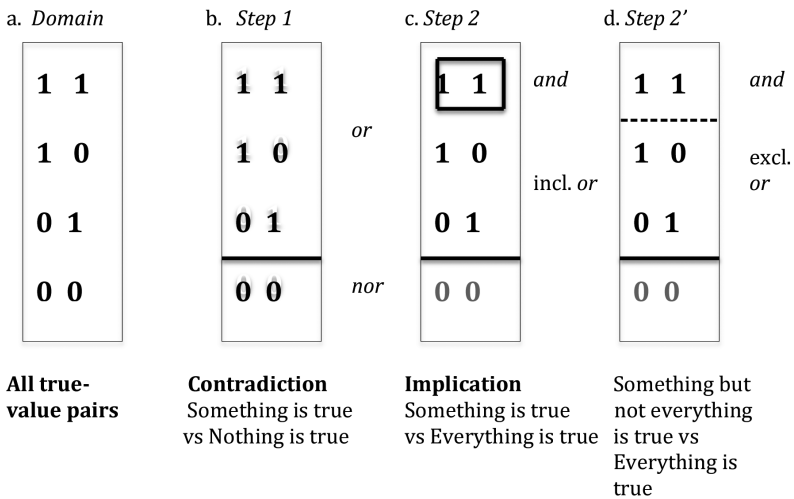(16) NONE>ALL>SOME in phonology (diagram from Dresher 2014: 167)

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

                 [syllabic]   [syllabic]

   [high]   *(non-high)*   [round]   *(non-round)*

[round]   *(non-round)*   [high]   *(non-high)*

(where marked values are indicated as [F] and unmarked values as (*non-F*). For expository purposes, we abstract away from the details of Dresher's markedness assumptions.)

Here we see that the three vowels /a/, /i/, and /u/ fall into different natural classes, depending on the way in which the vowel space that they occupy is divided. In both cases, the feature [syllabic] must initially be postulated to distinguish vowels from consonants: this is the basis for the ALL-division, which is universal, given that the sound spectrum does not have alternative "natural joints" (see Martí 2015). A range of SOME-division options follow, however. In the case of (16a), a first further distinction is drawn on the basis of the distinctive feature of vowel height ([high] vs *(non-high)* for Dresher), resulting in a natural-class distinction between high and non-high vowels. Phonological processes in this system (e.g. vowel harmony) will thus reference this high/non-high distinction, with /i/ and /u/ systematically exhibiting behaviour not shown by /a/. The rounding feature then serves to individuate the [high] vowels. In (16b), by contrast, the vowel space is initially sub-divided on the basis of roundness, with the height division being secondary, i.e. the basis for ultimate full individuation. In this case, phonological processes will therefore target /i/ together with /a/, excluding /u/. In each case, the vowel's systematically contrastive behaviour will alert acquirers to the nature of the successive divisions that are required – or, in our terms, to the form that the full NONE>ALL>SOME learning pathway should take. Strikingly, existing phoneme acquisition studies focusing on English and Dutch would appear to support the kind of learning pathways predicted by this approach (see i.a. Fikkert 1994; Stokes, Klee, Carson & Carson 2005; and also Dresher 2014 and Mobbs 2015 for discussion).

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. Consider (17) below in this connection:

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

| a. *Domain* | b. *Step 1* | c. *Step 2* | d. *Step 2'* |

a. *Domain*	b. *Step 1*	c. *Step 2*	d. *Step 2'*
All true-value pairs	**Contradiction** Something is true vs Nothing is true	**Implication** Something is true vs Everything is true	Something but not everything is true vs Everything is true

Here we see that successive divisions of the logical truth space necessarily begin with a separation of truth from falsity, i.e. Step 1 in (17b). As in the case of the vowel-space, further sub-division is then open to alternative possibilities: either we distinguish the case where something, possibly everything, is true from that where everything is true – Step 2 in (17c) – or we distinguish the case where everything is true from that where something, but not everything is true – Step 2' in (17d). As we will see in section 3.1.2 below, non-initial (i.e. SOME-) divisions more generally seem to open up a number of alternative possibilities at the same level of division ('subcategorisation').

Strikingly, Jaspers (2012) also shows how the (development of) human colour perception appears to follow the kind of successive division path MMM-driven development would predict. More generally, there is evidence from (developmental) cognitive psychology showing that object classification 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). Various child language acquisition phenomena also point in this direction. 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 (cf. Guasti 2017 for an overview) are all cases in point. And in the parsing domain, *good enough* parsing, in term of which humans preferentially operate with a shallow parse until it becomes clear that deeper parsing is required (Ferreira & Patson 2007) also looks like a reflex of MMM. The same is true for the evidence pointing to the use of fast and frugal

heuristics in decision-making, i.e. Daniel Kahnemann's (2001) *fast thinking* (see Gigerenzer & Todd 2000 for the seminal fast and frugal heuristics paper), and the picture that seems to be emerging from the study of writing systems: the majority of characters in writing systems are made of three strokes or less (Dehaene 2007), with cardinal orientations (horizontal and vertical) being vastly over-represented in the world's languages, compared to oblique ones, as one might expect, given humans' superior ability to compute the former (orientational anisotropy; Morin 2018).[19] We will discuss further linguistic 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 for overview discussion); and

---

19. Thanks to Daniel Harbour for discussion.

d. the various ways in which the earliest-acquired categories – centring on a basic predicate-/"archi"-**V** versus argument-/ "archi"-**N**-type category (cf. also Bouchard 2013; Douglas 2018; and Song 2019[20]) – are put to "extended" use in grammar structuring. Consider, for example, the varied evidence pointing to the existence of extended projections (Grimshaw 1991 et seq.), which are typically thought to be defined with reference to basic lexical categorial features (e.g. V, N, P, etc.); on the present account, these basic features may usefully be thought of in the kind of not fully fleshed-out "archi" terms discussed in Douglas (2018) and Song (2019). As we will see in section 3.1.2, extended-projection membership imposes structural constraints of different kinds. Another case in point is the ubiquity of verbalization and nominalization phenomena, where the latter seems to serve both a general "subordinating" function (e.g. in subordination and embedding structures; cf. Franco 2012 for discussion and reference, and Huddleston 1984: 379-380 for the distinction between these two), and – the opposite – a foregrounding purpose (as in VP topicalization/focus). Among finiteness-marking languages, we also see many languages which harness the distribution and inflectional marking of the verb to signal notions that can be lexically expressed too, e.g. declarative vs interrogative marking, and main- vs subordinate-clause status, as in (non-English) Germanic; or realis vs irrealis, as in some Romance. V also often acts as a reference point for focus (see recent work by Kriszta Szendrői and Fatima Hamlaoui, and Vieri Samek-Lodovici, and the more general existence of immediately-before- and immediately-after-verb focus systems – see Gibson, Kombarou, Marten & van der Wal 2017), 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).

