

# Going beyond the input: three factors and language acquisition (variation and change)<sup>1</sup>

Theresa Biberauer

University of Cambridge & Stellenbosch University

([samtb23@gmail.com](mailto:samtb23@gmail.com))

## 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 in (1):

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

Here, UG was thought to be “rich in structure” (Chomsky 1981:3), with the key consequence that the nurture component (the PLD) could be more restricted, providing – in the context of the classic Principles & Parameters (P&P) era of the 1980s and 1990s – “limited evidence, just sufficient to fix the parameters of UG [which could – TB] ... determine a grammar that may be very intricate and ... in general lack grounding in experience in the sense of an inductive bias”. (*ibid.*). During the Minimalist era, the rich UG assumption has, however, been drawn into question, the objective in this context being to populate UG with **only** the grammar-shaping content that cannot be ascribed to more general cognitive principles. More specifically, Chomsky (2005) proposes the so-called *Three Factors Model*, represented in (2):

- (2) UG + PLD + general cognitive factors → Adult Grammar (=I-language)  
F1      F2                      F3

Here, the additional factor may, for example, include language acquisition biases (‘principles of data analysis ... used in language acquisition and other domains’; Chomsky 2005:6), and constraints on the make-up and workings of the computational system underpinning human language (‘principles of structural architecture’ and ‘principles of efficient computation’; *ibid.*). On this view, UG is not typically assumed to be endowed with a rich parametric specification. To the extent that GB and earlier minimalist-era parameter-based accounts of language acquisition, variation, and change facilitated real insights into language structure, (2), then, means that we need a new account of how adult grammars come to give the appearance of being parametrically shaped systems. My purpose here will be to suggest what an account of this type might look like. Further, I will attempt to outline some of the novel advantages of a three-factors model in the domain of language acquisition, variation and change.

The paper is structured as follows: section 2 introduces the new model, section 3 considers some of its novel predictions, and section 4 concludes.

---

<sup>1</sup> The research reported here was funded by the European Research Council Advanced Grant No. 269752 “Rethinking Comparative Syntax” (ReCoS).

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

The 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 briefly discussed in the following sub-sections, but first a word on the “new” factor: Maximise Minimal Means.

This putative general cognitive factor is conceived as both a generally applicable learning bias harnessed by the acquirer during acquisition, and as a principle of structural architecture, facilitating the kind of efficient computation and also, crucially, the self-diversifying property that allows human language to be the powerful tool that it is. On this latter point, I follow Abler (1989), who highlights particulate structure as the basis of self-diversification, on account of the way it facilitates the creation of so-called *Humboldt systems*, namely those:

- (4) a. which ‘make[ ] infinite use of finite means’ (Humboldt 1836: 70)  
b. whose ‘synthesis creates something that is not present per se in any of the associated constituents’ (Humboldt 1836: 67)

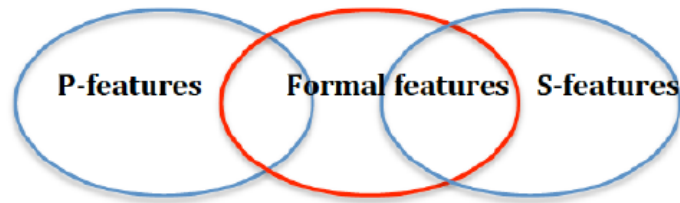
This, of course, calls to mind Hockett’s (1958) ‘duality of patterning’ to which we also return below (see section 2.2).

### 2.1. Factor 1: Universal Grammar

On the present model, UG is thought to provide the basic operations, feature-sensitive (as opposed to ‘blind’) Merge and Agree, plus 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. There may, additionally, be a very small set of universally specified formal features (= [F]s) not derivable from the input (see section 2.2); but **not** the full inventory from which acquirers make a one-time selection postulated in Chomsky (2001: 10): one of this model’s objectives is precisely to try to make progress on the question of what kinds of [F]s are required to characterize natural-language syntax.

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 mappings to produce the Saussurean arbitrariness familiar from the literature. Human language, however, (uniquely?) goes beyond this level of arbitrariness; it additionally involves a “higher” level of arbitrariness defined by Formal ([F]-) features, which map onto [P]- and [S]-features in systematic ways (see section 2.2 below for some discussion, and see also (5) for a rough schematization):

(5)



The proposal, then, is that there are degrees of arbitrariness in human language:

- (6) a. lexically stored, idiosyncratic conventionalized sound-meaning mappings involving just [P]- and [S]-features, and
- b. grammatically regulated and thus more systematically conventionalised sound-meaning mappings, involving [P]-, [S]- and [F]-features.

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

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

It is undeniably the case that generative work to date has not sought to establish a general theory clarifying the aspects of the input that are most crucial to acquirers in fixing upon the regularities (parametric specifications or otherwise) of their grammars. 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 – a goal that still needs to be pursued in the current context, given the seeming existence of regularities for which the input is either rare or non-existent (i.e. where acquisition would require negative evidence of a type not assumed to be available to the acquirer). There was, however, also a challenge that was quite widely acknowledged during the classic P&P era, namely the so-called *Linking Problem* (cf. i.a. Pinker 1984, Gervain & Mehler 2010, Ambridge, Pine & Lieven 2013, and Fasanella & Fortuny 2016 for discussion). This revolves around the question of how the contents of UG, rich or otherwise, is to be linked up to the actual linguistic input that acquirers are exposed to. From the classic P&P perspective, how do acquirers “recognize” the empirical facts that will allow them to set pre-specified parameters in the appropriate way (see Fodor & Sakas 2017 for overview discussion)?

In the present model, the proposal is that the child is specifically looking for what I will call *systematic departures from Saussurean arbitrariness*, i.e. for systematic departures from idiosyncratic one-to-one form-meaning mappings.<sup>2</sup> More specifically, these include:

- (6) a. Doubling/Agreement and expletives/dummy elements, i.e. cases where we have two/multiple forms and one meaning (cf. also Zeijlstra 2008), or one form with no meaning.

