

Less IS More: some thoughts on the Tolerance Principle in the context of the Maximise Minimal Means model *

THERESA BIBERAUER
UNIVERSITY OF CAMBRIDGE

1 INTRODUCTION

Recent work by Charles Yang seeks to probe the question of the circumstances under which humans form linguistic generalisations about the structure of the language input to which they are exposed. Yang (2016) presents an extended argument for the so-called *Tolerance Principle* (TP), a simplified version of which is given in (1) (see Yang 2016, 2018a/in press, 2018b) for discussion of the harmlessness of the simplification, and Yang (2016) for detailed motivation of the non-simplified form):

(i) The Tolerance Principle (TP)

Let a rule R be defined over a set of N items. R is productive if and only if e , the number of items not supporting R , does not exceed θ_N :

$$(1) \quad e \leq \theta_N = \frac{N}{\ln N}$$

If e exceeds θ_N , then the learner will ‘lexicalize’ only these and not generalize beyond them: that is, R is unproductive.

The purpose of this short paper is to consider the TP in the context of the so-called *Three Factors* approach to language design (Chomsky 2005), and, specifically, to examine how it might relate to a specific Three Factors proposal, namely the so-called *Maximise Minimal Means* (MMM) model of Biberauer (2011), (see Biberauer 2017, 2018c/in press for an overview). The paper is structured as follows: Section 2 introduces the Three Factors approach, and Biberauer’s MMM model, before Section 3 and Section 4 consider the kinds of insights that the TP and the MMM model can, respectively, give as to the ways in which “Less” in first language acquisition¹ can be “More”; Section 5 very briefly highlights the question of how the TP and MMM might be expected to operate in later language learning and language contact; and Section 6 concludes.

* This paper is a somewhat less terse version of a 1500-word commentary, solicited by the editors of a special edition of *Linguistic Approaches to Bilingualism* as a response to Charles Yang’s forthcoming Target Article, ‘A formalist perspective on language acquisition’ (Biberauer 2018b/in press). Thanks to Holger Hopp and Neal Snape for their input on the commentary, to Paula Buttery for initially drawing my attention to the ‘Goldilocks Effect’ (see section 3), and to the editors of the current volume for their very considerable patience. Usual disclaimers apply.

¹ By *first language acquisition*, I mean both native acquisition of a single language, and simultaneous acquisition of more than one language from (soon after) birth (so-called *simultaneous bilingualism*).

2 THREE FACTORS RATHER THAN TWO

Thirteen years since the publication of [Chomsky \(2005\)](#) and over a quarter of a century since the start of the Minimalist era in generative syntax and linguistics more generally, the model of grammar acquisition and structure that is most commonly attributed to generativists is still the “classic” one depicted in (2):

- (2) Universal Grammar/UG + Primary Linguistic Data/PLD (=input) → an adult (L1) grammar

In terms of this model, the formal specification of an adult L1 grammar (or I-language; cf. [Chomsky 1986](#)) is principally determined by the interaction of two components: an innately given, richly (i.a. also parametrically) specified UG on the one hand, and the input to which the contents of UG “steers” the acquirer’s attention - the so-called *PLD* - on the other (cf. [Chomsky 1981](#) for overview discussion). Crucially, the PLD was not assumed to be “everything the acquirer hears”, but, instead, only that part of the input that UG facilitated initial access to. [Evers & van Kampen \(2008\)](#) and [Gagliardi \(2012\)](#), among others, refer to this sub-part of the input that acquirers receive as the *intake*, a term that we will also adopt here on the grounds that it emphasises the need to think carefully about the components of the overall input that an acquirer might be able to access in the acquisition context.

The shortcomings of the “classic” Principles and Parameters model are by now well known (see [Newmeyer 2005](#), [Biberauer 2008](#), [Roberts & Holmberg 1994](#), and the discussions in [Picallo 2014](#), [Eguren, Fernández-Soriano & Mendikoetxea 2016](#)). These, combined with increasing understanding of the workings of human cognition - see i.a. [Piatelli-Palmarini, Uriagereka & Salaburu \(2010\)](#), [Berwick & Chomsky \(2013\)](#), [Gazzaniga, Ivry & Mangun \(2014\)](#) - led, at the start of the current millennium, to a revision of (2), the *Three Factors Model* schematised in (3):

- (3) UG + PLD (=intake) + general cognitive principles → an adult (L1) grammar

In terms of (3), UG is, crucially, assumed to be a minimally specified entity, i.a. lacking the rich parametric specification assumed during the “classic” Two Factors era. Against this backdrop, some, though certainly not all, 21st century work in the generative framework aims to ‘reduce[] the explanatory burden traditionally placed on innate linguistic parameters’ ([Yang 2018a/in press](#): p. 670) by trying to understand and explicate the grammar-shaping role of cognitive mechanisms that are not language-specific. In [Chomsky’s \(2005: p.6\)](#) terms, these include both ‘principles of data analysis ... used in language acquisition and other domains’, and ‘principles of efficient computation’.

