‘Functionalism’, the design of the language faculty, and (disharmonic) typology [*]

IAIN MOBBS

King’s College, Cambridge

Abstract

In an influential functionalist theory of grammar, Hawkins (1994, 2004) implicates performance principles of communicative *economy and efficiency in motivating structure. This contrasts with concerns for* computational *optimality expressed by the Minimalist Program (Chomsky 1995b), arguably the dominant framework for formal generativism. This article disputes the functionalist position and explores how we might fundamentally recast Hawkins’ principles as efficiency-oriented ‘third factors’ (Chomsky 2004). The computational optimality of these principles is made explicit in terms of the concepts of* time *and* space complexity*, and they are then used to frame a developing Minimalist picture of the language faculty; linguistic typology (including ‘over-represented’ disharmonic word orders); and the relationship between them. In doing so, adjusted views of markedness and linearisation emerge. Further dimensions of efficiency expressed by the FL are also indicated, introducing a novel approach to the computational optimality of* phases *(Chomsky 2000, 2001). Two recurrent themes are the impressive empirical coverage of a Minimalist theory of grammar, and the improved understanding that emerges from proper dialogue with computational science. An appropriately formulated system of (efficiency-oriented) third factors is suggested as a possible heuristic to guide future inquiry.*

[*] Many thanks to Theresa Biberauer and Ian Roberts for all their challenging and generous supervision over the past three years. My gratitude also to Noam Chomsky, Billy Clark, Kasia Jaszczolt, Derek Murray and Bert Vaux for helpful comments. Word count (incl. footnotes, excl. abstract , tables and bibliography): 19, 678

0. Introduction

In § 1, I introduce and assess the core thesis of Hawkins' (1994, 2004) functionalist theory of grammar, the *Performance-Grammar Correspondence Hypothesis*, and argue instead for an innate, formal grammatical competence. In § 2, I contrast Hawkins' concern for *communicative* efficiency with the concern for *computational* efficiency expressed by the *Minimalist Program* (Chomsky 1995b), reiterating the theoretical arguments adduced for the latter. § 3 recasts the performance principles Hawkins takes to motivate structure as efficiency-oriented 'third factors' (Chomsky 2004), and introduces basic principles of algorithmic science to make their computational optimality explicit. § 4 examines the literature for evidence of the implication of these third factors in the design of the language faculty (FL), during which adjusted views of markedness in acquisition and linearisation emerge. § 5 introduces a related third factor and assesses its role in the design of the FL. § 6 looks at certain aspects of how best to make sense of language typology within this framework. § 7 examines further dimensions of computational optimality seemingly implicated in the design of the FL, including efficiency of *scheduling* associated with phase-based derivation (Chomsky 2001). An expanded system of third factors is then offered as a possible heuristic to guide future Minimalist inquiry. In § 8, I summarise the two recurrent themes of the previous discussion: the impressive empirical coverage of a Minimalist theory of grammar, and the improved understanding that emerges from proper dialogue with computational science.

1. Hawkins' functionalism and its problems

Functionalist explanations of grammar all, to varying degrees, reject the *formalist* position that the structural properties of language may be formulated "without essential reference to matters […] outside of the system of language itself" (Anderson 1998: 11). Instead they pursue the idea that language structure reflects constraints on language use.

Hawkins (1994, 2004) presents an influential and "prototypical" (Haspelmath 2002: 1) example of functionalist thinking. He attempts to motivate a theory of grammar in which

"[e]ven highly abstract and fundamental properties of syntax" (Hawkins 2004: 2) are derived from pressure to maximize efficiency and reduce complexity in performance. This is enshrined in his *Performance-Grammar Correspondence Hypothesis* (PCGH), which holds that "grammars have conventionalised syntactic structures in proportion to their degree of preference in performance" (*ibid.*: 3). We will later expand upon the principles Hawkins takes to constitute performative pressure, but first justify our reasons for rejecting his central thesis in favour of an abstract, innate (potential for) grammatical competence.

First, Hawkins adopts a fluid conception of the traditional *competence-performance* distinction (Chomsky 1965). While he does not commit to the complete absence of abstract competence, much of what would standardly comprise it is instead derived from performance pressure. We take the standard distinction to be uncontroversial: Fromkin (1999) and Newmeyer (2003) provide a summary of just some of the extensive empirical support for this position, drawing on studies of: Specific Language Impairments in children; the asymmetry of abilities in 'savants'; aphasia; sign language; neuroimaging; the psychological status of full argument structure in partial utterances; etc..

Secondly, it is not at all clear how Hawkins' approach accounts for the learnability of grammars in light of *poverty-of-the-stimulus* (Chomsky 1980) considerations. Linguistic competence is radically underdetermined by the data a language acquirer is exposed to: infinitely many systems are consistent with a finite data set, yet inconsistent with one another. Given the speed and accuracy with which language acquirers converge on grammatical systems despite insufficient (and widely varying) input, whatever is responsible must be "biased or constrained in certain ways" (Hauser, Chomsky & Fitch 2002: 1577) to be able to account for it. An innate grammar is then motivated by this need for *explanatory adequacy* (Chomsky 1964).

While Hawkins (2004: 11) acknowledges learnability as a "real issue" that a theory of grammar must face, he argues that an abstract competence cannot be the solution. Addressing the issue through discussion of the 'negative evidence problem' (the question of how an acquirer "manages to infer ungrammaticality from the absence of certain linguistic data, while not doing so for others" (*ibid.*)), he restates the argument from Culicover (1999) that "negative evidence problems reduce to language-particular

idiosyncrasies", and as such, "the whole relevance of UG *[Universal Grammar]* to learnability must be considered moot" (Hawkins 2004: 11). [1] For Culicover (1999: 28) (and Hawkins), a language acquirer is "conservative" and "attentive" in the treatment of arbitrary lexical variation in the input: no underlying orderliness is discernible in the data bearing on a particular syntactic feature, so the acquirer must be attentive to the syntactic properties of individual words, and conservative in only assigning them properties on the basis of positive evidence. This contrasts with a UG-based picture of acquisition, which comprises a greatly simplified process of abstraction from (predominant) underlying orderliness, in accordance with general principles of UG. The language acquirer is (at least largely) *in*attentive to the syntax of individual words, and 'liberal' in assigning them properties despite the absence of outright positive evidence.

It is far from clear that Culicover and Hawkins' objection has any purchase. First, and most importantly, it neglects the fact (just hinted at) that non-UG-assisted learning is entirely compatible with the Chomskyan position. The Chomskyan claim is *not* that there are no aspects of language that are irregular and just have to be learned, but rather that if there is no underlying orderliness in the data, and there are no motivating principles, then the overall task of acquisition becomes impossible. While it may well be that Culicover presents some telling examples of precisely such "language-particular idiosyncrasies", there is more subtlety to the Chomskyan claim than Hawkins acknowledges.

As it is, many of the adduced cases of arbitrary lexical variations have by argued by Coppock (2007) to involve far less idiosyncrasy than at first seems the case. Culicover's (1999: 82) ideas are most clearly expounded with respect to six English prepositions and their putative "odd" behaviour with respect to head-complement order and the amount of material they piedpipe (although other cases are discussed). Coppock (2007: 2) disputes Culicover's claim that these data demonstrate "a number of possible patterns […] with no apparent generalization emerging among the exceptions", and instead suggests how they may in fact be explained in terms of four highly general principles. She goes on to offer (or reiterate) alternative systematic explanations for three other alleged cases of arbitrary lexical variation (further 'odd' prepositions (Huddleston & Pullum 2002, 2003), the dative alternation (Baker 1979), and the causative alternation (Bowerman 1988)). By contrast,

[1] Hawkins takes UG to be synonymous with abstract, innate grammatical competence, although strictly speaking, the term now refers only to its *dedicated* elements. (see § 3).

Coppock mentions just two cases from the literature which she is unable to explain in terms of broader postulates (the raising-alternation difference between *likely* and *probable*, and the category-of-complement distinction between *seems* and *happens* (Baker 1979)). [2]

It seems, then, that "language–particular idiosyncrasies" may be significantly less pervasive than alleged, and it is independently clear that, on a proper understanding of Chomsky's position, the existence of such idiosyncrasies cannot be understood as crucially undermining UG's potential for explanatory adequacy. Furthermore, although Hawkins attempts to justify a search for alternative sources of explanatory adequacy, at no point does he (explicitly or implicitly) demonstrate how exactly it is his approach offers this. And so even under the most generous interpretation, Hawkins' approach offers no better insight into learnability concerns than a UG-based approach.

It should be pointed out that these learnability concerns persist regardless of whether the functionalism is integrated atomistically or holistically (Newmeyer 1998, 2001); that is, regardless of whether the proposed model assumes a direct linkage between functional motivations and grammar, or assumes that "the influence of the former on the latter is played out [indirectly] in language use and acquisition and (therefore) language change" (Newmeyer 2001: 4). Newmeyer's (2003) support for holistic functionalism is motivated by the need to accommodate a conventional competence-performance distinction, as discussed above, but the associated mediation through a system of language-particular rules (Newmeyer 2004, 2006), as Roberts & Holmberg (2005) point out, simply begs the learnability question. Hawkins (2004: 5) acknowledges the alternative ways of assimilating functionalism and refuses to explicitly commit to either. However, at no stage does he implicate any of the rule-based machinery associated with a holistic approach; his approach seems fundamentally "prototypical" and atomistic, and so subject to the first objection as well as this one.

A third problem with the PCGH consists of a fairly fundamental flaw in the supporting argument. Following Hawkins (2004: 6), we take the PCGH to make the following explicit predictions:

[2] It should be noted that these two exmapless are in keeping with the high frequency of presentation we would expect of a learnable true exception. (Biberauer p.c.)

(1) *Grammatical predictions of the PCGH:*

(a) If a structure A is preferred over an A' of the same structural type in performance, then A will be more productively grammaticalized, in proportion to its degree of preference; if A and A' are more or less equally preferred, then A and A' will both be productive in grammars.

(b) If there is a preference ranking A>B>C>D among structures of a common type in performance, then there will be a corresponding hierarchy of grammatical conventions (with cut-off points and declining frequencies of languages).

(c) If two preferences P and P' are in (partial) opposition, then there will be variation in performance and grammars, with both P and P' being realized, each in proportion to its degree of motivation in a given language structure.

Hawkins takes (1a) to be illustrated by the tendency of heads consistently to precede or follow their complements. In a 'VO'-language such as English, heads typically come first – V-NP, P-NP, N-of-NP etc. – a 'lighter' head tending to precede a 'heavier' complement in each case. However, the 'light-before-heavy' tendency in English extends beyond the head-complement relation: the constituents of a VP, for instance, are ordered in the same way:

(2) [V – NP – PP – CP] (*convince my friends of the fact that mushrooms are an abomination*)

Similarly, lone adjectives and participles precede their heads:

(3) (a) *a silly proposal*
(b) *the ticking clock*

But if these adjectives and participles are associated with complements, they must then follow the head:

(4) (a) ** a sillier than any I've ever seen proposal*
(b) *a proposal sillier than any I've ever seen*

(5) (a) ** a ticking away the hours clock*

(b) *a clock ticking away the hours*

(Newmeyer 2003: 683)

Head-final languages, conversely, express a 'heavy-before-light' bias in performance. While these effects are consistent with the PGCH, they are equally consistent with its mirror-image, under which performative ability emerges out of grammatical competence – a *Grammar-Performance Correspondence Hypothesis* (GPCH). Hawkins extrapolates beyond the data, which merely show a correlation, but are mute with respect to causality.

A language acquirer might plausibly set a parameter ordering a head of a certain category and its complement, and the fact that this then tends to result in a light-heavy/heavy-light ordering in surface word order could then influence preferences in production more generally. Indeed this would fit more closely with what we know about the relationship between competence and performance; for example, according to Hale & Reiss (1998: 673):

> *the general conception is that the sensitivity of the speech perception system is generally reduced over time to attend only (or primarily) to those distinctions critical for parsing the target language, while the production system moves from a state of virtually complete inarticulateness to full competence in articulating the target language.*

The second prediction of the PGCH, (1b), is subject to exactly the same concerns as (1a), and the third, (1c), is logically dependent on it.

It seems then that we must question Hawkins's central hypothesis on empirical, (6a); theoretical, (6b); and logical, (6c), grounds:

(6) (a) inconsistence with the empirically-vindicated competence-performance distinction;

(b) failure to diagnose a real problem with a UG-based approach to learnability concerns, or to offer a viable alternative;
(c) failure to justify the choice of the PGCH over the GPCH, particularly pressing in light of some evidence favouring the latter.

2. Efficiency and economy in language

Hawkins's conception of efficiency and economy in language contrasts with the one expressed by the working hypothesis underlying the *Minimalist Program* (MP) (Chomsky 1995b), that "language is an optimal solution to legibility conditions" (the *Strong Minimalist Thesis* (SMT)) (Chomsky 2000: 97). The MP revolves around the desire to give "a principled account" of the linguistic apparatus. In terms of the SMT, the only "(virtual) conceptual necessity" a linguistic derivation must fulfil is legibility to both the *Conceptual-Intentional* (CI) and *Articulatory-Perceptual* (AP) systems. And so an account is considered principled in so far as it manages to relate sound and meaning in an optimal fashion. Concerns for computational efficiency and economy are therefore prior to ones of communicative ease. This is evidenced in the strictly cyclic nature of rule application in syntax – a natural principle of computational optimality in that it involves no 'look-back' into previous computation (see § 4.2.2). As a result of this however, operations tend to yield nested rather than crossing dependencies; in strings such as "the players who John coaches do well" agreement holds between *John* and *coaches*, and *the players* and *do*. Dependencies quickly overflow memory in such centre-embedding environments – "the mouse the cat the dog bit chased died" is more or less unparseable. Nested dependencies are clearly sub-optimal with respect to communicative efficiency. (Chomsky 2007: 4)

It is important to note that the SMT claims that *language* is an optimal solution to interface conditions, not that any single component of language is. For instance, the CI interface seems to be the primary concern for syntax, as the previous example demonstrates. This is also evident in other products of syntactic computation, such as island constructions, requiring circumlocution or special devices (e.g. resumptive pronouns) in usage. (Chomsky 2007: 7) The claim of the SMT is that each component of

the FL is computationally optimal with respect to its role in connecting the two interfaces, not that both interfaces are relevant to every component. The asymmetric implication of the two interfaces in core (syntactic) properties of language suggests that these were selected for on grounds of vastly expanded capacity for "abstract or productive thinking" (Luria 1974: 195), and that ''the role of language as a communication system between individuals would have come about only secondarily'' (Jacob 1982: 59). (cf. Chomsky 2007: 3) However, this claim is separate (though related – cf. n.5) to that made by the SMT, as is made explicitly clear (cf. Chomsky 2005b: 7).

The SMT asserts the "virtual truism [that] natural selection can only function within a 'channel' of options afforded by natural law" (Chomsky 2004: 105). As such, it is part of the broader biological programme investigating the role of a lineage's development in constraining its evolution – the so-called 'evo devo' revolution (cf. Carroll 2005). It has perhaps not always been made as clear as it might have been in what sense the SMT constitutes an evo devo hypothesis, and so it is worth making this explicit here. In claiming the FL optimally satisfies interface conditions, the SMT takes the FL to employ the brain's previously evolved/inherent computational resources, which might reasonably be expected to involve principles of computational efficiency and economy. The strongest interpretation of the SMT then is that the innate grammar required to address learnability concerns is exhausted by these 'cost-free' principles. A more plausible version is that the FL includes a limited amount of language-specific, genetically-encoded apparatus 'co-ordinating' general principles of computation in such a way as to meet conditions imposed by the interfaces. The term UG is now taken to refer only to these *dedicated* aspects of grammar. As such, the evolutionary burden is dramatically eased: only minor genetic changes (UG) are necessary to explain a linguistic phenotype – the central insight of 'evo devo' thought. [3]

[3] The role of evo devo thought in the development of Chomsky's work constitutes an interesting piece of intellectual history. Jacob's suggestions regarding the role of shared regulatory mechanisms in constraining the variety of possible organisms (as discussed in the Kant Lectures, cf. also Chomsky 1975) played a role by analogy in the development of the *Principles & Parameters* (P&P) framework. The parametric structure of innate grammar is compared to shared regulatory mechanisms, and languages themselves to the variety of possible phenotypes. Very small changes in the underlying parametric structure can result in pronounced differences in the resultant language, in the same way that a small change in transcription factor alleles can result in the difference, say, between an elephant and a fly. This analogical articulation of evo devo thought in fact paves the way for its direct implication. Under the P&P conception, it is no longer necessary that all principles of grammar are

Chomsky thereby presents the growth of the FL as akin to that of other biological systems, suggesting three factors are involved:

(7) (a) a genetic endowment specific to language, which interprets part of the environment as linguistic experience;
(b) external data, converted to the experience that selects one or another language from within a narrow range;
(c) principles not specific to the FL.
(adapted from Chomsky 2005a: 6, 2007: 2)

(7a) consists of UG and must include at least (unbounded) *Merge* and atomic lexical items, the minimum apparatus necessary to support recursion. (7b) consists of the *primary linguistic data* (PLD) interpreted as relevant to determining the final state of the grammar. (7c) consists of so-called 'third factors' – general cognitive principles that "enter into all facets of growth" (Chomsky 2007: 2). Chomsky (2005a: 6) categorizes these third factors as follows:

(8) (a) *principles of data analysis that might be used in acquisition;*
(b) *principles of structural architecture and developmental constraints;*

to which he adds;

(c) *properties of the human brain that determine what cognitive systems can exist*
(Chomsky 2007: n6).