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 3). In a system which maximizes minimal means, in which [F]s already serve as the basis on which the UG-given Agree operation 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 (see also Preminger 2018 on this).

20. Douglas (2018: 28, note 22), working within an MMM perspective, helpfully characterises the notions 'archi-V' and 'archi-N' as follows:

'We must think of the N/V distinction as distinguishing nominal features and verbal features (or nominal features and non-nominal features), which will eventually be successively subdivided into the finer-grained categories of the adult grammar (including [N] and [V]). The N/V distinction thus involves *archi-features* (by analogy with archi-phonemes): archi-N (**N**) and archi-V (**V**).'

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; Biberauer 2018; Roberts 2019):[21]

(19)

Does P(roperty) characterise L(anguage)?

NO: **macroparameter**   YES: All relevant heads?

   YES: **macroparameter**   NO: A natural-class subset of heads?

      YES: **mesoparameter**   NO: A further restricted natural-class subset of
                                                              heads?

         YES: **microparameter**   NO: Only lexically
                                                                  specified items?
                                                                  nanoparameter
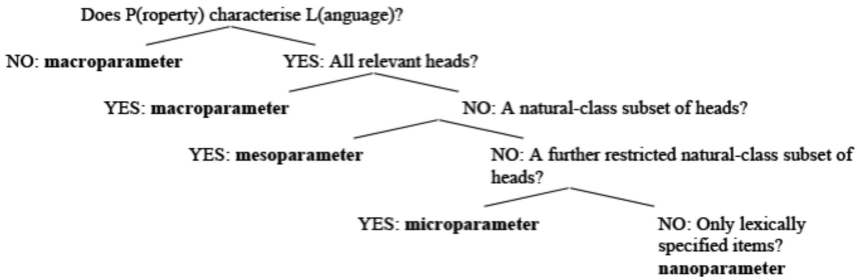
(20) For a given value v_i of a parametrically variant feature F:

   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;

   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.

Taking a specific example, the fact 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, 2018; Roberts 2019 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

---

21. 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.

    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 "archi"-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 "archi"-V and N as fulfilling **parallel roles** in **structuring different grammars** (just as [high] and [round] did in (16a) and (16b) above; cf. also Wiltschko 2014 on the distinct, but formally parallel choice of one of [tense], [person] and [location] as the designated 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 fixed sequence of any kind (*pace* the parameter hierarchies in i.a. Biberauer, Holmberg, Roberts & Sheehan 2014, and Roberts 2019). 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. Typologically equivalent SOME-choices, which are not successively considered in the acquisition context, are thus not typically in a featural superset/subset relationship; let us call these *SOME$_{Equivalent}$ choices*. By contrast, SOME-choices that are successively considered during acquisition are in a featural superset/subset relationship; let us call these *SOME$_{Subset}$ choices*. (22) illustrates the difference with reference to the typology of head-final systems presented in (21) above:

(22) a. "rigid" head-finality: Japanese, Malayalam, etc.   [**ALL**]

    b. clausal head-finality, nominal head-initiality, and *vice versa*: Chinese, Thai, Gungbe, etc.   [**SOME$_{Equivalent}$**]

    c. "leaking" OV of different kinds, e.g. West Germanic   [**SOME$_{Subset}$**]

    d. OVX, where O is the direct object   [**SOME$_{Subset}$**]

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

    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.

A further new possibility is highlighted in Biberauer & Roberts (2012, 2016, 2017). These works point out that the "size"-based parametric approach set out in (19-20) 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 (23) (Biberauer & Roberts 2008 show that OV-loss in the history of English appears to have followed the kind of "cascading" pathway sketched out in (23b,c):
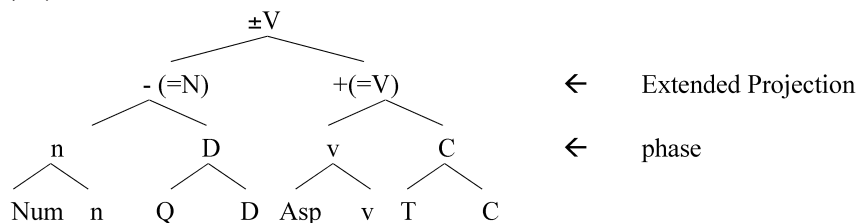
(23) (simplified) schema of potential changes in the **nature of the preverbal position** in an initially "rigidly" head-final OV system:

  a.   all Os > all non-clausal complement (DP, PP, etc.)[22]

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

  c.   all DPs (nominal objects only) > specific sub-types 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 (23), but clause-oriented, with the restriction referencing [tense], [aspect], [finiteness], etc. In this case, we would expect different diachronic possibilities, which need also not all go in the same direction (i.e. OV loss and VO gain; OV>VO is also diachronically attested, and the MMM system allows for changes in both directions, depending on how key aspects of the rest of the system are configured).

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 the acquisition of syntactic categories, for example, we might expect something like (24) 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 1993/1994 Truncation model; the ATOM model of Schütze & Wexler 1996; see Biberauer & Roberts 2015 for discussion of (24)):

22. 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).

(24)

```
                        ±V
              ┌──────────┴──────────┐
          - (=N)                  +(=V)              ←        Extended Projection
        ┌────┴────┐            ┌─────┴─────┐
        n         D           v           C          ←        phase
      ┌─┴─┐     ┌─┴─┐       ┌─┴─┐       ┌─┴─┐
    Num   n   Q     D    Asp     v     T     C
```

In terms of (24), we expect acquirers to want to utilize the (in part prosodi-cally mediated; see section 2.2) [F] facilitating the initial "archi"-V vs N distinc-tion (here: [$\pm V$]) 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 English-acquiring children appear to be confident about "basic" interrogative properties like *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), there seems to be good motivation for proposing that the (clause-typing-related) category C may define the second "archi"-V-based ([+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 learn-ing path in (14) 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 (e.g. V>v>Asp>T>C) often assumed, suggesting instead that acquirers will successively postulate ini-tially 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[23] (FOFC; see i.a. Biberauer, Holmberg & Roberts 2014; Sheehan 2013; Sheehan, Biberauer, Roberts & Holmberg in 2017). FOFC is stated in (25):

23. Note that, as of 2017, the *C* in *FOFC* stands for *Condition*. *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 initial, constraint-oriented acronym seemingly suggested; Final-over-Initial is what is barred.

