Minimalism and the Design of the Language Faculty

Iain Jerome Mobbs

Downing College

November 2014

This dissertation is submitted for the degree of

Doctor of Philosophy

University of Cambridge

Minimalism and the Design of the Language Faculty

Iain J. Mobbs

SUMMARY

This dissertation explores the cognition underlying the knowledge, use, and acquisition of human language – the *faculty for language*, or *language faculty* (FL). In particular, it examines the validity and utility of the set of related, but logically independent proposals comprising the argument of *linguistic Minimalism* (LM). These proposals define an atheoretic research program for linguistic enquiry – the *Minimalist Program for Linguistic Theory* (MP) (Chomsky 1993) – as well as introducing specific claims about the status of linguistic variation and the FL's broader role in human cognition. Chapter 1 defends the existence of a coherent subject matter for 'biolinguistic' enquiry of this kind. Chapter 2 attempts to clarify those Minimalist proposals which comprise the MP, and to present their compulsory logic; it offers an explicit (if partial) framework for Minimalist research based around principles of algorithmic science. Chapter 3 argues for the validity of the substantive hypotheses of LM: in particular, it defends the largely ancestral nature of humans' *conceptual-intentional* apparatus and proposes a particular characterization of permissible syntactic variation, introducing a novel analysis of *discourse pro-drop* as part of the argument. Chapters 4 and 5 examine the evidence for the implication of Minimalist notions of optimality and 'interface conditions' in the design of the FL. In doing so, these chapters offer new suggestions regarding the motivation for lexical categories, *phase*-based *spell-out* (Chomsky 2001), and abstract Case. Chapter 6 focuses on the status of the parser in LM, suggesting that parsing relies heavily on computationally-optimized cognition and defending a close relationship between syntactic grammar and the parser. The chapter goes on to motivate a (limited) parsing-based account of the linearisation of abstract syntactic structure and to propose an explanation for the *Final-over-Final-Constraint* (Biberauer, Holmberg & Roberts 2014). Chapter 7 suggests that we think about language acquisition as based on parsing with an over-complete grammar, rather than as determining the settings of an under-specified, defective grammar. It is suggested that diachronic evidence of 'conservative' acquisition (including the prevalence of grammaticalisation, "featural economy" (Clements 2003), and the gradual relaxation of the contexts for phonological rules) plausibly reflects concern for optimality in the process of acquisitional parsing. On the other hand, the prevalence of analogical forces in language change is presented as the footprint of 'liberal', computationally-constrained data analysis, attested in acquisitional tasks across cognitive domains. Chapter 8 offers an account of the evolutionary origins of non-referential, conceptually "promiscuous" lexical meaning. Finally, I suggest that the machinery grounding lexical semantics also affords our capacity for metaphorical thought, our rich theory of others, and our solipsistic nature (sometimes called 'the human condition'); I argue that the FL is even more intimately related with what it means to be human than already assumed. Chapter 9 concludes.

DECLARATION

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration with others, except where specifically stated in the text and acknowledgements.

This dissertation is not substantially the same as any submitted for a degree, diploma or other qualification at the University of Cambridge, or at any other university or similar institution.

This dissertation is 79,681 words long, including all footnotes, references, and appendices, but excluding the bibliography.

ACKNOWLEDGEMENTS

I've been given so much help by so many people over the years that I can't establish the distance to appreciate it any more. I do realise I've been unusually lucky though, and that I need other people. Language really isn't built for expressing gratitude, but it'll have to do.

Ian Roberts and Theresa Biberauer have been wonderful supervisors: they got me interested in linguistics, and every time I speak to them I become more interested. I remember Theresa's first lecture very clearly – understanding for the first time that 'knowledge' is something produced and insisted on by people, rather than a series of epiphanies. I guess that's what you call making a discipline come alive. The first time I met Ian, as a first-year undergraduate, was when he troubled to introduce himself to me; that first impression has been persistently reinforced. After years of their patient supervision, I occasionally feel able to work on my own, but I always (nervously) want to know what they think, because it's worth days of my fumbling; and they invariably spare the time to let me know, in the most constructive way. (The usual provisos apply about the content here of course.) They have been entirely supportive and very patient in the ways that matter most too. I'm deeply grateful.

The Department of Theoretical & Applied Linguistics at Cambridge is quite a disparate entity, but many extremely clever and decent people study, teach, and work in it. Bert Vaux (possibly without knowing) gave me a new perspective on many issues in linguistics: in particular, he helped me realise that you can study language in a number of different ways, with a number of different interests, and do so *without contradiction* – a pretty fundamental point, but often missed. I have also learnt a great deal from him about assessing the value of claims in terms of their predictions. James Clackson (actually a member of the Faculty of Classics) was my first DoS at

Jesus College and allowed me to change course from Natural Sciences to Linguistics. I suspect many other DoSes might not have been so supportive, and I will always be much indebted. I'm also very grateful to Francis Nolan, for his early support as the Head of Department when I first started linguistics, and for allowing me to study part-time early in my PhD. He also supervised my first attempt at independent research, and immediately made me feel like I had every right to try. Michelle Sheehan has never formally taught me, but every time I've sat in a room with her she's said something to me think harder. She's also been very kind in answering questions, sending me papers, explaining her thoughts, etc., showing me that the idealized notion of academic dialogue is possible. Jenneke van der Waal (again, possibly without realising) always reminds me that linguistics involves attention to language*s*, all of which are born equal, and therefore not to put theory before data. Her obvious sincerity and empathy always lift my spirits. I'm also very aware and appreciative of the support given by the department's non-teaching staff, particularly: Ulrike Balser, Alison Bingham, Siobhán Carew, Rachel Deadman, Ariel Knapman, and Louise Radok. Life would have been very much harder without them.

There has always been a number of intimidatingly bright (and very humble) students in DTAL at any given time, and it's been a privilege to be allowed to learn with/from them. Alison Biggs, Neil Myler and George Walkden are all obviously outstanding (though they wouldn't think that way of themselves) and I'm not remotely surprised by their early success. I like them all enormously and wish them all the very best of luck. If there is a 'field', with an overarching interest in progress toward answering linguistic questions, it would do well to hang onto them. (I'm not certain there is, but fortunately they are talented enough that it won't matter.) I've had many other peers whose thinking I'm not as familiar with, but who have enriched supervisions, seminars, reading groups, etc., and made for great company, including: Alastair Appleton, András Bárány, Tim Bazalgette, Alice Corr, Chi-Hé Elder, Chris Lucas, Elena Pala, Joe Perry, and Jeffrey Watumull. Emma Game was more or less the only linguistics student I had to talk to as an undergraduate, and although she left Cambridge six years ago, I still catch myself hoping she'll be in the library every time I go in.

Beyond its work, I don't know the linguistics community outside Cambridge particularly well. I have sent a lot of e-mails asking questions though, and received a lot of very helpful replies. Regarding the content included here, I'm very grateful for replies received from: Marijke de Belder, Matthew Dryer, Andrew Nevins, Luigi Rizzi, Bridget Samuels, and Charles Yang. The usual provisos apply.

Noam Chomsky also replied very helpfully to several e-mails, which, again, I'm very grateful for. I know he has no liking for veneration, but this is my acknowledgements section and his work has

made a big impact on me. It's not just that his arguments are so well-reasoned and that his conclusions are fascinating, but that they reflect an honest and humble approach to enquiry. There isn't any dissembling or ego in his work, nor any of the nervous arrogance or 'goal-hanging' of so much academic work. It's just someone thinking really hard because they choose to, reporting what they didn't learn as well as what they did, and never concealing the contribution of others. That's hard not to admire. It's easy to see why Chomsky's contribution is resented and how people become trapped in their instinctive opposition, but I don't think I'm alone in finding the amount of work dismissing his conclusions without engaging with his arguments embarrassing. I'm thankful for his insights and example.

Needless to say, most people I owe a debt of gratitude have no connection with linguistics. My mother and father have gone to enormous lengths to support me and my well-being. I desperately wish I were better at showing my appreciation for this directly, but I think they might finally be cottoning on. I love them very much – I always have and always will – and if any of us ever forgets, I hope that they'll be able to request this in the West Room of the Cambridge University Library. I've just as much love for my twin brother, Andrew, and older brother, Richard. They suffer endless self-centred rubbish from me, but respond with further support. What's more, I am very proud of them both: for who they are, and what they have made of themselves.

I have very kind friends, who also give much more than they receive; I hope to be able to change that some day. Almost all of them will never see this, and many of them go unmentioned here for reasons of space and forgetfulness. Though I have different reasons to be grateful to each of them, I'd like to say a sincere, undifferentiated thank you to: Laura Allsop, Miriam Arkush Lorie, Olivia Bampfylde, Magali Barraja, Alison Biggs, Ian Blaney, Ruth Davis, Catherine Dobson, Rob Foulkes, Olaf Henricson-Bell, Michiel Kamp, Charlotte Malcolm, Nathan Rasiah, Liz Rush, Teresa Segura-Garcia, Geoff Stanning, Tim Swain, and Matt Westlake.

And finally, there are two people who stand apart in my life. They are unrelated to me, and unrelated to each other; they live miles away from me, and miles away from each other; they are very different to me, and very different to each other. Neither of them has any good reason to do anything at all for me, but both do everything in their power for me. Neither has ever asked for anything in return, nor shown the slightest resentment or hesitation. And everything they've done, they've done when no-one is watching. They've given me a meaningful way to think about my own behaviour, and, most of all, they've shown me how to love other people. I've no idea what I did to deserve them. So for what it's worth, this dissertation is dedicated to Peter Batterham and David Sturgeon.

TABLE OF CONTENTS

0. Introduction

Human language differs fundamentally from any superficially similar activity of other organisms, in both its structural and its semantic properties. The initial research questions of linguistic inquiry are: what is contained in the knowledge of language; how is that knowledge acquired; and how is that knowledge put to use? (Chomsky 1986a) If we choose to be concerned with the cognition underlying these abilities – the *faculty for language*, or *language faculty* (FL) – then further questions arise, concerning: the cognitive content of this faculty; its evolution; its physical implementation; and its broader role (if any) in human cognition.

Chapter 1 of this dissertation defends our right to adopt this *bio-scientific* approach to linguistic enquiry, rejecting a range of proposals which attempt to trivialise the existence of the FL. This chapter categorizes these 'empiricist' approaches to language acquisition (FLA) and for each case summarizes the reasons to disregard them. It is suggested that there is instead good reason to posit a coherent object of *biolinguistic* enquiry, over which meaningful questions can be asked.

Over the course of the biolinguistic enterprise, various considerations appropriate to its scientific commitments have arisen. A collection of related, but logically independent proposals has coalesced in the literature of the past twenty-five odd years, particularly in the works of Noam Chomsky; together they comprise the argument of *linguistic Minimalism*. These proposals emerged organically, with the tightening and revision characteristic of nascent theory; as such they cannot be read off the primary literature entirely straightforwardly. The focus of Chapters 2 and 3, therefore, is to accurately characterize these 'Minimalist proposals' and the relationships between them, so that their (often truistic) insights can be properly appreciated and applied. In doing so, I offer some comments on underappreciated aspects of their worth.

Chapter 2 addresses that subset of the Minimalist proposals (the 'first', 'second' and 'third') which is neutral with respect to the design or origins of the FL, but instead outlines a framework for cogent linguistic research – *the Minimalist Program for Linguistic Theory* (MP) (Chomsky 1993). A proper understanding of the Minimalist proposals, and so meaningful pursuit of Minimalist research, relies on a nuanced appreciation of the *interface conditions*, which language must meet to be useable at all, and of two related aspects of cognitive optimality. In as far as it is consistent to present a precise formulation of these notions, I try to do so here. In particular, I co-

opt principles from algorithmic science to help characterize computational optimality. I conclude by presenting a (partial) framework for Minimalist enquiry, which it is hoped will help steer it away from too casual an approach, and protect against the misappropriation of Minimalist logic.

Chapter 3 covers the 'fourth' and 'fifth' Minimalist proposals, which make specific claims about the design and origin of the FL. They stand logically apart from, though in close association with, both the MP and each other. The fourth Minimalist proposal (*i.a.* Chomsky (2005a, b)) suggests that core computational abilities of the FL were selected for their role in grounding a recursive 'language of thought', with language's externalization being a secondary concern. I present a range of considerations which are consistent with this proposal, drawn from linguistic behaviour and further afield. In doing so, I defend the position that relatively rich *conceptual-intentional* and *articulatory-perceptual* interfaces are ancestral to language – something which has been challenged, particularly in the former case. I also present the predictions of the fourth Minimalist proposal for the results of the MP, so that it may be subject to the fullest evaluation possible.

The fifth Minimalist proposal (*i.a.* Berwick & Chomsky 2011) bears on the manner in which the FL permits variation between languages. In particular, it claims that language variation is highly restricted in the syntactic domain, but freer in domains relating to externalization. I attempt to demonstrate (admittedly from a restricted range of examples) that there is good reason to take this claim seriously. In particular, I argue that such unambiguous syntactic variation as there is tallies strikingly with the predictions of the fourth Minimalist proposal, offering a slightly richer characterization of the fifth Minimalist proposal. As part of this argument, I present a novel approach to the phenomenon of *discourse pro-drop*, arguing that it coat-tails on the presence of a wide range of null lexical items associated with 'bare' nominal syntax.

Chapters 4 and 5 begin the task of investigating the FL in light of the Minimalist proposals. In particular they set out to examine language for evidence of the principles of optimal cognition outlined in Chapter 2, while speculating to what extent breaches of optimality might be 'principled' in light of interface conditions. To the extent that they are related to the MP, the plausibility of the fourth and fifth Minimalist proposals is also at stake. Chapter 4 focuses on the principle of computational optimality which serves to minimize 'operational load', whereas Chapter 5 focuses on 'cache load'. The second notion of optimality, 'substantive' optimality, and principled breaches of optimality are addressed in both chapters.

Chapters 4 and 5 bring together existing models and Minimalist analyses of linguistic phenomena, as well as offering original analyses. In particular, in Chapter 4, I argue that lexical categories are a principled breach of substantive optimality: they have no (novel) semantic content in themselves, but allow for lexical items bearing conceptual content to be related together without licensing by complicated interpretative computation. I also dispute Nevins's (2010) claim that patterns of vowel harmony in the world's languages demonstrate the influence of the same 'minimal search' concerns evident in syntactic computation. Instead I argue that the locality of harmony arises from adjacency requirements stated explicitly in phonological rules, acquired through misinterpretation of (anticipatory or perseverative) co-articulation – the traditional position Nevins dismisses. This argument is consistent with the prediction of the fourth Minimalist proposal, that narrow syntax might make more *explicit* use of principles of computational optimality than phonology.

Chapter 5 also introduces an unfamiliar understanding of the computational optimality of *phase-based Spell-Out* (Chomsky 2001). Previous conceptions of the computational optimality relevant to Minimalist enquiry, including Chapter 2's framework, have been narrowly 'algorithmic', neglecting the fact that computational tasks often consist of consecutive algorithms, with the output of one being the input of the next. It is therefore possible to increase the rate of *throughput* (the overarching principle of computational optimality) by optimizing the 'scheduling' with which consecutive algorithms release output. I argued that this is the effect of phasal Spell-Out within the FL: a principle of optimal scheduling is added to the (partial) framework for Minimalist enquiry. Chapter 5 also presents a tentative account of the computational role of abstract Case in syntax. It is argued that Case-features are uninterpretable (meaningless) at the conceptual-intentional interface, but have utility in minimizing the burden placed on working memory by syntactic computation. This argument relies on a subtler understanding of the structure of working memory, motivated in studies by *i.a.* King & Just (1991).

Chapter 6 continues to track the cache burden imposed by the FL's computation, but focuses on an area of FL's activity largely unattended by Minimalist enquiry – parsing. Core results from parsing research are presented and assessed in these terms: 'left-to-right' incrementality, parsing domains, and the 'filler-first' approach to establishing filler-gap dependencies. Pritchett's (1992 a, b) finding that the parsing algorithm is thematically-oriented and head-driven is also introduced and assessed in Minimalist terms; I also suggest a correction to his model on empirical grounds. A picture of close identity between the abstract parser and the grammar seems to emerge, i.e.

evidence of substantive optimality within the FL. The revision made to Pritchett's model strengthens the argument for the 'monotonicity' of the abstract parser, in keeping with a *No Tampering Condition* reducing operational complexity. This chapter goes on to motivate the results of the parsing literature in limited exploration of Chomsky's (2007b) suggestion that there might be a parsing-based account of the linearisation of abstract syntactic structure. I suggest that the findings of research into linearisation are compatible with the implication of computational optimality in abstract parsing of this kind. Finally, I suggest that the *Final-over-Final-Constraint* (Biberauer, Holmberg & Roberts 2014) falls out from concerns for minimizing working memory when mapping surface strings onto the functional hierarchies of syntactic (cf. work in the 'cartographic' enterprise (*i.a.* Rizzi 1997)).

While Chapter 6 attempts to bring word order effects within Minimalist research, Chapter 7 tries to do the same for certain aspects of FLA. I take FLA to be the process of paring down an over-specified grammar to a smaller grammar, similar to the one(s) possessed by the users generating acquirers' input. This contrasts with the 'switchboard' metaphor commonly invoked to explain (syntactic) acquisition, which holds that learners start with an underspecified, defective grammar. I suggest that findings revealing 'conservative acquisition' reflect the activity of computational optimization in the (syntactic, morphological and phonological) parsing of the *learner-parser*. In particular, I argue that concern to reduce operational load in the parsing process results in the minimal form being postulated to explain input, reflected in diachrony by: the prevalence of grammaticalisation, the tendency to analyse separate morphemes fusionally, the tendency to relax constraints in phonological rule acquisition, and the fact that sound systems tend to use featural contrasts with "maximal efficiency", expressing "featural economy" (Clements 2003). I follow Vaux (2009) in associating this 'conservatism' with the same force underlying 'over-shadowing' in classic studies from experimental psychology. In this chapter, I also follow *i.a.* Ferdinand *et al.* (2013)'s suggestion that the over-generalization effects familiar from child FLA studies reflect a principle of data analysis imposed by the limits of working memory, suggesting that this tendency underlies the prevalence of analogical extension and levelling in the diachrony of morphology. I also discuss the clustering of word order systems around 'harmonic' poles in typology, and the problems associated with adopting an 'analogical' account of this phenomenon.

The final substantive chapter, Chapter 8, picks up earlier discussion concerning the evolution of the FL and its broader role in human cognition. The fourth Minimalist proposal suggests that narrow syntactic abilities were selected on grounds of vastly expanded capacity for abstract thought (the 'human capacity'). Chapter 3 points out that such thought relies not just on the

capacity for unbounded nesting recursion afforded by narrow syntax, but also on the abstract, 'conceptually eclectic' lexical items over which it operates, also seemingly unique in the animal world. This chapter argues that the machinery grounding syntactic computational abilities in actual fact more or less 'gifts' the "promiscuity of the interfaces" which underlies non-referential lexical meaning, *contra* Chomsky's (2012) claim that there are "no sensible origins" for this property of human language. I go on to suggest that the conceptual flexibility and compositionality of the lexicon grounds our capacity for metaphorical thought and extension, and so our ability to think about things beyond our innate conceptual domains. I also suggest that our rich theory of others and our solipsistic nature (commonly referred as the 'human condition') stems from lexically-based conceptual 'substitution' or 'denial', and so that the FL is even more intimately related with what it means to be human than already assumed.

Chapter 9 suggests that the prior dissertation demonstrates that, when pursued appropriately, the proposals of linguistic Minimalism are very powerful tools for investigating the design of the FL. It furthermore suggests that in as far as there are concrete Minimalist hypotheses, the findings of early research are supportive. A few closing comments are made consequent to this.

1. Biolinguistics and the Logical Conclusion

1.1 Biolinguistics

This dissertation attempts to contribute to the ongoing investigation of human language from a biological perspective. It addresses the knowledge, use and acquisition of language as reflexes of particular *cognitive* activity, as "emergent properties" of the human brain. These issues can be studied without contradiction from other perspectives, depending on one's interests: it would not be inconsistent, for instance, to study how speakers' knowledge and/or usage reflect social or historical factors (which, of course, it has been shown they do, in interesting ways.) The approach here, however, is a *biolinguistic* one.

The biolinguistic enterprise, or *biolinguistics*, adopts the standard ambitions of natural science when investigating a biological phenomenon: to develop an explanatory account of whatever lies behind the observed data – in this case, of the human *faculty for language*, or *language faculty* (FL) – and of its physical implementation and evolution (Chomsky 2010: 2, 5). This dissertation, however, makes no original observations whatsoever regarding the physical implementation of the FL, and restricts any comments on the FL's evolution to mechanisms of change, neglecting other dimensions, such as the genetic and the dynamic. Again though, that is not a statement of principle, merely of ignorance.

Biolinguistics holds that the language faculty is a mental "organ" and that it may be "abstracted for special investigation because [its] apparent internal integrity and special properties" (Chomsky 2007b: 2). The relevant "special properties" are that the FL appears to be both species-specific and common to more or less all humans. Language is acquired reflexively [1] except in cases of extreme pathology or extremely hostile learning conditions during a critical period. That is to say, biolinguists take it that there is a *coherent object of enquiry*, adopting the "modular view of learning" considered "the norm these days in neuroscience" (Gallistel 1997: 86).

[1] By analogy with other cognitive organs in the animal world, it is generally assumed that certain types of social interaction are instrumental in *triggering* this reflex. For example, it has been reported that mother-neonate contact is a prerequisite for the development of the mammalian visual system (Chomsky 1983). See also fn.3.

It is important to note immediately that a modular view of learning does not determine in what fashion a module is unique, leaving the whole spectrum of possibilities available: from the module's unique 'co-ordination' of external cognitive resources, to absolute organ-specificity. [2] The location of the FL on this spectrum is one of the chief concerns of *linguistic Minimalism* – a set of proposals regarding the proper pursuit of biolinguistic research and the design of the FL introduced in the next chapter.

1.2 The "poverty of the stimulus"

The term 'biolinguistics' is forty years old, having first been suggested by Massimo Piattelli-Palmarini, the organizer of a conference on the topic. The biolinguistic *enterprise*, however, began twenty years earlier, implicit in the very first works of generative grammar (Chomsky ([1955], 1975), 1957, 1959, 1964, 1965). These studies assumed a "realist interpretation" of the transformational generative grammars they proposed, arguing that they represented "the competence attained by the normal speaker-hearer … employed in the use and understanding of language … and specify[ing] the schematism brought to bear … in language acquisition" (Chomsky ([1955], 1975: 45)).

The term 'biolinguistics' itself is used explicitly with greater frequency in more recent literature, but the "generative grammar/biolinguistic amalgam" (Chomsky 2008: 134) has mainly been pursued through preoccupation with an implicit concern of all biological research: how to account for the huge discrepancy between the paucity of information available to an organism from its environment and the complexity of the structures/competencies it develops. The *argument from poverty of the stimulus* (APS) (Chomsky 1980) points out that linguistic competence is radically underdetermined by the data to which a language acquirer (LA-er) is exposed: infinitely many systems are consistent with a finite data set, yet inconsistent with one another. Given the speed and accuracy with which LA-ers converge on grammatical systems despite this insufficient [3] (and

[2] Generally taken to be impossible, because at some level – for example the cellular – everyone assumes that some machinery is held in common (Chomsky 2005b: 5).

[3] A particularly extreme example of this is the famous case of *Nicaraguan Sign Language*, a language acquired by a generation of deaf Nicaraguans despite the absence of any productions of a developed language in their *primary linguistic data* (PLD) (*i.a.* Senghas *et al.*1997, Senghas & Coppola 2001, Newport & Coppola 2005). Cf. also work on *Al-Sayyid Bedouin Sign Language* by Wendy Sandler and colleagues in Sandler *et al*. (2005) *et*

widely varying) input, whatever is responsible must be "biased or constrained in certain ways" (Hauser, Chomsky & Fitch (HCF hereafter) (2002: 1577)) to be able to account for it. A rich innate grammatical competence [4] as part of the FL is motivated by this need for *explanatory adequacy* (Chomsky 1964).

In other branches of biology, the poverty of the stimulus (PoS) is "taken to be so obvious that there is no need for a term" (Chomsky 2010: 6). For instance, there is clearly no way in which external data such as nutrition can determine why "humans develop arms instead of wings"; the fact that genetic information must guide development in relevant ways is self-evident and "no-one even bothers to argue about it". Variations on the theme arise for "every aspect of [the] growth and development" (*ibid.*) of every organism. [5]

And this same ubiquity applies to the development of language. The APS has traditionally been expressed by pointing out that it is logically impossible for induction alone to derive the observed systems of linguistic rules from *primary linguistic data* (PLD) – the presentation given above. However, an innate contribution on the part of the FL is in fact necessary to make sense of a wide range of acquisitional 'tasks'. [6]

Most fundamentally, any account of FLA must "presuppose (at least tacitly) that a child [is attentive to and] can somehow distinguish linguistic materials from the rest of the confusion around [them]" (Chomsky 2002: 85-6). The principles of the FL must include a definition of the relevant phonetic dimensions of language if a "problem of selective attention" (Gelman 2009: 228) is to be solved – that is, if we are to avoid "the common circular argument that selective attention is due to salience and salience directs attention" (*ibid.*).

---

seq.. These studies strengthen the assumption of fn.1 and argue for greater research into the precise nature of the environment required to trigger language acquisition (FLA).

[4] This has commonly been referred to as *Universal Grammar* (UG) (Chomsky 1957, 1965). We eschew this use of the term here for reasons of consistency: as will be seen in the following chapter, the term UG is now reserved for the *language-specific* parts of innate linguistic knowledge.

[5] Particularly striking for our purposes is the fact that (aspects of) the mature state of an animal communication system, the "waggle dance" of foraging bees, can be shown to develop in spite of the complete removal of relevant experience from juveniles' learning environments (Dyer & Dickinson 1994).

[6] Since this information must be brought to bear on PLD by language acquirers (LA-ers) reflexively, without conscious intervention, it follows that it will give rise to common features across attested languages ('language universals').

At a different level of abstraction, we might ask how it is that a LA-er comes to be attentive to and able to distinguish *words* within continuous strings. What's more, even if we decided to ignore this problem (without reason), another acquisitional conundrum would spring up immediately: (modelling) experiments have demonstrated that in order to successfully discern words, LA-ers must have access to more information than is available to them in the sound patterns of their PLD. The relevant facts in this case are the "transitional probabilities" across syllable pairs: the probability of syllables [7] within words being stream-adjacent is higher than the probability of syllables separated by word boundaries being stream-adjacent. However, Yang (2004), Gambell & Yang (2006) and Shukla, White & Aslin (2011) demonstrate, *contra* Chomsky ([1955], 1975), that bare (models of) domain-general statistical learning mechanisms (see §§2.3 & 7.1) fail to identify word boundaries accurately from this information alone. Once a 'prior', offering additional information about the nature of PLD, is made available to the (modelled) learner, however, results improve dramatically. [8] The prior in this case is knowledge of the prosodic structure of words, including the constraint that they can each bear at most one primary stress (Chomsky & Halle 1968). This work gives a particularly powerful demonstration of the scope of underdetermination, because few problems in FLA seem like they could be more tractable than the identification of words. This point is well made by the fact that Chomsky ([1955], 1975) was adapting suggestions made by Zellig Harris (1951), who was actually proposing that analysis of transitional probabilities might permit identification of *morphemes* within words. As it is, morphemic discrimination is even less amenable to a purely statistical analysis, as morphemes typically lack the "beads-on-a-string property" necessary for such a treatment (Chomsky 2011: fn.4).

It is worth pausing to emphasize that the logic employed in these modelling experiments is entirely sound: the argument for rich innate linguistic knowledge in no way precludes domain-general principles of data analysis from involvement in FLA (see especially Yang (2008) on this point). In just the same way, the genetic input determining the development of physical organs will not cause a suspension in principles of stem-cell embryology, or in mitotic cell division. [9]

[7] An independent poverty of stimulus (PoS) question arises with respect to the existence of syllables of course.

[8] Gambell & Yang (2006) go on to suggest that once we consider the possibility that LA-ers make use of their established knowledge alongside linguistic constraints, to 'bootstrap' the recognition of novel words from their recognition of old ones, there may be no need for probabilistic tracking in word segmentation at all.

[9] Domain-*specific* learning mechanisms for language have been proposed, such as Chomsky's (1965) *Evaluation Metric* and 'default' grammatical settings (*i.a.* Bickerton (1981, 1988), Clark & Roberts 1993,

FLA is best conceived of as nothing more than "innately guided learning" by children "innately equipped to recognize when [they] should learn, what cues [they] should attend to, how to store the new information, and how to [use] it in the future" (Gould & Marler 1987, cited in Legate & Yang (2005: 31)).

Before moving on, we introduce a final, especially egregious example of stimulus underdetermination, to give a true idea of the challenge presented by the *logical problem of language acquisition* (Baker & McCarthy 1981). The question in this case is: how do LA-ers know that lexical items (LIs) are being produced non-reflexively, with non-referential semantics, and moreover, non-referential semantics of a highly specific kind? Let us illustrate this point.

In learning the meaning of the word *book*, a LA-er must come to realise that while it can mean something concrete, such as the physical object beside me, it can also be used in an abstract sense, to mean the text the object contains. In this abstract sense, the two copies of *The Minimalist Program* in the room are the "same book". What's more, the LA-er must learn that these two meanings of *book* can be employed simultaneously: in usages such as "if you pick up a black and yellow book, you will likely find something good in it", the co-indexed pronoun *it* takes on both the concrete and abstract meanings of the word (Chomsky 1996).

It is important to note that even when we use *book* (in its concrete sense) to 'refer' to a particular book – that is, to pick out a particular extra-mental entity known to the speaker – the meaning of the word is still based in its relationship with concepts internal to the mind. If I tear the cover off my copy of *The Minimalist Program*, then, despite the material change, I can talk about the "same book" (concretely). In fact, I could replace each page, cover, piece of binding, etc. of the book in a piece-meal fashion, and at each stage in the process, even the end, I would still have the "same book". [10] The meaning of the word clearly cannot reside in the world itself.

On the other hand, if I had brought together exactly the same materials used in the piece-meal repair, only this time in one (impossible) fell swoop, then I would no longer have the "same book": I would have replaced it. Now suppose that I take the original book and tear out each page, placing them in a loose pile on my desk. If my intention in doing this is to make the book easier to read, then the book still exists; whereas if that happens to be my scrap-paper pile, then

---

Chomsky 1993, Roeper 2000). Their existence would be no contradiction, but, as it is, their role has at the very least dwindled in recent theories of the *faculty for language* (FL) (see discussion in §7.3.1 below).

[10] This is just a version of the classic 'Ship of Theseus Paradox' discussed in Chomsky (2009b: 381-2).

the book has just been destroyed (cf. Chomsky 1995a: 22). What these thought-experiments reveal is that successful acquisition of the (concrete) meaning of the word *book* does not consist of determining which set of material objects it picks out in the world, but rather of determining which mind-internal concepts compose to form its meaning. In this case, concepts something like [continuity with origin] and [intended usage] seem to be involved.

The abstract sense of *book* also has a specific, articulated conceptual structure. If I say "this book has sold 14 million copies worldwide", holding up a copy of *The Iliad*, I am likely to be including: copies published in a number of different languages, copies where different choices have been made over what constitutes the 'correct' original text (a lively topic of debate in many Classics departments), and maybe even abridgements. None of this is relevant to the abstract individuation of the book though, which relies on concepts something like [creative authorship] and [overall effect], rather than any notion of [string-specificity]. Every word in human language can be analysed in the same way as *book* to reveal (quasi-primitive approximations of) its rich, mind-internal conceptual structure. [11]

Given the near infinity of possible meanings an LA-er could attribute to any given word, [12] it goes without saying that the acquisition of lexical semantics would be an impossible challenge if it weren't for a specific mind-internal conceptual system, and the knowledge that this alone is used to construct words' meanings. The task of acquiring meaning without this knowledge would be made all the more intractable by the fact that the production of words is non-reflexive, and so even when a word is being used referentially, the object of reference might very well not be evident to the LA-er. I may refer to "my car" without my car being in sight, [13] or I may choose not to refer to my car even though it is in sight (indeed, I may refer to my car as "my bicycle" if I wish) (cf. Chomsky 2010: 14). The poverty of the stimulus available to determine lexical semantics is practically luminous, even for relatively simple concepts.

As we have seen then, as soon as you look at all carefully at FLA, logical problems proliferate in all parts of the task. Only the misguided intuition which frequently obstructs self-reflective thought, that "things couldn't be any other way" (cf. Chomsky 1996), would prevent us from

[11] For a thorough discussion of these issues, see Chomsky (1995a).

[12] This *assumes* their attentiveness to the possibility of arbitrary sound-meaning pairings.

[13] Indeed, research by Landau & Gleitman (1985) shows that the blind have no difficulty with the task of acquiring the full meaning of words for visible objects.

treating language in the same way we do other biological phenomena, and positing a rich, innate FL. [14]

1.3 The "misguided intuition"

Despite its essentially truistic nature, anti-nativist critics have challenged the APS in three main ways (cf. Matthews 2006: 87). Some critics of linguistic nativism have sought to challenge the very existence of any poverty in the stimulus, either claiming that, (i), PLD are in fact not as impoverished as assumed, or that, (ii), the language acquired is no more complex than can be handled by domain-general methods of learning. Other critics accept (some version of) the logical problem of language acquisition, but suggest that, (iii), innate abstract grammatical knowledge does not in fact provide a solution, and so different constraints on acquisition must be identified.

In practice, these so-called *empiricist* approaches have mainly sought to address the classic version of the APS, formulated with respect to the acquisition of (systems of) formal rules. They have been largely mute with respect to other core features of adult linguistic competence and what motivates the various selective attentions of their modelled acquirers. [15] As it is, even when we consider the three challenges in isolation from these problems, they still appear specious.

1.3.1 Challenge (i)

Challenges (i) and (ii) have focused their discussion on one property of adult grammar, the *structure dependence* of formal rules, particularly as demonstrated by the inversion of the auxiliary verb in English question formation:

(1) (a) The student has finished.
(b) Has the student *e* finished?

[14] Another product of cognitive activity that seems particularly well-insulated in this regard is our sense of morality. However recent studies (including *i.a.* Hauser 2006 and Mikhail 2007) have begun to break down this fallacy and investigate the cognitive constraints on possible moral systems – a 'biomoral' research enterprise.

[15] Some accounts (e.g. Reali & Christiansen 2005) make these issues explicit. These implicit or explicit assumptions render empiricist accounts indistinguishable from nativist ones except in matter of degree. In fact, this is a position any model of FLA finds itself in *ab initio*, unless it includes an account for why acquirers might think they ought to be looking for PLD in the first place – the overarching problem of selective attention.

Needless to say, any number of rules could be taken to generate interrogative strings such as (1b) from their declarative counterparts ((1a) in this instance). For example:

(2) (a) front the first auxiliary;
(b) front the auxiliary most closely following the first noun;
(c) front the auxiliary immediately preceding the matrix lexical verb;
(d) ….
(adapted from Legate & Yang (2002: 152-3))

The correct rule for forming questions makes crucial reference to the structural organisation of the sentence, holding that the auxiliary immediately following the subject NP should be fronted. In cases such as (3), only this rule, and none of (2a-c), would be appropriate. [16]

(3) Has [the student who is sitting at the computer] *e* been writing?

In their attack on linguistic nativism, Pullum & Scholz (2002) make the simplifying (and unjustified) assumption that (2a) is the only alternative hypothesis LA-ers have to rule out in converging on the correct rule for question formation. Under this assumption, examples such as (3) constitute *critical evidence*, allowing LA-ers to arrive at the correct rule by ruling out (2a). [17] Pullum and Scholz proceed to demonstrate the existence of such 'critical' evidence in a corpus of PLD, before concluding that constraint-free acquisition follows.

In their response, Legate & Yang (2002) make the point that the existence of critical evidence in PLD does not amount to its sufficiency for successful acquisition. They estimate that only *c.*0.05% of the questions in an *average* child's PLD are critical with respect to the simplified binary choice Pullum and Scholz propose. Other binary acquisitions successfully realized at the same age (around three and a half years old) include the development of non-*pro*-drop in English and V2 in German. Assuming similar attentiveness to all binary choices, the frequency of critical evidence bearing on these two choices ought (a) to be similar, and (b) to give us an indication of the amount of data necessary before a binary choice can be made in FLA, i.e. provide a benchmar for 'sufficiency'. Legate & Yang (2002: 156) document that in both cases the frequency of

[16] In fact, rules such as (2a-c) are never mistakenly postulated by acquirers, as demonstrated by Crain & Nakayama (1987) in elicitation tasks.

[17] We must also assume awareness of a relationship between declarative and interrogative sentences, something encoded by substantial machinery in 'blind' nativist models of syntactic acquisition / competence.

critical evidence is *c.*1.2%, significantly greater than 0.05%. In fact, as they point out the figure 0.05% is "low enough to be considered … not readily available for every human child" (*ibid.*: 158).