An account of grammar is now "principled" in so far as it implicates third factors in satisfying interface conditions. Principles of computational efficiency and economy fall into the second category, (8b). Superficially, Chomsky's categorization might be taken to divorce these principles from principles of data analysis, (8a). However, given their general cognitive status, we would expect principles of efficient and economic computation to be implicated in principles of data analysis in as far as data analysis is

involved in acquisition – acquisition merely involves parametrised principles – and so much of grammar may then be non-dedicated. (Chomsky 2005b: 3) Thereby, evo devo as *analogy* allows us to think of evo devo as *reality*.

computational. Similarly, the organic properties of the brain must be considered relevant to delimiting possible principles of data analysis and of structural architecture. The categories of third factors cannot be considered to be in some kind of complementary distribution (and at any rate this was not the intended reading (Chomsky p.c.)). All this classification reflects is that some principles of data analysis are not coextensive with efficiency-oriented principles of architecture (see n.14 below); some principles of structural architecture aren't directly determined by organic properties of the brain; and so on. The notion of third factors is a deliberately incoherent term, being a catch-all for anything at any level that is relevant to the growth of language that isn't language-specific apparatus or linguistic data. [4, 5]

Strangely, Hawkins (2004: 267) appears utterly unaware of these developments in the generative enterprise, enervating his further criticism. He asserts that there has been an "increasing disconnect between formalism and the search for explanatory principles", a statement utterly inconsistent with the Minimalist Program's search for "principled explanation." He further suggests that "the innateness of human language resides primarily in 'architectural innateness'", and that "many […] features of UG, possibly all, will be derivable from [communicative] efficiency and complexity considerations" (*ibid.*). Given that this claim is a direct functionalist translation of the SMT, it is hard to see how it squares with his previous censure. The problem is revealed by his (2004) bibliography, which lists nothing by Chomsky since *Barriers*, published eighteen years previously, and so nothing from his (then) decade's worth of writing on the MP. This is scarcely grounds for asserting anything is increasing – especially when it demonstrably is not.

[4] Aside from the fact that third factors may actually exist, this mirrors the terminological distinction between *E-* and *I-language*: E-language has no coherent characterization except with respect to not being I-language.

[5] It is of course possible to entertain a Hawkinsian notion of efficiency in explaining the evolution of the language faculty, invoking progressive natural selection in accordance with communicative demands, as pursued by Pinker & Jackendoff (2005). As suggested however, both the seeming primacy of the CI interface in core properties of language and concerns for the evolutionary burden seem to render the fundamentally non-adaptive approach preferable. This view relates the primacy of the CI-interface and computational optimality (see pp. 8), since if core properties of language placed equal emphasis on both interfaces, this would indicate evolutionary pressure toward successful communication. Given that there would clearly be more complex selective conditions involved in satisfying two interfaces than in satisfying one, we might expect to see a more idiosyncratic, adaptive design to the FL.

So, although the MP is part of a theoretically coherent biological enterprise, there still remain the empirical question of whether it is born out, and the methodological question of how best to approach it. These are issues I hope to address below.

3. Finding a use for Hawkins' principles

Despite the outlined difficulties with taking performance pressure to motivate grammatical competence, Hawkins' specific formulations for the principles taken to compose it might still prove instructive. As discussed, these principles are taken to maximize efficiency and reduce complexity in fulfilling "the basic function of language, to communicate information from the speaker to the hearer" (Hawkins 2007). He defines his three central principles as follows:

(9) ***Minimize Domains (MiD)***

The human processor prefers to minimize the connected sequences of linguistic forms and their conventionally associated syntactic and semantic properties in which relations of combination and/or dependency are processed. The degree of this preference is proportional to the number of relations whose domains can be minimized in competing sequences or structures, and to the extent of the minimization difference in each domain.

(Hawkins 2004: 32)

(10) ***Minimize Forms (MiF)***

The human processor prefers to minimize the formal complexity of each linguistic form F (its phoneme, morpheme, word or phrasal units) and the number of forms with unique conventionalised property assignments, thereby assigning more properties to fewer forms. These minimizations apply in proportion to the ease with which a given property P can be assigned in processing to a given F.

(*ibid.*: 38)

(11) ***Maximize Online Processing (MaOP)***

The human processor prefers to maximize the set of properties that are assignable to each item X as X is processed, thereby increasing O(nline) P(roperty) to U(ltimate) P(roperty) ratios. The maximization difference between competing orders and structures will be a function of the number of properties that are misassigned or unassigned to X in a structure/sequence S, compared with an alternative.

(*ibid.*: 51)

If Hawkins' proposed principles are to be of any relevance however, they must be radically reconceived. Instead of acting as constraints rooted in performance they must be recast as general computational constraints – tentatively explicit formulations of some 'efficiency-oriented' third factors. In order to do this, we must remove any suggestion of weighting/preference, or of performance/"processing" [6] from the definitions, since the principles will be expected to serve in a discrete, formal grammar. [7] We also remove any reference to language from the definitions, since third factors are taken to be general principles of cognition operating in the FL. This results in the following:

(12) **Minimize Domains$_1$ (MiD$_1$)**

Human computation minimizes the connected sequences of forms and associated properties in which relations of combination and/or dependency are considered.

(13) **Minimize Forms$_1$ (MiF$_1$)**

Human computation minimizes the formal complexity of each form F and the number of forms with unique property assignments, thereby assigning more properties to fewer forms.

[6] In its functionalist usage, the term 'processing' is inherently related to performance. We wish to reserve the term for the abstract computational sense of performing 'operations', in which it applies equally to any module involving computation.

[7] We exempt the name *Maximise Online Processing* itself to emphasize the continuity.

(14)**Maximize Online Processing$_I$ (MaOP$_I$)**

Human computation maximizes the set of properties that are assignable to each item X as it is considered, thereby increasing I(nitial) P(roperty) to U(ltimate) P(roperty) ratios. [8]

We now need to establish in what, if any, sense these proposed (efficiency-oriented) third factors in fact offer computational optimality. In other words, we must ask: *what does it mean to be computationally optimal?* To address this question, we introduce from computational science the two fundamental concepts of *complexity theory – time* and *space complexity*. The time complexity of an algorithm is the number of operations it must perform to complete its task. The space complexity of an algorithm is the amount of working memory it requires while running. (Manber 1989: 42-3) Reducing the number of operations an algorithm must perform reduces the time required by a given processor to complete the task. Reducing the burden an algorithm places on working memory allows more such algorithms to run in parallel, improving the speed with which a number of tasks can be performed. Therefore, there is computational optimality associated with reducing both the operational load (time complexity) and memory/cache load (space complexity) imposed by an algorithm.

In the above we make the somewhat simplified assumption that time and space complexity are entirely independent. In the design of most algorithms however, there is a *trade-off* of one for the other (Hellman 1980) – sacrificing operational complexity to reduce cache load, or, more commonly, vice versa. In practise, one of available processing power and available working memory must always limit the speed of computation. There will inevitably be a 'bottleneck' at which one of the two is maximally tasked, and so reducing the burden the algorithm places on the other will make no difference to overall efficiency. Either, algorithms are running to completion faster than the space in working memory can be cleared for a new algorithm to begin – in which case reducing the time complexity of the algorithm will have no effect on overall performance, and the available working memory/the algorithm's space complexity is the limiting factor. Or, sufficient memory is available for a new algorithm to begin, but the processor is already performing as many operations per second as it can – in which case reducing the space complexity of

[8] The use of 'initial' and 'ultimate' here refers to some abstract computational sense, which may or may not translate into a temporal one.

the algorithm will have no effect on overall performance, and the available processing power/the algorithm's time complexity is the limiting factor. [9] As a consequence, there is also computational optimality associated with striking the balance between time and space complexity in an algorithm's design in such a way as to maximally tax the available resources. The point of optimal balance depends of course on the relative 'cost' of working memory and processing power. However, investigation into the role of time and space complexity in the design of the FL's 'algorithms' is at such a nascent stage, and so little is known about the processing capacity and working memory available to them, that speculation regarding this source of optimality is hopelessly premature. For now, we pursue the more modest ambition of tracing the optimality associated with reducing each type of algorithmic complexity independently.

In what sense then do the third factors MiD_1, MiF_1 and $MaOP_1$ reduce the time and/or space complexity of the FL's algorithms? Let us consider each in turn. MiD_1 reduces the numbers of forms that computations performed by the FL must search before the appropriate 'feature' (some abstract 'unit' of formal complexity) is encountered. A reduction in the number of search operations directly reduces time complexity. MiF_1 reduces the number of features associated with each form and the number of forms postulated. There will an operation associated with processing each feature, so reducing the number of features reduces the number of operations that have to be performed. Similarly, there will an operation associated with 'stopping' one aggregation of features, or 'form', and 'starting' another. As such, reducing the number of forms also reduces the

[9] There is some debate as to whether working memory used in linguistic computation is specific to it (cf. Caplan & Waters 1999) or of a general cognitive character (cf. Baddeley 1986). Obviously this is fundamentally a question for neuro- and psycholinguistics, but if we were to treat the brain as a system of modular processors, dedicated to particular tasks, as is perhaps argued by Broca's agrammatism (although see Ingram (2007: 297-330) for a full appraisal of this as evidence), it is then analogous to a *multiprocessor computer structure*. There is then reason to expect *distributed memory* (i.e. language-specific working memory) to be the optimal scheme, as computer scientists have uncovered. Distributed memory is preferred to shared memory for at least three reasons. First, each processor is then able to get full access to its own local memory without interference from other processors. Secondly, the absence of common path into memory (or 'bus') means that there is "no inherent limit to the number of processors" (CSEP: § 3.3.3). Thirdly, "there are no cache coherency problems[: e]ach processor is in charge of its own data, and it does not have to worry about putting copies of it in its own local cache and having another processor reference the original." (*ibid.*)

algorithms' operational load. $MaOP_1$ essentially insists that properties be assigned as soon as possible in computation, minimising the number of underdetermined features that must be carried over until a later point. Any such features will of course have to be stored in working memory, and so reducing the number of them reduces the algorithms' space complexity. In summary then, MiD_1, MiF_1 and $MaOP_1$ each manifest computational optimality – the first two with respect to time complexity, and the third with respect to space complexity.

4. Efficiency-oriented third factors and the design of the language faculty

Using the efficiency-oriented third factors presented above as a framework, we now attempt to present the emerging Minimalist picture of the language faculty, rather than just Minimalist syntax. This is merely the logical conclusion of Chomsky's approach and indeed actively pursued in various of his observations. He suggests the paratactic nature of structural planning/parsing springs from the same minimal search concerns as evidenced in syntax (see § 4.2), and also that "mappings to SM *[i.e. phonology]* would be the 'best possible' way of satisfying the externalization conditions" (Chomsky 2007: 10) (after all the SMT refers to both interfaces). Given their status as general cognitive principles, we would expect efficiency-oriented third factors to be implicated in all components of the FL in as far they involve computation. [10] The faculty for language includes (but is not exhausted by): the formal competencies of syntax and phonology; acquisitional / 'data-analytic' faculties (on which more in § 4.1.5); and the performance faculties of pragmatic competence and structural planning/parsing. [11, 12] This mirrors Chomsky's (1986) guidelines regarding the initial research questions of linguistic inquiry:

[10] If, as Jacob suggests, the role of language in communication were evolved secondarily, we would expect the design of modules of externalisation to also reflect communicative concerns, that is, non-third factors. A well-known example is provided by the non-configurational effect of semantic plausibility on ease of sentence parsing (cf. Solomon & Perlmutter 2004). Another example is the putative role of 'naturalness' in the design of phonology (Chomsky & Halle 1968), as expressed, for instance, in Kenstowicz & Kisseberth's (1979: 251) suggestion that "phonological theory must have some apparatus for expressing the fact that neutralization to a glottal stop [in preconsonantal and final position] is a natural rule as opposed to, say, neutralization to *l*." (cf. Blevins 2004, especially Chapter 3, for a full discussion of the history of 'naturalness' in phonological theory)

namely, what is knowledge of language, how do we acquire it, and how is it put to use? I now attempt to demonstrate for each of the suggested efficiency-oriented third factors how it is implicated in the design of the FL, synthesizing existing findings of the literature and original observations. The order of consideration reflects nothing other than expositional concerns.

4.1 Minimise Forms$_1$

4.1.1 Syntax

The implication of formal economy considerations in syntactic computation was central to the development of the Minimalist Program, as part of *Bare Phrase Structure* (Chomsky 1995a) and the *Inclusiveness Condition* (Chomsky 1995b). Bare Phrase Structure precludes the projection of vacuous or redundant structure in the phrase marker, minimizing the number of (*Merge*) operations associated with introducing formal content. The Inclusiveness Condition holds that "[n]o new objects are added during the course of the computation apart from rearrangements of lexical properties (in particular, no indices, or levels in the sense of X-bar theory, etc.)" (Chomsky 1995b: 228) This also clearly serves to minimize the formal content of a derivation and associated operations.

[11] This is almost certainly a simplified conception of structurally oriented performative ability – the elements interacting with pragmatic competence and with syntax are perhaps better conceived as separate competencies. However, nothing in the argument below hinges on a precise formulation, so we treat them together without cost.

[12] The omission of phonetic modules of speech perception and production here is not a theoretical claim about whether or not "speech is special" (attributed to Alvin Liberman), but rather motivated by the observation that phonetic abilities are not computational, at least not in the relevant sense. As pointed out by Hale & Kissock (forthcoming), the mappings from the *S(urface) R(epresentation)* of phonology to gestural score, or from acoustic score to SR are merely *transductional*. That is to say they do not in any way change the computational 'structure' of the input, merely changing the 'medium', such that it is relevant to the articulators or phonology respectively. 'Ease of articulation' concerns like elision and assimilation instead reflect physical/communicative economy concerns applying to the articulation itself.

4.1.2 Phonology

(i) When *Internal Merge* (i.e. *Move*) acts in a syntactic derivation it creates multiple copies of an element, all of which are transmitted to the phonological component. However, phonology will (normally) only *Spell Out* the highest copy. This clearly minimises the number of operations associated with formal content in the phonological computation, in accordance with MiF_1. (It is also further evidence that language is not optimised with respect to communication, since by not spelling out all copies phonology creates *filler-gap* problems for the parser (cf. § 4.3.2 below). (Chomsky 2005a: 13))

(ii) Hale & Kissock (forthcoming: 7) discuss a type of phonological underspecification "which persists from underlying representation through phonetic representation, resulting in forms which are *never* fully specified featurally", so-called *perseverant underspecification*. In such cases, "articulation […] is determined in some relevant respect by context" (*ibid*.), rather than the fully specified *S(urface) R(epresenation)* output of phonology. This contrasts with the more commonly discussed type of underspecification in which featural underspecification at the underlying, lexical level is resolved in the course of phonology to result in full phonetic specification at the SR level (cf. § 4.2.2 and § 5.2 below). When perseverant underspecification obtains, the unspecified feature is simply never processed by phonological computation, clearly implicating MiF_1.

4.1.3 Structural planning / parsing

The drive toward brevity of utterance is well documented in the literature. Grice (1967, 1989: 27) formalises it in his third 'Maxim of Manner' as "be brief (avoid unnecessary prolixity)" – i.e. plan your utterance using as little form as possible. Levinson (2000) subsumes Grice's maxim under the notion of "marked messages" in his 'M-principle', where 'markedness' is increased by being "more prolix or periphrastic" (*ibid.*: 137):

(15) *The M-Principle:*

> *Speaker's maxim: Indicate an abnormal, nonstereotypical situation by using marked expressions that contrast with those you would use to describe the corresponding normal, stereotypical situation.*
> *Recipient's corollary: What is said in an abnormal way indicates an abnormal situation, or marked messages indicate marked situations.*
> (*ibid*.)

The Gricean reasoning then is that when speakers have gone out of their way to avoid the simpler expression, they must be trying to avoid whatever the simpler expression would suggest. A markedly long utterance has a marked interpretation. Various practises in usage demonstrate this effect. For instance, McCawley (1978) observes that a lexicalised causative is standardly associated with direct causation, and a periphrastic one with indirect causation. "Larry stopped the car" communicates 'Larry caused the car to stop in a normal fashion, by using the foot pedal', whereas "Larry got the car to stop", or "Larry caused the car to stop", communicates 'Larry caused the car to stop in a nonstereotypical way, e.g. using the emergency brake.' (Levinson 2000: 141) Similarly, the use of a double negative generally results in a marked interpretation with respect to the logically equivalent single positive. For instance, "it took not inconsiderable effort" communicates 'it took a close-to-considerable effort'. (*ibid*.: 144) Levinson (*ibid*.: 138-53) provides a summary of further empirical evidence for the economic use of (linguistic) form in utterances. Operations are minimized in the computation of the structural plan by reduced formal content, demonstrating the influence of MiF_1.

4.1.4 Pragmatic competence

The drive toward semantic minimality of utterance is also well documented in the pragmatic literature. Grice (1975) observes that speakers "do not make [their] contribution[s] more informative than required" (his second 'Maxim of Quantity'). Levinson (2000: 114) subsumes Grice's maxim under his 'I-principle':

(16) *I-Principle*

Speaker's maxim: the maxim of Minimization. "Say as little as necessary", that is, produce the minimal linguistic information sufficient to achieve your communication ends […]

Recipient's corollary: the Enrichment Rule. Amplify the informational content of the speaker's utterance, by finding the most specific interpretation[.]

Speakers then will incorporate as little 'informational form' as possible into their utterance, instead relying on general inferential abilities to enrich the content. A prominent example of 'I-enrichment' obtains in cases of event conjunction, which tend to be interpreted as temporally successive and, if plausible, causally related. For instance, "Amelia wrote a novel and made lots of money" communicates that Amelia made a lot of money from the sales of the novel she had written. This is significantly more information than is included in the logical form of the sentence, which is mute with respect to both the order of the events and whether they are related. Levinson (2000: 122-34) provides further exemplification of economic use of informational content in utterances. All other things being equal then, a semantically-impoverished utterance is preferred – a clear implication of MiF_1 minimizing operations associated with computation of informational form. [13]

4.1.5 'Data-analytic' / acquisitional faculties

We may think about the role of (certain) principles of data analysis in language acquisition as follows. Something must act in conjunction with innate grammatical knowledge to abstract from the primary linguistic data (PLD) the patterns that determine grammatical properties. We may think of the faculties that do so as general principles of data analysis, or, in Wexler (2004)'s terms, *general learning mechanism.* [14]

[13] It is suggestive in light of our argument here that Horn (1984) conflates the drives toward structural and semantic minimality into a single maxim: 'The R Principle' – "Make your contribution necessary […] say no more than you must".

