

24. Minimalism

Marc Richards

1. THE NATURE OF THE PROGRAM

Minimalism is a program for extending and developing linguistic theories beyond data coverage, holding them to a higher level of scrutiny and raising the bar for what counts as a genuine explanation of a linguistic phenomenon. Minimalist questions of the kind to be outlined below can be asked of any theoretical framework; the Minimalist Program (MP) itself, however, grew naturally out of the Principles and Parameters approach to transformational generative grammar (Chomsky 1981), and it is here that its greatest advances have been made and its most promising future prospects arguably lie. As the most recent incarnation of Principles and Parameters Theory (PPT), it shares with its predecessor, Government and Binding Theory (GB) the realist-mentalist conception of language fundamental to the Chomsky's work. That is, the object of study is *I-language*, a state of mind and a property of individual brains ('I' denoting internal, intensional, individual), and not language in its wider, social sense of a set of utterances or corpus of external tokens to which any individual has only partial access. The goal of rational linguistic inquiry is then to characterize the nature of that internal knowledge.

Traditionally, this characterization has sought to answer two questions: (i) What constitutes knowledge of language in the mind of a speaker-hearer (such that we have a correct description of the possible grammatical structures of a language)? And (ii) how does that knowledge arise in the mind of the speaker-hearer given the environmental stimulus (linguistic input) – how is it acquired, how much of it is acquired, and what is a possible (acquirable) human language? Insofar as a theory can provide satisfactory answers to these questions, it meets the respective benchmarks of descriptive and explanatory adequacy. Doing so is no mean feat, since there is a clear tension between the two desiderata, i.e. between reconciling a sufficiently restrictive and universal theory of the initial state of Universal Grammar (UG) – the genetic endowment which defines the range of possible, attainable human languages – with the diversity of variation attested across the world's languages. PPT offered the first real breakthrough in reconciling these antagonistic demands. By factoring out the universal (principles) from the variable (parameters), PPT simplified the acquisition procedure considerably – instead of selecting among entire, fully specified grammars (as on earlier generative approaches), acquisition now proceeds by the fixing of values of a limited set of open parameters on the basis of salient input data. In this way, PPT solves Plato's problem at least in principle (if not yet in practice).

The state of the field at the end of the eighties, then, could be roughly described as follows. A decade of research in GB theorizing had led to a richly specified view of UG as comprising a highly articulated modular architecture (Case Filter, Theta Criterion, X-bar Theory, Subjacency and the Empty Category Principle, Binding Theory, Control Theory, etc.), with each module composed of its own vocabulary, rules, conditions, constraints and principles. All of this technology was specific to UG, that is, to the faculty of language (FL), and it allowed for vast empirical coverage, thus attaining a high level of descriptive adequacy. Through parametrization of these principles under PPT, descriptive adequacy was attained for all possible human languages, as well as an account of their acquisition, thus reaching the level of explanatory adequacy. With GB thus yielding descriptive adequacy, and its umbrella framework PPT enabling explanatory adequacy, the stage was set for a

deeper level of explanatory adequacy to be sought. That is, PPT was now in a unique position to seriously address minimalist concerns of the kind that had already been raised in the earliest days of generative grammar but which had been beyond the scope of feasible inquiry at that stage.

Essentially, minimalist concerns are those which relate to a third research question regarding the nature of linguistic knowledge (beyond questions (i) and (ii) above). If we are taking the biological foundations of FL seriously, then to the ontogenetic question in (ii), that of explanatory adequacy and the logical problem of language acquisition, must be added the phylogenetic question of how UG arose in the species. That is, we have to meet the desideratum of what we might call ‘evolutionary adequacy’ (Boeckx & Uriagereka’s (2007) “natural adequacy”, Chomsky’s (2004a) “beyond explanatory adequacy) or Darwin’s problem (Boeckx 2009, Hornstein 2009) – the logical problem of language evolution. Just as knowledge exceeds experience, demanding innate, language-specific mental structures in order to fill in the gap between linguistic input and the acquired final state of linguistic competence, so it seems that that knowledge (the richly specified UG of GB) exceeds what we can reasonably or plausibly attribute to natural selection, given the tiny evolutionary window within which human language is assumed to have appeared (suddenly, around 60,000 years ago; Chomsky 2008). The time frame is simply too short for a highly articulated, richly specified UG replete with myriad FL-specific rules, constraints and principles to have emerged by selective pressure, with each piece of FL-specific technology (each rule, principle, module) separately selected for.¹ Every claim about the content of UG is a claim about the brain; therefore, the more we attribute to UG, the greater becomes the problem of accounting for how it all got there, evolutionarily speaking. This problem simply cannot be ignored if we are taking seriously the claim that language (syntax) is a biological entity, a natural object to be investigated as part of the natural sciences. Unfettered proliferation of FL-specific technology ushers us ever further from a naturalistic approach to language and a genuine understanding of the properties of FL.

We thus have another tension between levels of adequacy to grapple with (descriptive/explanatory versus evolutionary). In light of such considerations, the minimalist hypothesis is that UG must be maximally empty and maximally simple, perhaps containing just a single innovation that takes us from the pre-linguistic state to the fully functioning FL. A leading hypothesis, put forward by Hauser, Chomsky & Fitch (2002), is that this single innovation was (context-free) recursion, yielding phrase structure and providing a link between the two extralinguistic performance systems of sound (the articulatory-perceptual system, AP) and meaning (the conceptual-intentional system, CI). FL is thus reduced to its bare, conceptual minimum as a system for generating sound-meaning pairs. Clearly, a minimal UG leaves a lot of gaps to fill if we are to retain anything like the empirical coverage that the maximalist GB view of UG afforded. Minimalism attempts to fill those gaps by looking outside of FL itself to find general computational and cognitive rationalizations of the kind of FL-specific technology uncovered in GB studies, thus reducing that technology to the level of *principled explanation*.² The object of study

¹ Such a view is in any case implausible given that an FL containing, e.g., Principle A of the Binding Theory or a null-subject parameter would hardly be competitively fitter than one without. To the extent that FL accords a survival advantage, as it surely does, it does so as a whole, i.e. as a system for generating hierarchically structured sound-meaning pairs allowing thought, planning, evaluation and, ultimately perhaps, communication through externalization.

² On MP as a rationalization of GB, see also Hornstein (2009), Lasnik & Lohndal (to appear).

thus shifts from data coverage (explaining judgements and intuitions) to the properties and principles of the language system itself, those identified under earlier PPT/GB models. The question to be answered is why the computational system of human language has *these* properties and not other ones, by finding principled explanations for these properties in terms of general factors nonspecific to the language faculty.

Pursuing this line of enquiry, Chomsky (2005) identifies three factors that determine the growth and form of FL as a mental organ:

- (1) *Three factors in language design*
 - 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 that constitutes the linguistic environment in which language acquisition takes place; and Factor III comprises “general properties of organic systems” (Chomsky 2004a: 1), the result of physical constraints on the form and development of living organisms which define, limit and channel the range of evolutionary options. Minimalism is then an attempt to remove as much technology as possible from the first factor, either by showing it to be redundant and thus eliminable altogether, or by moving it to the third factor by finding cognitive correlates in other domains. Perhaps surprisingly, then, and somewhat controversially, the Minimalist Program is interested in what the study of language has traditionally abstracted away from, namely those properties of language which are *not* unique to this faculty.

In the case of FL, the computational system of human language, such third-factor constraints might plausibly include principles of efficient computation. Thus notions of *economy* (of derivations and representations) – least effort, shortest movement steps, and so on – have played a large role in the development of the MP, as has the notion of optimality. If every property of FL contributes to the efficiency of the mapping to the interfaces with the external systems of AP and CI, then FL is in that sense an *optimal* solution to the conditions imposed on it by these external systems. The strong minimalist thesis (SMT) then holds that language is indeed perfect in this way – UG is maximally empty; the only conditions on FL are interface conditions; and FL satisfies these conditions optimally. By emphasizing the role of third-factor (“principled”) explanations in linguistic science, the role of the first factor (UG) is thus correspondingly diminished, since this is where the imperfections reside (the unexplained, unprincipled residue). In this way, the SMT takes us from methodological minimalism – the search for elegant, simple, economical, nonredundant theories common to all scientific enterprise – to a substantive claim about the object of study itself: FL is itself elegant, simple, economical, and nonredundant.

The addition of economy considerations to explanations of linguistic phenomena (i.e. for determining convergent derivations) is one of the key developments that sets MP apart from its forebears. It has also attracted considerable criticism, most notably in a series of provocative articles by Lappin, Levine and Johnson (2000a, to which various prominent minimalists replied, with further responses by Lappin, Levine & Johnson (2000b, 2001)). Two such criticisms have particular merit. The first is that the formal economy principles of earlier minimalism (up to Chomsky 1995) involve the comparison of derivations, where the most economical derivation, such as that involving fewer or less costly operations (cf.

Chomsky 1991), would win out over its less economical competitors. Such comparison only adds to the complexity of computations, contra the minimalist desideratum of easing the computational burden. Secondly, many of the formal economy constraints postulated in early minimalism have a rather language-internal, FL-specific flavour (e.g. those imposing locality constraints on movement, as Freidin & Vergnaud (2001: 644 [fn. 10]) note, and *Procrastinate*, which avoids overt categorical movement as a more costly operation than covert feature movement and which requires for its implementation the postulation of descriptive ‘strong’ features to enforce PF-convergence). These constraints must therefore be encoded as principles of UG after all rather than following independently as ‘third-factor’ effects. Comparing the situation with the status of minima and maxima principles in physics, Lappin, Levine & Johnson (2000a: 666) write: “Minimization/maximization principles are derived from deeper physical properties of the particles (waves, vectors, etc.) which satisfy them. They follow from the subatomic structure and attributes of these particles, and are not themselves basic elements of the theory”.

Both of these criticisms have been addressed in more recent developments of the minimalist program (Chomsky 2000). *Procrastinate* is eliminated in favour of single-cycle generation and earliness of feature-checking (cf. Pesetsky 1989), and all that remains of comparison of alternatives is a fully localized preference for ‘Merge over Move’ that immediately rules out any non-convergent options (see section 4; also Groat 1999 on localizing Merge over Move, and Collins 1997 on the localization of economy in general). Formal economy principles that might have been taken to be FL-specific additions to UG are now replaced with general third-factor properties (for example, instead of a specific “Shortest Attract” economy metric for constraining movement paths, we have a general “minimal search” condition on probes, implementing minimality effects – see section 3).

The only FL-specific economy conditions that now remain are interface conditions, in line with the SMT, essentially just Full Interpretation (FI) and Last Resort (LR), corresponding to economy of representation and economy of derivation, respectively (cf. Chomsky 2000: 99). Both FI and LR militate against superfluous elements in syntactic structures – symbols and operations with no interpretive effect at the AP and CI interfaces. Informally, we can characterize LR as the requirement “don’t do too much” and FI as “don’t do too little”. FI is “the convergence condition” (Chomsky 1995: 194), ensuring the legibility of syntactic expressions at the interfaces by barring features that are without an interpretation at the two interface levels, Phonetic Form (PF) and Logical Form (LF). Such uninterpretable features include Case features on nouns and verbal agreement features. LR adds a kind of inertia to the system, ruling out vacuous steps in derivations – all operations must be triggered, again by uninterpretable features. The picture of syntax that emerges is one of an otherwise inert system spurred into action by (i.e. operating on) uninterpretable features (uF). Without uFs, LR dictates that nothing would happen; once uFs are introduced, FI ensures that the system acts immediately to get rid of them, by movement and agreement, like a ‘virus’ triggering the immune system into action (cf. Uriagereka 1998).³ The twin, antagonistic interface-economy principles FI and LR

³ Optionality of operations is thus largely eliminated through the conspiracy of FI and LR (a triggered operation is obligatory, an untriggered one excluded). However, optionality may still arise in (at least) two ways. Firstly, the trigger (uF) might itself be optional, added to some derivations but not to others, creating derivational minimal pairs. Since, by FI, any uF must have an effect on output (Chomsky (1995: 294)), these pairs must be distinguishable at the interface by leading to different interpretations (for scope, discourse structure, and other ‘surface semantic’ properties) – see, e.g., Chomsky 2001 on

thus provide a rationale for the existence of uFs: without LR, operations would be free, without the need for a formal uF trigger (cf. Lasnik & Saito's (1992) Affect- α model), but without FI, uFs would be impotent and ineffectual as triggers, since there would be no need to eliminate them – the interfaces would simply tolerate them.

The resulting system implies a version of the *Activity Condition* (see section 3) – items are only syntactically active for case/agreement purposes for as long as they contain uFs. This captures the old idea that case and agreement go hand in hand (Martin 1999): once the uninterpretable Case feature on a nominal has been eliminated (via agreement with a functional head), that nominal is no longer active for agreeing with further functional heads (thus ruling out structures such as **John is believed [that t is happy]*, attributed to the Empty Category Principle in GB). The effects of the Case Filter are also immediately derived, without appeal to a special GB-style module – nominals must be Case-licensed by having their Case features eliminated, for reasons of FI. The effects of the GB Theta Criterion are also subsumed: (a) a single nominal cannot be Case-valued more than once, due to Activity, and so cannot relate to two functional heads, thus ruling out structures such as **John hit t_{John}*, where *John* would receive both the agent and patient theta-roles; (b) cases where a nominal fails to receive a theta-role, such as *Mary* in **John danced Mary*, reduce to the Case Filter and thus FI. In this way, the interface economy principles FI and LR become general conditions on optimal derivations that do the work previously attributed to multiple GB modules (including the Case Filter, the Theta Criterion and some instances of the Empty Category Principle⁴), allowing such GB machinery to be dispensed with (cf. Freidin & Vergnaud (2001)) and thus bringing us closer to the minimalist ideal of a maximally empty UG.

