

Two Kinds of Variation in a Minimalist System

Marc Richards*

Abstract

This paper reconsiders the status of parameters and parametric variation from the perspective of recent developments within the minimalist program, in particular the problem that arises if parameters can no longer be part of UG, which must be maximally empty, and if variation is instead to be explained in terms of third-factor considerations (section 1). I pursue the idea that variation arises precisely because UG is maximally underspecified, thus leaving many options open (Biberauer & Richards 2006, Berwick & Chomsky 2008, Boeckx 2008). Adopting the asymmetric view of language design put forward in Chomsky 2005a, 2006 and Berwick & Chomsky 2008, whereby language is optimally designed only for satisfying conditions imposed by the semantic interface, I identify two kinds of variation corresponding to the two domains in which indeterminacies may arise in a minimalist architecture with a minimally specified UG: (i) within the narrow syntax or (ii) at the phonological (sensorimotor) interface ('externalization', the mapping to PF). In (i), the domain of the Strong Minimalist Thesis (SMT) and thus of a parameter-free UG, free variation is predicted to occur, with each 'competing option' a possible choice in every derivation (section 3). In (ii), to which the SMT does not apply, competing options are resolved consistently in a language through parametric choices, yielding macroparametric variation (directionality, polysynthesis) at the PF-interface (section 2).

1. Introduction: Parametric variation without UG

According to the biolinguistic perspective of Chomsky 2005a,b, 2006, three factors must be involved in the development of language in the individual, as given in (1).

*Various portions of this material, with varying emphases, have been presented at the Workshop on Universal Grammar, Language Acquisition and Change, University of Oslo (August 2008), the Workshop on Perspectives of Minimalist Syntax, University of Leipzig (October 2008), and at the Syntax & Morphology Colloquium, University of Leipzig (November 2008). I would like to thank those audiences for their many helpful comments and suggestions. Particular thanks are due to Petr Biskup, Gisbert Fanselow, Elly van Gelderen, Günther Grewendorf, Fabian Heck, Greg Kobele, David Lightfoot, Terje Lohndal, Gereon Müller, Patrick Schulz, Peter Svenonius, and Jochen Trommer.

(1) *Three factors in language design (Chomsky 2005b):*

- I. Genetic endowment
- II. Experience
- III. Principles not specific to the FL, the human faculty of language.

Factor I is the domain of Universal Grammar (UG); Factor II is the external data (E-language) that constitutes the linguistic environment in which language acquisition takes place; and Factor III comprises “general properties of organic systems” (Chomsky 2004: 1), the result of physical constraints on the form and development of living organisms. In the case of FL, a biological organ like any other (a ‘mental’ organ), such third-factor constraints might include principles of efficient computation and the interface conditions imposed from outside FL by the semantic (SEM) and phonological (PHON) systems with which it interacts. Factor III is what distinguishes minimalism from other approaches to FL, offering a different benchmark for what counts as a genuine explanation (taking us “beyond explanatory adequacy”, in Chomsky’s words). Accordingly, minimalism is characterized by a trend away from Factor I: UG must be as small, simple and empty as possible, on evolutionary grounds – not least, FL arose too quickly, suddenly and recently in evolutionary terms (perhaps around 60,000 years ago – Chomsky 2005a) for there to have been the historical ‘space’ for the development of a richly specified UG, replete with dozens of FL- and species-specific rules, principles, constraints and so forth. The more we ascribe to UG, the harder the question of its evolution becomes, i.e. the question of how it all got there (cf. Chomsky 2006:2-3). Hence the primary role that Factor III must have played in shaping FL. A highly structured UG must therefore be replaced by one with as few FL-specific principles as possible.

A rather less noted consequence of this approach is that what is true of principles must also be true of parameters. That is, not only must a maximally underspecified UG be devoid of language-specific principles (as far as possible), but also of parameters. This would appear to lead to a rather startling conclusion from the perspective of traditional principles-and-parameters theory: the range of variation across the world’s languages can no longer be taken to be part of the universal, genetic specification. Variation, and the forms it takes, is no longer determined by UG. How, then, are we to explain and accommodate language variation in a system based on third-factor explanation (i.e. minimalism)? If variation cannot be ascribed to UG, then where are we to locate it? Not only does variation lack a locus, but it would also now seem to lack a rationale given the uniformity of Factor I (see section 1.1) and the invariance of Factor III (that is, third-

factor constraints cannot be parametrized - section 1.2).¹ Let us briefly consider each of these in turn.

1.1. Uniformity of Factor I (UG)

A central tenet throughout the development of the minimalist program (MP) is that narrow syntax ('the computation of LF') is uniform. The quote in (2) is an explicit statement to this effect.

- (2) *Uniformity of Factor I (Chomsky 2001: 2)*
 "[A]ssume languages to be uniform, with variety restricted to easily detectable properties of utterances."

The MP thus abandons a direct parametrization of UG itself (Factor I) – i.e. the parametrized principles of the Government & Binding instantiation of Principles-and-Parameters Theory – in favour of a uniform FL, with crosslinguistic variation restricted to the lexicon (i.e. that which must in any case be learned), namely to the featural properties of lexical items. In particular, under the "Borer-Chomsky Conjecture" (as recently coined by Baker (2008)), variation is restricted to the properties and features of functional categories:

- (3) *Borer-Chomsky Conjecture (BCC; 'lexical parameters')*
- a. "Parametric variation is restricted to the lexicon, and insofar as syntactic computation is concerned, to a narrow category of morphological properties, primarily inflectional." (Chomsky 2001: 2)
 - b. "The availability of variation [is restricted] to the possibilities which are offered by one single component: the inflectional component." (Borer 1984: 3)

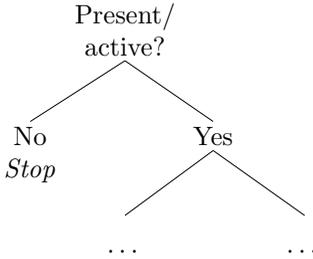
A simple illustration of this approach in action in 'classical' minimalism would be the distinction between overt and covert movement, which was determined by strong versus weak categorial features on functional heads. Thus the parametric difference between French and English in terms of finite verb movement was captured by ascribing a strong V-feature to French

¹Note that Factor II cannot be the locus of variation. Rather, Factor II is the *trigger* for variation, with different final states being acquired depending on the linguistic environment to which the child is exposed. Factor II provides the trigger experience (the 'primary linguistic data', PLD); but it cannot be the locus of variation per se, since Factor II is language- and organism-external (E-language, not I-language).

I/T, triggering overt V-to-I movement, and the weak version of this feature to English I/T. Such parameters, affecting the properties of lexical items, have come to be known as *microparameters*, since the points of variation that they instantiate are largely independent of each other, applying to individual formal features and yielding small-scale variation with only limited clustering effects.²

Such microparametric variation arises under the BCC through the interaction between Factors I and II in the process of language acquisition. According to Chomsky (2000: 100-101), Factor I (UG) provides a universal inventory of formal features **F** from which each language (learner) makes a one-time selection of a subset [F]. This subset then defines the range of features that play an active role in the operations of the language in question. If the universal inventory **F** defines the range of features that a child can postulate in constructing its lexicon, then variation in the actual features (and perhaps projections) present in a given language is an automatic consequence of a conservative learning algorithm of the kind postulated in such works as Thráinsson (1996), Bobaljik & Thráinsson (1998), and Koenenman & Neeleman (2001), whereby only those functional categories are postulated for which there is positive evidence in the input (cf. Thráinsson 1996:261, and Longobardi's (2001: 294) "Minimize feature content"). We can model this as the first step on a parameter schema of the kind familiar from such works as Uriagereka 1995, Roberts & Roussou 2003, Holmberg & Roberts 2008:

²This is not to say that clustering isn't possible on a microparametric approach to variation. Thus, for example, the effects of a large-scale macroparameter such as the Null Subject parameter may be reducible to a single lexical property on T, as in Holmberg & Roberts's (2008) formulation: "Does T bear a D-feature?". Bearing a D-feature is enough to derive all the syntactic consequences identified by Rizzi (1982): the D-feature is expressed morphologically as rich agreement and removes the need for EPP-movement of a subject DP, allowing the subject to stay low (yielding subject inversion and, from the low position, *that*-trace-circumventing subject-movement), etc.