[14] There may be other general principles of data analysis involved in acquisition that are 'non-abstractional', but govern the way in which the target grammar is selected from the range of possible grammars, such as Yang's (2002) variational approach. While this is not so obviously the reflex of

As discussed in § 2 above, efficiency-oriented third factors may be implicated in principles of data analysis in as far as they are computational; and mapping from PLD to (patterns determining) grammatical settings is clearly a complex computational task. By hypothesis therefore, a language acquirer will 'seek' to map their PLD onto grammatical settings in a computationally optimal fashion, using the fewest possible operations and the least possible amount of working memory. In other words, the acquisitional faculties are 'predisposed' to interpreting the PLD as reflecting whatever grammatical settings place the smallest possible computational burden on them. That is, *markedness* (Jakobson 1941) in acquisition is associated with economic data processing. [15]

This contrasts crucially with what has standardly been assumed about markedness since the seminal work of Chomsky & Halle (1968) in relation to their phonological distinctive-feature system. For Chomsky & Halle, the markedness asymmetry between the values of features relates to the *relative simplicity of the grammars acquired*, as determined by an evaluation metric they propose as part of synchronic grammatical knowledge; "the unmarked value of a feature [is] cost-free with respect to the evaluation metric, while the marked values [are] counted by the metric." (Battistella 1996: 75)

Much the same approach to markedness in acquisition is widely assumed in both the acquisitional and the diachronic literature. To take an example from the former, Roeper (1999) implicates "economy of representation" as a principle of UG determining the acceptance/rejection of grammars in acquisition. Similarly, Clark & Roberts (1993: 301) suggest acquisition is motivated by "elegance of derivations" and that "children avoid grammars that create inelegant representations" (*ibid*.: 342).

---

economy concerns, "if investigated further, [Yang's] suggestions might also [be found to] involve efficiency-oriented principles, as I suspect he'd agree." (Chomsky p.c.) That is, it may be that pure computation is employed more abstractly in a variational principle, so economic aspects are not writ as large.

[15] That is to say, complexity-oriented markedness. We are not denying the possibility of substantively motivated, or *natural*, markedness in the synchronic design of certain aspects of grammar (cf. the distinction between *formal* and *substantive* markedness in Roberts & Roussou 2003: 214). Chomsky & Halle (1968) introduce this as their second notion of markedness in phonology (cf. n. 11 above), and in this module at least, it is compatible with the approach we are taking here, since the modules of externalisation may reflect non-third-factor concerns in their design.

The diachronic literature that bears on this issue assumes an acquisition-based approach to language change (Paul 1920) that falls out of the indirect transmission of a grammar across generations: "grammars are mental entities and it is impossible to have direct access to the contents of another mind" (Roberts 2007: 124), and so "grammars must be constructed by the individuals of each generation" (Lightfoot 1979: 391).

Kiparsky (1965) was among the first to implicate conservative acquisition in explaining commonly observed patterns of language change. He adopts the evaluation metric proposed by Chomsky & Halle and takes it to motivate the "directionality" (*ibid.*: 54) of phonological change toward "*simplification* [emphasis PK's own] of the grammar" (*ibid.*: 53). Roberts & Roussou (2003) adapt Chomsky & Halle's notion of markedness to the syntactic domain. They propose that diachronic syntax reflects language acquirers' "general preference for simplicity of representations" (Roberts 2007: 208), evaluated in accordance with their feature-counting metric:

(17) *Given two structural representations R and R' for a substring of input text S, R is simpler than R' if R contains fewer formal features than R'.*
(as in Roberts 2007: 235 – slightly simplified from Roberts & Roussou 2003: 201)

As suggested, the approach to (complexity-oriented) markedness that is pursued in the literature is incoherent with the view of cognition we adopt here. Efficiency must be considered a property of *computation itself*, not of its *output*. There is no principled account available for an acquirer minimizing properties of the grammar postulated on the basis of elements of its innate knowledge. The SMT / 'evo devo' hypothesis holds that we 'count' features of computation, not features of whatever it generates. This is a subtle, but theoretically important difference.

Markedness conventions (complexity-oriented or otherwise) as part of innate knowledge have sometimes been construed as a design feature to further simplify the acquisition process in the case of pervasive ambiguity in the PLD. The suggestion is that grammatical knowledge encodes some kind 'default' settings that a language acquirer can revert to in cases where especially impoverished external evidence does not provide sufficient grounds for selecting one feature over others. Bickerton's (1981) *Language Bioprogram Hypothesis* of creole genesis holds that a new system is essentially 'invented' in cases of

simplified pidgin input on the basis of "an unmarked set of grammatical settings" (Bickerton 1988: 282) coded in UG. Similarly, Clark & Roberts (1993: 302) suggest that a learner is forced to "turn in on itself, abandoning external pressure" in cases where grammatical settings are "underdetermined by the linguistic environment" (*ibid*.).

However, the fact that certain 'unmarked' properties of grammars tend to appear more regularly when ambiguity in the input is prevalent should not be taken as evidence that these properties are innately specified and necessary to allow acquisition. The 'emergence of the unmarked' in such circumstances is exactly what would be expected under a view that associates markedness with the data processing. When presented with limited input, inaccurate matching of the target grammar is far more likely. The balance of misassignments would be expected to fall in the direction of properties favoured by the computation making them. While unmarked settings might superficially appear to operate more 'forcefully' over weak input, it is simply the case that the effect of economic computation is superficially manifest to a greater extent, because it isn't 'defied' by the unambiguous force of the data toward a particular grammatical setting. Just because you don't see the evidence of economic computation when there is strong determination of grammatical settings does not mean it is inoperative; just because unmarked settings 'emerge' in impoverished circumstances does not mean they were predetermined and necessary to 'bridge the gap'. There is no reason to assume any sort of dichotomy in the acquisition process whereby external evidence selects the grammar in some cases, but an innate grammatical setting does in the other. The 'emergence of the unmarked' in impoverished circumstances can therefore not be taken to argue that markedness (complexity-related or otherwise) should be construed as part of the innate knowledge. Apart from concrete suggestions as to how unmarked grammatical settings might arise in terms of efficient computation, this is approximately what is argued by DeGraff (2003, 2004, forthcoming) in his suggestion that Bickerton's *Language Bioprogram*-determined default grammatical settings are epiphenomenal and unnecessary to explain creole genesis once substrate effects are considered.

Returning now to strictly complexity-oriented markedness, we reiterate that associating complexity-oriented markedness with the simplicity of the grammar acquired is theoretically unsustainable, and at any rate unnecessary to explain the emergence of the unmarked. However, it should be noted that the confusion is perhaps understandable,

given that the empirical predictions of the two different approaches cannot be distinguished in this case. MiF_1 operating in the acquisitional computation will minimize the number of features/forms entertained in a possible grammar, in much the same way a feature/form-counting metric will. However, the two approaches do not always coincide in their predictions, and we will see in § 6 that the 'simple acquisition' view has the better empirical coverage.

We now follow up Clark & Roberts' (1993: 302) suggestion that "diachronic change can represent crucial information on those factors that learners rely on to select hypotheses", implicit in an acquisition-based approach to diachrony, and present some examples of the operation of MiF_1 in data processing. [16]

(i) Roberts & Roussou (2003) demonstrate how the directionality of *grammaticalisation* pathways can be characterised in terms of the loss of formal content – either movement-triggering EPP-features or substantive features. For instance, in the history of French, null indefinite determiners were lost, such that DPs with null Ds could no longer be referential. This created ambiguity as to the categorial status of certain nouns with 'generic' semantics, such as *point*, *rien* and *personne* ('little bit', 'thing' and 'person' respectively) often used referentially in expressions such as (18a), with the structure of DP as in (18b). This ambiguity was resolved by their reanalysis as non-referential quantifiers (the first stage of their development into clausal negators), assigning the DP structure in (18c).

(18a) *... ja por rien nel te deisse se* ***point de ton bien*** *i veisse*
… already for nothing not-it you I would-say if **bit of your goods** there would I see
'I would not tell you if I saw the smallest piece of your goods'
(*P.* 7261-3; Foulet 1990: 268)

(18b) $[_{DP}\ [_{D}\ Ø]\ [_{NumP}\ [_{Num}\ \textit{point}]\ [_{NP}\ t_{\textit{point}}\ ]]]$

(18c) $[_{DP}\ [_{D}\ Ø]\ [_{NumP}\ [_{Num}\ \textit{point}]\ [NP]]]$

[16] See Yang (2002) for an illustration of the role of his variational principle of data processing in historical explanation.

(*ibid*.: 196)

That the ambiguity was resolved in this fashion (instead perhaps of insisting on the inclusion of an indefinite or partitive article) is in accordance with the miniminisation of formal content entertained by the acquisitional faculties, since (18c) lacks a movement-triggering EPP-feature relative to (18b).

An example of a grammaticalisation pathway characterised by the loss of substantive features is afforded by the development of adverbal clitics from strong pronouns (via weak pronouns). (Roberts 2008b) Following Cardinaletti & Starke (1999), the structure of a pronominal DP may be represented as:

(19) $[_{DP}\ D\ [_{\varphi P}\ \varphi\ ([_{nP}\ n\ [_{NP}\ N]]])$ [17]

The grammaticalisation path from strong pronoun, to weak pronoun, to adverbal clitic is characterised by the loss of substantive formal content at each stage. The first stage involves the loss of D's [Case] feature, whereas the second involves the loss of the D-layer completely, leaving φP. Clitics that are D-elements are associated with C, whereas those that are φ-elements are associated with *v*. The pathway predicted by conservative acquisitional faculties should be manifest as a diachronic shift from 'second-position' (i.e. C-associated) to adverbial cliticisation. And this has indeed been observed for a range of languages, including:

(20) (a) *vemos que lo* ***non fazen*** *asi* [Old Spanish: Rivero 1991: 250]
we-see that it not they-do thus
(b) *vemos que no* ***lo hacen así*** [Modern Spanish]
we-see that not it they-do thus
"We see that they do not do it thus."

(21) (a) *uže* ***ti*** *neprijaznъ* ***ne*** *oudôbъjajetъ*
no-longer you disfavour not rules
"Disfavour no longer rules you."
[Old Church Slavonic, Pancheva 2004: 116]

[17] Weak pronouns and clitics might be taken not to include the *n*-phase, being entirely inflectional.

(b) ... *če nie* ***veče mu*** *pomognaxme*

… that we already him helped

"… that we already helped him"

[Modern Bulgarian, Pancheva 2004: 110]

(Roberts 2008b: 6-7)

This historical tendency to reduce the number of features used to characterize input structure clearly reflects the influence of MiF_1 in the acquisitional faculties.

(ii) Clements (2003, forthcoming) demonstrates that sound systems tend to use featural contrasts with "maximal efficiency", expressing "featural economy". This can be demonstrated using the consonant inventory of three different languages:

(22)

Three consonant systems

a. *Hawaiian: 8 consonants*
(after Elbert & Pukui 1979)

p	k	
m	n	
w	l	
	ʔ	
	h	

b. *French: 18 consonants*
(after Dell 1985)

p	t		k
b	d		g
f	s	ʃ	
v	z	ʒ	
m	n	ɲ	
	l		ʁ
		j	

c. *Nepali: 27 consonants*
(after Bandhu *et al.* 1971)

p	t	ts	ʈ	k
pʰ	tʰ	tsʰ	ʈʰ	kʰ
b	d	dz	ɖ	g
bʱ	dʱ	dzʱ	ɖʱ	gʱ
		s		ɦ
m	n			(ŋ)
	l,r			

(reproduced from Clements 2003: 288)

In Hawaiian, three manners of articulation (stop, nasal and approximant) cross-classify two places of articulation (labial vs. non-labial) *without gaps* to give six consonants. In French, voicing is fully exploited in stops and fricatives to double the number of obstruents. In Nepali, five places of articulation fully cross-classify four manners of articulation within its stop system. (*ibid*.)

Featural economy can be related to the "decreasing the number of features while holding the number of sounds constant" (Clements (forthcoming): 21), as reflected in "the frequent historical elimination of 'isolated' sounds that do not fall into regular patterns of correlation with other sounds" (*ibid*.: 21-2), such that "the feature that previously characterized them becomes redundant." (*ibid*.: 22) An example of this is provided by a development in realisation of Zulu plosives:

(23) Two stages in the realisation of Zulu plosives:

(a) stage 1

p'	*t*	*K'*
p^h	t^h	k^h
b	*d*	*G*
		K

(b) stage 2

p'	*t*	*K'*
p^h	t^h	k^h
p	*t*	*K*
ɓ		*G*

(Clements 2003: 317)

At stage 1, the Zulu plosive inventory contained two isolated stops, the implosive ɓ and the plain voiceless k, both of which were the sole members of their series. In the course of developing to stage 2, the voiced stops become devoiced and the two isolated sounds shifted into a single voiced series, shown in the last row. The feature that previously distinguished the implosive from its plain voiced counterpart (probably [-obstruent]) has been eliminated. (Clements 2003: 317-8)

Less form is postulated to capture distinctions if at all possible, reducing the operational complexity of acquisitional computation in accordance with MiF_1.

(iii) The reflex of MiF_1 in acquisition can be seen in the diachrony of morphology also. Schleicher (1861-2: 4, 342-3) proposes a general pathway of morphological change that

languages follow – from isolating, to agglutinating, to inflectional in a unidirectional fashion. We focus here on the trend from agglutinative synthesis toward inflectional/fusional synthesis, as perhaps evidenced in the development from Proto-Indo-European to its ancient world descendants, such as Greek, Latin and Sanskrit (Crowley 1992: 132-4; Lehmann 1962: 52). In as much as a language demonstrates this trend, [18] it is because of directionality at the level of the individual grammatical changes – in this case, the prevalent fusion of two morphemes into a single morpheme rendering the morphosyntactic features of both. The associated syncretism of morphologically-marked features is consistent with the activity of MiF_1 in the acquisitional computation. As discussed in § 3 every different aggregation of features, or *form*, that must be computed will be associated with an additional operation. Syncretism minimizes this effect, and so will be a favoured result of optimal data processing.

To take a concrete example of an individual change implicating MiF_1 in this way, let us look at a development in Proto-Slavic (Migdalski 2006: 14). Consider the reconstructed paradigm of the Proto-Slavic verb **nesti* 'to carry' (Długosz-Kurczabowa & Dubisz 2001: 265). The first element of the verb is the root; the second the thematic suffix, and the final element carries inflectional morphology.

(24) The paradigm of Proto-Slavic **nesti* 'to carry' in the present tense:

	SINGULAR	DUAL	PLURAL
1	*nes-ō-mь*	*nes-e-vě*	*nes-e-mъ*
2	*nes-e-šь*	*nes-e-ta*	*nes-e-te*
3	*nes-e-tь*	*nes-e-te*	*nes-o-nti*

A phonological change in Proto-Slavic nasalized the vowel *ō* when it preceded a nasal consonant, as in *nes-ō-mь* and *nes-o-nti*. This created ambiguity with regard to the presence of the two separate morphemes in the 1st pers. sing. and 3rd pers. pl. forms of the verb. All the other forms still suggest that two morphemes are present, but from a

[18] The pathway is clearly not exceptionlessly unidirectional: for instance Hindi, a descendant of Sanskrit, seems to have moved back toward agglutination (Hock & Joseph 1996: 183). However, absolute unidirectionality was never really to be expected, given the contingence of the *index of fusion* on other factors, being inversely related to the *index of synthesis* (Comrie 1989: 48-9). This synchronic relationship between the two must of course have a diachronic dimension, so unidirectionality could be perhaps breached, for instance, as a result of independent increase in the degree of synthesis.

phonological perspective, in these two cases it is no longer clear if the underlying form has a vowel and a nasal consonant, or just a nasal vowel. This ambiguity was resolved in favour of merging the thematic and inflectional features onto one morpheme, as shown in (25).

(25) The paradigm of **nesti* 'to carry' in the present tense (later version):

	SINGULAR	DUAL	PLURAL
1	***nes-õ***	*nes-e-vě*	*nes-e-mъ*
2	*nes-e-šь*	*nes-e-ta*	*nes-e-te*
3	*nes-e-tь*	*nes-e-te*	***nes-õtь***

(Długosz-Kurczabowa & Dubisz 2001: 265)

While this reanalysis is contrary to the evidential force of the rest of the paradigm, it is in accordance with the force of MiF_1 in acquisitional computation, since it removes the need for an unnecessary operation associated with postulating a null second morpheme.

Each of the above examples, then, demonstrates the minimization of operations performed in postulating features and/or featural aggregations to characterize the input, clearly implicating MiF_1 in the data-analytic computation.

4.2 Minimise Domains$_1$

4.2.1 Syntax

Minimal search considerations in syntax are familiar from the Minimalist literature:

(i) The *Extension Condition* (Chomsky 1995a: 248, 254) on syntactic derivation holds that (*Internal* and *External*) *Merge* must always be at the *root* of the tree, *extending* the phrase marker. This results in Strict Cyclicity of syntactic functions, as discussed in § 2. If there were more than one possible site of *Merge*, it would be necessary to search for the appropriate site in each case, increasing the operational load on computation.

(ii) The *goal* of a *probe* (a head) is limited to its c-command domain and therefore its complement; an *Agree* relation cannot be formed between a probe and an element in its specifier. (Chomsky 2007: 6) Furthermore, *Agree* is subject to strict locality considerations within this limited domain: the *Minimal Link Condition* (MLC) holds that the goal is the nearest feature that can enter into an agreement relation with the probe (Chomsky 1995b: 297, cf. Rizzi's 1991 *Relativised Minimality*). This could be formalized in the manner Rizzi (2001: 90) suggests:

(26) *Y* [a probe] *is in a Minimal Configuration (MC) with X* [a goal] *iff there is no Z such that:*
(i) *Z is of the same structural type as X, and*
(ii) *Z intervenes between X and Y.*

This effect blocks the raising of *wh*-element from an adverb position across a *wh*-subject in an indirect question:

(27) * How did you ask who could solve the problem <how>?
(*ibid*.)

The string in (27) is infelicitous unless the base-generated position of *how* is higher than *who*, in which case it qualifies *ask* rather than *solve*, not the intended reading.