Challenge (i) to the nativist position therefore seems entirely toothless: even when allowing a whole range of PoS questions to go unanswered (the existence of lexical categories, the identification of questions, etc.), when adopting inappropriate assumptions, and when restricting ourselves to a tiny amount of particularly amenable linguistic knowledge, [18] the contention remains highly problematic.

1.3.2 Challenge (ii)

Challenge (ii) has been more vigorously pursued in the literature. Especially well-known is the attempt by Rumelhart & McClelland (1986) to account for regular and irregular past tense morphology in English using solely a connectionist model of acquisition [19] – a model "which contains no explicit rules, only a set of neuron-style units which stand for trigrams of phonetic features of the stem, a set of units which stand for trigrams of phonetic features of the past form, and an array of connections between the two sets of units whose strengths are modified [by exposure to PLD]" (Pinker & Prince 1988: 73). Pinker & Prince (1988) demonstrate a whole range of problems which emerge from this bare probabilistic model. Most crucially, they point out that without any way for the model to know that you assign a particular *morpheme* to a *stem* as a matter of *rule*, it will fail to assign any past tense at all to newly-encountered verbs which do not share enough features with stems encountered in previous PLD, and so which have developed associations with particular realisations of the past tense. Without prior notions of morpheme, stem, and rule, acquisition would founder. [20] No inappropriate assumptions of the kind made by Pullum and Scholz are made by connectionist models, [21] but again, even when we beg a whole

[18] Pullum & Scholz (2002) offer the same argument under analogous assumptions with respect to three other non-primitive grammatical observations: the order of English auxiliaries, the absence of regular plurals in noun-noun compounds of English, and the correct anaphoric use of English *one*.

[19] Connectionism as a *model of learning* is not in contradiction with nativist approaches, which merely argue for constraints on hypothesis space (see Yang (2007) and discussion in §1.2).

[20] Similarly, Berent & Pinker (2008) demonstrate that the fact that English speakers favour noun-noun compounds containing irregular plurals over regular ones relies on the distinction between regular and irregular plurals being a real, rule-based one, not one hinging on phonological familiarity.

[21] Although there remains the more general question of why we should extend beliefs about the structure of the nervous system to the 'operational' level in the first place, seemingly a category error.

range of PoS questions (the existence of phonemes, semantic decomposition of words, etc.) and restrict ourselves to a tiny amount of particularly amenable linguistic knowledge, we find the challenge of the APS impossible to meet without (further) constraints on the hypothesis space.

As suggested in the previous section, recent years have also seen a series of attempts to motivate domain-general acquisitional mechanisms as a full account of the development of fronted auxiliary constructions. Berwick *et al.* (2011) summarise and dissect the problems with these studies. Space precludes me from presenting their reasoning in any useful way, but I summarise their findings briefly.

Berwick *et al.* (2011: 1223-4) find that models of the acquisition of auxiliary-inversion based on the substitutability of words within strings (*i.a.* Clark & Eyraud 2007, Clark, Eyraud & Habrard 2008) license illicit constructions, and that what knowledge they do allow LA-ers to develop is not equivalent to the principle of structure dependence. On the other hand, they find that approaches which rely on adjusting the probabilities associated with particular grammars in response to the operation of a particular evaluation metric over PLD (Perfors, Tenenbuam & Regier 2011) fail to account for the grammaticality of a range of examples (Berwick *et al.* 2011: 1229-31). Other models, based on learning by statistical reasoning over the frequency of word bigrams and trigrams (Reali & Christiansen 2005), are shown to rely on "the accidental homophony between pronouns and complementizers in English" (Berwick *et al.* 2011: 1234), producing no results otherwise. Finally, models of acquisition based on learning from connectionist networks equipped with information about parts of speech (also Reali & Christiansen 2005) fail to account for complex interrogatives such as (3), in which the matrix auxiliary differs from the embedded one.

If we accept Berwick *et al.*'s reasoning *in absentia* therefore, we find ourselves in a now familiar situation: despite admitting grammatical competence (to sample: knowledge of words, knowledge of whatever grammatical symbols are necessary for assessment by an evaluation metric, knowledge of parts of speech, etc.), and restricting ourselves to a tiny amount of particularly amenable linguistic knowledge, we find the challenge of the APS impossible to meet. Even if we decide (to my mind, without cause) to reject Berwick *et al.*'s arguments wholesale, the take home message would be basically the same: we would still be unable to account for the acquisition of the models' assumptions – all that allows them any purchase in the first place – and all the other facts about linguistic knowledge discussed in §1.2.

In conclusion, nothing in the literature suggests that challenge (ii) is anything of the kind.

1.3.3 Challenge (iii)

As a reminder, challenge (iii) is offered by critics who claim to accept the APS (or some version of it), but suggest that innate grammatical knowledge fails to resolve the issue; they propose alternative motivation for the constraints on hypothesis space which must exist. In practice, this refers to *functionalist* approaches to FLA.

1.3.3.1 Preferences in performance

Functionalist explanations of grammar all, to varying degrees, reject the *formalist* position that the structural properties of language [22] 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 *et seq.*) 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 the performance of "the basic function of language, to communicate information from the speaker to the hearer" (Hawkins 2007). [23] This is enshrined in his *Performance-Grammar Correspondence Hypothesis* (PGCH), [24] which holds that

[22] Functionalist accounts have actually reached a little wider than the domain of rule systems, in some cases taking frequency effects in language to suggest that we ought not to postulate certain abstract categories (e.g. Bybee's (2001) discussion of the phoneme, or Baayen's (2003) of the morpheme). These accounts raise sufficiently complicated issues to be beyond discussion here, but Yang (2008) robustly contests the logic of the conclusions they draw. It suffices for our purposes here to observe that these accounts: (a) are highly local in nature, so orthogonal to the discussion of the rest of the section, and (b) introduce new PoS questions: what would motivate the acquirer to repeatedly store phonetic realizations (taken to motivate frequency effects) in such a way to allow generalizations which resemble phonemic/morphemic behaviour?

[23] I am not going to discuss the extremely problematic nature of this assumption here, postponing discussion to §§2.2 & 6.1.1.1. One of the problems is noticed by Hawkins (2009: 70) himself, when he asks (but does not answer) the question "[t]o what extent do [performance] preferences result from parsing and comprehension, and to what extent are they production-driven?"

[24] This contrasts with the observation that the modules of performance may be modified in accordance with properties of the grammar over which they operate. As Hale & Reiss (1998: 73) point out, referring to work by *i.a.* Werker & Tees (1984), Werker & Lalonde (1988) and Jusczyk (1997), "the general conception is that the

"grammars have conventionalised syntactic structures in proportion to their degree of preference in performance" (Hawkins 2004: 3). [25]

One of the principles which Hawkins takes to motivate "preference in performance" is:

(4) *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.
(Hawkins 2004: 32)

Hawkins takes this principle to motivate variable word order effects in language, where particular orders of syntactically-equivalent [26] strings are preferred in performance:

(5) (a) [[The man $_{VP}$[waited $_{pp1}$[for his daughter] $_{pp2}$[in the late afternoon sun]]
1---------2---3-----4-----------5
(b) [[The man $_{VP}$[waited $_{pp1}$[in the late afternoon sun] $_{pp2}$[for his daughter]]]
1--------2---3---4------5--------6-------7
(Hawkins 2004: 32-3)

(5a) is overwhelmingly preferred to (5b) in performance because of the shorter distance between the furthest dependent heads, in accordance with *MiD*. The syntactic/semantic dependencies between the heads can be recognized on the basis of five words in (5a), as opposed to seven in (5b). This is not to say that (5b) is an impossible production – it is not – it is merely associated with a marked meaning, in this case perhaps emphasizing the reason the man is waiting.

---

sensitivity of the speech perception system is greatly 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." A particularly strong version of this 'Grammar-Performance Correspondence Hypothesis' would be identity between grammar and performance. We discuss in §6.1.4 the possibility that the parser and the grammar exploit the very same 'algorithms'. (Cf. Chomsky (2000: 90, fn.7).)

[25] Other versions of challenge (iii) exist, such as those seeking to motivate the acquisition of syntactic knowledge through its picture-like resemblance to external reality, the property of iconicity (e.g. Givón 1985). Given the abstract, formal nature of much syntactic knowledge, this is a position I don't pretend to understand.

[26] Abstracting away from any syntactic operations involved in generating different surface word orders.

Hawkins (2001: 7) takes *MiD* to be preferred in performance because it means that "fewer additional … decisions … need to be made simultaneously with phrase structure recognition" and "[l]ess [sic] demands are made on working memory". In (5a/b) for instance, *waited* is ambiguous as to whether it has benefactive semantics (and the relevant theta-grid) or not, and this core decision will be made at the same time as incorporation of the PP-complement headed by *for*. Greater demands on working memory (WM) are encountered in (5b) because the central ambiguity associated the V-head must be held over in WM for longer.

Hawkins (*ibid.*) takes the pressure to reduce the demands on WM to be "efficiency-based", rather than "capacity-constrained", in keeping with early resolution of central parsing decisions. A "capacity-constrained" account of the pressure to reduce WM cost recapitulates the insight of Chomsky (1961) and Miller & Chomsky (1963) that the relative unacceptability of centre-embedding constructions (among others) is "simply a consequence of the finiteness of memory" (Chomsky 1965: 14). This is the approach adopted by Gibson (1998) to the same locality preferences Hawkins seeks to explain: he suggests they can be explained purely by reference to the extra burden on WM [27] and pressure on that resource, without reference to efficiency in terms of early decision-making or a reduction in simultaneous decisions. In fact, this seems to be the only possible approach to the facts when you consider other maximally-local constructions which Hawkins (1994: 54*ff.*) and Gibson (1998: 51-3) seek to explain, such as *Heavy NP-Shift* ones:

(6) (a) [I $_{VP}$[waited $_{pp1}$[for the man with the ginger beard who had bought me a coffee earlier and really made my day] $_{pp2}$[in the late afternoon sun]]]

(b) [I $_{VP}$[waited $_{pp1}$[in the late afternoon sun] $_{pp2}$[for the man with the ginger beard who had bought me a coffee earlier and really made my day]]]

[27] Gibson has a nuanced approach to this cost. He distinguishes between the burden on working memory (WM) associated with the 'storage' of a head, and the WM burden associated with the operation the head is involved in when incorporating a later syntactic element into the phrase marker. Both of these burdens are taken to increase with the distance between the stored head and the head of the incorporated element. This is a reflex of Gibson's adoption of Just & Carpenter's (1992) proposal that both operations and recall from storage rely on the relevant items achieving a particular level of 'activation', a property which fades with further cognition. In this way, operations become a 'working memory' concern. Nothing hinges on this matter here however, although see §5.2 for further discussion of the evidence and implications of this model. I adopt the atheoretical definition that WM is "the limited-capacity system where information is stored for a short period of time [to be] manipulated during an ongoing cognitive activity" (Cecchetto & Papagno 2011: 441-2).

In this case, (6b) is overwhelmingly preferred to (6a), as predicted by *MiD*. However, in this case the PP-complement of V has been shifted to allow the adjunct PP to be closer to the head. In (6a), the core syntactic and semantic properties of the sentence are carried over for longer than they are in (6b), but (6b) significantly reduces the *maximum* distance over which an (optional and ambiguity-preserving) relationship of phrase structure must be established. This is indicative of concern for the limits of WM, rather than for the efficiency of the system in terms of decision-making. [28]

1.3.3.2 Fixed word order effects

In as far as Hawkins' principles of optimal performance are limited to the performance domain alone however, they have no bearing on the central PoS problem which concerns us here. In accordance with the PGCH, Hawkins suggests that his principles also motivate *fixed word properties* of languages, the principal example being head ordering effects: under the force of *MiD*, a VO-language such as English will tend to be prepositional as opposed to postpositional.

(7) (a) $_{VP}$[went $_{pp}$[to the movies]]
1------X
(b) $_{VP}$[went [the movies to]$_{pp}$]
1-----..…...-----X
(Hawkins 2009: 62)

Adjacency of V and P "minimizes the connected sequences of linguistic forms … in which relations of combination and/or dependency are processed"; a postposition would maximize the sequence of connected forms, X, with X equal to four in (7b), but two in (7a).

[28] Hawkins adopts two others central principles of optimal performance: *Minimize Forms* (*MiF*) and *Maximize On-line Processing* (*MaOP*). The former is essentially a recapitulation of Grice's (1967, (1989: 27)) uncontroversial third 'Maxim of Manner', "be brief (avoid prolixity)"; this increases speed and reduces complexity in performance by maximizing the use of inferential capacities. *MaOP* is taken to "maximize the set of properties that are assignable to each item X as X is processed" (Hawkins 2004: 51), and therefore the "speed and earliness of delivery of syntactic and semantic properties" (Hawkins 2007). Hawkins takes *MaOP* to motivate a range of left-right asymmetries, including the tendency of fillers to precede gaps. However, we will see when discussing this issue in §6.1.3 that this argument is unsuccessful.

Newmeyer (2001, 2003, 2004) clarifies the means through which preferences in performance could come to shape fixed word order properties of grammar; he points out that "the influence of the former on the latter [could be] played out in language use and acquisition and (therefore) language change" (Newmeyer 2001: 4). That is, preferences in performance (variable word order effects) will be reflected in PLD, which LA-ers will then interpret as reflecting grammatical rules (fixed word order effects); this is "holistic" functionalism in his terms.

The prediction of *MiD* for head ordering tendencies varies as a function of the property held constant across languages. The prediction of *MiD* for VO-languages is that they will be prepositional, but the opposite prediction is made for OV-languages.

(8) (a) $[_{pp}$[to the movies] went$]_{VP}$
1--2-----3------4
(b) [[the movies to]$_{pp}$ went]$_{VP}$
1------2
(Hawkins 2009: 62)

In OV-languages, adjacency of V and P results from postpositions, as shown in (8b). There is cross-linguistic support for the predicted relationship between verb-complement and adposition-complement ordering:

(9) Correlation between VO/OV and PrepN/NPost in the 981 languages showing dominant order for both, as surveyed in Haspelmath *et al.* (2013):

OV & NPost	472	(48.1%)
OV & PrepN	14	(1.4%)
VO & NPost	41	(4.2%)
VO & PrepN	454	(46.2%)

As can be seen, there is an overwhelming tendency toward *consistency* of V-Comp and Adposition-Comp ordering. As Hawkins points out, the same logic would explain the more general over-representation of harmonic word orders cross-linguistically. Although I adopt a 'capacity-constrained' explanation of the preference of the parser to minimize the length of dependency domains (and adopt a nativist position regarding linguistic competence), I believe that *holistic* functionalism is indeed the correct approach to explaining *cross-categorial harmony*

(Hawkins 1983): locality-oriented performance will result in sufficient ambiguity in PLD that LA-ers will reanalyse the Head-Comp ordering of one member of an inconsistent pair, resulting in grammatical harmony. We discuss this further in §7.4.2.

1.3.3.3 "Absolute universals"

When appropriately remotivated therefore, locality in performance not only has some value in explaining non-fixed word order properties of language, [29] but also (when combined with Newmeyer's proposal) the reflex of these properties in diachrony – certain elements of fixed word order typology. However, the problem with all this as a genuine alternative to innate linguistic knowledge is by now hopefully rather obvious. As well as motivating "variation-defining universals" (Hawkins 2009: 71), Hawkins must also motivate "absolute universals" (*ibid.*), as he acknowledges. His accounts of performance and the acquisition of harmonic word order rely on this (or innate grammar, which he disavows). He claims that "absolute universals can … be innately grounded as a result of processing constraints on grammars" (*ibid.*: 72). This is "atomistic" functionalism in the sense of Newmeyer (2001), assuming a *direct* linkage between functional motivations and grammar in 'real time', as part of the acquisitional process. "[W]ithin and beyond certain thresholds [of complexity and efficiency]" Hawkins (*ibid.*) anticipates LA-ers will develop "universals of the kind 'all languages have X' and 'no languages have X' respectively". However, Hawkins gives no indication of how this curious, introspective process actually works; it is completely mysterious how an understanding of complexity and efficiency in performance is expected to create a formal universal, and there is no evidence in the psycholinguistic literature of such a process taking place. Hawkins is mute on this crucial issue, merely observing that "[s]ystematic exploration of this idea is required in order to see to what extent absolute universals can be explained through processing" (*ibid.*). [30]

[29] Albeit somewhat trivially, so does *MiF*.

[30] Hawkins (2014)'s otherwise identical version of this paragraph substitutes this sentence with:

> *[s]ome interesting proposals have been made recently by Mobbs (2008) for incorporating the efficiency proposals of this book into Chomsky's (1995) Minimalist Program. The efficiency principles are now recast as general cognitive constraints on the 'internal computations' integrating linguistic and other mental entities, rather than as principles of performance as such, and are seen as having shaped cross-linguistic parameters in a way not unlike that proposed by Hawkins (2004). This proposal, which brings the two research traditions closer together, is discussed and critiqued [in pp. 62-72].*

In short, there is no actual account of learnability here. On the next page, Hawkins offers a second response to the problem, suggesting that *holistic* functionalism is in fact the motivation for absolute universals. He points out that "[t]he explanation for [the] cross-linguistic patterns that [he has] proposed … requires a theory of *diachrony [my emphasis]* that can translate the preferences of performance into fixed conventions of grammar". However, as we have seen, holistic functionalism is mediated by LA-ers interpreting their PLD using the grammatical machinery available to them, the same machinery Hawkins seeks to motivate. This renders redundant Hawkins' plea for further exploration of how "ease of processing drives the adaptation" by which "grammatical conventions" such as "categories and rule types" [31] "emerge" over time (*ibid.*). The reasoning here is circular, and so it is perhaps unsurprising that he does not pursue the matter himself, but again appeals to further research.

In summary, Hawkins doesn't seem clear where his real account of learnability is coming from, offering two suggestions without obviously being aware that they are different. The first account is empty, the second empty and impossible. Both require at least partial remotivation and ignore the majority of successfully acquired linguistic principles. Despite this, to the best of my knowledge, Hawkins' work is by far the most completely realised bid for explanatory adequacy in the challenge (iii) literature.

1.3.3.4 Why go outside?

We are now in a position to look at the reasons Hawkins (and others) feel the need to suggest an alternative to the nativist response to the PoS. 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" (Hawkins 2004: 11)), Hawkins restates the argument from Culicover (1999) that "negative evidence problems reduce to language-particular idiosyncrasies", and as such that "the whole relevance of [innate grammar] to learnability must be considered moot" (Hawkins 2004: *ibid.*). For Culicover (1999: 28), LA-ers are "attentive" and "conservative" in their treatment of arbitrary lexical variation in PLD: little

---

This is not my interpretation of what I wrote. To discuss Hawkins' claim here properly would require restating too much material, so I will let the reader read the relevant sections of Mobbs (2008) for themselves, if they choose to – or not, as it otherwise has no bearing on the above discussion. I simply note that Hawkins' 'promissory note' of future "systematic exploration" remains unfulfilled, and that functionalist and Minimalist research remain entirely distinct, although very much compatible when properly construed.

[31] As discussed, he makes no suggestions regarding how to overcome the PoS for other linguistic principles.

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 nativist picture of acquisition, which comprises a greatly simplified process of abstraction from (predominant) underlying orderliness, in accordance with general principles of the FL. LA-ers are (at least largely) inattentive 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 is meaningful however. Most importantly, it neglects the fact (hinted at parenthetically) that non-principled learning is entirely compatible with the nativist position. The nativist claim is not that there may be no arbitrary lexical exceptions to intra-linguistic, FL-based generalisations (a language system): [32] the claim is rather that the APS means that there must be prompted learning and tight constraints on the hypothesis space, resulting in the *finite* acquisition of a language system which will generate output with the underlying orderliness of FL principles. [33] While it may well be that Culicover is able to present some telling examples of precisely such "language-particular idiosyncrasies", there is more subtlety to the nativist claim than he acknowledges.

As it is, many of the alleged cases of arbitrary lexical variation appear to be far less idiosyncratic than claimed. Culicover's (1999: 82) expresses his ideas most forcefully with respect to six English prepositions and their putative "odd" behaviour – contrasted with the 'typical' behaviour of *to* in Table 1 (adapted from Coppock (2007: 3)):

[32] Indeed, given the pervasive role of attentiveness and conservatism in the acquisition of other aspects of language, e.g. form-meaning correlations, it would be a surprise if there were not.

[33] The issue of permissible variation at the systemic level will be discussed at greater length in §3.2 below; we will see that it bears a close resemblance to arbitrary lexical variation.

Table 1: Behaviour of Culicover's 'odd' prepositions

Preposition	Precede NP	Piedpipe (prec.)	Follow	Piedpipe (follow)	Strand
to	yes	yes	no	n/a	yes
notwithstanding	yes	yes	yes	no	no
ago	no	n/a	yes	yes	no
out	yes	no	no	n/a	yes
off	yes	no	no	n/a	yes
during	yes	yes	no	n/a	no
since	yes	yes	no	n/a	no

Like Coppock (2007), I believe that closer inspection reveals as specious Culicover's claim that these examples demonstrate unpredictable behaviour. Let us consider each preposition in turn.

The 'unusual' properties of *notwithstanding*, that it may follow its NP and never strands, are explicable in a similar way to the one Coppock (2007: 5-6) suggests. *Notwithstanding* constitutes two different LIs. *Notwithstanding*$_1$ is an intransitive participle heading a sentence adjunct, taking an NP as its subject. Its property of coming after a NP follows straightforwardly from the fact that specifiers precede their heads; its failure to license piedpiping from this position is a simple reflex of the *Adjunct Island Condition* (*AIC*) (Huang 1982, Chomsky 1981), which prohibits extraction from an adjunct. When it precedes NP, *notwithstanding* is a different LI entirely – *notwithstanding*$_2$ – a preposition [34] heading a sentence adjunct; its failure to strand again follows straightforwardly from the *AIC*. The 'exceptional' behaviour of *notwithstanding* is in fact the reflex of analysing two LIs as one.

The supposedly unusual properties of *ago* dissolve in a similar fashion. *Ago* is well analysed as an intransitive preposition, as suggested by Fillmore (2002) and Coppock (2007: 5), just like complement-less *before* and *after*, and approximately forty other prepositions listed in Huddleston & Pullum (2002). In this case the NP will be *ago*'s specifier, explaining its precedence. *Ago*'s failure to strand is a reflex of the fact that it always heads a sentence-level adjunct.

The supposedly unusual property of *out* and *off* is that they fail to piedpipe a following NP:

[34] Or perhaps, without significance for the rest of the analysis, as a transitive participle of a second verb with the 'reverse' lexical semantics of *nothwithstanding*$_1$.

(10) (a) Which window did it fall out?
(b) ??Out which window did it fall?

(11) (a) Which chair did he fall off?
(b) ??Off which chair did he fall?

This property seems to be a straightforward reflection of the fact that *out* and *off* are particles of complex verbs in such expressions, rather than prepositions. Particles 'strand' their NP complements only in the sense that they don't have NP complements. This is an impression reinforced by the observation that when *out* and *off* fail to strand, i.e. when they adopt a property of (non-adjunct-heading) prepositions, they require/are much improved by the support of *of* (or */ə/*), a full predicate in its own right. (10b) and (11b) become unambiguously grammatical as (12a) and (12b) respectively:

(12) (a) Out of which window did it fall?
(b) Off /ə/ which chair did he fall?

Again, allegedly exceptional prepositions seem to be nothing of the sort.

Finally, we turn to *since* and *during*, only exceptional in their inability to strand. Once again this is readily explicable by the fact that they only head adjunct PPs, [35] never complement PPs, and so the *AIC* precludes extraction of their NPs.

None of Culicover's showcase examples of 'rampant lexical exceptionalism' seem to be exceptional at all. We might ask what other examples have been put forward to sustain this position? Coppock (2007: 10-11) summarizes the headliners, each of which we discuss briefly.

"[T]he celebrity" (*ibid.*: 10) among these is the dative alternation, which some ditransitive verbs allow, while others do not.

(13) (a) I *gave* / *donated* the book to the student.
(b) I *gave* / **donated* the student the book.

[35] Observed for *during* by Hornstein & Weinberg (1981).

Close analysis of the alternation has revealed that the two constructions do not in fact mean the same thing, and are associated with different argument structure. Oehrle (1976) points out that the *double object construction* (*DOC*) seen in (13b), but not the *to-dative* of (13a), may have a causative interpretation. The *gave* version of (13b) emphasizes my role in the student getting the book relative to (13a). This is more clearly evidenced by (14a) and (14b), where the *to*-dative is odd because it cannot convey the sensible interpretation, that the article was directly responsible for my headache.

(14) (a) ??The article gave a headache to me
(b) The article gave me a headache.
(Miyagawa & Tsujioka 2004: 2)

The failure of certain verbs to allow the dative alternation is therefore a property of their lexical semantics. *Donate* simply does not have the semantics required to support a causative interpretation. [36, 37] The supposedly random distribution of the dative alternation is in fact principled.

It has also been suggested that arbitrary lexical variation is evident in the distribution of the causative alternation (Bowerman (1988: 84), cited in Coppock (2007: 11)):

(15) (a) That huge bite made her *choke/cough.*
(b) That huge bite *choked/*coughed* her.

Levin & Rappaport-Hovav (1995: 91) explain that when "some property inherent to the argument of the verb is 'responsible' for bringing about an eventuality", then introducing an external cause of the eventuality results in "compet[ition] for the single external argument slot [available] with the verb's own argument" (*ibid.*: 144). This prevents verbs associated with internally-caused eventualities, such as *cough* [38] from entering into lexical causative constructions such as (15b), in

[36] See Grimshaw (2005) (and references within) for discussion of the role of metrical considerations alongside semantic ones in explaining the distribution of the dative alternation.

[37] A second difference between double object constructions (*DOC*-s) and *to*-datives is associated with the nature of the goal phrase. In a *DOC*, the goal must be interpreted as the possessor of the theme, whereas in the *to*-dative the goal is interpreted as a locative (Bresnan 1982, Mazurkewich & White 1984). We will return in §4.1.7 to discuss the broader significance of ditransitives for syntactic theory.

[38] I.e. an unergative.

which two causes are part of the same theta-grid. This is not a problem for verbs with semantics of *external causation*, [39] such as *choke*, because their own arguments do not compete for the external argument slot. Again, 'arbitrary lexical variation' reduces to semantics.

Coppock (2007: 11) mentions two further examples of alleged arbitrary lexical variation from Baker (1979), which she does not feel able to explain. These are the distribution of raising constructions across adjectives:

(16) (a) It is likely / probable that Robin will succeed.
(b) Robin is likely / *probable to succeed.
(Coppock 2007: 11)

And the distribution of adjectival complements across verbs:

(17) (a) Michelle seems / happens to be happy.
(b) Michelle seems / *happens happy.

(16) becomes unproblematic once you abandon the unjustified assumption that the *likely* of (16b) is the same LI as the *likely* of (16a). In (16a), *likely* has essentially the same meaning as *probable*, referring simply to the degree of certainty of an *event*. In (16b) however, *likely* means that *Robin* has certain properties which make an event probable. [40] Although clearly etymologically related, these are not the same word. (17) becomes unproblematic when you note that *seems* has evidential semantics whereas *happens* does not. ('To me' can be added to the *seems* version of (17a) and (17b), but to neither of the *happens* versions.) Evidential semantics is consistent with a purely adjectival complement, whereas the 'event-related' semantics of *happens* is not. [41]

Not a single one of the examples of arbitrary lexical variation discussed by Coppock has survived closer analysis, and although there may be other examples to discuss, it does at least seem that "language-particular idiosyncrasies" may be significantly less pervasive than Culicover, Hawkins, and others allege. What's more, it is independently clear that on a proper understanding of the nativist position, the existence of such idiosyncrasies could not be understood as crucially

[39] I.e. unaccusatives.

[40] *Tough*-movement alternation ("John is impossible / *improbable to please") would seem amenable to a similar explanation.

[41] An observation due to Ian Roberts.

undermining an account of learnability based on abstract grammatical competence in the first place. Bearing in mind that the challenge (iii) literature offers no feasible alternative solution to the learnability problem it acknowledges, there doesn't seem much reason to abandon the standard assumptions of biological research.

1.4 'Language-external' considerations

We have now established the *logical* necessity of a biolinguistic / nativist account of FLA (and pursued a thorough review of attempts to circumvent it, none of which gets off the ground.) There is also, however, robust 'language-external' evidence for a coherent object of biolinguistic enquiry. I very briefly present the two most unambiguous sources of this here.

1.4.1 Studies of sign language

Extensive research has been conducted into natural languages rendered in signed gesture rather than speech (*i.a.* Petitto (1987, 2005), Petitto & Marentette 1991, Goldin-Meadow 2003). The findings of these studies are very clear: the time course, characteristic stages, and structural results of sign language acquisition are all very similar to those of spoken language acquisition. The linguistic use of gesture is sharply distinguished in this way from the broader gestural systems of signers and non-signers (even though gestural forms themselves are sometimes held in common across both systems).

One particularly striking finding of these studies is that children acquiring sign language go through the same 'babbling' phase of development as children acquiring spoken language. What's more, the babbling of signing children takes on the same syllabic structure as spoken babbling. The 'sign-syllable' consists of the rhythmic closing and opening, or movement and stasis, of the arms and hands. These are the same alternations found in the articulation of spoken language, with its close/open and stop/start alternations of the jaw and tongue in making consonant-vowel syllabic structure. Clearly something fundamentally *linguistic* underlies this patterning, not related to the properties of any particular articulatory-perceptual apparatus. [42]

[42] At least not directly related. It may well be that that syllabic organisation of externalised linguistic form is an evolutionary reflex of the nature of the articulatory-perceptual systems employed in speech: **con**sonants and vowels must be closely related to be able to be articulated and perceived (cf. Redford 1999). That is, syllabic structure reflects an *interface condition* (IC) of the type discussed in §2.2.

Similar identity with structural (and semantic) properties of spoken language is found at all levels of signed language, except where the differences in modality directly preclude it (see Swisher (1988) for interesting discussion.)

All this of course is unexpected under an empiricist account of FLA. The characteristic trajectory of development cannot be explained by the way in which PLD is presented, or the maturation of articulatory-perceptual faculties, since these differ across modalities, pointing instead to language-specific explanation. Similarly, if learning environments truly are the sole source of linguistic competence, then it is beyond improbable that signing acquirers would converge so unerringly on the same structural and semantic results as non-signing acquirers, as their environments differ dramatically.

Any suggestion that functionalist motivations might explain consistencies across modalities seems even more problematic. The challenge of communicating by gesture is very different to the challenge of communicating with speech, and so, again, it seems highly unlikely that identical solutions might arise. The existence of the syllable in sign language is particularly striking in this regard, because it is intuitively related to the 'jaw cycle', discussed in fn.42, a property of the *speech* apparatus. What signed language data reveal is that articulatory-perceptual concerns can only be considered an evolutionary motivation, not one active in 'real-time', atomistic functionalism; the rigid *competence-performance* distinction (Chomsky 1965) must be maintained. In fact, Petitto and colleagues (Petitto *et al.* 2000, Petitto *et al.* 2012) have conducted imaging studies of the brain in which they identify neurological correlates of this abstract competence, "tissue in the human brain dedicated to *function* of human language *structure*" (Petitto 2005: 97).

1.4.2 Prodigal knowledge

Many studies demonstrate that acquirers develop sophisticated appreciation of linguistic principles by an extremely young age. These findings border on complete inconsistency with accounts of acquisition not reliant on rich innate knowledge. They pose a particularly strong version of the APS, nevertheless met by LA-ers. Either empiricist models of acquisition must be inconceivably powerful, or we must accept the unthreatening conclusion that children really do start life knowing a great deal about the nature of language.

For example, Peter Jusczyk and his colleagues carried out a series of experiments investigating the varying attention of infants to different aural cues, based on assessment of the infants' physical orientation to the cues' sources. They found that by the age of four months infants are able to decipher and understand their own names (Mandel, Jusczyk & Pisoni 1995), and by the age of six months, very common words such as "mommy" and "daddy" (Tincoff & Juscyzk 1999). Full word segmentation seems to begin as early as seven and a half months old (Jusczyk & Aslin 1995). By roughly the same age infants also seem: to be sensitive to the perceptual properties of clauses (Jusczyk *et al.* 1992); to be able to locate the primary stress on the metrical structure of words; and to have identified the dominant stress pattern of their languages (Jusczyk, Cutler & Redanz 1993). [43]

Similarly prodigal knowledge is manifest in the syntactic/semantic domain. Using similar methods to Jusczyk and his colleagues, Hirsh-Pasek & Golinkoff (1996) uncovered that:

(18) (a) infants between thirteen and fifteen months know that words presented in strings are part of larger constituents;
(b) infants between sixteen and nineteen months appreciate that different word order may indicate a different underlying syntactic structure;
(c) twenty-eight month old children are able to appreciate at least some aspects of the relationship between a verb's syntactic frame and its semantic one (argument structure).
(adapted from Newmeyer 2003: 690)

Crain & McKee (1985) presented young children with scenarios and then elicited their truth-value judgments with respect to subsequent explanations of what (the experimenter claimed to think) had taken place. Using this technique they showed that even at the age of two children are sensitive the principles governing anaphoric binding (Chomsky 1981, 1986). The children responded appropriately to scenario explanations of the kind:

(19) (a) The Ninja Turtle danced while he ate pizza.
(b) While he danced, the Ninja Turtle ate pizza.
(c) His arch-rival danced while the Ninja Turtle ate pizza.
(d) He danced while the Ninja Turtle ate pizza.

[43] Indeed, many studies claim to have identified rich phonological competence in neonates. See *i.a.* Vouloumanos & Werker (2007) and Gervain *et al.* (2008), and references therein.

(reproduced from Newmeyer (2003: 690))

They correctly recognized that the pronoun and the "Ninja Turtle" can be co-referential in (19a-c), but not in (19d). [44]

Studies of the structure, acquisitional trajectory, and neurology of sign languages and studies of infant language development both offer robust, 'external' evidence that FLA "is a snap" (Crain & Pietroski 2002). [45] That is, FLA takes place as the reflexive action of an innate module of cognition, imposing a particular interpretation of experience.

1.5 Departing in peace

It must be admitted that this chapter has been somewhat negative, focusing on the inherent failings of partial (pseudo-)empiricist approaches to FLA, and on their inconsistency with external evidence. The purpose of this, however, is to demonstrate the value of much constructive investigation of the FL. The following chapters show that a genuine commitment to approaching linguistics with the ambitions and broader perspective of a biological scientist leads to a range of interesting conclusions about language's structure, variation, acquisition, use, evolution, and definitive role in human cognition. The rest of this dissertation will be as optimistic as this chapter was pessimistic.

[44] This would be a breach of *Principle C*: R-expressions cannot be bound.

[45] Other sources of persuasive 'language-external' evidence are discussed in *i.a.* Fromkin (1999) and Smith (1999). Particularly persuasive is the asymmetry of abilities in linguistic 'savants' (see Smith (1999: 24)), where both intellectual *and* communicative competence dissociate from linguistic abilities, posing problems for both bare and functionally-motivated empiricist accounts of FLA.

2. The Minimalist Proposals and the Minimalist Program

As discussed above, while the term 'biolinguistics' is forty years old, and the enterprise itself twenty years older, [46] it is only in recent years that a fully explicit appreciation of the field has emerged. With improved awareness of the (bio-)scientific nature of nativist linguistics, various appropriate discussions regarding the proper pursuit of linguistic research and the design of the FL have arisen. The instigation of these discussions has been facilitated by standard features of scientific progress: improvement in the analytic value of linguistic accounts and the corresponding disappearance of theory-internal obstacles to particular lines of thought. A collection of related (but logically independent) proposals can be discerned in the biolinguistic literature of the past twenty-five or so years, particularly in the works of Noam Chomsky. We call these here *the Minimalist proposals*, and together they comprise the concerns of *linguistic Minimalism*. The subset of these proposals which is neutral with respect to the design or origins of the FL, and solely defines a meaningful approach to linguistic research, constitutes *the Minimalist Program for Linguistic Theory* (MP) (Chomsky 1993).

The Minimalist proposals emerged and coalesced as part of the organic progress of a field, without too much in the way of ongoing reflection. While all the proposals are presented and argued for explicitly at some point in the literature, this is often as part of a technical analysis which they relate to, or without reference to their relationship with the other Minimalist proposals. As such, it is not an entirely straightforward task to appreciate the structure of linguistic Minimalism, or the (generally uncontroversial) reasoning behind it, from reading the primary literature. I believe a fully explicit presentation of these proposals and their rationale is warranted if their utility and insight is to be realised. This is the overarching task of the following two chapters; I will make some independent observations and comments in the process.

I will not attempt to give an intellectual history of linguistic Minimalism, making only such observations as seem pertinent. The order in which the Minimalist proposals are presented reflects purely expository concerns: to the extent that one proposal implies another, the premise

[46] This is not to discount the contribution of the precursors to nativist linguistics, including scholars of the "first cognitive revolution" and those significantly preceding it; Chomsky acknowledges their input throughout his work (see in particular Chomsky (1966)). This is merely the observation that Chomsky was the first to apply these ideas to language systematically and in an entirely *naturalistic* way.

precedes the entailment. As is to be expected, this does approximate the chronology of the Minimalist literature. However it is not identical with it: insights often outstrip reasoning, let alone evidence, and sometimes they present themselves in slightly disguised ways – somewhat apparent in the history of the biolinguistic 'insight' itself.