---

<sup>2</sup> Cf. also Fasanella (2014) and Fasanella & Fortuny (2016) on the so-called *Chunking Procedure*.

- b. Systematic silence, e.g. null exponence, null arguments, null complementisers, ellipsis, etc., i.e. cases where there appears to be meaning which arises systematically despite the absence of form.
- c. Multifunctionality, i.e. cases where there appears to be what we might think of as system-defining homophony, i.e. a pattern in terms of which a single form can contribute multiple meanings, depending on its placement/distribution (cf. also Duffield 2013, Wiltschko 2014).
- d. High-frequency recurring collocation, i.e. unduly frequent forms with a consistent meaning and position relative to contentful lexical items. 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.<sup>3</sup> Importantly, the difference between the two is signalled both prosodically – function words are shorter and more reduced than content words (lower number of syllables, less complex syllables, less diphthongization, shorter vowel duration, diminished amplitude, etc.) – and in frequency and, crucially, distribution terms – function words are more frequent, and occupy the edges of syntactic domains (see below).
- e. Movement, i.e. assuming Chomsky’s (2000) notion of ‘duality of semantics’ – i.e. roughly that human language expresses both thematic and discourse/scopal meaning – we can see that movement will often result in “extra” meaning. This would, for example, be true in topicalization- and focus-fronting cases. Also relevant here, however, is what we might think of as ‘higher-level duality of patterning’, deriving from the contrast between “neutral/basic” and “marked” orders. Just like Hockettian duality of patterning assumes two levels of structuring – meaningless phonemes which combine to create meaningful phoneme combinations – we might think of syntax as involving “meaningless” structuring that contrasts with meaningful structuring. More specifically, consider on the one hand meaningless “basic” word-order choices like OV vs VO – which are, crucially, known to be acquired early (cf. Tsimpli 2014 for overview discussion) – and meaningless obligatory filling choices like V’s spellout position or the need to fill Spec-TP or Spec-CP; on the other hand, we would have meaningful optional movements like T-to-C in English, or the nature of the XP that raises to Spec-CP. Here, the meaningless conventions require fixing – just like the contents of the phoneme inventory does – whereafter they can serve as the basis for further, potentially meaningful ordering patterns, which contrast with the “basic” one.<sup>4</sup>
- f. Recursion (cf. much work by Tom Roeper and William Snyder, i.a. Roeper 2011, Roeper & Snyder 2011). Recursion here involves

---

<sup>3</sup> Shi, Werker & Morgan (1999), for example, show that newborns can distinguish the prosodic cues associated with content and function words, respectively, while Shi & Werker (2001) demonstrate that a content-word preference already emerges at 6 months.

<sup>4</sup> Having both levels of duality of patterning allows the system to maximise the contribution of both the Lexical Items – i.e. the elements (containing the features) that are manipulated by the computational system – and that system’s structure-building operations, (External and Internal) Merge, as MMM would lead us to expect.

repetition patterns that cannot be ascribed to P- or S-properties. It differentiates “exceptional” domains from truly productive grammar (Roeper & Snyder 2011:158; cf. also Yang 2016).<sup>5</sup>

The driving intuition behind (6-f), then, 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. Strikingly, it appears to be the case that P-features alone (notably prosody) serve as the initial stepping-stone into grammar: much research during the past 20 years has demonstrated acquirers’ sensitivity from birth to the prosodic profile (e.g. strong-weak vs weak-strong) of their target language, and it has similarly been shown that children are able to pick up on the ‘edge-marking’ nature of function words during the pre-linguistic stage (cf. also note 2), a capacity which may, in turn, give them access to core properties like syntactic headedness. With basic, purely P-mediated regularities in place, the child can then proceed to draw on the cues provided by (6a-f)-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 & Shah 2017).

Evidently, the 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 crucially distinguishes the represent approach from “standard” emergentist approaches (e.g. those in the Construction Grammar tradition). Since both the [F]s and the categories they define will be emergent, however, 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, general cognitive bias, which I assume to play a key role in steering acquisition. In the linguistic context, I assume it to have – possibly

---

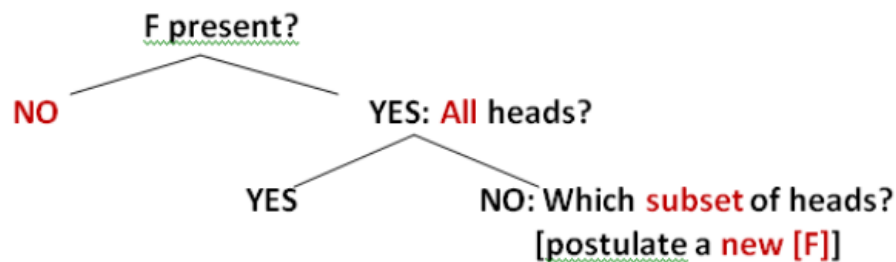
<sup>5</sup> Significantly, recursion also guarantees Distinctness in the sense of Richards (2010), i.e. the requirement that appears to characterize all components of language structure, and in terms of which formally identical elements, which compete for the same positions, cannot surface adjacent to each other within the same domain (cf. the diverse OCP effects that have been identified in phonology and morphosyntax). Cf. also D’Alessandro & van Oostendorp (2016) on so-called *Gravitational Grammar*. That we would see the kinds of repulsion and attraction effects highlighted in this work – and also properties like Relativized Minimality (see section 3) – follows quite directly from the approach outlined here: in systems that maximize minimal means, we expect the number of features and the composite objects constructed from them to be limited in such a way that attraction, repulsion, and intervention effects would be expected to emerge. In a system with too many distinct [F]s, the observed interactions could not be modelled as falling out from simple similarity and difference calculations.

among others – the language-specific manifestations in (7-8):

- (7) **Feature Economy (FE)**: postulate as few formal features as possible to account for the input (=intake) [generalised from Roberts & Roussou 2003]
- (8) **Input Generalisation (IG)**: maximise already-postulated features [generalised from Roberts 2007]