(4) *Partial parameter schema for feature F*

Whenever evidence for a feature F specified in \mathbf{F} (Factor I) is encountered in the linguistic input (Factor II), then the tree in (4) is activated and the ‘Yes’ branch pursued, whereupon further options then present themselves. These include: is F morphologically expressed?, does F Agree?, does F trigger movement?, how much structure does F pied-pipe? (see (21) in section 3, below).

In short, the assumption of uniformity of FL (Factor I), as expressed in (2) and the BCC, is inherently associated with variation of the lexical, microparametric kind. Such variation therefore presents no problem from the third-factor perspective of a maximally empty UG, as long as we assume that UG at least comprises the feature inventory \mathbf{F} (see next section). However, other kinds of variation remain problematic, for the reasons given in the previous section. In particular, the status of (nonlexical) macroparameters, the kinds of parameters which are responsible for the major typological divisions of the world’s languages and which were formerly conceived of as parametrized principles, is dubious. If Factor I can neither contain parameters as additional, FL-specific objects, nor admit of parametric options for whatever machinery it does contain (due to (2)), then macroparameters would still appear to lack both a locus and a rationale (cf. section 1 above; also Baker 1996, Richards 2004: chapter five).

This leaves us with one remaining possibility – since they cannot be Factor I, perhaps (macro)parameters can be ascribed to Factor III? As argued in the following section, however, a parametrization of Factor III principles is also excluded.

1.2. Uniformity of Factor III

Chomsky (2005, 2006) defines minimalist enquiry as the pursuit of third-factor (“principled”) explanations for linguistic phenomena and properties

of the language faculty. The goal of the MP is then to move descriptive technology from Factor I (the genetic endowment, UG) to Factor III by showing that that technology is dispensable, or reducible to third-factor effects. A guiding hypothesis that we can entertain in order to pursue this aim is the Strong Minimalist Thesis (SMT), which states that no aspect of FL is without a principled, third-factor explanation – specifically, we entertain the thesis that *all* properties of FL contribute to a computationally efficient satisfaction of interface conditions (IC). In that sense (i.e. if SMT were true), then FL would be a “perfect” solution to IC. Such a perfect solution would comprise an empty UG: logically, if everything is Factor III, then nothing is Factor I. Clearly, this is too strong a hypothesis – UG cannot be completely empty, otherwise there would be no FL (even in the ‘narrow’ sense of Hauser, Chomsky & Fitch 2002). The genetic endowment UG, then, should be maximally (but not completely) empty, consisting of a minimal unexplained residue. The question then arises as to what that minimal residue must be in order to account for the human capacity. In Chomsky’s words:

- (5) “Throughout the modern history of generative grammar, the problem of determining the character of FL has been approached ‘from top down’: How much must be attributed to UG to account for language acquisition? The MP seeks to approach the problem ‘from bottom up’: How little can be attributed to UG while accounting for the variety of I-languages attained, relying on third-factor principles?” (Chomsky 2006: 3)

Most of the work, then, is to be done by Factor III. The MP thus turns traditional linguistic inquiry on its head, focussing on that which the study of language has traditionally abstracted away from, namely those properties of language which are *not* unique to this faculty.

What, then, must a maximally empty UG minimally comprise? What is the indispensable, irreducible residue that makes up Factor I? As we have already discussed, a maximally empty UG cannot contain FL-specific principles and parameters (or rules, or constraints, etc.). One possibility that I would like to pursue here is that it contains only features, i.e. the inventory **F** of the previous section. Let us assume that there are just two fundamental feature types provided by UG on which the narrow syntax operates (i.e. formal features): firstly, there must be features for building structured expressions; and secondly, there must be features for connecting those expressions with the external systems of sound and meaning. This would seem the conceptual minimum for a system that generates expressions that pair sounds with meanings. The former, structure-building features are Chom-

sky's (2005, 2006) Edge Features (EF); the latter are the uninterpretable features (uF), which trigger Transfer to the interfaces for reasons of Full Interpretation (see next section, and Chomsky 2006).

These two feature types would seem sufficient to yield the full range of operations postulated in minimalist grammars and analyses. EF, on this view, would have been the evolutionary innovation taking us from the prelinguistic stage of inert, isolated lexical items (concepts) to allowing those items to be combined together into larger objects. That is, EF enables external Merge (and thus recursion; cf. Hauser, Chomsky & Fitch 2002) and, concomitantly, Move (qua internal Merge). Uninterpretable features yield Agree insofar as they lack values and must seek these in order to be deleted for Full Interpretation; for the same reason, they also trigger Transfer upon valuation (see next section), thus yielding a point of Transfer every time a uF is valued – that is, they yield phases. Thus from EF and uF, the minimal components of UG, we obtain Merge, Move, Agree and Transfer (spell-out), the principal mechanisms of FL.³

As we suggested at the end of the previous section, as long as UG contains features, microparametric variation (of the lexical kind) is possible. If uFs and EFs are provided by UG, then it is the job of the language learner to assemble these features into lexical items (see Richards 2008a for a proposal as to how uF-EF associations may differ across languages, deriving the difference between optional and obligatory EPP effects). The questions associated with parameter schemata like (4) are readily captured in these terms: thus, “does F trigger Agree?” becomes “is F a uF?”; “does F trigger Move?” becomes “does F bear (undeletable) EF?”. We return to this in section 3.1 below. For now, the main point is that, as far as lexical, microparameters are concerned, i.e. those conforming to the BCC, it is a case of ‘business as usual’ (contra Boeckx 2008): a third-factor-based system still admits of this kind of variation.

On the other hand, macroparameters, in the sense of parametrized principles, become even less tenable in such a system. Unlike the UG principles of the GB era, the principles that shape a minimalist syntax, i.e. third-factor principles, are unlikely candidates for variability. Inclusiveness, No Tampering, Phase Impenetrability and the like are invariant, language-external and language-independent principles of efficient computation that hold of FL, forming a single design solution (thus all languages satisfy IC equally well,

³Other minimalist principles, such as Inclusiveness, No Tampering, Phase Impenetrability, etc., are third-factor principles of efficient computation and thus not part of UG.

as instantiations of the single optimally designed FL). Boeckx 2008 calls this the “Strong Uniformity Thesis” (SUT) – essentially, it is not just Factor I that is invariant (cf. (2)), but also Factor III:

(6) *Uniformity of Factor III*

“There is simply no way for principles of efficient computation to be parametrized. [...]t strikes me as implausible to entertain the possibility that a principle like ‘Shortest Move’ could be active in some languages, but not in others. Put differently, [...] there can be no parameters within the statements of the general principles that shape natural language syntax. In other words, narrow syntax solves interface design specifications optimally in the same way in all languages.” (Boeckx 2008:5)

In sum, we have a uniform, invariant UG (Factor I), and a uniform, invariant optimal mapping to the interfaces (Factor III). SUT and SMT together would seem to leave no room for variation (beyond the lexical, microparametric variation reviewed above). In the remainder of this paper, I would like to explore and qualify this position somewhat, as outlined below.

Section 2 shows how a syntactic universal, namely the core sequence of functional categories C-T-v-(V), is amenable to parametric variation once it is reduced to a third-factor effect, thus illustrating how Factor III interacts with Factors I and II to yield crosslinguistic variations on a universal theme.

Section 3 argues that there is room for variation in the way that ICs are satisfied on the sensorimotor (‘PF’) side, so that the latter provides the previously elusive minimalist locus for macroparameters (cf. Richards 2004: Chapter 5). This possibility arises on account of recent speculations by Chomsky (2005a, 2006) and Berwick & Chomsky (2008) that the SMT holds only of the mapping to the semantic interface, SEM. That is, FL is optimally designed for the mapping to SEM, but not for the mapping to PF. Parametric variation is thus arguably to be expected as part of the imperfect satisfaction of ICs in the PF component, for which the syntax is not optimally designed. Section 3.1 illustrates this on the basis of the linearization of a symmetrical syntax (following Richards 2004), showing how the consistent resolution of competing linearization options at PF might yield the effect of a directionality macroparameter in such a system.