(25) **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/BHR 2008 *et seq.*, notably BHR 2014)

What (25) requires is that head-finality start at the bottom of an Extended
Projection, i.e. with a lexical 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 (^ signifies head-finality in each
case):

(25) Three very basic FOFC-respecting patterns:

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

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

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

   > **monotonicity**: structurally adjacent heads consistently bear ^

(26) Three basic FOFC-violating patterns:

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

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

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

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

As noted elsewhere (Biberauer, Holmberg & Roberts 2008; Biberauer, Newton
& Sheehan 2009; Biberauer, Sheehan & Newton 2010; BHR 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 sig-
nificantly 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 nomi-
nalized (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).[24]

a.   nominal   b. *   nominal

nominal   nominal   nominal   **verbal**

     nominal   **verbal**   nominal   **verbal**   **verb**

          verbal   **verb**   **verbal**   **verb**
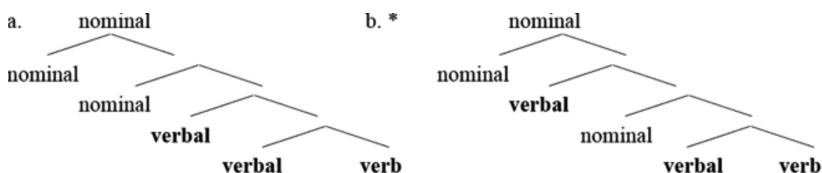
Similarly, in the domain of Agreement, we see (non)-agreement "cut-
off" effects exhibiting the same profile (see Biberauer 2017b for discussion).
Additionally, the various hierarchies proposed by typologists and others, and the
recently much-discussed *ABA syncretism constraint (cf. i.a. Caha 2009; Bobaljik
& Sauerland 2018 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).

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 been 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 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 has, as noted in the introduction, always been a clear sense
in which all acquisition requires the acquirer to go beyond the input: children end
up with knowledge of systematicities that simply aren't available to them via the
input. *That*-trace effects in languages that have them constitute one striking exam-
ple (Rizzi 1982, 1986). Here, we will briefly consider three further scenarios that
uncontroversially involve going beyond the input. One relates to artificial language
learning, and the other two 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).

---

24. 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.

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 of the time   (**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 examples both come from English. The first concerns 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. Before we move on to consider the patterns observed in the past tense, it is worth briefly considering why all three of the NONE>ALL>SOME patterns emerge in the present tense. To the extent that the ALL-choice rests on the postulation of featurally more complex phase-heads than the NONE-choice, we might, after all, expect there to have to be a further grammatical signal that this increased featural complication relative to the evidently available NONE-option is warranted. Importantly, however, NONE- and ALL-options can also be equally complex. Where an [F] is already part of a system, generalising it over a (novel) class of heads will, for example, conform to both IG and FE (cf. (12) and (13) above). Where the decision is simply a matter of spellout – consistently do/don't spell out a specified feature – there need also not be any complexity difference in play. Both considerations seem to hold for

the dialects that opt for ALL/NONE reanalyses of the verbal [number] marking. These reanalyses render a non-[number] analysis of some kind necessary. In the case of the ALL-*s*-realisation systems, -*s* arguably spells out only [present], which is simpler than the standard [3psg, present] specification; as we know that [tense] is already present in the verbal system at the stage at which [number] is extended to it from the nominal system (cf. i.a. Miller 2007; Miller & Schmitt 2012a,b), this [F]-attribution would not seem to entail the postulation of any new [F]s. In the case of the NO-*s*-realisation systems, the unmarked verbs once again need to be specified for [present], even in the absence of an overt spellout, to accommodate speakers' awareness of the [tense] specification, which is very evident in *do*-support contexts (interrogatives, tag structures, etc.). Whether -*s* is realised or not, then, an already-present feature [tense] will be ascribed to the consistently (un)inflected verb-forms in both the ALL- and the NONE-systems. And the same is true in the case of the Northern Subject Rule SOME-systems: here -*s* realisation always appears to be regulated by an already-present lexical-functional distinction (between full nominals and pronouns), potentially further mediated by "shallow"-seeming linear (i.e. PF-based) considerations. The NONE>ALL>SOME options in this case therefore seem to be comparable in "cost" terms. Since the regularity in question – what to make of -*s* – is known to be later-acquired (see again Miller 2008, Miller & Schmitt 2012a,b, and also Brown's classic (1973) Morpheme Order Study), this cost-equivalence is in fact unsurprising in the context of the present model: as already noted above, non-initial (i.e. later) choices do not necessarily take the form of featurally more or less complex options, or SOME$_{Subset}$ choices; instead, they may simply be alternative SOME or SOME$_{Equivalent}$ choices (cf. the discussion around (22)).

Turning to the past tense, we see that number marking in this domain in standard English is even more restricted than in the present, surfacing only on *BE* (i.e. *was/were*). In the vernacular varieties, we again see a number of 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 (see (31)).   **(SOME)**

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

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

As in the case of the present tense, the NONE>ALL>SOME options given in (30-31) can all be shown to be cost-equivalent. Thus the generalization options parallel the -*s*/-∅-generalisation options discussed for the present tense: both require the postulation of the feature [past], i.e. an instantiation of the [tense]-feature, which is demonstrably part of the English verbal system prior to verbal number marking. The grammatically defined SOME-choice that emerges in the past tense,

likewise, piggybacks on an [F] already present in the system, namely [polarity]. What determines this specific choice of [F]? One 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 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 Woods & Roeper in press for recent discussion and references; note also that this fits with the discussion surrounding (24) above).[25] [Polarity] then seems to be an early-acquired [F], at least in English, 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. Like [tense], this feature is already part of the grammar at the point where the acquirer is seeking a featural rationale for the singular-plural distinction on *BE*, meaning that this SOME$_{Subset}$ option is as "economical" as the options that, at first sight, appear to be "simpler" NONE- or ALL-options (the generalisation options in (30a)).

Our second real-life example comes from West Ulster English. As previously discussed in McCloskey (2000, 2016) and also Henry (2012, 2015), this variety of English permits unusually extensive quantifier-float options in A-bar contexts. Consider (33) in this regard; parentheses indicate the various *all*-placement options:

(33) What (all) did he (all) say (all) that he (all) bought (all)?

Henry (2015), however, shows that these options are not necessarily available to all West Ulster speakers; instead, it appears to be the case that different "floating" grammars exist, as illustrated in (34):

(34) a. What all did he say that he bought?
     b. What (all) did he (all) say (all) that he (all) bought (all)
     c. What (all) did he say (all) that he bought?
     d. What (all) did he say (all) that he bought (all)?
     e. What (all) did he (all) say that he (all) bought?
     f. What (all) did he (all) say that he (all) bought (all)?

---

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

(34a) is the standard English, no-floating grammar, while (34b) instantiates the grammar which permits stranding in all possible positions. (34c-f), in turn, represent grammars in which some natural-class subset of these options is available. The picture as a whole can be characterised as in (35):

(35) a. What all did he say that he bought?   NONE

b. What (all) did he (all) say (all) that he (all) bought (all)   ALL

           (*vP- & CP-edge plus base position*)

c. What (all) did he say (all) that he bought?   SOME$_{Equivalent}$

           (*CP-edge only*)

d. What (all) did he say (all) that he bought (all)?   SOME$_{Equivalent}$

           (*CP-edge plus base position*)

e. What (all) did he (all) say that he (all) bought?   SOME$_{Equivalent}$

           (*vP-edge only*)

f. What (all) did he (all) say that he (all) bought (all)?   SOME$_{Equivalent}$

           (*vP-edge plus base position*)

The pattern that we see, then, involves an across-the-board licensing or ban of stranding possibilities (35a,b), or the licensing of stranding options targeting one or other phase-edge with or without the quantifier's base position being a further possibility (35c-f).