Such results are a valid achievement in refining linguistic theory – essentially, we take an empirically motivated GB module, like the Case Filter, and explain it in deeper, language-independent terms, thereby rendering FL a more biologically plausible object (in the sense defined above). Nevertheless, the validity of this enterprise has been challenged by many on the basis that the shift from GB to MP was not empirically motivated (cf. Lappin, Levine & Johnson 2000a), bringing no new empirical facts under the scope of PPT (Newmeyer 2003). Such criticisms seem to overlook two facts: firstly, that there are other criteria for the advancement of theories than just empirical payoff, and secondly, that there are other empirical domains for linguistic inquiry and explanation than just natural language data.

On the first point, Freidin & Vergnaud (2001) cite Dirac's (1968) distinction between two procedures for the development of hypotheses in the natural sciences: the "experimental" and the "mathematical". In its focus on explanatory depth rather than data coverage, the MP is firmly aligned within the mathematical procedure, sometimes also referred to as the Galilean style of science (the search for mathematical perfection in nature). Chomsky, in a discussion held with Riny Huybregts and Henk van Riemsdijk between November 1979 and March 1980 (reproduced in Chomsky 2004b), had the following to say on the matter, responding to Katz & Bever's (1976) paper "The Fall and Rise of Empiricism":

Object Shift, and Reinhart 1995, Fox 2000 on interface economy. The other possibility for optionality arises where multiple derivational options exist for satisfying the same formal imperative (uF) – see Biberauer & Richards 2006 for discussion.

⁴ Other phenomena explained by the Empty Category Principle in GB receive minimalist reinterpretations in terms of the locality of movement (minimal search and the phase impenetrability condition – see sections 3-4); these include Rizzi's (1990) relativized minimality effects, such as superraising, the Head Movement Constraint, and superiority (*wh*-islands) – see Chomsky 1995: 181.

I think that the title of the paper captures a rather significant point about the nature of the field. There is a strong opposition to the idea that there might be abstract theories and *principled explanations* [emphasis mine] and deductive chains of arguments from principles that do not look like descriptions of phenomena. The almost inevitable tendency of people when any such proposal is made is to try to shoot it down [...]. That is usually pretty easy; there is just so much unexplained data. (Chomsky 2004b: 70)

As we have seen, the aim of MP is to do just this – to deduce GB-style machinery from deeper principles. Should this be achieved (or achievable), then the MP will actually be explaining *more* than GB, since not only will there be no loss of data coverage, but the tools used to explain those data will themselves be explained.⁵ Thus, by reconceptualizing barriers as phases (see section 4 below), movement as Internal Merge (see section 2), government as Probe-Goal (section 3), the Case Filter as (reducible to) FI, etc., we bring GB tools *and the phenomena they accounted for* under the fold of principled explanation – a significant result, but one that is apparently missed by those who dismiss phases (etc.) as just barriers *redux*. These are not merely cosmetic or terminological changes, but conceptual ones representing a considerable step forward in our understanding of these properties, especially if the restatement in minimalist terms allows problematic aspects of the original conceptions to be resolved (thus the stipulative voiding of barrierhood by adjunction receives a natural explanation in terms of the *edge* in phase theory, which is a notion independently implied on grounds of optimal design – see section 4.3 below).

The goal of the MP, then, is to reconceive as much descriptive technology as possible in this way, in order to meet more stringent explanatory demands. However, there is no expectation that we will be able to explain *everything* in these deeper terms. Nobody anticipates that the SMT should hold in its strongest form (a totally empty UG – see below). Rather, the SMT provides us with a heuristic that draws a principled line between those properties and phenomena that *are* amenable to principled explanation in terms of third-factor considerations, and those that are not (at our present level of understanding) and which must therefore be attributed to UG, complicating the evolutionary picture.

Returning to the second point mentioned above, that of the empirical basis of the MP, it should be borne in mind that the MP is no less empirically founded than its predecessors. The object of study might have changed from external phenomena (language data) to the properties of the mind responsible for those phenomena (the properties of I-language); however, the latter is still an empirical object in the natural world (indeed, more so than the tokens of E-language; see, e.g., Chomsky 2001: 41-42, Anderson & Lightfoot 2002). To properly understand this object, data can only take us so far; the MP investigates the possibility that a better theory of FL can be attained by taking conceptual, cognitive, and computational factors into account than that which is arrived at on the basis of data alone.

Nevertheless, it cannot be denied that the MP has had considerable success on the empirical stage, too. Not only has it yielded refined analyses of familiar phenomena (such as expletive-associate constructions, relativized minimality effects, crosslinguistic differences in verb placement, etc.), but it has also allowed new

⁵ As Hornstein, Nunes & Grohmann (2005: 256) neatly put it, “a minimalist reanalysis of the data [...] need not cover more empirical ground to be preferable: a tie goes to the minimalist!”

phenomena to be described and explained that were formerly beyond the scope of GB analysis (despite the critical assertions to the contrary mentioned above). Some of these uniquely MP-spawned empirical consequences will be reviewed in the following sections.

Given the impossibility of covering almost two decades' worth of wide-ranging minimalist research in a single, short survey, the remainder of this article must be necessarily superficial and selective in content, leaving out far more than it can include. In order to give just a sense of the fruitful directions in which the MP has taken us since Chomsky 1993, the focus will be on the major developments in Chomsky's own formulations and conceptions of the field (though given the richness of these works, the overview will still be far from exhaustive). To shape the presentation, we should first identify what must remain of UG (the first factor) under the reductive, SMT-driven approach. Here, considerations of *virtual conceptual necessity* come into play – what must a maximally simple FL minimally comprise in order to be usable at all as a system for generating convergent sound-meaning pairs?

In order for the syntax to provide legible expressions to the two external performance systems with which it interfaces, at least two primitive operations would seem indispensable. Firstly, there must be an operation for building structured expressions; secondly, there must be an operation connecting those expressions with the external systems. This would seem the conceptual bare minimum. The first operation has become known as *Merge*, the second as *Transfer*. Further, it is generally assumed that a third operation must exist, the dependency-forming operation *Agree*, which arguably plays a role both in Merge (e.g. for implementing movement) and in Transfer (for deleting uninterpretable features before they reach the interface). Each of these operations is associated with a particular type of formal feature that triggers it – Merge operates on Chomsky's (2007, 2008) structure-building Edge Feature (EF); Agree and Transfer are triggered by uninterpretable (ϕ -)features (case and agreement features), uFs.

In the following, we review each of these basic operations and feature types in turn: Merge and EF (section 2); Agree and uF (section 3); Transfer and phases (section 4). Although the emphasis will be on their conceptual motivation and development, major empirical consequences that go beyond the capabilities of GB will also be highlighted. Finally, section 5 concludes with a brief mention of some ongoing debates and current issues that are shaping the immediate development of the program and which bear on the future prospects of minimalism as providing a viable model of the human language faculty.

2. MERGE AND THE EDGE FEATURE

2.1 Bare phrase structure

By the GB era, the phrase-structure rules of earlier generative grammar had largely been eliminated as redundant, duplicating information already encoded in lexical information (e.g. subcategorization frames) and the X-bar schemata to which they conformed, principally those given in (2), where YP corresponds to the structural notion of *specifier* (of XP) and WP to *complement* (of X).

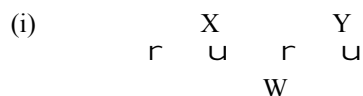
- (2) a. XP → (YP) X'
 b. X' → X (WP)

X-bar theory provided representational constraints on the form of phrase structure, a template into which lexical material could be inserted. What was missing, however, was a procedure for deriving those representations. Chomsky 1993 (1995: 189-191) sets the scene for a derivational, bottom-up view of phrase structure in line with minimalist desiderata by proposing a return to an elementary form of generalized transformations (albeit with an added notion of cyclicity – Chomsky 1995: 190; cf. below). A binary operation maps a pair of phrase markers (K, K^1) to the new phrase marker K^* ; a singulary operation maps the phrase marker K to K^* , where K^* includes K as a proper subpart (that is, we *extend* K). These operations would in later work become *Merge* – external and internal, respectively.

At its simplest, Merge takes two syntactic objects, X and Y , and combines them to form the unordered set $\{X, Y\}$, which can then provide the input to further instances of Merge (forming $\{W, \{X, Y\}\}$, and so on). In merging X to Y , there are two logical possibilities: either X originates inside Y (as a proper subpart) or else it originates outside Y (e.g. in the Lexicon or Numeration – see (3) below). The latter option does the work of erstwhile phrase-structure rules, forming predicate-argument structures and the like, whereas the former yields movement without further ado, thus implementing the transformational component of GB and earlier generative models.⁶ Interestingly, then, the displacement property of human language, long viewed as an imperfection of FL owing to its lack of conceptual necessity (cf. its absence from logical and artificial languages), is immediately given under Merge; in a twist of perspective, its absence or exclusion would now require departure from the SMT (Chomsky 2008: 7).

Since both kinds of Merge interleave freely throughout the derivation, the two internal syntactic levels of d-structure and s-structure, and the compositional cycle that mapped the former to the latter in GB, are eliminated as unformulable. This is a

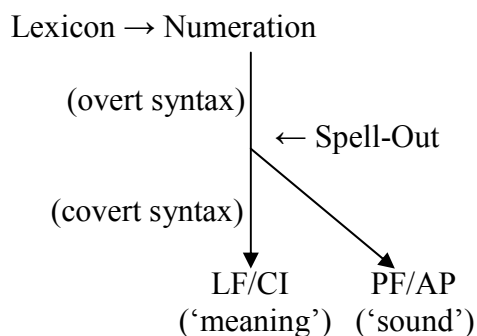
⁶ Other possible ways of combining X and Y are arguably excluded as departures from this minimal conception of Merge and the SMT (i.e. if they exist, they must be extensions to UG). For example, *Parallel Merge* (Citko 2005) would concatenate X with Y through a combination of External Merge (taking X and Y as separate objects) and Internal Merge (merging X with a subpart of Y), yielding structures like (i).



Parallel Merge/(i) is claimed by Chomsky (2007: fn. 10) to be a departure from the SMT: “Merge cannot create objects in which some object [W] is shared by the merged elements X, Y .” This exclusion is not uncontroversial, however. Such a variant of Merge has considerable empirical support; Citko (2005) shows how it yields an advantageous analysis of Across-The-Board *wh*-movement (as in *What did John recommend and Mary read?*), such that a single *wh*-item is extracted from multiple conjuncts. Further, insofar as it is X that is the selecting head here (i.e. the head whose Edge Feature is satisfied through Merge), Parallel Merge conforms to the Extension Condition and/or the No-Tampering Condition (see below for discussion of these concepts). However, Parallel Merge should perhaps still be excluded on the grounds that (i) is not strictly a combination of X and Y , locally speaking, but rather a combination of X and W , where W is contained in Y . It does not yield the set $\{X, Y\}$, and so is not an instance of Merge(X, Y). Internal Merge and External Merge thus remain the only two logical possibilities for Merge(X, Y) *per se* – either X is contained in Y or it is not. (This objection remains silent, however, as to whether (i) is a possible instance of Merge(X, W), which is all that is crucial for Citko. It is questionable, however, whether the double-rooted object in (i) that results from Parallel Merge is a viable syntactic object for further derivational operations – see Epstein et al 1998 and footnote 12 below.)

desirable result, since the only conceptually necessary levels of representation are those interfacing with the external systems of AP and CI, i.e. LF and PF, respectively.⁷ We thus arrive at the following architecture of the grammar, a refinement of the GB T-model.⁸

(3) The minimalist architecture of the language faculty



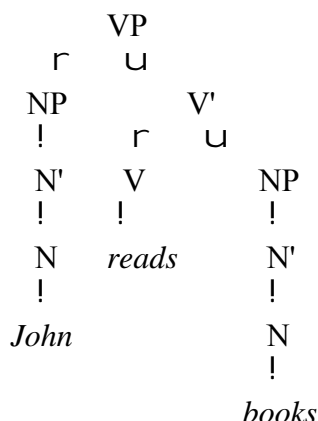
Furthermore, templatic X-bar restrictions are also eliminated by Merge – there are no obligatory intermediate projections, and so no trivial (unary) projections. Instead, what emerges from Merge is a *bare phrase structure* in which no structure exists independently of the lexical items that project it (thus structural artefacts of X-bar representations, such as bar levels, are also eliminated, as is the head–terminal distinction and the need for separate lexical insertion rules). Such a model thus conforms to a further minimalist desideratum, *Inclusiveness*, which requires that no new features be added in the course of the derivation beyond those already present on the lexical items in the Numeration – only rearrangements and deletions of the features of these items should be possible (Chomsky 2000: 113).

To illustrate the differences, the two trees in (4) represent the structure of the simple verb phrase *John reads books* under X-bar theory (4a) and under Merge/bare phrase structure (4b). (We assume here for purely expository purposes a naïve VP-internal subject position, and treat nominals as NPs.)