This chapter covers the three Minimalist proposals which comprise the MP. Proper clarification of the second and third Minimalist proposals requires detailing particular notions of optimality; in doing so, I present an explicit framework for the proper pursuit of (a major element of) the MP, which is referred to throughout much of the rest of the dissertation. I discuss the final two Minimalist proposals in the following chapter: these proposals make specific claims about the design and origins of the FL and stand apart from, though in close association with, the MP.

2.1 The first Minimalist proposal: methodological minimalism [47]

The first, and least controversial Minimalist proposal is purely methodological: it bears exclusively on the standards to be reached in our theorizing about the design of the FL.

After over a decade's work on generative grammar within the *Government & Binding* (GB) framework (Chomsky 1981), Chomsky & Lasnik (1993) performed a noted 'stock-take' of the grammatical machinery which had been proposed. This revealed that GB theory had developed "five relatively independent generative systems, each of them essentially a cycle … operating separately":

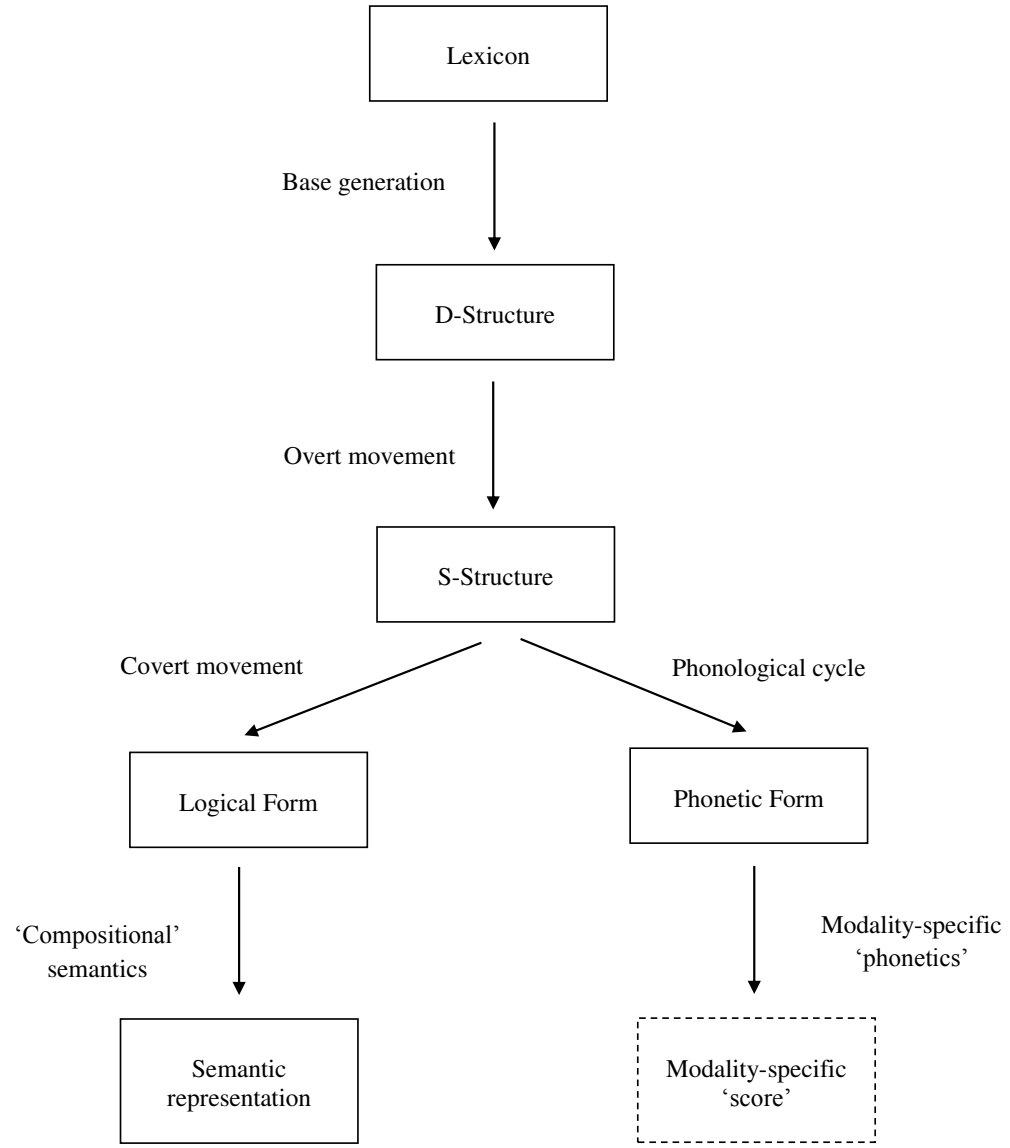
(i) "one that form[s] D[eep]-structure";
(ii) a "transformational cycle, ... mapping D-structure to S[urface]-structure";
(iii) a "covert transformational cycle … mapping S-structure to L[ogical] F[orm]";
(iv) "something mapping LF over to the semantic interface, a compositional semantics"; and

[47] A note on usage: I use **m**inimalism to refer to "the practice of using the minimum means necessary to achieve a desired result" (*OED*, 3rd edition), whereas the term **M**inimalism is reserved for the particular line of linguistic thought discussed here.

(v) "something mapping S-structure to P[honetic] F[orm]."
(Chomsky 2004a: 148). [48]

These cycles and the levels of linguistic representation formed have been commonly presented as part of an *Extended Standard Theory* (EST) (Chomsky 1965 *et seq*.) model (or *Y-model*) of the FL:

Figure 1: (Adapted) EST / Y-model of the Language Faculty:



[48] The system obviously missing from this classification is the mapping from PF to the 'score' for externalization (the 'phonetic' system) – omitted because it is modality-contingent and fundamentally extra-linguistic.

The syntactic cycles of the Y-model were complemented by a range of grammatical principles constraining possible D- and S-structures, including: the *Theta Criterion*, the *Projection Principle*, *X-Bar Theory*, etc. at DS; the *Visibility Condition*, the *Projection Principle*, *Binding Theory*, *X-Bar Theory*, etc. at SS; and the *Theta Criterion*, the *Visibility Condition*, *Binding Theory*, etc. at LF. These principles existed alongside independent conditions on displacement – the *Trace theory of movement*, *Relativized Minimality*, etc. – which constrained over-generation by a generalized transformational principle (*Move-α*) mapping between DS and SS.

The great proliferation of levels and principles postulated as part of syntactic competence naturally led to concerns about redundancy in the theory. Early Minimalist work (*i.a.* Chomsky & Lasnik 1993, Chomsky 1993) therefore began the task of "submit[ting] every assumption in that picture [the GB theory of syntactic competence] … to a critique which asked: can we really give a principled explanation for this, and if we can't, is there a way to get rid of it?" (Chomsky 2004a: 150-1).

For instance, it became clear on closer examination of the motivation for DS and SS that the evidence did not in fact necessitate two syntax-internal systems. There is no need to present these findings in full here (see instead Chomsky (1993: 19*ff.*) or Hornstein, Nunes & Grohmann. (2005: 24*ff.*)), however, in order to grasp the logic of the first Minimalist proposal it is important to follow a sample of the reasoning, taken from Chomsky & Lasnik (1993: 110-24) (here following the summary in Hornstein, Nunes & Grohmann (2005: 26-30)). [49]

The Visibility Condition holds that "[a] DP's [thematic] role is visible at LF only if it is Case-marked" (*ibid.*: 27); null operators / their chains (which have DP syntax) must therefore be assigned Case (despite being phonetically null). In passive constructions such as (20b), the verb *imagined* is considered to have been stripped of its (Accusative) Case-assigning capacity; similarly, the embedded *Infl*-element and the head of the infinitival VP, both being non-finite, are considered unable to assign (Nominative) Case. It is only once the verb's internal argument has been raised to the specifier position of a finite *Infl*-element that it can receive (Nominative) Case. In (20b), the expletive *it* fills this position, so the null operator receives no Case. In the active (20a), the null operator is assigned (Accusative) Case in its base-generated position.

[49] We do not challenge the GB machinery referred to as part of the explanation, returning to discuss it in Ch.4.

(20) (a) Columbus sailed the ocean [OP_i that people imagined t_i to be endless]

(b) *Columbus sailed the ocean [OP_i that it [$_{I'}$ was+*Infl* [$_{VP}$ imagined t_i to be endless]]]

A Visibility-based approach therefore analyzes (20b) as being in breach of the Theta Criterion, which holds that each argument must be associated with a theta-role, and that each theta-role must be associated with an argument. In (20b), the external argument of the embedded infinitive would seem at LF to have no theta role.

It will be noted from this discussion that Case-marking in GB theory must be assessed *after* displacement, that is, not at DS. Furthermore, the clear phonological implications of Case insist that it cannot be assigned at LF itself, i.e. after the output of *narrow syntax* (NS) has been sent to PF. There is therefore seemingly a very strong argument for a second NS-internal level of representation, SS.

As Chomsky (1993: 29) points out however, this argument "collapses" under a different, but equally adequate set of assumptions about the technical implementation of Case. As seen above, Case in GB is thought of as being *assigned* to NPs/DPs. Case-assigners, inserted in DS, are inherently specified for the relevant Case-feature, whereas Case-assignees are not. In a relationship of *government* [50] at SS, the assigner 'transfers' its Case-feature to the assignee, which then has a visible theta-role at LF and can be given an appropriate realization at PF.

The GB derivation of the simple passive construction, (21a), shows how the internal argument is inserted at DS as a bundle of (pro)nominal features, lacking any Case-feature, whereas the finite *Infl*-element bears a (Nominative) Case-feature. The DS, (21b), is then mapped to SS, (21c), in which the pronoun is governed by the finite *Infl* and adopts its Case-feature.

(21) (a) She was arrested.

(b) <u>DS</u>: [$_{IP}$ Δ was + $Infl_{CASE}$ [$_{VP}$ arrested [3P./SG./FEM. / [+pronominal, -anaphoric]]]

(c) <u>SS</u>: [$_{IP}$ [3P./SG./FEM. / [+pronominal, -anaphoric] / CASE]$_i$ was + *Infl* [$_{VP}$ arrested t_i]]

[50] A governs B *iff*: (i) A does not dominate B and every C that dominates A dominates B; and (ii) no barrier intervenes between A and B.

Under Chomsky's alternative proposal, both nominals and governing elements are inserted with their Case-features fully specified, and these features are then *checked* against each other in the course of the derivation. Still maintaining the DS/SS distinction, the derivation of (21a) under this approach is as follows:

(22) (a) DS: $[_{IP}\ \Delta$ was + $Infl_{NOM}\ [_{VP}$ arrested $she_{CASE}]]$

(b) SS: $[_{IP}\ [she_{CASE}]_i$ was + $Infl_{NOM}\ [_{VP}$ arrested $t_i]]$

She is fully specified at DS, (22a), bearing a Case-feature. For this feature to be licit in the derivation, it must at some stage be governed by another Case-bearing element. When the pronoun moves to Spec-IP, [51] its Case-feature is checked against the Case-feature on the *Infl*-element. At this point, the Case-marking on the pronoun can be confirmed [52] and the Visibility Condition becomes satisfied. As will be clear though, there is now no longer any reason why Case-marking cannot take place at LF itself, rather than before the output of NS has been sent to PF, i.e. at SS. PF receives the same feature-bundle either way, fully specified for the purposes of phonetic realization.

There is no clear empirical reason to favour one account of Case-marking over the other. In theoretical 'stand-offs' such as this, the first Minimalist proposal holds that we should adopt the more parsimonious explanation, that is, the one containing fewer ancillary claims. All other things being equal therefore, the checking solution ought to be favoured, as we need no longer appeal to the SS level of representation. Case-assignment is a prime example of an "[account] … offered in technical work [which] turn[s] out on investigation to be of roughly the order of complexity of what is to be explained, and involve assumptions that are not independently … well-grounded" (Chomsky [1995], 1998: 118).

As Boeckx (2006: 82 fn.14) points out, it is unsurprising to find that this "symmetry-seeking endeavour" was pursued in generative enquiry prior to linguistic Minimalism. Boeckx cites in this regard Chomsky's (1973) efforts toward reducing the existing principles of displacement to a single condition and Koster's (1978, 1987) attempts to unify movement and rules of construal. To

[51] Note that even in early Minimalism movement is not triggered by Case, but exclusively by categorical features, N/D-features here; Case instead is a 'free rider'. See §5.2.

[52] The Case-features of traditional Case-assigners also need to be checked (or in later theory, deleted) in the course of derivation, as demonstrated by the empirical evidence motivating the *Inverse Case Filter* (Bošković 1997) – see §5.2.

this we might add Rizzi's (1990) notion of *Relativized Minimality*, unifying existing locality conditions on head, A-. and A'-movement; and several others detailed by Chomsky (2005b: 7).

In practice, parsimony in pre-Minimalist theorizing was not motivated simply by sound reasoning, but also by an independent ontological concern – to relieve the "familiar tension between explanatory … and descriptive adequacy" (Crain & Pietroski 2002: 164). The APS insists that accounts of linguistic phenomena be contained within narrow limits, rather than proliferating into "complex and varied rule-systems" (Chomsky 2005b: 7). The conflation of rule systems postulated to meet descriptive adequacy into more general principles eases this tension somewhat.

The first Minimalist proposal abstracts away from any ulterior motives such as these, however, and simply reiterates the "straightforward methodological question[s]: can we make our theories better, can we eliminate redundancies, can we show that our principles are more general than we thought … ?" (Chomsky 2002: 96). These are questions that are arguably "always appropriate", as a matter of "good science" (*ibid.*: 97).

2.2 The second Minimalist proposal: the Strong Minimalist Thesis as methodology

Like the first Minimalist proposal, the second Minimalist proposal is fundamentally methodological. Unlike the first proposal however, it is couched in terms of a specific claim about the nature of language. The second Minimalist proposal points to the value of adopting a particular *ontological* position regarding the FL for guiding productive enquiry into its properties.

The *Strong Minimalist Thesis* (SMT) holds that "language is an optimal solution to legibility conditions" (Chomsky 2000: 97). In the first instance, therefore, the SMT notes that one of the few things that can be said of language as a matter of "virtual conceptual necessity" (Chomsky 1993: 2), i.e. in advance of serious enquiry, is that it allows us to connect form and meaning. The FL must produce structures which are "legible" to the cognitive apparatus dealing with form and meaning, the *Articulatory-Perceptual* (AP) [53] and *Conceptual-Intentional* (CI) systems respectively. These "legibility conditions" have also been referred to "bare output conditions" (Chomsky 1995c: 221), but are now more commonly referred to as the "interface conditions".

[53] Also referred to as the *Sensory-Motor* (SM) system to emphasize that the particular modality of externalization is arbitrary.

This latter term is the most appropriate to my mind, since it does not exclude other conditions language must satisfy in its interaction with other systems. These include the requirement that language in use should be parseable. Closely related is the requirement that underlying grammatical competence be of such a type that when it is specified as part of a particular language system, that language system be recoverable by FLA: if LA-ers cannot converge on (approximately) the same system as the previous generation, then communication with that generation (and their peers) will be impossible. This is crucially different from the burden of explanatory adequacy, which solely insists that a coherent language system be acquirable on the basis of impoverished PLD. In referring to the SMT hereafter, we assume the substitution of the term "interface conditions" (ICs) for "legibility conditions".

The SMT further specifies that the FL satisfies these ICs in "an optimal" [54] fashion. Two independent types of optimality [55] are referred to here. The first is what we will call *substantive* optimality. Language solves ICs more optimally in this regard the smaller the range of 'tools' it used in doing so. In §2.1, we saw how Case-checking diminished the need for SS; [56] all other things being equal, Case-checking is a more optimal solution to ICs in *substantive* terms than Case-assignment.

The second fashion of optimality referred to by the SMT is the abstract computational one. Chomsky commits to the classical *Computational Theory of Mind* (*i.a.* Turing 1950, Marr 1982, Gallistel 1990), which holds that the brain is an organ of computation, computing mental representations comprised of *symbolic* structures. Mental processes are rule-governed manipulations of mental representations, sensitive to their constituent structure (*computational operations* (COs)). [57] Symbols and symbolic structures are semantically-evaluable: they are causally connected to states of the world, carrying information about them. They are also the way this information is carried forward in time once *written* into memory (either by COs or *signals* from organs interfacing with the world).

[54] We discuss the use of the indefinite article in §2.4.

[55] Although perhaps not strictly correctly, I use "optimal(ity)" to refer both to state of being optimal and to the degree to which something approaches that state; the expression 'more optimal' *will* appear.

[56] Assuming for our purposes that Case-related arguments are the only ones bearing on the existence of *S(urface)-Structure* (SS).

[57] Substantive optimality can be thought of as the number of different types of symbol and computational operation (CO) the FL employs.

This information may be retrieved, or *read*, from symbolic memory for use in future COs, or for conversion into signals giving rise to responses to the world. [58] There is compelling behavioural evidence that the brain does indeed make use of symbolic manipulation and a *read-write* memory of this kind, but this is not the place to present it; see instead Gallistel (1990, 1998, 2000) and Gallistel & King (2009). An account of a linguistic phenomenon is therefore "principled" in terms of the SMT in so far as it reflects ICs, or observes principles of substantive or computational optimality, *viz.* economy in the range of computational symbols and operations necessary and 'efficiency' in the way they are used. [59, 60]

The computational nature of linguistic cognition, at least its grammatical component, is expressed by Chomsky's (1986a, 1990) distinction between *I-language* and *E-language*. An *I-language* (or grammar) is a "function-in-*intension*" (set of COs) *internal* to an *individual*, which "strongly generates" a set of "structural descriptions" (symbolic structures). There is no non-arbitrary definition of E-language, but it can be thought of here as whichever phonetic representations can be associated with structural descriptions of the I-language. Chomsky makes it clear that it is only I-language which is of concern in biolinguistic enquiry: notions of 'well-formedness', utterance, and extensionality "have no status in the theory of language" (Chomsky 1987: 181).

It is important at this stage to be clear about the remit of the SMT. The SMT holds of the abstract linguistic computation that creates outputs legible to the AP- and CI-interfaces. It therefore applies not only to NS's computation, but also to the mappings to the AP- and CI-interfaces. [61]

---

[58] This contrasts with *connectionist* theories of mental computation, such as Rumelhart & McClelland's (1986), discussed in §1.3.2. These theories hold that mental representations are nothing but particular patterns of activation in a network of 'nodes', with mental processes consisting of the spreading activation of these patterns. There is no *read-write* memory in such systems, merely the modifiability of connections in the patterns, and neither the nodes, nor 'constituents' of activation patterns are semantically-evaluable in their own right.

[59] Since the availability of particular symbols and COs is part of our computational abilities, the terms substantive optimality and computational optimality are no longer mutually exclusive (see fn.57). We might profitably rename the latter *operational* optimality, but we stick to the more established term for the sake of consistency with the literature. Nothing hinges on the matter so long as we are clear that 'computational optimality' refers to the *manner of use* of computational machinery.

[60] This contrasts rather starkly with Hawkins' (2004: 267) claim that there has been an "increasing disconnect between formalism and the search for explanatory principles".

[61] Chomsky himself has always been clear on this matter, stating that "mappings to [the sensory-motor interface] would be the 'best possible' way of satisfying the externalization conditions" (Chomsky 2007b: 14).

To the extent that it is pursued in analysis, the SMT constitutes a refusal on the part of the scientist to make pre-theoretic assumptions about the design or 'purpose' of language – a sound basis for enquiry. All the SMT takes as read is that language meets certain ICs and that its grammatical component is computational. The SMT does *not* amount to the claim that the use of language in externalization is merely epiphenomenal: rather it adopts an entirely neutral position about the 'function' of language. Far from excluding the conclusion that utility in externalization is a crucial factor in explaining the design of FL, the SMT in fact facilitates the substantiation of such a claim. Systematic deviation from cognitive optimality in linguistic phenomena will provide the necessary evidence that 'ulterior motives' enter into the design of (relevant areas of) the FL, and, subject to interpretation, will provide information about the nature of these forces. The SMT helps us move "beyond explanatory adequacy" (Chomsky 2004c) and ask *why* the FL is the way it is.

A teleologically-oriented research program, on the other hand, would be guided by a non-trivial heuristic, say, "language is the perfect means of communication". It is far from clear that such a program would make the full range of conclusions about the design of the FL available. In adopting a pre-theoretic position regarding the overall function of language, we all but preclude ourselves from discovering discrepancies between the current utility and the functional origins of different aspects of language – a potential (set of) conclusion(s).

Furthermore, as Berwick & Chomsky (2011: 25) observe, "the inference of a biological trait's 'purpose' or 'function' from its surface form is always rife with difficulties", citing Lewontin's (2001: 79) discussion of the many different roles played by bones – in structure, locomotion, storage, etc. This is particularly pertinent for the observable uses of language. [62] Functional assumptions are widely adopted: many linguists who adopt a formal, nativist theory of competence, only admitting functional explanation at the level of evolution (*i.a.* Pinker & Bloom 1990, Jackendoff (1992 *et seq.*), Pinker (1994 *et seq.*)) believe that "the language faculty evolved in the human lineage for the communication of complex propositions" (Pinker & Jackendoff 2005: 204). However, even restricting ourselves to its operation in the social space, language clearly has other uses than the *narrow* informative one: it may be used more freely in communication, with much information transmitted, but little of it in the content of what is said. This is particularly characteristic of *expressive/affective* speech acts, which express emotion, or

[62] The possibility of 'hidden' function is a particular problem in the cognitive domain, a matter pursued in §3.1.

attempt to evoke some feeling in others. The same is true of much *directive* language use, where the aim is to cause or prevent the particular action of another.

Social uses of language are far from the only ones: externalization (or at least 'internal externalization') has a range of private uses. Inner speech changes our perspective on our thoughts, putting us in the position of observer rather than producer, so affording us 'critical distance' to reflect on and clarify our ideas. Other important private uses of language are associated with 'rehearsing' (unproduced) phonetic form, which can be used to help with memory (try remembering a telephone number) or focussing attention (sometimes called 'self-talk', often visible in token articulation). Hawkins, Pinker, Jackendoff *et al*. focus on one of the uses of language, without offering justification. [63]

A further problem faced by such a teleological heuristic for linguistic research is the difficult question of what it means to be well-designed for externalization, which can scarcely be answered in advance of evidence. To the extent that the informational content of an utterance is relevant at all, one might reasonably ask, for instance, whether its efficient transmission is determined with respect to ease of production or ease of comprehension, or whether we might expect a 'trade-off' between the two, or something of the like. [64] Presumably the answer to this question depends, at least in part, on what is thought of as the pressure to externalize in the first place. We have seen a number of possible uses for externalization already, but the situation could be nuanced further, by considering to what end each use might be (or originally have been) put. With regard to the narrow informative use of language, a number of suggestions has been proposed, including pressure to: gossip, co-operate while hunting, exchange information about social status, etc., as summarised by Számado & Szathmary (2006) (see also Fitch, Hauser & Chomsky (FHC hereafter) (2005: 186)).

It goes without saying that systematic deviation from an uncertain basis can scarcely be determined, and a teleologically-oriented research program is therefore inappropriate. If a teleological heuristic were precisely formulated, deviation from it could just as easily be interpreted as evidence for the need to revise the heuristic as it could evidence of alternative design factors. This renders the guiding contention in danger of becoming an unfalsifiable hypothesis rather than a framework for rational enquiry. The methodological argument for

[63] A spoiler: the fourth Minimalist proposal suggests that externalization is a secondary concern in the design of the FL.

[64] This question is revisited in §6.1.1.1.

adopting the SMT as a rallying point is that it allows us to entertain all possible conclusions about the design of the FL, treating the matter as one of ongoing enquiry.

The above discussion does not present the SMT's ontological claim as entirely settled: this is far from the case. The notion of what it means to be substantively and computationally optimal will be discussed at greater length in §2.4 and elaborated throughout the dissertation, but it suffices here to note that common sense and established principles of algorithmic science enable us to substantiate cognitive optimality in an uncontentious way. However, the notion of good design with respect to ICs remains almost entirely an empirical issue. Little is known about the AP- and CI-interfaces in particular, and most of our current knowledge of them stems from study of the FL itself. [65] To pursue the MP, we are forced to make assumptions which potentially over-simplify the complexity of the ICs. In particular, we assume that the articulatory and perceptual interfaces access the same information, and the production and comprehension systems likewise (Chomsky 2000: 91). In taking the SMT as a heuristic for research, we therefore "face the daunting task of simultaneously setting the conditions of the problem and trying to satisfy them, with the conditions changing as we learn more about how we satisfy them" (Chomsky 2004b: 396), "not an unfamiliar feature of rational inquiry" (Chomsky 2005b: 10).

It will not have gone unremarked that this difficulty runs along the very much same lines as one raised above in objection to a teleologically-oriented research program. However, it must equally be noted that the SMT introduces only that uncertainty which is (virtually) conceptually necessary, retaining the most 'vanilla' heuristic possible. Another way of understanding the methodological argument for pursuing the SMT, therefore, is that it avoids unnecessary proliferation of unknown variables – again, a sound basis for scientific enquiry.

Similarly, it will have been noted that ontological minimalism (OM) as methodology and methodological minimalism (MM) will often conspire to the same end. Indeed, a further methodological argument for pursuit of the SMT is that it encodes MM as a first principle of research. For instance, once you accept a system is computational, MM alone argues for the pursuit of computational optimality in theories, a point apparently over-looked in the literature.

[65] Our only other knowledge about these systems comes from study of pre-linguistic humans and our nearest evolutionary relatives, via the *comparative method* discussed in §§2.3 & 3.1 below. While this is a fruitful research program, in many areas it is very difficult to pursue and it relies on not unproblematic assumptions. Its achievements are not to be minimized, merely kept in perspective given the obvious richness of our *articulatory-perceptual* (AP) and *conceptual-intentional* (CI) abilities.

This methodological argument is evident when comparing Chomsky's assessments of the value of the SMT and MM. OM as methodology relieves the researcher of "the temptation to offer a purported explanation for some phenomenon on the basis of assumptions that are of roughly the order of complexity of what is to be explained" (Chomsky 1995c: 233-4), and "sharpen[s] the question of whether we have a genuine explanation or restatement of a problem in other terms" (*ibid*.). This bears more than a passing resemblance to his point that MM helps us avoid "[accounts] … of roughly the order of complexity of what is to be explained, and involv[ing] assumptions that are not independently … well-grounded" (Chomsky [1995], 1998: 118). In short, pursuit of the SMT has the same "therapeutic" (Chomsky 1995c: 233) quality as MM.

Importantly though, the second Minimalist proposal goes much further than the first, offering more extensive and explicit guidance in respect of biolinguistic enquiry; the proposals are far from co-extensive, as has sometimes been suggested.

2.3 The third Minimalist proposal: the Strong Minimalist Thesis and the evo-devo hypothesis for language

The first two Minimalist proposals extend general principles of good scientific practice to biolinguistics; the third Minimalist proposal extends to it the broader concerns of biological science.

The third Minimalist proposal (*i.a.* HCF, Chomsky (2004c, 2005a, 2005b)) holds that the design of the FL ought to reflect the "virtual truism" (Chomsky 2004c: 105) that previous products of evolution constrain the possible outcomes of future evolution by natural selection. The exploration of such 'channels' in natural selection, and the associated possibility that they "may be even more deeply grounded in physical law" (Chomsky 1965: 59), has been pursued as part of *evolutionary-developmental* biology ('evo-devo'). [66] The third Minimalist proposal is that evo-devo concerns might be reflected in the growth and development of language.

Perhaps the classic example of lineage constraining evolution is afforded by the genetic 'toolkits' controlling organisms' morphological development. The best-studied of these is the *Hox* gene

[66] For recent examples of thought in *evolutionary-developmental* biology ('evo-devo'), see *i.a.* Goodman & Coughlin (2000), Carroll *et al.* (2001), Gould (2002: 1025*ff.*) and Carroll (2005). For discussion of the origins of evo-devo, see Maynard Smith *et al.* (1985), Chomsky (2005a, b) and Berwick & Chomsky (2011).

cluster of metazoans. These genes encode protein molecules which regulate the expression of genes involved in the embryonic processes specifying the identity of body segments and the orientation of the anterior-posterior axis. *Hox* genes (or more precisely the protein motifs they encode) are highly conserved across phyla and evolutionary time: the development of a fruit fly with a defective *Hox* gene can be rescued by the protein encoded by the homologous gene in chickens (Lutz *et al.* 1996).

Changes in the embryonic expression of 'toolkit' genes are associated with significant changes in body morphology, and this seems to be the basis of much of the morphological innovation in evolution. For instance, vertebrae vary along the length of the vertebrate spine, from cervical vertebrae in the neck, to the thoracic vertebrae down the back. The evolution of the backbone from, say, python to mouse is associated with a change in the boundary between cervical and thoracic vertebrae. In a python, the change-over takes place at the sixth vertebra, whereas in a mouse it takes place at the twelfth. The boundary between cervical and thoracic vertebrae correlates with the anterior boundary of expression of the *hoxc6* gene in embryonic development (Burke *et al.* (1995), as reported in Carroll *et al.* (2001: 131-3)). This transformation in body plan may be largely attributed, therefore, to a change in a higher gene in the cluster, controlling the spatial expression of the *hoxc6* in the embryo. This is a relatively small evolutionary event compared to the alternative: a piece-meal, adaptive account of how natural selection favoured (necessarily gradual) re-design of the anterior vertebrae. Radical evolutionary novelty, such as the mammalian spine, [67] may in fact be a product of comparatively minor changes in the organisation of otherwise conserved genetics.

The substantive claim of the SMT locates the study of grammar within evo-devo. In claiming that abstract linguistic computation optimally satisfies ICs, the SMT takes I-language to employ inherent principles of cognition, pre-existing the development of language, which it must be expected make optimal use of the brain's computational resources (Chomsky 2004c: 105). (That is, we make the self-evident assumption that a major pressure shaping the evolution of cognition was the reproductive/survival utility of being able to process information.) The strongest interpretation of the SMT, then, is that the innate grammar necessary for explanatory adequacy is exhausted by 'cost-free' principles of cognition and whatever is required to 'co-ordinate' them to meet ICs.

[67] Or the echinoderm water vascular system, or the tailless ascidian body plan, etc. (Carroll *et al*. 2001: 163-7).

As it is, certain core facts about language clearly cannot be explained in terms of optimality, and we are forced to propose further innate competence. Whatever is postulated must account for "the most elementary property of language … [:]that it is a system of discrete infinity consisting of hierarchically organized objects" (Chomsky 2008: 137). Language takes a finite number of elements and forms structured expressions composed of a discrete number of elements: "there are no 6.5 word sentences" (HCF: 1571). Furthermore, these structured expressions are potentially infinite in number – there is no "non-arbitrary upper bound to sentence length" (*ibid*.). Any sentence can be lengthened by making it the subordinate clause of a construction ascribing, say, belief: "it is my conviction that it is Jack's opinion that it is Mary's understanding that … that X."

And so innate grammar must at least include a CO mathematically equivalent to set-formation, which takes at least two existing *syntactic objects* (SOs) [68] and forms a new SO (i.e. a self-similar structure): this is the minimum required to "get recursion off the ground". This operation is commonly called *Merge*. We must also postulate a symbol which identifies the mental representations over which *Merge* operates, allowing new LIs to be incorporated into the phrase marker [69] and capturing the unbounded scope of *Merge*. This has been designated the *edge-feature* (EF) of a LI (Chomsky 2008: 139). The fact of embedding ("putting one phrase inside another of the same type or lower level, e.g., noun phrases in noun phrases, sentences in sentences, etc." (Everett 2005: 622)) shows that merged SOs may be either complex or lexical, and so the EF of LIs must be of a kind to permit "free *Merge* to the edge, indefinitely" (Chomsky 2008: 144).

On this view then, general cognitive principles of optimality entering into innate grammar are akin to conserved genetic machinery instantiating principles of spine development; whereas those aspects of innate grammar not attributable to principles of optimality (including at least *Merge*, an undeletable EF, and whatever is required to meet ICs) are akin to the minor genetic innovation altering the expression of the *hoxc6* gene. [70] The substantive claim of the SMT therefore suggests that focus on the rich adaptive history of (core facts of) language may in fact be misguided (*contra i.a.* Pinker & Bloom 1990, Jackendoff (1992 *et seq*.), Pinker (1994 *et seq.*), Pinker &

[68] We discuss the possibility of *n*-ary *Merge* with $n > 2$ in §4.1.7 below.

[69] Itself a *lexical item* (LI) at the 'base' of a derivation.

[70] I ignore (for the time being) the possibility that *Merge* or an EF may themselves be existing principles of cognition, co-opted from an extant computational module for use in the FL (see Hauser, Chomsky & Fitch (HCF hereafter) (2002: 1570), Fitch, Hauser & Chomsky (FHC hereafter) (2005: 201), Chomsky (2007b: 7)). In this case, their co-option for language would constitute the genetic innovation.

Jackendoff 2005, Jackendoff & Pinker 2005); grammar may be significantly more "evolvable" (Ridley 2004: 589) than supposed – a central insight of evo-devo thought.

In fact, the role of evo-devo thought in the development of Chomsky's work is an interesting piece of intellectual history. The development of the *Principles & Parameters* (P&P) framework for language structure and acquisition [71] was conceived under analogy with François Jacob and Jacques Monod's suggestions regarding the role of underlying developmental pathways in constraining the variety of possible organisms (discussed in Chomsky (1980: 67)). The principles and parameters (parameterized principles) thought to comprise innate grammar are analogous to the genetic machinery determining developmental pathways, conserved across phylogenetic space. The parametric variation of human I-languages compares to variation in (the expression of) these 'toolkit' genes across species / clades. Small changes in parametric choices can result in pronounced difference between the resultant I-languages, just as small changes in (the expression of) genes coding for a transcription factors explain (major elements of) the difference between a snake and a mouse.

The development of the P&P framework removed the "conceptual barrier to shifting the burden of explanation from … the genetic endowment [to] language-independent principles … [and] thereby providing some answers to the fundamental questions of biology of language, its nature and use, and perhaps even its evolution" (Chomsky 2005b: 9). This is because, under the P&P conception, acquisition is merely selection from among pre-determined parametric options; it no longer necessarily invokes a "rich and highly articulated" "format" for acquisition, provided by innate grammar to bring explanatory adequacy within "reach" (*ibid*.). Aside from resolving the aforementioned tension between explanatory and descriptive adequacy, this means that much of grammar is now 'free' to be non-dedicated, and so is amenable to explanation in terms of broader cognitive principles. An entirely *analogical* application of evo-devo insights in linguistics paved the way for their *direct* application to language in the third Minimalist proposal.

I have so far presented only a partial characterization of evo-devo, neglecting its more abstract proposal. As mentioned, work in this framework also explores the possibility that local canalization in natural selection "may be even more deeply grounded in physical *[or natural]* law". This alludes to the further "truism" that "there are the physio-chemical constraints of the

[71] The current formulation of this idea and its own place in linguistic Minimalism are discussed in §3.2.

world, necessities that delimit biological possibilities" (Berwick & Chomsky 2011: 22). [72] That is, there are practically infinite logically possible 'solutions' to a given selective pressure, [73] but given the challenges of establishing these solutions in organic matter, only a few will be "evolutionarily stable" outcomes. In the morphological domain, Berwick & Chomsky (*ibid.*) cite the failure of any known organism to meet selection pressure for locomotion by developing wheels (a highly energy-efficient solution) and attribute this to the physical difficulty of "providing nerve control and a blood supply to a rotating object".

It may be presumed that it is analogously difficult – "in some unknown way" (Chomsky 2002: 57) – to render computational abilities [74] in organic matter (in this case, neuron cells). [75] As such, brains instantiating principles of substantive and computational optimality may be the only evolutionarily stable (set of) solution(s) to the pressure of processing information. This fleshes out (no pun intended) the observation above that "principles of cognition … must be expected to make optimal use of the brain's capacity for computation". Suggestive work in this regard has been carried out on the structural architecture of nervous systems by Cherniak (1995) and Cherniak *et al*. (2004). This work demonstrates a cross-species [76] trend toward "component placement optimization" (Cherniak *et al*. 2004: 1081): functional areas of nervous systems seem to be positioned so as to maximize the adjacency of those areas which regularly relate to each other – "down to a best-in-a-billion optimality level" (*ibid.*). This "save wire" principle may plausibly be attributed to the difficulty of instantiating long-distance interconnections between modules in neuronal tissue.

---

[72] Chomsky (*i.a.* 2007b: 5-6) has regularly cited the emphasis placed on these factors in work by Alan Turing (1952) and D'Arcy Thompson ([1917], 1992).

[73] As Berwick & Chomsky (2011: 21) point out, the orthodox position in biology has been that organisms instantiate a "near infinitude of variety that ha[s] to be sorted out case by case" (quoting Gunther Stent from Carroll (2005: 24)). Note the very suggestive analogy with Joos's (1957: 96) often cited observation that "(l)anguages can differ from each other without limit and in unpredictable ways" – a line of thought to which the *Principles & Parameters* (P&P) framework responds. (See Thomas (2002) for discussion of this comment in historical context.)