Yang’s TP offers an explicit formal proposal as to what a potential third-factor principle - which rests on the well established (general cognitive) fact that humans are curiously sensitive to statistics (cf. i.a. [Yang 2004](#), [Lidz & Gagliardi 2015](#), [Pearl & Goldwater 2016](#)) - might look like. Additionally, the TP also makes reference to the nature and role of the linguistic input (Factor 2): it provides ‘elements’ (cf. (1)

above), which Yang (2018a/in press) strongly identifies with lexical items, which then provide evidence either for or against the postulation of a rule-based generalisation. In relation to Factor 1, Yang is less transparent about his assumptions: while, as noted above, referencing the Three Factors imperative to ‘reduce[] the explanatory burden traditionally placed on innate linguistic parameters’ (Yang 2018a/in press: p.670), he also evidently relies on a parametrically specified UG in associated work, e.g. that done in the context of his well-known Variational Learning Model (Yang 2002), which is also endorsed in Yang (2018a/in press). Here, and elsewhere, the kinds of parameters assumed in the “classic” Principles and Parameters - or “rich UG” - era serve as the locus for the initially equal probabilities or weights that are adjusted as the acquirer is exposed to the PLD. Regardless of the precise UG commitment Yang wishes to make, what is very clear is that his approach ‘doesn’t eliminate the parameter *per se*, which determines the options’ (Rothman & Chomsky 2018/in press); nor, crucially, does it suggest the format for the rules to which the TP is applied, or give us any basis for adjudicating how different the format for these rules might be to that of the parameters that generativists might be wanting to postulate in the current minimalist era.² In other words, what still requires spelling out in the context of Yang’s TP proposals is the nature of the basis for the generalisations - be they rules or parameters - that the TP regulates (see also Rothman & Chomsky 2018/in press: p.768 for the same observation).

Like Yang, Biberauer (2017, 2018c/in press) aims to put forward an explicit Three Factors model: the MMM model. In terms of this model, UG is assumed to lack all but a very basic specification (the operations Merge and Agree, and a formal feature template, in terms of which grammaticalised properties are represented), with formerly UG-contained content like parameters emerging as a result of this interaction of the minimal UG specification with Factors 2 and 3. In the absence of UG “steering”, the aspects of the PLD that are defined as being of particular relevance to the acquirer, in the sense of constituting their hypothesised ‘intake’ (see above), are what Biberauer (2017, 2018c/in press) describes as *systematic departures from Saussurean arbitrariness*. The leading idea here is that children encounter two types of arbitrariness in the course of acquisition: the arbitrary and thus unpredictable form-meaning mappings that define ‘content words’ (or ‘roots’ in Distributed Morphology terms), and the still arbitrary, but more systematic and

² This matter is of particular interest in view of the fact that parameter-sceptics like Newmeyer (2004, 2005) reject parameters while in fact accepting two types of “language-particular rule”: the first is the non-UG mediated, idiosyncratic type that Yang’s TP has so often been tested in relation to; the second, however, he characterises as ‘parameter-settings “detached” from the parameters themselves’ (Newmeyer 2004: p.183). For Newmeyer, this latter type is indeed UG-mediated, but that UG mediation takes the form of general, crucially parameter-free principles like ‘All phrases are headed’ and ‘There are limits on the structural distance between the launching and landing site of a moved element’ (Newmeyer 2004: p.184). In emergentist approaches like that outlined in Biberauer (2017, 2018c/in press) (see also Biberauer, Roberts & Sheehan 2014, Longobardi 2018, Epstein, Obata & Seely 2018 and section 4 below), this second type of “rule” would, in fact, chiefly result from the interaction of third-factor considerations with the PLD, with UG playing only a minor role. See Section 4 below for some further discussion of the ways in which more idiosyncratic, “periphery”-oriented rule-type generalisations and more “core”, parameter-type generalisations might relate to each other on an emergentist approach.

thus less unpredictable form-meaning mappings that define functional items, or the locus of (parametric) variation in Borer-Chomsky Conjecture (BCC; [Baker 2008](#)) terms. In essence, the proposal is that any systematic departure from one-to-one form-meaning mapping will - at the appropriate stage of the acquirer's development (see section 3 below) - alert the child to the need to analyse the input not just in terms of the phonological and semantic features defining the Saussurean form-meaning mappings, but, additionally, in terms of grammar-defining formal features (cf. [Chomsky 1995](#) on the three-way feature distinction between phonological, semantic and formal features). Systematic departures from one-to-one form-meaning mappings include:

- (4) a. doubling phenomena where there appears to be “extra” form in play (e.g. agreement or concord);
- b. “silence” where there appears to be missing form (e.g. null arguments, ellipsis, and null exponence of different kinds);
- c. multifunctionality/syncretism, where the same form appears to have distinct functions, depending on its distribution, suggesting an underspecification analysis (cf. [Wiltschko 2014](#), [Duffield 2017](#), [Baunaz, de Clercq, Haegeman & Lander 2018](#) for discussion of relevant case studies);
- d. movement of the kind that facilitates [Chomsky's \(2001\)](#) ‘duality of semantics’, or the use of a single element to express both first-phase (i.e. broadly thematic) and second-phase (i.e. broadly, discourse-oriented) meaning; and
- e. consistent ‘edge’ distribution of the kind that characterises the distribution of functional elements (see section 3 below for further discussion).