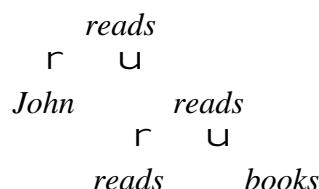
⁷ Indeed, even these should be eliminated from a strictly derivational system, in favour of constant, unrestricted access by the interfaces, as Epstein et al 1998 argue, paving the way for phase-cyclic computation (see section 4).

⁸ The Numeration (lexical array) in (3) is a collection of lexical items selected from the lexicon that defines the workspace of a derivation. It is motivated on numerous grounds in Chomsky 1995 (Chapter 4): it ensures that PF and LF expressions/representations are drawn from the same vocabulary and are thus compatible (so that we do not get arbitrary mismatches between sound and meaning); it defines the *reference set* of derivations that compete for economy purposes (for example, we do not want the availability of less costly expletive constructions (*there was heard a loud explosion*) to uniformly block overt subject-raising derivations (*a loud explosion was heard*); therefore it must be that only those derivations are compared that proceed from the same numeration – in this case, a Numeration containing the expletive is not comparable with one that does not); it allows for a type–token distinction to be made, thus distinguishing multiple selections of the same lexical item (e.g. coreferential *he* in *he thinks he is happy*) from multiple occurrences arising through movement (e.g. *he seems t_{he} to be happy*), which are clearly treated differently at the AP interface. Furthermore, it is assumed that, for a derivation to be sent to the interfaces for interpretation (to AP at the point marked ‘Spell-Out’ in (3), and to CI at the end of the covert cycle), the Numeration must be exhausted (Chomsky 1995: 226). Spelling-out before the Numeration is exhausted would again compromise compatibility between PF and LF representations, since lexical items could be added in the covert computational cycle, yielding, e.g. a PF *John left* with an LF interpretation *They wonder whether John left before finishing his work* (Chomsky 1993 (1995: 189)).

(4) a. X-bar



b. Merge (Bare Phrase Structure)⁹



The bar levels of X-bar theory can be read off the structure in (3b) as derived, relational properties (maximal/nonmaximal, minimal/nonminimal). Similarly, *complement* and *specifier* now simply refer to first-merged and second-merged items.

2.2 The Extension Condition

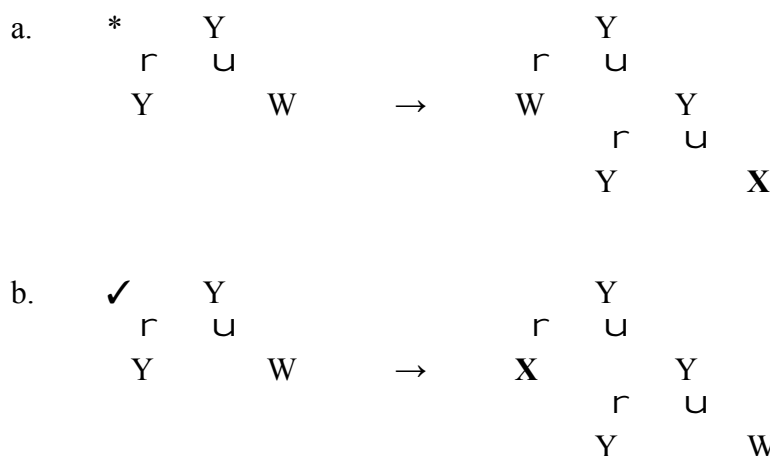
Given Last Resort (see section 1), all operations require a formal trigger in the form of an uninterpretable feature. Chomsky (2007, 2008) proposes that the feature responsible for Merge is the *Edge Feature* (EF), a property of lexical items. Arguably,

⁹ As numerous researchers have pointed out, bare phrase structures such as (4b) are more problematic than the richer representations in (4a) in one key area: linearization. The base pair $\{\alpha, \beta\}$ of every subtree stands in a symmetric c-command relation (sisterhood) that cannot be ordered under Kayne's (1994) *Linear Correspondence Axiom* (LCA), which essentially maps asymmetric c-command (sister-of-contain) onto precedence. Thus *reads* precedes *books* in (4a) by virtue of the asymmetric c-command relation between the V and N' (/N) nodes, but no such asymmetry holds between the heads *reads* and *books* in (4b). If we adopt Bare Phrase Structure, then we are forced to conclude that the LCA cannot be a constraint on phrase structure itself as a property of narrow syntax, but must be a linearization strategy operative only after Spell-Out in the mapping of syntactic hierarchy onto phonotemporal order in the PF wing of the grammar. As Chomsky 1995: 340 puts it: "We take the LCA to be a principle of the phonological component that applies to the output of Morphology". This unlinearizable point of symmetry must then be resolved at (or by) PF, to which end numerous proposals have been made in the literature. If movement leaves behind an empty category (trace) that does not need linearizing, then movement of one of the two offending sisters is one option (Kayne 1994: 133, Chomsky 1995: 337, Moro 2000). Other possibilities include *cliticization* of one of the sisters to the other (via head-adjunction in the syntax, leading to word-internal restructuring in the morphological component – Chomsky 1995: 337, Uriagereka 1998, Nunes 1999), or a head-complement directionality parameter of the kind familiar from GB (see Saito & Fukui 1998 for a syntactic parametrization of Merge, and Groat 1997, Epstein et al 1998, Richards 2004 for parametric approaches located at the PF-interface).

EF is the single evolutionary innovation that takes us from the pre-linguistic stage of inert, isolated lexical items (concepts) to a system that enables their combination into larger objects, and those larger objects with further lexical items, and so on – that is, EF yields iterative Merge, and thus recursion (“a discrete infinity of structured expressions”, Chomsky 2007: 5).

The Edge Feature captures a further important property of Merge, namely its monotonicity: whilst Merge adds new structural properties (relations) to the two objects merged together (X and Y), namely sisterhood and c-command (as a result of Merge, X and Y are sisters, and X c-commands into Y and vice versa), all previously existing structural properties and relations remain unchanged. Such conservation of information is arguably a natural computational principle (a third-factor consideration; see Lasnik, Uriagereka & Boeckx 2005 on conservation laws); Chomsky (2007, 2008) calls it the *No-Tampering Condition*, NTC.¹⁰ It follows from the NTC that Merge involving the syntactic object X (either a lexical item or the output of previous applications of Merge) must always apply to the ‘edge’ of X, extending the tree upwards, hence the *Edge Feature*. In this way, EF captures the *Extension Condition* on Merge, barring countercyclic Merge operations that would take X and Y and merge X to W contained in Y, replacing W with {X, W} (Chomsky 1995: 248). More generally, X cannot merge *inside* Y, altering the sisterhood relations of Y, as in (5a); rather it must merge *outside* Y (i.e. to its edge), as in (5b).

(5) Merge X to the Edge:



[based on Chomsky 2000: 136 (57)]

Empirically, the Extension Condition is borne out in the form of Relativized Minimality effects (Rizzi 1990), since without it, countercyclic derivations would be possible in which the intervener is merged *after* the movement operation takes place which the intervener is meant to block. Chomsky 1993 (1995: 190) illustrates with the following examples:

- (6) a. [_T seems [_T is certain [_{JP} John to be here]]]
 b. [_{C'} C [_{VP} fix the car]]
 c. [_{C'} C [John wondered [_{C'} C [_{IP} Mary fixed what how]]]

¹⁰ An explicit formulation is given in Chomsky 2000: 137 (59): “Given a choice of operations applying to α and projecting its label L, select one that preserves $R(L, \gamma)$ ”.

Without the Extension Condition (cyclicity of Merge), the availability of the intermediate structures in (6) would allow violations of superraising, the head movement constraint (HMC), and the *wh*-island constraint, respectively, to be derived. Thus, in (6a), *John* could move to matrix Spec-IP across the empty embedded finite subject position; merging *it* to that position after movement of *John* to the matrix clause would then yield the illicit superraising violation **John seems it is certain t to be here*. Similarly, moving *fix* to C followed by merger of the auxiliary *can* in (6b) would yield the HMC-violating question **Fix John can t the car?*, and movement of *how* to matrix Spec-CP could precede movement of *what* to embedded Spec-CP in (6c), falsely deriving a *wh*-island violation (**How did John wonder what Mary fixed t_{what} t_{how}?*). The minimalist ‘shortest move’ economy principles (Minimal Link Condition) that would rule out such minimality violations must therefore be bolstered by the notion of the *strict cycle* (Chomsky 1973), applied to Merge in the form of an extension condition, without which they would be ineffective.¹¹ The Edge Feature immediately ensures that Merge conforms to the NTC, thus barring countercyclic Merge. Furthermore, in excluding countercyclic Merge to a position inside the complement of a head, the EF/NTC also subsumes much of the work done by the Projection Principle and Theta Criterion in GB (the ban on raising to complement positions); Chomsky 1995:191. In this way, we once again see the SMT in action: heterogeneous and overlapping GB technology is shown to be redundant, reducible to a much smaller set of general, simple, FL-independent computational principles (here, the NTC).

Before turning to a further advantageous result of the NTC in the next section (the copy theory of movement), two additional empirical consequences of EF should be remarked. The first is that, in order to allow for specifiers to be projected (second-merge), EF must remain undeleted in the syntax, so that it can be satisfied a second time (Chomsky 2007: 11). However, since it remains undeleted, restriction to just a single specifier would require a stipulation. Thus, under EF, if one specifier is possible, then any number of specifiers are possible. Multiple specifiers therefore come for free from undeletable EF (in sharp contrast to X-bar theory and approaches based on Kayne’s (1994) LCA; see footnote 9 on the latter). This is a desirable consequence, as multiple specifiers have played an important part in minimalist analyses of various phenomena, including Object Shift, in which the light verb head *v* projects both its usual thematic specifier, to which the external argument is merged, and a nonthematic specifier to which the object is raised (Chomsky 1995: 352); Transitive Expletive Constructions, analysed in Chomsky 1995:342-344 as involving multiple specifiers of T, one hosting the raised external argument and the other the expletive; and, under phase theory (section 4), successive-cyclic movement of items out of phases via multiple specifiers of CP and *v*P.

Secondly, although Merge via EF sharply constrains the range of possible Merge sites for X merging to Y, there is room for at least a little flexibility precisely in the case of multiple specifiers. This wiggle room is exploited by N. Richards 1999, who argues that a conspiracy of economy considerations (Attract Closest and Shortest Move) should result in the “tucking in” of an outer specifier under an inner specifier when both are the product of movement. That is, via tucking-in, multiple XP-

¹¹ Brody 2002 invokes this reliance on the supplementary notion of cycle/extensionality as an argument against the derivational approach of minimalism and in favour of purely representational alternatives. However, as argued above, the cyclicity of Merge follows from independent principles of structural optimality (the NTC), and so would not be an addition to UG but a third-factor effect.

follows from this that the operation of movement (internal Merge) cannot insert traces into base positions, as it did under GB's Move- α . Nor can the base position simply be entirely vacated, since this would tamper with the structural relations established by the moved item in its original position (sisterhood with its original merge-partner). Rather, movement must leave the original position unchanged, implying that when an item moves, it also stays behind. That is, under NTC, we arrive at a *copy theory of movement*, such that a single item becomes associated with multiple positions, leaving copies of itself in each position through (to, from) which it moves. Since this has important empirical consequences that extend the range of data coverage in MP vis-à-vis GB, let us consider how copies work in some more detail.

2.3 Empirical advantages: Copy deletion and Spell-Out

The traces left by movement in GB violate Inclusiveness/NTC in at least three ways: (i) they modify the structure, replacing the lexical material of the moved item with a new kind of element (an empty category, of which there are various kinds depending on the type of movement); (ii) they are coindexed with the moved item, and this index is itself a further violation of Inclusiveness; (iii) they are not part of the Numeration but rather are introduced only in the course of the derivation, generated by the movement operation itself. Instead, as described above, the copy theory of movement emerges as the 'null assumption': "Internal Merge leaves a 'copy' in place" (Chomsky 2004a: 8), creating multiple occurrences of the same item. In (8), *John* merges (at least) twice: once as V's complement, once as T's specifier.

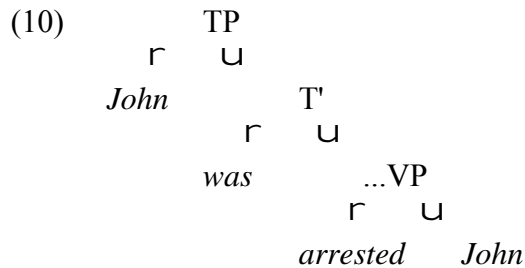
- (8) a. John was arrested
 b. [John was [arrested (John)]]

In the case of *overt* movement, it is assumed that the lower copy is deleted in the phonological component (PF/AP); however, it remains available for interpretation at the semantic interface, yielding an advantageous implementation of *reconstruction* without the need for lowering and trace-replacement operations:

- (9) a. John_i wondered [which picture of himself_{i/j}] Bill_j saw t_{wh}
 b. John wondered (which picture of himself) Bill saw (which picture of himself)

Setting aside certain details (see Chomsky 1993 (1995: 202-210)), the two interpretations of *himself* indexed in (9a) correspond to interpretation of different copies of *pictures of himself* at LF. (Reference to syntax-internal levels, such as S-structure or van Riemsdijk & Williams's (1981) NP-Structure, is thus no longer necessary, in line with minimalist desiderata and the architecture in (3).)