Section 4 turns to the narrow syntax, where variation is also expected to arise in a third-factor-based system, albeit in a different form. Due to the maximally underspecified nature of UG, inherent points of optionality should emerge wherever UG lacks specified instructions for their resolution, such as the manner in which a given feature is satisfied or lexicalized. Unlike the PF system, the narrow syntax is subject to the SMT, and so

these choice points are simply left open, as to specify or define them one way or another (e.g. by means of parameters, yielding consistent choices) would be to add to UG, contra the SMT (and SUT). Instead, the grammar (UG) simply shouldn't care (cf. Biberauer & Richards 2006, Boeckx 2008:13, Holmberg & Roberts 2008: 67). The prediction, then, is that unlike at PF, variation in the syntax is left unresolved, yielding true optionality – that is, either of the two competing options is a possible choice in every derivation. Variability, in the form of competing options, thus emerges from the SMT as a third-factor effect, indeterminacies being the direct result of a minimally specified, maximally empty UG. Section 4.1 illustrates this on the basis of optional pied-piping in Russian *wh*-movement (following Biberauer & Richards 2006), highlighting the role that such competing options might play in providing the requisite variable trigger experience (Factor II) for the resetting of lexical microparameters (language change).

2. Variation and universals: Interaction between Factors I, II and III

A well-established candidate for a syntactic universal is the core functional sequence C-T-v-(V), known as the Core Functional Categories (CFCs). The CFCs C-T-v would seem to constitute the minimal 'backbone' of the clause, always present in every language, including at the earliest stages of acquisition (cf. Poeppel & Wexler 1993). As an apparent universal, the CFC sequence is something which we might want to attribute to UG; however, to do so would be a departure from the SMT, on the grounds reviewed in the previous section (it would increase the size of Factor I). The SMT thus prompts us to seek a third-factor explanation of this universal property. At the same time, it is also well established that there is crosslinguistic variation in the elaboration of this core functional sequence, with each CFC defining a region of the clause that may be 'expanded' in certain languages (such as Rizzi's (1997) articulated left periphery), in ways identified in 'cartographic' studies. Thus we are in the following situation: we must provide a third-factor explanation for a universal property of FL, thus allowing us to remove it from UG; but at the same time this third-factor property must admit of microparametric variation of the kind described in section 1.2 (that is, the projection of additional heads where there is positive evidence for such in the input, as modelled in (4)).

Let us begin with the first of these two tasks. For Chomsky (2005a, 2006), a property receives a 'principled' (i.e. third-factor) explanation insofar as it follows from the SMT, i.e. contributes to the efficient satisfaction of ICs. For

the purposes of this paper, I will assume that there are just two relevant ICs from the SMT perspective (i.e. those that shape the syntax), namely Full Interpretation (FI) holding at each interface. (This would also appear to be Chomsky's (2006: 13) line of thinking, since he crucially refers there to "both interface conditions", implying just two of them.) FI requires that there be no redundant, i.e. uninterpreted, symbols at either interface, in line with our fundamental assumption, the SMT, which holds that the syntax affords a perfect, nonredundant mapping to the interfaces (or at least to SEM, if Chomsky 2006, Berwick & Chomsky 2008 are correct). Thus, adopting the SMT, we assume that the syntax is nonredundant and that this is a reflection of the ICs that it satisfies, hence FI.

With FI in place as the IC *par excellence*, efficient satisfaction of ICs is achieved if we can ensure that FI is met in an optimal way. Central to ensuring this is phase-cyclic computation (Chomsky 2000, 2001, 2004, 2005a, 2006): To satisfy FI efficiently, uFs must be deleted as soon as they are valued (cf. Epstein & Seely 2002). This is because interpretable features (i.e. those features which contain a value throughout the derivation, as part of their lexical specification) and valued uninterpretable features (i.e. those which only acquire a value in the course of the derivation) are indistinguishable at the interface (Chomsky 2001, 2006). FI tells us that those valued uFs cannot simply be ignored; however, it is an empirical fact that they cannot be interpreted either (like lexically valued, interpretable features). Thus, for example, verbal phi-features, unlike nominal phi-features, do not affect interpretation: plural agreement on verbs does not (necessarily) result in the interpretation of a plural event. Similarly, valued Case remains uninterpretable, with no semantic (thematic) difference between the nominative subject of a finite embedded clause and the accusative-marked subject of an ECM equivalent (i.e. *he/him* in *I expect that he will arrive soon* versus *I expect him to arrive soon*), or between the accusative-marked theme of a transitive and the nominative-marked theme of the equivalent passive (e.g. *Someone hit him* versus *He was hit*). For this very reason, Case is assumed to have no interpretable counterpart in minimalist checking/Agree theory, so that there can be no question of simply interpreting a valued Case uF at the semantic interface as if it were an interpretable feature.

Since neither ignoring nor interpreting valued uF is an option, we conclude that valued uFs must be deleted. Further, this deletion must happen immediately so that no additional mechanisms must be added in order to tell apart these valued uFs from iFs. Let us consider two such additional mechanisms. Firstly, instead of phases (immediate Transfer), FI could be ensured by adding a *delay* to valuation, so that valuation of uFs waits until the immediate deletion of those valued uFs becomes possible, i.e. at the

very end of the derivation. However, this added delay would be a departure from SMT (by adding to UG and/or increasing the amount of structure that has to be kept track of). Alternatively, FI could be ensured with neither phases nor a delay by reconstructing the derivation upon its completion to see which features entered the derivation already valued and which were valued along the way. However, such unlimited lookback would also increase operative memory as the entire derivational history would have to be stored until the end. It would seem, then, that the most efficient way to solve the distinguishability problem and thus to meet FI optimally is to simply transfer these features immediately, as soon as they are valued, thus yielding phases (multiple transfer points). Let us refer to this as ‘Value-Transfer Simultaneity’ (VTS): Transfer must occur every time a uF is valued, in order to ensure efficient satisfaction of FI.

However, VTS still does not suffice to ensure that FI is met. If the entire structure is transferred at each uF site (the phases CP/vP), phases will fail to meet FI, due to the existence of unvalued features within them, such as Case on the external argument at the vP level. Every derivation would therefore crash; clearly, this cannot constitute an efficient satisfaction of FI. To achieve the latter, some version of the *Phase Impenetrability Condition* (PIC) is thus also required (Chomsky 2000, 2001): only a subpart of the phase can be transferred (the complement of the phase head). In sum, efficient satisfaction of FI requires both VTS and PIC.

Unfortunately, VTS and PIC are incompatible as they stand (Richards 2007b), so that FI is still not ensured. VTS requires that valued uFs, located on the phase head, be transferred immediately; PIC, on the other hand, demands that the phase head (and its specifiers) must remain untransferred until the *next* phase level. Since we can dispense with neither VTS nor PIC (both are required to ensure efficient satisfaction of FI, as argued above), the only solution is for valued uFs to descend from the phase head onto that part of the phase which actually gets transferred (and deleted), namely the complement domain:

(7) uF must descend from edge to nonedge (i.e. from C to T, v* to V, etc.).

That is, we have derived Chomsky’s (2005a) mechanism of feature inheritance, which now follows from ‘good design’ (the SMT): it ensures that FI is met by enabling Agree-features to be valued and deleted immediately, as part of Transfer, at the phase level.

This line of third-factor reasoning, proceeding from the premise that FI must be optimally met (i.e. without delay or lookback), thus yields the conclusion that the efficient satisfaction of IC requires three components: VTS, PIC, and feature inheritance.

Further, it now follows that every phase head must be followed by a non-phase head (in order to receive its inherited features). Any further nonphase heads beyond this do not follow from the SMT – that is, Factor III provides only for sequences of functional heads of the kind in (8-e), comprising pairs of phase heads (P) and nonphase heads (N).

- (8) a. * P – P – P – P ...
 b. * P – P – N – P ...
 c. * P – N – N – P ...
 d. * N – N – N – N ...
 e. P – N – P – N ...

A phase consisting of two heads, P-N, is thus the minimum phase in conformance with the SMT. The universal sequence of CFCs, C-T-v(-V), can now be viewed as the lexicalization of the basic template which Factor III provides (P-N-P-N); as such, this sequence is obligatorily present in every language. We have therefore reduced this syntactic universal to a Factor III effect, as desired.

As a Factor III effect, no positive evidence is required for children to acquire the basic sequence P-N-P-N (C-T-v-V) in (8-e): children minimally expect {P-N} pairs. However, sequences such as (8-c), with additional N heads, may still be acquired, as long as there is positive evidence for them in the input (i.e. Factor II). This point is made by Gallego (2008), who observes that whilst (8-e) is a third-factor effect (if the above reasoning is correct), (8-c) may emerge as a ‘second-factor’ effect. Thus Factor III interacts with Factor II to yield richer functional sequences in accordance with the parameter schema in (4), drawing on the innate pool of features provided by Factor I. All three factors thus interact to provide a point of microparametric (feature-based) variation which, as we saw in sections 1.1–1.2, is entirely compatible with the kind of minimalist system we are developing here (i.e. one with a maximally empty UG).