As noted above, these properties lead to the postulation of formal features ([F]s), which define the structure of the grammar. Crucially, [F]-postulation is constrained by a Factor Three constraint, namely a general cognitive bias to Maximise Minimal Means. In the domain of [F]-postulation, MMM drives acquirers to restrict their postulation of [F]s, in part by prioritising the (re-)use of already-postulated [F]s over the postulation of new [F]s. In terms of [F] postulation, then, ‘Less is More’ (cf. i.a. [Newport 1988, 1990](#)). More generally, as we will see in Section 3, MMM predicts that acquirers will always seek to make the most of the input – or, more accurately, intake – to which they have access.

Strikingly, Yang’s TP also has the potential to facilitate insight into one aspect of how ‘Less’ can be ‘More’ for child-acquirers. Consider (5) from Yang (2018b) to see why this may be so:

(5) The maximum number of exceptions for a productive rule over N items

| N | θ_N | % |
|-------|------------|------|
| 10 | 4 | 40.0 |
| 20 | 7 | 35.0 |
| 50 | 13 | 26.0 |
| 100 | 22 | 22.0 |
| 200 | 38 | 19.0 |
| 500 | 80 | 16.0 |
| 1,000 | 145 | 14.5 |
| 5,000 | 587 | 11.7 |

As (5) clearly shows, the TP as formulated in (1) leads us to expect that the number of exceptions tolerated by a rule increases as the number of lexical items known (N above) increases. To the extent that this prediction is on the right track (see i.a. the commentaries by Cécile de Cat, Christine Dimroth and Roumyana Slabakova in the volume of *Linguistic Approaches to Bilingualism* in which the shorter version of this commentary is due to appear - Biberauer 2018b/in press), it can thus be easier to learn a rule while one has a limited vocabulary than it is when one’s vocabulary is more extensive: less vocabulary may quite literally be “more” in the sense of facilitating quicker rule formulation, or, in Yang’s formulation ‘If children’s vocabulary is smaller, the odds of acquiring productive rules improve considerably.’ (Yang 2018a/in press: p.690).

In what follows, I would like to strongly endorse the ‘Less is More’ spirit of Yang’s model (Section 3), and to suggest how the more “periphery”-oriented TP (cf. Yang 2016: p.218-23) may be integrated with a similarly ‘Less is More’-oriented three-factors approach to “core” grammar (Section 4), which potentially also sheds light on an important difference between L1 and L2 acquirers (Section 5).

3 ‘LESS IS MORE’ IN LANGUAGE ACQUISITION - EVEN PRE-LEXICALLY

As its formulation makes very clear, the TP is strongly lexically oriented: to do the tolerance calculation, it is necessary to identify a domain associated with a specific number of lexical items to which a hypothesised generalisation (=rule) could apply. For the tolerance calculation to be meaningful - i.e. for the TP actually to **drive** acquisition - we additionally need there to be exceptions to the hypothesised rule. These requirements conveniently allow us to situate the TP in relation to the MMM model introduced above.

In terms of the MMM-model, child acquirers aim to maximally exploit the knowledge at their disposal at all stages of the acquisition process, including the pre-lexical stage. As indicated in Section 2, MMM itself is conceived as a general, non-language-specific learning bias, which drives acquirers to construct their grammars incre-

mentally on the basis of knowledge accessible to them at a given point. The idea is essentially that lesser linguistic access to begin with - i.e. less input which actually qualifies as intake - allows acquirers to focus their attention on a sub-component of the input, which they are then able to master and harness as the basis for access to more complex, previously inaccessible aspects of the input.

And there is considerable evidence suggesting that L1 acquisition does indeed progress in this “Goldilocks” manner, with acquirers systematically attending to input that is neither too simple nor too complex, but “just right” - thus paralleling what we see in other domains (cf. i.a. Kidd, Piantadosi & Aslin 2012, 2014 on vision and audition respectively).³ 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 foetal auditory system is functional from around 6 months’ gestation (Mehler & Dupoux 1994, Moore 2003). 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, Jusczyk, Lambertz, Halsted, Bertoncini & Amiel-Tison 1988, Nazzi, Bertoncini & Mehler 1998, Gervain, Nespor, Mazuka, Horie & Mehler 2007), and also, subsequently, their strikingly early ability to establish the “basic” (i.e. 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 (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 formal properties (they are, in general, shorter [fewer syllables], 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

³ As Elman (1993) had already indicated, this “Goldilocks” pattern may give us insight into why human beings, who are born so helpless and who reach maturity so slowly in comparison to other species in the living universe, are such exceptionally good learners. In Elman’s words, ‘Maturational changes may provide the **enabling conditions** which allow learning to be most effective.’ (my emphasis - TB; Elman 1993: p.71). Strikingly, the ‘Less is More’ effect has also been demonstrated for i.a. correlation detection (Kareev 1995), and contingency assessment (Fiedler & Kareev 2006) by humans. Crucially, Fiedler & Kareev (2006) and Anderson & Doherty (2007) further show that the circumstances under which this small-sample advantage reliably holds are in fact heavily constrained, requiring conservative and, objectively, suboptimal decision criteria, i.e. the kind that we might expect young children with limited experience to employ. No less strikingly, a ‘Less is More’/‘Start Small’ effect has also been observed in computational modelling, starting with Elman’s 1993 neural network models.