Comparing the different treatment of copies at the two interface levels, the question arises as to *why* only a single copy of the moved item is pronounced (i.e. why the phonological features of all other copies are deleted at PF). The most successful and influential account of why more copies are not pronounced is that of Nunes (1999, 2004), who approaches the problem from the perspective of linearization and Kayne's (1994) *Linear Correspondence Axiom* (LCA; see footnote 9). The LCA maps asymmetric c-command onto precedence. To illustrate, the simplified post-movement structure of (8) is given in (10).



The top copy of *John* asymmetrically c-commands *was* (since *was* is contained inside *John*'s sister, T'); in turn, *was* asymmetrically c-commands the lower copy of *John*. By the LCA, the ordering instructions determined at PF therefore include:

- (11) <John, was>, <was, John>

This means that *John* must both precede *was* and follow *was*, an ordering paradox violating the asymmetry requirement on linear ordering. Nunes's suggestion is that the phonological features of one of the copies of *John* must be deleted, thus exempting it from the need to be linearized.¹³

A unique prediction of the copy theory of movement under Nunes's analysis is that, should the LCA be overridden or not apply (e.g. because an alternative linearization strategy is available), more than one copy may be realized. This sets it sharply apart from the trace theory of movement, in which lower chain links are intrinsically devoid of phonological content by virtue of being empty categories. The copy theory would therefore find empirical confirmation over the trace theory if evidence of multiple spell-out of chain links (realization of multiple copies) could be found.

Such evidence, Nunes suggests, comes from *wh*-copying in a variety of languages, in which intermediate 'traces' of successive-cyclic *wh*-movement may be overtly realized:

- (12) a. ***Wen*** glaubst du, ***wen*** sie getroffen hat? [German]
 Who-ACC believe you, who-ACC she met has
 b. ***Mit wem*** glaubst du, ***mit wem*** Hans spricht?
 with who-DAT believe you with who-DAT Hans talks
 c. ***Wovon*** glaubst du, ***wovon*** sie träumt?
 Whereof believe you whereof she dreams
- (13) ***Wêr*** tinke jo ***wêr***'t Jan wennet? [Frisian]
 Where think you where-that Jan resides
- (14) a. ***Waarvoor*** dink julle ***waarvoor*** werk ons? [Afrikaans]
 Why think you why work we
 b. ***Met wie*** het jy nou weer gesê ***met wie*** het Sarie gedog ***met wie*** gaan
Jan trou

¹³ Nunes (1999: 229) proposes a metric of Formal Feature Elimination to explain why it is usually the lower copy that is deleted; essentially, the copy with the most unvalued features is deleted. Note that this implies a dedicated operation, Copy, that creates new copies that can be distinguished between on the basis of their individual featural properties, rather than treating copies as multiple instances of the self-same item with identical features in every position. (See Chomsky 2008: fn. 16 on the distinction.)

With who have you now again said with who has Sarie thought with
who go Jan marry

- (15) *Kas misline kas o Demiri dikhlâ?* [Romani]
whom you-think whom Demir saw

Crucially for Nunes's linearization account, such *wh*-copying is only possible with simple pronominal forms (*wh*-pronouns) and not with morphologically complex, full *wh*-phrases:

- (16) a. **Wessen Buch glaubst du wessen Buch Hans liest?* [German]
Whose book believe you whose book Hans reads
b. **Welchen Mann glaubst du welchen Mann sie liebt?*
Which man believe you which man she loves
c. **Save éhave mislinea save éhave o Demiri dikhlâ?*

Assuming that the simplex intermediate *wh*-copies in (12)-(15) can undergo morphological reanalysis under adjacency or via head-adjunction in the syntax, yielding a [wh+C°] adjunction complex that is restructured into a single phonological word, such *wh*-copies are no longer subject to the LCA, which does not apply word-internally (Chomsky 1995: 337). The illegitimacy of (16) then follows if we assume that XPs (maximal projections) cannot undergo adjunction to heads in the syntax (Chomsky 1995: 319), hence no morphological reanalysis is possible in the case of complex *wh*-copies.

Converging evidence of multiple copy realization in cases where independent PF requirements override the LCA comes from the domain of verb copying (see, e.g., Aboh 2004, Landau 2007). Many languages exhibit 'verb-focusing' ('predicate cleft') structures in which the verb is fronted and interpreted contrastively.

- (17) a. **Ðù** Séná **ɖù** blédjì ló bléblé [Gungbe]
eat Sena eat bread DET quickly
'Sena ATE the bread quickly.'
b. **Fón** Séná **fón** bléblé
stand Sena stand quickly
'Sena STOOD UP quickly.'
c. **Ná** Séná **ná** kwé vǐ ló
give Sena give money child DET
'Sena GAVE the child some money.'
d. **Wá** Séná **wá**
arrive Sena arrive
'Sena ARRIVED.'
e. **Nyòn** náwè ló **nyòn** hwànpè
know woman DET know beauty
'The woman IS beautiful'

- (18) **Fifún** ni Tolú **fún** mi ní igbá [Yoruba]
giving COP Tolu gave me CASE calabash
'Tolu GAVE me the calabash.'

- (19) **Dumat'** o ženit'be on **dumaet**, no nikogda on ne ženitsja [Russian]
 Think-_{INF} about marriage he thinks, but never he not get-married
 'He does THINK about marriage, but he'll never marry.'
- (20) **liknot**, hi **kanta** et ha-praxim [Hebrew]
 buy-_{INF} she bought ACC the-flowers

Child language 'errors' display a similar phenomenon, in the form of auxiliary doubling (Nunes 1999, Landau 2007):

- (21) a. Why **did** the farmer **didn't** brush his dog?
 b. What kind of bread **do** you **don't** like?
 c. Why **could** Snoopy **couldn't** fit in the boat?

Landau (2007) proposes that both copies have to be pronounced in cases such as (17)-(20) as each satisfies a distinct PF requirement: the top copy bears a high-pitch accent, signalling focus, whilst the lower copy (in T) bears inflection (cf. *do*-support in English). Whatever the correct analysis turns out to be, and whether or not V-copying can be unified with *wh*-copying, it is clear that such phenomena become far more transparent under the copy theory of movement than they would be under trace theory.

With multiple copies realizable under certain PF-defined circumstances, a further possibility that arises once movement is viewed as copying is for different parts of each copy to be realized (so-called *scattered deletion*; cf. Bošković 2001), rather than one single entire copy. Fanselow & Ćavar (2002) propose such an analysis for certain Left-Branch Extraction effects ('split constituents') in Slavic and Germanic (see also Bošković 2005):

- (22) a. [**Crveno auto**] je on [~~erveno~~ **auto**] kupio [Croatian]
 Red car is he red car bought
 b. [**Na kakov krov**] je Ivan [~~na kakov~~ **krov**] bacio loptu [~~na kakov~~ **krov**]
 On what-kind-of roof is Ivan on what-kind-of roof ball thrown ...

Of course, scattered deletion must be constrained so as not to overgenerate and yield unacceptable left-branch extractions such as those in (23b-d).

- (23) a. [**Visoke djevojke**] je on vidio [~~visoke~~ **djevokje**]
 Tall girls is he seen tall girls
 b. *? [**Visoke djevojke**] je on vidio [~~visoke~~ **djevokje**]
 c. * [**Visoke lijepe djevojke**] je on vidio [~~visoke~~ **lijepe djevokje**]
 Tall beautiful girls he is seen tall beautiful girls
 d. * [**Visoke lijepe djevojke**] je on vidio [~~visoke~~ **lijepe djevokje**]

To this end, Bošković (2001) proposes that scattered deletion is characterized by the kind of inertia we attributed to the syntax in section 1: operations do not apply freely, but only if forced to do so. Thus scattered deletion might best be viewed as a last-resort PF strategy that applies only if independent PF constraints block full pronunciation of the top copy (i.e. full deletion of lower copies). Such a constrained approach has been usefully applied to a range of languages and phenomena (cf. Lambova 2004 on participle-auxiliary orders in Bulgarian).

In sum, considerations of structural optimality and computational simplicity (monotonicity, the NTC) lead to a simpler approach to movement (as copying) that finds compelling empirical substantiation. Other notable empirical applications of the copy theory of movement include Grohmann 2003 on resumption as lower-copy spell-out, Fujii 2005 on copy-raising in English, and Bobaljik 2002 on Holmberg's Generalization and covert movement as lower-copy spell-out in a "single output" model of the syntax.

3. AGREE AND UNINTERPRETABLE FEATURES

3.1 Probe-Goal

A unified picture of subject and object agreement had started to emerge by the end of the GB era. In earlier GB, subjects were assigned nominative case by the finite inflectional head I (combining Tense and Agreement) in the Spec-IP position; this Specifier-Head relation was then also responsible for subject-verb agreement. Objects, on the other hand, received case via government by the lexical verb. These differing configurations for subject versus object case corresponded to the difference between subjects and objects with regard to verbal agreement in languages like English. However, work such as Baker (1988) on Chichewa and other languages showed that object agreement was also a possibility; around the same time, Kayne (1989) and Christensen & Taraldsen (1988) had argued that participle agreement with objects in French and Scandinavian was contingent on movement, indicating a specifier-head relation for object agreement lower in the clause. In order to unify participle agreement with finite verb agreement, Kayne (1989) postulated an Agr head for the former. Pollock (1989) also identified a low Agr head as the landing site of short-distance infinitive raising in French. This projection was adopted as AgrOP in Chomsky 1991 (Chapter 2 of Chomsky 1995). Completing the symmetry, subject agreement was now attributed to an AgrS head, following Pollock's (1989) splitting of IP into AgrS and Tense.

Since case and agreement were now licensed together, for objects and subjects alike, in uniform specifier-head configurations, the GB notion of case assignment was replaced in early minimalism with one of checking (of case and agreement features) with functional heads. Thus nouns and verbs enter the derivation fully inflected for case and agreement features: nouns bear uninterpretable Case features as well as their inherent, interpretable phi-features (person and number), whilst verbal features include uninterpretable phi-features.¹⁴ Agreement then becomes a process of checking these uninterpretable features with matching ones in a local checking configuration defined by a functional head. Movement therefore feeds agreement, since checkees have to move into a relevant checking domain in order to check their features. This movement could be either overt or covert, depending on whether it took place before or after Spell-Out (cf. (3)). This in turn was determined via the postulation of 'strong' categorial features on functional heads. Since these were stipulated to be PF-uninterpretable, such features must be checked pre-Spell-Out, resulting in overt movement of verbs and arguments into specifier-head checking configurations. In the absence of a strong feature, checking would be delayed until the covert component,

¹⁴ This lexicalist view of Case/agreement-licensing has the additional advantage of eliminating lowering operations, such as Affix-Hopping in English.

where formal features alone would raise and adjoin to the relevant functional head (in accordance with the least effort principle *Procrastinate*; see section 1).

In this way, differences in overt word order could be captured as parametric differences in the featural composition of functional heads. For example, the requirement in languages like English that the derived subject position (Spec-IP) be overtly filled in finite clauses (cf. (24)), known as the *Extended Projection Principle* (EPP), could be formally stated as a strong D-feature on the head I.¹⁵ Similarly, the difference between English and French in terms of verb movement to I (cf. (26)) could be captured by saying that English I has a weak V-feature and French a strong V-feature. (See Adger 2003 for a summary of many further crosslinguistic parametric differences stated in terms of feature strength.)

- (24) a. There appeared **a face** at the window
 b. **A face** appeared at the window
 c. * Appeared **a face** at the window
- (25) a. John often [_{VP} **kisses** Mary]
 b. *John **kisses** often [_{VP} *t_V* Mary]
- (26) a. *Jean souvent [_{VP} **embrasse** Marie] (French)
 b. Jean **embrasse** souvent [_{VP} *t_V* Marie]

The alternation in (24a-b) has had a particular impact on the development of minimalist models of checking and agreement (see Boeckx 2006: 186-190 for a survey of the earliest and intermediate stages). The question of greatest concern here is how subject agreement is established between the associate argument (*a face* in (24)) and the agreement head (I/AgrS/T). (That the verb agrees with the associate can be seen in, e.g., *There seem(*s) to be several men in the garden.*) In (24b), agreement is established overtly in a specifier-head configuration through raising of the associate the EPP/subject position. In (24a), it is established covertly, via raising of agreement features at LF. However, as pointed out by den Dikken (1995), the raised features do not yield new binding possibilities at LF (**There seem to each other to be some applicants eligible for the job*), nor can they license negative polarity items (**There seem to any of the deans to be no applicants eligible for the job*). This perhaps suggests that no raising in fact takes place, even covertly.