[74] Loosely speaking, the types of symbol available, the contents of the COs available, an ability to perform COs ('processing power'), and an ability to carry information forward in time in an accessible manner ('working memory').

[75] Although we can say little more than this – see §2.4.

[76] Cherniak (1995) studies a nematode ganglion system, while Cherniak *et al.* (2004) discuss the cat and macaque cortices.

Chomsky has suggested a framework for thinking about the role of evo-devo concerns in the study of the FL. Treating the FL akin to other biological systems, [77] he points out that three factors are involved in its growth and development:

(23) (a) a genetic endowment specific to language;
(b) external data, converted to experience that selects one or other language from within a narrow range;
(c) principles not specific to the FL.
(adapted from Chomsky (2005a: 6), (2007b: 2))

The grammatical component of (23a) is now the sole topic of *Universal Grammar* (UG) (Chomsky 1957, 1965), a term previously synonymous with all innate grammatical competence; the rest of the burden of learnability is carried by (23c). (23b) amounts to PLD.

(23c) incorporates a range of so-called 'third factors', principles that "enter into all facets of [the] growth and evolution [of cognitive systems]" (Chomsky 2007b: 3). Chomsky presents the following characterization of third factors:

(24) (a) *principles of data analysis that might be used in acquisition;*
(b) *principles of structural architecture and developmental constraints;*
(Chomsky 2005a: 6)
(c) *properties of the human brain that determine what cognitive systems can exist.*
(Chomsky 2007b: fn.4)

Principles of substantive and computational optimality fall into the second category, (24b), and the subset of principles of physio-chemical law discussed is part of (24c).

(24a) includes domain-general principles of statistical learning, with which acquirers track the probabilities associated with particular analyses of their PLD. Yang (1999, 2000, 2002, 2004) has shown involvement of statistical learning in syntactic development; Legate & Yang (2007) in morphological development; Legate & Yang (2012) in phonological development; Shulka, White

[77] Boeckx (2006: 10) notes that Chomsky's framework mirrors that proposed by Lewontin (2001) for explanation in biological science more generally: genes, environment, and organism.

& Aslin (2011) (partly reviewing previous work) in speech segmentation. [78] Probabilistic methods also seem to be active in non-linguistic tasks in: deciphering melodies in streams of tones (Saffran *et al.* 1999); picking out higher-order spatial structures from sequences of visual elements (Fiser & Aslin 2001; cf. Kirkham, Slemmer & Johnson 2002); and improving performance in motor response to visual cues presented in a stream (Hunt & Aslin 2001). Hauser, Newport & Aslin (2001) even demonstrate that, like humans, cotton-top tamarins discriminate between sequences of syllables in accordance with their input frequency. There seems to be robust evidence that a domain-general, primitive principle of data analysis operates in FLA, discussed further in §7.1.

At least two other domain-general principles of data analysis also seem to be implicit in FLA, both discussed in Ch.7. We will see that learners tend to generalize variable input (*i.a.* Ferdinand *et al.* 2013). This effect is most familiar in the linguistic literature from the well-known tendency of children to over-regularize morphological inflection (*i.a.* G. Marcus *et al.* 1992). We will also see that LA-ers tend to characterize ('parse') their input using the minimum set of postulates (see Gallistel & Gibbon (2000, 2002) and Gallistel (2002, 2003), and their discussion of 'over-shadowing' effects in conditioning studies from 1960s). Much attention has been paid to the 'conservatism' of LA-ers in the linguistic literature, starting with Kiparsky's (1965) adaptation of Halle (1961, 1962) and Chomsky's (1965) "simplicity metric", from its original use in capturing linguistically significant generalisations, to a *literal* preference for simplicity of grammar or representation (see also Clark & Roberts (1993), Kiparsky (1995), Roberts & Roussou (2003), Roberts (2007)). However, none of this discussion has focused on the acquisitional *process*, and so the wider significance of studies into grammaticalisation, changes in phonological rules, etc. has been missed, including their indication of cognitive optimality and domain-general principles of data analysis.

As is clear, the categories of third factors are not in complementary distribution (and at any rate this was not the intended reading (Chomsky p.c.)). As we saw, principles of efficient computation may be consequent to physio-chemical law. Similarly, in as far as it relies on abstract computation, principles of optimality will constrain data analysis, as just intimated. Generalization effects may stem from the limited capacity of WM (see *i.a.* Ferdinand *et al.* 2013) – a developmental constraint of a (24b)-kind (although not a principle of computational optimality *per se*). The limits of WM are in turn imposed by the nature of organic matter, a (24c)

[78] Reiterating a point from §1.2: as a means of learning, probabilistic 'training' is in no way inconsistent with (or an alternative to) innate linguistic competence.

concern already discussed. The three categories of 'third factors' are given in order of increasing generality: (24a) reflecting (24b) reflecting (24c).

While we have seen how (the substantive claim of) the SMT expresses an evo-devo hypothesis for grammar, the third Minimalist proposal indicates the relevance of third factors for the design and evolution of the *whole* FL. In particular, we expect principles of optimality to be implicated in all components of the FL (grammatical or otherwise) in as far as they involve computation, including the faculties of performance.

We must be careful, however, not to neglect the significance of the third Minimalist proposal for non-computational aspects of the FL. The FL is far from exhausted by computation, and developmental constraints are not exhausted by principles of computation: they also include, say, the physiology of the vocal [79] and auditory systems, and existing hominid conceptual apparatus. A useful distinction for thinking further about the role of non-computational (and computational) factors in the evolution of language has been introduced by HCF and FHC. These papers distinguish between the *faculty of language in the broad sense* (FLB) and the *faculty of language in the narrow sense* (FLN). The FLB comprises all abilities and structures necessary and sufficient to explain human language, "independent ... of whether they are specific to language or uniquely human" (FHC: 181). The FLB includes the grammatical computation, but also the operations of "at least two other organism-internal systems" (HCF: 1570-1) – the AP- and CI-systems – and perhaps other faculties, a position left open. The FLN is defined as that "subset of mechanisms of FLB [that] is both unique to humans, and to language itself" (FHC: 181). The relevance of this distinction springs from the observation that "[s]omething about the faculty of language must be unique in order to explain the difference between humans and other animals – if only the particular combination of mechanisms in FLB" (*ibid.*: 182).

Investigation of the contents of the FLN is investigation of the contents of (23a), the genetic innovations required for the FL phenotype. The third Minimalist proposal would have us examine how much needs to be put in this category, and how much of the FLB (previously referred to as just the FL) is inherited from elsewhere, i.e. attributable to third factors. One approach to this question is the strictly language- and computation-internal one outlined above. Under this

[79] In fact we can see third-factor principles at work *within* the production system. Concern for 'least effort' in the design of serial motor activity, itself cognitively and physically costly, underlies the prevalence of effects such as elision and co-articulation. Indeed, as would be anticipated, sensitivity of hand-signs to surrounding gestures is also well attested in the phonetics of signed language – see Cheek (2001) and Mauk (2003).

approach, the computations involved in the use, structure and acquisition of language are inspected for evidence of optimality. To the extent that this can be discerned, we can narrow the FLN. To the extent that optimality is breached, we have *candidates* for the FLN: *Merge* and an undeletable EF are already under consideration.

As suggested, there are limits to this approach, even for computational abilities. As HCF (1572) remind us, "if the language evolution researcher wishes to make the claim that a trait evolved uniquely in humans for the function of language processing, data indicating that no other animal has this particular trait are required". This is a matter of logical necessity. Likewise, it is necessary to assess the FLB species-*internally*, with respect to non-linguistic cognitive domains, in order to determine whether any of its features are shared in this way. Only this *comparative method* allows us to decide among the candidate structures provided by the language-internal method, and only the comparative method can speak to non-computational components of the FLB. [80]

If we are interested in the evolution of the FLN, however, not just its contents, we must compare the FLB with the closest relatives to the *Homo* genus possible, ideally the great apes: this reduces the possibility that any common features are the product of convergent evolution, and so similar by dint of analogy rather than homology. [81] Yet even if we can compare a trait directly with our nearest relative (arguably the chimpanzee), we must accept that there have been five to six million years of divergent evolution, and so ten to twelve million years of relative evolution, since we shared a common ancestor. Aside from this limitation, for many relevant traits it is extremely difficult to design and interpret studies which allow us to determine the abstract cognitive capacities of non-linguistic organisms.

With respect to traits shared across human cognitive domains, the obvious confound is how to determine which version of the trait is ancestral, usually impossible given the absence of relevant

[80] We actually saw the comparative method in action just now, in determining the non-FLN status of three principles of data analysis.

[81] This is not to dismiss the value of uncovering analogies in identifying principles of biological design relevant to evolution. For instance, songbirds demonstrate sophisticated skills of vocal imitation, showing a similar acquisitional trajectory to human speech (a 'babbling' phase) and employing similar genetic and neurological machinery (Bolhuis & Everaert 2013). HCF (2002: 1572) point out that this implies "important constraints on how vertebrate brains can acquire large vocabularies of complex, learned sounds", constraints which can properly be investigated further, through study of comparable situations (if available). It must also be noted that the comparative method itself is ultimately the only means of distinguishing homology and analogy.

paleontological or archaeological records. Finally (and trivially), the comparative method has no purchase with respect to the evolution of human traits secondary to the existence of a unique trait. The 'advantage' of the purely linguistic approach is that it allows careful delineation of what might be specific within computational components of the FLB without direct comparators, such as the parser or phonology. The language-internal and comparative approaches are entirely complementary, and in fact both are necessary in investigating the evolution of language.

An excellent demonstration of both the value and difficulty of the comparative method is provided by investigation into the computational capacity of infinitely embedding recursion, characterized by *Merge* and an undeletable EF. This appears to be a species- and domain-specific property of language (specifically syntax). In determining this, the comparative method has dismissed claims that recursion is evident in *i.a.*: birdsong (see Beckers *et al.* (2012), Bolhuis & Everaert (2013)); artificial language learning in animals, including baboons (Rey *et al*. 2012) and finches (Abe & Watanabe 2011) (see Watumull, Hauser & Berwick (2014)); and action grammar (Fujita 2007) (see Samuels, Hauser & Boeckx (in press)).

Species-internally, a unary version of *Merge* is a natural source of the successor-function [82] that grounds arithmetic number and mathematical competence (see Chomsky (2007b, 2008)). Infinite nesting recursion also seems to be evident in the human faculty for music (see Katz & Pesetsky (2009)). In terms of the *origins* of recursion in the FL, various possibilities have been suggested, including our aforementioned capacity for numerical reasoning, but also, to the extent that there are genuine parallels in these domains, our capacities for: spatial reasoning/navigation, serial motor activity, tool-making, social planning, etc. (see HFC and FHC for discussion). As it is, "there are not and probably never will be data capable of discriminating among the many plausible speculations that have been offered about the original function(s) of [recursion]" (FHC: 185).

In summary, the comparative method has taken candidate structures for the FLN (in this case suggested by the language-internal method) and found that they are not satisfactory. With respect to evolutionary origins of recursion, however, the comparative method cannot adjudicate. The genetic novelty of NS may encode for *Merge* and an undeletable EF, or simply their co-option for language. As we will see though, further considerations strongly suggest the linguistic primacy of unbounded embedding recursion, at least among its unambiguous instantiations: the adaptive

[82] If $LI_1 = 1$, then unary *Merge* of LI_1 gives $\{LI_1\} = 2$, and unary *Merge* of $\{LI_1\}$ gives $\{\{LI_1\}\} = 3$, and so on.

value of recursion as part of language seems incredibly powerful relative to its value as part of a mathematical or musical faculty. This is discussed at greater length in §3.1 and §8.1. [83]

This is not the place to consider the findings of the comparative method at length, but it is important to present some of the 'highlights' to give a sense of its value and of the state of progress in pursuing the third Minimalist proposal. [84] First, I believe it might be useful to present a somewhat enriched characterization of the FLB:

(25) (a) FL (Human-specific, Language-specific) = FLN
(b) FL (Narrow evolutionary burden) = FLN + originally language-specific resources exapted by other modules + non-ancestral abilities, developed for use in the FL, but with analogues in the animal kingdom
(c) FL (Human-specific, Not language-specific)
(d) FL (Not human-specific, Not language-specific)
(e) FL (Not human-specific, Language-specific) = Ø

The content of these categories is a matter for empirical investigation. Third-factor principles and the ICs are divided between FL(HS,NLS) and FL(NHS,NLS). Early biolinguistic enquiry assumed that FL(HS,LS) (and therefore FL(NEB)) had rich content, but the tenor of the finding of recent comparative research (and language-internal Minimalist enquiry) suggests the opposite.

So far, all we attributed with confidence to FLN is the genetic machinery incorporating *Merge* and an undeletable EF into NS: we are necessarily vacillating between putting infinite recursive

[83] Now is a good moment to reiterate the value of an agnostic position regarding the overall function of language, only this time expressed with respect to the comparative method:

> *By separating the internal computations from what they might or might not be used for, we open the door to a new line of inquiry, and especially, the possibility that computations deployed by humans for language may be deployed by animals in other domains such as spatial navigation and social interaction that have no communicative expression … [W]e can [also] begin to look at the possibility that some of the computations subserving language in humans are actually deployed in other domains in both humans and other animals, and thus, did not evolve for language.*
>
> (Hauser, Barner & O'Donnell 2007: 106)

[84] It is also important to remember that the comparative method is the best available language-independent evidence for the nature of ICs (see fn.65).

hierarchy in FLN and putting it in FL(HS,NLS). However, something else which appears to belong in FLN is non-referential, non-reflexive lexical meaning, discussed in §1.2. Productions of animal communication seem to rely entirely on one-to-one relationships with aspects of the environment for their meaning. Both spontaneous and trained productions are apparently just labels of "association" with the world, "with no sensitivity to differences among natural kinds": for instance, "[a] chimp will use the same label *apple* to refer to the action of eating apples, the location where apples are kept, events and locations of objects other than apples that happened to be stored with an apple" (Petitto 2005: 39). There is no evidence that animal productions systematically supervene on mind-internal concepts (see also Seidenberg & Petitto (1979, 1987) and Terrace *et al.* (1979)), and furthermore, they seem to be directly connected of the mind-brain's recognition of particular states of the world: Tattersall (2002) reports Jane Goodall's observation that for chimpanzees "the production of a sound in the *absence* of the appropriate emotional state seems to be an almost impossible task" (cited in Berwick & Chomsky (2011: 39)). This is all in sharp contrast to human LIs, which are produced voluntarily, having been disconnected from the source of their meaning and stored in a separate, abstract lexicon – a learnt list of pairings between *instructions* to the CI-systems and (abstract) instructions to the AP-system. [85, 86] Similarly, human LIs are distinctive in being in a one-to-many, many-to-one relationship with the world: one LI can take on many meanings, and one meaning can be given by many different LIs.

A second capacity which seems destined for FLN is the human prosodic hierarchy (although not the use of rhythm itself). Kitahara (2003) reports that cotton-top tamarins are unable to discriminate between human languages on the basis of their prosody. This is an unsurprising finding, as the "syntactic resources that require such prosodic-sensitive system [sic] might not have evolved for them" (*ibid.*: 38). That is, prosody gives information about the underlying syntactic structure of a sentence, which would otherwise be difficult (or even impossible) to parse successfully, as discussed above in §1.2 and below in §6.1.4. We have already seen word-prosody's crucial role in word segmentation; prosody seems to satisfy the IC of useability/'inheritability', and so be a principled element of FLN.

[85] Assuming the existence of abstract underlying representations, which we do for well-documented empirical reasons (see Krämer (2012) for a summary).

[86] Strictly speaking, the abstract input to the serial set of rules which generate the articulatory/acoustic score of instructions to the AP-system, following the classic presentation of Kenstowicz (1994), argued for robustly by Vaux (2008).

A capacity which seems to be part of FL(NEB) yet excluded from FLN is our ability for auditory learning and vocal imitation, necessary to support a rich lexicon. A similar ability is demonstrated by songbirds (see fn.81) and distantly-related mammals (Janik & Slater 1997), but not by our nearest relatives, the great apes (*i.a.* Crockford *et al.* 2004). This strongly suggests an adaptive origin for this (analogous) ability, under pressure to support the FL.

We have already discussed one human ability which is part of the FL(HS,NLS), namely the fully-generalized successor-function which gives rise to the arithmetic concept of number employed by language. Also into this category fall distinctive properties of human social and personal cognition. We reserve discussion of this matter to §8.2, but it is sufficient here to note that human-unique conceptualisation of self and others exists and is used in lexical meaning. [87] Another (trivial) component of FL(HS,NLS) will be the *finer* details of our AP-apparatus.

Another way of thinking about the FL(NEB) and FL(HS,NLS) together is as the FL (Wider evolutionary burden). We have already seen that there is little evolutionary burden imposed by arithmetic number, and Chapter 8 proposes that recursion, non-referential semantics, and distinctive properties of humans' personal and social cognition more or less collapse together in evolutionary terms. If there is anything to this suggestion, it would not just be interesting because of the dearth of plausible accounts of the evolution of lexical meaning and human social / personal cognition, but also because part of the third Minimalist proposal is that language might be part of a 'channel' in evolutionary space.

The comparative method has revealed rich contents to FL(NHS,NLS), which, again, is as expected if evo-devo concerns apply to the FL. Language makes use of a range of ancestral cognitive and physiological machinery, including but *by no means* exhausted by:

[87] It will be noted that we have not yet postulated as part of FLN any genetic machinery relating to the externalization of language. While this may at some stage be proven necessary, one distinct possibility, as Chomsky (2010: 27) suggests, is that externalization arose out of our rich theory of others: we understood others were processing information in the same way as us, through "normal empathy", and so recognized the possibility of meaningful externalization. This has intuitive appeal, because it's hard to imagine what selective advantage would be associated with an isolated mutation favouring externalization, such that it might spread through a scattered population.

Table 2: Some conserved machinery co-opted by the FLB

(i) the singular-plural distinction (Barner *et al.* 2007);
(ii) categorical perception (Cutting 1982, Cutting & Rosner 1974, Rosen & Howell 1981);
(iii) a finite repertoire of species-particular articulatory productions (Michael Coen, cited in Berwick & Chomsky (2011: 34));
(iv) algebraic rule learning (Hauser & Glynn 2009);
(v) temporal indexing of information (Gallistel 2009);
(vi) spatial indexing of information (Gallistel 2009);
(vii) the notion of bounded entities moving in a continuous fashion (see Spelke (1998) for a review);
(viii) rhythmic articulation (Ramus *et al.* 2000, Yip 2006);
(ix) the animate-inanimate distinction (*i.a.* Gelman 1990);
(x) a largely conserved anatomy of vocal production (see FHC (198-9) for review). [88]

Noticeably absent from the above classification are principles of substantive and computational optimality. We very plausibly assume these to be ancestral properties of cognition, part of FL(NHS,NLS). However, to the best of my knowledge, the matter has undergone little empirical investigation in non-linguistic cognition, in humans or animals. This is at least partly attributable to the difficulty of studying the computational processes with output harder to detect than language. [89] One example cited as external evidence of computational optimality (*i.a.* HCF (1578), Chomsky (2012: 60)) is the foraging strategy of certain insects, birds and primates (*i.a.* Stephens & Krebs 1986, Gallistel & Cramer 1996, Shettleworth 1998). As it is, these studies merely reveal that optimal results are reached in spite of the great complexity of the task. While this is compelling evidence of some "hardwired disposition selected over the eons" (Piattelli-Palmarini & Uriagereka 2004: 352), i.e. a third factor, it does not obviously speak to the issue of computational efficiency. The analogy with the principle of 'minimal search' from the algorithmic domain is just that, an analogy, confusing the result and the cognitive process, the latter of which we know very little about. As more detailed understanding of the computation underlying different cognitive tasks becomes available, it is anticipated that the comparative

[88] Several items are deliberately left off this list for discussion in §3.1.3.

[89] Also relevant, as Chomsky (2012: 60) points out, is the remaining adaptionist intuition that "simplicity is the last thing you'd look for in a biological organism".

method will reveal the domain-generality and ancestral nature of cognitive optimality, but at the moment we are in a position of (highly plausible) assumption.

We conclude our discussion of the third Minimalist proposal by returning to an earlier refrain. The MP raises concerns that apply independent of commitments (best avoided) regarding the purpose of language. Despite persistent misunderstanding of this issue, [90] the third Minimalist proposal does not preclude natural selection from playing a role in the evolution of the FL. It merely argues that matters of "correlation and balance" (Darwin, 1856 letter to W.D. Fox, cited in Berwick & Chomsky (2011)) with existing resources be considered in the evolutionary narrative, as they must. This point is made forcefully by the chronology of the FL's emergence.

The archaeological record suggest that language arose within the last 100,000 years, some 5-6 million years after the emergence of the *Homo* genus (and *c.*100,000 years after the appearance of anatomically recognizable *Homo sapiens*). [91] Furthermore, there is compelling evidence that the FL has not evolved in any significant way since our ancestors left Africa between 50,000 and 80,000 years ago (Tattersall 2010). The FL emerged in the "blink of an evolutionary eye", and the question of how this is compatible with "the notion that language evolved piecemeal in the human lineage" [92] (Pinker & Jackendoff 2005: 218) in a "gradual" (*ibid.*: 223) fashion arises whether you want it to or not. As it is, evolution is necessarily 'conservative', a "truism" the third Minimalist proposal extends to biolinguistics.

[90] See Boden (2005) and Pinker & Jackendoff (2005), replied to by Chomsky (2007a) and Fitch, Hauser & Chomsky (2005) (particularly the online appendix) respectively; see also Kinsella & G. Marcus (2009).

[91] See §3.1.1 for further discussion.

[92] As Chomsky (2005a) points out, incremental accounts of the evolution of recursion – ones in which "later stages ha[ve] to build on earlier ones" (Pinker & Jackendoff (2005: 223) and references therein) – seem inherently nonsensical. Any intermediate stage in evolution (for which there is no evidence) would involve the imposition of a constraint on recursion; the end is reached before the middle. Chomsky (*ibid.*) points out that the same paradox also arises in FLA. While there is *prima facie* evidence that a LA-er adopts intermediate positions, developing progressively complex productions, in actual fact, studies have long demonstrated that children's computational abilities are dramatically underemployed by the faculties of production: "the telegraphic speech of … children does not reflect the fact that they [can] discriminate telegraphic syntax from … adult syntax" and that they prefer "well-formed sentences that they themselves never produc[e] at this stage … as indicated by their tendency to act on [these alone]" (Gleitman, Gleitman & Shipley 1973: 142).

2.4 A framework for the Minimalist Program

2.4.1 "An optimal solution"

We have now outlined the three Minimalist proposals which are a necessary consequence of studying language as a biological object, and together comprise the MP. Unifying all three proposals, but playing a different role in each, is concern for substantive and computational optimality. For the MP to be pursued with integrity therefore, there must be clarity over what it means to be optimal in these ways. In this section, I present the nearest thing to a rigorous framework for Minimalist enquiry possible. I say the nearest thing possible, because, as Chomsky (2001: 1-2) points out, "'good design' conditions are in part a matter of empirical discovery". These conditions are unknown in at least two related ways.

First, very little is known about how cognition is implemented by the brain, and, by extension, about what constraints the brain imposes on the FL. While it is clear that symbols, COs, a processor and WM are instantiated in some fashion by the brain, it is not clear how this is done: "relating mental computations to analysis at the cellular level is currently a distant goal" (Chomsky 2005b: 2). [93] Indeed, Gallistel & King (2009: xvi) are profoundly sceptical about the only hypotheses that have "ever been seriously considered by the neuroscience community", arguing that we may need to fundamentally reconsider our overall approach to the question. [94] We cannot therefore know which symbols and functions are possible descriptions of linguistic tasks; computationally sub-optimal solutions may be favoured by neurology for currently unknown reasons. (As we saw, (24c) is a more general constraint on growth and development than (24b).)

To take a concrete example, it is commonly assumed that the operation *Merge* is instantiated by a simple function of set-formation, yielding 'unordered' pairs: *Merge* (A, B) = {A, B}. This is the 'smallest' function compatible with recursion. However, there is no empirical reason to assume

[93] Although there have been interesting findings about the neurology of the FL at an architectural level (see Berwick *et al.* (2013) and references within).

[94] Cf. Chomsky's (2000b: 106) observation that:

> *Large scale reduction is rare in the history of the sciences. Commonly the more "fundamental" science has to undergo radical revision for unification to proceed. The case of chemistry and physics is a recent example: Pauling's account of the chemical bond unified the disciplines, but only after the quantum revolution in physics made these steps possible.*

that a larger function, yielding 'ordered' pairs, is not used to generate recursion, such that *Merge* (A, B) = {A, {A, B}} or {B, {A, B}}. It is known that this function (or some equivalent function associated with greater computational complexity) is employed in NS non-cyclically to generate adjoined structures (see §4.1.5), but as yet we have no empirical reason to assume it is not also the basic operation of structure-building. (Chomsky p.c. to Jeffrey Watumull) The assumption of a 'set-*Merge*' function over a 'pair-*Merge*' function in cyclic derivation is based on MM alone.

Uncertainty surrounding implementation is one of the reasons we only explore the hypothesis that "language is **an** optimal solution to [interface] conditions". (The second reason will be clarified when we introduce computational principles relevant to the proposed research heuristic.) Our framework for optimality must therefore be based on nothing more than general guidelines of an *a priori* nature. [95]

In short, we restrict ourselves to commenting on "the goal of the computation … and … the logic of the strategy by which it [is] carried out" (Marr 1982: 25), refraining from commenting on "how this computational theory [is] implemented" (*ibid.*) or how the computation might "be realized physically" (*ibid.*).

I begin by proposing a self-explanatory, trivial notion of substantive optimality: *The Less, The Better* (*TLTB*). This principle merely observes that a proliferation of types of symbols, constraints on types of symbols, types of CO, and constraints on the output of COs [96] (taken to be) used in the FL runs contrary to MM, OM as methodology, and the evo-devo hypothesis for language. [97]

[95] It should be noted that a teleologically-based research program for language, besides introducing a wide range of unnecessary uncertainties, suffers from this more fundamental uncertainty just the same: there is the unacknowledged question of what is 'easy' for the production and comprehension systems. This question can only be answered with certainty commensurate with our concrete knowledge of these systems.

[96] Incorporated as increased complexity of COs, potentially involving reference-set computation of the kind discussed above.

[97] The final point also merits clarification in light of the 'indefinite article issue'. First, it would be necessary to account for the development of symbols / constraints on symbols / operations in some way, an evolutionary burden. Secondly, there will be some cognitive cost simply to having any symbol/constraint/operation 'in the toolkit'. That is, some "principle of structural architecture" will plausibly constrain the instantiation of new computational machinery – although we have little idea of its character, in accordance with the obscurity of neuronal implementation. For instance, does it 'discriminate' against symbols more than operations? Is it only relevant beyond a certain level of complexity? It is for this reason that we cannot propose any more nuanced a principle of substantive optimality than *TLTB*.

Pursuit of analyses observing *TLTB* is therefore central to the MP. The only principled breaches of *TLTB* will be those required to meet ICs, or which afford a 'compensatory' saving, by allowing more efficient computation. We will see examples of both below.

TLTB has been previously expressed in Minimalist work in the *Inclusiveness Condition* (*IncCon*) and the principle of *Full Interpretation* (*FI*). The *IncCon* holds that "[n]o new objects are added during the course of the computation apart from rearrangements of lexical properties" (Chomsky 1995c: 228). *FI* holds that features can only appear in a derivation if they are already interpretable to the interfaces, or can be properly licensed for deletion [98] before the interfaces (cf. Chomsky 1991: 437-8); features which are redundant with respect to the interfaces or the derivation are prohibited. [99] Taken together they preclude redundant symbolic content in the derivation *with respect to the broader postulates.* The *IncCon* and *FI* are consistent with *TLTB*, but do not exhaust it: *TLTB* also speaks to the number of different types of symbols in need of licensing and number of operations used in meeting ICs, i.e. the "broader postulates".

I turn now to the second 'branch' of Minimalist optimality. I take the overarching principle of computational optimality to be *Maximize Throughput*: the overall aim of any algorithm [100] is to turn input into output as quickly as possible. To explain how this aim is achieved, we introduce two fundamental concepts of *complexity theory – time* and *space complexity*. The time complexity of an algorithm is the number of operations it must perform (and intermediate memory states it must pass through) to complete its task. The space complexity of an algorithm is the amount of WM it requires while running (Manber 1989: 42-3). Reducing the number of operations which must be performed as part of an algorithm (trivially) reduces the time taken by a given processor to run it to completion. In practice, the speed with which an algorithm can produce output will often be constrained by the fact that operations are running to completion faster than space can be cleared in WM for their results to be stored, and a 'bottle-neck' arises. [101] In this case, reducing the space complexity of an algorithm will improve its throughput, but

[98] "Checked" in earlier theory.

[99] Uninterpretable versions surviving to the interface will in cause a derivation to 'crash'.

[100] Marr (1982) uses the term *algorithm* to refer to his second level of abstraction – "functional" implementation. My use of the term is synonymous with his "computational theory", the highest level of abstraction, and the only one over which we currently have the right to formulate principles of optimality.

[101] In organic memory, this may not be quite so black-and-white: the ease of retrieval of particular objects may 'decay' with the amount of time they have been stored, in which case it is really the process of retrieval (or perhaps 'reactivation' – see fn.27) that causes the bottleneck, not limited space *per se*.

reducing its time complexity will have no effect. The opposite situation also arises: processing power may be the resource which is maximally tasked, such that there is a 'backlog' of symbols stored in WM awaiting future operations. In this case, reducing the time complexity of an algorithm will improve its throughput, but reducing its space complexity will have no effect. Throughput is maximized by achieving the optimal 'trade-off' (see e.g. Hellman (1980)) between time and space complexity in the design of the algorithm, such that as many operations as possible can run to completion before processing power or WM becomes a limiting factor. Of course, the point of optimal balance depends on the relative cost of these two resources; however, as discussed, next to nothing is known about this issue, and so we restrict ourselves to postulating principles which reduce operational load (time complexity) or cache load (space complexity). We refrain from speculating on precisely how these principles contribute to maximizing throughput, while remaining aware that in "some unknown way" they must. [102] This is another way in which we are forced to generalize our framework in accordance with what is known about conditions of 'good design'.

I therefore base a framework for computational optimality around principles of *Minimize Time Complexity* (*MinTC*) and *Minimize Space Complexity* (*MinSC*). [103] I divide each of these principles into sub-principles describing the various different ways of reducing operational load and cache load. I begin with *MinTC*.

The most trivial way of reducing time complexity is to *Minimize Redundant Operations* (*MinRO*) – that is, to prohibit operations which have no effect on the (relevant) structure of the output. That means ruling out operations which inherently fail to develop additional relevant structure – *Minimize Vacuous Operations* (*MinVO*) – and prohibiting the reversal of previously performed operations – the *No Tampering Condition* (*NTC*) (Chomsky 2005b: 13). [104] Redundant

[102] Even this is too strong. It may be that reducing space complexity is never relevant to maximizing throughput, and that reducing space complexity is simply an independent principle of optimality, minimizing the cognitive cost of storage (and perhaps retrieval) from WM.

[103] Strictly speaking, these are merely principles of *algorithmic* optimality. In practice, a computational task may consist of a chain of algorithms, such that the output of one algorithm is the input of the next. Overall throughput can therefore be improved by optimizing the way in which algorithms in a chain release their output, a matter discussed in greater length in §5.1 below, but put to one side for the time being.

[104] (Sub-)principles which reduce redundant operations will also reduce the space complexity to the extent that derivational history must be held in WM. The same can be said for *TLTB*: if fewer different types of symbols are used in meeting ICs, then it is likely that fewer will have to be stored in WM as part of derivational history (and in fact the fewer operations will have to be performed to introduce them, reducing operational load). For the

operations are also performed every time a symbol in WM is read but deemed 'unsuitable' for use by the algorithm motivating this search: a natural sub-principle of *MinTC* would be to design algorithms so as to *Minimize Search* (*MinSearch*).

The other logical means of reducing time complexity is to design algorithms in such a way that the same information does not have to be established more than once for independent reasons during the course of a derivation. This is slightly different from redundancy, because relevant structure might be developed in each case; but it would obviously be preferable for the processor not to recapitulate previous computation *in toto* to do so, in keeping with *Minimize Reduplication* (*MinRedup*).

MinSC appears to consist of three sub-principles. First, an algorithm ought not to introduce a symbol (or complex symbolic structure) into WM before an operation can relate it to the existing derivation: optimal computation will *Minimize Caching of Unintroduced Items* (*MinCUI*). Secondly, optimal computation will not allow processing to reach a point at which a relevant symbolic structure is available for an operation to be performed, only to postpone that operation: a commitment would then have to be carried over in WM. Optimally, an algorithm ought to *Minimize Caching of Incomplete Derivation* (*MinCID*). Finally, an optimal algorithm will 'page out' complete derivation to less pressured memory resources ('secondary storage') at the earliest possible opportunity, reducing the load on WM in accordance with *Minimize Caching of Completed Derivation* (*MinCCD*).

One thing apparent about the above proposals is that they are only meaningful if assessed in a *local* fashion. Principles of economy invoked in early Minimalist work [105] were essentially specious (Lappin & Johnson 1999, Lappin, Levine & Johnson (2000a, 2000b, 2001)), as they required a large number of derivations to be performed, saved in WM, and then compared with respect to relevant properties. The difficulties were recognized almost as soon as the first proposals were made, and efforts immediately made to render a "more "local" interpretation of reference sets" (Chomsky 1995c: 227) (*i.a.* Chomsky (1993: 32-44), Collins (1997) and Epstein *et al.* (1998)). However, as Brody (1997: 140) points out, this still left intractable "complexity":

---

purposes of simplicity, however, we present matters of substantive, operational, and cache-related optimality as independent.

[105] For example, the *Shortest Derivation Constraint* – "[m]inimize the number of operations necessary for convergence" (Chomsky 1991: 427) – or *Shortest Move* – "the operation [of movement] should always try to construct the "shortest link"" (Chomsky & Lasnik 1991: 89).

by "looking ahead" in this way, any computational optimality of principles of economy is nullified.

It was only with the development of empirically and theoretically [106] preferable models of syntax, which took feature satisfaction to motivate structure, as part of *Bare Phrase Structure* [107] (Chomsky 1995b) and *Attract*-based theories of movement [108] (Chomsky 1995c: ch.4), that comparison-free principles of economy began to fall out. It then began to become clear that more general principles of economy might be at play, *not* in "their formulation … specific to the language faculty" (Chomsky 1991: 447), as previously supposed, and that the FL might *not* "be unique among cognitive systems, or even in the organic world, in that it satisfies minimalist assumptions" (Chomsky 1995c: 221). The third Minimalist proposal emerged later in the MP, capitalising on progress made in pursuit of the earlier two proposals (and still earlier progress in resolving the tension between descriptive and explanatory adequacy).

2.4.2 An aside on reference-set computation and CI-mapping

As Chomsky (1995c: 222) points out, "[i]t is important to distinguish the topic of inquiry here from a different one" – the question of "whether [grammatical computation] is derivational or representational". It is, of course, an empirical question whether the principles of grammar necessary to meet ICs are encoded as filters/bans on the output of computation, or whether convergent derivation is licensed by successive operations alone. It should be borne in mind, however, that the comparison of outputs under filters/bans constitutes non-local licensing of operations, requiring "reference-set computation" and vacuous operations of comparison, as discussed above. But this is not to say that any comparative computation will not itself be subject to principles of *local* economy, prohibiting redundant operations, etc.; it is for this reason that the issues of local vs. global economy and representational vs. derivational computation are orthogonal. However, from the point of view of *overall* computational optimality, it is preferable that the linguistic principles necessary to satisfy ICs be instantiated as unfulfilled properties of symbols. It is worth noting that Chomsky (1995c: 223-4) gives some empirical arguments for the fundamentally derivational nature of linguistic computation, [109] and recent syntactic proposals for how to meet explanatory and descriptive adequacy are at least less reliant on output conditions.

[106] With respect to the first two Minimalist proposals.

[107] See §4.1.3.

[108] See §4.1.4.

[109] See also Vaux (2008), Embick (2010), Heck & Müller (2007), and Müller (2004, 2009).

And so, while avoidance of reference-set computation is not itself a principle of computational optimality of the kind we propose, which are *inherently* local in their concerns, it could be considered a 'meta-principle' of computational optimality. A breach of this meta-principle is justified on familiar grounds, except adjusted to the higher level of abstraction: reference-set computation is admissible when derivational approaches are unable to satisfy ICs. This is precisely the argument made by Reinhart (*i.a.* 2004, 2006) in her discussion of optimal design and the peripheries of grammatical computation. For instance, multiple quantifier constructions produced by NS may manifest scopal ambiguity, and reference-set computation may be required to license the quantifier raising/lowering necessary to produce the intended meaning (Fox 1995, 2000). Similarly, the main sentence-stress assignment rule on its own does not to allow us to assign focus-stress to every sentence constituent which may assume discourse focus, and global licensing may have to be invoked to assign it appropriately (Reinhart 2006: ch.3). On such occasions when local computation fails to satisfy ICs, only then is global computation legitimate, in accordance with evo-devo expectations.