Crucially, McCloskey (2016) observes that the input for these structures will be very scarce indeed, raising the question of how the variant stranding grammars are acquired: they will clearly fall beyond the prescriptive radar, and it is also not the case that the A- and A-bar stranding patterns in a given system necessarily overlap in any way. Here, then, we undoubtedly face another "going beyond the input" scenario, where acquirers are converging on grammars that conform to the NONE>ALL>SOME expectations that an MMM-mediated model would predict for input-poor scenarios generally. What I would like to suggest – in advance of fieldwork to establish the actual facts – is that input from other components of the grammar that are already in place will enable the acquirer to converge on an appropriate grammar. Data alerting the child to the need or not to distinguish between different clausal phase heads (C, v) could, for example, (help to) determine the size and composition of the class of stranding-permitting heads. One type of data that might be relevant in this regard – particularly also if we bear in mind the need to pinpoint structures that could plausibly be salient enough to supply the acquirer with the relevant input at a suitably early stage – is the inverted-subject imperative. The examples in (36) demonstrate the fact that these are not equally readily available in all varieties of Ulster English (data from Henry 1995, 2015):

(36) a. Sit you down!   [Dialect A: ✔; Dialect B: ✔]

　　 b. Go you away!   [Dialect A: ✔; Dialect B: ✔]

　　 c. Run you!   [Dialect A: ✔; Dialect B: ✗]

　　 d. Read you that book!   [Dialect A: ✔; Dialect B: ✗]

As (36) shows, some varieties permit inverted-subject imperatives, regardless of verb-type (Henry's *Dialect A*), while others exhibit argument-structure-based constraints on the availability of this imperative-type. Henry's *Dialect B*, for example, only permits inverted-subject imperatives with telic intransitives; thus transitive (36d) and the atelic intransitive in (36c) are both ruled out. In terms of a fairly standard minimalist view, v is the phase-head that regulates argument-structure and so-called *first-phase* syntax more generally (cf. i.a. Ramchand 2008, and D'Alessandro, Franco & Gallego 2017), while C is the phasal locus of clause-typing and (at least some – see Heim & Wiltschko 2017) discourse-related properties. Accepting this view, we see that acquirers of Dialect A-type systems will receive evidence from a high-frequency – and presumably also highly salient – input structure that discourse-marked (i.e. non-neutral declarative) v and C phase-heads can be generalised across, i.e. IG as in (13) can apply. In this case, then, we might expect NONE or ALL stranding grammars to be postulated as there is another well attested non-neutral, A-bar-structure where the relevant clausal phase heads can all be treated identically: **all** vs are compatible with the inverted-subject-associated imperative C, i.e. any v can match up with the relevant type of C, and so we might also expect all vs and Cs to behave identically in relation to quantifier stranding. Acquirers exposed to Dialect B-type systems, by contrast, will receive imperative evidence that the v and C phase-heads cannot simply be treated as a natural class in the context of discourse-marked (i.e. non-neutral/non-declarative) structures: transitive and atelic vPs need to be distinguished to capture the constraint on the distribution of inverted-subject imperatives. In these grammars, then, we might expect acquirers not to generalise across v and C to produce either a NONE or ALL grammar; instead, postulation of one of the SOME grammars presumably allows them to exploit the already-present featural discrepancies between phase heads in their target variety. If this kind of approach to the quantifier-stranding possibilities depicted in (34/35) is on the right track, we again, as in the case of verbal number-marking, see that apparent NONE>ALL>SOME options in fact constitute $SOME_{Equivalent}$ options, with the result that acquirers have a number of equally MMM-compatible options for resolving a poverty-of-the-stimulus-type indeterminacy.

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 was ascribed a role in the context of the model presented here, with the general cognitive factor, Maximise Minimal Means, being argued to 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):[26]

> 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, William. 1989. On the particulate principle of self-diversifying systems. *Journal of Social Biological Structure* 12: 1-13.

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

Ambridge, Ben, Pine, Jonathan & Lieven, Elena. 2013. Child language acquisition: why Universal Grammar doesn't help. *Language* 90(3): e53-e90.

---

26. Thanks to Itziar Laka for drawing attention to this important extract during a generative linguistics event in Reading (May 2017).

Baunaz, Lena, De Clercq, Karen, Haegeman, Liliane & Lander, Eric. 2018. *Exploring Nanosyntax.* Oxford: Oxford University Press.

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

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

Biberauer, Theresa. 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, Theresa. 2017a. Particles and the Final-over-Final Constraint. In Sheehan, Michelle, Biberauer, Theresa, Holmberg, Anders & Roberts, Ian (eds.). *The Final-over-Final Condition*, 187-296. Cambridge, MA: MIT Press.

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

Biberauer, Theresa. 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, Theresa. 2017e. Factors 2 and 3: a principled approach. *Cambridge Occasional Papers in Linguistics* 10: 38-65. [Available on-line: <http://www.ling. cam.ac.uk/COPIL/Articles/COPiL_10_3_Biberauer.pdf>].

Biberauer, Theresa. 2018. Pro-drop and emergent parameter hierarchies. In Cognola, Federica & Casalicchio, Jan (eds.). *Null Subjects in Generative Grammar: A Synchronic and Diachronic Perspective*, 94-135. Oxford: Oxford University Press.

Biberauer, Theresa, Bockmühl, Juliane, Herrmann, Erika & Shah, Sheena. 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, Theresa, Holmberg, Anders & Roberts, Ian. 2008. Disharmonic word-order systems and the Final-over-Final-Constraint (FOFC). In Bisetto, Antonetta & Barbieri, Francisco (eds). *Proceedings of XXXIII Incontro di Grammatica Generativa*. [Available on-line at: <http://amsacta.cib.unibo.it/archive/00002397/01/ PROCEEDINGS_IGG33.pdf>].

Biberauer, Theresa, Holmberg, Anders & Roberts, Ian. 2014. A syntactic universal and its consequences. *Linguistic Inquiry* 45(2): 169-225.

Biberauer, Theresa, Holmberg, Anders, Roberts, Ian & Sheehan, Michelle. 2014. Complexity in comparative syntax: the view from modern parametric theory. In Newmeyer, Frederick J. & Laurel Preston (eds.). *Measuring Linguistic Complexity*, 103-127. Oxford: Oxford University Press.

Biberauer, Theresa, Newton, Glenda & Sheehan, Michelle. 2009. Limiting synchronic and diachronic variation and change: the Final-over-Final Constraint. *Language and Linguistics* 10(4): 699-741.