At the same time, these structures throw up questions as to where the trigger for movement lies. Last Resort (section 1) dictates that there must be an uninterpretable feature (a morphological deficiency) somewhere that drives the movement. Given the well-formedness of (24a) with the associate *a face* remaining in situ (overtly), it cannot be any deficiency on the associate that forces it to move in (24b). Rather, as we have seen, movement in (24b) is driven by the EPP – the associate satisfies a strong feature of the *target* head; it thus moves for ‘altruistic’

¹⁵ The EPP of GB remains a poorly understood condition, resisting principled explanation. Various researchers have tried to eliminate it (cf. Castillo, Drury & Grohmann 1999) or reduce it to other mechanisms, such as Case (Martin 1999, Epstein & Seely 2006). Perhaps the greatest insight and innovation in minimalist work on the EPP is to be found in Alexiadou & Anagnostopoulou (1998), who propose that rich verbal agreement in null-subject languages like Greek and Italian is essentially nominal (bears a D-feature), enabling T’s EPP feature (strong D-feature) to be checked via movement of the finite verb (V-to-I) in these languages, extending the phenomenology and parametrization possibilities of the minimalist EPP in interesting new directions (some of which are explored in Richards & Biberauer 2005).

reasons. On the other hand, Chomsky (1993) argues on the basis of examples like (27) that movement must be for reasons of *Greed*, i.e. to satisfy a deficiency on the moving XP.

(27) *There seems to [a strange man] [that it is raining outside]

Case on *a strange man* is licensed by the prepositional head *to*; it therefore has no further need to move or do anything. Consequently, it is unable to raise overtly to check agreement with the matrix T/Infl head. Were properties of the target alone enough to license movement (and thus agreement) at LF, this sentence should converge, wrongly. This suggests that Greed is at stake. Yet Greed alone cannot explain (24b). Lasnik (1993) suggests a compromise view, *Enlightened Self-Interest* (ESI), in which the uninterpretable feature that is checked through movement may be located either on the target or on the moving item. The illegitimacy of (27) then suggests a further restriction such that case-checked nominals are inert for further movement and agreement operations (Lasnik 1995).

These empirical considerations, combined with the conceptual problems associated with the computational complexities of Greed and Procrastinate (lookahead and global comparison of derivations; cf. Chomsky 1995: 201, 261), indicated that a rethink of checking theory was in order. The simple system that replaced checking – Probe-Goal Agree of Chomsky 2000 – addresses all the above problems, dispensing with associate movement in expletive constructions and replacing ESI with the Activity Condition: *both* items entering an agreement relation must have as yet unsatisfied featural requirements (i.e. both must at least have the *potential* of having a feature satisfied by the agreement operation in question), allowing agreement to be asymmetric (as under Greed and ESI) but also symmetric (as under ESI).

Under Probe-Goal Agree, uninterpretable features are modelled as features that lack a value (addressing problems raised by Epstein & Seely (2002)). Unvalued phi-features on functional heads, called *probes*, then seek to find a matching set of interpretable (valued) phi-features, a *goal*, inside the existing structure, i.e. inside their complement (thus Probe-Goal Agree has the additional advantage of dispensing with special checking relations such as specifier-head and replacing them with the general, independent relation of c-command/sisterhood). Should the closest matching goal be active (by virtue of having an unvalued feature of its own – Case), then Agree(Probe, Goal) takes place, as a result of which the unvalued phi-features of the probe receive values from the goal, and the unvalued Case feature of the goal is valued by the probe (nominative by T, accusative by transitive *v*). Once valued, probe and goal are no longer active, and so cannot participate in any further Agree operations.¹⁶ Case- and agreement-valuation thus go hand in hand: nominals can agree only once, capturing (27) and ‘Inverse Case Filter’ effects (Bošković 2002) formerly attributable to the ECP (**John is believed that t_{John} is happy*); and the same holds for probes, yielding Case Filter effects, such as **It seems John to be happy*: the matrix probe is valued by

¹⁶ This Activity Condition on Agree has been disputed by many, who take the null assumption to be that any valued/interpretable phi-set should be sufficient to value a probe (see, e.g., Rezac 2004 for such a view, and Nevins 2005 for an influential attempt to eliminate the Activity Condition). Even with the Activity Condition, it is still assumed that inactive phi-sets can intervene for Agree between a probe and a more remote, active goal (so-called *defective intervention*: Chomsky 2000: 123, 129, Boeckx 2003, Hiraiwa 2005), yielding certain Minimal Link Condition effects (superraising, wh-islands, etc.; cf. (6) above).

it and so is inactive for valuing Case on the lower argument, *John*, leading to a violation of FI.

3.2 Empirical advantages

Probe-Goal Agree severs the tie between movement and agreement. The configuration in (24a), in which the goal remains in situ, becomes primary, and movement to (spec-)probe obtains only in the presence of an additional movement trigger (the generalized ‘EPP-feature’ of Chomsky 2000, or else the ‘OCC’-feature of Chomsky 2001; the Edge Feature in Chomsky 2007). In terms of the empirical payoff that this affords, we have seen that a superior account of expletive-associate constructions emerges, as well as (Inverse) Case Filter effects in terms of FI. However, the price we pay for this is the loss of the connection between agreement and movement witnessed in the kinds of Romance and Germanic participle agreement facts that originally motivated the postulation of Agr(O) and spec-head object agreement in works such as Kayne (1989).¹⁷ Nevertheless, probe-goal Agree attains an overall net increase in empirical coverage, allowing a considerable range of new agreement phenomena to be accounted for that were beyond the scope of earlier minimalist and GB approaches to agreement. Space permits just the briefest mention of a few of them.

Firstly, since all agreement under Probe-Goal Agree is ‘at a distance’, nonlocal agreement phenomena are much more readily captured. Such long-distance, cross-clausal agreement, in which the matrix verb registers agreement with an in-situ embedded argument, can be found in such languages as Itelmen (Bobaljik & Wurmbrand 2005), Chukchee (Stjepanović & Takahashi 2001, Bošković 2007), Blackfoot (Legate 2005), Tsez (Polinsky & Potsdam 2001) and Hindi (Boeckx 2004, Bhatt 2005) – examples of the latter two are given in (28) – and all have been given Agree-based analyses in the above-cited works.

- (28) a. Tsez
 Eni-r [už-ā magalu bāc’ruŋi] b-iyxo
Mother.DAT [boy.ERG bread.III.ABS III.ate] III-know
 ‘The mother knows the boy ate the bread.’ (Polinsky & Potsdam 2001: 584)
- b. Hindi
 Vivek-ne [kitaab parh-nii] chaah-ii
Vivek.ERG book.F read-INF.F want-PFV.F
 ‘Vivek wants to read the book.’ (Boeckx 2004: 25 (5))

Agree has also opened up a well-known class of agreement restrictions, the *Person-Case Constraint* (PCC; Bonet 1991, 1994), to perspicuous analyses in terms of probe-sharing (multiple Agree). The PCC bans combinations of dative arguments with local-person (1/2) direct objects (where both arguments agree with the verb); the French *me-lui* restriction is an instance of this.¹⁸

¹⁷ See Svenonius 2001a, Holmberg 2002, D’Alessandro & Roberts 2008 for attempts to revive this connection within a phase-based probe-goal system.

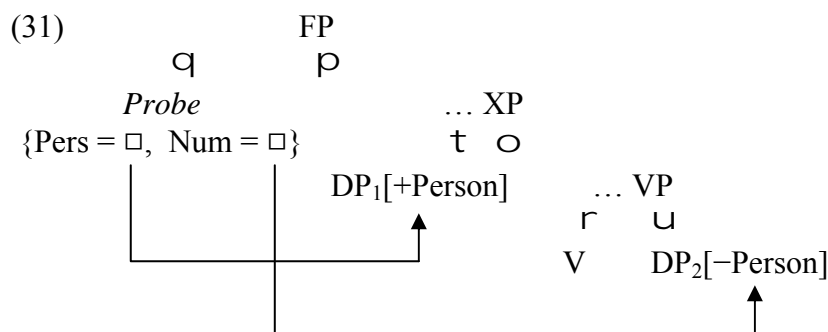
¹⁸ Other variants of the PCC are found, such as a weaker constraint allowing combinations of two local-person arguments and barring just the combination 3—1/2 (see Nevins 2007). Boeckx 2000, Anagnostopoulou 2005 equate the Icelandic restriction against 1/2-person nominative objects in the presence of quirky dative subjects (Taraldsen 1995, Sigurðsson 1996) with this weaker PCC. The weak PCC also bears a clear resemblance to direct–inverse alternations (subject–object agreement

- (29) Jean le/*me lui a recommandé (French)
 Jean it/me him has recommended

The key insight afforded by a Probe-Goal perspective on agreement is that PCC restrictions obtain where two goals relate to a single functional head (probe) for Case-valuation (Anagnostopoulou 2003, Rezac 2004):

- (30) *PCC: single probe, multiple goals*
 [P ... G_{DAT/ERG} ... G_{NOM/ACC/ABS}] → *NOM/ACC/ABS-1/2

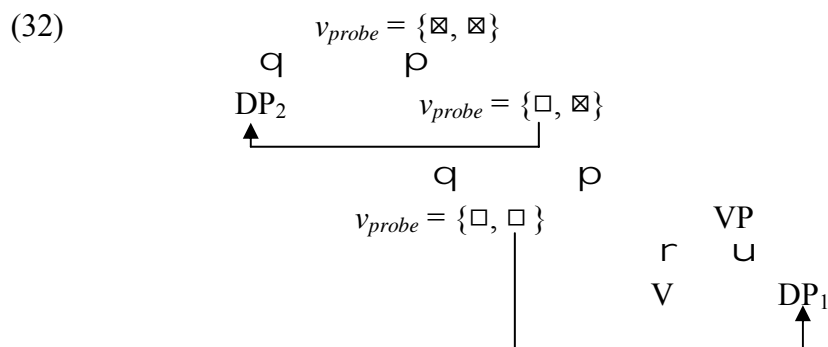
Although the Activity Condition (see previous section) prevents a probe from entering a *full* agreement relation with more than one goal (since once its features are valued, they are no longer active), partial or split agreement is possible wherever the functional head bears a composite probe. Thus Person and Number may probe separately (Béjar 2003, Rezac 2003) and value distinct arguments. By minimal search, the first (closest) goal encountered by the probe gets the first bite of the cherry, valuing as many features on the probe as it can. Any residue left over may then probe further, valuing a second, more remote argument; however, because the remaining probe is diminished, there will be fewer agreement (matching) possibilities for that second argument, which places restrictions on its featural composition (if its Case is to be valued and thus FI to be met). PCC effects then arise when the first goal values Person on the probe, leaving only Number to probe, match and value the second goal. The second goal must therefore lack a Person feature, if it is to fully match the remaining probe. On the common assumption that third-person is the absence of Person (variants of this assumption are made in most of the above-cited works), and that dative arguments are obligatorily [+Person] (see, e.g., Adger & Harbour 2007), it follows that the remoter goal (object) is barred from being 1/2-person. The general configuration is depicted in (31).



Many interesting variants exist, such as those based on the cyclic expansion of search space for probes (Rezac 2003, Béjar & Rezac 2009). Here, the probing head that values the two goals is situated between them, rather than above them (as in (32)). The probing head (e.g. *v*) first seeks a goal in its sister (complement) domain; should its features not be fully valued by the object it finds there, projection of the probe under bare phrase structure allows it then to search inside its specifier, so that the head

interactions in which the verb agrees with the highest ranked argument, where 1/2-person outranks 3-person, again barring 3—1/2 combinations) in Algonquian and other languages (cf. Anagnostopoulou 2005, Rezac 2008, Heck & Richards 2009, all of whom offer analyses in terms of Probe-Goal Agree).

that normally agrees with the internal argument (valuing internal case) may exceptionally agree with its external argument for certain features.



Béjar & Rezac (2009) show how the Probe-Goal configuration in (32) yields a transparent analysis of ergative displacement in Basque (where the external argument controls absolutive agreement just in case the internal argument is third-person) and related direct-inverse phenomena (cf. footnote 18). More generally, the idea that the functional head v may, under certain circumstances, probe into its specifier has been argued by Müller (2008) to yield the basic difference between ergative-absolutive languages, on the one hand, and those with nominative-accusative alignment on the other.

In sum, fine-grained approaches to agreement are a unique empirical selling point of minimalist Probe-Goal-Agree, allowing more exotic and complex agreement patterns and restrictions to be derived relatively straightforwardly.

4. TRANSFER AND PHASES

4.1 The cyclicity of Agree

In section 3 we saw how the Edge Feature captures the cyclicity of Merge, conforming to the NTC. Every head is a Merge cycle, with no merger possible to heads already passed in the derivation. Similar conclusions arise regarding the cyclicity of Agree and the uninterpretable features that drive it. Thus, in order to ensure that subjects are islands (Condition on Extraction Domain [CED] effects, Huang (1982)), passive must precede *wh*-movement in (33) (Chomsky 1995: 328).

(33) *Who was [a picture of t_{who}]_i taken t_i ?

In terms of Agree, the Probe-Goal relation between T and the subject (yielding subject-verb agreement and raising to Spec-TP) must precede that between C and *who*, in order to rule out a countercyclic derivation of (34b) in which the subject-island effect would be bled (by first moving *who* to Spec-CP then the remnant subject to Spec-TP).

(34) a. [CP Who was [TP T [VP taken [DP a picture of t_{who}]]]? →
 b. [CP Who was [TP [DP a picture of t_{who}]_i T [VP taken t_i]]?

Every head must define a cycle not only for EFs, then, but also for uFs, so that the phi-probe and strong-D (or EPP) feature on T must be satisfied before the derivation

proceeds to the next head (C). In terms of the original Strict Cycle Condition (Chomsky 1973), every head with featural requirements must now count as a ‘cyclic node’. The resultant *featural cyclicity* (N. Richards 1999) is formulated by Chomsky (1995: 234 (3)) as in (35), where α is an unvalued (active) feature.