Mapping to the CI-interface has other unusual properties in respect of the optimality considerations presented above. First, even operations which involve movement, i.e. which seem to use syntactic machinery, such as antecedent-contained deletion (May 1985), are *tampering* operations, applying only to void containment relationships giving rise to erroneous readings. Binding and quantifier raising breach the strong cyclicity of NS (Chomsky 2002: 158-9) and the principles of (locally-assessed) computational optimality motivating it (see §§4.1.2 & 5.1). Similarly, CI-mapping (arguably only in regard of scope, following Chomsky's (2008) discussion of binding) seems to compute *c-command* [110] relationships, apparently unnecessary in NS, where the only relationships recognized are set-membership and *probe-goal* (see §4.1.4). This is a breach of *TLTB*. Given our location, immediately at the CI-interface, it appears likely that these breaches of optimality are principled ones. In fact, Chomsky (2002: 159) even suggests that we might think of ourselves as being on other side of the interface, with the above operations being "the interpretive systems on the meaning side, the analogue to articulatory and acoustic phonetics, what is going on right outside the language faculty … using operations similar to internal operations but probably not the same."

[110] Reinhart (1979) formulates *c-command* as follows: A c-commands B *iff*: (i) the first branching node dominating A dominates B, (ii) A does not dominate B, and (iii) A does not equal B.

2.4.3 The framework

I am now in a position to present a tentative framework for how best to think about the Minimalist notions of optimality necessary for meaningful pursuit of the MP (*pace* the need for comparative research). This will be referred to throughout the rest of this dissertation.

Figure 2: A framework for (non-comparative) pursuit of the Minimalist Program

Substantive optimality:

- *The Less, The Better* (*TLTB*)

Computational optimality:

- *Maximize Throughput* (*MaxTP*)
    - *Minimize Time Complexity* (*MinTC*)
        - *Minimize Redundant Operations* (*MinRO*)
            - *Minimize Vacuous Operations* (*MinVO*)
            - *No Tampering Condition* (*NTC*)
            - *Minimize Search* (*MinSearch*)
        - *Minimize Reduplication* (*MinRedup*)
    - *Minimize Space Complexity* (*MinSC*)
        - *Minimize Caching of Unintroduced Items* (*MinCUI*)
        - *Minimize Caching of Incomplete Derivation* (*MinCID*)
        - *Minimize Caching of Completed Derivation* (*MinCCD*)

The final two Minimalist proposals, which together with the MP establish the argument of linguistic Minimalism, are presented in the following chapter. Here, however, we have examined the logic of the first three Minimalist proposals, before demonstrating the possibility of an explicit (albeit partial) heuristic for pursuit of the MP. It is hoped that the discussion as a whole will assist in meaningful biolinguistic enquiry and protect against the misappropriation of Minimalist logic.

3. Rounding Out Linguistic Minimalism

The first three Minimalist proposals are an immediate reflex of studying language from a biological perspective, and apply regardless of any particular ideas about the nature of the FL. The fourth and fifth Minimalist proposals, however, *are* particular ideas about the nature of the FL. They are intimately connected with the proposals of the MP, but logically independent, and require independent argumentation.

3.1 The fourth Minimalist proposal: the 'primacy of the CI-interface'

The fourth Minimalist proposal (*i.a.* Chomsky (2005a, b) and Berwick & Chomsky (2011)) holds that core properties of the FL may have been 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 [may have come] about only secondarily" (Jacob 1982: 59). The corollary of the fourth Minimalist proposal is that NS and the mapping to the CI-interface (referred to collectively in the rest of the section as syntax) would be "an optimal solution" to CI-ICs alone, and ought not to breach the SMT to accommodate any other ICs.

3.1.1 'Language-external' considerations

The fourth Minimalist proposal gains its initial plausibility from the observation that language constructs an infinity of hierarchically-organized, embedding structures, using LIs with non-referential semantics of great sophistication. This allows 'reason' to work with internal representations of much greater semantic complexity and abstraction, and so to operate in completely new arenas. The utility is obvious, allowing a range of clearly advantageous behaviours: planning sequences of events; solving complex problems with consideration for contingencies; reflecting on past events; rehearsal of thoughts for memory purposes; etc. [111]

It should be noted that these abilities rely on a capacity for abstraction emerging from the 'conceptual eclecticism' of the underlying LIs: productive thought is not restricted to referential

[111] We reserve detailed discussion of the ability for metaphorical extension of thought over issues otherwise 'unthinkable' with ancestral CI-apparatus until §8.2.

items [112] or abstract LIs of one conceptual type. [113] There would perhaps be some utility to a *Merge*/EF conglomerate operating within a single conceptual domain – this is the intuition behind suggestions that recursion may have been exapted by the FL from another faculty, discussed in §2.3. As we saw in §1.1, however, the LI *book* is composed *i.a.* of concepts relating to: persistence in time / identity; intention; creation; and accomplishment. There is a certain disconnect, therefore, between the claim that *core syntactic properties* of language were selected for on grounds of "abstract or productive thinking" and the nature of such thought. Indeed, "[t]he fact that there even is a problem has barely been recognized, as a result of the powerful grip of the doctrines of referentialism" (Berwick & Chomsky 2011: 40). [114]

Nonetheless syntax is also a *sine qua non* of productive abstract thought. [115] It is far from clear that the ability to form an isolated mental representation with 'cross-conceptual' semantics is of any utility at all. [116] This is revealed to a certain extent by the very fact that only by using examples combining LIs with similar semantic objects are we able to reveal the subtleties of words' meanings. Furthermore, the scope and complexity of abstract thought is a direct reflex of unbounded *Merge*'s ability to produce infinitely long and infinitely structured mental representations – limited only by the capacity of computational and CI resources. We will argue in Ch.8 that the machinery of unbounded embedding recursion more or less 'gifts' the "promiscuity" of the interfaces underlying lexical semantics (and much else unique to humans), *contra* Chomsky's (2012: 40) claim that "there are no sensible origins" for this property of human

[112] Assuming referentiality is separable from the property of reflexivity, which given the cognition behind animal 'words', seems unlikely: animal productions seem to be properties of the same perceptual systems which give them meaning. In this case, 'referential items' would not be available for hierarchical organisation in the first place.

[113] Be they simplex or complex. The *Merge*/EF amalgam is not taken to be the only means of building mental representations of symbolic complexity (i.e. thought): for example, bees (and of course other animals, including humans) are able to compute representations of courses in 'spatial' cognition which combine the notions of **bearing* and **distance* (Gallistel 1998).

[114] It must be acknowledged that Luria and Jacob's comments refer to language as a whole.

[115] This takes language to *be* the "language of thought" (LoT), not as providing the machinery to instantiate an independent, recursive LoT (J.A.Fodor 1975), which would otherwise suffer from an "explanatory regress" (Berwick & Chomsky 2011: fn.6). As it is, there does not seem to be any empirical reason to postulate a system of this kind. The fourth Minimalist proposal is also distinct from the observation that inner speech has a range of cognitive uses (see §2.2): there is no perceivable relic whatsoever of purely syntactic thought, consistent with Berwick & Chomsky's (2011: 25) observation that "the inference of a biological trait's 'purpose' or 'function' from its surface form is always rife with difficulties."

[116] *Contra* Hauser (2013), as discussed in §8.1.

language. It should be noted that the evolutionary chronology of language discussed above argues for the catastrophic emergence of lexical semantics every bit as much as it does so for unbounded *Merge*, unless there is some mysterious reason to believe LIs were present prior to the development/co-option of unbounded *Merge* (having developed via a mechanism different to the one proposed below). However, even if this is all beside the point, it would remain plausible that core computational properties of language were selected for mind-internal reasons.

We can begin to move beyond initial intuition by looking at properties of the world. One relevant consideration already encountered is the result of studies into the structure, acquisitional trajectory, and neurology of sign languages. Core facts about language present as entirely independent from externalization, as would be anticipated if syntax emerged without regard for and prior to externalization, but instead evolved under selective pressure for thought. This contrasts with the previously discussed commitment that all aspects of the FL evolved "in response to the adaptive value of more precise and efficient communication" (Pinker & Jackendoff (2005: 223) and references within).

Other data consistent with the fourth Minimalist proposal come from archaeology. Evidence of symbolic activity, including representational engraving, music [117] and personal ornamentation (see d'Errico *et al.* (2009) for summary), appears suddenly in the artifactual record, around 100,000 years ago, *c.*100,000 years after the first anatomically recognizable *H. sapiens*. No sign of such activity is noticeable in the (deduced) behaviour of earlier *H. sapiens* or more ancient, similarly large-brained hominids, such as *H. neanderthalensis* (Klein 2009, Bar-Yosef & Bordes 2010, Higham *et al.* 2010). At the *same time*, the archaeological records begin to indicate a range of other cognitive abilities, including complex planning, steady improvement of technology, counting, [118] and the interpretation of natural phenomena (see Tattersall (2012) for discussion). Complex planning for the future, innovation, and interpretation of the past all come from reason interacting with the 'elaborative rehearsal' of complex, non-referential thoughts. The simultaneous development of these clearly advantageous 'thought-behaviours' with language seems unlikely to be coincidence, strongly suggestive of a selective relationship.

The progression of tool technology apparent in the artifactual record is particularly persuasive in this regard. Hominid tool manufacture is first attested *c.*2.5 million years ago (Semaw *et al.*

[117] The faculties for art and music include the FL's capacity to relate non-iconic form with non-referential meaning. (As discussed in §2.3, the latter seems to supervene on linguistic computational capacities also.)

[118] Reliant on unary *Merge*, as discussed in §2.3.

1997), and technology has become more sophisticated ever since (Tattersall 2012). There is a sharp distinction between the pattern of innovation over most of this period, and that of the last *c.*100,000 years. The earlier period shows sporadic refinement of technology, and only recently do we see a "pattern of continual enhancement" (Hauser *et al.* 2014: 6). The early record expresses the absence of any systematic approach to innovation, with developments contingent on trial-and-error or accidental discovery. The recent pattern, however, indicates an active, 'problem-solving' approach to technological improvement, "typical of modern linguistic *Homo sapiens*" (*ibid*.).

3.1.2 'Language-internal' considerations

The case for the fourth Minimalist proposal is bolstered by consideration of the concrete output of grammar, which has many features inconsistent with a syntax produced by adaptive response to pressure for "precise and efficient communication". Chomsky makes this point repeatedly in his work, *i.a.* Chomsky (2002, 2005a, 2005b, 2007, 2008) and Berwick & Chomsky (2011), from which the following examples are drawn.

First, and most obvious, is the fact of ineffability: "a lot of the things that we would like to say may be very hard to express, maybe even impossible to express" (Chomsky 2002: 107). What's more, this is particularly true for significant feelings and pressing communicative concerns; often there simply are not "the words to say it". For a supposed adaptation for communication to be so divorced from its subject matter that "a lot of personal interactions collapse" (*ibid.*: 108) as a result would be surprising to say the least. [119]

Also relevant is the fact that our articulacy and success in communication (at least in particular domains) can vary significantly as a function of personality, education, and other contingencies. This would make the FL somewhat anomalous among core domains of cognition, such as vision, memory, motor control, etc., in which the overwhelmingly better part of competence is taken to develop irrespective of anything but major deficiencies in input or other abilities (see *i.a.* Gelman (2009) and references within). It is also noticeable that impossible outputs of grammar are perfectly comprehensible – a strange property of a system serving communication – e.g. prohibited double-object constructions such as:

[119] Thoughts which are "impossible to express" will include non-linguistic/pre-linguistic thoughts, which fall outside the remit of the LoT entirely. See fn.115.

(26) *I donated the library a book.

Closely related to ineffability is the fact of *island constraints* (Ross 1967) on constructions. Typically, structures breaching these constraints will be "fine thought[s]" (Chomsky 2013: 41), yet require awkward circumlocution to express. Take, for instance, the *wh*-island constraint, [120] breached by (27b):

(27) (a) I wonder [where$_i$ [he went to find a present t_i]]
(b) *What$_j$ do I wonder [where$_i$ [he went to find t_i t_j]]?

To be grammatical, this must be re-expressed using a cleft construction:

(28) What is it that I wonder where he went to find?

This is a particularly interesting case, because *wh*-islands arise as a reflex of the *Minimal Link Condition* (Chomsky 1995c: 311): [121]

(29) *The Minimal Link Condition* (MLC)
K attracts α only if there is no β, β closer to K than α, such that K attracts β.

Where is frozen in place by the selectional properties of *wonder*, which obligatorily takes a sentential complement (an indirect question), and so the MLC blocks raising of *what* to matrix C. This is a show-case example of a principle of computational optimality, *MinSearch*, operating in the computation of NS, without breaking stride for communicative concerns.

What's more, as seen in §1.3.3.1, the core property of syntactic computation, unbounded embedding recursion leads to structures which despite being perfectly grammatical and semantically sound are unparseable. Notoriously, unbounded *Merge* yields nested dependencies which quickly overflow WM. To be communicatively viable, (30a) must be re-expressed more paratactically, as (30b), or better still, completely paratactically, as (30c).

[120] This forbids extraction of a *wh*-phrase embedded in a CP which has a *wh*-phase in its specifier.

[121] See §4.1.6.

(30) (a) The mouse the cat the dog bit chased died.
(b) The mouse died, which the cat the dog bit chased.
(c) The mouse died; the cat had chased the mouse; the dog had bitten the cat.

For the core property of language to be directly implicated in grammatical, meaningful nonsense of this kind would be unexpected if language stems from pressure to make sense. And there are other interactions of production with parsing that lead to the same conclusion: language displays rampant lexical and structural ambiguity in usage, including ambiguity yielding significant cognitive effort to overcome, *viz. garden path* sentences (Bever 1970), and even fully global, unresolvable ambiguity. These are matters discussed at greater length in §6.1.4.

3.1.3 The status of the interface conditions

We have seen that various reasons to suspect there is something to the fourth Minimalist proposal. It might well be the case that language initially interfaced with the CI-system alone, and that components of the FL involved in externalization were arose only subsequently. One means of further scrutinizing this possibility is closely aligned with the pursuit of the MP.

If the structure-generating properties of language evolved under motivation from externalization, they would be expected to reflect CI-ICs, AP-ICs, and useability/inheritability conditions, as well as general properties of cognition. That is, a communicative account of the evolution of the FL would lead us to expect greater 'idiosyncrasy' in the design of syntax, on account of higher demand for disruption of general cognition. Finding principles of optimality active in syntax would by no means be grounds alone for accepting the primacy of the CI-interface, but their greater 'visibility' in syntax would be suggestive, particularly in light of syntax's 'failings' in regard of externalization. An intriguing example in this regard was the case of *wh*-islands.

Hinzen (2006, 2008), Sheehan & Hinzen (2011) and Hinzen & Sheehan (2013) have gone so far as to suggest that it is syntax itself which 'naturalizes' much of our semantic interpretation. One submission is that the "specific compositionality of semantics" (Hinzen 2008: 354) may be a direct reflex of the structures generated by syntax. Hinzen (*ibid.*) suggests that propositionality (in this sense, the existence of primary bearers of truth-value), event structure, and the notion of argumenthood fall out from the CP-*v*P-DP alternation of syntactic structures and have no place in a theory of our language-independent CI-abilities independent of language. Similarly, Sheehan &

Hinzen (2011) argue that the degrees of specificity with which we refer are the result of the format of our grammar, not independent properties of cognition. Rigid reference / truth is associated with the substitution of N/T to the 'edge' of the DP/CP phase, such that only the edge is responsible for determining reference. Definite reference sensitive to descriptive specification, or 'facthood', is associated with a filled D/C position, but no substitution, so NP and TP contribute to reference alongside the properties of D/C. Indefinite existential nominal / propositions have empty or underspecified D/C positions, with their reference specified exclusively by the contents of the phase. This is not the place to discuss the parallelism of the nominal and clausal domains in depth; suffice it to say though that Sheehan & Hinzen (2011) take the ways we may refer to the world to be grammatical in origin.

From my perspective, there is a certain circularity to these proposals: the CI-system is still expected to have the resources necessary to "exploit the property of [these] generated expressions" (Chomsky 2008: 15); and language must still postulate appropriate semantic content to C/*v*/D-heads, content revealed by work within the 'cartographic enterprise' (see §3.2.2.1) to be highly sophisticated. It's not easy to understand how the ability to interpret propositions differ from the notion of propositionality itself, nor to understand where the content of heads for, say, evaluative mood or epistemic modality comes from if objects bearing truth values are syntax-internal notions. Analogous considerations apply at other levels of clause structure. Another way of putting this objection is to ask what gives non-linguistic thought (see fn.115, 119) its conceptual status: this approach seems to strip such thought of mind-internal significance, a somewhat strange conclusion.

Sheehan & Hinzen (2011: 4) observe that the existence of propositionality, event structure, argumenthood, particular modes of reference, etc. "should be a central question from a biolinguistic perspective", and that a fully "explanatory approach to semantic theory … thus remains unaddressed". They seem to take the parallels between linguistic structure and our conception of the world as evidence for the source of this explanation. [122] This seems to approach the issue backwards: the relevant question to ask really is why syntax is constrained

[122] Or in the case of modes of reference, they take the identity between referential possibilities across the clausal and nominal domains to indicate their grammatical origin. This seems somewhat specious, because it is only when we get recursive linguistic structure that we can possibly distinguish 'sentential' from purely 'nominal' meaning in the first place. For instance, when a vervet monkey makes an alarm call for *leopard* (Cheney & Seyfarth 1990), for all we know this has sentential meaning equivalent to "(It's true that) I am looking at that leopard", just as much as it means "leopard".

to generate the structures it does, given its infinite generative capacity. It is preferable to at least start from the position that the categorisation of syntax reflects a CI-interface system designed to meet *pre-linguistic* thought.

As it is, closer scrutiny reveals that the comparative method has gone some way toward answering Sheehan and Hinzen's "central question" along these lines. Studies of the social cognition of scrub-jays (*i.a.* Emery & Clayton (2004), Dally, Emery & Clayton (2006)) and rhesus monkeys (*i.a.* Flombaum & Santos 2005) reveal that these animals have some understanding of conspecifics' / humans' knowledge-states, at least within certain domains. For instance, Flombaum and Santos found that rhesus monkeys were more likely to steal food from experimenters who couldn't see them, or from a box without bells attached. The monkeys were sensitive to the equivalence of knowledge-states arising from the different types of perception: when a monkey was obviously being observed, it demonstrated no preference for stealing quietly or stealing noisily (Santos, Nissen & Ferrugia 2006). Monkeys are able to reason over knowledge-states, so it is clear that some notion of proposition is ancestral. [123]

Studies also testify to an [actor-[action-goal]] schema in animal cognition, i.e. a sense of event structure. It is clear that rhesus monkeys understand the relationship between direction of gaze and paying attention; a monkey who can see will also be acquiring information – 'acting' in an event. The same finding is attested in studies of new-born chicks. Regolin, Tommasi & Vallortigara (2000) showed chicks a video display of two static balls, one red and one blue; the red ball was then shown to move and bump into the blue ball. After habituation in this way, the chicks were presented with the objects shown on the screen and were found to imprint to the red ball rather than the blue one. This indicates the chicks were sensitive to its 'activity' in events, conceptualizing an [actor-action] schema. This impression is reinforced by the finding that the imprinting preference did not arise when the initial movement of the red ball was occluded, so that the chicks could not have known whether it had started moving of its own accord or under external influence.

[123] This may be too strong a claim if knowledge relies on belief in the way Andrews (2012) suggests. Discussion in §8.2 suggests that only humans truly master belief-states, and in this case propositionality (even in the sense assumed here, and constrained to particular domains) is only explicable through the evolutionary innovation underlying belief-attribution. Monkeys are then taken to master the *implication of perceptual states* within particular domains, using domain-specific cognitive apparatus.

An understanding of goal-directed/intentional behaviour has been established in chimpanzees in a study by Uller (2004). Uller employs a dishabituation paradigm designed by Gergely *et al.* (1995) to test the equivalent property of pre-linguistic human infants. Chimpanzees were habituated to a video showing a small ball jumping over an obstacle to get to a larger ball. After habituation, the chimpanzees were shown one of two videos, both showing the small ball moving toward the larger ball, only without there being an intervening obstacle. In one video the small ball jumped in the air before reaching the larger ball regardless, whereas in the other it moved directly to the larger ball. The chimpanzees displayed greater attention to the former video than the latter, indicating that they had interpreted the ball's behaviour in the habituation video in terms of its ultimate aim. It seems, therefore, that both the [actor-[action]] and [action-goal] schemata, notions of *event structure*, are ancestral.

Furthermore, Gallistel (2011) argues robustly that events are represented ancestrally as functions/predicates acting over variables/arguments, just as verbs act in language. In the scrub-jay studies mentioned above, it was found that *experienced* jays noticed the attention of other jays when caching, returning to retrieve and re-cache these items. An experienced, or suspicious jay is one which has itself stolen from the caches of other jays in the past: innocent jays are also naïve jays (Emery & Clayton 2001). As Gallistel (2011: 259) points out, "[t]he generalization from a jay's own behaviour" seems to show that "the symbol for an action is independent of the agent and the direct and indirect objects". The experienced jay would seem to be in control of the something like the following representations – <*I*> **take** < *food*> <*from you*> and <*you*> **take** < *food*> <*from me*> – acquiring the second on the basis of the first. It seems that predicate-argument representation of event structure / *argumenthood* is also ancestral. (Cf. also Carruthers (2006) and Hurford (2007).)

Finally, we turn to the issue of modes of reference, finding again that animals show similar behaviour to humans. It is well-documented (see Marler, Evans & Hauser (1992) for summary) that referential alarm calls reliably fall into one of two classes: some refer to a specific creature, such as the 'martial eagle', 'leopard', or 'python' calls of vervet monkeys, whereas others refer to the general category of the predator, such as the 'aerial predator' or 'ground predator' calls of bantam chickens (Evans & Marler 1995). The former case is arguably equivalent to rigid/'proper name' reference, and the latter to definite reference sensitive to descriptive specification. Also in the former category would fall bottlenose dolphins' signature whistles, with which they identify themselves (Janik, Sayigh & Wells 2006). It is not surprising that nothing akin to the indefinite existential reference (associated with bare nouns and embedded propositions in human language)

is manifest in animal communication, given its reflexive nature. We are now in a position where we could consider adding: propositionality, event structure, predicate-argument structure, and modes of reference to Table 2 above.

In conclusion, while Sheehan and Hinzen raise a good question about the nature of the CI-interface, there seem to be logical and empirical grounds to question their answer. The comparative method has made significant progress toward explaining why the functional architecture of syntax is the way it is, answering their question in a meaningful way. Finally, returning to the main track, we observe that in as far as Hinzen and Sheehan's suggestions are true however, the CI-ICs are diminished further, and principles of general cognition expected to be even more plainly manifest in syntax than is already anticipated. [124]

The fourth Minimalist proposal leads to a different expectation in the case of phonology and mapping to the PF. The imposition of ICs from NS, the AP-system, and useability/inheritability predicts that the principles of general cognition will be somewhat less easy to discern in the activity of modules of externalization. Importantly, this does not amount to the claim that evo-devo concerns do not apply to these modules: it is a "truism" that subsequent evolutionary development is constrained by previous development. Furthermore, the FL's modules of externalization emerged at least as recently as syntax, and so the question of 'evolvability' is equally pressing. It will be noted that studies using the comparative method, whose results are summarised in Table 2, show that much of the machinery involved in externalization is conserved from ancestry, including: categorical perception, a finite repertoire of species-particular articulatory productions, algebraic rule learning, rhythmic articulation, and the anatomy of vocal production. [125] In fact, in one way we have discussed, they may be *less* dedicated content to modules of externalization, since a rich theory of others may be sufficient to motivate externalization, with further genetic instruction unnecessary (see fn.87). The implication of the fourth Minimalist proposal for externalization is best investigated by pursuit of the MP and

[124] One entirely atheoretical point that emerges from these considerations is the need to be cautious about the use of the term 'interface condition'. A CI-IC is something which must be satisfied to allow us to have conceptual sophistication of a kind already realized (albeit perhaps in limited domains). In as far as language creates novel conceptual complexity / typology, we must make the distinction between a CI interface condition and a CI *concern.* It has already been argued that this distinction is necessary in the case of complex abstract thought and arithmetic number, and §8.2 will argue the same is true for metaphor and aspects of our personal and social cognition.

[125] For more detailed discussion of this matter, see Samuels (2011).

comparing its results for syntax (NS and mapping to the CI-interface) and phonology/mapping to PF. A relative lack of visible optimality in the design of phonology/mapping to PF, when compared to syntax, would be suggestive evidence for Berwick, Chomsky, *et al.*'s claim.

3.2 The fifth Minimalist proposal: variation and linguistic Minimalism

The fifth and, in my classification, final Minimalist proposal is another claim regarding the nature of the FL itself. In this case, the proposal bears on the manner in which the FL permits variation between language systems.

3.2.1 The history of the syntactic 'parameter'

The issue of how to explain linguistic variation has been dominated by the question of how to reconcile descriptive and explanatory adequacy. Descriptive adequacy promotes increasingly complex and varied formulations of grammar to accommodate diversity, whereas the need for explanatory adequacy argues the exact opposite, that languages must essentially be "cast in the same mold" (Chomsky 2004a: 148), or the acquisition problem cannot be overcome.

Early attempts at defining permissible variation began with the question of descriptive adequacy, deriving structure via essentially language-specific phrase structure rules and transformations, constrained in various ways (Chomsky 1973). It soon became clear that explanatory adequacy could not be met on those terms, and focus turned toward the means of constraint, ultimately generalizing the construction of phrase structure as part of X-Bar Theory (Chomsky 1973, Jackendoff 1977, Chomsky 1981), and conflating all transformational rules into one generalized principle, *Move-α* (Chomsky 1981, 1986b). Under the new perspective of this *Government & Binding* (GB) Theory (Chomsky 1981), grammar was conceived of as a range of innate principles, and variation between languages could therefore be captured by the underspecification, or *parameterization*, of certain principles, leaving (binary) grammatical options to be specified by LA-ers on the basis of PLD. This *Principles & Parameters* (P&P) approach (Chomsky 1981) resolved the tension between descriptive and explanatory adequacy in the way discussed in §2.3: P&P acquisition merely consists of selecting from among pre-determined options and no longer (necessarily) invokes a "rich and highly articulated" "format" for grammar, required to satisfy explanatory adequacy (Chomsky 2005b: 9).

GB parameters were associated with the deep formulation of NS, defining the fundamental orientation of the system. Classic examples, illustrating the wide scope of permissible variation, include the *Head Parameter* (HP), the *Wh-movement Parameter* (WHP), and the *Null Subject Parameter* (NSP).

(31) *Head Parameter* (cf. Chomsky 1986)
Heads X PRECEDE/FOLLOW their complements.

(32) *Wh-Movement Parameter* (cf. Huang 1982)
Wh-movement is either overt (within NS) or covert (after NS but before LF).

(33) *Null Subject Parameter* (cf. Rizzi 1982, 1986)
Subjects can be *pro*, where *pro* is a null, [+pronominal, -anaphoric] entity which is licensed by X^0_y where X^0_y is a subset of the heads in a given language. The grammatical specification of *pro* is recovered from the features of the licensing head.
(reproduced from Biberauer (2008: 8))

Most GB parameters, including these, have proved empirically inadequate. We will return in §6..2 to discuss the status of linearisation within grammar, merely noting here that a category-insensitive HP, (31), runs into the well-known difficulty that the vast majority of languages are not fully harmonic, but exhibit mixed headedness. A WHP taking the form of (32) is similarly problematic. It leads us to expect only two types of language: *wh*-movement languages and *wh*-in-situ languages. This binary choice cannot account for a language such as Duala however, which allows *wh*-movement and *wh*-in-situ in matrix clauses, without any (necessary) interpretive difference, while requiring movement in embedded clauses (Biberauer 2008: 18).

More recent thinking about null-subject languages (NSLs) is discussed in greater detail shortly, but here we present the 'clustering' prediction of Rizzi's NSP. Following work by Perlmutter (1971), Rizzi suggests that the following collection of properties is associated with a positive setting of the NSP, properties predicted to be collectively absent from non-null-subject languages:

(34) (a) the possibility of a silent, referential, definite subject of finite clauses;
(b) 'free subject inversion';
(c) the lack of complementizer-trace effects;

(d) the availability of expletive null subjects;
(e) "rich" agreement inflection on finite verbs.

However, when Gilligan (1987) tested the predicted correlations, he found many were not borne out. For instance, many morphologically-rich languages, such as Russian, are not NSLs; many languages morphologically impoverished without respect to agreement, such as Japanese, are NSLs; some languages lack null subjects, but permit that-trace violations, such as Icelandic; and so on. In fact of all the possible correlations among the properties, only four were unchallenged by Gilligan's hundred-language sample (Newmeyer 2004: 202-6):

(35) (a) free subject inversion → expletive null subjects;
(b) free subject inversion → complementizer-trace violations;
(c) referential null subjects → expletive null subjects;
(d) complementizer-trace violations → expletive null subjects.

The onset of the MP altered the landscape for parametric variation dramatically. Grammatical knowledge (like the rest of the FL) is now taken to be shaped by three factors, restated here for convenience:

(23) (a) a genetic endowment specific to language;
(b) external data, converted to experience that selects one or other language from within a narrow range;
(c) principles not specific to the FL.

We have seen that Minimalist considerations indicate a relatively small UG, (23a), and emphasize the contribution of domain-general cognitive principles, part of (23c), to grammatical competence. As M. Richards (2008a: 134) points out, variation cannot be associated with the overall architecture of cognition, but, equally, where there is little content, there can be little differentiation, and so associating variation with UG is also problematic. This raises the question of how to derive linguistic variation in a way consistent with these constraints.

One way of achieving descriptive adequacy within this new Minimalist framework is to restrict variation to that which must be learned anyway, the lexicon, and in particular to the featural properties of (LIs instantiating) functional categories. This 'lexico-centric' approach to syntactic variation can be expressed as the "Borer-Chomsky Conjecture" (BCC) (coined by Baker 2008):

(36) Borer-Chomsky Conjecture (BCC)

(a) "Parametric variation is restricted to the lexicon, and in so far as syntactic computation is concerned, to a narrow category of morphological properties, primarily inflectional." (Chomsky 2001: 2)

(b) "The availability of variation [is restricted] to the possibilities which are offered by one single component: the inflectional component." (Borer 1984: 3)

The BCC gives rise to variation associated with the input to NS alone, rather than its deeper formulation. UG is taken to provide an inventory of formal features, **F**, from which acquirers make a one-off selection, defining a subset, **[F]**, of features which play an active role in their language; the other formal features are 'discarded' (Chomsky 2000a: 100-1). Features may vary in terms of whether they enter the derivation *unvalued* (hence uninterpretable at the interfaces) or *valued*; an unvalued feature, *uF*, will act as a *probe*, searching the complement of its head (see §4.1.4) for a *goal* bearing a valued version of that feature, (*i*)*F*, with which it can *Agree* and attain a value. Each *uF* may also be paired with an *EPP-feature*, which dictates that the *Agree* relationship induces movement of the goal to the specifier of the probe. [126]

Following M. Richards (2008a: fn.1), it is important to note that the second factor, external data, is not the "locus of variation", merely the "trigger for variation". It cannot of course be the locus for variation, since it is "language- and organism-external (E-language, not I-language)" (*ibid.*). The logic of the BCC is that it restricts the variation in NS to its input, the LIs, other properties of which must already be taken to vary and be learned (as emphasized by Borer (1984)). In this way, the role of the second factor is expanded with respect to previous accounts of parametric variation, consistent with the Minimalist proposals.

3.2.2 Language uniformity and permissible syntactic variation

3.2.2.1 Language uniformity

The fifth Minimalist proposal is significantly more far-reaching than the BCC, submitting that "parametrization and diversity [are] mostly – possibly entirely – restricted to externalization" (Berwick & Chomsky 2011: 37). This relates to the claim of the fourth Minimalist proposal, that

[126] A'-movement is taken to be motivated by *edge-features* (EFs) of phase-heads (see §5.1) alone (Chomsky 2008) and not mediated by uninterpretable feature valuation.

"externalization is not a simple task, [having] to relate two quite distinct systems", leading us to expect that "morphology and phonology … might turn out to be quite intricate [and] varied" (*ibid*.). This contrasts with expectations for NS, anticipated to be more fully satisfied by general cognitive principles alone. The asymmetric distribution of content within UG is taken to be associated with the asymmetric distribution of opportunities for variation – with variation arising under the influence of PLD ambiguity in acquisition, *viz.* "accidental historical events" (*ibid*.).

The fourth Minimalist proposal gives rise to a second, related motivation for this sceptical position regarding NS variation. If NS (and mapping to the CI-interface) did indeed develop independently of externalization for use in abstract thought, then there would have been *no relevant data* available to the 'language acquirer' to determine any points of grammatical underspecification (regardless of whether these are identified with the featural input to NS or overarching syntactic principles). NS must be (close to) fully deterministic if language was, and continues to be, an unacquired tool for abstract thought. For reasons of learnability in respect of its primary purpose, NS is expected be invariant in the face of acquisitional contingency / PLD ambiguity, even when language comes to be externalized. Essentially the same point is made by the observation that formal features are interpretable/meaningful, and so variation in their activity inevitably gives rise Whorfian conclusions, which seems undesirable. Ultimately of course, the fifth Minimalist proposal must be assessed empirically, and so we examine the evidence for lexico-centric NS parameters. Before doing so however, we must comment briefly on the overarching question of why the FL does not fully determine grammar, whichever way it may vary.

As made clear, wherever parameters reside, they are associated with *underspecification* of UG. Further specification of UG would eliminate optionality, but would be associated with further genetic content, a further evolutionary burden. The expectation of the third Minimalist proposal is that this will be avoided. Of course, underspecification is constrained by the limits of our acquisitional faculties, so cannot be too far-reaching, demanding at least partial specification of grammatical principles and points of parametric variation as we understand them.

Returning now to the assessment of the fifth Minimalist proposal, we find similar observations about the plausibility of the BCC had been made prior to Berwick & Chomsky's (2011) comments. Sigurðsson (2004 *et seq*.) recognizes the difficulty of assuming that CI-interpretable features can be "disregarded in the use of [a language]" (Chomsky 2001: 10), when languages

seem to encode the same semantic distinctions. [127] He points out that when a language does not phonetically realize a certain feature, a theory involving "L-selection" would predict its absence from that language's semantics. However, this would lead us to conclude that: definiteness ought not to be expressed in Japanese and Russian; that Arabic copula constructions ought to lack tense and finiteness; that Icelandic ought to convey logophoric meanings absent in English, etc. (Sigurðsson 2004). In short, Sigurðsson points out the dangers of taking superficial evidence of featural differences too literally, arguing that they do not seem to pattern with underlying semantic differences. He argues for predominant "L-uniformity" on anti-Whorfian grounds, more or less the argument which arises from the fourth Minimalist proposal.

The findings of the cartographic enterprise are also suggestive of genuine consistency in formal feature inventories across languages (cf. Miyagawa (2004: 7) and Cinque & Rizzi (2009: 45)). Closer investigation of the "core functional categories" (Chomsky 2000a: 102) (C, T, *v*, D, and perhaps P) has uncovered that they are best viewed as a cover term for "richer systems" (Chomsky 2001: 43, fn.8). Specifically, cartographic research suggests there is a universal design for the clause and major phrases, each consisting of a uniform hierarchy of functional heads, [128] fixed in number and order, even when languages differ in the extent to which they overtly realize each category. Furthermore, only a small subset of possible conceptual distinctions seems to be employed by language in framing propositions/events/etc. – see in particular Kayne (2005) and Cinque (2013) for discussion. There seems to be a mandatory framework of interpretable formal features for building phrase structure; this is very plausibly imposed by the nature of the CI-interface and "the way we conceive the world". That is, it seems likely that only quite particular semantic structures make sense mind-internally, and so NS must be tailored/constrained accordingly. The argument that certain interpretable features can be somehow omitted by a language is problematic from this perspective: it is reasonable to suggest that the default position ought to be that every language conceptualizes the world in the same way, using the same machinery (cf. Kayne (2005b: 12), Cinque & Rizzi (2009: 45)), if only for the familiar reasons of methodological minimalism. In fact, this might well be a useful heuristic for research in comparative syntax – not precluding the postulation of variation in NS machinery if there is compelling evidence, but, like the SMT itself, insisting that we rule out more parsimonious

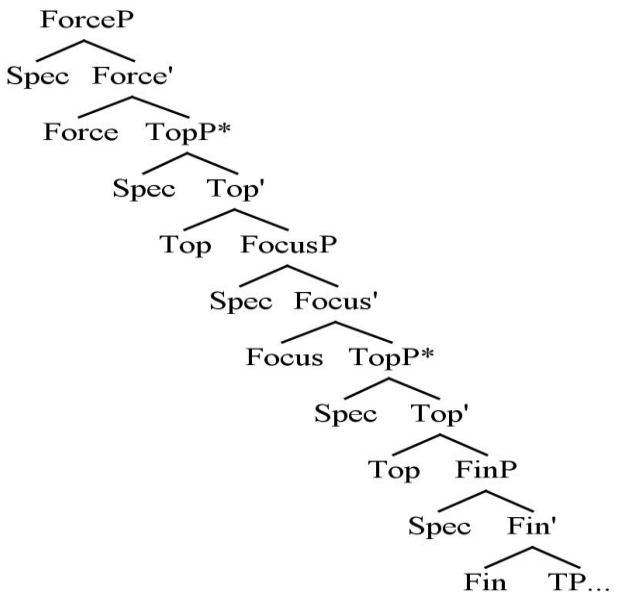
[127] See *i.a.* Longobardi (2008) (discussed below), Ramchand & Svenonius (2008) and Wiltschko (2014) for different takes on this matter.