Biberauer, Theresa & Roberts, Ian. 2008. Cascading parameter changes: internally driven change in Middle and Early Modern English. In Þórhallur Eythórsson (ed.). *Grammatical Change and Linguistic Theory: The Rosendal Papers*, 79-113. Amsterdam: Benjamins.

Biberauer, Theresa & Roberts, Ian. 2009. The Return of the Subset Principle. In Crisma, Paola & Longobardi, Giuseppe (eds.). *Historical Syntax and Linguistic Theory*, 58-74. Oxford: Oxford University Press.

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

Biberauer, Theresa & Roberts, Ian. 2015. Rethinking formal hierarchies: a proposed unification. *Cambridge Occasional Papers in Linguistics* 7: 1-31.

Biberauer, Theresa & Roberts, Ian. 2016. Parameter typology from a diachronic perspective: the case of Conditional Inversion. In Bidese, Ermenegildo, Cognola, Federica & Moroni, Manuela C. (eds.). *Theoretical Approaches to Linguistic Variation*, 259-291. Amsterdam: Benjamins.

Biberauer, Theresa & Roberts, Ian. 2017. Parameter setting. In Ledgeway, Adam & Roberts, Ian (eds.). *The Cambridge Handbook of Historical Syntax*, 134-162. Cambridge: Cambridge University Press.

Biberauer, Theresa & Sheehan, Michelle. 2012. Disharmony, antisymmetry, and the Final-over-Final Constraint. In Etxebarria, Myriam & Valmala, Vidal (eds.). *Ways of Structure Building*, 206-244. Oxford: Oxford University Press.

Biberauer, Theresa & Sheehan, Michelle. 2013. Introduction. In Biberauer, Theresa & Sheehan, Michelle (eds.). *Theoretical Approaches to Disharmonic Word Order*, 1-46. Oxford: Oxford University Press.

Biberauer, Theresa, Sheehan, Michelle & Newton, Glenda. 2010. Impossible changes and impossible borrowings: the Final-over-Final Constraint. In Breitbarth, Anne, Lucas, Christopher, Watts, Sheila & Willis, David (eds.). *Continuity and Change in Grammar*, 35-60. Amsterdam: John Benjamins.

Bobaljik, Jonathan. 2002. A-Chains at the PF Interface: copies and covert movement. *Natural Language and Linguistic Theory* 20(2): 197-267.

Bobaljik, Jonathan & Sauerland, Uli. 2018. *ABA and the combinatorics of morphological rules. *Glossa: A Journal of General Linguistics* 3(1): 15.

Boeckx, Cedric. 2015. *Elementary Syntactic Structures: Prospects of a Feature-Free Syntax*. Cambridge: Cambridge University Press.

Bond, Oliver, Corbett, Greville, Chumakina, Maria & Brown, Dunstan. 2016. *Archi: Complexities of Agreement in Cross-theoretical Perspective*. Oxford: Oxford University Press.

Bornstein, Marc & Arterberry, Martha. 2010. The development of object categorization in young children: Hierarchical inclusiveness, age, perceptual attribute and group versus individual analyses. *Developmental Psychology* 46: 350-365.

Bouchard, Denis. 2013. *The Nature and Origin of Language*. Oxford: Oxford University Press.

Branigan, Phil. 2012. Macroparameter learnability: an Algonquian Case Study. Unpublished ms: Memorial University of Newfoundland.

Brown, Roger. 1973. *A First Language: The Early Stages*. Cambridge, MA: Harvard University Press.

Caha, Pavel. 2009. *The Nanosyntax of Case*. PhD dissertation, CASTL Tromsø.

Chomsky, Noam. 1959. Review of *Verbal Behaviour* by B.F. Skinner. *Language* 35: 26-58.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, Noam. 1986. *Knowledge of Language. Its Nature, Origin, and Use.* Westport, CT: Praeger.

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In Hale, Ken & Keyser, S. Jay (eds.). *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, 1-52. Cambridge, MA: MIT Press.

Chomsky, Noam. 1995. *The Minimalist Program.* Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: the framework. In Martin, Roger, Michaels, David & Uriagereka, Juan (eds.). *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, 89-156. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In Kenstowicz, Michael (ed.). *Ken Hale: A Life in Language*, 1-50. Cambridge, MA.: MIT Press.

Chomsky, Noam. 2005. Three factors in Language Design. *Linguistic Inquiry* 36: 1-22.

Chomsky, Noam. 2007. Of minds and language. *Biolinguistics* 1: 9-27.

Chung, Sandy. 2012. Are lexical categories universal? The view from Chamorro. *Theoretical Linguistics* 38: 1-56.

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

Cinque, Guglielmo. 2013. Cognition, Universal Grammar, and typological generalizations. *Lingua* 130: 50-65.

Cinque, Guglielmo. 2018. A microparametric approach to the head-initial/head-final parameter. *Linguistic Analysis* 41(3/4): 309-366.

Clark, Eve. 1993. *The Lexicon in Acquisition.* Cambridge: Cambridge University Press.

Crain, Stephen & Pietroski, Paul. 2001. Nature, nurture and Universal Grammar. *Linguistics and Philosophy* 24(2): 139-186.

D'Alessandro, Roberta, Franco, Irene & Gallego, Ángel. 2017. *The Verbal Domain.* Oxford: Oxford University Press.

D'Alessandro Roberta & van Oostendorp, Marc (2018). Magnetic Grammar. Unpublished ms: Utrecht and Radboud Universities. (<https://ling.auf.net/lingbuzz/004312>; last accessed 5 December 2018)

Dehaene, Stanislas. 2007. *Reading in the Brain: The New Science of How We Read.* London: Penguin.

Demuth, Katherine. 1994. On the 'underspecification' of functional categories in early grammars. In Lust, Barbara, Suñer, Margarita & Whitman, John (eds.). *Syntactic Theory and First Language Acquisition: Cross-Linguistic Perspectives*, 119-134. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Demuth, Katherine. 2003. The acquisition of Bantu languages (eds). In *The Bantu Languages*. Nurse, Derek & Philippson, Gerard (eds.), 209-222. Surrey: Curzon Press.

Douglas, Jamie. 2018. Maori subject extraction. *Glossa* 3(1): 110.

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

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

Dryer, Matthew. 2009. Verb-object-negative order in Central Africa. In Cyffer, Norbert, Ebermann, Ewald & Ziegelmeyer, Georg (eds.). *Negation Patterns in West African Languages and Beyond*, 307-362. Amsterdam: Benjamins.

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

Duffield, Nigel. 2017. On what projects in Vietnamese. *Journal of East Asian Linguistics* 26(4): 351-387.

Eguren, Luis, Fernandez-Soriano, Olga & Mendikoetxea, Amaya. 2016. *Rethinking Parameters*. Oxford: Oxford University Press.