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

- (9) The **NONE>ALL>SOME** learning path



Here, the idea is that (9) models the interaction between the three factors in (2) as follows: an acquirer who does not pick up on a systematic departure from Saussurean arbitrariness in the input will not pose the ‘F present?’ question, with the result that the initial NO is a default which the comparatively oriented linguist can juxtapose with the initial YES, the answer produced by posing this question. The initial NO (or the NONE-system), then, respects both FE and IG. The initial YES (or the ALL-system) necessarily violates FE – as all [F]-postulation and thus, (further) grammar construction, will – but respects IG as the newly identified [F] is assumed to be present on all heads in the relevant domain (all heads in the case of headedness; all argument-licensing heads in the case of null-argument phenomena; all verbal heads in the case of finiteness marking, etc.). Should it emerge that the postulated [F] is not sufficient to delineate the domain over which the property in question is distributed, a further [F] will be postulated, thus producing a SOME-system (at later acquisition stages, this [F] may already be part of the system). If the relevant regularity is still not suitably demarcated, a further [F] is postulated, as before, producing another SOME-system. And so on until the relevant regularity has been appropriately characterized.<sup>6</sup>

Importantly, there appears to be non-syntactic evidence in favour of the validity of postulating MMM and, more specifically, the NONE>ALL>SOME learning path it gives rise to. Dresher’s (2009) Successive Division Algorithm approaches the acquisition of phonology, and thus, by extension, phonological typology in the same way (see (10) below), while the work of Dany Jaspers (cf. i.a. Jaspers 2013, Seuren &

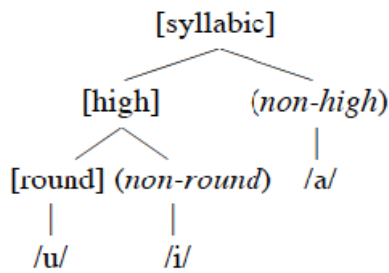
---

<sup>6</sup> The proposed learning path thus progresses from super- to subset, which might at first sight suggest a ‘superset trap’ problem. Since the supersets in play here plausibly follow from the acquirer’s initial ‘ignorance’, however, with subsets being postulated precisely because it is clear that the existing superset grammar is deficient, the classic Subset Principle reasoning does not apply here (see also Branigan 2012 on this). A superset ‘grammar’ is always defeasible by the input. Additionally, see i.a. Fodor & Sakas (2005, 2017) and Biberauer & Roberts (2009) for critical discussion of the extent to which ‘grammar size’ can meaningfully be translated into super- and subset relations.

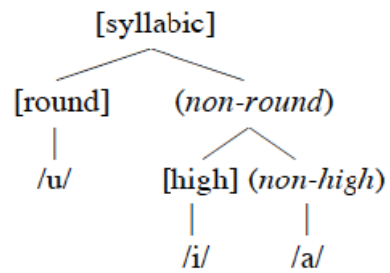
Jaspers 2014) independently postulates a NONE>ALL>SOME algorithm in the domain of logico-cognitive concept formation (see (11) below)) and also to account for human colour perception (Jaspers 2012). More generally, there is evidence from (developmental) cognitive psychology showing that object classification also seems to develop on the basis of ‘hierarchical inclusiveness’, with superordinate/more inclusive/less specified categories being acquired before subordinate/less inclusive/more specified categories (cf. i.a. Bornstein & Arterberry 2010).

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

a. [high] > [round]

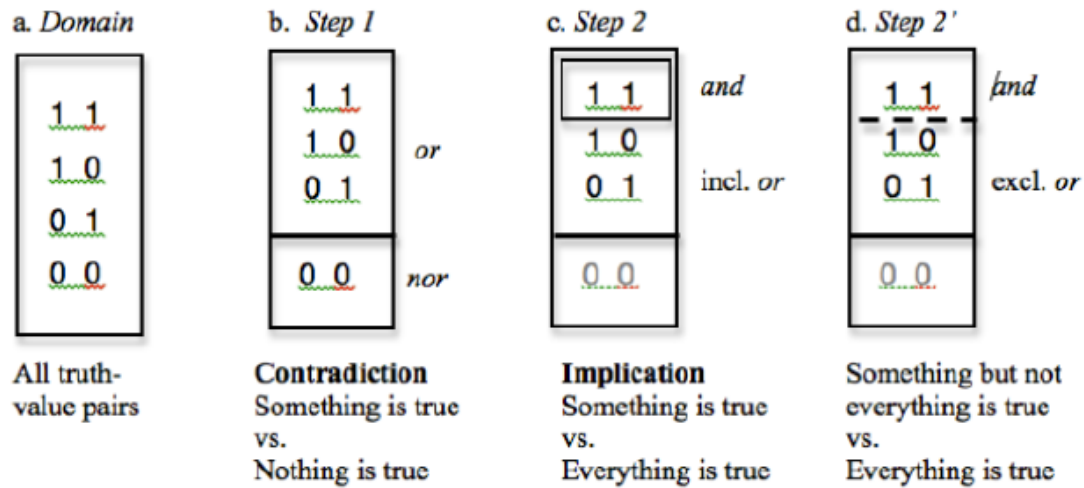


b. [round] > [high]



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

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



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

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

### 3. Novel predictions of the model

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

#### 3.1. General formal properties

##### 3.1.1. *Recycling*

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

- (12) 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 morphosyntactic regularities; and
- d. the various ways in which the earliest-acquired categories (V and N) are put to “extended” use in grammar structuring: V often acts as a reference point for focus (see recent work by Kriszta Szendrői & Fatima Hamlaoui, and Vieri Samek-Lodovici), or for the A’-domain (as in V2 systems, and Hungarian – cf. Kiss 2006, who distinguishes a “nonconfigurational” post-V zone from a configurational pre-V zone; a similar, apparently “configurationality”-distinguishing pre- and post-V zone is found in Kiowa – Adger, Harbour & Watkins 2009); the existence of extended projections (Grimshaw 1991 et seq.), typically thought to be defined by lexical categorial features (e.g. V, N, P, etc.); verbalization and nominalization, where the latter also seems to serve both a general “subordinating” function, e.g. in relation to subordination and embedding (cf. Franco 2012 for discussion and references; and Huddleston 1984:379-380 for the distinction between these two), but also for the reverse foregrounding purpose (as in VP topicalization/focus).