[128] Or possibly the uniform linearisation of features within a phase label, allowing multiple specifiers of a single head (Chomsky 2005b: 18).

explanations first, guarding against theory-internal 'engineering solutions' with potentially problematic implications.

The seminal work in cartography involved investigation of the left periphery of the clause (*i.a.* Rizzi 1997) – a sample of these findings is shown in Fig. 3. However, similar discoveries have been made within: the core functional structure of the clause (*i.a.* Cinque 1999), the DP (*i.a.* Cinque 1994, 2010), and the PP (see contributions in Asbury *et al.* (2008)). (See Cinque & Rizzi (2009) for an overall survey of the results and methodology of cartography.)

Figure 3: Section of the articulated CP (Rizzi 1997)



The seeming consistency and rigidity of these hierarchies across languages invites us to consider the motivation for their ordering. It seems very plausible that ordering within these hierarchies (and, in fact, the universal C-T-*v* composition of clauses) reflects semantic properties of human cognition and our particular way of viewing the world, i.e. CI-ICs. For instance, Chomsky (2009a: 219-220), following Bever (1970), comments that the adjective hierarchy may follow from how 'noun-like' adjectives are, with more 'nominal' adjectives appearing closer to the noun.

The other hierarchies may follow from similar implicational relationships, when considered from the point of view of human cognition. [129]

However, regardless of the motivation for the hierarchies, there is strong evidence for invariance in the identity and ordering of the 'vertebrae' of the clausal/phrasal spines, as is to be expected under an approach treating NS primarily as a tool for thought. This promotes the expectation that the featural composition of the categories of the hierarchies will be invariant as well. For language to be useable in real-time as a system of thought, without consistent reference to external systems of interpretation, it must be the case that its clausal and phrasal frameworks are largely deterministic in all respects; we can say this much even without complete knowledge of the contents of these systems.

How then are we to assess the evidence that the syntaxes of different languages seem to vary? First, let us look at the ways in which syntactic variation remains permissible under the above understanding of the fifth Minimalist proposal. We note that it remains the case that there is no constraint on the syncretism of functional categories or their features on the LIs which enter NS. NS is 'concerned' only with fulfilling its formal commitments (the functional categories, their formal composition, and their order of *Merge*-r); it is unconcerned with the lexical accretion of what is necessary to fulfil these commitments, and so the make-up of LIs may vary freely, subject to acquisitional contingency.

One of the first suggestions along these lines comes from Ritter (1993). She points out that while Romance languages and Hebrew both (as expected) seem to realize a D-Num-N hierarchy in their DP, the [gender] feature 'bundles' with 'noun' LIs in Hebrew, but with [number]-bearing LIs in Romance. Evidence for this comes, *i.a.*, from the relative ordering of number and gender suffixes (*ibid.*: 799-801). Assuming number suffixes are attached to the head noun as a consequence of syntactic head movement, then gender marking in Hebrew ought to be closer to the noun stem than number marking. In most cases, the gender suffix of Hebrew nouns deletes when the Num-head raises; however, one feminine suffix, *-it*, merely reduces when the plural suffix, *-ot*, is added, and can indeed be seen to surface closer to the noun stem. In most Romance languages, word order considerations suggest that N raises to Num, such that the nominal stem, number marking and gender will be realized on the same head, regardless of their base-generated

[129] From discussion in §3.1.3, cf. Sheehan & Hinzen's (2011) observation that higher heads in the functional spine are associated with more specific reference that lower ones. For further discussion of these issues, see *i.a.* Sigurðsson (2004b), Ramchand (2008), and Ramchand & Svenonius (2014) (and references within).

position. Walloon, however, does not seem to exhibit N-to-Num movement (Bernstein 1991). Walloon can be seen to manifest a feminine plural marker *-ès* co-occuring with prenominal adjectives. Attempts to explain this as something other than a LI bearing [gender] and [number], i.e. to explain it as two LIs (a feminine affix *-è-* and a plural affix *-s*), predict that the feminine marker will surface in (at least some) feminine singular noun phrases, which is not the case. It seems that Walloon differs from Hebrew in having [number, gender]-syncretic LIs, whereas Hebrew differs from Walloon in having [nominal, gender]-syncretic LIs. (See Ritter (1993) for the further discussion necessary to generalize the findings to other Romance languages.) [130]

Other examples of 'bundling' effects come from work in "nanosyntax" by Starke (*i.a.* 2002, 2006, 2011). Without getting into too much syntactic detail, it seems that Germanic 'verb' LIs (mostly) 'spell-out' a smaller area of the functional hierarchy than Latinate 'verbs' (Starke 2011: 9-10). A high functional head, *α,* of the verbal hierachy is associated with the 'enrichment' of the V-head itself, spelt-out by the particle of particle-verb constructions, e.g. (37a), and the adjective of resultative constructions, e.g. (37b):

(37) (a) He [$_{\alpha P}$ [$_{VP}$ picked the chair] [$_{\alpha}$ up]];
(b) The barber [$_{\alpha P}$ [$_{VP}$ cut his hair] [$_{\alpha}$ short]].

Particle-verbs and resultative constructions are absent in languages such as Italian or French because the V-head and α-head accrete. [131] Analogous explanations can be offered for the different behaviour of indefinite pronouns and *wh*-elements in French when compared to English (Starke 2011).

We can see, therefore, the variation in the existence of lexical items does have at least some purchase in explaining the syntactic differences between languages, in a way consistent with our understanding of the fifth Minimalist proposal. However, various scholars have proposed fully systematic, 'lexico-centric' parameters of the kind anticipated by the BCC. For instance, much attention has been paid to the distribution of null subjects within languages. We saw the difficulties associated with the classic NSP, and now we briefly discuss the principal BCC-

[130] Cf. Giorgi & Pianesi's (1997) notion of "feature-scattering".

[131] Săvescu Ciucivara & Wood (2011) make a similar suggestion regarding Talmy's (1985) distinction between the 'path'-incorporating 'put' verbs of Romance languages, and the 'satellite-framed' 'put' verbs of Germanic.

compatible (set of) account(s) of *consistent* NSLs (CNSLs), [132] offered in Holmberg (2005) and the papers in Biberauer *et al*. (2010). A fully detailed account of the varied proposals made in these papers is impossible, but it is possible to give a flavour of the problems posed by a BCC account of consistent *pro*-drop. In these accounts, both the make-up of the null element itself [133] and the featural specification of T are crucial components. Unifying these accounts is the presence of a D-feature on T, which can therefore license referential null subjects, absent in non-CNSLs. The following formulation of the NSP is suggested:

(38) The Null Subject Parameter
Does T bear a D-feature?
(Roberts & Holmberg 2010: 14)

In Holmberg (2005) and Roberts (2010), the *pro*-s of CNSLs are specified as 'weak pronouns' (in Cardinaletti & Starke's (1999) terms) and are "licensed as referential" by the D-feature on T. This successfully accounts for the distribution of null subjects and captures the fact that generic pronouns of CNSLs are overt, since covert versions would be specified as referential. The problem arises, however, that *pro* can surface in the same position as (referential) *overt* weak

[132] Consistent null-subject languages (CNSLs) are those languages in which: (i) it is possible to leave the definite subject pronoun unexpressed in any person-number combination in any finite context; and (ii) there is rich agreement inflection on the verb. These are paradigmatic *pro*-drop languages discussed by Rizzi (1986), and they include Italian, Spanish and Modern Greek. They are distinct from 'partial null-subject languages' and 'semi-*pro*-drop languages', which space precludes us from discussing here (see Holmberg & Sheehan (2010) and Biberauer (2010)). A very different kind of null-subject system, 'discourse *pro*-drop', is discussed shortly.

[133] Holmberg (2005) demonstrates that *pro* is indeed a *Merge*-d element (albeit using evidence from partial null-subject languages), inconsistent with approaches taking the pre-verbal subject position as unprojected in CNSLs (*i.a.* Borer (1986), Alexiadou & Anagnostopoulou (1998); cf. also Holmberg (2010), Sigurðsson (2011), Holmberg & Roberts (2013a)). In these approaches, the *Extended Projection Principle* (EPP) – the requirement that Spec-TP be filled by a nominal element – is reduced to the requirement that T be [+pronominal]. Chomsky (2014), in fact, argues against the traditional EPP holding at all in CSNLs, and associates rich agreement with the ability of TP to project a label in the absence of a specifier. He does not clarify if he takes this to reflect featural variation of T or not, but since labelling is no longer considered to take place syntax-internally, but as part of interpretation (see §4.1.8), it seems likely that he does not, with the ability to label independently merely reflecting 'recoverability' concerns of the kind discussed immediately below. That is, "labelling algorithms" can now also use phonological information to identify relevant syntactic relationships, and it is plausibly easier for them to identify T as the probe (the labelling element) if it is phonologically-marked for the *Agree* operation it instigates. This gives substance to the otherwise undefined notions of "richness" and "weakness" of syntactic objects (SOs) for labelling purposes which Chomsky (2014) assumes. (See fn.137 also.)

pronouns (Roberts 2010: 72, fn.14), such as *egli* and *esso* of Italian. This raises the question of why these overt pronouns are not licensed as a referential *pro* (or, in Roberts' terms, "deleted") by the D-feature on T. We are forced to classify overt weak pronouns as a *third* category of pronouns in Italian, alongside strong pronouns and *pro*, with some relevant distinction arising between *pro* and these overt weak pronouns. Roberts (*ibid*.: fn.14) suggests it might be that overt weak pronouns are inherently specified for Case, and hence not licensed as *pro* / undeletable: this would prevent them from being "defective goals" relative to their T-probes (which following Chomsky (2000, 2001) lack Case features), a requirement for deletion in Roberts' account. It is also essential for Roberts' account, therefore, that we take *pro* to lack Case, a situation called into question by Holmberg's (2010) discussion of generic null subjects in Finnish. [134]

These issues disappear, however, if we abandon the notion of special *pro*-licensing which follows from a commitment to 'extra' T-content active in the derivation – a commitment which is based on the observation of rich agreement morphology alone. *Pro* and *egli*/*esso*-type pronouns do not need to be categorically distinguished, because *pro*, like these overt weak pronouns, does not need to be specially licensed to be referential: it is an *independent* LI – simply a null weak pronoun – governed by the same rules of distribution as other LIs.

The intimate connection between rich verbal agreement morphology and consistent *pro*-drop [135] gives a straightforward explanation of the 'recoverability' of referential null subjects in CNSLs. In order for a LA-er to infer referential *pro* LIs, there must be sufficient evidence for them in the input, i.e. rich agreement morphology on the verb indicating the φ-features of the covert pronoun. Similarly, language users will only drop overt subjects when verbal morphology means that it is comprehensible to do so. [136] The same claim is made independently by Faarlund (2013: 257) in his discussion of the history of *pro* in Scandinavian languages, where he suggests that "loss of *pro* from the lexicon … can be explained as caused by a lack of sufficient input data during

[134] See Roberts (2010: 79-80, fn.22) for a response.

[135] In accounts such as Roberts (2010), the rich agreement morphology is taken to indicate that T has a full set of φ-features, necessary for it to bear a D-feature; it is taken that T may vary with respect to φ-content also.

[136] This is not to say that syncretisms preclude the availability of thematic null subjects of the kind found in CSNLs – see Cole (2009) for relevant discussion. Nor is it to say that verbal agreement and the availability of null subjects will always co-vary diachronically – see Axel (2005) for relevant discussion. The connection between verbal agreement morphology and definite null subjects in CSNLs, nonetheless, remains robust.

acquisition; at a certain stage the necessary cues for a phonologically empty item is [sic] insufficient" (*ibid.*: 281) (cf. also Holmberg (2005: 560)). [137]

Somewhat counterintuitively, I propose that similar logic can be extended to explain the free distribution of null arguments in a very different set of languages, 'discourse-*pro*-drop languages' (DPDLs), which have little if any person-agreement marking, but allow null arguments quite freely in various functions; Japanese and Korean are classic examples. Naturally, BCC-compatible models of these languages have been somewhat different to those proposed for CNSLs. I can neither present nor analyse these accounts here (*i.a.* Tomioka (2003), Neeleman & Szendrői (2007, 2008) and Saito (2007); see Roberts & Holmberg (2010: 9-10) for a summary), but I suggest that discourse *pro*-drop amounts to nothing more than the availability of particular feature-bundles (LIs) in the lexica of DPDLs.

I pursue Tomioka's (2003: 336) observation that "all languages which allow discourse pro-drop allow (robust) bare NP arguments", without following his suggestion that a null argument is merely an instantiation of one of these bare NPs (formed by ellipsis, or a true *pro*) (cf. also Barbosa (2011, 2013)).

Taking Japanese as a representative example of a DPDL, let us first examine the distribution of its "bare NP arguments". Japanese can use a bare noun with kind-referring, generic semantics:

(39) Kuzira-wa honyuurui da
Whale-TOP mammal-COP
'Whales are mammals'
(Nemoto 2005: 392)

Japanese bare nouns can also be used with indefinite or definite semantics:

[137] If Chomsky (2014) is correct, then the loss of rich verbal agreement is associated with the rise of the EPP and the loss of the ability of TP to label, rather than the loss of a LI. This does not involve the loss of substantive content of T (see fn.133), merely a movement diacritic, so the matter is of no broader significance to the argument here (see §3.2.2.2 below).

(40) John-ga hon-o yon-da
John-NOM book-ACC read-PAST
'John read a book/the book'
(Kurafuji 2004: 212)

Japanese bare nouns may take on singular or plural (non-generic) readings:

(41) John-ga hon-o yon-da
John-NOM book-ACC read-PAST
'John read some books/the books'
(*ibid*.)

Japanese bare nouns can be associated with either structural Case:

(42) Gakusei-wa hon-o yomimasita
Student-TOP book-ACC read
'A student/students read (a) book(s)'
(Nemoto 2005: 384)

In order to be given specific number, Japanese bare nouns must be associated with a classifier construction:

(43) Hanako-wa san satsu-no hon-o katta
Hanako-TOP three (bound-object-classifier)-POSS book-ACC bought
'Hanako bought three books'
(*ibid.*: 403)

Japanese 'bare nouns' are clearly not a uniform set. They *can* be "robust bare NPs" without the ability to refer, merely denoting the kind semantics of their stem, e.g. (39). That is, Japanese possesses LIs which instantiate the *n*P-phase alone. [138, 139] When the same LI surfaces in the same

[138] Nothing hinges on precisely how we formulate the functional hierarchy and its 'phasing'. I am simply referring to the section of syntactic structure which excludes everything higher than the lexical semantics of the root and the categorial feature of the noun. Following Borer (2005a, b), there is clearly some functional structure associated with the property of individuation, between the N-head and Num-head. The lexical semantics of nouns only includes what Baker (2003) calls a "criterion of identity"; the "inner aspect" of "boundedness" is not

form, but with non-generic, numerated semantics, we cannot deny the existence of the outer D-φ-Num-'boundedness' phase (see fn.138), because its semantics is evident. Japanese (along with Korean, Chinese etc.) postulates a null LI on which [boundedness], [Num], [φ] and [D] are bundled (or rather a set of such LIs). [140]

Longobardi (2008) does in fact seek to deny the existence of the D-phase in DPDLs, however, assuming that Japanese *et al.* fail to grammaticalise φ-features at all. He takes a literal interpretation of the absence of visible person-number agreement on verbs. [141] This leads him to postulate a pragmatically-constrained 'type-shifting' operation specifically to reconstruct D-phase properties when they are semantically evident. Without going into the details of his account, we note that Longobardi makes four predictions (independent of the absence of visible person agreement), which, if they hold, he claims would be evidence in favour of his assumption. Taking two of these together, his account predicts that Japanese expressions translating Indo-European pronouns (DPs or φPs) and proper nouns (DPs) will have the same distribution as Japanese 'bare nouns'. Needless to say, these facts are also explained by the fact that Japanese actually *has* a (null) D-phase, instantiating definiteness and φ-features. His third prediction, that "bare nouns will be able to achieve kind-referential interpretation", is nothing more than a restatement of the fact that Japanese has nominal LIs which don't encode higher grammatical structure. His final prediction is that all quantifiers will "appear in floating positions outside the nominal phrase", i.e. not in D, where they would surface in most European languages (according to Guardiano & Longobardi (2006)) – the implication being that this would show that D does not exist. While this prediction is borne out in Japanese, it follows directly from an independent property of the Japanese lexicon, before consideration of grammatical deficiency is even relevant. Quantifiers in Japanese, Korean, Chinese, etc. can't surface in D because they don't have deterministic

---

included, as Wiltschko (2012) also observes (within a very different argument). Nouns may denote kinds or individuals, but the distinction is not semantically inherent. The distinction between 'mass' and 'count' nouns however *is* purely semantic: when 'bounded' (associated with a [boundedness]-feature), mass nouns may be countable (have a Num-head) and determinable (have a D-head), i.e. be individuals, while retaining mass semantics – "a cup of the salt", "a glass of those two wines", "I'll lie on any sand", etc. As is expected in a rigid, conceptually-determined hierarchy, the 'route' to the higher functional categories is through the lower ones.

[139] Chinese at least has analogous verbal LIs, giving rise to the verbal "High Analyticity" properties of Chinese discussed in Huang (2013).

[140] Languages such as Latin and Russian seem to bundle [Num], [φ] and [D] on the same (typically) overt head (although see Bošković (2008) for a dissenting view.)

[141] Cf. Saito's (2007) account of discourse *pro*-drop languages (DPDLs); also Roberts (2010: 85-7) and Holmberg & Roberts (2013a).

semantics, but rather are modified noun constructions. For instance, the parse of *san satsu-no hon-o* from (43) is something like "three bound object of book". It is semantically analogous to English expressions such as "three head of cattle" or "ten stem of roses", which clearly require an independent D: "the three head of cattle", "those ten stem of roses" etc. A Japanese quantifier would not therefore be expected to surface in D anyhow (nor would the lexical N-head be expected to project a D in this construction.) [142] In summary, there is no reason to suppose that the D-phase is absent or otherwise anomalous in Japanese therefore; in fact there is robust evidence that it exists, merely as part of a null LI.

Miyagawa (2010: 23-5) offers independent evidence to suppose that Japanese does grammaticalise φ-features, even if they are unmarked on the verb. In particular, he points out that Japanese has D(iscourse)-modal (Inoue 2006) elements, which likely occur in the C domain [143] and express the attitude of the speaker toward the utterance and the hearer. These D-modal elements appear to restrict the person of permissible subjects, i.e. agree with the subject. Exhortative D-modals, for example, only allow the first person:

(44) {Watasi/*Anata/*Yamada-sensei}-ga Taroo-ni tegami-o
I/*you/*Prof. Yamada-NOM Taro-DAT letter-ACC
okuri-MASYOO
send-let's
'Let's have me/*you/*Prof. Yamada send Taro a letter.'
(Ueda (2006: 168), cited in Miyagawa (2010: 24))

On the other hand, prohibitive D-modals only allow second person subjects, while negative supposition and assertion modals only allow first and third person subjects.

It is worth noting that Miyagawa (2010: 28) makes a suggestion regarding DPDLs' nominals themselves which points in the direction of the account suggested here. Chinese, another DPDL with 'bare nouns', also has robust person agreement of the "Indo-European" kind (see Miyagawa (*ibid*.: 46-52) for evidence), but, as Migayawa points out, this does "not negate the idea that the "kind" nominal cannot carry φ-feature agreement", suggesting that "the actual goal of the φ-probe in Chinese is an empty agreement head". He does not, however, go on to dismiss

[142] See Kobuchi-Philip (2010) for discussion of the status non-numerical quantifiers in Japanese, many of which clearly have the same underlying structure as numerical quantifiers.

[143] NB, there is good reason to suppose that C is the source of T's φ-features – see §5.1.

Chierchia's (1998) *nominal mapping parameter* (NMP), which parameterizes what bare arguments 'let' their languages denote (as assumed by Longobardi (2008)), leaving the "issue open".

Let us now summarise salient aspects of the distribution of null pronouns in Japanese.

Japanese may use a definite referential null pronoun: in any person, in the singular or the plural; and with either personal gender, or impersonally:

(45) Ø siken-ni otita
I/you/he/she/they/you/we exam-DAT failed
'I/you/he/she/they/you/we failed the exam'
(Neeleman & Szendrői 2008: 332)

(46) Haha-ga ataraii tokei-o katte-kureta-ga boku-wa
Mother-NOM new watch-ACC buy-gave-but I-TOP
sugu-ni Ø nukusit-simatta
soon it lose-PERF
'My mother bought me a new watch but I soon lost it'
(Tomioka 2003: 322)

Japanese null pronouns can also be used with indefinite referential semantics:

(47) Ken-wa kuruma-o kat-ta Erika-mo Ø kat-tta
Ken-TOP car-ACC buy-PERF Erika-also (a car) buy-PERF
'Ken bought a car and Erika bought one too'
(*ibid.*: 323) [144]

[144] Some scholars (*i.a.* Oku 1998) would analyse this as an example of DP-ellipsis and part of an alternative account of null arguments in DPDLs entirely. Their evidence would come from the possibility of 'sloppy' readings of certain null arguments, which are absent when those arguments are overtly realized. We cannot discuss this issue in detail here, but merely observe that the possibility of DP-ellipsis isn't inconsistent with the particular version of the *pro*-based account of discourse *pro*-drop presented here (not all null arguments need be of the same type), and also remind the reader of Tomioka's observation that discourse *pro*-drop and 'bare nominals' are intimately related, which a DP-ellipsis account offers no explanation for. As it is, Ahn & Cho (2012) (and papers referenced within) argue convincingly against DP-ellipsis accounts of sloppy readings in

Japanese null pronouns can surface just as easily as internal arguments as they can as external arguments, as shown in the previous three examples.

Japanese null pronouns can be part of a quantificational construction assigning them specific number:

(48) Ken-wa tabako-o inchinichi futa-hako suu-ga Erika-wa
Ken-TOP cigarette-ACC a day two packs smoke-but Erika-TOP
Ø ip-pon-shika suwan-ai.
(cigarette) 1-[long, thin object classifier]-except smoke-NEG.
'Ken smokes two packs of cigarettes a day, but Erika smokes just one cigarette.'
(*ibid.*)

The order of presentation of the properties of Japanese *pro* was intended to illustrate the close parallels between its behaviour and that of Japanese nominal phrases. Following the logic of Raposo's (1998) proposal for null objects in European Portuguese, [145] I suggest that the fact that *pro* surfaces in the same contexts as *apparently* bare LIs is an immediate reflex of the rich set of null [D, φ, Num, boundedness] LIs necessary to explain the latter behaviour. Apart from the lexical semantics of the root, an N-feature and a Case-feature, most possible combinations of DP-properties already exist as null feature-bundles as part of bare nominal syntax. For such a language to develop the equivalent set of null pronouns (i.e. discourse *pro*-drop) would entail nothing more than LA-ers postulating a null placeholder with trivial lexical semantics (say, 'person'-N or 'thing'-N) [146] and a Case-feature. The acquirer has very little extra 'distance to

---

DPDLs, and in favour of "deep anaphora" *pro* even in these crucial cases. (Tomioka's account on the basis of *NP*-ellipsis relies on the absence of the D-phase and the existence of 'type-shifting' operations of the Longobardi-type, both of which are disputed above – see also fn.146.)

[145] Rephrasing slightly, Raposo argues that European Portuguese has a null article LI equivalent to the 'definite' article used to mark type-semantics in other Romance languages. This can then take a null NP (his *pro*) as a complement to give a definite null object pronoun (when under anaphora).

[146] These elements of null arguments are also suggested in Barbosa (2011, 2013), following Tomioka (2003), although her account takes them as co-extensive with the null arguments of DPDLs, failing to license their semantic distribution without appeal to 'type-shifting' operations of the kind suggested by Tomioka (2003) and Longobardi (2008). In fact, the account here can be seen as a response to her closing observation that: "[a]s acknowledged by Tomioka (2003) this *[IJM: her/their]* hypothesis faces challenges. In particular, it requires a detailed examination of the distribution of bare nouns in a given language in relation to the conditions on the

cover' to acquire the full set of *pro* LI-bundles seen in DPDLs, specified as: [person-N or thing-N; bounded; singular or plural; 1st, 2nd or 3rd person; masculine, feminine, or impersonal; indefinite or definite; Case], when it already has the null LI-bundles specified as: [bounded; singular or plural; 3rd person; masculine, feminine, or impersonal; [147] indefinite or definite; Case]. [148]

There is no need to appeal to accounts of DPDLs which ignore independently-evidenced empirical facts about such languages, or which resort to engineering tricks to fill semantic gaps. This account merely observes extensive featural overlap between the LIs supporting discourse *pro*-drop and bare nominal syntax, universally present in DPDLs, and suggests a plausible acquisitional connection. This account is highly explanatory and shows the predictions of the fifth Minimalist proposal surviving a seemingly very tough test.

3.2.2.2 Permissible syntactic variation

An obvious problem for the strictest interpretation of the fifth Minimalist proposal is the apparent cross-linguistic variation in A-movement [149, 150] within syntactic structures. For instance, it seems difficult to deny that at least some languages' basic OV-order involves movement to a specifier position. Korean for instance has the canonical word order SONegV. As Whitman (2005) argues, assuming that Neg is merged outside VP, then the pre-Neg position of O must have been

---

licensing of NP ellipsis as well as *pro*-drop, a task that goes well beyond the scope of the present paper, but which we believe is worth pursuing" (Barbosa 2013: 49).

[147] Gender is null in Japanese DPs also: for instance, *isha* refers to a male or female doctor equally; even the literal translation of the formal terms for (someone else's) wife, *okusan*, and husband, *goshujin*, are gender-neutral, meaning 'the person inside' and 'the master' respectively.

[148] I won't speculate on exactly what the null D-phase-bundle would be in Japanese quantification expressions, possibly smaller than in unquantified, non-bare expressions, but the issue is not important, since the null *pro* for quantification expressions would be related to whatever this bundle is in the same way full *pro*-s would be related to full null D-phase-bundles.

[149] The status of head-movement in the FL remains vexed. We follow Chomsky (2001: 37-8) in his arguments that head-movement "fall[s] within the phonological component", although this commitment is not of significance to the broader claim here. However, for a summary of the arguments and a dissenting view, see Roberts (2010b). For an ambivalent position, see Chomsky (2014).

[150] Seemingly systematic variation in A'-movement is confounded by the existence of covert movement, introduced in §4.1.4 (see *i.a.* Cheng (2009) and Erlewine & Kotek (2014) for relevant discussion, and other references in fn.177), and by variation in feature inheritance, discussed later in this section.

produced by movement rather than underlying OV-ordering in Neg's complement (not necessarily precluded, but which would give NegOV ordering alone). Credible NS variation with respect to A-movement [151] does not accord with the fifth Minimalist proposal.

In the context of the argument here however, it is important to note that an A-movement operation *itself* is inherently vacuous semantically. [152] That is not to say that A-movement of an XP has no semantic implications: it will constrain reconstructive interpretations for instance. *Himself* may be construed as co-referential with *John* when in the raising construction (49b), but not in (49a), in accordance with Binding Principle A: [153]

(49) (a) *The problem$_j$ seems to himself$_i$ t_j to have been solved by John$_i$
(b) John$_i$ seems to himself$_i$ t_i to have solved the problem.

However, movement to the subject position is motivated by an EPP-feature on the unvalued φ-features of T, *not* by the lack of value of the φ-features themselves, which is resolved by the *Agree* relationship alone. (Reconstructive interpretation takes place after NS.) A-movement is "blind to semantic motivation, although it is not immune to semantic consequence" (Uriagereka 2002: 212). From the point of view of NS as grounding an internal 'system of thought', A-movement is permissible syntactic variation between externalized languages, because it has no impact on the satisfaction of interpretable features which ground the system's semantic productivity: EPP-features are movement 'diacritics' [154] which are postulated by LA-ers to make sense of their PLD, without bearing on their underlying syntactic commitments.

Further NS variation of the same 'permissible' kind is associated with the status of languages as *subject-* or *topic-prominent*. We have just seen English examples of movement to Spec-TP associated with φ-feature valuation, such that the subject systematically surfaces in the highest position in TP. In a range of other languages however, including Japanese, the sentence's topic/focus systematically surfaces as the highest non-null element of a clause. As Miyagawa

[151] Exactly how much depends upon one's commitments regarding the linearisation of syntactic structure. Linearisation will be discussed in §6.2, but the relevant issue, whether there is a universal underlying order, will necessarily be left unresolved.

[152] The semantic motivation for / role of *A'*-movement is discussed in §4.1.4. The diachronic origins of A-movement are discussed in §7.4.2.

[153] Anaphors must be locally bound.

[154] "Features-of-a-feature" in Pesetsky & Torrego's (2001) terms.

(2010) argues, these languages seem to allow [topic] and [focus] features to act as probes in various constructions in just the same way as φ-features. [Topic] and [focus] features are not restricted to a head in the C-periphery (as shown in Fig. 3), from where they can only trigger A'-movement (through association with an EF, see fn.126), but may move to a clause-internal position from where they can trigger A-movement. This is a possibility afforded by the same mechanism of feature inheritance under which T obtains its φ-features from C (see §5.1): there is no constraint against C 'lowering' other features, and so the LA-er may postulate [topic]- and [focus]-lowering to make sense of PLD.

Evidence for this claim is that fronted phrases suffixed by *-mo* ('also'), carrying focus stress, disallow reconstruction (*ibid.*: 64-5), suppress weak-crossover violations (*ibid.*: 67), and may bind lower anaphors (*ibid.*), all of which are all properties indicative of A-movement. Migayawa goes on to argue that [topic]/[focus]-features are inherited by a distinct head in the clause-internal structure, α, in some topic-prominent languages (e.g. Japanese), unlike others (e.g. Finnish) in which they are bundled together with φ-features on T. He even suggests that α may inherit both [topic]/[focus] and φ-features in some languages, such as Kinande and Kilega. The important point to extract for our purposes here though, is that NS variation may arise in connection with *how* a LA-er takes a substantive feature of syntactic structure to be satisfied, but not *whether* it is satisfied. The associated machinery affords different routes to conceptual equivalence when it interacts with the acquisition process. [155]

Another instance of genuine NS variation is suggested by Hinzen & Sheehan (2013: 197-201), who propose that the structural Case system of NS can be parameterized. They suggest that in a subset of syntactically-ergative languages, including Dyirbal, ergative is an inherent Case and *all* absolutive-bearing DPs (intransitive verb subjects *and* transitive verb objects) pattern in the Nominative Case position. This contrasts with the 'accusative' structural Case system displayed by other syntactically-ergative languages, such as West Greenlandic, whereby ergative DPs and absolutive-bearing subjects *alone* pattern in Nominative Case positions, with absolutive objects patterning in the Accusative Case position. In the former subset of languages, it is suggested that the *v*-head fails to assign any structural Case *at all*.

I am not in a position to comment on the accuracy of Hinzen and Sheehan's analysis and whether they identify a dimension of freedom actually available to LA-ers. I simply observe that such

[155] It is important to bear in mind that this is absolute conceptual equivalence, not merely having the "same thought" (Holmberg & Roberts 2013b: 20).

reanalysis would at least be plausible: ergative-absolutive marking provides the necessary ambiguity over whether the absolutive is associated with Nominative Case positions or Accusative Case positions. In as far as it is accurate, Hinzen and Sheehan's proposal fits the pattern of the findings for systematic NS parameterization: it is localized to non-semantic elements of NS apparatus, with variation arising under interaction with FLA.

In conclusion, I have suggested there are grounds to adopt a version of the fifth Minimalist proposal which emphasizes its continuity with the fourth Minimalist proposal. Narrow syntactic variation seems to be highly restricted, if not entirely precluded, with respect to CI-interpretable features. [156] Putative examples of non-universality seem to dissolve into variation in the availability of LIs ('bundling' parameters). Such NS variation as there is where you would expect to find it if language was selected to support an internal system of thought, in the absence of external data to license any underdetermination.

We have not discussed proposals regarding variation in externalization strategies so far (*pace* fn.149, 150 and 156), although we will be doing so at various points below. One thing that stands out in this regard though is that the scale of phonological variation is mostly 'sub-micro-parametric', associated with language-specific rules and phoneme inventories. The processes involved in externalizing language turn out to be so "intricate" that scope for variation under PLD ambiguity is too vast for generalizations of the kind grounding 'parameters' *per se* to emerge (except at the immediate interface with NS, where we find head-movement parameters, covert-movement parameters, and the Polysynthesis Parameter). Again this is in accordance with the expectations of the fifth Minimalist proposal, which I conclude has a good deal to recommend it.

We are now in a position to examine the various tasks of the FL for evidence of substantive and computational optimality and, by extension, evidence of domain-general cognition. That is, we

[156] Under the BCC-approach to syntactic variation, macro-parametric variation has received little attention, as variation on that scale seems incompatible with a limited UG (although see the approach discussed in *i.a.* Roberts & Holmberg (2010) and Roberts (2012)). The most notable exception to this is Baker's (1996*)* *Polysynthesis Parameter*. Space precludes us from discussing Baker's suggestion in detail, but the crucial observation is made by Chomsky (2012: 55), pointing out that the Polysynthesis Parameter seems to be fundamentally a parameter of linearisation: "it has to do with whether a sentence's arguments – subject, object, and so on – are internal to the syntactic structure, or [whether they] are marked in the syntactic structure … kind of like pronouns, and [so] hang around on the outside" (cf. Kayne 2005b: 7).

will now pursue the framework for pursuit of the MP presented in Fig. 2. [157] In as far as they relate to the MP, we will also be investigating the plausibility of the fourth and fifth Minimalist proposals.

[157] Which it must be remembered does not include the crucial comparative approach to the third Minimalist proposal.

4. *Minimize Time Complexity*

This chapter begins assessing the workings of the FL for evidence of optimal, domain-general cognition and principled divergence. In doing so, it synthesizes existing models of linguistic phenomena (and existing Minimalist assessments), as well as offering some original analyses. This investigation is structured around the framework of Fig. 2. We begin with the (sub-)principle of computational optimality, *Minimize Time Complexity* (*MinTC*), before moving onto the role of *Minimize Space Complexity* (*MinSC*) in following chapters. The principle of substantive optimality – *The Less, The Better* – and putatively principled breaches of optimality will be dealt with in discussion as they arise.

4.1 Minimalist phrase structure and 'spring cleaning'

We start by examining recent models of phrase structure, adopting an approximately chronological presentation for the purpose of structure. We discuss any potential implication of optimality or principled imperfection as we reach surviving theory, even though they manifest also be manifest in superceded proposals presented in passing.

4.1.1 Moving on and the Copy Theory of Movement

The section picks up §2.1's discussion of the earliest work in the MP. It will be remembered that a crucial piece of the GB argument against Case-marking as an LF phenomenon was that it must take place before Spell-Out, so that Case can be afforded an appropriate representation at PF. This issue disappears under the assumption that DPs have their Case-features specified *ab initio* and merely *checked* as part of a derivation. While the correct Case-feature must still be associated with the correct syntactic position, this may equally well be verified at LF as at the pre-Spell-Out level of syntactic representation, SS: either way, PF receives the same feature-bundle, fully-specified for the purposes of phonetic realization. Separate concerns diminished motivation for a DS level of representation (see Chomsky (1993: 20-1)), and so a way was sought to do without DS and SS in a new theory of phrase structure.

A new system was proposed in which an item was selected from the lexicon and projected to one of three possible X-bar templates: [X], [$_{X'}$ X], or [$_{XP}$ [$_{X'}$ X]]. This was the "sole residue" (Chomsky 1993: 21) of the Projection Principle [158] of GB theory, a natural development of the loss of its principal locus of operation. In abandoning DS, it also became necessary to find a way to account for the constraint against substitution into theta-positions, which was enforced in GB by the joint effect of the Theta Criterion and the Projection Principle. An independently-motivated [159] principle of phrase structure building was proposed which covers the same ground – the *Extension Condition* (EC). Phrase structure was built in this model by a generalized transformational operation, which added an empty position marker, Ø, to the overall phrase marker, before substituting it with an external SO, or a *copy* of an internal SO, introduced by *Move-α* (the lower copy being deleted in the phonological component). Movement built structure in just the same fashion as base-generation. The EC holds that the empty position added under a transformation must be external to the targeted phrase marker, that is, substitution operations must extend their target. A consequence of this is that SOs cannot be inserted into complement positions, barring raising into a theta-position; what remains of the Theta Criterion is now free to operate at LF alone.