Elman, Jeffrey. 1993. Learning and development in neural networks: the importance of starting small. *Cognition* 48(1): 71-99.

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

Epstein, Samuel D., Kitahara, Hisatsugu & Seely, T. Daniel. 2013. Simplest Merge generates set intersection: Implications for complementizer-trace explanation. In Goto, Nobu, Otaki, Koichi, Sato, Atsushi & Takita, Kensuke (eds.). *Proceedings of GLOW in Asia IX.*, 77-92. Mie University. Reprinted in: Epstein, Samuel D., Kitahara, Hisatsugu & Seely, T. Daniel. 2015. *Explorations in Maximizing Syntactic Minimization*. London: Routledge, 175-94.

Evers, Arnold & van Kampen, Jacqueline. 2008. Parameter setting and input reduction. In Biberauer, Theresa (ed.). *The Limits of Syntactic Variation*, 483-515. Amsterdam: Benjamins.

Fasanella, Adriana. 2014. *On how Learning Mechanisms shape Natural Languages*. Ph.D. dissertation, UAB. (<http://filcat.uab.cat/clt/publicacions/tesis/pdf/fasanella_14.pdf>)

Fasanella, Adriana & Fortuny, Fortuny. 2016. Deriving linguistic variation from learnability conditions: the Chunking Procedure. In Eguren, Luis, Fernandez-Soriano, Olga & Mendikoetxea, Amaya. *Rethinking Parameters*, 105-132. Oxford: Oxford University Press.

Ferreira, Fernanda & Patson, Nikole. 2007. The good enough approach to language comprehension. *Language and Linguistics Compass* 1: 71-83.

Fikkert, Paula. 1994. *On the Acquisition of Prosodic Structure*. Dordrecht: ICG Printing.

Fodor, Janet D. & Sakas, William G. 2005. The Subset Principle in syntax: costs of compliance. *Journal of Linguistics* 41: 513-569.

Foley, Claire, Núñez del Prado, Zelmira, Barbiers, Isabella & Lust, Barbara. 2003. Knowledge of variable binding in VP-ellipsis: language acquisition research and theory converge. *Syntax* 6: 1-52.

Fodor, Janet D. & Sakas, William G. 2017. Learnability. In Roberts, Ian (ed.). *The Oxford Handbook of Universal Grammar*, 249-269. Oxford: Oxford University Press.

Fortuny, Jordi. 2010. On the duality of patterning. In Zwart, Jan-Wouter (ed.). *Structure Preserved: Studies in Syntax for Jan Koster*, 131-140. Amsterdam: Benjamins.

Franco, Ludovico. 2012. Against the identity of complementizers and (demonstrative) pronouns. *Poznan Studies in Contemporary Linguistics* 48: 565-596.

Freeman, Thomas. 2016. *Metaphysical Syntax.* MPhil dissertation, University of Cambridge.

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

Gagliardi, Annie. 2012. *Input and Intake in Language Acquisition.* PhD dissertation, Maryland.

Gallego, Ángel. 2011. Parameters. In Boeckx, Cedric (ed.). *Oxford Handbook of Linguistic Minimalism*, 523-550. Oxford: Oxford University Press.

Gass, Susan. 1997. *Input, Interaction, and the Second Language Learner.* Mahwah, NJ: Erlbaum.

Gervain, Judit, Nespor, Marina, Mazuka, Reiko, Horie, Ryota & Mehler, Jacques (2007). Bootstrapping word order in prelexical infants: a Japanese-Italian cross-linguistic study. *Cognitive Psychology* 57: 56-74.

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

Gervain, Judit & Werker, Janet. 2008. How infant speech perception contributes to language acquisition. *Language and Linguistics Compass* 2(6): 1149-1170.

Gianollo, Chiara, Guardiano, Cristina & Longobardi, Giuseppe. 2008. Three fundamental issues in parametric linguistics. In Biberauer, Theresa (ed.). *The Limits of Syntactic Variation*, 109-142. Amsterdam: Benjamins.

Gibson, Hannah, Koumbarou, Andriana, Marten, Lutz & van der Wal, Jenneke. 2017. Locating the Bantu conjoint/disjoint alteration in a typology of focus marking. In Van der Wal, Jenneke & Hyman, Larry (eds.). *The Conjoint/Disjoint Alternation in Bantu*, 61-100. Berlin: Mouton de Gruyter.

Gigerenzer, Gerd & Todd, Peter. 2000. Fast and frugal heuristics: the adaptive toolbox. In Gigerenzer, Gerd, Todd, Peter & The ABC Research Group (eds.). *Evolution and Cognition. Simple Heuristics that make us Smart*, 3-34. Oxford: Oxford University Press.

Göksun, Tilbe, Roeper, Tom, Hirsh-Pasek, Kathy & Golinkoff, Roberta. 2011. From nounphrase ellipsis to verbphrase ellipsis: The acquisition path from context to abstract reconstruction. In Harris, Jesse & Grant, Margaret (eds.). *Occasional Working Papers in Linguistics 38: Processing Linguistic Structure*, 53-74. Amherst, MA: GLSA.

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

Guasti, Maria-Teresa. 2017. *Language Acquisition. The Growth of Grammar.* Cambridge, MA: MIT Press.

Guardiano, Cristina & Longobardi, Giuseppe. 2017. Parameter theory and parametric comparison. In Roberts, Ian (ed.). *The Oxford Handbook of Universal Grammar*, 377-398. Oxford: Oxford University Press.

Hall, Daniel C. 2007. *The Role and Representation of Contrast in Phonological Theory.* PhD dissertation, University of Toronto.

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

Hauser, Marc, Chomsky, Noam & Fitch, Tecumseh. 2002. The faculty of language: what is it, who has it, and how did it evolve? *Science* 298(5598): 1569-1579.

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

Heim, Johannes & Wiltschko, Martina. 2017. The complexity of speech acts. Evidence from speech act modifiers. Unpublished ms: UBC.

Henry, Alison. 1995. *Belfast English and Standard English: Dialect Variation and Parameter Setting.* Oxford: Oxford University Press.

Henry, Alison. 2012. Phase edges, quantifier float and the nature of (micro-) variation. *IBERIA* 4(1): 23-39.

Henry, Alison. 2015. Morphosyntactic variation and finding the open values in syntax. Paper presented at the Formal Approaches to Morphosyntactic Variation conference (Vitoria).

Hockett, Charles. 1958. *A Course in Modern Linguistics.* New York: Macmillan Company.

Hornstein, Norbert & Nunes, Jairo. 2008. Adjunction, labeling, and bare phrase structure. *Biolinguistics* 2: 57-86.

Hornstein, Norbert & Pietroski, Paul. 2009. Basic operations: Minimal syntax-semantics. *Catalan Journal of Linguistics* 8: 113-39.