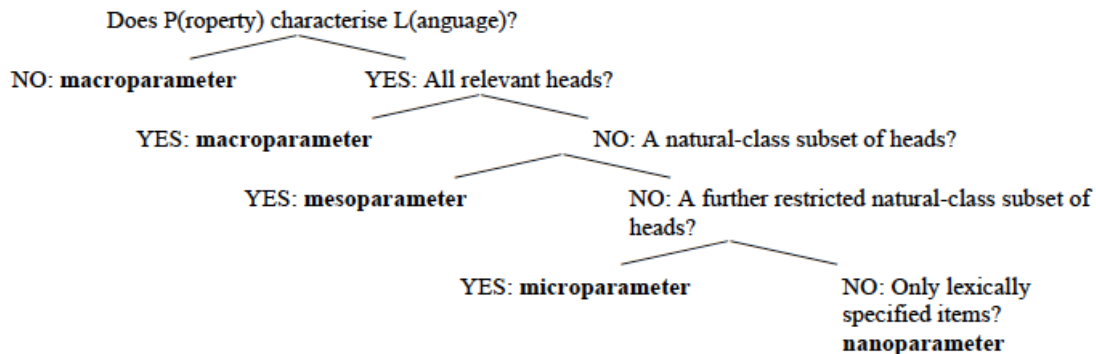
##### 3.1.2. *The shape of grammatical (parametric) variation and its connection to the course of acquisition*

The NONE>ALL>SOME learning path also leads us to expect “the same” phenomenon to surface across languages in different sized versions. (13) schematises



one way of thinking about this, with (14) attempting a rough characterization of what is at stake (cf. also Biberauer & Roberts 2016, 2017):<sup>7</sup>

(13)



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

That the types of head-final systems that can be identified crosslinguistically can be (partially) distinguished along the lines in (15) thus fits with the expectations of the model (see i.a. Cinque 2005, 2017, Biberauer 2008, Biberauer & Sheehan 2013 for discussion):

- (15) a. “rigid” head-finality: Japanese, Malayalam, etc.  
 b. clausal head-finality, nominal head-initiality, and vice versa: Chinese, Thai, Gungbe, etc.  
 c. “leaking” OV of different kinds, e.g. West Germanic  
 d. OVX, where O is the direct object (Hawkins 2009)  
 e. O<sub>[F]</sub>VX, where O<sub>[F]</sub> is a restricted object-type (e.g. Neg, Focused, Specific, etc.)

Here it is worth highlighting the SOME-options reflected in (15), i.e. the systems for which the original head-initial/-final decision did not go all in one or other direction (see Biberauer & Roberts 2017 for simplified discussion, and Biberauer 2017b for more detailed consideration). That uniformly head-initial/-final clausal or nominal structures should occur once again reflects the expectation that early-acquired V and N will play a key structuring role in natural-language grammars (cf. (12d) above). Importantly, we can, from a typological perspective, think of V and N fulfilling **parallel roles in structuring different grammars** (just as [high] and [round] did in (10) above; cf. also Wiltschko 2014 on the distinct, but formally parallel choice of one of [tense], [person] and [location] as the substantive content for INFL). More

<sup>7</sup> 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.

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. As [F]-postulation is assumed to be driven by regularities in the input (section 2.2), and as there is no innately specified learning path, there is no expectation that these [F]s will be “tested” in a sequence of any kind (*pace* the earliest parameter hierarchies proposed within the ReCoS project; cf. Biberauer, Holmberg, Roberts & Sheehan 2014 for some exemplars). Instead, a linguists’ (typologically oriented) amalgamated representation of the potential learning paths would indicate that these SOME-options are typologically equivalent (i.e. choices made at the same stage of the learning path). The possibility of thinking about typological equivalence in this in part acquisition-oriented way is a new one, which arises directly from the way the present model is constructed.

As also pointed out by Biberauer & Roberts (2012, 2016, 2017), the “size”-based parametric approach set out in (13-14) also leads to novel diachronic predictions. The expectation would, for example, be that “larger” (more macro) choices which require fewer [F]s exhibit greater stability over time. And this seems to be true: rigid head-finality, for example, seems very stable, whereas West-Germanic-style OV is far less so. Furthermore, we predict that change in the direction of “smaller” (more micro) choices will exhibit a particular character, namely one which references [F]s that are already present in the system. Again, this seems to be correct. If we consider the case of OV-loss/restriction, it does seem that what we observe is a process along the lines of (16) (Biberauer & Roberts 2008 show that OV-loss in the history of English appears to have followed the kind of “cascading” pathway sketched out in (16b,c):

(16) (simplified) schema of potential changes in the **nature of the preverbal position** in an initially “rigidly” head-final OV system:

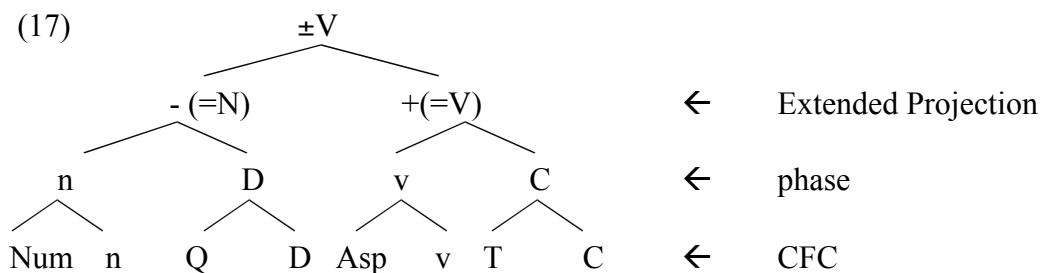
- a. all Os preverbal > all non-clausal complements (DP, PP, etc.)<sup>8</sup>
- b. all non-clausal complements (DP, PP, etc.) > all DPs (nominal objects only)
- c. all DPs (nominal objects only) > specific sub-type of DP (e.g. DP<sub>[negative]</sub>, DP<sub>[focus]</sub>, DP<sub>[topic]</sub>) > pronominal object > clitic pronominal object, etc.