3. Macroparameters and Externalization

As we saw in section 1, parametrized principles are unformulable under minimalist assumptions (owing to the SMT and SUT), and the status of macroparameters is therefore dubious. For such reasons, work on macroparameters has largely been abandoned in the minimalist era (cf. Baker 1996, 2001).

If the only parameters possible in a minimalist system with a maximally un(der)specified UG are lexical, featural ones (i.e. microparameters), then

one strategy for reinstating macroparametric variation would be to derive the effects of macroparameters from a conspiracy amongst smaller, microparameters. Thus, clustering effects may emerge from a preference for identical or ‘harmonic’ settings to hold amongst a subset of microparameters (Boeckx 2008, Holmberg & Roberts 2008). As such, clustering would itself become a potential third-factor effect, attributable to general, FL-independent learning strategies that reflect the conservatism of the learner and a bias towards simplicity. To this end, Boeckx (2008:12) proposes a ‘Superset Bias’ which directs learners to strive for parametric value consistency among similar parameters, thus economizing on memory (a single parameter setting can be generalized across multiple parameters). Holmberg & Roberts (2008:52) suggest a similar learning strategy, a markedness preference which they call ‘Generalization of the Input’: “If acquirers assign a marked value to H, they will assign the same value to all comparable heads”. Hence, for example, the preference for cross-categorical harmony across directionality parameters (head-first/head-final).

Whilst this might well allow us to dispense with many former macroparameters, I would like to explore a potentially more interesting possibility here: at least some macroparameters might arise as PF-mapping strategies. In particular, macroparameters emerge from the need to linearize a symmetrical syntax at the syntax-PF interface. Such a view was already proposed and developed in Richards 2004 (see section 3.1 below); however, its plausibility has been strengthened more recently by a conjecture put forward in Chomsky 2005a, 2006 and Berwick & Chomsky 2008, namely that language design is asymmetric with respect to the two interfaces: the SMT holds only of the mapping to the SEM-interface, with “externalization [mapping to PHON] a secondary process”. Essentially, what this means for the relation between syntax and the interfaces is that, whereas SEM dictates (i.e. shapes the syntax according to its own demands – perhaps just FI, as we saw in the previous section), PF has to ‘make do’ with what the syntax gives it. That is, the mapping to PF is imperfect, which leaves it open to variation. In Berwick & Chomsky’s words:

- (9) “Parametrization and diversity, then, would be mostly – possibly entirely – restricted to externalization. That is pretty much what we seem to find: a computational system efficiently generating expressions interpretable at the semantic/pragmatic interface, with diversity resulting from complex and highly varied modes of externalization, which, furthermore, are readily susceptible to historical change.” (Berwick & Chomsky 2008:15)

We thus identify one source of variation in a minimalist, third-factor-based,

SMT-conforming system: imperfect externalization, leaving PHON to do the best job it can of adapting to its requirements the syntactic expressions which it receives from spell-out/Transfer and which are not primarily designed to meet these requirements (and so may not meet them at all). An FL not designed for externalization leaves many such options open in the morphophonological component, foremost amongst which is linearity/ordering (including head-complement ordering, copy spell-out, etc. – all rich sources of crosslinguistic variation); cf. Boeckx 2008.

That the syntax is imperfectly designed for meeting linearization demands at the PF-interface is immediately apparent from the principal property of Merge: it is symmetrical (see in particular Chomsky 2000 on Set-Merge). The symmetry of Merge may be seen as a third-factor property (cf. Brody 2000 on symmetry as the ‘default’ in nature), and is in line with our pursuit of a maximally empty UG – Merge must be maximally unspecified, and therefore it specifies no particular ordering (contra Kayne 1994), creating unordered sets. However, the physics of speech demand that one or other logical order (precedence/subsequence) be imposed between merge pairs, such as heads and complements – a third-factor consideration *par excellence*, but crucially not one which has the status of an IC (i.e. a condition to whose satisfaction the syntax must be tailored), owing to the aforesaid asymmetry of language design. In other words, if language were perfectly designed for PF (as well as SEM), then we might expect this asymmetry requirement to be imposed on the syntax as an IC, requiring syntax to yield asymmetric expressions to the PF-interface. That this does not happen and Merge is, instead, symmetrical, lends support to Chomsky’s position that “externalization is secondary” and not relevant for optimal design.

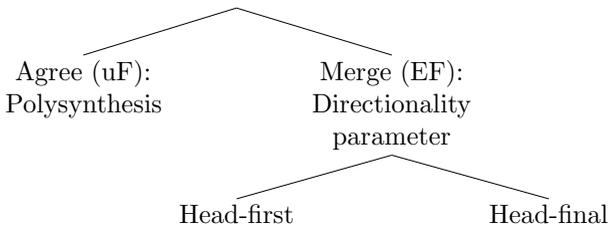
The symmetrical expressions transferred to PF are thus illegible at that interface, i.e. FI is not met at PF. The sensorimotor interface has to ‘make do’ with what the syntax gives it, which means that it must find its own solution to this legibility problem. Such solutions may vary from language to language. We would thus arrive at what Boeckx (2008) calls “variation in the externalized aspects of language”. Macroparameters might then find both a locus and a rationale as PF-repair strategies in an SMT-conforming system.⁴

⁴Admittedly, the conceptual and ontological status of these macroparameters – i.e., are they themselves Factor III effects, or must they be part of the Factor I genetic specification? – is unclear. Thus, whilst variation in the form of competing linearization options may be a third-factor effect (owing to the symmetry of Merge), the actual mechanisms (‘parameters’) that select among the possible options at the PF-interface and ensure a consistent resolution in a given language must likely be encoded. This then leads to the

What form might these macroparameters take? It is tempting to speculate that our two basic feature types from section 1.2, i.e. EFs and uFs, again provide the two basic options from which linearization at PF might proceed. Given a minimal UG comprising the operations Merge (from EFs) and Agree (from uFs), we might expect two basic macroparametric options, depending on whether linearization is determined by Merge information or by Agree information. As we will see in the following section, the former readily yields the (effect of the) traditional head-complement parameter, in that the intolerable optionality in head-complement ordering that results from an underspecified UG with symmetrical Merge must be resolved one way or the other, with two logical possibilities – head-first or head-final. Potentially, the latter option (linearization by Agree) might yield something akin to Baker’s (1996) polysynthesis parameter, insofar as Agree-based linearization would involve arguments being realized at the probe (instead of at the goal – the Merge position). Such expression on the probe would plausibly take the form of agreement morphemes and/or incorporation (see Roberts 2006 for an analysis of head-movement and cliticization in these terms, and Holmberg & Roberts 2008 for a similar conception of the polysynthesis parameter).

We arrive at a single, linearization-based macroparameter in which the initial step on the parameter schema (depicted in (10)) encompasses and subsumes the two major macroparameters from the pre-minimalism era: directionality and polysynthesis, corresponding to whether a language linearizes by Merge (EF) or by Agree (uF), respectively.

(10) *Linearization (desymmetrization) strategy at PF*



In sum, whilst the majority of macroparameters might indeed be better

question: do these mechanisms add to UG (or are they part of some FL-external specification?), and, if so, does this matter from the perspective of the SMT (if the latter holds only of the SEM-interface, so that only the mapping to SEM requires a maximally empty UG)? I leave these rather nebulous questions unresolved here.

understood in terms of conspiratorial settings among microparameters (as outlined at the start of this section), there may be room in a minimalist system for one genuine macroparameter to be accommodated, at the PF-interface, as a variable linearization (desymmetrization) strategy, tentatively given in (10). Interestingly, the first step in this PF-based schema – the split between polysynthesis and directionality – resembles the top of Baker’s (2001) independently motivated hierarchy of syntactic macroparameters. In the following section, we turn to the second step in (10), namely the implementation of the Merge-based linearization strategy, and consider how the competing options delivered by a symmetrical syntax might be resolved into consistent head-first/head-final orderings at PF.

3.1. VO/OV and Holmberg’s Generalization

If the hallmark of a macroparameter is a clustering of grammatical properties around a single parametric setting, then we would expect a Merge-based linearization strategy located at the PF-interface to yield additional effects beyond just directionality. Richards 2004 offers a concrete attempt to implement a linearization parameter with exactly these properties, delivering certain restricted order-preservation effects in addition to basic head-first/-final orderings from the desymmetrization of Merge at PF. This section summarizes the main workings of that parameter.