By the MLC, narrow constraints exist on the extent of possible search operations instigated by a probe, limiting the time complexity of the syntactic algorithm in accordance with MiD_1.

(iii) Chomsky (2000, 2001) introduces the notion of *phase*-based derivation. Phases are cyclic domains of derivation, each associated with a separate "subarray of the numeration." Structure is derived on a 'subarray-by-subarray' basis: when one subarray is exhausted, the next is accessed, and derivation continues. Each phase (or rather the complement of a *phase head*) is sent to the interfaces (*Spelt Out*) independently as soon as it is complete, yielding the *Phase Impenetrability Condition* (PIC). The PIC holds

therefore that only the previous *phase head* and its *edge* are visible to higher syntactic structure. Given the structure in (28):

(28) $[_{ZP}$ Z … $[_{HP}$ α [H YP]]]

where H and Z are the heads of phases, only α and H may be probed by operations at the ZP level.

Phase-based derivation insists that successive cyclic A'-movement proceed via phase edges, and empirical evidence in favour of this is provided by Legate (2003: 507).

(29) [Which of the papers that he_i gave $Mary_j]_k$ did every $student_i$ [vP t_k ask her_j [to read t_k carefully]]?

The *wh*-phrase in (29) must reconstruct to a position below *every student*, such that *he* may be bound by it, but above *her*, such that *Mary*, an R-expression, is not c-commanded by a co-referent pronoun. This clearly shows that A'-movement proceeds successively via the *v*P-phase edge *as well as* the specifier position of CP, a necessary corollary of a phase-based derivation.

Phase-based derivation clearly limits the extent of possible search operations, minimizing operational load in accordance with MiD_1. (Chomsky 2000: 108, Chomsky 2004)

4.2.2. Phonology

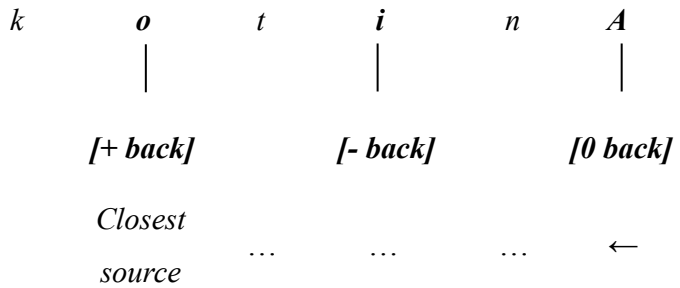
The minimal search considerations that obtain in phonological component are more or less congruent to those in syntax:

(i) Strict Cyclicity in phonology mirrors that in syntax. (Cyclic) phonological rules are associated with affixation and may only apply within that cycle, at the 'root' of the derivation. (Chomsky & Halle's (1968) discussion of how to distinguish the stress on *compensation* and *condensation* provides the classic evidence; cf. also Kenstowicz 1994: 205). Regardless of how cyclicity is accommodated by the phonological architecture – whether through interleaving (cyclic) phonology and morphology (Kiparsky 1982), or by affixal marking in a serial system (Halle & Vergnaud 1987, Marvin 2003) – the central

premise is the same. As in syntax, there is no search for a rule's locus of operation, diminishing time complexity under MiD_1.

(ii) Nevins (2005) demonstrates the parallel between the locality considerations bearing on featural assimilation and dissimilation in phonology and on *Agree* in syntax. He explicitly positions his observation as part of the Minimalist endeavour to determine what in language reflects "general cognitive processes" (*ibid*.: 16) and so might be taken out of UG. As introduced in § 4.2.2, featural underspecification may exist at the underlying, lexical level and then be resolved in the course of phonology to produce a fully specified SR. Such *non-perseverant* underspecification obtains in cases of feature harmony and disharmony, such as vowel harmony for the feature [back] of Finnish. Certain Finnish suffixes encode an unspecified [back] feature, which is valued by a fully specified [back] feature in the stem. In this case, the change is assimilatory, i.e. the unspecified feature adopts the same value as expressed by its source. Just as is imposed by the MLC in syntax, the defective feature's search for a valuing feature is limited to the *closest* possible source. For instance, when the underspecified essive suffix /nA/ [19] is added to the stem /koti/ ('house'), the [0 back] feature of the /A/ must be valued by the nearest specified [back] feature in the stem, relativised to contrastive segments. A contrastive segment for a given feature is one that has a counterpart in the language's inventory which is featurally identical except for an opposite value of the relevant feature. /i/ is not contrastive for [back] in Finnish, and so the first source reached by /A/'s search is /o/'s [+ back] feature, as shown in (30):

(30) *Search procedure*



[19] /A/ represents the feature bundle specified as [-high], but unspecified for [back].

(Nevins 2005: 36)

/A/'s [0 back] feature is thereby valued as [+ back] using the minimum possible number of search operations.

(iii) The cyclic *Spell-Out* of phases simplifies phonological computation in much the same way it does syntactic. Phonology can "forget about" (Chomsky 2004: 4) the derivational results of previous phases, and so phonological rules associated with affixes in a higher phase may not refer to any of this material. (cf. Chomsky 2008: 9) Assuming a syntactic model of morphology (Halle & Marantz 1993, Marantz 1997), this effect is manifest in the action of the stress assignment and vowel reduction rules of English (Marvin 2003). The details of the stress assignment rule need not concern us; it is sufficient for our purposes to recognise that it is associated with appropriately marked affixes (it is a 'cyclic' rule) and assigns stress to a *single* vowel. The vowel reduction rule takes place at the level of 'prosodic word', after cyclic rules have taken place, and 'reduces' unstressed vowels. The effects of phases in the phonological derivation are best seen with respect to an example, the word 'governmental', with the structure shown in (31) (movement rules, which we will not show, establish the correct linear order):

(31)[*aP* [*a* al] [n_1*P* [*n* ment] [*vP* [*v* Ø] [√ govern]]]]]

(Marvin 2003: 52)

Let us consider the stages of the derivation in turn. The stress assignment rule is triggered by the null *v* affix, assigning stress to the *o* of *govern*. As *v*P is a phase, *góvern* is then sent to *Spell-Out*. *-ment* is an unmarked (non-cyclic) affix and does not trigger a stress rule. *-al* is a cyclic affix and so triggers a stress rule. The stress previously assigned to the *o* of *govern* is inaccessible to *-al*'s stress rule, and so cannot be reassigned to a different vowel as it would if it were accessible. The affix *-ment* however is still accessible to *-al*'s stress rule, which thereby assigns its vowel a stress. As a result, when the level of 'prosodic word' is reached, two vowels are stressed: *góvernméntal*. The vowel reduction rule then reduces the unstressed vowels to schwas to result in the SR [g v rnm nt], stressed on the [ ] and the [ ]. Were multiple *Spell-Out* not implicated in the derivation, stress would not be preserved from previous cycles, and just a single vowel would be stressed and unreduced. (Marvin 2003: 51-54) As discussed, by restricting the search

possibilities of cyclic rules in this way, phases reduce the time complexity of the phonological algorithms.

4.2.3 Structural planning / parsing:

(i) Language users prefer to limit the range over which syntactic/semantic dependencies must be entertained in parsing. This can be seen in the usage of rightward extraposition:

(32) (a) [[That their time should not be wasted] [is important]]
↑----------↑
(b) [It [is important [that their time not be wasted]]]
↑----------↑

(32b) is overwhelmingly preferred in performance because the number of words that must be searched before the necessary syntactic/semantic relationship between *that* and its predicate is established in just three words, compared to eight for (32a). (Hawkins 2004: 107) As seen in § 4.1.3 and § 4.1.4, the more computationally marked string, (32a), may be used for pragmatic reasons, perhaps in this case emphasizing the *importance* of not wasting time over the importance of not *wasting time*. All other things being equal though, an utterance will be planned to minimise the number of search operations the parser must perform to establish syntactic/semantic connections within it.
(ii) See § 4.3.2 below.

4.3 Maximize Online-Processing$_1$

4.3.1 Syntax

MaOP$_1$ is reflected in syntax by way of the *Earliness Principle* (Pesetsky 1989), which holds that featural needs of a lexical item must be met at the earliest possible stage in the derivation, before another lexical item is introduced. Evidence demonstrating this effect is provided by Japanese *honorific constructions*. (Kiguchi 2006a) In such constructions, an active ditransitive verb cannot agree with the direct object (DO). In (33b), only the

indirect object (IO) is in a position to be honorified, but 'a new student' is an inappropriate candidate.

(33) (a) *Posudoku-ga Honsutein-sensei-ni shinnyuusei-wo go-syokai-sita*
Postdoc-Nom Hornstein-Prof -Dat new student-Acc HO-introduce-past
'The postdoc introduced Prof. Hornstein to a new student.'
(b) * *Posudoku-ga shinnyuusei-ni Honsutein-sensei-wo go-syokai-sita*
Postdoc-Nom new student-Dat Hornstein-Prof-Acc HO-introduce-past
'The postdoc introduced a new student to Prof. Hornstein.'
(Kiguchi 2006a: 128)

However (33b) is improved when the DO is raised via passivization as in (33c):

(33) (c) *?Honsutein-sensei-ga (posudoku-niyotte) shinnyuusei-ni go-syokai-sare ta.*
Hornstein-Prof -Nom (postdoc-by) new student-Dat HO-introduce-pass-past.
'Prof. Hornstein was introduced to a new student by the postdoc.'
(Kiguchi 2006a: 131)

Kiguchi (2006b) proposes that Japanese ditransitives are 'High Applicatives' consisting of a *HighAppl* phase head (McGinnis 2001) as well as a C and a *v*, giving (33a) and (33b) the structure in (34):

(34)$[_{TP}$ [Subj] $[_{T'}$ [T] $[_{vP}$ [*v*] $[_{ApplP}$ [IO] $[_{Appl'}$ [Appl] $[_{VP}$ [V] [DO]]]]]]

The IO intervenes between *v* and DO and so blocks honorific agreement between the two.

Kiguchi (2006a: 130) proposes that in the case of long passives, the involvement of [Case] features resolves this tension. Because DO still bears an unvalued [Case] feature upon completion of the HighApplP phase, it moves to the outer specifier of HighApplP under a phasal *Edge-Feature* (Chomsky 2008), giving the derivation of (33c) at the *v*P phase the structure in (35):

(35)… [$_{vP}$ [*v*] [ApplP [DO$_1$ [$_{Appl'}$ [IO] [$_{Appl'}$ [Appl] [$_{VP}$ [V] [DO]]]]]

DO in this intermediate position can now form an honorific *Agree* relation with *v*, explaining the improved felicity of (33c) relative to (33b). This analysis suggests that the featural requirements of *v* must be exhausted as soon as possible, before T is introduced into the phrase marker and attracts DO to its specifier. For it is only in DO's intermediate position that it is c-commanded by *v*. Once it is in the specifier of TP, it is too late for honorific agreement between the two to obtain.

As this example demonstrates, the syntactic algorithm is designed not to store unvalued features in working memory to be valued at the later stage in the derivation. This is in accordance with the demands of MaOP$_1$.

4.3.2 Structural planning / parsing

Following Chomsky's (2007: 7) suggestion of a parsing-based account for the linearisation of symmetric syntactic derivations, it seems that MaOP$_1$ offers a principled reason for the widely assumed 'SPEC-initial' [20] principle. SPEC is the target position for (phrasal) *Internal Merge*, i.e. the position in which moved phrases are merged. While a few SPEC positions may be projected under *External Merge* (e.g. subjects, and perhaps also adverbs and indirect objects under the cartographic approach), every case of *Internal Merge* uncontroversially part of syntax (the status of head movement is somewhat vexed, cf. Chomsky 2001: 37-8, Roberts 2008a) projects a SPEC position. This of course results in a *filler-gap* relation for the parser. Invoking the standardly assumed left-to-right parsing procedure, fillers should precede gaps under MaOP$_1$, such that the syntactic and semantic features of gaps may be assigned straight away. Should a gap come first, various features would have to be held over in working memory until the appropriate filler was reached and could value them – a clear increase in cache load. Therefore, since SPEC is so frequently occupied by a filler antecedent to a gap, SPEC-initial becomes computationally optimal in the context of a parsing-based account of linearisation.

[20] We take upper-case lettering to refer to the abstract, relationally-defined *position* in the syntactic structure, rather than an element occupying that position. SPEC refers to the specifier position, COMP to the complement position, and H to the head position.

This argument of course relies on the idea that movement should be (at least predominantly) leftward. This point has been argued by Ackema & Neeleman (2002). They assume a *filler-driven* parsing procedure in accordance with experimental evidence presented by *i.a.* Frazier & Flores D'Arcais (1989). This means that a gap may only be postulated after an antecedent/filler has been identified. The asymmetry between leftward and rightward movement emerges from this approach when it is considered in tandem with the periodic "closing off" of already-parsed structure at appropriate points, as argued *i.a.* by Just & Carpenter (1992) and Gibson (1998). Rightward movement presents a problem to a filler-driven, interim-storing parsing strategy, since left-to-right parsing must then be suspended to insert a trace into a partially analysed string. If this string has already been "closed off" to the parser, such insertion should be impossible, or at least difficult. [21]

While the observation that fillers should precede gaps is hardly new, it has generally been discussed in terms of the increased likelihood of misassignments by a 'surface word order'-oriented parser (Fodor 1978), rather than considered in terms of working memory cost in the context of parsing-based linearisation of underlying syntactic structures. Hawkins (2004: 204) associates his MaOP with the former, and actively distances it from any working memory concerns (*ibid.*: 232-5). His functionalist commitments of course also distance him from the whole concept of abstract syntactic derivation (involving movement operations) and lead him to suggest that any theory of linearisation that relies on it is wrong-headed. Whereas under Hawkins' view MaOP is mute on the issue of linearisation (*ibid.*: 252), if we entertain a fully-articulated formal model and a computational view of efficiency, as I have argued we must, interesting and far-reaching explanatory uses arise for his own principles.

Hawkins (*ibid.*: 252) does, however, observe that consistent H-COMP adjacency would be compatible with his MiD. This observation can be transferred into a computational efficiency-oriented account: establishing the syntactic relationship between head and complement in parsing involves one instigating a search operation for the other (it makes

[21] Ackema & Neeleman (2002) use this reasoning to explain the clause-internal locality restriction on rightward movement (the so-called *Right Roof Constraint* (Ross 1967)). Assuming that distinct clauses constitute distinct parsing domains, rightward movement beyond the clause to which the moved element belongs requires multiple "reopenings" of the type required in monoclausal rightward movement structures (even if movement is successive-cyclic, as is generally assumed of long-distance leftward movement.)

no odds for our concerns whether head does so for its complement, or complement for its head), and so if they are restricted to adjacency then the number of search operations required is kept to a constant minimum under MiD_1. Strangely though, Hawkins fails to observe that the same argument applies equally to SPEC-H adjacency, suggesting (*ibid.*: 253) that a [SPEC [COMP-H]] order is as well supported as [SPEC [H-COMP]] under $MiD_{(1)}$. This neglects the importance of the syntactic and semantic relationship between a specifier and its head: for instance, the theta-role/theta-assigner relationship, or, under Rizzi/Cinque-style assumptions, the head carrying information about the status of the specifier (cf. Chomsky 2007: 7). Given this, the constant minimisation of search operations in parsing computation renders SPEC-H adjacency optimal under MiD_1. The synthesis of these two arguments is that [SPEC [H-COMP]] and [[COMP-H] SPEC] may be considered optimal, and the four other possible orders sub-optimal.

We may now re-introduce the argument from the beginning of this section, namely that the general cognitive principle of $MaOP_1$ mandates that SPEC precede its sister. Therefore, of the two orders optimal under MiD_1, only one is optimal under $MaOP_1$: [SPEC [H-COMP]]. We have therefore arrived at a principled account for a particular linear order being imposed on a syntactic derivation. Essentially, this result constitutes SPEC-initial and COMP-final principle emerging from third factors interacting in computation of linearisation.

It should be noted that this approach allows us to retain what has proved useful in the *Antisymmetry*-based account to linear order (Kayne 1994), while allowing us to avoid its problems. Being situated beyond *Spell-Out*, it still allows us to entertain Bare Phrase Structure (BPS) and avoid the problems associated with Antisymmetry as a theory of phrase structure, including: the unminimal double marking of linear order in both Narrow Syntax and at the interface; the projection of redundant structure; the proliferation of functional architecture owing to the absence of multiple SPEC positions; the reintroduction of the category-segment distinction etc.. All these problems may be avoided by reframing the *Linear Correspondence Axiom* (LCA) as a post-syntactic mapping strategy on a BPS derivation (Chomsky 1995b: 340); but two problems that persist are the incomplete ordering of the base pair (i.e. when two heads are merged) owing to mutual *c-command* (Nunes 1999, Moro 2000, Richards 2004a, b) and the fact that the LCA cannot be shown to emerge from concerns for computational optimality, and

so constitutes a breach of the SMT (Chomsky 2007: 7). Both these problems are resolved under the view that linear ordering is imposed on abstract syntactic structure by a computationally optimal parser, resulting in SPEC-initial and COMP-final principles.

Not only does our approach offer a principled way of avoiding the problems associated with the LCA, either in the syntax or beyond, it retains all the empirical coverage that has made it attractive. This is not the place to present this evidence at length, but it is worth reiterating the point made in Richards (2004b: 186) (regarding the LCA) that "the 'Spec-Head-Comp' [22] hypothesis should […] be broken down into […] two separate claims, each of which needs to be independently justified by abstracting away from the other." To a certain extent, arguments for an LCA-based approach to linearisation have taken evidence in favour of SPEC being initial [23] (a fact at any rate largely unchallenged in the literature) as evidence in favour of an asymmetric *c-command*-based approach, clearly extrapolating beyond the data. To justify the SPEC-HEAD-COMP hypothesis properly, we then need independent evidence to support the COMP-final principle. One example of synchronic empirical evidence that bears properly on this component of the claim is Hinterhölzl's (2000) demonstration that internal arguments preceding the infinitival marker in West Germanic languages, as in (36), have moved out of the VP, and so cannot be in their base-generated position ordered in a COMP-initial fashion.