(35) D[erivation] is cancelled if α is in a category not headed by α .

Active features are thus possible only on the root node (the ‘locus’ in Collins’s (2002) development of this idea).

In GB, the island effect in the cyclic derivation of (33) (also (6c) above) was captured by *Subjacency*, a locality condition on movement: Two blocking categories, DP and IP, are crossed in a single step in the cyclic version of (34b). Subjacency also yielded the effect of the Strict Cycle Condition, forcing long-distance, ‘unbounded’ movement dependencies to proceed via shorter, bounded steps through intermediate CPs. In Minimalism, the relevant cycles for successive cyclic movement are defined as *phases* (C and transitive v). A phase is both a point at which the syntax is accessed and evaluated by the interfaces (‘transferred’), and the entity transferred at that point. Like Subjacency, phases combine cyclicity and locality. Transferred material becomes opaque to further syntactic computation, in accordance with the *Phase Impenetrability Condition*, yielding absolute locality effects – the phase boundary represents an upper limit on the length of movement and agreement dependencies. This in turn forces movement to proceed cyclically via an escape hatch at the ‘edge’ of the phase (spec-CP, spec- v P).¹⁹ The resultant phase cyclicity is not just relevant for movement, however. That phases define cycles for Agree(ment) too is apparent from the problem in (36) (cf. Chomsky 2004a, Anagnostopoulou 2003, Müller 2004).

(36) a. What did John read?
 b. [_{CP} What did [_{TP} John T [_{vP} (*what*) [_{vP} (*John*) [_{VP} read (*what*)]]]]]

The problem is how T can probe the subject *John* across the *wh*-copy/trace *what* in Spec- v P. The *wh*-copy contains the kind of features that the T probe is looking for (i.e. phi-features) and should therefore block Agree(T, *John*) as a closer potential goal (by defective intervention; see footnote 16). The solution proposed by Chomsky (and elaborated in Anagnostopoulou 2003) is that phases are cyclic domains (i.e. domains of rule ordering), so that “all operations within the phase are in effect simultaneous”, taking place at the phase level (at the point of Transfer to the interfaces). Therefore, C can *wh*-probe *what* and displace it to Spec-CP ‘before’ T probes *John*, thus removing the intervener out of the way.

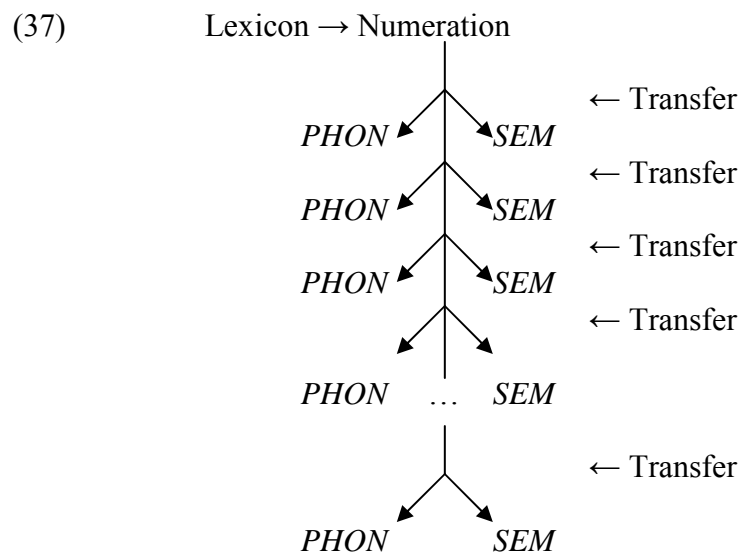
The upshot is that phase-internal countercyclicity effects are expected, such as those holding between T and C (also V and v). However, this means that the featural/locus cyclicity needed to account for the CED effect in (33), where every head is a cycle, is lost. Further, as Epstein & Seely (2002) discuss at length, delaying the valuation and deletion of T’s uFs until the phase level is conceptually problematic. Chomsky (2007, 2008) addresses the latter problems by making the phase heads

¹⁹ In this way, phases “yield a strong form of subjacency” (Chomsky 2000: 108). Nevertheless, phases are still wanting as a theory of locality – they clearly do not make for very good islands, precisely because they are designed to be escapable, via the phase edge (see also Boeckx & Grohmann 2007 on this point). What is required to bring phases up to the level of descriptive adequacy attained by barriers in GB, then, is a theory of precisely when the phase edge (specifier region) is and is not available (i.e. able to be projected). Müller 2009 is an important step in this direction.

themselves the locus of the agreement probes (with these features being passed down onto their complements, T and V, by *feature inheritance*). It is thereby the phase heads that define agreement cycles. The facts in (33) and (6), however, indicate that we must maintain that *every* head (phase and nonphase) is still a separate *Merge* cycle. I would therefore like to suggest that what we arrive at is a kind of *relativized featural cyclicity* in which each primitive feature type defines its own cycle: EFs define Merge cycles and uFs define phase cycles; every head is an EF cycle (since every head has EF, in order to be able to Merge), whilst every phase head is a uF cycle (since phase heads are the uF sites).

4.2 Conceptions of phases

The model of the grammar given in (3) has each interface accessing the syntax only once, at the end of the derivation and after the Numeration has been exhausted. However, to restrict the operation Transfer in this way is arbitrary and stipulative. Further, as argued by Epstein et al (1998), Epstein & Seely (2002), eliminating the internal levels of d-structure and s-structure is not enough from the minimalist perspective. Rather, LF and PF should also be replaced as integral levels of representation (as should the special PF-only operation Spell-Out, a residue of s-structure), with every derivational step (transformation) being accessed and evaluated by the interfaces as the computation proceeds (“invasive”, online, dynamic interpretation). That is, we should aim for a multiple-Transfer model of the grammar:



Breaking down the numeration into smaller subarrays would be one way to achieve such multiple Transfer – the derivation proceeds on a subarray-by-subarray basis, with Transfer occurring upon exhaustion of each subarray. Alternatively, uFs might act as the trigger for Transfer – as argued by Epstein & Seely (2002), valued uFs must be immediately transferred and deleted, without delay, in order for convergence to be possible (the CI interface must be able to distinguish between valued uninterpretable and interpretable features in order to delete the former). The former view of phases, as *lexical subarrays*, is put forward in Chomsky 2000, where each subarray contains exactly one instance of one of the phase heads (C, v). The latter conception of phases might be called the *convergence* view, since it is premised on the idea that cyclic

Transfer is required in order to produce legitimate interface objects²⁰ – either legitimate PF-objects (Uriagereka 1999, who proposes that complex left branches must be spelled out separately for linearization reasons, in order to integrate them with the main tree by the LCA) or legitimate CI-interpretable objects (Epstein & Seely 2002, Chomsky 2007, 2008).

Whether conceived of as subarrays or as uF-sites, phases (and cyclic Transfer) have been argued by Chomsky to facilitate the syntactic computation and to economize on cognitive resources in various ways, and thus to be a central third-factor property in an optimally designed FL conforming to the SMT. For example, under the subarray conception of Chomsky 2000:99-101, phases minimize the workspace (the amount of lexical information that has to be “carried along” at any given point of the derivation); in Chomsky 2001:5, phases minimize the delay between valuation and deletion of valued uFs, in the manner reviewed above; in Chomsky 2004a:4, phases minimize working memory and the search space available to probes through the periodic ‘forgetting’ of structure (interpreted material cannot be further modified by Merge or Agree); Chomsky 2007:5,16 argues that phases contribute to optimal design by implementing strict cyclicity; and Chomsky 2008: 4,8-9 emphasizes the elimination of redundant internal levels and cycles (LF and the mappings to the interfaces) that phases afford, yielding ‘single cycle generation’ (cf. (37)).

4.3 Empirical advantages

Chomsky’s (2000) original formulation of phases and its immediate development (the revised Phase Impenetrability Condition in Chomsky 2001) were guided by empirical considerations. Lexical subarrays are first motivated as a way to localize the domain in which Merge of an expletive preempts movement of a noun phrase.

- (38) a. There is likely [_{α} to be a proof discovered]
 b. *There is likely [_{α} a proof to be (*a proof*) discovered]
 c. There is a possibility [_{α} that proofs will be (*proofs*) discovered]

The pair in (38a-b) shows the relevant effect (Merge-over-Move): *there* must merge to the embedded subject position (subsequently raising to the matrix subject position), blocking movement of *a proof* inside the embedded clause α . The question is why the presence of *there* in the numeration does not similarly block raising of *proofs* in (38c). The problem is solved if α in (38c), but not in (38a-b), is built from a distinct subarray of the numeration, from which it is then possible to exclude the expletive. Finite clauses (CPs) are thus phases (the product of a separate subarray); non-finite TPs are not.

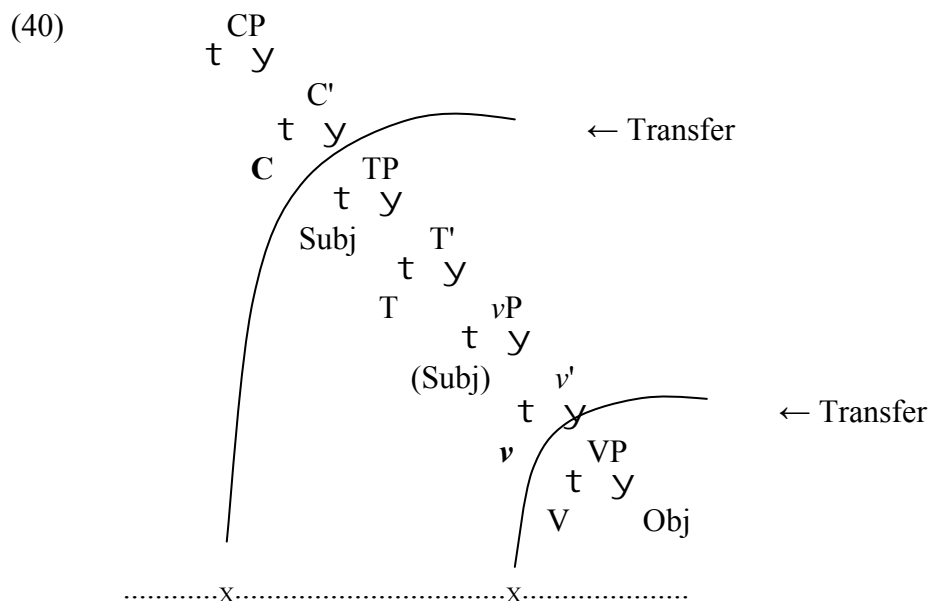
To capture the effects of successive cyclicity, Chomsky (2000: 108) proposes that the *complement* of the phase head is transferred, becoming inaccessible to further operations – TP in the case of the C-phase, and VP for the ν -phase. Transfer of the complement leaves the specifier(s)/‘edge’ of the phase available as an escape hatch to which active material inside the complement can move, thereby allowing the derivation to continue and converge (this property of the phase edge thus follows from optimal design). Some of the most convincing evidence that phase-cyclic computation

²⁰ For a different convergence-based notion of phases, in which structure is transferred as soon as it is internally convergent (i.e. with all features valued), see Svenonius (2001a,b).

does indeed proceed in this way, and thus for the reality of phases as units of interpretation, has arguably come from the PF-interface. In addition to Uriagereka's (1999) reduction of CED effects to the workings of PF-motivated multiple spell-out, several researchers have derived syntactic order preservation constraints such as Holmberg's Generalization (Holmberg 1986, 1999) from phase-cyclic linearization (Richards 2004, Fox & Pesetsky 2005). Others have explored the possibility that phase-cyclic mapping to the PF-interface should be detectable in the form of phonological phrase boundaries coinciding with syntactic phase boundaries. Thus Franks & Bošković (2001) show how the placement of second-position clitics in Bulgarian is sensitive to the phase boundary that occurs between C and its complement, TP. In a similar vein, Richards (2004) proposes that the distribution of weak pronouns in VO (head-initial) versus OV (head-final) Germanic provides evidence for the reality of the phase boundary between *v* and its complement, VP. Consider the paradigm in (39).

- (39) a. *Nemandinn las (hana) ekki (*hana)* (Icelandic/VO)
 Student-the read (it) not (it)
- b. *Nemandinn hefur (*hana) ekki lesið (hana)*
 Student-the has (it) not read (it)
- c. *Der Student las (es) nicht (*es)* (German/OV)
 The student read (it) not (it)
- d. *Der Student hat (es) nicht (*es) gelesen*
 The student has (it) not (it) read

Weak pronouns are forced to undergo *obligatory* object shift in VO Scandinavian (39a), but only in those environments in which the finite verb raises out of VP (39b), in accordance with Holmberg's Generalization. However, in OV Germanic (German, Dutch, Afrikaans, etc.), this shifting of weak pronouns is obligatory irrespective of the position of the finite verb (the object must shift in (39d) no less than in (39c)). If weak pronouns are phonologically enclitic, requiring incorporation into a leftward-neighbouring prosodic word at PF, then the facts in (39) follow immediately: Transfer imposes a phase boundary across which cliticization is impossible (since phases are separate spell-out units).