distinction between content and functional items, producing 6-month-olds with a preference for the former (see again 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 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. 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 become "bored" by them: they are no longer "just right" in the Goldilocks sense, and we see a shift in interest to more fine-grained, but still not inaccessibly complex 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 is at least partly "steered" by what the existing grammatical specification makes accessible to the acquirer.⁴ 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 that prefigures what Yang argues to be so important in the context of lexical learning. This might lead us to ask whether the TP - which is argued to underlie acquirers' **lexically based** 'Less is More' successes - can be understood as a further manifestation of the more general cognitive bias to Maximise Minimal Means.

In the following section, I will further suggest that an MMM perspective may allow us to refine our understanding of how what we might think of as "core"

⁴ Importantly, the acquirer's formal knowledge is, on the MMM approach, assumed to be represented in terms of feature-based representations which draw on the UG-given feature template mentioned in Section 2. These feature-based representations are assumed to arise as a consequence of (i) an acquirer registering a formal regularity - conceived of as a systematic departure from Saussurean arbitrariness; see Sections 2 and 4 - in the intake (Factor 2), (ii) UG dictating that this regularity should be encoded in terms of the universal feature template, and (iii) the third-factor bias to Maximise Minimal Means leading the acquirer to try to harness an already-postulated formal feature to account for the new regularity, only postulating a new formal feature where this is not feasible. The formal specifications that drive ever more detailed acquisition of an L1 grammar are thus the outcome of the interaction of all three of the factors in (3). As UG remains one of the core grammar-structuring elements, it can also be seen, on the MMM model, to retain some of the "steering" role that it had in the "classic" Principles and Parameters period (see Section 2).

and “peripheral” components of the grammar (cf. [Chomsky 1981](#)) relate to one another. This is relevant to our understanding of the potential relationship between the TP and MMM because [Yang \(2016: p.218-23\)](#) explicitly relates the TP to the “periphery”. This is, however, evidently not the perspective adopted by all those who endorse the TP: [Rothman & Chomsky \(2018/in press: p.768\)](#), for example, identify as ‘the greatest contribution of Yang’s model ... its overcoming the stipulation of values for innate linguistic parameters (i.e. the key shapers of the “core” during the classic Principles and Parameters period - TB)’. Section 4 picks up where the present section left off in considering how MMM might shape acquisition of “core” grammar, before considering whether it might be possible to integrate Yang’s TP-sanctioned peripheral rules with an MMM-shaped grammar.

4 ‘LESS IS MORE’ WITH SOME LEXICON IN PLACE: INTEGRATING PARAMETRIC VARIATION?

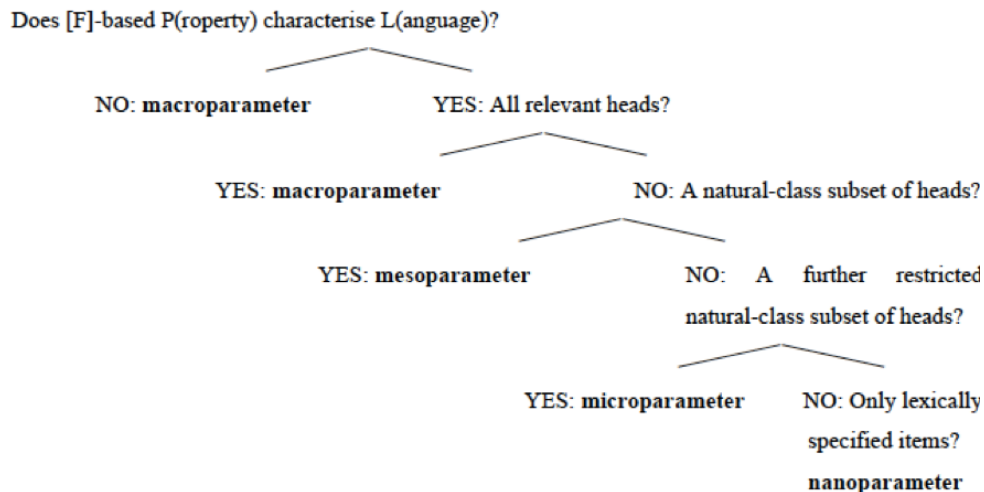
In MMM terms, what is expected once acquirers have some vocabulary in place is that they will discover the key distinction between the truly arbitrary, memorization-requiring form-meaning mappings that typically define content items - classic Saussurean arbitrariness - and the “higher-level” arbitrariness that defines the recurring form and distribution of functional elements within a system, the systematic departures from Saussurean arbitrariness introduced in Section 2 above. As indicated there, the MMM proposal is that acquirers will latch onto systematic cues of the kind in (4) to drive the postulation of grammar-defining formal/[F]-features alongside content-item-defining semantic/[S]- and phonological/[P]-features ([Chomsky 1995](#)).