Linearization by Merge (i.e. the right-hand path in (10)) takes Merge-pairs (i.e. sisters) to be the basic unit of linearization: Merge partners in the syntax become ordering partners at PF. A good candidate for such a sisterhood-based linearization strategy is that offered in Epstein et al (1998: Chapter 5). They propose that it is simple c-command, rather than the arguably more complex notion of asymmetric c-command, that translates to precedence at PF. Since Merge-pairs (such as head-complement) mutually c-command each other, they overdetermine linearization, providing contradictory instructions to PF such that each sister must precede the other.⁵ Linearization-by-sisterhood must therefore proceed via a strategy of deleting superfluous (contradictory, ambiguous, symmetrical) information at the

⁵As such, a linearization strategy based on sisterhood and mutual c-command is already a more viable candidate than one based on the LCA, since the whole tree can be linearized on the basis of sisterhood, whereas the LCA underdetermines linearization under Bare Phrase Structure, where the symmetrical base pair of every sub-tree fails to be linearized.

interface, so as to resolve the paradox of mutual precedence. To this end, Epstein et al propose a Precedence Resolution Principle (PRP), essentially as given in (11).

(11) *PRP*

If two categories symmetrically c-command each other, ignore all c-command relations of one of the categories to the other. [Based on Epstein et al 1998:152]

From this interface deletion strategy, the effects of a head-complement directionality parameter are immediately implied: by ignoring the c-command relation from head to complement, head-finality ensues; by ignoring the opposite c-command relation, from complement to head, head-initiality emerges:

$$(12) \quad \begin{array}{c} \text{VP} \\ \swarrow \quad \searrow \\ \text{V} \quad \text{DP} \end{array} \quad \begin{array}{l} \text{C-command relations} \\ \rightarrow \{V > \text{DP}, \text{DP} > V\} \end{array} \quad \begin{array}{l} \text{Ignore} \\ \rightarrow V > \text{DP} \\ \rightarrow \text{DP} > V \end{array} \quad \begin{array}{l} \text{PF-order} \\ \rightarrow \text{DP} > V \quad (= \text{OV}) \\ \rightarrow V > \text{DP} \quad (= \text{VO}) \end{array}$$

Expanding on this, Richards 2004 proposes that (11)/(12) should be generalized to hold of internal as well as external Merge, thus imposing VO/OV 'shape' on Move and Merge alike. That is, the PRP should be amended so that it deletes a consistent subset of c-command instructions in any given language (rather than allowing the c-command relation $\text{DP} > V$ to always win out in the case of Move, irrespective of base order, the way it does in Epstein et al).

We could express this roughly as in (13). Assume that the unordered set created by $\text{Merge}(\alpha, \beta)$ delivers contradictory instructions to PF in the form of two contradictory ordered pairs $\{\langle \alpha, \beta \rangle, \langle \beta, \alpha \rangle\}$, formalizing the above idea that unordered sets translate to mutual precedence at PF. Then, the two competing linearization options are resolved through the consistent deletion of either $\langle \alpha, \beta \rangle$ (yielding head-finality, (13-b)) or $\langle \beta, \alpha \rangle$ (yielding head-initiality, (13-a)).

(13) *Parametrized desymmetrization:*

Given $\text{Merge}(\alpha, \beta) \rightarrow \{\langle \alpha, \beta \rangle, \langle \beta, \alpha \rangle\}$:

- a. Head-initial = Delete all $\text{Comp} > \text{Head}$ [i.e. $\{\langle \alpha, \beta \rangle, \langle \beta, \alpha \rangle\} \rightarrow \{\langle \alpha, \beta \rangle\}$]
- b. Head-final = Delete all $\text{Head} > \text{Comp}$ ⁶ [i.e. $\{\langle \alpha, \beta \rangle, \langle \beta, \alpha \rangle\} \rightarrow \{\langle \beta, \alpha \rangle\}$]

⁶The "all" in (13-a,b) will be relativized to the phase level below.

Let us take (13) to be the parameter constituting the second step along the right-hand branch in (10). That is, (13) operates at the syntax-PF interface to ensure the Merge-based linearization of a symmetrical syntax. It has the effect of a head parameter at the interface, but maintains a uniform UG/syntax, in conformance with the SUT and SMT (section 1.2).

However, as desired from a macroparameter, (13) immediately implies further consequences. In particular, the linear order-preservation effect known as Holmberg's Generalization (HG) that constrains Germanic Object Shift is directly entailed by the 'head-initial' (VO) setting (13-a) of this parameter.

A full summary of the immense literature on this topic will not be attempted here. It suffices for our purposes to characterize Object Shift (OS) as the short leftward displacement of weak (destressed) objects in Germanic (Holmberg 1986, 1999, Vikner 1995, Collins & Thráinsson 1996, Thráinsson 2001, and many others) and HG as the constraint on this operation such that the shifted object cannot cross an in-situ (nonfinite) lexical verb.

As a syntactic constraint on movement, HG is dubious from the minimalist perspective adopted here, since the technology usually used to account for it (invoking such notions as equidistance, Greed, government, phonological visibility, topic/focus-features, etc.) exceed a minimally specified UG and thus depart from the SMT. Instead, I align myself with many other researchers⁷ in taking HG to be best (and most simply) viewed as a verb-object order preservation constraint: i.e., the derived order must reinstate the base order (VO). This constraint famously holds only of VO languages (cf. (14)); the equivalent short object movement in OV Germanic ('scrambling') may occur irrespective of the finiteness of the main verb, yielding both OV (= (15-b)) and, with V2-raising, VO orders (= (15-a)) alike, an apparent 'anti-HG' effect:

(14) VO (Icelandic)

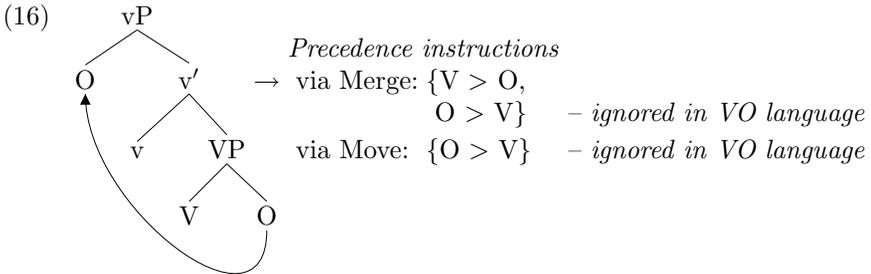
- a. Nemandinn las (bókina) ekki (bókina)
The-student read (the-book) not (the-book)
"The student didn't read the book."
- b. Nemandinn hefur (*bókina) ekki lesið (bókina)
The-student has (the-book) not read (the-book)
"The student hasn't read the book."

⁷These include: Müller 2000, 2001, 2006; Fox & Pesetsky 2003, 2005; Williams 2003; Richards 2004; Koenenman 2005.

(15) OV (German)

- a. Der Student las (das Buch) nicht (das Buch)
 The student read (the book) not (the book)
 “The student didn’t read the book.”
- b. Der Student hat (das Buch) nicht (das Buch) gelesen
 The student has (the book) not (the book) read
 “The student hasn’t read the book.”

HG and its confinement to VO languages emerges straightforwardly from the linearization algorithm in (13), since object displacement over V in a VO language is only orderable by (13-a) where further movement of V over O takes place:



The displaced O in (16) must be the tail of a V > O chain, and not the head of an O > V chain, in order to be PF-realized in its derived position, since the latter instruction (O > V) is deleted in VO/(13-a) languages. Object movement is thus strongly tied to verb movement, hence its further restriction to those VO languages with independent movement of the verb (V-to-T, e.g. Icelandic). In this way, HG is derived for exactly that subset of languages in which it holds (namely, those set to (13-a), i.e. VO languages). Scrambling is free in OV languages without any further V movement (i.e. the anti-HG effect in (15-b)) as the resulting O > V instructions are uniformly legible at PF in languages set to (13-b).

The ease with which this system derives order-preserving movement, such as OS/Scrambling in (14)-(15-b), could actually be a disadvantage, however. A major problem faced by all such PF/linearization-based accounts of HG is that order preservation is by far the exception rather than the rule. The majority of movement types are order-distorting, freely inverting basic VO to yield surface OV, as in the standard A/A-bar movements in (17) – passivization, topicalization, *wh*-movement, etc. (Consider also the VO order that arises in OV German in (15-a).) Clearly, HG does not hold of all movements, even in VO languages.