Chomsky's (1993) initial revision of the GB picture clearly made progress in respect of substantive optimality, dispensing with DS and SS, the first two compositional cycles of syntax, [160] much of the Projection Principle, and the Trace theory of movement (TTM). NS continued, however, to rely on an X-bar-theoretic template [161] for projection of LIs (coupled with a reduced Projection Principle), a complex transformational operation introducing empty positions into the syntactic derivation, Relativised Minimality, [162] the EC, [163] etc. [164] The Visibility Condition, Theta Theory and the Binding Theory continued to apply at LF.

---

[158] The principle that sub-categorization information, i.e. information determining the way external SOs slot into phrase structure, must be represented throughout a syntactic derivation.

[159] See Chomsky (1993: 22-3) and §4.1.2 below.

[160] The covert movement cycle persisted under the then-current notion that movement was more 'complicated' than base-generation, and so had to be postponed until necessary.

[161] It is worth noting that the development of X-Bar Theory itself is an example of 'pre-Minimalist' methodological minimalism, seeking to eliminate the redundancy between lexical specification and phrase structure rules.

[162] See §4.1.6.

[163] See §4.1.2.

[164] Complemented by specious principles of global economy of the kind discussed in § 2.4.2, which we need not consider further.

Despite this complicated picture, it is not difficult to discern Minimalist themes emerging. We begin with consideration of the *Copy Theory of Movement* (CTM) (Chomsky 1993: 34-5), which supplants the previous *Trace Theory of Movement* (TTM). Under the TTM, the relevant element was raised and associated with a referential index, and a place-holder element (trace) identified with the same referential index was inserted in the phrase marker to mark its base-generated position, so that its interpretation could be 'reconstructed' at LF.

It is important to note that there are independent grounds to support the CTM. Most obviously, we find many constructions across languages in which lower copies are phonetically realized. Traces are null by definition, so this constitutes a convincing argument. Following Nunes (2011: 154), we take an example from Romanian, which fronts multiple *wh*-words, illustrated in (50):

(50) (a) cine ce precede?
who what precedes
(b) *cine precede ce?
who precedes what
'Who precedes what?'
(*ibid*., after Bošković 2002a)

However, this requirement seems to be waived when it would give rise to identical words adjacent within the same phonological phrase, a result disfavoured by the AP-apparatus. [165] In this case, the phonological system may elect to realize the object *wh*-word in its base-generated position, a readily available solution if it would ordinarily be deleting this copy anyhow: [166]

(51) (a) *ce ce precede?
what what precedes
(b) ce precede ce?
what precedes what
'What precedes what?'
(*ibid*.)

Bošković (2002a) goes on to provide independent reason to believe Romanian is not adopting *wh*-in-situ syntax (via some unknown mechanism) in such examples and really is spelling out a

[165] See further discussion of the tendency to avoid haplology in §4.3.4.

[166] See further discussion of why this is the ordinary state of affairs in §4.3.1.

lower copy. He shows that the *wh*-object in constructions such as (51b) is able to license a parasitic gap, just like moved *wh*-objects, but unlike in-situ *wh*-objects:

(52) Ce precede ce fără să influenteze?
What precedes what without SUBJ.PRT influence.3SG
'What precedes what$_i$ without influencing it$_i$?'
(Nunes (2011: 155), after Bošković (2002a))

Nunes (2011: 155-167) adduces further empirical support for the CTM from cases showing the 'scattered deletion' of multiple copies and full realization of multiple copies.

As suggested, the CTM eschews any need for a reconstruction operation of the kind found in trace-based approaches. Reconstruction is essentially a lowering application of movement, matching referential indexes; it does nothing to add information to the syntactic structure, in breach of *MinVO*, and so *MinTC*. The CTM introduces no vacuous operations of this kind, plausibly implicating *MinVO* and domain-general cognition in the NS. It must also be noted that the absence of traces, indices, and a reconstruction operation reduces the number of different symbolic structures and COs employed by NS, in accordance with *TLTB*.

We would also do well to look at the process necessary to insert traces into syntactic structure. This involves 'unmerging' a SO and 'remerging' a trace. This clearly tampers with existing derivation, in breach of the *NTC*, and hence *MinTC*. The CTM also means we are able to 'do without' these operations, a substantive saving in accordance with *TLTB*.

A further substantive saving is the absence of any restriction on 'copying' under the CTM, something like which motivates the TTM. It should be pointed out that the 'copying' operation of the CTM does not create separate, independent instantiations of the base-generated element: it merely creates an 'instruction' to refer to the previous 'link in the chain', with the lowest copy being 'labelled' as a point of reference / 'link in the chain' (cf. Chomsky 2008: fn. 17). This captures the results discussed in Sigurðsson & Holmberg (2008), whereby only the head of a movement chain, i.e. the *whole discontinuous element*, induces intervention effects, whereas the lower copy does not (Chomsky 2013: 44). It should be pointed out that this is economical from a

computational point of view, since otherwise movement would involve repeated generation of the same symbolic structures; [167] the CTM thus accords with *MinRedup* in this regard.

4.1.2 The Extension Condition

The *Extension Condition* (EC) (Chomsky 1993: 22-3) on syntactic derivation holds that all structure-building operations (base-generation or displacement) take place at the *root* of the tree, *extending* the phrase marker.

Evidence for the EC comes from the absence of super-raising constructions, violating Relativised Minimality constraints, as discussed by Chomsky (*ibid*.):

(53) (a) is certain [John to be here]
(b) seems [is certain [John to be here]]
(c) **John**$_i$ seems [is certain [__$_i$ to be here]]
(d) John seems [**it** is certain [__ to be here]]
(reproduced from N. Richards (2011: 176))

In (53a), a raising infinitive is created as the complement of a raising predicate *be certain*. In (53b), a secondary raising predicate is added on top of this predicate construction, without raising taking place. In (53c), *John* is raised to the subject of the matrix predicate, as it can be in the absence of an intervening DP. [168] In (53d), an expletive is inserted counter-cyclically in the subject position of the embedded predicate. This kind of derivation is unattested in language however, suggesting that the counter-cyclic structure-building is prohibited, and that *John* must instead raise to the root at (53a), with the expletive being added at the final stage of the derivation, generating the permissible 'It seems John is certain to be here'.

The EC is a reflex of two independent principles. First, it relies on the constraint that branching be *binary*, a matter discussed in §4.1.7 below and plausibly reflecting concern for computational optimality. Secondly, it relies on the *No Tampering Condition* (Chomsky 2001), which we have borrowed as the name of a general principle of computational optimality. The EC means there is

[167] In particular, it would then be anticipated that 'pied-piping' would be a marked option, and avoided unless necessary. As Donati (2006: fn.7) points out, this is empirically problematic for a range of constructions.

[168] Under *phase*-based derivation (see §5.1), necessarily via the specifier of the embedded CP (raising the question of improper movement here).

no need to 'undo' previous structure-building operations in order to open a space for a new object to be introduced. The EC serves to *Minimize Redundant Operations* in accordance with the *NTC*. It should also be noted that the EC means there is only one possible location at which operations building phrase-structure can apply, meaning there is no search for their locus, minimizing time complexity in accordance with *MinSearch* (Chomsky 2004c: 109). [169, 170]

4.1.3 Bare Phrase Structure

Chomsky's (1995b) second Minimalist account of phrase structure – so-called *Bare Phrase Structure* (BPS) – does away with much of the machinery of his 1993 model. In BPS, a LI is selected from the lexicon and added to the *Numeration*, the 'pre-syntactic workspace' in which LIs are associated with their optional formal features (φ-features, Case-features, Tense-features, etc.); LIs begin the derivation fully-inflected. These LIs are no longer projected onto X-bar templates, but project through the action of the structure-building operations themselves, which are motivated by the featural demands of LIs. A LI's sub-categorization information is only represented when the relevant operation takes place, and no sooner: the Projection Principle is now entirely superfluous. The X-bar template for projection is replaced by an 'empty' principle of grammar – the *Inclusiveness Condition* (*IncCon*) – which holds that the LF output of NS must consist solely of the features of the LIs found in the *Numeration*. The structure-building operations reflect the two "irreducible facts" of syntactic derivation. *Merge*, as we have seen, is necessary for unbounded nesting recursion to 'get off the ground'; in BPS it takes a LI from the Numeration and incorporates it at the root of the phrase marker. A second operation, *Move*, instantiates displacement in language and, again, is an operation which copies a SO from within the phrase marker, incorporating this copy in accordance with the EC.

---

[169] It should be pointed out that we have adopted the 'Strong Extension Condition' (Chomsky 2004c: 109) here. Another version of 'extension' has been suggested by N. Richards (1997, 2001), who argues on empirical grounds that the 'root' of the derivation for new SOs is the position as close to the head as possible, yielding "tucking in". This would *Minimize Search*, but lead to a narrow breach of *NTC*, although this would arguably be principled if the undeletable EF defines the edge as Richards proposes. The matter is ultimately an empirical one, which I am not in a position to resolve.

[170] The inheritance of features in *phase*-based derivation affords the possibility that structure-building operations motivated by the phase head and the non-phase-head apply in parallel, potentially yielding bifurcation of the root of a derivation. See §5.1 for further discussion.

BPS eliminates the previous redundancy between terminal elements and LIs, [171] as lexical entries already include categorial information. The indices associated with the different levels of X-bar templates can also be dispensed with: the syntactic notions of 'minimal' and 'maximal projection' accessed in the course of the syntactic computation fall out from structural context, as "relational" properties. A minimal projection is simply a LI selected from the Numeration; a maximal projection is a SO that doesn't project any further; intermediate projections are those SOs which are neither maximal nor minimal. The notions of complement and specifier are also reducible to relational properties of *Merge*/*Move* operations – being the directly incorporated element in the first instance, and a subsequently incorporated SO in the latter. We also no longer need to postulate separate nodes to stand for features relevant to future computation. [172] Finally, unlike under templatic approaches, vacuous intermediate projections are not projected at all: structure-building operations must be motivated by properties of LIs.

There are independent empirical grounds to support these suggestions, alongside the argument from methodological minimalism. Chomsky (1995b: 402-3) points out that clitics behave as both heads and maximal projections, a situation prohibited under templatic approaches, but trivial under a relational one. In its base-generated, thematic position, a clitic is an XP, whereas in its final position it is adjoined to a head as a head. [173] Chomsky (1995b: 399) also points out that BPS fails to distinguish between unergative intransitive verbs, such as *cough*, and unaccusative intransitive verbs, such as *choke* (see §1.3.3.4). Under templatic approaches these would have been assigned the structures (54a) and (54b) respectively.

(54) (a) $[_{VP}$ [NP] $[_{V'}$ $[_{V}$ cough]]]
(b) $[_{VP}$ $[_{V'}$ $[_{V}$ choke] [NP]]]

Whereas under BPS both become:

[171] Which in some sense preserved the redundancy noted in fn.161.

[172] With projection now being a theory-internal notion, the labelling of SOs must now be established by independent labelling algorithms. There are independent empirical grounds to suppose this is the case. At this stage, it was still considered necessary for labels to be established within the syntactic derivation, providing information required by future syntactic computation, as well as by the AP- and CI-interfaces. This position has been revised such that labelling is now considered part of the process of interpretation itself. See fn.133 and §4.1.8.

[173] See also Bošković (2002b).

(55) $[_{VP}$ V NP]

This loss of distinction is a virtue, since, as Hale & Keyser (1993, 2002) have argued, unergatives appear to be a kind of transitive predicate with a null internal argument: *cough* can be used in the construction *to cough a cough*. This is not the case with unaccusative verbs such as *choke*: you cannot **choke a choke*. It is therefore advantageous to be able to represent one sort of intransitive verb (unergatives) as a regular transitive.

BPS is highly economic, invoking no symbols other than those already present on LIs, and abandoning the X-bar templates and empty position markers, Ø, of previous approaches. The Projection Principle and complex transformational operations are abandoned completely, although a *Move* operation is retained. The fact we are now able to do away with many assumptions clearly accords with expectations of *TLTB*. [174] What's more, the fact that *Merge* doesn't appear to apply vacuously, projecting empty intermediate structure, is consistent with the force of *MinVO* in syntactic computation.

4.1.4 The Probe-Goal system of agreement and movement

The status of agreement and movement in early theories of Minimalist phrase structure was somewhat vexed. The 'checking' theory of movement (Chomsky 1993, 1995c) took movement as motivated by checking agreement relationships; it was considered that features could only be checked in a Spec-Head configuration, regardless of whether that configuration obtains overtly. Overt *Move* was considered a more complicated operation than *Merge*, consisting of *Move-F(eature)* and *Piedpipe*, and also more complicated than covert *Move*, taken to be simply *Move-F*. These considerations led to: (a) the 'procrastination' of 'costly' agreement by overt movement for economy reasons, creating a very rich covert movement cycle; and (b) the conclusion that languages vary in how 'economic' they are, since they use different amounts of overt movement to express a given meaning. The notion in (b) in turn suggested that displacement was a true imperfection of NS – "something beyond the conceptually necessary, as its absence from other symbolic systems would seem to confirm" (M. Richards 2009: 57).

[174] It is worth pointing out (perhaps slightly belatedly) that *Merge* (of something functionally equivalent) underlies any approach to phrase structure building, regardless of any richer assumptions made. At some level of abstraction, set-formation is required to construct the templates of X-Bar Theory.

Chomsky (2000) abandons the requirement that agreement must take place in a Spec-Head configuration, realizing that the same empirical coverage can be captured without this assumption, while avoiding: (a) the need for a separate covert movement cycle, (b) a principle of 'procrastination', and (c) the somewhat implausible conclusion that languages differ in efficiency. Checking was replaced with *valuation*, and a separation was created between the operation executing this, *Agree*, and movement itself. A *probe* bearing an unvalued feature will search its c-command domain [175] for a valued counterpart of that feature, the *goal*. If this search is successful an *Agree* operation may take place, valuing the unvalued feature and creating an agreement relationship between probe and goal. Movement only takes place if the probing feature is associated with a movement diacritic, or *EPP-feature*. [176] EPP-feature-less *Agree* now gives rise to long-distance *Agree*, largely replacing the rich 'theory-internal' cycle of covert movement motivated by the insistence on a Spec-Head agreement configuration. What's more, such covert movement as is independently-motivated [177] can now be captured simply by ordering Spell-Out before the relevant *Move* operations (see Nissenbaum 2000). This removes any motivation for the final NS-*internal* level of representation, LF, a saving in terms of *TLTB*. The remaining principles of LF can now be construed as aspects of interpretation. In particular, what remains of the Theta Criterion is simply the requirement that the CI-interface be presented with valid predicate-argument structures. The Visibility Condition / Case Filter amount to the requirement that uninterpretable features (Case-features in this case) be deleted before the interfaces. Similarly, Binding Theory is the process by which indices are referentially specified in the mapping to the CI-interface, at which they will otherwise be uninterpretable.

The next salient observation for our purposes is that *Agree* seems very similar in operational content to *Merge* (Chomsky 2000: 134). *Merge* is really something like: *Select* + *Match* + *Merge* + *Value (*c-*selection feature)*; *Agree* is something like: *Select/Search* [178] + *Match* + *Value (uF)*.

[175] Note, this is not the same as saying that c-command is a relationship computed in the derivation (cf. §2.4.2) – see discussion below.

[176] A property subject to parameterization – see §3.2.1. We are abstracting away from A'-movement here, since discussion of its instantiation prior to Chomsky's (2008) suggestion (see fn.126) takes us too far afield.

[177] For example: in relation to *wh*-questions, see *i.a.* Watanabe (2001) and Cheng (2009); in relation to scope, see *i.a.* Szabolsci (2001); and in relation to focus, see *i.a.* Erlewine & Kotek (2014) (and references within). As will be clear, covert A'-movement is widely proposed / attested (see fn.150) and covert A-movement less so (if at all – although see Polinsky & Potsdam (2013) for a putative example).

[178] As Chomsky (2000: 134) notes, "the selector F for Merge is analogous to the probe for Agree". See also: Rizzi (2008), for whom *Agree* may be thought of as *Internal Search*, while *Merge*'s *Select* may be thought of as

NS seems to make maximal use of operational content in accordance with the influence of *TLTB*. The lack of explanatory value in a separate covert movement cycle or a principle of procrastination is also consistent with *TLTB*.

With languages no longer taken to use different amounts of movement to express the same meaning, it became possible to think of *Move*, still an independent operation, as a principled imperfection (Chomsky 2000: 117-26). As Chomsky notes, displacement affords language greater expressive power. In particular, it allows language to provide the second of the "two kinds of information … required by the 'thought system' with which it interacts, namely (i) locally-required thematic relations (Theme, Patient, etc.) and (ii) edge-related information (such as new/old information, topic, focus, specificity, etc.) which are traditionally called 'deep' and 'surface' properties, respectively" (Chomsky 2004a: 164). That is, if language must satisfy the CI-IC of "duality of semantics" (*ibid.*: 165), *Move* can be thought of as a principled breach of *TLTB*. [179, 180]

The motivation for *Move* can plausibly be enriched by the speculation that it is the optimal way of encoding the duality of semantics. Other ways of marking this property can be countenanced, but they seem to involve more stipulations and greater computational complexity than *Move*. For instance, some kind of 'cyclic' *Merge* operation could instantiate dislocation, in which case we would have to generate phrases independently rather than through the CTM (many times in cases of successive-cyclic movement), identifying them with each other using indices and related operations, clearly substantively and computationally complex. Furthermore, some device would be required to distinguish these 'copies' from unrelated occurrences (Chomsky 2008: 145-6). *Move* may be the optimal solution to the duality of semantics, and so a 'doubly' principled breach of optimality.

External Search; and Pesetsky & Torrego (2006), who suppose a *Vehicle Requirement on Merge*, under which "if α and β merge, some feature F of α must probe F on β".

[179] Chomsky (2000) extends this argument to unvalued interpretable features also, but, as it is, these are only involved in semantically-vacuous A-movement, so require independent motivation. See §5.1 for such an account.

[180] Chomsky (2008: fn.29) offers an intriguing empirical argument in favour of the position that the duality of semantics really is a CI-IC (rather than an 'interface concern' – see fn.124). He points out the relative ease with which we come to understand non-linguistic duality of semantics, e.g. quantifier-variable notation for quantificational logic.

As it is, many elements of this discussion are rendered otiose by Chomsky's (2004c) reconceptualization of *Move* as *Internal Merge*. In the absence of any stipulation to the contrary, *Merge* ought to be allowed to 're-introduce' SOs internal to the derivation, as well as introducing LIs from the Numeration or external SOs (*External Merge*). There is no empirical reason to suppose such a stipulation, in accordance with minimization of the evolutionary burden and *TLTB*, and so displacement need neither be more complex than, nor distinct from base-generation. Nor must we postulate a *Move* operation to capture the duality of semantics, also consistent with *TLTB*.

We now return to the claim made above that probes / A'-movement-triggering phase-heads are limited to searching their c-command domains (Chomsky 2007b: 9). [181] The empirical motivation for this claim comes from the seeming absence of constructions which would be expected if 'upward search' were permissible. The grammaticality of (56a) can be explained by the fact that the *wh*-phrase is in the c-command domain of *v**, and so can be raised to the edge of *v**-phase unproblematically (a requirement of successive-cyclic movement – see §5.1). The same is not true in (56b) in which the *wh*-phrase is in the PP-complement of the specifier of *v**, and so not accessible to *v** for raising to the edge. C cannot extract the PP-complement from its existing specifier position directly because subjects are islands for extraction (Chomsky 1973), for reasons not fully understood but plausibly related to locality concerns (and so to *MinSearch*) (Chomsky 2008: 147).

(56) (a) [Of which car]$_i$ did [$_{v*P}$ they find [the driver t$_i$]]?
(b) *[Of which car]$_i$ did [$_{v*P}$ [the driver t$_i$] cause a scandal]?
(*ibid.*: adapted)

Further arguments against 'upward search' are discussed in Chomsky (*ibid.*: fn.36).

There are, however, various arguments in the literature that Spec-Head agreement remains necessary – see *i.a.* Baker (2008), Kayne (1989), Koopman (2006), N. Richards (2004), Zeijlstra (2012) – but it is not possible to do these claims justice and attempt a conclusion on this issue here. One passing observation regarding classic evidence for upward search is in order however. Kayne (1989, [2000]) and Baker (2008: 75-6) both discuss data from Kimball & Aissen (1971)

[181] We abstracted away from A'-movement above.

which they claim shows evidence of productive "bottom up" (*ibid.*: 243) verbal agreement in certain dialects of English:

(57) (a) the people *who* Clark *think* are in the garden
(b) the person who Clark think*(s) are in the garden

However, there seems to be a very plausible parsing account of such data, to be preferred in light of the fact that (57a) is only a *possible* production, alternating with the standard inflection. As discussed in Rizzi (2013: 175), a significant number of studies (*i.a.* Bock & Miller (1991), Franck *et al.* (2006), Franck, Frauenfelder & Rizzi (2007)) has shown that the plural specification of nominal adjuncts of singular subjects will be represented on verbal morphology as *misproduction* in a small, but stable proportion of elicited cases. This readily explains why in the same dialects discussed in Kimball & Aissen (1971) upward agreement with a singular *who* over a plural subject fails:

(58) *the man who the girls thinks is in the garden

Baker (2008: 76) resorts to an account involving markedness among φ-features to capture this distinction, a complication which is essentially 'theory-internal'. It seems far more plausible that the natural misproduction has become more acceptable (though notably not systematically grammatical) in certain dialects of English as result of sociolinguistic forces. These studies collect no relevant data in regard, and therefore classic evidence for upward agreement seems fairly unconvincing. (Cf. also discussion in Bošković (2007b), Preminger (2012), Van Koppen (2011), Zeijlstra (2012)).

However, without being able to conclusively resolve the empirical debate myself, I merely note that if Chomsky is correct in his claim that probes/phase-heads are limited to searching their c-command domains, this would be consistent with *MinSearch* (Chomsky 2008: 146).

4.1.5 Adjunction

As Chomsky (2004c: 117-8) observes, there is an operation building phrase structure which takes one SO and composes it with another, such that the latter retains all of its syntactic properties and the former has no role except at the semantic interface. Such *adjuncts* are, for instance, not subject to c-command relations for binding purposes, unlike their non-adjunct partners (see

Chomsky (2004c: 117)). This "separate plane" (*ibid.*: 118) property of adjuncts must be captured in the nature of the adjunction operation. *Pair-Merge* (discussed in §2.4.1) is an asymmetric operation, yielding *ordered* pairs, in which only one member is visible for future operations: *pair-Merge* (A, B) = {A, {A, B}}. B, the adjunct, is 'buried' by the operation and only visible at the interface. *Pair-Merge* is a more complicated CO than symmetric *set-Merge*, which has no 'discriminatory' content, but it is necessary to explain the empirical facts.

The 'extra' content of *pair-Merge* over set-formation is a breach of *TLTB*. [182] Chomsky (*ibid.*) suggests that there may be a CI-IC which requires "the existence of a device to yield predicate composition". He notes, however, that this is merely "a promissory note, given the limitations of understanding of C-I". Some suggestive work in this regard comes from studies of animal communication systems. For instance, suricates will give different alarm calls for mammalian, avian and reptilian predators, but vary their calls depending on the proximity of the threat (Manser, Bell & Fletcher 2001); the same signal conveys information about external reality and motivational factors. There seems to be some kind of asymmetric conceptualization to the signal, akin to an adjunction structure, since the behaviour elicited by the signal is constant (appropriate to the type of threat), but varies in commitment in accordance with specification for urgency (under experimental conditions in which the threat isn't actually present, the behaviour lasts longer the greater the urgency encoded in the call). [183] This perhaps cashes out Chomsky's "promissory note" a little. Either way, it seems highly plausible that *pair-Merge* is a principled breach of the *TLTB*.

4.1.6 The Minimal Link Condition

Chomsky (1995c: 297-312) adapts Rizzi's (1990) representational notion of *Relativised Minimality*, proposing a derivational, movement-oriented counterpart:

(29) *Minimal Link Condition* (MLC)
K attracts α only if there is no β, β closer to K than α, such K attracts β.
(Chomsky 1995c: 311)

[182] As of course is *set-Merge* itself, principled in that it grounds recursive hierarchy in conceptual structure (see §3.1.1).

[183] Similarly, it is thought that calls of chimpanzees indicate both the presence of food and its quality (*i.a.* Slocombe & Zuberbühler (2005, 2006)).

The notion of intervention referred to in this principle is 'relativised' to the featural content of SOs, such that we can restate the MLC as follows:

(59) *In the configuration:*
... X_{+F} ... Z_{+F} ... Y_{+F} ...
X_{+F} cannot attract Y_{+F} if there is an element Z_{+F} specified with the same feature +F and closer to X than Y.
(Rizzi 2011: 223)

Chomsky (2001) revises the MLC further, as a locality condition holding on *Agree* (and by extension the equivalent operation for A'-bar movement). For example, as shown in §3.1.2, it is this effect that blocks a SO bearing a quantificational feature from raising across a higher SO also bearing a quantificational feature.

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

The string in (60) 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.

As pointed out by Chomsky (2001), the MLC imposes narrow constraints on the extent of possible search operations instigated by a probe/phase-head, and this minimizes the time complexity of the syntactic algorithm in accordance with *MinSearch*.

4.1.7 Binary branching

As discussed in §2.3, *Merge* must be able to put *at least* two SOs into a set together to form a new SO (or self-similar structure) – this being the minimum required to establish recursion. There is empirical reason to believe that the 'arity' of *Merge* is limited such that no more than two SOs are ever joined in one set.

Evidence comes from a range of constructions across languages (see in particular Kayne (1984)); however, one construction which seems particularly amenable to a ternary branching analysis, and so proves very instructive to examine, is the ditransitive construction of English, discussed in §1.3.3.4.

(61) (a) I sent the book to Peter.
(b) I sent Peter the book.
(c) I [sent [the book] [to Peter]] / I [sent [Peter] [the book]]

It turns out that a binary branching analysis of these constructions has significant empirical advantage over the ternary branching analysis in (61c). As Barss & Lasnik (1986) argue, in the double object construction (DOC), (61b), the goal (*Peter*) asymmetrically c-commands the theme (*the book*), as revealed by anaphor binding: [184]

(62) (a) I sent Peter himself.
(b) *I sent himself Peter.

This observation led Larson (1988) to propose his famous, binary "VP-shell" analysis of DOCs, with the goal asymmetrically c-commanding the theme. A version of this idea was adopted by Marantz (1993) (and, subsequently, Harley (1995), Bruening (2001) and Pylkkänen (2002)) in a way that also captures the difference in argument structures between DOCs and *to*-dative constructions (TDCs), discussed in §1.3.3.4.

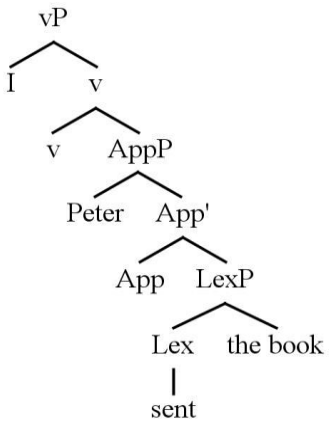
DOCs, but not TDCs, may have a causative interpretation. Relative to (61a), (61b) emphasizes my role in Peter getting the book. This effect is evidenced more clearly in (14a) and (14b), repeated from above, where the *to*-dative is odd, because it cannot convey the clearly intended meaning, that the article was directly responsible for my headache.

(14) (a) ??The article gave a headache to me
(b) The article gave me a headache.
(Miyagawa & Tsujioka 2004: 2)

Using evidence from the surface realizations of other languages, including Bantu, Marantz (1993) argues that DOCs have a second verbal head, an applicative (V_{Appl}), dominating the lexical verbal head (V_{Lex}). *Give* begins in V_{Lex}P, selected by the applicative head V_{Appl}, which relates the event of V_{Lex}P to the goal *Peter*.

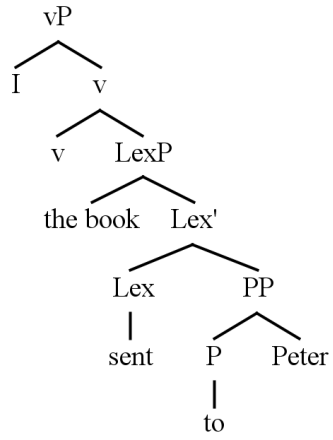
[184] Also by licensing of negative polarity items (Barss & Lasnik (1986), discussed in Miyagawa & Tsujioka (2004: 3)).

(63)



In the TDC, however, V_{Appl} does not appear and the simpler structure does not give rise to the causative and possessive interpretations.

(64)



A third consideration in favour of a binary branching analysis of ditransitive constructions is pointed out by Radford (2004: 263). (61c) wrongly predicts that *Peter the book* and *the book to Peter* will not form constituents and that it will be impossible to co-ordinate them with analogous strings. This prediction is not borne out:

(65) (a) I sent Peter the book and John the violin
(b) I sent the book to Peter and the violin to John.

Having established some grounds for assuming branching is obligatorily binary in language we now ask why, since *prima facie*, this constraint is a breach of *TLTB*. Several suggestions have been made in this regard.

One argument is that languages avoid branching of greater 'arity', because under probe-goal agreement this would greatly complicate search. Under *multiple-Merge*, there would be more than one possible 'path' down which to search for the goal, and each would have to be pursued, multiplying redundant search operations. Binary branching plausibly, therefore, reflects *MinSearch* (Chomsky (2004a: 168), (2004c: 115)).

Chomsky (2005b: 6) also notes a personal communication from Luigi Rizzi, who suggests that compulsory binarity might result from "minimization of search in working memory", also expressed by Chomsky (2008: 138) as the "restriction of computational resources". Rizzi clarified this suggestion in a response to my own communication: his point is that "binary Merge needs two cells in operative memory, [whereas] ternary Merge would need three, etc.". In other words, he is suggesting that *multiple-Merge* would breach the principle of *MinCUI*.

A final optimality-motivated proposal is made in Watumull (2010, 2012) (in the context of a radically different account), where he points out that *multiple-Merge* introduces the sub-operation of "bracket erasure" to ensure *associativity*. That is, at an 'early' stage of the *multiple-Merge* operation a set including two SOs will exist, but this set will have to 're-opened' (bracket erasure will have to take place) to allow the incorporation of another SO; its status as a set must be destroyed, a clear violation of the *NTC*.

Some possible IC-motivations for binarity are put forward by Chomsky (2005b, 2008). He suggests that our CI-apparatus may be attuned to input introducing one conceptual relationship at a time. We have seen some suggestive evidence in this regard in §3.1.3's discussion of the ancestry of event structure, [185] but for reasons of ignorance it can be considered no more than that. Chomsky also suggests that requirements of linearisation might impose binarity, with the modules of linearisation possibly being unable to compute order over non-binary structure. There

[185] Seemingly adopting an [actor-[action-goal]] schema (see §3.1.3).

is some plausibility to this suggestion, since the larger the set that must be linearised, the more complicated it will be to identify set members consistently, the basis of consistent linearisation. However, again, this must remain speculation, given our (relative) ignorance regarding the nature of linearisation.

A very different approach to this issue is taken in Watumull (2010, 2012). Watumull contends that binary *Merge* is favoured because any alternative operation would require a mathematically larger function – specifically, a "counting" function. Binary set-formation is "effectively unary" in that it need only "saturate" its "domain": its "co-domain" (the existing derivation, or empty set at the base) is "antecedentally saturated". He argues that *Merge* is binary "by design", and *multiple-Merge* is ruled out by constraints on possible mathematical functions, rather than by constraints on computation.

While it is a point well-taken that a more complex function would be associated with greater computational complexity, I think it would be misguided to take this account as *explanatory*. There simply isn't any relevant evidence to enable us to determine whether the function instantiating the *Merge* operation is underlyingly unary, or reduced to a non-counting function by other concerns, "ex post facto". It would not be a "violation of virtual conceptual necessity" for the latter to the case, as Watumull suggests, because we have no way of knowing what the neurological and cognitive content of "necessity" is in this instance. This is same uncertainty discussed in §2.4.1, regarding the constraints imposed on the FL by the "properties of the human brain that determine what cognitive systems can exist". To fly just one of the (currently more or less infinite) unfalsifiable kites, there may be some key cognitive utility associated with a function which operates over more than two objects, and so cognition may be optimized for, and obligatorily consist of, counting functions. Or perhaps there is some difficulty instantiating non-derived "saturation" of a "co-domain" in the brain, on account of some unknown property of neurological matter. And so on. [186] The Minimalist argument is not that any kind of imperfection is somehow disfavoured: instead it emphasizes the role of continuity with existing machinery, whatever that may be.

In summary, however, there is substantial reason to believe that binary branching in NS is at least partially a reflection of concern for computational optimality, possibly in terms of space

[186] For these kinds of reasons, Chomsky (2004a: 172-4) cautions against seeking scientific explanation in terms of mathematical "beauty" except for the most basic science, dealing with the very simplest objects, i.e. particle physics, suggesting that mathematics "doesn't work for anything [else :] there's no mathematics in biology".

complexity as well as time complexity if Rizzi is correct. It may plausibly also reflect ICs. As such, the constraint that *Merge* be binary (if it exists – we do not know) is a principled breach of *TLTB*.

4.1.8 Labelling

As discussed in §4.1.3 (in particular fn.172), for a SO to be interpreted at the interfaces it must be possible to identify what kind of object it is. [187] Under templatic approaches, this process of 'labelling' SOs was previously part of the operation generating phrase structure itself. However, BPS (and its empirical basis) suggests this is not the case and that labelling must be established by independent processes.

Chomsky (*i.a.* 2008: 14) suggests that the "labelling algorithms" are as follows:

(66) The labelling algorithms:
  (a) *In {H, α}, H an LI, H is the label.*
  (b) *If α is internally merged to β, forming {α, β} then the label of β is the label of {α, β}.*

If valid, these algorithms are consistent with the involvement of *MinSearch*: they are the simplest algorithms possible, never looking further than the previous operation for the relevant information, and always searching for a single item. Some evidence for this account comes from the 'problematic' nature of constructions over which these minimal algorithms are indeterminate.

Minimally-searching labelling algorithms are inconclusive when two complex SOs are *Merge*-d, creating SO = {XP,YP} (Chomsky (2008: 146-7, fn.34), (2013: 43-6)). The prediction of (66), therefore, is that either (a) such constructions cannot remain intact and must be modified to leave one visible head (reinterpreting Moro's (2000) *dynamic antisymmetry*), or (b) that X and Y must be relevantly identical, such that they provide the same label. The evidence is that this prediction is borne out, providing strong evidence for the minimally-searching algorithms and implicating *MinSearch* (Chomsky 2013: 43).

[187] In terms of semantic properties at the CI-interface, and categorical properties at the AP-interface.

An example of the modification alternative is provided by 'split topic' constructions (STCs) from German and other languages, shown in (67):

(67) Nagetiere hat Peter nur zwei Eichhőrnchen gesehen
rodents has Peter only two squirrels seen
In terms of rodents, Peter's only seen two squirrels.
(*ibid*.: 44)

Ott (2011) demonstrates that STCs seem to be derived by the raising of the "Nagetiere" part of the topic from the small clause SC = {zwei Eichhőrnchen, Nagetiere} (DP, NP), rendering the SC labellable. He motivates the necessity of this analysis with various properties of STCs. To take one example, it is not necessary for the DP and NP to match in number, suggesting they are generated separately, rather than as discontinuous parts of the same constituent. For the purposes of predication it is not relevant whether the NP is singular or plural, as it denotes a property, and so the following is possible:

(68) Zeitungen kenn ich nur eine gute
newspapers know I only one good
'In terms of newpapers, I only know one good one.'
(Ott 2011: 111)

The prediction that the {XP,YP}-structure is impermissible (without extraction) unless X and Y are identical in the relevant sense can be seen in the behaviour of *wh*-in-situ constructions of English, following Chomsky (2013: 44-5). In marked contexts, the verb *think* may take a complement with low *wh*-in-situ phrase as part of a direct question:

(69) They think Brentford is found in which city?

The verb *wonder* may take an interrogative CP complement with a *wh*-in-situ phrase as the specifier of CP:

(70) They wonder [$_{\alpha}$ in which city [$_{\beta}$ C [Brentford is found]]]

In this case, a {*wh*-P, CP} complex obtains, which begs the question of how it is labelled. It is plausible to suggest that interrogative *wh*-phrases bear a Q-feature (Cable (2007, 2010), Narita

2011), shared by interrogative C. As such, minimally-searching labelling algorithms encounter no ambiguity and can label complex indirect questions for interpretation. This also explains why the *wh*-phrase is obligatorily low in (71)-type constructions: if the *wh*-phrase were in the Spec-CP it would be unlabelled, because *think* does not select an interrogative CP to match its Q-feature:

(71) *They think [$_{\alpha}$ in which city [$_{\beta}$ $C_{indicative}$ [Brentford is found]]]

It seems, therefore, that the predictions of the minimally-searching labelling algorithms proposed in (66) are borne out, suggesting that *MinSearch* is indeed engaged.