Alternatively, it could also be that the OV-constraining factor is not nominal-oriented, as in (16), but clause-oriented, with the restriction referencing [tense], [aspect], [finiteness], etc.

A key feature of the NONE>ALL>SOME learning paths is that they lead us to expect natural classes constructed on the basis of “nested” featural specifications. Thinking of syntactic category formation, for example, we would expect (17) 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, 1992 Small Clause Hypothesis, Rizzi’s 1993/1994 Truncation model, the ATOM model of Schütze & Wexler 1996; see Biberauer & Roberts 2015 for discussion of (17)):

---

<sup>8</sup> 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-final nominal/D so that CPs can begin to undergo extraposition (see Biberauer & Sheehan 2012 on this).



In terms of (17), we expect acquirers to want to utilize the (in part prosodically mediated) [F] facilitating the initial V vs N distinction (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 children are confident about “basic” interrogative properties like *wh*-movement because they have grasped the workings of the auxiliary system or, indeed, all the specifics of the C-system (cf. i.a. Thornton 1995 for discussion and references relating to English), there seems to be good motivation for proposing that the (clause-typing-related) category C may define the second [ $+V$ ] category-type acquired by children. In phase-based systems (Chomsky 2001 *et seq.*), this head instantiates a phase-head, whose properties further determine the properties of T (cf. again Chomsky 2001). Similar reasoning can be applied in relation to *v* and associated non-phase heads, and to the corresponding heads in the nominal domain.

What is important for our purposes here is that the NONE>ALL>SOME learning path in (9) assumes an acquirer keen to generalize over as large a domain as possible to create formally defined domains sharing a particular property. This works against the kind of incremental upwards learning often assumed, suggesting instead that acquirers will successively postulate initially underspecified elements which can then be fleshed out to create sub-types of different kinds, each building upon the [F]s of the initially underspecified category, which, in turn, builds on that of the first category; and so on. 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 original Greenbergian correlations – V-O, Aux-V, C-TP, P-NP, and vice versa for OV – which inspired Hawkins’s (1983) postulation of ‘categorical harmony’ effects. And consider the Final-over-Final Condition (FOFC; see i.a. Biberauer, Holmberg & Roberts 2014, Sheehan 2013, Sheehan, Biberauer, Roberts & Holmberg in press/2017), stated in (18):

(18) **The Final-over-Final Condition (FOFC)**

A head-final phrase  $\alpha P$  cannot dominate a head-initial phrase  $\beta P$  where  $\alpha$  and  $\beta$  are heads in the same Extended Projection.

(cf. Biberauer, Holmberg & Roberts/BHR 2008 *et seq.*, notably BHR 2014)

What (18) requires is that head-finality start at the bottom of an Extended Projection, i.e. with V or N (see Grimshaw 1991 *et seq.*), and that once a head-final sequence has “stopped”, it cannot restart within the same EP. Contrast the structures in (19) and (20) (^ signifies head-finality in each case):

(19) Three very basic FOFC-respecting pattern:

- a. [CP C<sup>^</sup> [TP T<sup>^</sup> [VP V<sup>^</sup>]]]
  - b. [CP C [TP T<sup>^</sup> [VP V<sup>^</sup>]]]
  - c. [CP C [TP T [VP V<sup>^</sup>]]]
- > **monotonicity**: structurally adjacent heads consistently bear <sup>^</sup>

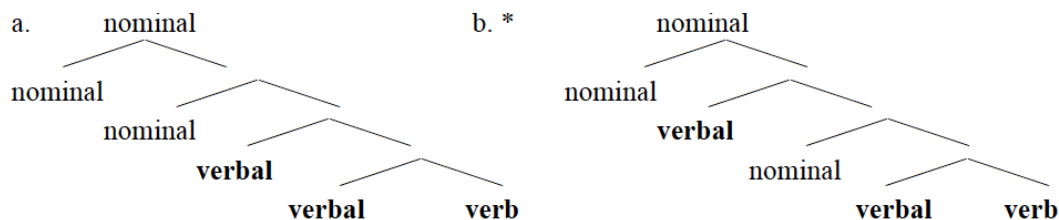
(20) Three basic FOFC-violating patterns:

- a. \*[CP C<sup>^</sup> [TP T [VP V<sup>^</sup>]]]
- b. \*[CP C<sup>^</sup> [TP T [VP V]]]
- c. \*[CP C<sup>^</sup> [TP T<sup>^</sup> [VP V]]]

> **non-monotonicity**: structurally adjacent heads vary in their <sup>^</sup>-specification; an “on-off” pattern

As noted elsewhere (Biberauer, Newton & Sheehan 2009, Biberauer, Sheehan & Newton 2010, BHR 2014, Sheehan et al. in press/2017), this requirement as diachronic implications: OV>VO changes must proceed top-down, and VO>OV changes bottom-up, which seems to be correct. Very significantly for our current purposes, however, FOFC-style monotonicity effects are not restricted to the domain of word order. Something strikingly similar emerges in relation to categorization: see Panagiotidis (2014) and references therein on so-called *Phrasal Coherence*, which is illustrated in (21)

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



And similarly, in the domain of Agreement, we see (non)-agreement “cut-off” effects exhibiting the same profile (see Biberauer 2017b for discussion). Additionally, the various hierarchies proposed by typologists and others, and the recently much-discussed \*ABA syncretism constraint instantiate further examples.

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 during the classic P&P era. These, we contend, are precisely the kinds of patterns that are best understood as the product of the kind of three-way interaction between UG, the input and MMM proposed here.

### 3.1.3. Going Beyond the Input scenarios

For Chomskyans, there is a sense in which all acquisition involves going beyond the input. Here, though, we will briefly consider only two scenarios that rather

uncontroversially involve this. One relates to artificial language learning, and the other to real-life learning.