- (17) a. A man arrived (*a man*)
 b. John was rescued (*John*)
 c. John, I like (*John*)
 d. Which book did you read (*which book*)

Thus, whilst some types of movement conserve head-complement ‘shape’, many others do not. In fact, as argued in Richards 2004, the head-complement order-preservation property is unique to OS/Scrambling, and is its principal, defining property.

The question, then, is how a single grammatical system (e.g. VO Icelandic) can allow both shape-breaking (passive, *wh*-movement) and shape-preserving (HG/OS) movement operations, given that (13) should be responsible for linearizing the entire tree.

The system presented above allows a simple resolution to this problem, drawing a principled line between shape-breaking and shape-conserving movement in exactly the right place. First we must make two reasonable assumptions, motivated elsewhere in the literature:

- (18) All varieties of *v* are phase heads –i.e. both transitive *v** and passive/unaccusative ‘defective’ *v* (*v*_{def}); cf. Legate 2003, Richards 2004.
 (19) OS/Scrambling targets spec-*v*P, as in (16); cf. Neeleman & Weerman 1999, Chomsky 2001, Kitahara 2002.

Given (19), it is immediately clear that order-preserving object movement is ‘shorter-distance’ than order-disrupting object movement: the former targets spec-*v*P; the latter targets spec-TP (A-movement / raising-to-subject) or spec-CP (A-bar movement). Given (18), this means that only the order-preserving kind of movement involves movement within a single phase; order-disrupting movement targets positions outside of the original phase. This leads to a strikingly simple generalization covering all of (14)-(15), (17):⁸

- (20) a. Order-preserving movement is phase-internal.
 b. Order-disrupting movement is cross-phasal.

⁸Note that the phase-based approach to linearization and order-preservation of Fox & Pesetsky (2003, 2005) cannot make such a generalization, since all the movements in question (OS, passivization, *wh*-movement, V2) are cross-phasal on their assumptions (VP, rather than *v*P, being the relevant linearization domain for them). For a detailed comparison between the system proposed here and Fox & Pesetsky, see Richards 2004, Richards 2007a.

All that is required in order to derive (20) is for (13) to operate on a phase-by-phase basis. That is, (13) holds up to the phase level, at which point the phase is transferred and linearized and the relevant ordering information is lost to subsequent phases, in accordance with Chomsky's complexity-reducing view of phases as the unit of derivational memory – derivational information, plausibly including ordering information (c-command relations), is simply 'forgotten' at the end of the phase. It follows that any constraints on linearization, such as the requirement of consistent precedence instructions (imposed by (13)) that yields order preservation of the HG kind, can only hold within a single phase (and not beyond, contra Fox & Pesetsky 2003/5). In the case of order-inverting object movement across the verb, there is no longer any memory that the object DP was originally merged with V in the lower phase (vP) at the point where the higher phase (CP) is transferred and linearized. Instead, the object DP is 'relinearized' in the higher phase with its derived sister (a projection of T/C).

Summarizing this section, we have sketched out Richards's (2004) proposal for a Merge-based linearization algorithm that conforms to the requirements of the previous section, desymmetrizing syntactic information at the PF-interface and yielding a macroparametric cluster that includes basic directionality (VO/OV) and the \pm HG property. Further, it allows a principled line to be drawn between shape-breaking and shape-preserving movement, such that escaping a phase implies escaping the order-preservation effects imposed by (13) on account of the cyclic purging ('forgetting') of derivational information at the phase level. VO/OV 'shape' is imposed at the phase level by (13), but beyond this level, the system no longer cares, on third-factor grounds (reduction of operative memory through phase-cyclic computation), hence the ordering freedom associated with longer-distance movement.

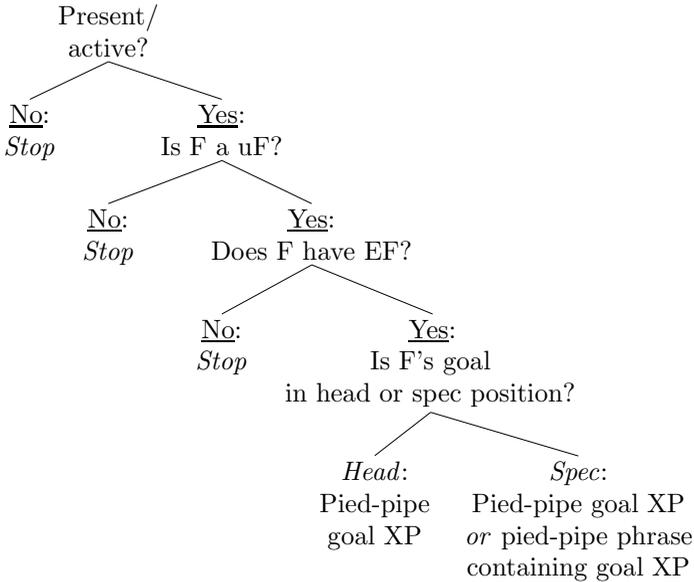
This suggests a further possible source for variability in a minimalist system: an underspecified UG relying on third-factor principles yields points of indeterminacy where the system 'no longer cares', owing to a loss of information or a lack of specification. In the case above, the result is ordering freedom beyond the phase level. In the next section, we consider the result of such indeterminacies in the syntax.

4. Microparameters, Narrow Syntax, and Tolerable Optionality

In section 3, we explored one source of variation in a minimalist system: imperfect mapping to PF. At the end of the previous section, we suggested a second source: the underspecification of UG, leaving options open. The

prediction for narrow syntax, the domain of the SMT, is that such indeterminacies should give rise to free variation, with competing options existing side by side in a single grammar. This is because the consistent resolution of this optionality in favour of one or other option in the syntax would require the existence of a special, FL-specific device (a parameter), enlarging the content of UG. A maximally empty UG thus leads us to expect ‘true’ (semantically vacuous) syntactic optionality to emerge in the satisfaction of certain featural requirements. As we saw in section 2, FI demands that featural requirements be met (i.e. uFs must be valued and deleted); furthermore, third-factor considerations of efficient computation, such as minimal search and the PIC, will serve to constrain the manner in which those requirements are met (thus uF will be satisfied by the closest matching goal that the probe encounters, and the latter cannot seek a probe inside a phase that has already been transferred). However, beyond these considerations and restrictions, the manner in which featural requirements are met is left unspecified, opening up multiple options in certain configurations. Let us consider the configurations in which such optionality might arise.

According to the ‘Yes-No’ format of (featural, micro)parameter theory developed in Roberts 2007, all parameters must be fixed one way or another – thus for all features F, the first step in the schema in (4) must ultimately be set to either ‘Yes’ or ‘No’. Further, there can be no optional associations of a feature F with the movement-triggering feature EF in order to yield optional movements. Either F is associated with EF in a given language or it is not; to allow for optional EF-associations is tantamount to allowing for optional, competing grammars (cf. Kroch 1989), since it is this very feature (EF, its presence or absence) which defines a parameter on this approach. (See Koopman & van der Wurff 2000 for arguments against competing grammars.) This is shown in (21), a more articulated and specified version of the parameter schema in (4), based on Holmberg & Roberts 2008 and slightly modified in light of our claim in section 1.2 that Merge and Agree reduce to the two feature types, EF and uF.

(21) *Partial parameter schema for feature F*

The lexical, microparametric options left open by UG (i.e. whether a feature F is present in a language or not, whether it is associated with EF or not, etc.) yields a parameter that must be fixed one way or another (e.g., in terms of (21), a feature F may be set to Yes-Yes-Yes-Spec). However, certain settings may yield systems with further indeterminacies left open by the underspecified, maximally empty UG. The Yes-Yes-Yes-Spec setting of (21) yields such a scenario.⁹ Where the probe finds its goal in a left branch (specifier), multiple pied-piping options emerge. Biberauer & Richards 2006 discuss numerous empirical cases, including optional embedded verb-second in Modern Afrikaans. The following section illustrates the relevant configuration and resulting optionality on the basis of Russian *wh*-movement.

⁹The ‘Spec’ (versus ‘Head’) setting, as given at the bottom of (21), is included for expository convenience only; it does not need stating in the grammar (that is, it is not a parameter per se). Rather, the probe finds its goal wherever the latter is located, which is determined by independent factors in the language in question (see the difference between Russian and English in section 4.1: in Russian, the *wh*-goal is located in spec-DP, whereas in English it is the (head of) that DP itself).