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

Hudson Kam, Carla & Newport, Elissa. 2005. Regularizing unpredictable variation: the roles of adult and child learners in language formation and change. *Language Learning and Development* 1: 151-95.

Humboldt, Wilhelm von. 1836. *Über die Verschiedenheit des menschlichen Sprachbaues.* Paderborn: Verlag Ferdinanc Schöningh.

Jaspers, Dany. 2012. Logic and colour. *Logica Universalis* 6: 227-248.

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

Kahnemann, Daniel. 2011. *Thinking, Fast and Slow.* London: Penguin.

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

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

Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1: 199-244.

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

Leffel, Timothy, Šimik, Radek & Wierzba, Marta. 2013. Pronominal F-markers in Basaa. *Proceedings of NELS 43*: 265-276. GLSA Publications.

Leivada, Evelina. 2017. What's in (a) Label? Neural origins and behavioral manifestations of Identity Avoidance in language and cognition. *Biolinguistics* 11: 1-30.

Levin, Theodore. 2016. Successive-cyclic case assignment: Korean nominative-nominative case-stacking. *Natural Language and Linguistic Theory* 35(2): 447-498.

Lidz, Jeff & Gagliardi, Annie. 2015. How nature meets nurture: Universal Grammar and statistical learning. *Annual Review of Linguistics* 1: 333-353.

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

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

Lleó, Conxita & Demuth, Katherine. 1999. Prosodic constraints on the emergence of grammatical morphemes: crosslinguistic evidence from Germanic and Romance languages. In: Greenhill, Annabel, Littlefield, Heather & Tano, Cheryl (eds.). *Proceedings of the 23rd Annual Boston University Conference on Language Development (BUCLD 23)*, volume 2: 407-418. Somerville, MA: Cascadilla Press.

Longobardi, Giuseppe. 2018. Principles, Parameters, and Schemata: a radically underspecified UG. *Linguistic Analysis* 41(3-4): 517-558.

Martí, Luisa. 2015. Grammar versus Pragmatics: Carving Nature at the Joints. *Mind and Language* 30(4): 437-473.

McCloskey, Jim. 2000. Quantifier float and *wh*-movement in an Irish English. *Linguistic Inquiry* 31: 57-84.

McCloskey, Jim. 2006. Questions and questioning in a local English. In Zanuttini, Raffaella, Campos, Hector, Herburger, Elena & Portner, Paul (eds.). *Crosslinguistic Research in Syntax and Semantics: Negation, Tense, and Clausal Architecture*, 87-126. Washington, DC: Georgetown University Press.

McCloskey, Jim. 2016. Micro-parameters in a tiny space: stranding at the edge. Paper presented at CamCoS 5 (Cambridge, UK).

Mehler, Jacques, Jusczyk, Peter, Lambertz, Ghislane, Halsted, Nilofar, Bertoncini, Josiane & Amiel-Tison, Claudine. 1988. How infant speech perception contributes to language acquisition. A precursor of language acquisition in young infants. *Cognition* 29: 143-178.

Mehler, Jacques & Dupoux, Emile. 1994. *What Infants Know: The New Cognitive Science of Early Development*. Cambridge, MA: Blackwell.

Miller, Karen. 2007. *Variable Input and the Acquisition of Plurality in Two Varieties of Spanish.* PhD dissertation, Michigan State University.

Miller, Karen & Schmitt, Christina. 2012a. Not all children agree: acquisition of agreement when the input is variable. *Language Learning and Development* 8(3): 255-277.

Miller, Karen & Schmitt, Christina. 2012b. Variable input and the acquisition of plural morphology. *Language Acquisition* 19(3): 223-261.

Milsark, Gary. 1974. *Existential sentences in English*. PhD dissertation, MIT.

Mobbs, Iain. 2015. *Minimalism and the Design of the Language Faculty*. PhD dissertation, University of Cambridge.

Moore, David. 2002. Auditory development and the role of experience. *British Medical Bulletin* 63: 171-181.

Morin, Olivier. 2018. Spontaneous emergence of legibility in writing systems: the case of orientation anisotropy. *Cognitive Science* 42(2): 664-677.

Nazzi, Thierry, Bertoncini, Josiane & Mehler, Jacques. 1998. Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance* 24: 756-766.

Nespor, Marina, Peña, Marina & Mehler, Jacques. 2003. On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio* 2: 203-229.

Newmeyer, Frederick J. 2004. Against a parameter-setting approach to language variation. In Pica, Pierre, Rooryck, Johan & van Craenenbroeck, Jeroen (eds.). *Language Variation Yearbook Volume 4*, 181-234. Amsterdam: Benjamins.

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

Newport, Elissa. 1990. Maturational constraints on language learning. *Cognitive Science* 14: 11-28.

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

Pearl, Lisa. in press. Modelling syntactic acquisition. In J. Sprouse (ed.). *Oxford Handbook of Experimental Syntax*. Oxford: Oxford University Press.

Pearl, Lisa & Sprouse, Jon. in press. Comparing solutions to the Linking Problem using an integrated quantitative framework of language acquisition. To appear in *Language*. (<https://ling.auf.net/lingbuzz/003913>; last accessed 5 December 2018).

Pérez-Leroux, Ana T., Castilla-Earls, Anny P., Bejar, Susana & Massam, Diane. 2012. Elmo's sister's ball. The development of nominal recursion in children. *Language Acquisition* 19(4), 301-311.

Pérez-Leroux, Ana T., Peterson, Tyler, Bejar, Susana, Castilla-Earls, Anny P., Massam, Diane & Roberge, Yves. 2018. The acquisition of recursive modification in NPs. *Language* 94(2): 332-359.

Pesetsky, David. 2014. *Russian Case Morphology and the Syntactic Categories*. Cambridge, MA: MIT Press.

Picallo, Carme (ed.). 2014. *Linguistic Variation in the Minimalist Framework*. Oxford: Oxford University Press.

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

Preminger, Omer. 2018. Back to the Future: non-generation, filtration, and the heartbreak of interface-driven minimalism. In Hornstein, Norbert, Lasnik, Howard, Patel-Grosz, Pritty & Yang, Charles (eds.). *Syntactic Structures after 60 Years: The Impact of the Chomskyan Revolution in Linguistics*, 355-380. Berlin: de Gruyter.

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

Ramchand, Gillian. 2008. *Verb Meaning and the Lexicon. A First Phase Syntax*. Cambridge: Cambridge University Press.

Ramchand, Gilllan & Svenonius, Peter. 2014. Deriving the functional hierarchy. *Language Sciences* 46: 152-174.

Richards, Marc. 2014. Defective Agree, Case alternations, and the prominence of Person. In Bornkessel-Schlesewsky, Ina, Malchukov, Andrej & Richards, Marc (eds.). *Scales and Hierarchies: A Cross-Disciplinary Perspective*, 173-194. Berlin: Mouton de Gruyter.

Richards, Marc. 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, Norvin. 2010. *Uttering Trees*. Cambridge: MIT Press.