Experimental work by i.a. Hudson Kam & Newport (2005) has revealed that ‘children learn unpredictable variation differently than adults. They have a stronger tendency to impose **systematicity** on inconsistent input ... (my emphasis; TB)’ (Hudson Kam & Newport 2005:184; see Mobbs 2015 for overview discussion). In particular, while adults demonstrate frequency-matching, approximately replicating the variability in the original input, child acquirers employ regularization strategies. The nature of these strategies is of particular interest here. Consider (22) in this connection:

- (22) The types of regularization that children impose on the input:
- a. **minimization**: use the variable form none of the time (NONE)
  - b. **maximization**: use the variable form all the time<sup>6</sup> (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 (22c) was the most rarely used strategy; nevertheless, child acquirers appear to appeal MMM-driven regularization strategies of the kind assumed in this model.

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

- (23) 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 (SOME).

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

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

- (24) a. generalization throughout the paradigm, either to all *was* or all *were* (ALL/NONE)
- b. specialization relative to **polarity**: *were* (i.e. *weren't*) in negative clauses, regardless of person and number, with *was* occurring in affirmative clauses, regardless of number. (SOME)

- (25) a. They **was** writing a lot of tests that time.
- b. He **weren't** doing much else.

The grammatically defined SOME-choices that emerge in the past tense thus centre on [polarity]. The question is why? A highly plausible conditioning factor here would be the evidence that acquirers get from interrogative structures that auxiliaries are fundamentally concerned with polarity. Consider (26) in this regard:

- (26) a. They **were** all picnicking in the sunshine.  
b. **Were** they all picnicking in the sunshine?  
c. They **ate** a lot of cake.  
d. **Did** they all eat a lot of cake?

Here we see a very fundamental declarative-interrogative contrast in respect of auxiliary positioning (cf. (6e) above) and realization (cf. (6b) 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 Roeper & Woods 2016 for recent discussion and references). [Polarity] then seems to be an early-acquired [F], which, in the context of our model, would therefore be expected to serve as the basis for input structuring in cases where the input is in some way compromised.

#### 4. Conclusions

Our objective here has been to introduce a neo-emergentist model of language acquisition, variation, and change that, like its classic P&P predecessor, seeks to understand language variation (and change) as a reflex of the way in which language is acquired. Where the explanatory burden previously rested largely on UG and its hypothetically rich parametric content, we have instead considered how parametrically shaped adult grammars might arise in the absence of a UG-given parametric endowment. Each of the three factors in Chomsky's (2005) model were ascribed a role in the context of the model presented here, with the general cognitive factor, Maximise Minimal Means, being argued to 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):<sup>9</sup>

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**

---

<sup>9</sup> Thanks to Itziar Laka for drawing attention to this extract during a recent generative linguistics event in Reading.

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.