(36) *(...weil) er sie wagte zu küssen* [German]
(…because) he her dared to kiss
(…because) he dared to kiss her

Non-finite *Infinitivus Pro Participio* (IPP) constructions from West Flemish and Afrikaans provide evidence that the infinitival marker is outside VP, not affixed to V.

(37) *mee Valère <u>te</u> [willen [dienen boek kuopen]] <u>een</u>* [West Flemish]
with Valere to want that book buy have
'with Valere having wanted to buy that book'
(*ibid*.: 297)

[22] Translating this into our nomenclature, this is the 'SPEC-HEAD-COMP hypothesis'.

[23] For instance, the absence of reverse V2 word order in the world's languages. (Kayne 2003: 2)

Since the material in square brackets can intervene between the infinitival marker and the corresponding infinitival verb in these constructions, the infinitival marker in West Germanic languages cannot be taken as a verbal affix, but rather must be occupying a functional position in the TP domain. As such, arguments preceding the infinitival marker like in (36) must also be outside their base-generated VP domain.

LCA-based approaches employ a rigidly H-initial underlying word order, and so necessarily derive head-final word orders by movement out of COMP. Biberauer & Roberts (2008) further demonstrate the power of this idea, introducing as it does an added dimension of variation in the amount of material piedpiped (Biberauer & Richards 2006) – using it to explain the synchronic word patterns of Old English and Middle English and give a finely-grained analysis of their diachrony. Similarly, an LCA-based approach and its reliance on the distribution of movement diacritics in deriving word order is fundamental to Biberauer, Holmberg & Roberts' (2008) explanation of the *Final-over-Final-Constraint* on word order (and word order changes) across the world's languages, which centres on a principle generalising movement diacritics across domains (see § 5.1 below for a fuller summary).
SPEC-initial and COMP-final principles imposed on symmetric syntactic structure by a computationally optimal parser offer precisely the same empirical advantages as the LCA, arguing for the same rigid, underlying word order, but do so without the associated theoretical difficulties and in proper accordance with the SMT. This seems to me a significant result.

4.3.3 Pragmatic competence

Experimental pragmatics has stressed the 'automaticity' of enriching the informational content of a sentence. The semantic, or *plain*, meaning of a word seems to be enriched as soon as possible in the recursive computation of a string's interpretation, i.e. in a *local* fashion. The *global* (neo-Gricean) view (cf. Horn 1989, Levinson 2000) is that the truth conditions of a string must be retained until the end of string presentation before enrichment, clearly posing a greater burden on working memory. The *local* view has received empirical support from Chierchia's (2004) observations regarding implicatures in sentences with adjoined clauses. Consider (38), and its implication, (39):

(38) Mary is either seeing some students or reading a paper.

(39) Mary is either seeing some students or reading a paper, but she is not seeing every student.

(Chierchia *et al.* 2004: 286)

The intended meaning of (38), i.e (39), seems to be computed with respect to the first disjunct only. This is incompatible with the global view, under which enrichment is held over until the completion of the retained, unenriched truth conditions of the whole sentence. Such a view would lead us to expect the intended reading of (38) is:

(40) Mary is either seeing some students or reading a paper and it is not the case that (Mary is seeing every student or reading a paper).

However, from (40) we get (41), clearly not what is stated:

(41)Mary is not reading a paper.

(*ibid*.)

This example seems to demonstrate the immediate enrichment of logical form, rather than its retention until a global enrichment process, consistent with the force of $MaOP_1$ in minimizing cache load.

5. Generalise Features

I restate the definition of MiF_1 from (13) here for convenience:

(13) **Minimize Forms$_1$ (MiF$_1$)**

Human computation minimizes the formal complexity of each form F and the number of forms with unique property assignments, thereby assigning more properties to fewer forms.

Taken in isolation, an element of this definition is ambiguous. There are of course two ways of "[m]inimiz[ing] … the number of forms with unique property assignments." One is the way pursued by MiF_1 – simply rendering "fewer forms" through increasing feature syncretism. The other is to minimize the unique property assignments among the forms rendered – that is, to have featural content of some forms determined by featural content of other forms, rather than attributed independently.

Would there be any computational advantage associated with the latter of these two? When we stated on pp. 15 that "[t]here will an operation associated with processing each feature, so reducing the number of features reduces the number of operations that have to be performed", we were simplifying slightly. There will be an operation associated with processing a feature, unless the algorithm, by design, takes the result of previous operation to apply instead. Designing the algorithm to determine featural content of some forms on the basis of that possessed by others therefore reduces the number of operations associated with 'changing' featural content between one form and the next.
An alternative interpretation of an element of MiF_1's definition has therefore pointed us toward a further potential efficiency-oriented third factor, which I state in (42):

(42) **Generalise Features (GenF)**
Human computation generalises features over forms in the same domain

We now examine if and how GenF might be implicated in the design of the FL's algorithms.

5.1 Phonology

The notion of *non-perseverant* underspecification in phonology was introduced above in § 4.1.2 and § 4.2.2. It is extensively empirically motivated by data from alternations in affix segments (such as the Finnish essive suffix seen above), demonstrating their

dependence on the value of a feature on a root segment to determine their value for that feature. For instance, data from Kirghiz shows the ordinal suffix always adopting the [round] feature value of the root (Comrie 1981: 62):

(43) Alternations in the Kirghiz ordinal suffix /-InčI/: [24]

Numeral	Ordinal	Gloss
bir	*birinči*	one
beš	*bešinči*	five
altı	*altınčı*	six
žıyırmı	*žıyırmınčı*	twenty
üč	*üčünčü*	three
tört	*törtünčü*	four
toguz	*toguzunču*	nine
on	*onunču*	ten

The 'inheritance' of feature values by underspecified segments from a fully specified segment, thereby generating a complete phonetic representation, clearly minimizes the operations that must be performed in differentiating featural content between segments in a phonological domain, in direct accordance with the force of GenF.

Non-perseverant underspecification represents a particularly interesting phenomenon in the light of our concerns, as it seems to demonstrate two different efficiency-oriented third factors (MiD_1 and GenF) acting in conjunction to perform a linguistic task, as anticipated by the SMT.

5.3 'Data-analytic' / acquisitional faculties

(i) The force of GenF in the acquisitional modules is evidenced by analogical processes in the diachrony of morphology, specifically analogical extension across paradigms, or *proportional* analogy. A classic example is the generalisation of the <-s> plural affix in

[24] /I/ represents a segment specified only for [+high].

English. Old English had various different declension classes for nouns, each associated with a different means of forming the plural, for instance: *sunne*/*sunnan* 'sun/s'; *scip*/*scipu* 'ship/s'; *brother*/*brethren* 'brother/s'. One paradigm developed <-s> as a plural marker and this was generalised to the other paradigms, which had not used it previously. A plural marker historically only appropriate to one paradigm became characteristic of almost all forms. The over-generalisation in acquisition that is implicated in this type of historical process is particularly prevalent in child language. Children frequently produce regularized forms such as *foot ~ foots* or *bring ~ brang ~ brung*. (McMahon 1994: 71-2)

Kiparsky (1965) again associates over-generalisation in acquisition with a drive toward "simplification of the grammar" in terms of a loss of "exception features" marking a morpheme as inflecting in an irregular fashion. As discussed as length in § 4.1.5, we take this view of markedness to be theoretically unsustainable because it takes efficiency concerns to be 'forward-looking', referring to the product of computation rather than process itself – counting features, rather than time or space complexity. [25]

Analogical tendencies fall naturally out of an approach that takes GenF to be implicated in the acquisitional faculties. In processing primary linguistic data, the computation will prefer to characterise the morphological feature of a presented morpheme by reusing the result of an operation performed to classify previously presented morphemes, resulting in the generalisation of 'dominant' (perhaps simply most common) morphological features across the lexicon. This 'emergence of the unmarked' would be expected to be particularly prevalent in the face of limited evidence to defy it (see pp. 22 above). Such is the case early on in acquisition, resulting in the anticipated over-regularisation of child language. The diachronic and synchronic data are both consistent with the operation of GenF in the computational faculties responsible for the acquisition of morphology.

[25] It is far from clear that 'exception features' necessarily have any psychological status. Instead we might model a morphological class as a list under a common 'heading', which provides the inflectional information. In this case, moving a morpheme from one list to another under analogical extension would have no net effect on the simplicity of grammar until a list is completely empty and the heading can be removed. A single case of analogical extension could not then be taken to reflect "simplification of the grammar" (unless it removed a previously unique case). A seemingly unambiguous case of over-generalisation in the acquisition of syntax outright *complicating* the grammar is provided in § 6, the empirical reflex of the theoretical difficulties associated with the feature-counting approach.

(ii) We will deal with the reflex of GenF in the acquisition of syntax in the following section.

6. Typology

I now wish to comment briefly on certain aspects of how best to understand language typology under the approach pursued above.

As we have seen, Hawkins' appraisal of structure and acquisition largely eschews abstract grammatical explanation in favour of a performance-based account. Surface word order is therefore central to his account, and his theory of typology is necessarily more or less coextensive with his theory of grammar. In § 1, I suggested that his hypothesis of 'performance-grammar correspondence' was unsustainable on various grounds, restated here:

(6) (a) inconsistence with the empirically-vindicated competence-performance distinction;
(b) failure to diagnose a real problem with a UG-based approach to learnability concerns, or to offer a viable alternative;
(c) failure to justify the choice of the PGCH over the GPCH, particularly pressing in light of some evidence favouring the latter.

Each of these problems bears equally on Hawkins' theory of typology and his theory of grammar. However, a further problem specific to the former obtains. Hawkins fails to distinguish properly between *synchronic universals* of language and the *contingent results of historical processes*, taking the whole explanatory burden to be carried by the latter. As Kiparsky (2006: 4) points out, this is a distinction that must be accommodated regardless of commitments as to what motivates grammar. While Hawkins and other *panchronic* theorists (cf. Blevins 2004 and her *Evolutionary Phonology*) are right to recognise the role of language change in explaining typological distribution, it is a logical requirement that absolute constraints do exist on a possible language, and these must therefore be properly spelled out by a theory of typology.

Instead, I intend to pursue the properly nuanced, formal approach to language typology Kiparsky (2008) advocates; this recognises and avoids the difficulties with a panchronic, functionalist course. The empirical reflex of this theoretical necessity will be evidenced in the power of the approach, particularly in capturing asymmetrical attestation among disharmonic word orders.

6.1 Synchronic explanation

The innate grammatical knowledge necessary to address poverty-of-the-stimulus concerns places inviolable constraints on what constitutes a possible human language. While a given language may not necessarily exhaust the possibilities presented by grammar (Fitch, Hauser & Chomsky 2005: 202-3), it cannot connect form and meaning by means that violate them. The state of the human mind does not permit language acquirers to postulate grammars that "differ from each other without limit and in unpredictable ways" (Joos 1957: 96), [26] since this would render the acquisition problem insoluble.

Some of the inviolable grammatical machinery has been introduced above. Examples of grammars debarred by this machinery would include: a grammar allowing a syntactic/phonological 'probe' to be valued by a 'goal' beyond the nearest possible one (cf. § 4.2.1 and § 4.2.2); a grammar allowing lexical items to merged in positions other than the root of the derivation (cf. § 4.2.1), etc..

Variation between grammars has long been thought to be encoded in innate competence through the binary *parameterisation* of certain grammatical principles. (Chomsky 1981) Such variation is nonetheless restricted by the nature of the grammatical principle with which it is associated (and other principles of the system). Underspecification merely serves to *extend possible language space* beyond that which would be associated with exclusively complete principles.

[26] It will equally, of course, constrain the *use* of a given language. In fact, in as much as grammatical competence and performance modules comprise non-dedicated principles of computation, the same factors will constrain both possible grammar and possible usage.

As a result of parameterisation, we may see regular *clustering* of superficial grammatical properties on languages in the typological record. For instance, it has suggested that the presence of a [person]-feature on the Agr-functional head in I can be related to the clustering of: null non-referential subjects, non-nominative subjects, stylistic fronting, and relatively rich subject verb agreement, as in Insular Scandinavian, Old French, Middle English and Yiddish. The absence of this feature is suggested to result in the clustered absence of these properties, as in Mainland Scandinavian and Modern English. (Holmberg & Platzack 2005, Roberts & Holmberg 2005) The grammatical variation associated with this parametric setting is narrowly constrained by the nature of *Agree* relations, the hierarchy of functional categories, and many other grammatical principles. [27] A full discussion of the role of parameters in explaining clustering in the typological record is beyond our remit here (see instead Biberauer (forthcoming) and papers in the volume it introduces); it is enough to note that typological explanations in terms of such synchronic universals are still statements about what constitutes a possible language.

Coming back to unparameterised synchronic universals, we turn our attention to one proposed by Biberauer, Holmberg & Roberts (2008: 101), who suggest that "[i]f a phase-head has an EPP-feature, then all the heads in its complement domain from which it is non-distinct in categorial features have an EPP-feature." They employ this synchronic universal to explain the seeming absence of a particular disharmonic word order from the world's languages – namely that word order in which a superficially head-final phrase dominates a superficially head-initial phrase of the same category (i.e. nominal or verbal.) A head-final phrase is constrained to dominating another head-final phrase, and cannot dominate a head-initial phrase, in accordance with a descriptive generalisation noted by Holmberg (2000) – the *Final-over-Final Constraint* (FOFC). However, under the usually assumed view of syntactic apparatus, the derivation of FOFC-violating structures should be entirely possible.

There are thought to be two types of syntactic head, *phase* heads and *non-phase* heads (see § 4.2.2). When a non-phase head dominates a phase head, a FOFC-violating structure is prevented by the *Phase Impenetrability Condition* (PIC). The appropriate derivation would require the raising of the phase head and its complement to the specifier of the non-phase head, as in (44):

[27] Including other parametrised principles, resulting in interaction possibly masking clustering effects.

(44) $[_{\text{Non-PH Phrase}}\,[_{\text{PH Phrase}}\,\text{PH}\,[\text{Complement}]]_1\,[_{\text{Non-PH'}}\,\text{Non-PH}\,[\,t_1\,]]]$

However, by the PIC, the complement of PH is inaccessible to Non-PH since it is has already been sent to *Spell-Out*. Therefore the surface word order generated will be PH NPH Comp_{PH}, not in fact a FOFC-violation.

When a phase head dominates a non-phase head however, there is no such bar on the raising of the lower phrase's complement, as it has not been sent to *Spell-Out*:

(45) $[_{\text{PH Phrase}}\,[_{\text{Non-PH Phrase}}\,\text{NPH}\,[\text{Complement}]]_1\,[_{\text{PH'}}\,\text{PH}\,[\,t_1\,]]]$

The derivation in (45) should therefore result in a word order of NPH Comp_{NPH} PH, a superficially head-final phrase dominating a head-initial one, in breach of FOFC.

However, the empirical record demonstrates the striking absence of this word order even when a phase head does dominate a non-phase head. To take a single example (Biberauer, Holmberg & Roberts present many), adpositional phrases in Finnish display each of the four possible orderings of P, N and O (O being the nominal complement) except the one which violates FOFC, (46d). This is despite the fact that P is thought to be a phase head, and N a non-phase head of the same category:

(46) (a) P N O:
kohti kuvaa Stalinista
towards picture Stalin-ABL
"towards a picture of Stalin"

(b) P O N:
kohti Stalinin kuvaa
towards Stalin-GEN picture
"towards a picture of Stalin"

(c) O N P:
Stalinin kuvaa kohti
Stalin-GEN picture towards
"towards a picture of Stalin"

(d) N O P:

**kuvaa Stalinista kohti*
picture Stalin-ABL towards

(Biberauer *et al.* 2008: 91)

Biberauer *et al.*'s proposal though leads us to expect just such gaps. To derive N O P surface word order, P must possess an EPP-feature to move the NP, while N must not have an EPP-feature, in order to retain N O order. Yet being a categorially non-distinct non-phase head dominated by a phase head possessing an EPP-feature, N should have inherited an EPP-feature, rendering the derivation impossible.

As Biberauer *et al.* note, their suggestion is in the spirit of Chomsky's (2008) empirically-supported proposal regards the inheritance of *phi*- and tense-features by non-phase heads from the head of their phase – cf. the putative 'C-T relationship' (which can clearly be generalised to other phase head-non-phase head pairs.) As such, it merely extends the action of previously assumed syntactic machinery from substantive features to non-substantive ones. While the CI-interface motivation Chomsky proposes for this inheritance relationship applies only to substantive features, it would be expected under Minimalist assumptions that (presumably subsequently evolved) linearisation features would utilise the same pathways. [28] Biberauer, Holmberg & Roberts' proposal therefore constitutes a 'cost-free' explanation of the empirical gap in disharmonic word orders, and demonstrates the real force of formal, synchronic universals in typological explanation. [29]

[28] It is important to note that this is not the same as the claim that EPP-features are dependent on *Agree* features in syntax. Indeed, Biberauer *et al.* (2008: 98) specifically note that EPP-features in their approach are "purely linearisation instructions" and do not need to "piggyback" (*ibid*.) on *Agree* relations.

[29] It might be taken that the inheritance of linearisation features constitutes an implication of GenF in the syntactic computation, since inherited movement diacritics lower down in the phase need not be stated by independent operations as they would if there were no generalisation. However, the fact that the phase head only seems to generalise the *presence* of an EPP-feature, not the *absence*, suggests otherwise. When it does not possess a movement diacritic, extra operations may still be entertained in the postulation of independent EPP-features in its domain; there is no *Initial-over-Initial Constraint* (IOIC), as we would expect if the inheritance relationship were truly motivated by GenF. The compulsory inheritance of certain features of a phase head by a non-phase head, but the possibility of extra features not encoded on the phase head being added is in entirely keeping with CI-motivated machinery, but much less so with GenF motivation.

6.2 Diachronic explanation

If synchronic forces define a possible language space, then diachronic forces determine the distribution of languages within this space. Together they conspire to produce the patterns seen in the typological record. [30]

As I see it, there are three logically possible ways of effecting typological skew by diachronic / acquisitional means.

6.2.1 The nature of the acquisitional modules

-

As we have seen above, the abstractional faculties are 'biased' toward interpreting PLD in whatever way imposes the smallest computational burden on them. This predisposes language change to following certain pathways, creating "basins of attraction" (Roberts & Roussou 2003: 235) within possible language space.