Two points are crucial: firstly, the [F]s under discussion here are those defining the formal make-up of functional categories, and, hence, in the context of the so-called *Borer-Chomsky Conjecture* regarding the nature of syntactic variation ([Baker 2008](#)), parametric variation, or “core grammar”. Secondly, the postulation of grammatical [F]s, in accordance with MMM-driven Feature Economy (‘Postulate as few [F]s as possible to account for observed regularities’) and Input Generalisation (‘Maximise the use of already-postulated [F]s’), allows the acquirer to optimise their knowledge of system-defining regularities in a manner entirely parallel to that underlying the TP-governed postulation of “periphery”-oriented rules: both reflect a response to the acquirer’s ‘search for [memorization-limiting - TB] productive generalisations’ ([Yang 2018a/in press: p.679](#)); and this search is, I argue, driven by MMM. Further, to the extent that [F]s have both grammar-(Narrow Syntax)internal and realisational (PF) consequences - triggering Agree and Move operations on the one hand, and feeding into placement and morphological realisation on the other - we can begin to refine our understanding of the way in which “core” [F]-mediated regularities and TP-sanctioned peripheral rules might relate to one another.

The conclusion that peripheral rules are more “surfacey”, with the TP merely regulating realisational options, seems too simple. Consider [Biberauer & Roberts \(2015, 2016\)](#)’s “size”-based emergent (i.e. non-UG-given) parameter typology (6), and the kind of ‘learning pathway’ it gives rise to (7):

- (6) For a given value v_i of a parametrically variant feature [F]:
- Macroparameters:** all heads of the relevant type, e.g. all probes/phase heads, etc. share v_i . E.g. harmonic head-finality, radical pro-drop.
 - Mesoparameters:** all heads in a given natural class, e.g. [+V] or all C-/T-heads, share v_i . E.g. Chinese-style head-finality⁵, Romance-style pro-drop, Verb-Second.
 - Microparameters:** a small, lexically definable sub-class of functional heads (e.g. (modal) auxiliaries, subject clitics) show v_i . E.g. Partial pro-drop, English-style Verb-Second.
 - Nanoparameters:** one or more individual lexical items is/are specified for v_i . E.g. Bavarian-style pro-drop, English-style Conditional Inversion.⁶

(7)



Key here is the idea that macro-, meso- and microparameters are formulated over natural classes, i.e. categories defined on the basis of more and less detailed [F]-specifications. As outlined in Biberauer & Roberts (2015, 2016), (L1) acquisition is, in the general case, assumed to proceed “top-down”, with the acquirer’s initially underspecified grammar progressively becoming more articulated in [F]-terms as more of the input becomes accessible. This assumption seems to be supported by the fact that we do seem to observe a range of so-called *shadow* phenomena at earlier acquisition stages, where an incompletely specified item stands in for what should really be a range of forms, or where a form is initially restricted to a given environment before being articulated in various ways to perform a range of related functions (cf. i.a. Demuth 1994, 2003, Lleó 1998, 2001, Lleó & Demuth 1999, van Kampen 2004, 2007, Roeper 2007 for discussion). More restricted/specialised natural

⁵ Head-initial clausal domain; head-final nominal domain.

⁶ For “size”-oriented discussion of pro-drop and V/Aux-to-C phenomena, see Biberauer (2018a) and Biberauer & Roberts (2016) respectively.

classes are featurally more complex, meaning that we expect “smaller” [F]-regulated phenomena to be fully acquired later than “bigger” ones.

Nanoparameters are different, however. They govern non-[F]-unified individual lexical items, are acquired bottom-up, independently of the (timing of the) general top-down process, and may therefore be acquired either early or late, depending on the complexity of the [F]s encoded on the items associated with the nanoparameter. Because nanoparameters don’t target an entire natural class - consider (standard) English Conditional Inversion, which affects only a subset of the language’s seven past-marked auxiliaries, *had*, *were*, *should* - and because the class is quantifiable, there is a sense in which they exhibit the same profile as TP-sanctioned “peripheral” rules. And, English CI in fact satisfies the TP ($N/\ln N = 1.946$). In principle, given the definition of nanoparameter, this need not always be the case, however, leading us to expect TP-regulated instability in this corner of grammar, which seems correct (see [Biberauer & Roberts 2016](#) on the loss of Conditional Inversion in modern and, particularly, contact Englishes). The TP may, then, regulate not just the productivity of regular rules such as those discussed in Yang’s work, but also the stability of certain types of exceptions, the nanoparameters under discussion here. If this is correct, there may in fact be different types of peripheral rule: one governing regular behaviour with exceptions (Yang’s central focus), and one governing exceptional behaviour itself. Both types interface with “core” grammar, with the latter (additionally) qualifying as “(nano)parametric” wherever it encodes item-specific instantiations of a parametric schema (see [Biberauer & Roberts 2016](#), and [Longobardi 2018](#) for discussion). If this is correct, there is a real sense in which MMM and the TP may allow us to understand how the “core” and the “periphery” are in fact contiguous, as originally foreseen in [Chomsky \(1981: p.8\)](#).