## References

- Abler, W. (1989). On the particulate principle of self-diversifying systems. *Journal of Social Biological Structure* 12:1-13.
- Adger, D, D. Harbour & L. Watkins (2009). *Mirrors and Microparameters: Phrase Structure beyond Free Word Order*. Cambridge: CUP.
- Ambridge, B., J. Pine & E. Lieven (2013). Child language acquisition: Why Universal Grammar doesn't help. *Language* 90(3): e53-e90.
- Biberauer, T. (2008). Introduction. In: T. Biberauer (ed.). *The Limits of Syntactic Variation*. Amsterdam: Benjamins, 1-72.
- Biberauer, T. (2011). In defence of lexico-centric parametric variation: two 3<sup>rd</sup> factor-constrained case studies. Paper presented at the Workshop on Formal Grammar and Syntactic Variation: Rethinking Parameters (Madrid).
- Biberauer, T. (2015). Learning from questions and commands: probing the nature and origins of native-speaker knowledge. Cambridge Humanities Research Grant Proposal.
- Biberauer, T. (2016). Going beyond the input (and UG): an emergentist generative perspective on syntactic variation, stability and change'. Invited talk at the Language Contact, Continuity and Change in the Emergence of Modern Hebrew conference (Jerusalem)
- Biberauer, T. (2017a). Particles and the Final-over-Final Constraint. To appear in: M. Sheehan, T. Biberauer, A. Holmberg & I. Roberts (eds). *The Final-over-Final Condition*. Cambridge, MA: MIT Press.
- Biberauer (2017b). Maximising Minimal Means: typological, acquisitional and diachronic perspectives. Invited Lecture Series at the Center for Research in Syntax, Semantics and Phonology (CriSSP) (Brussels)
- Biberauer, T. (2017c). Learning from commands in contact situations: some Southern African case studies. Talk given at the Stellenbosch Linguistic Research Seminar (Stellenbosch, 5 April 2017)
- Biberauer, T., J. Bockmühl & S. Shah (2017). Learning from commands: the case of Afrikaans and Namibian German. Paper presented at the 11th Creolistics Workshop (Giessen, 23 March 2017)
- Biberauer, T., A. Holmberg & I. Roberts (2008). Disharmonic word-order systems and the Final-over-Final-Constraint (FOFC). In Bisetto, A. & F. Barbieri (eds). *Proceedings of XXXIII Incontro di Grammatica Generativa*.
- [Available on-line at:  
[http://amsacta.cib.unibo.it/archive/00002397/01/PROCEEDINGS\\_IGG33.pdf](http://amsacta.cib.unibo.it/archive/00002397/01/PROCEEDINGS_IGG33.pdf)]
- Biberauer, T., A. Holmberg & I. Roberts (2014). A syntactic universal and its consequences. *Linguistic Inquiry* 45(2): 169-225.
- Biberauer, T., A. Holmberg, I. Roberts & M. Sheehan (2014). Complexity in comparative syntax: the view from modern parametric theory. In: F. Newmeyer & L. Preston (eds.). *Measuring Linguistic Complexity*. Oxford: Oxford University Press, 103-127.
- Biberauer, T., G. Newton & M. Sheehan (2009). Limiting synchronic and diachronic variation and change: the Final-Over-Final Constraint. *Language and Linguistics* 10(4): 699-741.
- Biberauer, T. & I. Roberts (2008). Cascading parameter changes: internally driven change in Middle and Early Modern English. In: T. Eythórsson (ed). *Grammatical Change and Linguistic Theory: The Rosendal Papers*. Amsterdam: Benjamins, 79-113.
- Biberauer, T. & I. Roberts (2009). The Return of the Subset Principle. In: P. Crisma & G. Longobardi (eds.). *Historical Syntax and Linguistic Theory*. Oxford: OUP, 58-74.
- Biberauer, T. & I. Roberts (2012). On the significance of what hasn't happened. Paper presented at the DiGS 14 Conference (Lisbon)
- Biberauer, T. & I. Roberts (2015). Rethinking formal hierarchies: a proposed unification. In: J. Chancharu, X. Hu & M. Mitrović (eds.). *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*. Amsterdam: Benjamins, 259-291.
- Biberauer, T. & I. Roberts (2017). Parameter setting. In: A. Ledgeway & I. Roberts (eds). *The Cambridge Handbook of Historical Syntax*. Cambridge: Cambridge University Press, 134-162.
- Biberauer, T. & M. Sheehan. (2012). Disharmony, antisymmetry, and the Final-over-Final Constraint. In: V. Elguea & M. Etzebarria (eds.). *Ways of Structure Building*. Oxford: Oxford University Press, 206-244.
- Biberauer, T. & M. Sheehan (2013). Introduction. In: T. Biberauer & M. Sheehan (eds.). *Theoretical Approaches to Disharmonic Word Order*. Oxford: Oxford University Press, 1-46.
- Biberauer, T., M. Sheehan & G. Newton. (2010). Impossible Changes and Impossible Borrowings: the Final-over-Final Constraint. In: A. Breitbarth, C. Lucas, S. Watts & D. Willis (eds.). *Continuity and Change in Grammar*. Amsterdam: John Benjamins, 35-60.
- Bond, O., G. Corbett, M. Chumakina & D. Brown (2016). *Archi: Complexities of Agreement in Cross-theoretical Perspective*. Oxford: OUP.
- Bornstein, M. & M. Arterberry (2010). The development of object categorization in young children: Hierarchical inclusiveness, age, perceptual attribute and group versus individual analyses. *Developmental Psychology* 46: 350-365.
- Branigan, P. (2012). Macroparameter learnability: an Algonquian Case Study. Unpublished ms: Memorial University of Newfoundland.
- Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, N. (2000). Minimalist inquiries: the framework. In: R. Martin, D. Michaels & J. Uriagereka (eds). *Step by step: essays on minimalist syntax in honor of Howard Lasnik*. Cambridge, MA: MIT Press, 89-156.
- Chomsky, N. (2001). Derivation by phase. In: M. Kenstowicz (ed.). *Ken Hale: a life in language*. Cambridge, MA.: MIT Press, 1-50.
- Chomsky, N. (2005). Three factors in Language Design. *Linguistic Inquiry* 36: 1-22.
- Chung, S. (2012). Are lexical categories universal? The view from Chamorro. *Theoretical Linguistics* 38:1-56.
- Cinque, G. (2005). A note on Verb/Object order and Head/Relative clause order. *University of Venice Working Papers in Linguistics* 15: 49-104.
- Cinque, G. (2017). A microparametric approach to the head-initial/head-final parameter. To appear in *Linguistic Analysis*.
- D'Alessandro R. & M. van Oostendorp (2016). Gravitational Grammar. Talk given in Stuttgart.
- 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*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 119-134.
- Demuth, K. (2003). The acquisition of Bantu languages. In: D. Nurse & G. Philippson (eds). *The Bantu Languages*. Surrey: Curzon Press, 209-222.
- Dresher, E. (2009). *The Contrastive Hierarchy in Phonology*. Cambridge: CUP.
- Dresher, E. (2014). The arch not the stones: universal feature theory without universal features. In: M. Krämer, S. Ronai and P. Svenonius (eds). *Nordlyd* 41(2): 165-181.
- Duffield, N. (2013). Minimalism and semantic syntax: interpreting multifunctionality in Vietnamese. Unpublished ms: University of Konan ([LingBuzz/001919](http://LingBuzz/001919); last accessed 21 September 2015)
- Fasanella, A. (2014). *On how Learning Mechanisms shape Natural Languages*. Ph.D. dissertation: UAB. ([filcat.uab.cat/clt/publicacions/tesis/pdf/fasanella\\_14.pdf](http://filcat.uab.cat/clt/publicacions/tesis/pdf/fasanella_14.pdf))
- Fasanella, A. & J. Fortuny (2016). Deriving linguistic variation from learnability conditions: the Chunking Procedure. In: L. Eguren, O. Fernandez-Soriano & A. Mendikoetxea (2016). *Rethinking Parameters*. Oxford: Oxford University Press,.
- Fodor, J. & W. Sakas (2005). The Subset Principle in syntax: costs of compliance. *Journal of Linguistics* 41: 513-569
- Fodor, J.D. & W.G. Sakas (2017). Learnability. In: I. Roberts (ed.) *The Oxford Handbook of Universal Grammar*. Oxford: OUP, 249-269.
- Franco, L. (2012). Against the identity of complementizers and (demonstrative) pronouns. In: K. Dziubalska-Kolaczyk (ed.). *Poznan Studies in Contemporary Linguistics* 48: 565-596
- Gervain, J. & J. Mehler (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology* 61: 191-218.
- Grimshaw, J. (1991). Extended Projection. Unpublished ms: Brandeis.