4.1. True optionality in Russian *wh*-movement (Biberauer & Richards 2006)

As captured by the Left Branch Constraint (LBC), movement of a *wh*-element to C[+*wh*] in English requires the whole DP to be pied-piped along to spec-CP (cf. (22)). Famously, the LBC does not appear to be operative in all languages. Thus, as (23-b) shows, the non-pied-piped counterpart of (23-a) is grammatical in Russian, unlike in English. Furthermore, pied-piping is not barred in Russian, but is simply optional (as (23) shows). We thus have a case of semantically vacuous, ‘true’ optionality (i.e. optionality which has no interpretive effect on outcome and so does not conform to the Chomsky-Reinhart-Fox rationale for optional operations; see Chomsky 2001: 34 (60) for a formulation thereof).

- (22) a. Whose book did you read?
 b. *Whose did you read book?

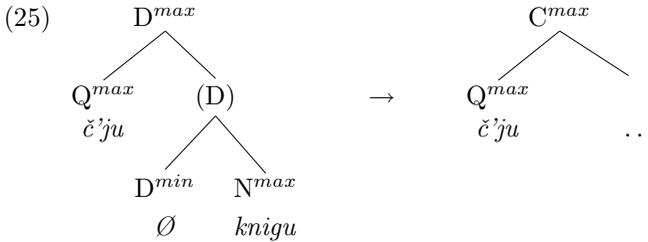
- (23) a. Č’ju knigu ty č’ital [Russian]
 Whose book you read
 b. Č’ju ty č’ital knigu
 Whose you read book

Biberauer & Richards (2006) account for the difference between Russian and English as follows. Pied-piping is forced in English if we assume that (i) *wh*-elements such as *what* and *which* are *wh*-determiners and thus the head of their respective *wh*-DPs, and (ii) heads cannot move to specifier positions (for whatever reason; Biberauer & Richards invoke Chomsky’s (1995) Chain Uniformity Condition, the status of which is unclear in the current system). If heads cannot undergo *wh*-movement, then these elements cannot raise independently to spec-CP. Pied-piping of the whole *wh*-DP is thus forced in English (Biberauer & Richards refer to this as ‘head-piedpiping’), as illustrated in (24).

- (24)
- | | | |
|---|---|-------------------------------|
| D^{max}
 | ↯ | C^{max}
 |
| D^{min} N^{max}
<i>which</i> <i>book</i> | | D^{max}
<i>which</i> ... |

Russian, on the other hand, lacks overt articles, determiners, and many other exponents of the category D. It is therefore reasonable to treat Russian *wh*-elements, such as *kakoj* (‘what’, ‘which’), *kotoryj* (‘which’, ‘what’), and *čej* (‘whose’), as quantifiers heading quantificational phrases. These QPs would

then arguably occupy the specifier of the relevant DPs. As such, they have the status of maximal projections and are thus able to raise independently to spec-CP without incurring a uniformity violation; this derives the lack of LBC effects in (23-b), as illustrated in (25), an instance of ‘spec-piedpiping’.



However, a further option is available here: the entire DP, containing the goal QP in its specifier, may also be pied-piped, since this too conforms to uniformity and allows satisfaction of the relevant feature (EF on C). Thus as long as the C probe’s EF is satisfied, the grammar doesn’t mind how. Such true optionality is characteristic of spec-piedpiping.

Essentially, we can characterize the relevant configuration in which such system-internal, formal optionality occurs as in (26), adapted from Roberts (2007:308):



To satisfy EF on uF, either YP may raise or else the containing XP may do so – the grammar doesn’t care. The result is tolerable optionality – free variation within a single language, as expected from a minimally specified UG in which there is no room for parameters or any other devices for resolving such indeterminacies.

This approach has further interesting consequences in the domain of diachronic syntax. In particular, Biberauer & Roberts (2005, 2007) and Roberts (2007) apply it to language change in the history of English (and elsewhere). As Roberts (2007:309, 332-3) points out, such instances of formal optionality provide a possible solution to the logical problem of language change in an approach to the latter that is based on imperfect acquisition (i.e. parameter missetting and abductive change; cf. Lightfoot 1979, 1991, 1999). Put simply, the problem faced by parameter resetting models of change is to explain how the adult population (Generation 1) can produce a trigger experience sufficient for the language-acquiring generation (Generation 2) to set a different parameter value from Generation 1, without Generation 1 having to have the alternative, innovative setting themselves in order to produce that PLD in the first place. The notion of an indifferent,

underspecified UG, i.e. one that allows for true optionality wherever (26) obtains, has the potential to cut this Gordian knot. As long as doublets (as in (23)) are generable by a single grammar (i.e. a single generation of speakers), there is variation inherent in the PLD. All that is then required, Roberts observes, is for one of the doublets to become more frequent than the other – e.g. by taking on sociolinguistic, pragmatic prominence – and it will oust the other, insofar as a retreat to the non-doublet-generating grammar is possible.

Whether or not such a retreat is possible is dependent on whether a grammar that just generates the prominent alternant is permitted by UG. Thus a grammar that just generates option (23-b) and not (23-a) is impossible in this system, since (26) always allows for the XP-pied-piping option in addition to YP-pied-piping. On the other hand, a grammar just generating option (23-a) is fine (English being such a language – cf. (22)). Such a retreat also depends on psycholinguistic factors, in particular the learning strategies deployed in the acquisition process, of which the Subset Principle (Berwick 1985) is of relevance here. Thus, as Biberauer & Roberts (2007) point out, the two grammars involved in cases of true optionality stand in a subset-superset relation – in the case of Russian and English *wh*-movement, English (22) is a subset of Russian (23). Should the non-pied-piping doublet (23-b) lose sociolinguistic value and thus disappear sufficiently from the PLD, then a retreat to the subset grammar, generating only (23-a) as in English, is predicted (syntactically, this would involve a categorial reanalysis of Russian *wh*-words from Q heads to D heads), since there would be no positive evidence for the superset grammar. Indeed, exactly such a change has arguably taken place in the history of Greek (see Biberauer & Richards 2006 for details).

In sum, language change converges with language acquisition in those configurations in which true optionality arises from an underspecified UG in accordance with the SMT. Whilst the competing options are tolerable in a synchronic grammar, yielding stable variation as in Russian (23), they may be resolved diachronically in the process of language change in favour of a subset grammar generating only a single option, as in English (22).

5. Summary

We have identified and exemplified two kinds of variation that may arise in an SMT-conforming FL in which Factor I is maximally empty and Factor III carries the bulk of the burden of explanation: (i) variation due to the imperfect mapping to PF (section 3), in which the resolution of competing

linearization options takes the form of PF-based macroparameters (section 3.1); and (ii) variation due to an underspecified UG (section 4), in which competing options are tolerated in the syntax, with no parametric resolution, in conformance with the SMT. In both cases, variation arises from points of indeterminacy, be it the manner in which a feature is satisfied or the manner of externalization (such as how the symmetrical syntax is linearized). Crucially, none of these choice points affects meaning, lending support to Chomsky's (2005a, 2006) and Berwick & Chomsky's (2008) position that language is optimally designed to meet IC only at SEM. The mapping to SEM is thus uniform, with a single, optimal design solution that admits of no variability.

6. References

- Baker, M. (1996): *The Polysynthesis Parameter*. Oxford University Press, Oxford/New York.
- Baker, M. (2001): *The Atoms of Language*. Basic Books, New York.
- Baker, M. (2008): *The Syntax of Agreement and Concord*. Cambridge University Press, Cambridge.
- Berwick, R. (1985): *The Acquisition of Syntactic Knowledge*. MIT Press, Cambridge, Mass.
- Berwick, R. & N. Chomsky (2008): The Biolinguistic Program: The Current State of its Evolution and Development. Ms., MIT. Forthcoming in: A.M. di Sciulo & C. Aguero, eds., *Biolinguistic Investigations*. MIT Press, Cambridge, Mass.
- Biberauer, T. & M. Richards (2006): True Optionality: When the grammar doesn't mind. In: C. Boeckx, ed., *Minimalist Essays*. Benjamins, Amsterdam, pp. 35-67.
- Biberauer, T. & I. Roberts (2005): Changing EPP-parameters in the history of English: accounting for variation and change. *English Language and Linguistics* 9: 5-46.
- Biberauer, T. & I. Roberts (2007): The Return of the Subset Principle. Ms., University of Cambridge.
- Bobaljik, J. & H. Thráinsson (1998): Two heads aren't always better than one. *Syntax* 1: 37-71.
- Boeckx, C. (2008): Approaching Parameters from Below. Ms., Harvard.
- Borer, H. (1984): *Parametric Syntax: Case Studies in Semitic and Romance Languages*. Foris, Dordrecht.
- Brody, M. (2000): Mirror Theory: Syntactic Representation in Perfect Syntax. *Linguistic Inquiry* 31: 29-56.