An example of this was seen in § 4.1.5. The over-representation of featurally-economic phoneme inventories was related to "the frequent historical elimination of 'isolated' sounds that do not fall into regular patterns of correlation with other sounds" (Clements (forthcoming): 21-2), such that "the feature that previously characterized them becomes redundant." (*ibid*.: 22) This pathway was in turn related to the force of MiF_1 in the acquisitional modules, minimizing the operations associated with features postulated to capture distinctions.

A second example is afforded by the classic typological observation: harmonic word order systems, those demonstrating consistence of head-complement order across the categories, are disproportionately represented relative to disharmonic systems (Greenberg 1963). A case in point is the correlation between VO/OV and NGen/GenN ordering:

[30] Along with areal and other *arbitrary* factors.

(47) Correlation between OV/VO and NGen/GenN in the 933 languages showing dominant order for both surveyed in Haspelmath *et al.* (2005):

OV & GenN	434 (46.5%)
VO & NGen	352 (37.7%)
OV & NGen	113 (12.1%)
VO & GenN	34 (3.7%)

84.2% of the languages surveyed are harmonic in respect of this correlation, and just 16.8% disharmonic.

The first attempt to explain the skew toward harmony centred on a synchronic universal of grammar. Vennemann's (1974) *Natural Serialisation Principle* (NSP) requires operators and operands to be serialized in a consistent order. Since objects and genitive qualifiers are operators, and verbs and nouns are operands, this predicts correlations such as the one above. Vennemann (cf. also Lehmann 1973) then proposes that word order conformity is a *causal* factor in language change: once a change introduces an initial violation, the NSP will re-introduce consistency by demanding the subsequent acquisition of word orders harmonic with this change. This dictates that languages 'drift' uniformly toward one of the two harmonic systems – head-final or head-initial.

There is a glaring theoretical problem with this approach however, first pointed out in Hawkins (1979), but summarised here by Song (2001: 304):

> *[t]ypological consistency must at the same time be considered to be strong and weak. […] It must be weak enough to permit incongruous word order properties to be incorporated into typologically consistent languages in the first place, and it must also be strong enough to remedy the resulting situation by bringing all remaining word order properties into line with the new ones.*

Hawkins (1983: 41) likewise criticizes the empirical predictions of the NSP on the grounds of the prevalence of disharmonic systems, observing that up to 77% of Greenberg's initial 30-language survey does not conform to its synchronic predictions.

Hawkins (1979, 1983) proposes an alternative diachronic explanation for the over-representation of harmonic word orders. He instead suggests a principle of acquisition, *Cross-Categorial Harmony* (CCH), motivated by language user-acquirers preference for simpler, more usable grammars:

(48) *There is a quantifiable preference for the ratio of preposed to postposed operators within one phrasal category [...] to generalize to others.*
(Hawkins 1983: 134)

By taking a principle of acquisition rather than a synchronic universal to motivate the skew, Hawkins avoids the major theoretical difficulty associated with Vennemann and Lehmann's approach, and also accommodates the empirical fact that there are many disharmonic languages. If CCH is simply a principle of language acquisition/change, it is necessarily in competition with other forces that might serve to disrupt harmony, such as ambiguity of the input provoking disharmonic reanalyses.

However, Hawkins' panchronic, functionalist account here suffers from all the same problems as his more recent approach. Furthermore, there is no psycholinguistic evidence whatsoever to suggest that disharmonic systems are any more problematic in usage than harmonic ones. Such an approach also fails to accommodate the persistence of disharmonic word orders across time. Old Icelandic, for instance, maintained a consistent mixed system (Adposition-Noun, Object-Verb, Complementiser-initial, Adjective-Noun etc.) for all of half a millennium, between the 13th and 18th centuries (Hróarsdóttir 2000). Such persistence would be unexpected if disharmonic systems really were functionally deficient, as language users tend to innovate to bridge gaps of this kind, as evidenced by prevalence of neologism.

Roberts (2007: 273-5) reinterprets CCH as a principle of acquisition within a formal framework: a "markedness convention in terms of the conservatism of the learner", which holds that "if acquirers assign a marked value to H they will assign the same value to all comparable heads." The 'marked value of H' here is assessed relative to Roberts & Roussou's (2003) feature-minimizing metric in (17) (or in our terms, its MiF_1-motivated reinterpretation). Accordingly, if a language acquirer postulates an EPP-feature on a head, marked with respect to feature-minimization/MiF_1, he/she will then prefer to postulate an

EPP-feature across the heads of the other categories. The effect of this generalisation would be to correlate OV word order with other head-final word orders by a diachronic pathway. This is despite the force of the feature-minimization metric/MiF_1, which tends to eliminate EPP-features and promote diachronic drift toward head-initial harmony (assuming a SPEC-HEAD-COMP underlying word order). The claim here then is not that certain grammatical systems are in any way disprefered, merely that the nature of acquisition conspires to make them less likely to be postulated; language space itself is *homogenous*. This avoids all the problems associated with Hawkins' approach.

However, such a markedness convention is inconsistent with an approach that associates conservative acquisition with the output of computation, rather than the process itself. Drift toward head-finality involves the postulation of extra movement features and extra SPEC-positions in the phrase marker; the postulated grammar is in fact formally more complex, rather than simpler. Roberts (2007: 234) thus acknowledges that "a simple feature-counting approach to markedness no longer suffices". While he still maintains that the learner's tendency "to assign the simplest representation or derivation possible to the PLD it is exposed to" (*ibid*.) motivates the markedness metric in (17), he suggests that this second notion is motivated by "a different aspect of the overall conservatism of the learner" – a drive toward "set[ting] parameters in the most efficient way possible" (*ibid*.: 275).

My suggestion here is that both notions of markedness derive from this latter motivation, and that a feature-counting approach is both theoretically incoherent and redundant. We have already argued that Roberts (& Roussou)'s first notion of markedness falls directly out of the operation of MiF_1 in the acquisitional computation. The "generalisation of the input" (Roberts 2007: 275) expressed by the second is simply the reflex of GenF. [31] If an operation assigns an EPP-feature to a given head, the computation will prefer to reuse the result of this operation in characterizing future heads, rather than perform an operation to 'cancel it out'. In this way, we arrive at a principled account of why harmonic systems are favoured by diachrony. When taken together with FOFC, which removes certain disharmonic systems from possible language space, we see why the typological record reflects a skew toward harmony.

[31] Roberts (2007: 299-300) draws the analogy between his second notion of markedness and *lexical diffusion*, a phenomenon closely related to analogical extension, which we have also taken as motivated by GenF in data processing (see § 5.2).

While this approach eschews any difficulty with the persistence of disharmonic systems in the face of functional deficiency, we have still to understand how such systems might resist the force of GenF in acquisition (as seen with Old Icelandic). This is naturally accounted for in a formal approach; the word order of a language is stabilized by at least two factors. First, the mechanistic demands of reanalysis must be met: in the absence of ambiguity in the input, reanalysis may not occur and so economic computation not be expressed (cf. the *Inertia Principle* of Keenan 2002 and Longobardi 2001). And secondly, FOFC (or rather the synchronic universals behind it) constrains the structural positions at which a change in linearisation is possible at any given time.

Under FOFC, change from head-final to head-initial word order (within a given categorial domain) must proceed 'top-down', and change in the opposite direction must be 'bottom-up', as shown in (49) for the verbal domain:

(49) (a) [[[O V] T] C] > [C [[O V] T]] > [C [T [O V]]] > [C [T [V O]]]
(b) [C [T [V O]]] > [C [T [O V]]] > [C [[O V] T]] > [[[O V] T] C]
(Biberauer, Holmberg & Roberts 2008: 92)

In change from head-final to head-initial, if change were to start at VP (i.e. with the loss of the EPP-feature from *v*), the resulting word order would have a head-final CP/TP dominating a head-initial VP, violating FOFC. If TP were the first to change, we would have a period during which a head-final CP dominates a head-initial TP, again a FOFC violation. The only FOFC-respecting possibility is for CP to change first, such that a head-initial CP dominates a head-final TP. Thereupon, TP may change, and only then is VP free to change. The change from head-initial to head-final word order requires the opposite: if TP changes before VP, a head-final TP will dominate a head-initial VP, in breach of FOFC; and so on. (*ibid*.)

At any given stage therefore, change in word order within each categorial domain is only possible at one position, stabilizing languages against change. Together with the mechanistic demands of reanalysis, this explains the resistance of disharmonic systems to levelling.

Finally, while we have addressed the attraction of the harmonic poles, we have not yet understood the approximately even number of head-final-type and head-initial-type languages attested. The effect of MiF_1 in acquisition is to remove EPP-features, and so we would expect the balance to lie toward the head-initial pole. An opposing force is indicated, promoting the introduction of movement features into systems. This force is provided by demand for dual semantics – the CI-condition taken to motivate the possibility of movement in the first place (Chomsky 2000). Rephrasing Kiparsky's (1996) speculations regarding the origin of OV syntax, optional movement might be postulated in syntax to mark a discourse or scopal effect (cf. Chomsky 2008 on the role of *edge-features*) and become obligatory through the subsequent conventionalisation and loss of this effect. (cf. Simpson 2004 for discussion of a case study) [32]

We have now arrived at a position where we can offer a measure of principled explanation for the overall distribution of word order systems within possible language space, addressing both their clustering toward the harmonic poles and the approximate balance of each type. The course pursued eschews any of the difficulties with earlier approaches to the asymmetric attestation of harmonic and disharmonic languages and, uniquely, is able to accommodate the persistence of disharmonic word orders across time – another (anticipated) empirical advantage of a formal approach.

The phonological and syntactic examples presented above demonstrate how the nature of acquisitional computation predisposes language change to following certain pathways, forming "basins of attraction" within possible language space.

6.2 'Ease of development' / Natural pathways

The second logical means by which diachrony may effect typological skew is through the *direct* relationship between linguistic features (or combinations of linguistic features) as

[32] The syncretism of the edge-feature onto the *Agree*-feature of the verb, as an EPP "feature-of-feature" (Pesetsky & Torrego 2001) would be promoted in such cases by the force of MiF_1 in the acquisitional modules. Given the loss of semantic evidence for the edge-feature that would be effected by conventionalisation, acquisitional computation will prefer to characterise the data using just the *Agree*-feature.

reanalysable sources for one another. The circumstances necessary for the development of a particular feature may be less probable than those for others.

The scarcity of infinitival subjects in the world's languages can be related to the fact that (at least one pathway of) their development relies on the existence of an infinitival marker (such as *to*) and the use of active infinitives in a 'passive sense' (Lightfoot 1991: 15), shown with structure in (50):

(50) He let$_i$ $_{S'}$[COMP $_S$[PRO $_{VP}$[turf tredan]]] t_i [Old English]
'He ordered turf to be trodden'

(Making GB assumptions in this case, PRO is an arbitrary subject, so may be unlicensed following Hornstein & Lightfoot 1987.)

The use of active infinitives in this way must then be precluded by an independent development such that they become fully blown passive constructions. The change from OV to VO word order creates ambiguity as to the status of *turf*; since *turf* is now in a subject position, but cannot logically *tread*, it must be interpreted as a raised object. The presentation of strings with a *to*-marker ('he let turf to tred' etc.) allows *to* to be interpreted as the T-head necessary to license an external argument, so the equivalent production of (50) in the new grammar is (51):

(51) He let$_i$ $_{S'}$[COMP $_S$[turf$_j$ to $_{VP}$[be trodden t_j]]] t_i

However, in the history of most languages, the co-occurrence of active infinitives used in a 'passive sense' and an infinitival marker never obtains, and so infinitival subjects never develop, as in the closely related French. (Lightfoot 1991: 15)

Certain aspects of the interface systems with which language interacts may constrain possible reanalyses. For instance, the correlation between VO/OV and PrepositionN/NPostposition word order is exceptionally strong, even when compared to other harmonic correlations:

(52) Correlation between VO/OV and PrepN/NPost in the 892 languages showing dominant order for both surveyed in Haspelmath *et al.* (2005):

OV & NPost	427 (47.9%)
OV & PrepN	10 (1.1%)
VO & NPost	38 (4.3%)
VO & PrepN	417 (46.7%)

Harmonic word orders obtain in 94.6% of the sample, and disharmonic ones in just 5.4%, three times fewer than in (47). The particular strength of this connection may be related to a common grammaticalisation pathway for adpositions, the reanalysis of serial verbs, as seen with Lahkota adpositions such as *iyu'wex* 'across' vs. *iyu'weya* 'to cross'; *kaxldg* 'through' vs. *kaxldka* 'to pierce' (Pustet 2000: 158). When adpositions arise from verbs in this way, we would expect their position relative to the noun phrase to be preserved: Lakhota for instance is a OV/NPost language. We therefore anticipate that the ordering of adpositions and verbs with respect to their complements be strongly harmonic. The conceptual relationship between serial verbs and adpositions, which renders reanalyses plausible, is therefore reflected in the typological record.

Furthermore, if there is a second common pathway for the grammaticalisation of adpositions, one might reasonably expect it to influence how the languages that are inconsistent with respect to the first pathway are distributed within disharmonic space. Such a pathway is afforded by the development of adpositions from Genitive-Noun/Noun-Genitive locative or body part expressions, for instance, *ni'ih* 'under' vs. *ni'ih* 'foot' in Zapotec languages. (Lillehaugen 2004) In this case, one anticipates the position of adposition to mirror that of the noun (Dryer 2005: 387): Zapotec languages are NGen and PrepN.

(53) Correlation between VO/OV, PrepN/NPost and NGen/GenN in the 37 languages disharmonic with respect to the VO/OV-PrepN/NPost correlation and showing dominant word order for all three, as surveyed in Haspelmath *et al.* (2005):

VO & NPost & NGen	1

OV & PrepN & GenN 1
VO & NPost & GenN 28
OV & PrepN & NGen 7

Of the 29 VO-NPost languages surveyed then, 28 (96.6%) are harmonic with respect to the PrepN/PostN-NGen/GenN correlation. Only 1 language (Majang), just 3.4% of the sample, is disharmonic in both respects. Similarly, out of the 8 OV-PrepN languages surveyed, 7 (87.5%) are harmonic with respect to the prediction of our second pathway, and 1 (12.5%) is doubly disharmonic (Tobelo). The conceptual relationship between locative/body part nouns and adpositions seems to distinguish between the disharmonic word orders of the first pathway very strongly.

The above shows the power of conceptual plausibility in offering finely-grained explanation of asymmetric attestation among disharmonic word orders, just as offered by FOFC in the synchronic domain. Disharmonic word orders illustrate the force of a properly balanced approach to typology particularly effectively, allowing syntactic and diachronic factors to collaborate in producing a more insightful result.

The influence of conceptual plausibility on morphological diachrony is traced by Anderson (2005: 4-5). He demonstrates that some of the burden in explaining the non-random distribution of two case systems across the different past aspects of *split ergative* languages is carried by the semantic relation between possession and completed past action. In such languages, nominative/accusative marking tends to be associated with an imperfective main verb, while ergative/absolutive marking with perfective aspect.

One significant source of innovative perfect forms is documented by Benveniste (1960). The construction that serves to express possession is often used as a marker of the perfect. The verbal expression of possession is often a transitive verb (like English *have,* Spanish tener), but in many languages a distinct prepositional construction is used:

(54) (a) *U menya Ø kniga* [Russian]
at me (is) book
'I have a book'

(b) *Eur velo c'hlas am eus* [Breton]
A bicycle blue at-me is
'I have a blue bicycle'
(Anderson 2005: 4)

Should a construction of this type come to mark the perfect, the subject of the perfect verb will be marked with an oblique (originally locative) case, while the object will be marked like the subject in copular constructions, as a nominative. The result is that new perfects are associated with what is formally an ergative construction, while non-perfects are associated with the original accusative construction. Benveniste argues this is what happened in the development of the Armenian perfect. The subject appears in the genitive, betraying the possessive origin of the construction, while the object appears in the accusative, by a later extension of this case to all objects.

(55) *zayn nšan arareal ēr nora* [Armenian]
that miracle-ACC performed AUX he-GEN
'He performed that miracle'
(Anderson 2005: 4)

Although other pathways are implicated in the development of split ergative licensing, the pathway 'licensed' by the conceptual relationship between possession and completed past action goes some way to explaining the non-random distribution of cases across past aspect in these systems.

Finally, let us turn to the phonological domain. The role of articulatory and perceptual factors in determining directionality of sound change has been extensively traced by *i.a.* Lindblom (1992) and Ohala (1989, 1993). The typological impact of such 'natural' pathways has been formalized by Blevins (2004) as part of the program of *Evolutionary Phonology*. To take just a single example, the relative scarcity of voiced velar stops in the obstruent inventories of the world's languages can be related to certain properties of articulation expressed by Ohala's (1997) *Aerodynamic Voicing Constraint*:

(56) *Voicing requires: (i) vocal cords adducted (lightly approximated at midlines) and (ii) air flowing through the vocal folds. Consequences of this are:*

(a) Voicing is inhibited on obstruents

(b) Factors favoring obstruent voicing are:

(i) shorter closure duration

(ii) larger oral cavity

(iii) active expansion of oral cavity (via larynx lowering, jaw lowering, augmenting velum elevation)

(iv) velic leakage

(Blevins 2006: 139)

The fact that velar leakage favours voicing means that voiced velar stops are relatively unstable with respect to articulation. Their misarticulation as a voiced continuant or devoiced stop is common, promoting their reanalysis and loss from sound systems.

In the above, we have seen that the prevalence of certain diachronic pathways as a result of syntactic, semantic and phonetic 'naturalness' offers real insight into the distribution of languages across possible language space.