‘Marked structures have to be learned on the basis of slender evidence too, so **there should be further structure to the system outside of core grammar**. We might expect that the structure of these further systems relates to the theory of core grammar by such devices as **relaxing certain conditions of core grammar, processes of analogy** in some sense to be made precise, and so on, though there will presumably be independent structure as well.’ (emphasis mine, TB)

5 WHAT ABOUT LATER LANGUAGE LEARNING?

To conclude, let us consider some consequences of the present discussion for our understanding of the differences between L1 and L2 acquisition.

[Yang \(2018a/in press\)](#) suggests that adult learners’ large vocabularies effectively undermine their ability to harness the TP as effectively as L1 acquirers do, leaving them unable to reap ‘Less is More’-type benefits in the context of their language learning. From the perspective of an MMM-based approach, this may well be part of the problem. More generally, however, we would expect both adults’ abundance of highly-developed resources - language-specific and general-cognitive - and their (consequent) inability to “get back to basics” so as to build up their grammars in a

“Goldilocks”-mediated way to fundamentally affect the course and outcome of their language-learning.

Adults, then, have both supra- and sub-word-level knowledge that gets in the way of their replicating L1 acquisition; see again section 3 on the apparently exclusively prosodic focus of the earliest stages of L1 acquisition, and the way in which linguistic knowledge appears to be successfully fleshed out on the basis of properties that become available to the L1 acquirer in a successive fashion. L2 systems are therefore necessarily constructed in a very different way to L1 systems.

The MMM model does lead us to have certain expectations about L2 learning, however. For example, the specifications of the L1 grammar necessarily constitute one of the ‘means’ that are at the L2 learner’s disposal in the context of the L2 learning task. Precisely how they are deployed (at different stages of the learning process - the classic *L1 Access and L1 Transfer* questions; cf. [Foley & Flynn 2013](#) for an overview) might be expected to vary depending on considerations such as whether the L2 learning is tutored or not, the way in which the learner - consciously or unconsciously - preferentially harnesses the other (linguistic and general-cognitive) means at their disposal, and the extent to which the learner is motivated to develop a coherent, fully elaborated L2 system. What one might expect, though, is that surface parallelisms (e.g. shared “basic” word order, *wh*-movement or subject-realisation behaviour) would - at least initially - be encoded on the basis of the formal features ([F]s) that define the (apparently) corresponding property in the L1; and one might also expect acquirers to have a - sometimes unjustified - preference for using [F]s that are already part of their L1 repertoire when structuring evidently L2-specific components of grammar (cf. also i.a. the work of Donna Lardi re on problems of this kind). Transfer, in short, is predicted on the MMM approach; but what that transfer produces is expected to be more varied than is typically predicted by generatively oriented L1-transfer theorists, given the crucially emergent nature of the [F]s in the MMM approach and, consequently, of the categories that they define. The details here, as elsewhere, remain to be explored in future research.

6 CONCLUSION

In this short paper, I have attempted to offer a brief consideration of how we might think about Charles Yang’s already very influential Tolerance Principle (TP) in the context of [Chomsky’s 2005](#) Three Factors approach to language and linguistic variation, and, more specifically, [Biberauer’s \(2017, 2018c/in press\)](#) Maximise Minimal Means model. Viewed against this backdrop, the TP would seem to instantiate a reflex of the more general cognitive bias to Maximise Minimal Means. Regardless of the correctness of this conjecture or of the validity of either the TP or MMM, however, I hope to have shown two things here, namely that: (i) attempts to engage in more precise terms with the content of the notion ‘third factor’ can suggest new, potentially productive perspectives on a range of new and also unresolved old questions, and (ii) three-factors-inspired formal work may have a lot to offer to deepen our understanding of the surprising properties of L1 acquisition, mono- and bilingual, and of the nature of the challenge facing L2 learners.