- Chomsky, N. (1995): *The Minimalist Program*. MIT Press, Cambridge, Mass.
- Chomsky, N. (2000): Minimalist Inquiries: the Framework. In: R. Martin, D. Michaels, & J. Uriagereka, eds, *Step by step*. MIT Press, Cambridge, Mass., pp. 89-156.
- Chomsky, N. (2001): Derivation by Phase. In: M. Kenstowicz, ed., *Ken Hale: A Life in Language*. MIT Press, Cambridge, Mass, pp. 1-52.
- Chomsky, N. (2004): Beyond Explanatory Adequacy. In: A. Belletti, ed., *Structures and Beyond. The Cartography of Syntactic Structures (volume 3)*. Oxford, Oxford University Press.
- Chomsky, N. (2005a): On Phases. Ms., MIT.
- Chomsky, N. (2005b): Three Factors in Language Design. *Linguistic Inquiry* 36: 1-22.
- Chomsky, N. (2006): Approaching UG from below. Ms., MIT. Published in: H.-M. Gärtner & U. Sauerland, eds., *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax-Semantics*. De Gruyter, Berlin, pp. 1-30.
- Collins, C. & H. Thráinsson (1996): VP-internal Structure and Object Shift in Icelandic. *Linguistic Inquiry* 24: 391-444.
- Epstein, S. D., E. Groat, R. Kawashima & H. Kitahara (1998): *A Derivational Approach to Syntactic Relations*. Oxford University Press, New York.
- Epstein, S. D. & T. D. Seely (2002): Rule Applications as Cycles in a Level-free Syntax. In: S. D. Epstein & T. Daniel Seely, eds, *Derivation and Explanation in the Minimalist Program*. Blackwell, Oxford, pp. 65-89.
- Fox, D. & D. Pesetsky (2003): Cyclic Linearization and the Typology of Movement. Handout, paper presented at GLOW 2003 (Lund).
- Fox, D. & D. Pesetsky (2005): Cyclic Linearization of Syntactic Structure. *Theoretical Linguistics* 31: 1-46.
- Gallego, Á. (2008): The Second Factor and Phase Theory. Ms., Universitat Autnoma de Barcelona.
- Hauser, M., N. Chomsky & W. T. Fitch (2002): The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science* 298: 1569-79.
- Holmberg, A. (1986): Word Order and Syntactic Features in the Scandinavian Languages and English. Doctoral dissertation, University of Stockholm.
- Holmberg, A. (1999): Remarks on Holmberg's Generalization. *Studia Linguistica* 53: 1-39.
- Holmberg, A. & I. Roberts (2008): Introduction. *Parametric variation: null subjects in minimalist theory*. Cambridge University Press, Cambridge. (Forthcoming.)

- Kayne, R. (1994): *The Antisymmetry of Syntax*. MIT Press, Cambridge, Mass.
- Kitahara, H. (2002): Scrambling, Case, and Interpretability. In: S. D. Epstein & T. D. Seely, eds., *Derivation and Explanation in the Minimalist Program*. Blackwell, Oxford, pp. 167-83.
- Koenenman, O. (2005): Shape conservation, Holmberg's generalization, and predication. Ms., Amsterdam.
- Koenenman, O. & A. Neeleman (2001): Predication, Verb Movement and the Distribution of Expletives. *Lingua* 111: 189-233.
- Koopman, W. & W. van der Wurff (2000): Two Word Order Patterns in the History of English: Stability, Variation and Change. In: R. Sornicola, E. Poppe & A. Shisha-Halevy, eds., *Stability, Variation and Change of Word-Order Patterns over Time*. John Benjamins, Amsterdam/Philadelphia, pp. 259-83.
- Kroch, A. (1989): Reflexes of grammar in patterns of language change. *Language Variation and Change* 1: 199-244.
- Legate, J.A. (2003): Some Interface Properties of the Phase. *Linguistic Inquiry* 34: 506-16.
- Lightfoot, D. (1979): *Principles of Diachronic Syntax*. Cambridge University Press, Cambridge.
- Lightfoot, D. (1991): *How To Set Parameters: Arguments from Language Change*. MIT Press, Cambridge, Mass.
- Lightfoot, D. (1999): *The Development of Language*. Blackwell, Oxford.
- Longobardi, G. (2001): Formal Syntax, Diachronic Minimalism, and Etymology: The History of French *Chez*. *Linguistic Inquiry* 32: 275-302.
- Müller, G. (2000): Shape Conservation and Remnant Movement. Ms., University of Tübingen.
- Müller, G. (2001): Order Preservation, Parallel Movement, and the Emergence of the Unmarked. In: G. Legendre, J. Grimshaw & S. Vikner, eds., *Optimality-Theoretic Syntax*. MIT Press, Cambridge, Mass, pp. 279-313.
- Müller, G. (2006): Towards a Relativized Concept of Cyclic Linearization. Ms., University of Leipzig. Published in: H.-M. Gärtner & U. Sauerland, eds., *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax-Semantics*. De Gruyter, Berlin, pp. 61-114.
- Neeleman, A. (1994): Scrambling as a D-structure phenomenon. In: N. Corver & H. van Riemsdijk, eds., *Studies in Scrambling: Movement and Non-Movement Approaches to Free Word-Order Phenomena*. De Gruyter, Berlin, pp. 387-429.
- Neeleman, A. & F. Weerman (1999): *Flexible Syntax: A Theory of Case and Arguments*. Kluwer, Dordrecht.

- Poepfel, D. & K. Wexler (1993): The Full Competence Hypothesis of Clause Structure in Early German. *Language* 69: 1-33.
- Richards, M. (2004): Object Shift and Scrambling in North and West Germanic: A Case Study in Symmetrical Syntax. Doctoral dissertation, University of Cambridge.
- Richards, M. (2007a): Dynamic Linearization and the Shape of Phases. *Linguistic Analysis* 33: 209-237 (Special Issue: Kleanthes Grohmann, ed., *Dynamic Interfaces*, Vol. 2).
- Richards, M. (2007b): On feature-inheritance: An argument from the Phase Impenetrability Condition. *Linguistic Inquiry* 38: 563-72.
- Richards, M. (2008a): Defective Agree, Case Alternations, and the Prominence of Person. In: M. Richards & A. Malchukov, eds., *Scales. Linguistische Arbeitsberichte* 86: 137-161.
- Richards, M. (2008b): Deriving The Edge: What's In A Phase? Ms., University of Leipzig.
- Richards, M. (2008c): Quirky Expletives. In: R. d'Alessandro, G. Hrafnbjargarson & S. Fischer, eds, *Agreement Restrictions*. De Gruyter, Berlin, pp. 181-213.
- Rizzi, L. (1982): *Issues in Italian Syntax*. Foris, Dordrecht.
- Rizzi, L. (1997): The Fine Structure of the Left Periphery. In: L. Haegeman, ed., *Elements of Grammar*. Kluwer, Dordrecht.
- Roberts, I. (2006): Clitics, Head Movement, and Defective Goals. Ms., University of Cambridge.
- Roberts, I. (2007): *Diachronic Syntax*. Oxford University Press, Oxford.
- Roberts, I. & A. Roussou (2003): *Syntactic Change: A Minimalist Approach to Grammaticalization*. Cambridge University Press, Cambridge.
- Thráinsson, H. (1996): On the (Non-)Universality of Functional Categories. In: W. Abraham, S.D. Epstein, H. Thráinsson & J.-W. Zwart, eds., *Minimal Ideas: Syntactic Studies in the Minimalist Framework*. Benjamins, Amsterdam, pp. 253-81.
- Thráinsson, H. (2001): Object Shift and Scrambling. In: M. Baltin & C. Collins, eds., *The Handbook of Contemporary Syntactic Theory*. Blackwell, Oxford, pp. 148-202.
- Uriagereka, J. (1995): An F Position in Western Romance. In: K. Kiss, ed., *Discourse Configurational languages*. Oxford University Press, New York/Oxford, pp. 153-75.
- Vikner, S. (1995): *Verb Movement and Expletive Subjects in the Germanic Languages*. Oxford University Press, New York.
- Williams, E. (2003): *Representation Theory*. MIT Press, Cambridge, Mass.