Due to the impossibility of cross-phasal cliticization, a weak pronoun cannot be the leftmost element inside a phasal domain. Its host must be a phase-mate; for in-situ weak pronouns, this means the host must be inside VP. In VO languages (41), if the verb raises, so must the weak pronominal object, as only an in-situ verb can meet this requirement. In OV languages (42), nothing can meet it, since the verb is to the right; therefore, movement of the pronoun out of VP is always forced, no matter whether the verb moves out of VP or not.

- (41) Icelandic (VO) (= (39a))
- a. Unshifted pronoun
 * $[_{CP} [Nemandinn] [_{C'} [las] \dots [_{vP} ekki [_{VP} tv hana]]]]$
 () () $_{\varphi}$
- b. Shifted pronoun
 $[_{CP} [Nemandinn] [_{C'} [las] \dots [_{vP} hana [_{vP} ekki [_{VP} tv t_{Obj}]]]]$
 () () $_{\varphi}$

- (42) German (OV) (= (39d))
- a. Unshifted pronoun
 * $[_{CP} [Der Student] [_{C'} [hat] \dots [_{vP} nicht [_{VP} es gelesen]]]]$
 (...) () $_{\varphi}$
- b. Shifted pronoun
 $[_{CP} [Der Student] [_{C'} [hat] \dots [_{vP} es [_{vP} nicht [_{VP} t_{Obj} gelesen]]]]$
 (...) () $_{\varphi}$

5. OUTLOOK

In terms of longevity, Minimalism has already surpassed all previous transformational–generative models (Standard and Extended Standard Theory, GB). The results of nearly twenty years of minimalist research show that the MP’s attempt to reduce GB-style descriptive technology to interface conditions and general computational and cognitive principles is a viable, realistic and worthwhile enterprise. The list in (43) summarizes some of the results reviewed above, which are themselves a far from comprehensive selection.

- | | | | |
|------|----------------------------------|---|--|
| (43) | <u>Minimalism (third factor)</u> | | <u>GB (first factor)</u> |
| | Full Interpretation | ← | Case Filter, Theta-Criterion, ECP |
| | No-Tampering Condition | ← | Projection Principle, Theta-Criterion, strict cycle, extension condition |
| | Phases | ← | Barriers (Subjacency, CED), strict cycle |

In addition, we saw that Merge eliminates d- and s-structure, and phases eliminate LF and PF as representational levels. What remains is a largely empty UG, perhaps comprising just the basic inventory of formal features (uFs and EFs) required to yield the three minimal operations, Merge, Agree and Transfer. The biolinguistic ideal of the SMT would therefore seem a serious prospect after twenty years of progress. Of course, innumerable challenges still stand in its path, and not just the obvious empirical ones. Among the questions currently shaping the minimalist program are the following:

If UG is maximally devoid of FL-specific principles, then what about parameters? Chomskyan minimalism has always adopted the view that variation must be limited to the properties of lexical items and to the mapping to the AP interface (Chomsky 1995: 169, 221, etc.), domains outside the purview of the SMT; nevertheless, the role of the third factor in determining the form and fixing of parameters is ripe for investigation (cf. Boeckx 2008, Holmberg & Roberts 2008). A related question is whether the interfaces are ‘created equal’. Chomsky 2007, 2008, Berwick & Chomsky 2008 suggest that language design is asymmetric, with SMT holding only between syntax and CI. In support of this idea, we might add that the very fact that Merge is symmetrical (see footnote 9) indicates that syntax is indeed not perfectly designed to meet conditions imposed by the AP interface – AP requires asymmetry (for linear order), yet the syntax is not geared towards providing this. Consequently, the AP interface has to do the best it can, allowing head-directionality macroparameters to naturally arise here (cf. Richards 2004). Things are very different at the CI interface, where the SMT holds. But many questions arise here too: most intriguingly, should the nature of the optimal mapping to CI be viewed functionally in terms of the CI system bending the syntax to its will (the “I-functional” view of the SMT; cf. Epstein 2007), or is the semantics in fact shaped in the image of the syntax, with perfection at the CI interface really a reflection of optimal syntactic computation (Hinzen 2006, Uriagereka 2008, Narita 2009)? If interface conditions do hold sway (i.e. the former view), then does that imply that syntax should be crash-proof at the level of individual derivations (Frampton & Gutmann 2002), or simply at the level of FL (Chomsky’s “usable at all” criterion)?

An example of the latter kind of ‘weak’ crash-proofing is the need for uFs on phase heads to be offloaded onto their complements in order to prevent every phase from crashing (and thus non-usability of FL); see Richards (2007a). However, this *conceptually* superior view of phases (as the locus of uFs) is incompatible with the *empirically* superior version of the Phase Impenetrability Condition which Chomsky (2001:13-14) proposes in order to allow for agreement relations between T and internal arguments, as demanded by facts from Icelandic and, if passive/unaccusative *v* is also a phase (cf. Legate 2003), by expletive-associate constructions.²¹

This last point demonstrates particularly acutely the tension between data coverage and explanatory depth that lies at the heart of the minimalist program. The minimalist instinct is to favour the view of phases that better conforms to the SMT, and thus to seek alternative explanations for the recalcitrant data rather than abandoning the enterprise. To echo Chomsky (1995: 10), whether such instincts are on the right track, only time will tell.

²¹ See Richards (2007b) for details of this incompatibility.

REFERENCES

- Aboh, Enoch Oladé. 2004. Snowballing Movement and Generalized Pied-Piping. In: Anne Breitbarth and Henk van Riemsdijk (eds) *Triggers*, 15-47. Berlin: de Gruyter.
- Adger, David. 2003. *Core Syntax. A Minimalist Approach*. Oxford: Oxford University Press.
- Adger, David and Daniel Harbour. 2007. Syntax and Syncretisms of the Person Case Constraint. *Syntax* 10: 2–37.
- Alexiadou, Artemis and Elena Anagnostopoulou. 1998. Parametrizing Agr: Word Order, V-Movement and EPP-Checking. *Natural Language and Linguistic Theory* 16: 491-539.
- Anagnostopoulou, Elena. 2003. *The Syntax of Ditransitives. Evidence from Clitics*. Berlin: de Gruyter.
- Anagnostopoulou, Elena. 2005. Strong and Weak Person Restrictions: a Feature-Checking analysis. In: Lorie Heggie and Fransisco Ordóñez (eds.), *Clitic and Affix Combinations. A Theoretical Perspective*. Amsterdam: Benjamins.
- Anderson, Stephen and David Lightfoot. 2002. *The Language Organ. Linguistics as Cognitive Physiology*. Cambridge: Cambridge University Press.
- Baker, Mark. 1988. *Incorporation: A Theory of Grammatical Function Changing*. University of Chicago Press, Chicago
- Béjar, Susana. 2003. Phi-syntax: A theory of agreement. Ph.D. dissertation. University of Toronto.
- Béjar, Susana and Milan Rezac. 2009. Cyclic Agree. *Linguistic Inquiry* 40: 35-73.
- Berwick, Robert and Noam Chomsky. 2008. The Bilingual Program: The Current State of its Evolution and Development. Ms., MIT. Forthcoming in: Anna-Maria di Sciullo and Calixto Aguero (eds.), *Bilingual Investigations*. Cambridge, Mass.: MIT Press.
- Bhatt, Rajesh. 2005. Long Distance Agreement in Hindi-Urdu. *Natural Language and Linguistic Theory* 23: 757-807.
- Biberauer, Theresa and Marc Richards. 2006. True Optionality: When the grammar doesn't mind. In: Cedric Boeckx (ed.) *Minimalist Essays*, 35-67. Amsterdam: Benjamins.
- Bobaljik, Jonathan. 2002. A-chains at the PF-interface: Copies and “covert” movement. *Natural Language and Linguistic Theory* 20: 197-267.
- Bobaljik, Jonathan and Susi Wurmbrand. 2005. The Domain of Agreement. *Natural Language and Linguistic Theory* 23: 809-865.

- Boeckx, Cedric. 2000. Quirky Agreement. *Studia Linguistica* 54: 354-380.
- Boeckx, Cedric. 2003. *Islands and Chains. Resumption as Stranding*. Amsterdam: Benjamins.
- Boeckx, Cedric. 2004. Long-distance Agreement in Hindi: Some Theoretical Implications. *Studia Linguistica* 58: 23-36.
- Boeckx, Cedric. 2006. *Linguistic Minimalism. Origins, Concepts, Methods, Aims*. Oxford/New York: Oxford University Press.
- Boeckx, Cedric. 2009. Approaching parameters from below. In: Anna-Maria di Sciullo and Cedric Boeckx (eds.) *Biolinguistics: Language evolution and variation*. Oxford: Oxford University Press.
- Boeckx, Cedric and Kleanthes Grohmann. 2007. Putting phases in perspective. *Syntax* 10: 204-222.
- Boeckx, Cedric and Juan Uriagereka. 2007. Minimalism. In: Gillian Ramchand and Charles Reiss (eds.), *The Oxford Handbook of Linguistic Interfaces*. Oxford/New York: Oxford University Press.
- Bonet, Eulàlia. 1991. Morphology after syntax: Pronominal clitics in Romance. PhD dissertation, MIT.
- Bonet, Eulàlia. 1994. The Person-Case Constraint: a morphological approach. In: Heidi Harley and Colin Phillips (eds.) *The morphology-syntax connection* (MITWPL 22), 33-52. Cambridge, Mass.: MIT Press.
- Bošković, Željko. 2001. *On the Nature of the Syntax-Phonology Interface: Cliticization and related phenomena*. Amsterdam: Elsevier.
- Bošković, Željko. 2002. A-movement and the EPP. *Syntax* 5: 167-218.
- Bošković, Željko. 2007. Agree, Phases, and Intervention Effects. *Linguistic Analysis* 33: 54-96.
- Bošković, Željko. 2005. On the locality of left-branch extraction and the structure of NP. *Studia Linguistica* 59: 1-45.
- Brody, Michael. 2002. On the Status of Representations and Derivations. In: Samuel David Epstein and T. Daniel Seely (eds.) *Derivation and Explanation in the Minimalist Program*, 19-41. Oxford: Blackwell.
- Castillo, Juan Carlos, John Drury, and Kleanthes K. Grohmann. 1999. No more EPP. In *Proceedings of WCCFL 19*, 153-166. Somerville: Cascadilla Press.

- Chomsky, Noam. 1973. Conditions on Transformations. In: Stephen Anderson and Paul Kiparsky (eds.) *A Festschrift for Morris Halle*, 232-286. New York: Holt, Rinehart & Winston.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, Noam. 1991. Some notes on economy of derivation and representation. In: Robert Freidin (ed.), *Principles and parameters in comparative grammar*, 417-454. Cambridge, Mass.: MIT Press. [Reprinted in: Noam Chomsky. 1995. *The minimalist program*, 129-166. Cambridge, Mass.: MIT Press.]
- Chomsky, Noam. 1993. A minimalist program for linguistic theory. In: Ken Hale and Samuel J. Keyser (eds.), *The view from Building 20*, 1-52. Cambridge, Mass.: MIT Press. [Reprinted in: Noam Chomsky. 1995. *The Minimalist Program*, 167-217.]
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, Mass.: MIT Press.
- Chomsky, Noam. 2000. Minimalist inquiries: the framework. In: Roger Martin, David Michaels and Juan Uriagereka (eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, 89-156. Cambridge, Mass.: MIT Press.
- Chomsky, Noam. 2001. Derivation by phase. In: Michael Kenstowicz (ed.), *Ken Hale: a life in language*, 1-50. Cambridge, Mass.: MIT Press.
- Chomsky, Noam. 2004a. Beyond explanatory adequacy. In Adriana Belletti (ed.), *Structures and beyond. The cartography of syntactic structures*, Volume 3, 104-131. Oxford: Oxford University Press.
- Chomsky, Noam. 2004b. *The Generative Enterprise Revisited*. Berlin: de Gruyter.
- Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36: 1-22.
- Chomsky, Noam. 2007. Approaching UG from below. In: Uli Sauerland & Hans-Martin Gärtner (eds.) *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax-Semantics*, 1-30. Berlin: Mouton de Gruyter.
- Chomsky, Noam. 2008. On phases. In: Robert Freidin, Carlos P. Otero and Maria-Luisa Zubizarreta (eds.), *Foundational Issues in Linguistic Theory*, 133-166. Cambridge, MA: MIT Press.
- Christensen, Kirsti and Knut Tarald Taraldsen. 1988. Expletive Chain Formation and Past Participle Agreement in Scandinavian Dialects. In: Paola Benincà (ed.) *Dialect Variation in the Theory of Grammar*, 53-84. Dordrecht: Foris.
- Citko, Barbara. 2005. On the Nature of Merge: External Merge, Internal Merge, and Parallel Merge. *Linguistic Inquiry* 36: 475-496.
- Collins, Chris. 1997. *Local economy*. Cambridge, Mass.: MIT Press.