REFERENCES

- Anderson, R. & M. Doherty. 2007. Sample size and the detection of means: a signal detection account. *Mind and Cognition* 35(1). 50–58.
- Baker, M. 2008. The macroparameter in a microparametric world. In T. Biberauer (ed.), *The limits of syntactic variation*, 351–374. Amsterdam: Benjamins.
- Baunaz, L., K. de Clercq, L. Haegeman & E. Lander (eds.). 2018. *Exploring nanosyntax*. Oxford: Oxford University Press.
- Berwick, R. & N. Chomsky. 2013. *Why only us*. Cambridge, MA: MIT Press.
- Biberauer, T. 2008. *The limits of syntactic variation*. Amsterdam: Benjamins.
- Biberauer, T. 2011. In defence of lexico-centric parametric variation: two 3rd factor-constrained case studies. Paper presented at the Workshop on Formal Grammar and Syntactic Variation: Rethinking Parameters (Madrid).
- Biberauer, T. 2017. Factors 2 and 3: a principled approach. *Cambridge Occasional Papers in Linguistics* 10. 38–65.
- Biberauer, T. 2018a. Pro-drop and emergent parameter hierarchies. In F. Cognola & J. Casalicchio (eds.), *Null subjects in generative grammar: A synchronic and diachronic perspective*, 94–135. Oxford: Oxford University Press.
- Biberauer, T. 2018b/in press. Less IS more: on the Tolerance Principle as a manifestation of Maximise Minimal Means. *Linguistic Approaches to Bilingualism* 8(6). 707–711.
- Biberauer, T. 2018c/in press. Factors 2 and 3: towards a principled explanation. *Catalan Journal of Linguistics* 7.
- Biberauer, T. & I. Roberts. 2015. Rethinking formal hierarchies: a proposed unification. *Cambridge Occasional Papers in Linguistics* 7. 1–31.
- Biberauer, T. & I. Roberts. 2016. Parameter typology from a diachronic perspective: the case of Conditional Inversion. In E. Bidese, F. Cognola & M. Moroni (eds.), *Theoretical approaches to linguistic variation*, 259–291. Amsterdam: John Benjamins.
- Biberauer, T., I. Roberts & M. Sheehan. 2014. No-choice parameters and the limits of syntactic variation. In R. Santana-LaBarge (ed.), *Proceedings of the 31st West Coast Conference on Formal Linguistics (WCCFL 31)*, 46–55. Somerville, MA: Cascadilla Press.
- Chomsky, N. 1981. *Lectures in Government and Binding*. Dordrecht: Foris.
- Chomsky, N. 1986. *Knowledge of language: Its nature, origin and use*. New York: Praeger.
- Chomsky, N. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, N. 2001. Derivation by phase. In M. Kenstowicz (ed.), *Ken Hale: A life in language*, 1–50. Cambridge, MA: MIT Press.
- Chomsky, N. 2005. Three factors in language design. *Linguistic Inquiry* 36. 1–22.
- Demuth, K. 1994. On the ‘underspecification’ of functional categories in early grammars. In B. Lust, M. Suñer & J. Whitman (eds.), *Syntactic theory and first language acquisition: Cross-linguistic perspectives*, 119–134. Hillsdale, N.J.: Lawrence Erlbaum Associates.

- Demuth, K. 2003. The acquisition of Bantu languages. In D. Nurse & G. Philippson (eds.), *The Bantu languages*, 209–222. Surrey: Curzon Press.
- Duffield, N. 2017. On what projects in Vietnamese. *Journal of East Asian Linguistics* 26(4). 351–387.
- Eguren, L., O. Fernández-Soriano & A. Mendikoetxea. 2016. *Rethinking parameters*. Oxford: Oxford University Press.
- Elman, J. 1993. Learning and development in neural networks: the importance of starting small. *Cognition* 48(1). 71–99.
- Epstein, S., M. Obata & D. Seely. 2018. Is linguistic variation entirely linguistic? *Linguistic Analysis* 41(3/4). 481–516.
- Evers, A. & J. van Kampen. 2008. Parameter setting and input reduction. In T. Biberauer (ed.), *The limits of syntactic variation*, 483–515. Amsterdam: Benjamins.
- Fiedler, K. & Y. Kareev. 2006. Does decision quality (always) increase with the size of information samples? Some vicissitudes in applying the law of large numbers. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32. 883–903.
- Foley, C. & S. Flynn. 2013. The role of the native language. In J. Herschensohn & M. Young-Scholten (eds.), *The Cambridge handbook of second language acquisition*, 97–113. Cambridge: Cambridge University Press.
- Gagliardi, A. 2012. *Input and intake in language acquisition*: University of Maryland dissertation.
- Gazzaniga, M., R. Ivry & G. Mangun. 2014. *Cognitive neuroscience. The biology of the mind*. New York: W.W. Norton & Company.
- Gervain, J., M. Nespor, R. Mazuka, R. Horie & J. Mehler. 2007. Bootstrapping word order in prelexical infants: a Japanese-Italian cross-linguistic study. *Cognitive Psychology* 57. 56–74.
- Gervain, J. & J. Werker. 2008. How infant speech perception contributes to language acquisition. *Language and Linguistics Compass* 2(6). 1149–1170.
- van Kampen, J. 2004. Learnability order in the French pronominal system. In R. Bok-Bennema, B. Hollebrandse, B. Kampers-Manhe & P. Sleeman (eds.), *Romance languages and linguistic theory 2002: Selected papers from 'Going Romance', Groningen 28–30 November 2002*, 163–182. Amsterdam: Benjamins.
- van Kampen, J. 2007. A perspective on doubling constructions in child Dutch. *Nordlyd* 34(3). 131–152.
- Kareev, Y. 1995. Through a narrow window: working memory capacity and the detection of covariation. *Cognition* 56(3). 263–269.
- Kidd, C., S. Piantadosi & R. Aslin. 2012. The Goldilocks Effect: human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE* 7(5). e36399.
- Kidd, C., S. Piantadosi & R. Aslin. 2014. The Goldilocks Effect in infant auditory attention. *Child Development* 85(5). 1795–1804.
- Lidz, J. & A. Gagliardi. 2015. How nature meets nurture: Universal Grammar and statistical learning. *Annual Review of Linguistics* 1. 333–353.
- Lleó, C. 1998. Proto-articles in the acquisition of Spanish: Interface between phonology and morphology. In R. Fabri, A. Ortman & T. Parodi (eds.), *Modelle*