6.3 Acquisitional 'contingency'

As I see it, the final logical way of affecting the typological record by diachronic means is through the *indirect* relationship between linguistic features in the acquisitional process. That is, if one certain linguistic feature is a necessary part of the acquisitional evidence for another (as a *cue* or *trigger*), loss of the former will result in ambiguity in the expression of the latter, and plausibly its subsequent loss. A possible example of this is offered by Biberauer & Roberts (2005), who suggest that crucial evidence for V-to-T movement is provided by the rich tense morphology of the verb (contrary to earlier suggestions by *i.a.* Roberts (1985), Vikner (1997) that this is role is performed by rich agreement-morphology) (cf. Chomsky's 1957 'Affix hopping'). As a result of the loss of synthetic tenses, English, Swedish, Norwegian etc. cannot support V-to-T movement, whereas richly inflected Romance languages (Spanish, Italian etc.) retain it. Acquisitional 'contingency' may therefore correlate rich tense morphology and V-to-T movement in the typological record.

A further means of introducing ambiguity in the expression of a syntactic setting is as the recursive result of earlier syntactic changes. (Longobardi 2001: 278) In such cases, a *parametric cascade* obtains. Biberauer & Roberts (2008) present a detailed example of such a cascade in the history of English, relating each step to ambiguity introducing by from the previous one:

(57) i) Loss of VP-to-Spec-*v*P movement (late 12th/early 13th century)
ii) Restriction of object shift to negative and quantified objects (1400)
iii) Loss of *v*P-movement to Spec-TP (early 15th century)
iv) Loss of V2 (1450)
v) Development of lexical T (modals and *do*) (1525)
vi) Loss of V-to-T (1575)
vii) Contraction of negation (1600)
viii) Development of negative auxiliaries (1630)
ix) Development of *do*-support (later 17th century)
(*ibid*.: 26-7)

It is tempting to suggest that such parametric cascades might contribute to an explanation of the typological record. If there are certain cascades that languages 'fall down' then that should be reflected in "basins of attraction" in possible language space. However, as Biberauer & Roberts demonstrate, parametric cascades are far too contingent on language-specific conditions to be anything like general:

> *[t]hese changes were a consequence of certain aspects of the initial conditions and the intermediate stages. A minor difference at any stage could have prevented further changes from taking place, or led to different changes.*
> (*ibid*.: 25)

Specifically, they show that: change (i) is contingent on the loss of particle verbs and ditransitive constructions (possibly under the influence of borrowing from French); (iv) on the decliticisation of subject clitics; and (ix) on existence of a potential dummy auxiliary.

As a result of such contingency only very closely related languages have any hope of following the same cascades, yet even the closely related Scandinavian languages only take the first two steps in the above cascade, because of the total absence of subject clitics from North Germanic. Parametric cascades therefore likely have little explanatory value in explaining typological patterns.

We have seen in the above that a fully nuanced, formal approach has significant force in explaining the typological record. The advantages of an approach that allows abstract, synchronic explanation alongside diachronic was particularly with respect to disharmonic word orders. And in as much as grammatical competence and abstractional modules are implicated in the above, we therefore have a theory of typology based partly on efficiency-oriented third factors.

7. A fuller picture

In this final section, I wish to highlight certain aspects of computational optimality implicated in the design of the FL, but not covered by the third factors already presented. Each is associated with the phase-based nature of derivation.

(i) We have discussed how the FL minimises the memory burden associated with storing incompletely fulfilled features of derivation through the proposed third factor of $MaOP_1$. However, we have not discussed the two other possible burdens on the cache – storing *unintroduced* elements, or storing *completed* derivation. In both cases, the burden is alleviated by derivation by phase.

Phases are subarrays of the numeration, and so reduce the number of lexical items (LIs) that must be available in active memory at any given stage. (Chomsky 2000) Similarly, the periodic "purging" (Richards 2004b: 72) of the results of completed derivations reduces the space complexity of both the syntactic (Chomsky 2001) and phonological (Chomsky 2004) algorithms.

(ii) We have discussed the minimization of search operations associated with valuing features in the course of the derivation, through discussion of MiD_1; but not those associated with *Select*-ing a Lexical Item from the Numeration. In Rizzi's (2007) terms, we have discussed the operations of *Internal Search*, but not those of *External Search*. As *sub*-arrays of the Numeration, phases reduce the average number of unintroduced Lexical Items searched before one with the relevant features is encountered. This offers a clear reduction in the time complexity of the algorithm. Precisely this insight is pursued by simple *binary-splitting*, or *divide-and-conquer*, list sorting algorithms (such as 'Merge Sort') in computer science. These algorithms (recursively) *partition* an unsorted list into two sub-lists of roughly half the size and then sort these sub-lists, before merging the two sub-lists into one sorted list. This procedure provides a significant minimization in time complexity, since the extra operational cost associated with merging two sorted lists is insignificant – each list must be traversed only once – and it is greatly outweighed by the operational saving in comparisons performed sorting each sub-list. (Cormen *et al.* 1990: 12-13) Phase-based syntax derivation is equivalent to sorting sub-arrays of the Numeration into the right structural relations by search operations. Although extra *Spell-Out* operations must be performed as a result, this is more than compensated for by the reduction in operations performed searching unintroduced Lexical Items.

(iii) I wish to introduce an aspect of the optimality of phases to the best of my knowledge not previously discussed in the literature, again drawing on properties of optimal design documented in computer science. The relevant concept here is that of *throughput*. When a computational task is divided into stages, the output of one element becomes the input of the next. The 'throughput' of computation is a measure of how often the final processing element produces an output, how often the task is completed. *Pipelining* is an implementation technique that allows serial processing elements to operate in parallel. Each stage in the *pipeline* will periodically pass on a completed section of computation to the next stage, which can then perform its processing of it in parallel with the continuing computation of the previous stage. In turn, it will periodically avail some output to the next stage in the pipeline, before resuming its processing. And so on. This allows multiple processing elements to be working on a complex task in parallel. While it does not decrease the time each stage takes to perform a given amount of computation, it increases

the instruction throughput of the series as a whole. And as such, the rate at which tasks are completed is increased. (Hennessy & Patterson 2006: A2-A77)

A precise analogy to this is afforded by the language faculty. The complex task of connecting form and meaning is undertaken by a series of processing elements performing different elements of this task, for instance: the mapping between CI and *partially ordered* 'mental form' (syntax), the complete ordering of 'mental form' (the parser), and the mapping from completely ordered 'mental form' to the AP interface (phonology). Phase-based derivation holds that each stage performs computation in a 'piece-meal' fashion, periodically availing a section of its output to the next stage of the pipeline. As such, syntactic, linearising, phonological etc. processing may operate in parallel improving the throughput of the FL itself: meaning is related to form more efficiently. Phase-based derivation therefore offers optimal *scheduling* of the modules comprising the FL.

The upshot of the above is that we may now present an extended system of efficiency-oriented third factors implicated in the design of the FL. This is structured around the three observed means of improving the efficiency with which meaning and form are related: reducing the time complexity of the FL's algorithms, reducing the space complexity of the FL's algorithms and maximizing the throughput of the FL's algorithmic chain.

(58) A system of efficiency-oriented third factors operative in the FL:

Minimize Time Complexity

i) Minimize Forms$_1$
ii) Generalise Features
iii) Minimize Search
    (a) Minimize Internal Search, i.e. Minimize Domains$_1$
    (b) Minimize External Search, i.e. Partition Arrays

Minimize Space Complexity

i) Minimize Caching of Incomplete Derivation, i.e. Maximize Online Processing$_1$

ii) Minimize Caching of Unintroduced Elements
iii) Minimize Caching of Completed Derivation

Maximize Throughput
Optimise Scheduling

I think it is possible that something like the above system of efficiency-oriented third factors may offer a useful heuristic by which the aims of the Minimalist Program may be pursued more rigorously.

8. Conclusion

In the above, I have critically assessed Hawkins' prototypical functionalist theory of grammar and argued instead for an innate, formal (potential for) grammatical competence. I reformulated the performance principles Hawkins takes to motivate structure as principles of optimal computation, which then guided an examination of the empirical evidence for Chomsky's evo devo hypothesis for language. The findings contend that the Minimalist Program is indeed an extremely fruitful approach: computational efficiency seems to be widely implicated in the design of the language faculty, relating form and meaning in a near-optimal fashion. It also forms a productive element in a properly balanced explanation of linguistic typology.

I also attempted to expand a clearer framework for the pursuit of Minimalist inquiry. We saw that greater understanding emerged through increased dialogue with computational science. Notions such as time complexity, space complexity and pipelining allowed for improved insight into, among other things, the linearisation of syntactic structure, markedness in acquisition, and the co-operation of the different modules. If the faculty of language is indeed structured in accordance with general cognitive fundaments, this renews linguistics' role in offering a privileged window onto the human mind. As such, a deeper scrutiny of what it really means to be computationally optimal is of general significance to cognitive science.

Bibliography

Ackema, P. & A. **Neeleman**. (2002). 'Effects of short-term storage in processing rightward movement', in S. Nooteboom, F. Weerman & F. Wijnen (eds.), *Storage and Computation in the Language Faculty*, pp. 219-256. Dordrecht: Kluwer.

Anderson, S. (1998). 'A Formalist's Reading of Some Functional Work in Syntax', in F. Newmeyer, M. Darnell, E. Moravcsik, M. Noonan & K. Wheatley (eds.), *Functionalism and Formalism in Linguistics, Volume I: General Papers*, pp. 111-36. Amsterdam: John Benjamins.

Anderson, S. (2005). 'Morphological Universals and Diachrony', in G. Booij & J. van Marle (eds.), *Yearbook of Morphology 2004*, pp. 1-17. Dordrecht: Springer.

Baddeley, A. (1986). *Working memory.* Oxford: Oxford University Press.

Bandhu, C., B. **Dahal**, A. **Holzhausen** & A. **Hale**. (1971). *Nepali segmental phonology*. Kirtipur: Tribhuvan University.

Baker, C. (1979). 'Syntactic theory and the projection problem', in *Linguistic Inquiry* 10, pp.533-581.

Battistella, E. (1996). *The logic of markedness*. Oxford: Oxford University Press.

Benveniste, E. (1960). '"Être" et "avoir" dans leurs fonctions linguistiques', in *Bulletin de la Société de Linguistique de Paris*.

Biberauer, M.T. (forthcoming). 'Introduction' to T. Biberauer (ed.), *The Limits of Syntactic Variation*. Amsterdam: John Benjamins.

Biberauer, M.T., A. **Holmberg** & I. **Roberts**. (2008). 'Disharmonic word-order systems and the Final-over-Final-Constraint (FOFC)', in A. Bisetto & F. Barbieri (eds.), *Proceedings of IGG 33*. [As online as http://amsacta.cib.unibo.it/archive/00002397/01/PROCEEDINGS_IGG33.pdf].

Biberauer, M.T. & M. **Richards**. (2006). 'True Optionality: when the grammar doesn't mind', in C. Boeckx (ed.), *Minimalist Essays*, pp. 35-67. Amsterdam: John Benjamins.

Biberauer, M.T. & I. **Roberts**. (2005). 'Subjects, Tense and Verb-movement in Germanic and Romance'. Paper presented at GLOW in Asia V, New Delhi.

Biberauer, M.T. & I. **Roberts**. (2008). 'Cascading Parameter Changes: Internally-driven Change in Middle and Early Modern English', in T. Eythórssen (ed.), *Grammatical Change and Linguistic Theory: The Rosendal Papers*. Amsterdam: John Benjamins. [Page numbers above refer to version online at http://people.pwf.cam.ac.uk/mtb23/NSP/]

Bickerton, D. (1981). *Roots of Language*. Ann Arbor: Karoma.

Bickerton, D. (1988). 'Creole Languages and the Bioprogram', in F. Newmeyer (ed.), *Linguistics: The Cambridge Survey*. Cambridge: Cambridge University Press, pp. 267-84.

Blevins, J. (2004). *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.

Blevins, J. (2006). 'A theoretical synopsis of Evolutionary Phonology', in *Theoretical Linguistics* 32, pp. 117-65.

Bowerman, M. (1988). 'The 'no negative evidence' problem: How do children avoid constructing an overly general grammar?', in J. Hawkins (ed.), *Explaining language universals*, pp. 73-101. Oxford: Blackwell.

Caplan, D. & G. **Waters**. (1999). 'Verbal working memory and sentence comprehension', in *Behavioral and Brain Sciences* 22, pp. 77-126.

Carroll, S. (2005). *Endless Forms Most Beautiful: The New Science of Evo Devo and the Making of the Animal Kingdom*. New York: Norton.

Cardinaleti, A. & M. **Starke**. (1999). 'The typology of structural deficiency: A case study of the three classes of pronouns', in H. van Riemsdijk (ed.), *Clitics in the Languages of Europe: Empirical Approaches to Language Typology*, pp. 145-233. New York: Mouton de Gruyter.

Chierchia, G. (2004). 'Scalar Implicatures, Polarity Phenomena, and the Syntax/Pragmatics Interface', in A. Belletti (ed.), *Structures and Beyond*, pp. 39-103. Oxford: Oxford University Press.

Chierchia, G., M.T. **Guasti**, A. **Gualmini**, L. **Meroni**, S. **Crain** & F. **Foppolo**. (2004). 'Implicature, Relevance, and Default Pragmatic Inference; Semantic and Pragmatic Competence in Children's and Adults' Comprehension of 'or'', in I. Noveck & D. Sperber (eds.), *Experimental Pragmatics*, pp. 301-21. Basingstoke: Palgrave MacMillan.

Chomsky, N. (1957). *Syntactic Structures.* The Hague: Mouton.

Chomsky, N. (1964). *Current Issues in Linguistic Theory*, in J. Fodor & J. Katz (eds.), *The Structure of Language: Readings in the Philosophy of Language*, pp 50-118. Englewood Cliffs, Prentice Hall Inc..

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA.: MIT Press.

Chomsky, N. (1975). *Reflections on Language*. New York: Pantheon Books.

Chomsky, N. (1980). *Rules and representations*. Oxford: Blackwell.

Chomsky, N. (1981). *Lectures of Government and Binding*. Dordrecht: Foris.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin and Use.* London: Praeger.

Chomsky, N. (1993). 'A Minimalist Program For Linguistic Theory', in K. Hale & S. Keyser (eds.), *The View from Building 20*, pp. 1-52. Cambridge, MA.: MIT Press.

Chomsky, N. (1995a). 'Bare Phrase Structure', in G. Webelhuth (ed.), *Government and Binding theory and the Minimalist Program*, pp. 383-439. Oxford: Blackwell.

Chomsky, N. (1995b). *The Minimalist Program*. Cambridge, MA.: MIT Press.

Chomsky, N. (2000). 'Minimalist Inquiries: The Framework', in R. Martin, D. Michaels & J. Uriagereka (eds.), *Step by step: Essays in Minimalist Syntax in Honor of Howard Lasnik*, pp. 89-156. Cambridge, MA.: MIT Press.

Chomsky, N. (2001). 'Derivation by Phase', in M. Kenstowicz (ed.), *Ken Hale: A Life in Language*, pp. 1-52. Cambridge, MA.: MIT Press.

Chomsky, N. (2004). 'Beyond Explanatory Adequacy', in A. Belletti (ed.), *The Cartography of Syntactic Structures, Vol. 4: Structures and Beyond*, pp. 104-31. Oxford: Oxford University Press.

Chomsky, N. (2005a). 'Three Factors in Language Design', in *Linguistic Inquiry* 36, pp. 1-22.

Chomsky, N. (2005b). 'Some simple evo-devo theses: how true might they be for

language?'. Talk given at the Alice V. and David H. Morris Symposium on Language and Communication, Stony Brook University, New York, October 14th 2005. [Page numbers above refer to version online at http://www.punksinscience.org/kleanthes/courses/MATERIALS/Chomsky_Evo-Devo.doc].

Chomsky, N. (2007). 'Approaching UG from Below', in U. Sauerland & H. Gartner (eds.), *Interfaces + Recursion = Language?: Chomsky's Minimalism and the View from Syntax-Semantics*, pp. 1-29. New York: Mouton de Gruyter. [Page numbers above refer to version online at https://netfiles.uiuc.edu/bhatia/www/synsem/Chomsky.pdf].

Chomsky, N. (2008). 'On Phases', in R. Freidin, C. Otero & M. Zubizaretta (eds.), *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*, Cambridge, MA.: MIT Press. [Page numbers above refer to version online at http://www.phon.ucl.ac.uk/home/hans/mrg/chomsky_onphases_1204.pdf].

Chomsky, N. & M. **Halle**. (1968). *The Sound Pattern of English*. New York: Harper & Row.

Clark, R. & I. **Roberts**. (1993). 'A Computational Approach to Language Learnability and Language Change', in *Linguistic Inquiry* 24, pp. 299-345.

Clements, G. (2003). 'Feature economy in sound systems', in *Phonology* 20, pp. 287-333.

Clements, G. (forthcoming). 'The Role of Features in Phonological Inventories', to appear in E. Raims & C Cairns (eds.), *Contemporary Views on Architecture and Representations in Phonological Theory*. Cambridge, MA.: MIT Press.

Computer Science Education Project. (1995). *Computer Architecture*. US Dept. of Energy. [Online at http://www.rzg.mpg.de/~rfs/comas/Helsinki/helsinki04/CompScience/csep/csep1.phy.ornl.gov/ca/ca.html].

Comrie, B. (1981). *The Languages of the Soviet Union*. New York: Cambridge University Press.

Comrie, B. (1989). *Language Universals and Linguistic Typology*. Oxford: Blackwell.

Cormen, C., C. **Leiserson** & R. **Rivest**. (1990). *Introduction to Algorithms*. McGraw-Hill.

Coppock, E. (2007). 'Towards a True Theory of the Periphery'. Presented at the 33rd Annual Meeting of the Berkeley Linguistic Society 2007, Berkeley, CA. [As online at http://www.stanford.edu/~coppock/].

Crowley, T. (1992). *An Introduction to Historical Linguistics*. New York: Oxford University Press.

Culicover, P. (1999). *Syntactic Nuts*. Oxford: Oxford University Press.

DeGraff, M. (2003). 'Against Creole Exceptionalism', in *Language* 79, pp. 391-410.

DeGraff, M. (2004). 'Against Creole Exceptionalism (redux)', in *Language* 80, pp. 834-9.

DeGraff, M. (forthcoming). 'Language Acquisition in Creolization: Uniformitarian Approaches', to appear in S. Kouwenberg & J. Singler (eds.), *Handbook of Pidgin and Creole Linguistics*. Oxford: Blackwell.