- Collins, Chris. 2002. Eliminating labels. In: Samuel David Epstein and T. Daniel Seely (eds.) *Derivation and explanation in the minimalist program*, 42-64. Oxford: Blackwell.
- D'Alessandro, Roberta and Ian Roberts. 2008. Movement and Agreement in Italian Past Participles and Defective Phases. *Linguistic Inquiry* 39: 477-491.
- Dikken, Marcel den. 1995. Binding, Expletives, and Levels. *Linguistic Inquiry* 26: 347-354.
- Dirac, Paul A. M. 1968. Methods in theoretical physics. In: *From a life of physics: Evening lectures at the International Center for Theoretical Physics*, Trieste, Italy. A special supplement of the International Atomic Energy Agency Bulletin, Austria. [Reprinted in: Abdus Salam (ed.) *Unification of Fundamental Forces*, 125-143. Cambridge: Cambridge University Press.]
- Epstein, Samuel David. 2007. On I(nternalist)-functional Explanation in Minimalism. *Linguistic Analysis* 33(1-2): 20-53.
- Epstein, Samuel David, Erich Groat, Ruriko Kawashima and Hisatsugu Kitahara. 1998. *A Derivational Approach to Syntactic Relations*. New York: Oxford University Press.
- Epstein, Samuel David, Hisatsugu Kitahara and T. Daniel Seely. 2009. Deducing Extraction Constraints and Transfer-Application from 3rd Factor Considerations on Language Design. *GLOW Newsletter* 62.
- Epstein, Samuel David and T. Daniel Seely. 2002. Rule Applications as Cycles in a Level-free Syntax. In: Samuel David Epstein and T. Daniel Seely (eds.) *Derivation and explanation in the minimalist program*, 65-89. Oxford: Blackwell.
- Epstein, Samuel David and T. Daniel Seely. 2006. *Derivations in Minimalism*. Cambridge/New York: Cambridge University Press.
- Fanselow, Gisbert and Damir Ćavar. 2002. Distributed Deletion. In: Artemis Alexiadou (ed.) *Theoretical Approaches to Universals*, 66-107. Amsterdam: John Benjamins.
- Fox, Danny. 2000. *Economy and Semantic Interpretation*. Cambridge, Mass: MIT Press.
- Fox, Danny and David Pesetsky. 2005. Cyclic Linearization of Syntactic Structure. *Theoretical Linguistics* 31: 1-46.
- Frampton, John and Sam Gutmann. 2002. Crash-proof Syntax. In: Samuel David Epstein and T. Daniel Seely (eds.) *Derivation and Explanation in the Minimalist Program*, 90-105. Oxford: Blackwell.
- Franks, Steven and Željko Bošković. 2001. An argument for multiple spell-out. *Linguistic Inquiry* 32: 174-183.

- Freidin, Robert and Jean-Roger Vergnaud. 2001. Exquisite connections: some remarks on the evolution of linguistic theory. *Lingua* 111: 639-666.
- Fujii, Tomohiro. 2005. Cycle, Linearization of Chains, and Multiple Case Checking. *Proceedings of ConSOLE XIII*: 39-65.
- Gallego, Ángel. 2006. Instability. Ms., Universitat Autònoma de Barcelona.
- Groat, Erich. 1997. A Derivational Program for Syntactic Theory. Ph.D. dissertation. Harvard University.
- Groat, Erich. 1999. Raising the Case of Expletives. In: Samuel David Epstein and Norbert Hornstein (eds.) *Working Minimalism*, 27-43. Cambridge, Mass: MIT Press.
- Grohmann, Kleanthes K. 2003. *Prolific Domains: On the Anti-Locality of Movement Dependencies*. Amsterdam: Benjamins.
- Hauser, Marc, Noam Chomsky and W. Tecumseh Fitch. 2002. The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science* 298: 1569-79.
- Heck, Fabian and Marc Richards. 2009. A Probe-Goal Approach to Agreement and Non-incorporation Restrictions in Southern Tiwa. To appear in *Natural Language and Linguistic Inquiry*.
- Hinzen, Wolfram. 2006. *Mind design and minimal syntax*. Oxford: Oxford University Press.
- Hiraiwa, Ken. 2005. Dimensions of Symmetry in Syntax: Agreement and Clausal Architecture. Ph.D. dissertation, MIT.
- Holmberg, Anders. 1986. *Word Order and Syntactic Features in the Scandinavian Languages and English*. Ph.D. dissertation, University of Stockholm.
- Holmberg, Anders. 1999. Remarks on Holmberg's Generalization. *Studia Linguistica* 53: 1-39.
- Holmberg, Anders. 2002. Expletives and Agreement in Scandinavian Passives. *Journal of Comparative Germanic Linguistics* 4: 85-128.
- Holmberg, Anders and Ian Roberts. 2008. Introduction. To appear in: *Parametric variation: null subjects in minimalist theory*. Cambridge: Cambridge University Press.
- Hornstein, Norbert. 2009. *A Theory of Syntax. Minimal Operations and Universal Grammar*. Cambridge: Cambridge University Press.
- Hornstein, Norbert, Jairo Nunes and Kleanthes K. Grohmann. 2005. *Understanding Minimalism*. Cambridge: Cambridge University Press.

- Huang, C.-T. James. 1982. Logical relations in Chinese and the theory of grammar. Ph.D. dissertation, MIT.
- Katz, Jerry and Thomas Bever. 1976. The fall and rise of empiricism. In: Thomas Bever, Jerry Katz and D. T. Langendoen (eds.), *An Integrated Theory of Linguistic Ability*, 11-64.. New York: Thomas Y. Crowell Company.
- Kayne, Richard. 1989. Facets of Romance Past Participle Agreement. [Reprinted in: Richard Kayne. 2000. *Parameters and Universals*, 25-39. Oxford: Oxford University Press.]
- Kayne, Richard. 1994. *The Antisymmetry of Syntax*. Cambridge, Mass.: MIT Press.
- Lamova, Mariana. 2004. On Triggers of Movement and Effects at the Interfaces. In: Anne Breitbarth and Henk van Riemsdijk (eds.) *Triggers*, 231-258. Berlin: de Gruyter.
- Landau, Idan. 2007. Constraints on Partial VP-fronting. *Syntax* 10: 127-164.
- Lappin, Shalom, Robert Levine and David Johnson. 2000a. The structure of unscientific revolutions. *Natural Language and Linguistic Theory* 18: 665-671.
- Lappin, Shalom, Robert Levine and David Johnson. 2000b. The revolution confused: a response to our critics. *Natural Language and Linguistic Theory* 18: 873-890.
- Lappin, Shalom, Robert Levine, and David Johnson. 2001. The revolution maximally confused. *Natural Language and Linguistic Theory* 19: 901-919.
- Lasnik, Howard. 1993. Lectures on Minimalist Syntax. *University of Connecticut Occasional papers in Linguistics* 1. [Reprinted in: Howard Lasnik. 1999. *Minimalist Analysis*. Oxford: Blackwell.]
- Lasnik, Howard. 1995. Last Resort. In: Haraguchi Shosuke and Michio Funaki (eds.) *Minimalism and Linguistic Theory*, 1-32. Tokyo: Hituzi Syobo. [Reprinted in: Howard Lasnik. 1999. *Minimalist Analysis*. Oxford: Blackwell.]
- Lasnik, Howard and Terje Lohndal. To appear. Government-Binding/Principles and Parameters Theory. In: Lynn Nadel (ed.) *Wiley Interdisciplinary Reviews: Cognitive Science*. Wiley & Sons.
- Lasnik, Howard and Mamoru Saito. 1992. *Move α . Conditions on Its Application and Output*. Cambridge, Mass.: MIT Press.
- Lasnik, Howard, Juan Uriagereka and Cedric Boeckx. 2004. *A course in minimalist syntax. Foundations and Prospects*. Oxford: Blackwell.
- Legate, Julie Anne. 2003. Some Interface Properties of the Phase. *Linguistic Inquiry* 34: 506-16.

Legate, Julie Anne. 2005. Phases and Cyclic Agreement. In: Martha McGinnis and Norvin Richards (eds.), *Perspectives on Phases* (MITWPL 49), 147-156. Cambridge, Mass.: MIT Press.

Martin, Roger. 1999. Case, the Extended Projection Principle, and Minimalism. In: Samuel David Epstein and Norbert Hornstein (eds.) *Working Minimalism*, 1-25. Cambridge, Mass: MIT Press.

Moro, Andrea. 2000. *Dynamic Antisymmetry*. Cambridge, Mass: MIT Press.

Müller, Gereon. 2004. Phase impenetrability and wh-intervention. In: Arthur Stepanov, Gosbert Fanselow and Ralf Vogel (eds.), *Minimality Effects in Syntax*, 289-325. Berlin: de Gruyter.

Müller, Gereon. 2008. Ergativity, Accusativity, and the Order of Merge and Agree. Ms., Universität Leipzig.

Müller, Gereon. 2009. On Deriving CED Effects from the PIC. To appear in *Linguistic Inquiry*.

Narita, Hiroki. 2009. How Syntax Naturalizes Semantics. Review of Uriagereka (2008) *Syntactic Anchors: On Semantic Structuring*. To appear in *Lingua*.

Newmeyer, Frederick J. 2003. Review of Chomsky, "On nature and language"; Anderson & Lightfoot, "The language organ"; Bichakjian, "Language in a Darwinian perspective". *Language* 79, 583-599.

Nevins, Andrew. 2005. Derivations without the Activity Condition. In: Martha McGinnis and Norvin Richards (eds.), *Perspectives on Phases* (MITWPL 49), 283-306. Cambridge, Mass.: MIT Press.

Nevins, Andrew. 2007. The Representation of Third Person and its Consequences for Person-Case Effects. *Natural Language and Linguistic Theory* 25: 273-313.

Nunes, Jairo. 1999. Linearization of Chains and Phonetic Realization of Chain Links. In: Samuel David Epstein and Norbert Hornstein (eds.) *Working Minimalism*, 217-249. Cambridge, Mass: MIT Press.

Nunes, Jairo. 2004. *Linearization of Chains and Sideward Movement*. Cambridge, Mass.: MIT Press.

Pesetsky, David. 1989. Language Particular Processes and the Earliness Principle. Ms., MIT.

Polinsky, Maria and Eric Potsdam. 2001. Long-distance Agreement and Topic in Tsez. *Natural Language and Linguistic Theory* 19: 583-646.

Pollock, Jean-Yves. 1989. Verb movement, Universal Grammar, and the structure of IP. *Linguistic Inquiry* 20: 365-424.

- Reinhart, Tanya. 1995. Interface Strategies. *OTS Working Papers in Linguistics*. Utrecht.
- Rezac, Milan. 2003. The Fine Structure of Cyclic Agree. *Syntax* 6: 156–182.
- Rezac, Milan. 2004. Elements of Cyclic Syntax: Agree and Merge. Ph.D. dissertation, University of Toronto.
- Rezac, Milan. 2008. Phi Across Modules. Ms., CNRS UMR 7023, Université de Paris 8.
- Richards, Marc. 2004. Object Shift and Scrambling in North and West Germanic: A Case Study in Symmetrical Syntax. Ph.D. dissertation. University of Cambridge
- Richards, Marc. 2007a. Deriving the Edge: What's in a phase? Ms., University of Leipzig.
- Richards, Marc. 2007b. On feature-inheritance: an argument from the Phase Impenetrability Condition. *Linguistic Inquiry* 38: 563-572.
- Richards, Marc and Theresa Biberauer. 2005. Explaining *Expl*. In: Marcel den Dikken and Chritina Tortora (eds.), *The Function of Function Words and Functional Categories*, 115-153. Amsterdam/New York: John Benjamins.
- Richards, Norvin. 1999. Featural Cyclicity and the Ordering of Multiple Specifiers. In: Samuel David Epstein and Norbert Hornstein (eds.) *Working Minimalism*, 127-158. Cambridge, Mass: MIT Press.
- Riemsdijk, Henk van, and Edwin Williams. 1981. NP-structure. *Linguistic Review* 1: 171-217.
- Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, Mass.: MIT Press.
- Saito, Mamoru & Naoki Fukui. 1998. Order in Phrase Structure and Movement. *Linguistic Inquiry* 29: 439-74.
- Sigurðsson, Halldór. 1996. Icelandic finite verb agreement. *Working Papers in Scandinavian Syntax* 57: 1-46.
- Stjepanović, Sandra & Shoichi Takahashi. 2001. Eliminating the Phase Impenetrability Condition. Ms., Kanda University of International Studies.
- Svenonius, Peter. 2001a. Impersonal passives: A phase-based analysis. *Proceedings of the 18th Scandinavian Conference of Linguistics* 39.2: 109-125. Lund: Travaux de l'Institut de Linguistique de Lund.
- Svenonius, Peter 2001b. On Object Shift, Scrambling, and the PIC. In: Elena Guerzoni and Ora Matushansky (eds.) *A Few from Building E39* (MITWPL 39), 267-289. Cambridge, Mass.: MIT Press.

Taraldsen, Knut Tarald. 1995. On agreement and nominative objects in Icelandic. In: Hubert Haider, Susan Olsen and Sten Vikner (eds.), *Studies in Comparative Germanic Syntax*, 307-327. Dordrecht: Kluwer.

Uriagereka, Juan. 1998. *Rhyme and reason*. Cambridge, Mass.: MIT Press.

Uriagereka, Juan. 1999. Multiple spell-out. In: Samuel David Epstein and Norbert Hornstein (eds.) *Working Minimalism* 251-282. Cambridge, Mass.: MIT Press.

Uriagereka, Juan. 2008. *Syntactic Anchors: On Semantic Structuring*. Cambridge: Cambridge University Press.