- der Flexion: 18. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft*, 407–418. Tübingen: Niemeyer.
- Lleó, C. 2001. Early fillers: undoubtedly more than phonological stuffing. *Journal of Child Language* 28. 262–265.
- Lleó, C. & K. Demuth. 1999. Prosodic constraints on the emergence of grammatical morphemes: crosslinguistic evidence from Germanic and Romance languages. In A. Greenhill, H. Littlefield & C. Tano (eds.), *Proceedings of the 23rd annual Boston University Conference on Language Development (BUCLD 23)*, vol. 2, 407–418. Somerville, MA: Cascadilla Press.
- Longobardi, G. 2018. Principles, parameters, and schemata: a radically underspecified UG. *Linguistic Analysis* 41(3/4). 517–557.
- Mehler, J. & E. Dupoux. 1994. *What infants know: The new cognitive science of early development*. Cambridge, MA: Blackwell.
- Mehler, J., P. Jusczyk, G. Lambertz, N. Halsted, J. Bertoncini & C. Amiel-Tison. 1988. How infant speech perception contributes to language acquisition. a precursor of language acquisition in young infants. *Cognition* 29. 143–178.
- Moore, D. 2003. Auditory development and the role of experience. *British Medical Bulletin* 63. 171–181.
- Nazzi, T., J. Bertoncini & J. Mehler. 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, M., M. Peña & J. Mehler. 2003. On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio* 2. 203–229.
- Newmeyer, F. 2004. Against a parameter-setting approach to language variation. In P. Pica, J. Rooryck & J. van Craenenbroeck (eds.), *Language Variation Yearbook*, vol. 4, 181–234. John Benjamins.
- Newmeyer, F. 2005. *Possible and probable languages*. Oxford: Oxford University Press.
- Newport, E. 1988. Constraints on learning and their role in language acquisition: Studies of the acquisition of American Sign Language. *Cognitive Science* 10. 147–172.
- Newport, E. 1990. maturational constraints on language learning. *Cognitive Science* 14. 11–28.
- Pearl, L. & S. Goldwater. 2016. Statistical learning, inductive bias, and Bayesian inference in language acquisition. In J. Lidz, W. Snyder & C. Pater (eds.), *The Oxford handbook of developmental linguistics*, 664–695. Oxford: Oxford University Press.
- Piatelli-Palmarini, M., J. Uriagereka & P. Salaburu. 2010. *Of minds and language. A dialogue with Noam Chomsky in the Basque Country*. Oxford: Oxford University Press.
- Picallo, C. 2014. *Linguistic variation in the minimalist framework*. Oxford: Oxford University Press.
- Roberts, I. & A. Holmberg. 1994. Introduction: parameters in minimalist theory. In T. Biberauer, A. Holmberg, I. Roberts & M. Sheehan (eds.), *Parametric variation*.

- Null subjects in Minimalist theory*, 1–57. Cambridge: Cambridge University Press.
- Roeper, T. 2007. *The prism of grammar: How child language illuminates humanism*. Cambridge, MA: MIT Press.
- Rothman, J. & N. Chomsky. 2018/in press. Towards eliminating arbitrary stipulations related to parameters: linguistic innateness and that variational model. *Linguistic Approaches to Bilingualism* 41(3/4). 764–769.
- Shi, R. & J. Werker. 2001. Six-month-old infants' preference for lexical words. *Psychological Science* 12(1). 70–75.
- Shi, R., J. Werker & J. Morgan. 1999. Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition* 72(2). B11–B21.
- Tsimpli, I. 2014. Early, late or very late? Timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism* 4(3). 283–313.
- Wexler, K. 1998. Very early parameter setting and the unique checking constraint: a new explanation of the optional infinitive stage. *Lingua* 106. 23–79.
- Wiltschko, M. 2014. *The universal structure of categories. Towards a formal typology*. Cambridge: Cambridge University Press.
- Yang, C. 2002. *Knowledge and learning in natural language*. Oxford: Oxford University Press.
- Yang, C. 2004. Universal grammar, statistics, or both? *Trends in Cognitive Science* 10. 451–456.
- Yang, C. 2016. *The price of linguistic productivity. How children learn to break the rules of language*. Cambridge, MA: MIT Press.
- Yang, C. 2018a/in press. A formalist perspective on language acquisition. *Linguistic Approaches to Bilingualism* 8(6). 665–706.
- Yang, C. 2018b. A user's guide to the Tolerance Principle. Unpublished ms: University of Pennsylvania (<https://ling.auf.net/lingbuzz/004146>; last accessed 15 October 2018).

Theresa Biberauer
 University of Cambridge
mtb23@cam.ac.uk