Riemsdijk, Henk van. 2008. Identity avoidance: OCP effects in Swiss relatives. In Freidin, Robert, Otero, Carlos P. & Zubizarreta, Maria Luisa (eds.). *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*, 227-250. Cambridge, MA: MIT Press.

Ritter, Elisabeth & Wiltschko, Martina. 2009. Varieties of INFL: TENSE, LOCATION and PERSON. In van Craenenbroeck, Jeroen (ed.). *Alternatives to Cartography*, 153-201. Berlin: Mouton de Gruyter.

Ritter, Elisabeth & Wiltschko, Martina. 2014. The composition of INFL. *Natural Language and Linguistic Theory* 32(4): 1331-1386.

Rizzi, Luigi. 1982. *Issues in Italian Syntax*. Dordrecht: Foris.

Rizzi, Luigi. 1986. Null objects in Italian and the theory of *pro*. *Linguistic Inquiry* 17: 501-557.

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

Rizzi, Luigi. 2014. On the elements of syntactic variation. In Picallo, Carme (ed.). *Linguistic Variation in the Minimalist Framework*, 13-45. Oxford: Oxford University Press.

Rizzi, Luigi. 2018. On the format and locus of parameters: the role of morphosyntactic features. *The Linguistic Review* 41(3-4): 159-190.

Rizzi, Luigi & Cinque, Guglielmo. 2016. Functional categories and syntactic theory. *Annual Review of Linguistics* 2: 139-163.

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

Roberts, Ian. 2012. On the nature of syntactic parameters: a programme for research. In Galves, Charlotte, Cyrino, Sonia, Lopes, Ruth, Sandalo, Filomena & Avelar, Juanito (eds.). *Parameter Theory and Linguistic Change*, 319-334. Oxford: Oxford University Press.

Roberts, Ian. 2019. *Parameter Hierarchies and Universal Grammar*. Oxford: Oxford University Press.

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

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

Roeper, Tom & Snyder, William. 2004. Recursion as an analytic device in acquisition. In van Kampen, Jacqueline & Coopmans, Peter (eds.). *Proceedings of GALA 2003*, 401-408. Utrecht: LOT.

Roeper, Tom & Snyder, William. 2005. Language learnability and the forms of recursion. In Di Sciullo, Anna-Maria (ed.). *UG and External Systems: Language, Brain and Computation*, 155-169. Amsterdam: Benjamins.

Sandler, Wendy. 2010. Prosody and syntax in sign language. *Transactions of the Philological Society* 108(3): 298-328.

Sandler, Wendy. 2012. The phonological organization of sign languages. *Language and Linguistics Compass* 6(3): 162-182.

Santos, Ana. 2009. *Minimal Answers: Ellipsis, Syntax, and Discourse in the Acquisition of European Portuguese*. Amsterdam: Benjamins.

Schütze, Carson T. & Wexler, Kenneth. 1996. Subject case licensing and English root infinitives. In Stringfellow, Andy, Cahana-Amitay, Dalia, Hughes, Elizabeth & Zukowski, Andrea (eds.). *BUCLD 20 Proceedings*, 670-681. Cambridge, MA: Cascadilla Press.

Schuler, Katherine, Yang, Charles & Newport, Elissa. 2016. Testing the Tolerance Principle: children form productive rules when it is more computationally efficient to do so. *Proceedings of the 38th Annual Conference of the Cognitive Science Society*: 2321-2326.

Seuren, Pieter & Jaspers, Dany. 2014. Logico-cognitive structure in the lexicon. *Language* 90: 607-643.

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

Sheehan, Michelle, Biberauer, Theresa, Holmberg, Anders & Roberts, Ian (2017). *The Final-over-Final Condition. A Syntactic Universal*. Cambridge, MA: MIT Press.

Shi, Rushen & Werker, Janet F. 2001. Six-month-old infants' preference for lexical words. *Psychological Science* 12(1): 70-75.

Shi, Rushen, Werker, Janet F. & Morgan, James L. 1999. Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition* 72(2): B11-B21.

Shlonsky, Ur. 2010. The cartographic enterprise in syntax. *Language and Linguistics Compass* 4(6): 417-429.

Song, C. (2019). *On the formal flexibility of syntactic categories*. PhD dissertation, University of Cambridge.

Starke, Michal. 2009. Nanosytax. A short primer to a new approach to language. *Nordlyd* 36(1): 1-6.

Starke, Michal. 2014. Towards elegant parameters: language variation reduces to the size of lexicaly-stored trees. In Picallo, Carme (ed.). *Linguistic Variation in the Minimalist Framework*, 140-154. Oxford: Oxford University Press.

Stokes, Stephanie, Klee, Thomas, Perry Carson, Cecyle & Carson, David. 2005. A phonemic implicational feature hierarchy of phonological contrasts for English-speaking children. *Journal of Speech, Language, and Hearing Research* 48: 817-833.

Svenonius, Peter. 2002. Introduction. In Svenonius, Peter (ed.). *Subjects, Expletives and the EPP*, 3-28. Oxford: Oxford University Press.

Thornton, Rosalind. 1995. Referentiality and *wh*-movement in child English: Juvenile D-Linkuency. *Language Acquisition* 4: 139-175.

Tsimpli, Ianthi. 2014. Early, late or very late? Timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism* 4(3): 283-313.

Vikner, Sten. 1995. *Verb Movement and Expletive Subjects in the Germanic Languages*. Oxford: Oxford University Press.

Wallage, Phillip. 2017. *Negation in Early English: Grammatical and Functional Change*. Cambridge: Cambridge University Press.

Westergaard, Marit. 2009. *The Acquisition of word order: Micro-cues, Information Structure and Economy*. Amsterdam: Benjamins.

Wexler, Kenneth. 1998.Very early parameter setting and the unique checking constraint: a new explanation of the optional infinitive stage. *Lingua* 106: 23-79.

Willis, David. 2016. Exaptation and degrammaticalization within an acquisition-based model of abductive reanalysis. In Norde, Muriel & van de Velde, Freek (eds.). *Exaptation in Language Change*, 227-260. Amsterdam: Benjamins.

Wiltschko, Martina. 2014. *The Universal Structure of Categories. Towards a Formal Typology.* Cambridge: Cambridge University Press.

Woods, Rebecca & Roeper, Tom. in press. Rethinking auxiliary doubling in adult and child language. To appear in: Woods, Rebecca & Wolfe, Sam (eds.). *Rethinking Verb Second*. Oxford: Oxford University Press.

Yang, Charles. 2016. *Price of Productivity. How Children Learn and Break Rules of Language.* Cambridge, MA: MIT Press.

Zeijlstra, Hedde. 2008. On the syntactic flexibility of formal features. In Biberauer, Theresa (ed.). *The Limits of Syntactic Variation*, 143-174. Amsterdam: Benjamins.