- Haspelmath, M. (2010). Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3): 663-687.
- Hawkins, J. (1983). *Word Order Universals*. New York NY: Academic Press.
- Hawkins, J. (2009). An asymmetry between VO and OV languages: the ordering of obliques. In: G. Corbett & M. Noonan (eds). *Case and Grammatical Relations: Studies in Honor of Bernard Comrie*. Amsterdam: Benjamins, 167-190.
- Hockett, C. (1958). *A Course in Modern Linguistics*. New York: Macmillan Company.
- Huddleston, G. (1984). *Introduction to the Grammar of English*. Cambridge: CUP.
- Hudson Kam, C. & E. Newport (2005). Regularizing unpredictable variation: the roles of adult and child learners in language formation and change. *Language Learning and Development* 1: 151-95.
- Humboldt, W. (1836). *Über die Verschiedenheit des menschlichen Sprachbaues*. Paderborn: Verlag Ferdinand Schöningh.
- Jaspers, D. (2012). Logic and colour. *Logica Universalis* 6: 227-248.
- Jaspers, D. (2013). Constraints on Concept Formation. Poster presented at GLOW 36 (Lund).
- van Kampen, J. (2004). Learnability order in the French pronominal system. In: R. Bok-Bennema, B. Hollebrandse, B. Kampers-Manhe & P. Sleeman (eds). *Romance Languages and Linguistic Theory 2002: Selected Papers from 'Going Romance', Groningen 28-30 November 2002*. Amsterdam: Benjamins, 163-182.
- Lleó, C. (1998). Proto-articles in the acquisition of Spanish: Interface between Phonology and Morphology. In: R. Fabri, A. Ortmann & T. Parodi (eds). *Modelle der Flexion: 18. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft*. Tübingen: Niemeyer.
- Lleó, C. (2001). Early fillers: undoubtedly more than phonological stuffing. *Journal of Child Language* 28: 262-265.
- Lleó, C. & K. Demuth (1999). Prosodic constraints on the emergence of grammatical morphemes: crosslinguistic evidence from Germanic and Romance languages. In: A. Greenhill, H. Littlefield & C. Tano (eds). *Proceedings of the 23<sup>rd</sup> Annual Boston University Conference on Language Development (BUCLD 23)*, volume 2: 407-418. Somerville, MA: Cascadilla Press.
- Mobbs, I. (2015). *Minimalism and the Design of the Language Faculty*. Ph.D. dissertation: University of Cambridge.
- Panagiotidis, P. (2014). *Categorial Features: a generative theory of word class categories*. Cambridge: CUP.
- Pesetsky, D. (2013). ["Что дѣлать? What is to be done?"](#) Plenary talk at the LSA annual conference.
- Pinker, S. (1984). *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.
- Radford, A. (1990). *Syntactic Theory and the Acquisition of English Syntax: The Nature of Early Child Grammars of English*. Oxford: Blackwell.
- Richards, N. (2010). *Uttering Trees*. Cambridge: MIT Press.
- Ritter, E. & M. Wiltschko (2009). Varieties of INFL: TENSE, LOCATION and PERSON. In: J. van Craenenbroeck (ed.). *Alternatives to cartography*. Berlin: Mouton de Gruyter, 153-201
- Ritter, E. & M. Wiltschko (2014). The composition of INFL. *Natural Language and Linguistic Theory* 32(4): 1331-1386.
- Rizzi, L. (1993/4) Some notes on linguistic theory and language development: the case of Root Infinitives. *Language Acquisition* 3: 341-393
- Roberts, I. (2007). *Diachronic Syntax*. Oxford: OUP.
- Roberts, I. & A. Roussou (2003). *Syntactic Change. A Minimalist Approach to Grammaticalization*. Cambridge: CUP.
- Roeper, T. (2011). The acquisition of recursion: how formalism articulates the acquisition path. *Biolinguistics* 5: 57-86.
- Roeper, T. & W. Snyder (2002). Recursion as an analytic device in acquisition. In: J. van Kampen & P. Coopmans (eds). *Proceedings of GALA 2003*. Utrecht: LOT, 158-168.
- Roeper, T., W. Snyder & K. Hiramatsu (2002). Learnability in a minimalist framework: root compounds, merger, and the syntax-morphology interface. In: I. Lasser (ed.). *The Process of Language Acquisition*. Frankfurt: Peter Lang.

- Roeper, T. & R. Woods (2016). Separating Tense and Assertion: Evidence from embedded V2 and Child Language. Paper presented at the DGfS Workshop Verb Second in Grammar and Processing: its Causes and Consequences (DGfS, Konstanz)
- Schütze, C. T. and K. Wexler (1996) Subject case licensing and English root infinitives. In: A. Stringfellow et al (eds.). *BUCLD 20*, Cascadilla Press, 670-681.
- Seuren, P. & Jaspers, D. (2014). Logico-cognitive structure in the lexicon. *Language* 90: 607-643.
- Sheehan, M. (2013). Explaining the Final-over-Final Constraint: formal and functional approaches. In: T. Biberauer & M. Sheehan (eds). *Theoretical Approaches to Disharmonic Word Order*. Oxford: OUP, 407-444.
- Sheehan, M., T. Biberauer, A. Holmberg & I. Roberts (in press/2017). *The Final-over-Final Condition. A Syntactic Universal*. Cambridge, MA: MIT Press.
- Shi, R. & J. Werker (2001). Six-month-old infants' preference for lexical words. *Psychological Science* 12(1): 70-75.
- Shi, R., J. Werker, & J. Morgan (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition* 72(2): B11-B21.
- Tsimpli, I. (2014). Early, late or very late? Timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism* 4(3): 283-313.
- Willis, D. (2016). Exaptation and degrammaticalization within an acquisition-based model of abductive reanalysis. In: M. Norde & F. van de Velde (eds). *Exaptation in Language Change*. Amsterdam: Benjamins.
- Wiltschko, M. (2014). *The Universal Structure of Categories. Towards a Formal Typology*. Cambridge: CUP.
- Thornton, R. (1995). Referentiality and *wh*-movement in child English: Juvenile D-linkuency. *Language Acquisition* 4: 139-175.
- Yang, C. (2016). *Price of Productivity. How Children Learn and Break Rules of Language*. To appear with MIT Press.
- Zeijlstra, H. (2008). On the syntactic flexibility of formal features. In: T. Biberauer, (ed.). *The Limits of Syntactic Variation*. Amsterdam: Benjamins, 143-174.