Dell, F. (1985). *Les règles et les sons*. 2nd edition. Paris: Hermann.

Długosz-Kurczabowa, K. & S. **Dubisz**. (2001). *Gramatyka historyczna je,zyka polskiego*. Warszawa: Wydawnictwo Uniwersytetu Warszawskiego.

Dryer, M. (2005). 'Relationship between the Order of Object and Verb and the Order of Adposition and Noun Phrase', in M. Haspelmath, M. Dryer, D. Gil & B. Comrie (eds.), The World Atlas of Language Structures, pp. 382-7. Oxford: Oxford University Press.

Elbert, S. & M. **Pukui**. (1979). *Hawaiian grammar*. Honolulu: University Press of Hawaii.

Fitch, W.T., M. **Hauser** & N. **Chomsky**. (2005). 'The evolution of the language faculty: Clarifications and implications', in *Cognition* 97, pp. 179-210.

Fodor, J. (1978). 'Parsing strategies and constraints on transformations', in *Linguistic Inquiry* 9, pp. 427-93.

Foulet, L. (1990). *Petite syntaxe de l'ancien français*. Paris: Champion.

Franks, S. & Ž. **Bošković**. (2001). 'An argument for multiple spell-out', in *Linguistic Inquiry* 32: 174-83.

Frazier, L. & G. **Flores D'Arcais**. (1989). 'Filler driven parsing: a study of gap-filling in Dutch', in *Journal of Memory and Language* 28, pp. 331-42.

Fromkin, V. (1999). 'Some thoughts about the mind/brain/language interface', in *Lingua* 100, pp. 3-27.

Gibson, E. (1998). 'Linguistic complexity: locality of syntactic dependencies', in C*ognition* 68, pp. 1-76.

Greenberg, J. (1963). 'Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements', in J. Greenberg (ed.), *Universals of Grammar*, pp. 73-113. Cambridge, MA.: MIT Press.

Grice, H. P. (1969). 'Utterer's meaning and intentions', in *Philosophical Review* 78, pp. 147-77.

Grice, H. P. (1975). 'Logic and conversation', in P. Cole & J. Morgan (eds.), *Syntax and semantics, Vol. 3: Speech Acts*. New York: Academic Press.

Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA.: Harvard University Press.

Hale, M. & M. **Kissock**. (forthcoming). 'The Phonetics-Phonology Interface and the Acquisition of Underspecification', to appear in G. Ramchand & C. Reiss (eds.). *Handbook on Interface Research in Linguistics*. Oxford: Oxford University Press. [As online at http://modlang-hale.concordia.ca/Ramchfinal.pdf].

Hale, M. & C. **Reiss**. (1998). 'Formal and Empirical Arguments Concerning Language Acquisition', in *Linguistic Inquiry* 29, pp. 656-83.

Halle, M. & A. **Marantz**. (1993). 'Distributed Morphology and the pieces of inflection', in K. Hale & J. Keyser (eds.), *The View from Building 20*, pp. 111-76. Cambridge, MA.: MIT Press.

Halle, M. & J.-R. **Vergnaud**. (1987). *An Essay on Stress*. Cambridge, MA.: MIT Press.

Haspelmath, M. (2002). 'Functionalist linguistics: usage-based explanations of language structure'. Handout from talk given at Düsseldorf Summer School, 15th July 2002. [As online as http://email.eva.mpg.de/~haspelmt/papers.html].

Haspelmath, M., M. **Dryer**, D. **Gil** & B. **Comrie**. (2005). *The World Atlas of Language Structures.* Oxford: Oxford University Press.

Hauser, M., N. **Chomsky** & W. T. **Fitch**. (2002). 'The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?', in *Science* 298, pp. 1569-79.

Hawkins, J. (1979). 'Implicational universals as predictors of word order change', in *Language* 55, pp. 618-48

Hawkins, J. (1983). *Word Order Universals*. New York: Academic Press.

Hawkins, J. (1994). *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Hawkins, J. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Hawkins, J. (2007). 'An efficiency theory of complexity and related phenomena'. Handout from a talk presented to the RCEAL Colloquium, University of Cambridge, 27th November 2007.

Hellman, M. (1980). 'A Cryptanalytic Time-Memory Trade-Off', in *IEEE Transactions on Information Theory* 26, pp. 401-6.

Hennessy, J. & D. **Patterson**. (2006). *Computer Architecture: A Quantitative Approach*.

Morgan Kaufmann.

Hinterhölzl, R. (2000). Licensing movement and stranding in the West Germanic OV languages', in P. Svenonius (ed.), *The Derivation of VO and OV*, pp. 293-326. Amsterdam: John Benjamins.

Hock, H. & B. **Joseph**. (1996). *Language history, language change, and language relationship: An introduction to historical and comparative linguistics*. Berlin: Mouton de Gruyter.

Holmberg, A. (2000). 'Deriving OV order in Finnish', in P. Svenonius (ed.), *The Derivation of VO and OV*, pp. 123-152 Amsterdam: Benjamins.

Holmberg, A. & I. **Roberts**. (forthcoming). 'Introduction' to A. Holmberg & I. Roberts (eds.), *Null subjects: the structure of parametric variation.* Cambridge: Cambridge University Press.

Horn, L. (1984). 'Toward a new taxonomy for pragmatic inference: Q-based and R-based implicatures, in D. Schiffrin (ed.), *Meaning, Form, and Use in Context*, pp 11-42. Washington: Georgetown University Press.

Horn, L. (1989). *A natural history of negation*. Chicago: University of Chicago Press.

Hornstein, N. & D. **Lightfoot**. (1987). 'Predication and PRO', in *Language* 63, pp. 23-52.

Huddleston, R. & G. **Pullum**. (2002). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Huddleston, R. & G. **Pullum**. (2003). 'Anomalous adjectives and prepositions in English'. Handout from talk given at the 2003 LSA Annual Meeting.

Ingram, R. (2007). *Neurolinguistics: A Introduction to Spoken Language Processing*. Cambridge: Cambridge University Press.

Jacob, F. (1982). *The possible and the actual*. New York: Pantheon.

Jakobson, R. (1941). *Kindersprache, Aphasie und allgemenie Lautgesetze*. English translation: A. Kieler, *Child Language, Aphasia and Phonological Universals*. The Hague: Mouton.

Joos, M. (1957). *Readings in Linguistics: the development of descriptive linguistics in America since 1925.* Washington, D.C.: American Council of Learned Societies.

Just, M. & P. **Carpenter** (1992). 'A capacity theory of comprehension: Individual differences in working memory', in *The Psychological Review* 99, pp. 122-49.

Kayne, R. (1994). *The Antisymmetry of Syntax*. Cambridge, MA.: MIT Press.

Kayne, R. (2003). 'Antisymmetry and Japanese'. Unpublished ms., New York University.

Keenan, E. (2002). 'Explaining the Creation of Reflexive Pronouns in English', in D. Minkova & R. Stockwell (eds.), *Studies in the History of English: A Millennial Perspective*, pp. 325-55. Berlin: Mouton de Gruyter.

Kenstowicz, M. & C. **Kisseberth**. (1979). *Generative Phonology.* San Diego: Academic.

Kenstowiz, M. (1994). *Phonology in Generative Grammar*. Oxford: Blackwell.

Kiguchi, H. (2006a). 'Cyclic A-movement and Agree in Japanese', in *University of Maryland Working Papers in Linguistics* 14, pp. 127-42.

Kiguchi, H. (2006b). 'Phases and locality constraints on A-movement in Japanese', in *Studia Linguistica* 60, pp. 34-63.

Kiparsky, P. (1965). *Phonological Change*. Ph.D. thesis, MIT.

Kiparsky, P. (1982). 'From Cyclic Phonology to Lexical Phonology', in H. van der Hulst & N. Smith (eds.), *The Structure of Phonological Representations*, pp. 130-75. Dordrecht: Foris.

Kiparsky, P. (1996). 'The Shift to Head-Initial VP in Germanic', in H. Thráinsson, J. Peter & S. Epstein (eds.), *Studies in Comparative Germanic Syntax: Vol.* II, pp. 140-79. Dordrecht: Kluwer,

Kiparsky, P. (2006). 'The Amphichronic Program vs. Evolutionary Phonology', in *Theoretical Linguistics* 32, pp. 217-36

Kiparsky, P. (2008). 'Universals constrain change; change results in typological generalizations', in J. Good (ed.), *Language Universals and Language Change*, pp. 23-53. Oxford: Oxford University Press. [Page numbers above refer to version online at www.stanford.edu/~kiparsky/Papers/cornell.pdf].

Legate, J. (2003). 'Some Interface Properties of the Phase', in *Linguistic Inquiry* 34: 506-16.

Lehmann, W. (1962). *Historical Linguistics: An Introduction*. New York: Holt, Rinehart & Winston.

Lehmann, W. (1973). 'A structural principle of language and its implications', in *Language* 49, pp.47-66.

Levinson, S. (2000). *Presumptive Meanings*. Cambridge, MA., MIT Press.

Lightfoot, D. (1979). *Principles of Diachronic Syntax*. Cambridge: Cambridge University Press.

Lightfoot, D. (1991). *How to set parameters: arguments from language change.* Cambridge, MA.: MIT Press.

Lillehaugen, B. (2004). 'The Acquisition of Body Part Prepositions in Valley Zapotec

Languages', in *Proceedings of The First Conference on Indigenous Languages of Latin America Proceedings*. University of Texas, Austin. [As online at http://www.ailla.utexas.org/site/cilla1_toc.html].

Lindblom, B. (1992). 'Phonological units as adaptive emergents of lexical development', in C. A. Ferguson, L. Menn & C. Stoel-Gammon (eds.), *Phonological development: models, Research, Implications*, pp. 565-604. Timmonium: York Press.

Longobardi, G. (2001). 'Formal syntax, diachronic minimalism, and etymology: the history of French *chez*', in *Linguistic Inquiry* 32, pp. 275-302.

Luria, S. (1974). Transcript of remarks at 'A Debate on Bio-Linguistics,' a conference organized by the Centre Royaumont pour une Science de l'Homme, Endicott House, Dedham, MA., 20th - 21st May 1974.

Manber, U. (1989). *Introduction to algorithms.* Reading, MA.: Addison-Wesley.

Marantz, A. (1997). 'No Escape from Syntax: Don't Try Morphological Analysis in the Privacy of Your Own Lexicon', in *University of Pennsylvania Working Papers in Linguistics* 4, pp. 201-25.

Marvin, T. (2003). *Topics in the Stress and Syntax of Words*. Ph.D. thesis, MIT.

Moro, A. (2000). *Dynamic Antisymmetry*. Cambridge, MA.: MIT Press.

McCawley, J. (1978). 'Conversational implicature and the lexicon', in P. Cole (ed.), *Syntax and Semantics, Vol. 9: Pragmatics*. New York: Academic Press.

McGinnis, M. (2001). 'Variation in the phase structure of applicatives', in P. Pica & J. Rooryck (eds.), *Linguistic Variations Yearbook 1*, pp. 101-42. Amsterdam: John Benjamins.

Migdalski, K. (2006). *The Syntax of Compound Tenses in Slavic*. Utrecht: LOT Publications.

Nevins, A. (2005). *Conditions on (Dis)Harmony*. Ph.D. thesis, MIT.

Newmeyer, F. (1998). *Language Form and Function*. Cambridge, MA.: MIT Press.

Newmeyer, F. (2001). 'Where is Functional Explanation?', in M. Andronis, C. Ball, H. Elston & S. Neuvel (eds.), *Papers from the Thirty-Seventh Meeting of the Chicago Linguistic Society*, *Part 2: The Panels*, pp. 99-122. Chicago: Chicago Linguistic Society.

Newmeyer, F. (2003). 'Grammar is Grammar and Usage is Usage', in *Language* 79, pp. 682-707.

Newmeyer, F. (2004). 'Against a parameter-setting approach to language variation', in P. Pica, J. Rooryck & J. van Craenenbroek (eds.), *Language Variation Yearbook, Vol. 4*, pp. 181-234. Amsterdam: John Benjamins.

Newmeyer, F. (2006). 'A rejoinder to *On the role of parameters in Universal Grammar: a reply to Newmeyer* by Ian Roberts and Anders Holmberg'. [As online at http://people.pwf.cam.ac.uk/mtb23/NSP/].

Nunes, J. (1999) 'Linearization of Chains and Phonetic Realization of Chain Links', in S. Epstein & N. Hornstein (eds.), *Working Minimalism*, pp. 217-49. Cambridge, MA.: MIT Press.

Ohala, J. (1989). 'Sound change is drawn from a pool of synchronic variation', in L. Breivik & E. H. Jahr (eds.), *Language Change: Contributions to the study of its causes*, pp. 173-98. Berlin: Mouton de Gruyter.

Ohala, J. (1993). 'The phonetics of sound change', in C. Jones (ed.), *Historical Linguistics: Problems and Perspectives*, pp. 237-78. London: Longman.

Ohala, J. (1997). 'Aerodynamics of phonology', in *Proceedings of the Seoul International Conference on Linguistics*, pp. 92-7. Seoul: Linguistic Society of Korea.

Pancheva, R. (2004). 'The Rise and Fall of Second-Position Clitics', in *Natural Language and Linguistic Theory* 23, pp. 103-67.

Paul, H. (1920). *Prinzipien der Sprachgeschichte*. Tübingen: Niemeyer.

Pesetsky, D. (1989). *Language-particular processes and the Earliness Principle*. Unpublished ms., MIT.

Pesetsky, D. & E. **Torrego**. (2001). 'T-to-C movement: causes and consequences', in M. Kenstowicz (ed.), *Ken Hale: A Life in language*, pp. 355-426. Cambridge, MA.: MIT Press.

Pinker, S. & R. **Jackendoff**. (2005). 'The faculty of language: what's special about it?', in *Cognition* 95, pp 201-36.

Holmberg, A. & C. **Platzack**. (2005). 'The Scandinavian languages', in G. Cinque & R. Kayne (eds.), *The Oxford handbook of comparative syntax*, pp. 420-458. Oxford: Oxford University Press.

***Pustet**, R. (2000). 'Lahkotan postpositions', in International Journal of American Linguistics* 66, pp. 157-180.

Rackowski, A. & N. **Richards**. (2005). 'Phase Edge and Extraction: A Tagalog Case Study', in *Linguistic Inquiry* 36, pp. 565-99.

Richards, M. (2004a). *Object Shift and Scrambling in North and West Germanic: A Case Study in Symmetrical Syntax.* Ph.D. thesis, University of Cambridge.

Richards, M. (2004b). 'Keeping Order at PF', in *Cambridge Occasional Papers in Linguistcs* 1, pp. 183-203.

Rivero, M.-L. (1991). 'Patterns of V-raising in Long Head Movement, and Negation', in *The Linguistic Review* 8.

Rizzi, L. (1991). *Relativised Minimality*. Cambridge, MA.: MIT Press.

Rizzi, L. (2001). 'Relativised Minimality Effects', in M. Baltin & C. Collins (eds.), *The Handbook of Contemporary Syntactic Theory*, pp. 89-110. Oxford: Blackwell.

Rizzi, L. (2007). 'On Delimiting Movement'. Talk given at the Cambridge-Conneticut-Hyderabad-Nanzan-Siena-Tsing Hua Consortium for Linguistics, Tsing Hua University, 16th December 2007.

Roberts, I. (1985). 'Agreement Parameters and the Development of the English Modal Auxiliaries', in *Natural Language and Linguistic Theory* 3, pp. 21-58

Roberts, I. (2007). *Diachronic Syntax*. Oxford: Oxford University Press.

Roberts, I. (2008a). Clitics, Head Movement and Incorporation. Unpublished ms., University of Cambridge.

Roberts, I. (2008b). 'Diachrony and Deficiency'. Talk given to *SyntaxLab*, University of Cambridge, February 2008.

Roberts, I. & A. **Holmberg**. (2005). 'On the role of parameters in Universal Grammar: A reply to Newmeyer', in H. Broekhuis, N. Corver, R. Huybregts, U. Kleinhenz & J. Koster (eds.), *Organizing Grammar: Linguistic Studies in Honor of Henk van Riemsdijk*, pp.538-53. Berlin: Mouton de Gruyter.

Roberts, I. & A. **Roussou** (2003). *Grammaticalisation: A Minimalist Approach to Syntactic Change*. Cambridge: Cambridge University Press.

Roeper, T. (1999). 'Universal Bilingualism', in *Language and Cognition* 2, pp. 169-86.

Ross, J. (1967). *Constraints on Variables in Syntax*. Ph.D. thesis, MIT.

Schleicher, A. (1861-2). *Compendium der vergleichenden Grammatik der indogermanischen Sprachen. Kurzer Abriss einer Laut- und Formenlehre der indogermanischen Ursprache, des Altindischen, Altiranischen, Altgriechischen, Altitalischen, Altkeltischen, Altslawischen, Litauischen und Altdeutschen*. Weimar: Böhlau.

Simpson, A. (2004). 'The EPP, fossilized movement and reanalysis', in E. Fuss & C. Trips (eds.), *Diachronic Clues to Synchronic Grammar*, pp. 161-90. Amsterdam: John Benjamins.

Solomon, E. & N. **Perlmutter**. (2004). 'Semantic integration and syntactic planning in language production', in *Cognitive Psychology* 49, pp. 1-46.

Song, J. (2001). *Linguistic Typology: Morphology and Syntax*. London: Longman.

Venneman, T. (1974). 'Topics, subjects, and word order: From SXV to SVX via TVX', in J. Anderson & C. Jones (eds.), *Historical Linguistics: Proceedings of the First International Congress of Historical Linguistics*, pp. 339-76. Amsterdam: North-Holland.

Vikner, S. (1997). 'V-to-I movement and inflection for person in all tenses', in L. Haegeman (ed.), *The New Comparative Syntax*, pp. 189-213. London: Longman.

Wexler, K. (2004). 'Lenneberg's dream: Learning, normal language development and specific language impairment', in L. Jenkins (ed.), *Variation and Universals in Biolinguistics*, pp. 239-84. Amsterdam: Elsevier.

Yang, C. (2002). *Knowledge and learning in natural language*. Oxford: Oxford University Press.