Two final points are worth noting in regard of labelling. As observed in fn.133, it must now be taken to be a part of interpretation in itself, rather than an NS operation licensing future syntactic derivation involving the labelled phrase (as in *i.a.* Chomsky (1995c: 256-60)). As reported in Chomsky (2013: 30), Samuel Epstein observes that this cannot be the case if constructions such as (67) are to be permissible at all, because their generation involves the operation *Merge* (Z, {DP, NP}) = {Z, {DP, NP}}, i.e. *Merge* to an as-yet unlabellable SO. This is a favourable result in terms of concern for optimality, because labels need not be carried over in WM during NS derivation, in keeping with *MinCCD*. There is also a sense in which the need for a label to license a syntactic operation would duplicate the activity involved in constructing the appropriate array of particularly-specified LIs relevant for a particular derivation. Some activity of the utterance planner (part of the mysterious question of "the 'creative' aspect of language use" (Chomsky 1972: 11) – see fn.273) has caused NS to 'countenance' only particular operations in the first place, rendering licensing by a label somewhat redundant. The locus of labelling accords with *MinVO* (assessed locally, as it must be, in NS computation).

4.2 Lexical categories

From the earliest days of generative grammar investigation seemed to reveal "reason[s] for suspecting that grammar cannot be effectively developed on a semantic basis" (Chomsky 1957: 101 fn.9) and that "only a purely formal basis can provide a firm and productive foundation for the construction of grammatical theory" (*ibid.*: 100). Chomsky (1957) does not discuss the difficulties associated with semantically-oriented proposals regarding all elements of grammatical form, observing it would be "rather pointless" (*ibid.*). He restricts himself to refuting the intuitions that "the grammatical relation subject-verb … corresponds to the general 'structural

meaning' actor-action" and that "the grammatical relation verb-object … corresponds to the structural meaning action-goal or action-object of action" (*ibid.*: 94). In the former case, he points to sentences such as "John received a letter" or "the fighting stopped", and in the latter case to sentences such as "I will disregard his incompetence" or "I missed a train" (*ibid.*: 100).

Although Chomsky (1957) does not discuss lexical categories, the understanding that they do not reflect particular semantic distinctions is implicit. [188] Attempts have persisted to the present day, however, to motivate the semantic significance of lexical category features (LCFs). [189] In particular, it has been claimed that LCFs afford "particular *interpretive perspectives* [italics author's own] on concepts" (Panagiotidis 2011: 371), differentiating noun-verb "pairs" such as $_{N}$*work* – $_{V}$*work* without reference to "*inherent properties of the concept itself as expressed by the root* [italics author's own]" (*ibid.*: 373). Similarly, Embick (2012: 82) claims that a "[r]oot is interpreted in the context of a category-defining morpheme *x*", which instigates a "semantic operation" (*ibid.*: 83). (Cf. Borer (2005a, b) for a similar approach.) Again, the upshot is that homophonous verbal-nominal(-adjectival) sets must employ the same root underlyingly.

This is not an appealing assumption. Semantic approaches to LCFs would have us postulate radical polysemy of the underlying conceptual roots and extremely rich content of the LCF-triggered 'operation'. Take the noun – (denominal) verb pair $_{N}$*stone* – $_{V}$*stone* for instance. There are infinite possible predicates based on $_{N}$*stone*: not just 'to throw/hurl a stone at someone as punishment', but also 'to carve someone out of stone', 'to put someone on a stone', 'to balance a stone on someone's head', etc. The suggestion of semantically-oriented approaches is the rather strange one that the root of $_{N}$*stone* contains the semantics 'a stone when hurled at someone', which never surfaces except through the imposition of the "interpretive perspective" associated with verbalization. This operation must, therefore, be as rich and varied in character as attested relationships between nouns and their denominal verbs, that is, *very* rich and varied. With $_{N}$bronze – $_{V}$bronze, the operation imposes a reading something like 'to coat with [the noun]', whereas with $_{N}$*air* – $_{V}$*air*, the verb means something like 'to expose to [the noun]'. And so on.

[188] The matter is also discussed indirectly in 'Remarks on Nominalization' where Chomsky observes that derived nominals have meaning of an "idiosyncratic character" in relation to their associated verb (Chomsky 1970: 189).

[189] Whether encoded on a separate syntactic head, as in Hale & Keyser (1993), Marantz (1997), etc., or otherwise.

Embick (2012) is forced to appeal to extraneous notions of convention and use to account for the absence of the particular readings of verbalized concrete nouns which accrue in other cases. For instance, $_{N}$*cup* – $_{V}$*cup* demonstrates an instrumental reading of the noun in English, which is impossible for $_{N}$*bowl*, which 'ought' to be interpretable in a similar way. To explain this fact, Embick (*ibid*.: 85) points to the absence of any "convention that associates a "canonical use"" of $_{N}$*bowl* with a verb.

This solution has the distinct flavour of an engineering trick to preserve an account. These problems essentially arise from a commitment to interpreting undoubted etymological (hence diachronic) relationships as part of the synchrony of grammar. Other than intuition, there is no reason to assume that the root (and underlying conceptual content) is consistent across homophonous sets and necessitates a complex systematic account of meaning creation / filtering. It seems much more plausible to suggest that the root in $_{V}$*stone* is conceptually richer than that in $_{N}$*stone*, and actually contains the semantics of hurling and punishment, with there being no nominal stage in the derivation whatsoever. *Category-changing* morphemes, [190] such as *-al*/*-ity* in sets such as *tribe*/*tribal*/*tribality*, must have their own semantic content to determine what is added to the meaning of the noun/adjective to make it a modifier/potential argument – in this case, something like 'having the property of X' and 'the property of being Y'. [191, 192]

So, if we adopt the position that modifier/argument/predicate semantics is a property of roots (and semantically-enriched, category-changing morphemes), what then is the role of LCFs in syntax? It is clear that they exist, because we see evidence for them at PF and in NS in c-selection violations, etc. How do we justify this breach of *TLTB*, given that it is clear we cannot appeal to the CI-interface? Predicate/argument/adjunct-roots do not need further specification to allow

[190] When merged immediately to the root itself, category-designating morphemes. A further problem for approaches that assume semantically-meaningful null morphemes derive denominal verbs of the kind discussed above is that, under common assumptions, the nominal phase will already have been dismissed under phasal Spell-Out (see §5.1) (following Marantz (2007)), and so be inaccessible for direct semantic manipulation (see Panagiotidis (2011: 367 fn.1), Embick (2012: 80-1)). We would therefore have to assume that the full set of meanings of the necessarily-polysemous root is preserved by the null nominalizing head and sent to the CI-interface for interpretation before further specification by the null *v*. This is again somewhat suspicious.

[191] For discussion of different 'flavours' of *v* and *n*, see Folli & Harley (2005) and Lowenstamm (2008) respectively.

[192] It is worth pointing out that this is all orthogonal to the notion of "idiomaticity" discussed by Marantz (1997), which takes syntactically-derived meanings and reinterprets them in accordance with an "Encyclopedia" of non-compositional readings (cf. Embick (2012: 83 fn.6) and Marantz (2010)).

them to be *Merge*-d in meaningful structures: s-selection alone may license *Merge*. [193] However, s-selection-based licensing of *Merge* would be a complex operation, involving detailed inspection and comparison of the semantic properties of roots. In other words, to motivate a *Merge* operation without c-selection, we would essentially have to perform the same interpretive computation as must be independently performed at the CI-interface, clearly a radical breach of *MinRedup*. This, of course, is avoided under c-selection, with the licensing of a *Merge* operation only involving checking a single feature of SOs, a huge saving in terms of TC.

Essentially following the suggestions of Baker (2003), I suggest that LCFs recapitulate the most fundamental semantic properties of their associated roots. The V-feature designates a LI as a predicate with the power to license an argument (*ibid.*: 23*ff.*). An N-feature designates a LI as fulfilling the 'criterion of identity', and so able to support the 'referential index' required to assume an argumental/thematic role (*ibid.*: 95*ff.*). [194] Adjectives/adverbs are neither nominal nor verbal, and A-features indicate their role in secondary composition with an argument, predicate, event, or other secondary composers (*ibid.*: 190*ff.*). [195] As well as being used attributively, adjectives, like nouns, may be used predicatively. Baker (*ibid.*: 34-46) proposes a further syntactic category *Pred*, which when composed with adjectives (and nouns) gives them the ability to predicate, in line with the argument of *i.a.* Bowers (1993). To this end, Baker presents empirical evidence from Edo and Chichewa, in which he suggests *Pred* is overt. [196] Baker (2003: 88-94, 169-89, 238-63) and Chung (2012) argue convincingly that these categories are universal. This is unsurprising given our assumptions about the nature of the CI-ICs: the lexical categories of language merely mirror the demands of the interface – providing predicate-argument structure (including arguments which are also predicates) and secondary composition. This is a straightforward response to Embick (2012: 77) and others' question as to why languages converge on the same set of lexical categories cross-linguistically.

[193] On current assumptions, roots can already select arguments and serve as arguments (see Marantz (2001)).

[194] Phrases of *functional* syntactic categories, which have non-reduplicative, 'actual' semantic content, may also bear referential indices and be arguments, such as DP, CP and TP (cf. Baker (2003: 98 fn.3)).

[195] Whether we need to recognize separate lexical categories for adjective and adverb, or for the different kinds of 'adverb', composing with TP/VP/AP etc., is a topic we leave to one side (see Baker (*ibid.*: 230-7)).

[196] Baker (*ibid.*: 303-25) suggests that adpositions take arguments, but are not true predicates (nor arguments). Instead he suggests that they are a particular way of forming a relational adjunct, and so constitute a relational functional category.

This section has argued that lexical LIs have their primary semantic characteristic reduplicated syntactically for reasons of economic computation and, as such, that LFCs are a principled breach of *TLTB*. The upshot of this of course is the possibility of grammatical nonsense of the *colorless green ideas sleep furiously* variety.

4.3 Mapping to PF

4.3.1 The highest copy

Chomsky (*i.a.* 2005b: 13) frequently notes that the *Internal Merge*-based displacement by which NS instantiates duality of semantics creates multiple copies from the view of mapping to the AP-interface. FL may either elect to realize only one of these copies, minimizing the operational load in phonological computation and cognitive / physical cost of serial motor activity. [197] Or it may elect to realize all of these copies to avoid the difficulties associated with establishing *filler-gap* dependencies in parsing, [198] but maximizing these other costs. In a system optimized for externalization, the latter choice would be anticipated. As it is, the former choice is the one overwhelmingly preferred, with the highest copy being the copy realized. There are principled exceptions to this, of the kind discussed or referenced in §4.1.1 (cf. also Chomsky's (*ibid.*) discussion of Landau (2006)), but the choice made is consistent with the role of *MinRedup* in the phonological computation and the third-factor constraint minimizing motor activity costs. One further comment is warranted in this regard. It is noticeable that (except in unusual circumstances) it is always the *highest* copy of the displaced SO which is phonetically realized. This is a trivial reflection of the IC of useability: it is important that it be the highest copy that is realized to indicate the identity of the probe – the filler for parsing purposes. In other words, the highest copy must be realized to show what actually took place syntactically (cf. Chomsky (2007b: fn.17)). [199]

[197] See fn.79. The cognitive / physical cost of co-ordinated motor activity may be added to the third-factor 'economy' constraints putatively active in the FL.

[198] See §6.1.3.

[199] Cf. Nunes' (2004) suggestion that chain reduction takes place for the purposes of unambiguous linearisation instructions under Kayne's (1994) *Linear Correspondence Axiom* (see discussion in §6.2), and that the highest copy of the chain is preserved to reduce the number of unvalued PF-interpretable features which would otherwise have to be deleted before the derivation is presented. This approach seems to run into 'look-ahead' problems of the kind discussed in §2.4.2.

4.3.2 Strict cyclicity

The *strict cyclicity* (Chomsky 1973) of phonological computation mirrors the EC of NS. (Cyclic) phonological rules may only apply at the 'root' of the derivation / within their cyclic domain. The classic demonstration of this property of phonological rule application comes from Chomsky & Halle's (1968) discussion of the different second syllables of *compensation* and *condensation* in certain dialects of English, presented succinctly in Kenstowicz (1994: 204-5), whose exposition I follow here.

In the relevant dialects of English, the second syllable of *comp[ə]nsation* is unstressed and so its vowel is reduced to a schwa. On the other hand, vowel reduction fails to obtain on the second syllable of *cond[ɛ]nsation*, its pronunciation with a schwa being unacceptable. It would be impossible to explain the distinction between the two surface representations if their derivations began with [compensat + ion] and [condens + ation], since these representations are essentially equivalent. An explanation is afforded, however, by the observation that *cond[ɛ]nsation* derives from *condense*, whereas *comp[ə]nsation* derives from *compensate*. The cyclic application of the stress rule, shown below, means that *condensation* will have stress placed on its second syllable in an earlier cycle, blocking the reduction of its vowel. However, since the stress rule does not apply stress to the medial syllable of *compensate*, its vowel may reduce. Subsequent stress neutralization then leaves this as the only contrastive feature.

(72)

[condens]ation	[compensat]ion	
		first cycle
[condense]	[compensate]	
condénse	cómpensàte	stress rule
		second cycle
[condéns]ation	[cómpensàt]ion	
condènsátion	còmpensátion	stress rule
		later rules
----------	còmpənsátion	vowel reduction
còndɛnsátion	----------	stress neutralization

(reproduced from Kenstowicz (1994: 204))

Cyclic application of phonological rules is the essence of this account, and regardless of how the domains are 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: there is no tampering with the output of previous rules, nor any search for the site of rule application, in accordance with the demands of the *NTC* and *MinSearch*.

4.3.3 Nevins (2010)

Nevins (2010) attempts to motivate a new account of vowel harmony patterning in the worlds' languages. Nevins rejects the standard position that vowel harmony is instantiated by language-specific, contextually-driven phonological rules, such as the Turkish labial/roundedness and backness harmony rules:

(73) $\underset{[+high]}{V} \rightarrow [\alpha round] / \underset{[\alpha round]}{V} \; C_0 \_\_\_\_\_$

(74) $\underset{[+high]}{V} \rightarrow [\alpha back] / \underset{[\alpha back]}{V} \; C_0 \_\_\_\_\_$

He also rejects the associated position that rules of vowel harmony are intimately related with the interaction between "certain phonetic tendencies of the articulators" (*ibid.*: 9) and the process of acquisition. Anticipatory and perseverative co-articulation are well-established phenomena in phonetic production, and so Ohala (1994: 491) (and others) adopt the intuitive position that "[v]owel harmony, phonological co-occurrence constraints between the features of vowels in polysyllabic words, is a fossilized remnant of an earlier process involving vowel-to-vowel assimilation." I will address Nevins' argument against this position before moving on to discuss his alternative account.

Nevins (*ibid.*: 9-10) presents the findings of a study by Przezdiecki (2005) into co-articulatory effects in Central Ede dialects of Yoruba. In some of these dialects, e.g. Ijẹṣa, the high vowels, /i, u/, participate in tense/lax harmony, whereas in others, e.g. Standard Yoruba, they do not. Przezdiecki (2005) found that in the latter dialects there is a "statistically significant trend toward the vowel [i] being more laxed when it precedes a lax vowel" (Nevins (*ibid.*: 9)), whereas for [u] he did not find such a trend. Nevins interprets this as problematic for an account in which

harmony in Ijẹṣa-type dialects has phonetic origins, since [i] and [u] do not pattern together. This seems an unduly pessimistic interpretation of the results when you consider the nature of the acquisition process which interacts with such PLD. Przezdiecki in fact demonstrates the ambiguity of the input with respect to whether high vowels take part in a phonological process of harmony. The issue of why /i/ and /u/ might pattern together in a (mistakenly postulated) rule has a natural account in terms of the independently well-attested tendency toward 'conservative' characterization of the featural content of phonological rules in acquisition (see §7.3.4 below, and earlier discussion in §2.3). Far from presenting a "dilemma" for phonetically-based accounts of the origins of vowel harmony, Nevins actually presents rather persuasive support.

In lieu of phonological rules with standard linearly-assessed contexts, Nevins proposes that vowel harmony is based on a search-and-copy procedure triggered by the idiosyncratic marking of a segment as lacking specification for a distinctive feature. This search procedure moves mechanically through each segment in turn, checking it for a fully-specified copy of this feature. Search is set rule-specifically to be leftward, rightward, or both. When a relevantly-specified feature is encountered, it will be assumed/copied by the underspecified vowel. This approach is presented as akin to the *Agree* operation of NS discussed in §4.1.4. Nevins takes it that the same *MinSearch* considerations which motivate the MLC in NS (see §4.1.6) motivate the adjacency effects observed by vowel harmony.

The logic of his argument is well demonstrated by his exposition of Turkish labial and back harmonies, the rule-based version of which were given in (73) and (74). This discussion demonstrates clearly that it is the nearest possible 'source' of a feature that is operative in vowel harmony. Consider the starting representation in (75), in which the underspecified vowels /A/ and /I/ are part of the plural and genitive suffixes on (part of) the stem /pul/ ('stamp'):

(75)	+voc	-voc	-voc	+voc	-voc	+voc
	+high	+son	+son	-high	+son	+high
	+rd	-nas	-nas	-rd	-nas	
	+back	+lat	+lat		-lat	
	….u	l	l	A	r	I

In Nevins' model, the first search begins from /A/ and initially examines the /l/ to its left for a potential source of [+/-back]. Having rejected this segment, the search continues to the preceding

/l/, which is also examined and rejected. Then /u/ is inspected, which satisfies the target's requirements, and [+back] feature is assumed by /A/ to form the fully-specified /a/.

(76)	+voc	-voc	-voc	+voc	-voc	+voc
	+high	+son	+son	-high	+son	+high
	+rd	-nas	-nas	-rd	-nas	
	+back	+lat	+lat	**+back**	-lat	
	….u	l	l	a	r	I

An analogous procedure is then initiated by the underspecified segment /I/, which picks up its [-round] and [+back] features from the newly specified /a/.

(77)	+voc	-voc	-voc	+voc	-voc	+voc
	+high	+son	+son	-high	+son	+high
	+rd	-nas	-nas	-rd	-nas	**-rd**
	+back	+lat	+lat	**+back**	-lat	**+back**
	….u	l	l	a	r	i

As will be clear, as of yet, there are no empirical grounds to think this minimal search procedure is preferable to cyclic application of the rules in (73) and (74), which specify adjacency in their contexts.

Nevins attempts to further motivate his account by suggesting that it shares a property of syntactic agreement – being subject to *defective intervention*. Defective intervention is discussed in greater detail in §5.2, but the logic of Nevins' argument is that there may be elements which match the requirements of the search operation, but do not provide values for the unvalued feature / 'probe'. Since the search operation is constrained by minimal search, it cannot proceed beyond these 'defective interveners' to find a non-defective 'goal', and default specification must be adopted. If a *defective intervention constraint* (DIC) can be shown to apply, it would present compelling evidence for the operational minimality of vowel harmony.

Nevins (*ibid.*: 122) presents an example of defective intervention from syntax (drawn from Chomsky (2001)):

(78) Það finnst/*finnast mörgum stúdentum tölvurnar ljótar
there find.SG./*PL. some student.DAT.PL. computer.NOM.PL. ugly
'Some students find the computers ugly.'
(Holmberg & Hróarsdóttir 2004: 654)

Nevins suggests that the probe T needs to value its φ-features, with the closest goal being *some students*, bearing a [+plural] feature it could copy. He further proposes that there is an additional condition on the goal of agreement, that it be [+Nominative], which is obviously not met by *some students*. T has encountered a 'defective' intervener, unable to transmit its φ-values. Under minimal search requirements, T cannot probe *the computers* to value its φ-features, and, as such, a "last resort" operation takes place, ascribing the probe the default value of [-plural].

However, this is not an accurate characterization of the derivation. T's uφ-probe does not relativize its search to [+Nominative] elements; [200] *some students* is a perfectly valid goal. The reason agreement does not obtain, however, is because its φ-features are not visible to future operations: *some students* is not associated with any uninterpretable features, with its Case-feature being specified *inherently* as Dative. Dative Case in Icelandic is *quirky*, keeping the SO itself active for *Match*, even though the φ-set itself not available for *Agree*, and so preventing probing beyond the SO under the MLC. This is a special property of quirky Case relative to non-quirky, structural Case, which deletes with φ-checking, leaving SOs invisible for intervention purposes. We will not present full details of an account of this phenomenon here (instead see M. Richards (2008b)), but quirky Case-bearing 'interveners' can be thought of as inherently Case-marked DPs associated with additional [3rd person, Case]-feature bundles – 'hidden' expletive 'shells', active for *Agree*. These 'hidden expletives' agree with the T-element, yielding the familiar 'default' third-person agreement familiar from 'partial' *Agree* in overt expletive constructions (e.g. 'There is/*am only me'). [201]

[200] In the Probe-Goal-Agree framework, no explanatory ground is covered by differentiating abstract Case-features, as they do not motivate operations. The differential realization of structural abstract Cases (nominative and accusative) at PF in some languages merely registers the identity of the Probe.

[201] See M. Richards (2008b) for further empirical arguments in favour of this formulation of quirky Case intervention effects. Note that this account actually removes the need for any *Defective Intervention Constraint* (DIC) whatsoever: a second φ-set, visible for Agree since it bears an undeleted Case-feature, is the true intervener. See M. Richards (*ibid.*) for discussion (and dismissal) of other examples (Chomsky 2000, 2001), not involving inherent Case, taken to motivate a DIC.

Nevins compares his understanding of the DIC to the situation that arises in cases of "parasitic" vowel harmony, when a vowel underspecified for a particular feature may only be specified by a vowel with which it shares the value for another feature. Segments with which it shares this feature-value will prevent the underspecified vowel from adopting a value for its unvalued feature from a more distant segment. Parasitic vowel harmony is found in Yawelmani, in which [αhigh] suffix vowels may only copy [+/-round] from segments also specified as [αhigh]. In cases where there is an intervening [αhigh]-bearing segment, unspecified for [+/-round], the default [-round] feature is inserted.

Nevins fails to motivate a characterization of vowel harmony by minimal search since his allusion to the DIC is specious. He also fails to offer a robust argument for rejecting the traditional position, under which vowel harmony is instantiated by phonological rules motivated by phonetic factors acting through (mis)acquisition. [202] Furthermore, the example of parasitic harmony he presents is also extremely suggestive of a phonetic origin for (Yawelmani) vowel harmony. It is well-known that articulatory-perceptual events are related at a phonetic level, and so finding them tied together is harmonic processes is a natural expectation of a physiologically-*oriented*, rule-based account. In particular, the retraction of the lips (roundedness) is articulatorily (and perceptually) related with the height of tongue body; it is therefore to be anticipated that a segment non-distinct in height will disrupt anticipatory/perseverative rounding co-articulation, giving rise to the likelihood of a secondary requirement on any vowel harmony rule posited by misinterpretation. In fact, this connection is captured in one of the original markedness conventions of Chomsky & Halle (1968: 405), by which they tried to specify 'natural' phonological processes:

(79) (XIb) [*u* round] → [-round] / [___, +low]

(79)/(XIb) points out that retraction is *positively* correlated with tongue body-raising (as you can easily justify to yourself by alternately retracting and rounding your lips). In summary, the particular conditioning of vowel harmony rules in fact offers highly suggestive grounds to suppose that adjacency effects stem from the misinterpretation of co-articulation within surface realizations, rather than evidence for the role of minimal search requirements, as suggested.

[202] See the following section for the detailed characterization of this mechanism – entirely consistent with the competence-performance distinction.

There is one other obvious reason to doubt Nevins' conclusions. He presents cases in which search for vowel harmony is able to travel in both directions, e.g. Woleaian low harmony (Nevins (*ibid.*: 42)), clearly not minimal from an operational point of view, but entirely consistent with the action of co-articulation (which may be articulatory and perseverative) on PLD. These patterns in the PLD may then be misinterpreted as bidirectional, adjacency-constrained harmony rules of phonological grammar. (This is *holistic* functionalism in Newmeyer's sense – see §1.3.3.2.)

In summary, there seems little reason to suppose that Nevins (2010) uncovers the implication of *MinSearch* in the phonological algorithms. This is no disaster of course, since there are perfectly adequate, though more intricate accounts of vowel harmony effects available. This finding is, however, consistent with the prediction of the fourth Minimalist proposal that domain-general cognitive principles of optimality might not satisfy the needs of the mappings to PF / the CI-interface so fully as they do the needs of NS. [203]

4.3.4 Representations and rules

A good deal of machinery is implicated in mapping-to-PF, and I attempt nothing other than the most cursory of examinations in this section, merely to illustrate the very powerful role ICs play in motivating complexity and sub-optimal computation in this area of grammar. Following on from the previous section, these are findings consistent with the fourth Minimalist proposal.

Phonology radically breaches *Full Interpretation*, requiring abstract underlying representations (URs) composed of *i.a.* phonemes and features, which have to be appropriately enriched, in breach of the Inclusiveness Condition, to give surface representations (SRs) interpretable by the AP-interface. In short, phonology radically flouts the principle of *TLTB*.

The motivation for URs is the useability/inheritability IC. Every lexical entry must contain the information enabling it be accurately realized in all phonetic contexts. As observed by Chomsky & Halle (1968: 296) however, we cannot take the option of storing every realization of a word, because every case is an exception, being different in some way. The grammar would therefore be intractable in production or comprehension, as well imposing an enormous burden on memory. Instead we must represent each LI as a single, abstract, phonetic representation, without specification for how it varies in realization.

[203] We have already discussed the need for indices, c-command relations, and reference-set computation in the mapping-to-CI, and for reference-set computation in the mapping-to-PF (see §2.4.2).

A second requirement imposed by useability is that we be able to express / pass on a potentially unbounded number of LIs in a distinguishable fashion. We are unable to accurately distinguish and reproduce an infinity of possible sound-meaning pairings unless they are encoded in a finite number of segments (Hockett 1960: 12). What's more, if a finite number of segments – an 'alphabet' – is to encode a potentially infinite number of meanings, these segments must (at least predominantly) be arbitrary signifiers, not meaningful in themselves, giving rise to "l'arbitraire du signe" (De Saussure ([1915], 1971)) (cf. Hockett (1960: 6), Fortuny (2010: 134)). The IC of useability/inheritability therefore gives rise to the phonemic, abstract, arbitrary level of phonological representation we are familiar with, and so to the "duality of patterning".

The upshot of URs is that the LA-er must postulate rules which 'parse' the relationships URs have with their varying SRs, the realizations of which comprise their PLD. However, this is not to say that phonological rules are restricted to straightforwardly detailing the ways in which sounds must vary in accordance with phonetic context. Acquirers do not have access to SRs themselves (the instructions to the articulatory apparatus), but rather to the phonetic realizations (PRs) of SRs. Misinterpretations of the influence of the AP-apparatus (both articulatory and perceptual effects) will take place in the absence of any means to differentiate PRs and SRs, giving rise to a rich rule system. These rules will often reflect properties of the AP-apparatus therefore (cf. Blevins (2004) and references within), as we saw with rules motivated by co-articulation effects in the previous section. A second set of rules with a transparent phonetic motivation are those which fall under the *Obligatory Contour Principle* of phonology [204] (Leben 1973): Walter (2007) discusses how adjacency of identical elements (or elements sharing particular features) is plausibly a reflection of the articulatory difficulty of rapidly repeating particular gestures and/or the perceptual difficulty of distinguishing similar sounds in close proximity. [205] It is equally important to note, however, that since the arbitrary PLD to which LA-ers are exposed "encode layers of historical change" (Vaux 2008: 27) and there is no 'functionalist' motivation to phonological acquisition, much of this data will give rise to phonetically "unnatural" (mis)interpretations. See Vaux (*ibid*.: 44-7) for discussion and references.

From the point of view of computational optimality, what is striking about the rule systems acquired in phonology is their vacuity. They are not triggered by issues of interface-

[204] A set of independent rules with common motivation, rather than a principle of grammar in itself (see Odden (1986, 1988) for evidence of exceptions to a general constraint on haplology).

[205] As Samuels (2009: 368) notes, "[t]hese are both extremely general properties of articulatory and perceptual systems which we have no reason to expect would be unique to language or to humans."

defectiveness, and apply to URs regardless of whether they actually make any difference or not. In fact, given the arbitrary historical forces under which rules and their linearisation develop, [206] the effect of one rule may in fact remove the context in which another can apply, a so-called 'bleeding' relationship (Kiparsky 1968). Both these effects can be illustrated by English pluralization:

(80) (a) the underlying form of the plural morpheme, /z/, is adjoined;
(b) an epenthesis rule applies first:
ø → [ɪ] / [+ strident] ___ [+ strident]
giving: <pɪgz>, <dʌkz>, <fɪʃɪz>;
(c) the assimilation rule applies second:
[+obstruent] → [-voice] / [- voice] __)σ
giving: [pɪgz], [dʌks], [fɪʃɪz].

As will be clear, in the case of stems ending in non-strident, non-obstruent consonants, both rules have no effect. Furthermore, when the stem ends in a strident obstruent, such as /s/ in /meɪs/, the epenthesis rule gives rise to <meɪsɪz>, blocking the devoicing effect of the assimilation rule, resulting in [meɪsɪz], rather than *[meɪsɪs]. The second rule has been rendered vacuous by the first rule. It is clear that phonological rule systems breach *MinVO* radically; but it is important to

[206] It has been suggested to me that the so-called *Elsewhere Condition* (Kiparsky 1973) of morpho-phonology constitutes evidence of computational optimality in the mapping to PF. The Elsewhere Condition imposes disjunctive ordering between a pair of rules instead of the normal conjunctive relationship whereby both rules apply. In particular, when two rules are in "competition", the more "specific" rule is ordered before the "more general" one, which only applies if the former fails to obtain. This prevents the vacuous application of the more "specific" rule. We can again use English pluralization to illustrate the logic. The Elsewhere Condition promotes the 'narrower', irregular rule of plural formation, such as /tooth/ > /teeth/ over the more general addition of the /-z/ suffix, which is blocked from applying to /tooth/. There are two problems with the argument that this implies computational optimality. First, for the Elsewhere Condition to apply would require 'look-ahead' and global computation of optimality of the kind discussed in §2.4.2, which is not meaningfully optimal. Secondly, there is a natural acquisition account of these ordering effects. Only the disjunctive ordering is a successful parse of the PLD: any alternative ordering would be inconsistent with the visibility of the irregular plurals. There seems little reason to suppose the Elsewhere Condition is anything other than a reflex of acquisitional logic. (This position is accordant with the observation by Zeevat (1995) that the irregular form and the regular form may in fact co-exist: this would be inconsistent with any sort of absolute grammatical constraint, but is consistent with indeterminate ordering by LA-ers.)

note that they exist to satisfy the IC of useability, making these breaches principled in the relevant sense.

As well-known, phonemes are not themselves the atoms of the 'phonetic alphabet': there is substantial evidence that phonemes have rich internal structures, and can be analysed as complexes of *distinctive features* (Trubetzkoy ([1939], 1949), Jakobson 1941). This, of course, introduces another type of symbol into phonological computation. The motivation for cognitively decomposing speech sounds still further, in breach of *TLTB*, again seems to be the IC of useability/inheritability. Featural decomposition allows the language user/acquirer to postulate rules of greater abstraction and scope. Instead of postulating rules for each phoneme, we can now postulate rules relative to a set of phonemes and phonemic contexts, reducing their number to a learnable and useable level.

There is the separate question of the parameters along which phonology allows sound to vary to fulfil this IC. These parameters are very plausibly an immediate reflex of properties of the AP-apparatus. As alluded to in Table 2, point (iii), Michael Coen's research (discussed in Berwick & Chomsky (2011: 34)) suggests that "each species has some finite number of articulatory productions, e.g., phonemes, that are genetically constrained by its physiology, according to principles such as minimization of energy during vocalization, physical constraints, and the like", a position they compare to "Kenneth Stevens' picture of the quantal nature of speech production (Stevens, 1972; 1989)." Phonetic features are those properties of speech events which can be independently controlled by human AP-apparatus, partly defined by their accordance with the third-factor constraints imposed by the physical / cognitive costs of different motor activities (see §4.3.1, fn.79, fn.197 for previous discussion.) This property of articulation is 'co-opted' by phonological design to fulfil an IC. [207]

Other levels of symbolic structure (or the "prosodic hierarchy") employed in phonology, in breach of *TLTB*, can also be afforded principled explanations in terms of useability/inheritability. As discussed §1.2, word stress seems to have a crucial role in the identification of word boundaries in continuous speech. As discussed in §1.4.1, in particular fn.42, it seems likely the syllabic organisation of speech reflects the requirement that consonants and vowels be found in close association, if non-sonorant *con*sonants are to be producible, or to 'sound' in a perceivable

[207] Along with the precise nature of categorical perception, a property of the perceptual apparatus which itself reflects third-factor concerns for minimizing cognitive cost, by reducing sensitivity such that ranges of values on a phonetic scale are treated as equivalent.

manner. Finally, in §2.3, we discussed the possibility that higher prosody structure reflects the needs of the parser when determining the underlying syntactic structure of a sentence. There is a large body of psycholinguistic literature demonstrating just this: see Cutler, Oahan & Donselaar (1997) and Wagner & Watson (2010) for a review.

In summary, it seems that the mapping-to-PF does manifest some evidence of computational optimality, but perhaps not as much has been argued by Nevins (2010). However, mapping-to-PF also radically breaches *TLTB* and *MinVO*, but in ways that can be given principled explanations in terms of ICs. Minimalist enquiry is, therefore, far from mute with respect to modules of externalization, but, when applied appropriately has a broad explanatory scope. What's more, the results of this chapter's enquiry are consistent with the prediction of the fourth Minimalist proposal.

5. *Minimize Space Complexity*

This chapter continues to explore the possibility that the FL makes use of domain-general optimal cognition and domain-specific cognition satisfying ICs. Our focus in terms of computational optimality switches from operational load to cache load. Discussion of the implication of *Minimize Space Complexity* in the computation of the parser is hived off into the following chapter, which therefore must be considered alongside the discussion here. [208]

5.1 Phase-based derivation and phasal Spell-Out

Chomsky (2000, 2001) introduces the notions of *phase*-based derivation and phasal Spell-Out. The formulation and understanding of these ideas have undergone considerable development in subsequent papers, particularly Chomsky (2004c, 2007b, 2008), incorporating insights from many other scholars and their work, referenced within. I discuss only (my understanding of) the current account here, however, rather than giving a narrative presentation. [209]

Phases are cyclic domains of derivation, within which all operations other than *External Merge* (including *Transfer* to PF/the CI-interface) are motivated by features of the *phase-head* (PH), and so take place effectively simultaneously: probing operations of other heads within the phase (*non-phase-heads* (NPHs)) are motivated by features transmitted to them from the phase-head by an *inheritance* operation. In the clausal domain, PH-NPH pairs uncontroversially include C-T and *v**-V, with the probing features of T and V being derivative of C and *v**.

[208] One potentially relevant issue has already be introduced (counter-cyclically) in §4.1.7, where it was suggested that *multiple-Merge* might breach the principle of *Minimize the Caching of Unintroduced Items* (*MinCUI*).

[209] The original formulation of phases involved *sub-arrays* within the Numeration and was underpinned by empirical arguments from *Merge*-over-*Move* effects, which we saw in §4.1.4 are empirically and theoretically untenable.

There are various sources of evidence in support of this conception of clausal structure. [210] Using the following derivation (which abstracts away from T-to-C movement and *do*-support), Chomsky (2004c: 123) gives a simple demonstration of the effective simultaneity of structure-building operations at the phase level, yielding apparently counter-cyclic effects (cf. fn.170):

(81) What C [he T [~~what~~ [~~he~~ see ~~what~~]]]
'What did he see?'

If cyclicity is observed, this derivation gives rise to an apparent violation of the MLC: T would raise *he* to Spec-TP, skipping the copy of *what* in a higher position. However, under feature inheritance, T is inactive until C is merged, at which point *what* is raised to Spec-CP, voiding this breach of the MLC. It is as if the raising of *what* were simultaneous with a 'previous' operation (when considered from the point of view of cyclicity), the raising of *he*. This is the natural expectation if all probing operations stem from the PH, but impossible if there are independent probes on T.

The derivative nature of T's φ- and T-features is also suggested by the properties of bare-TP constructions (*exceptional case marking* (ECM) infinitives), which are unmarked for tense and unable to delete the Case-feature of the lexical verb's argument (a capacity dependent on probing unvalued φ-features) (Chomsky 2008: 143-4). In fact, as Chomsky (*ibid*.: fn.26) observes, the C-origin of φ-features is even sometimes marked morphologically, as in West Flemish (see Miyagawa (2005) for a summary of such cases).

A final point, made in Chomsky (2008: 148), is that the raising of objects to Spec-V (see Lasnik & Saito (1991) and Lasnik (2003) for evidence and discussion) receives a natural explanation under the inheritance of a probing feature of *v** by V. This phenomenon is particularly odd because its effects are never visible, since V raises to *v**.

We can now begin to speculate the motivation for this property of phrase structure. The operation of inheritance seems to be a breach of *TLTB* and *MinVO*. [211] The first place to look is the

[210] Counter-proposals have also been made (see, for example, Gallego (2014) and references within), but I am not in a position to discuss them in detail here, merely presenting Chomsky's arguments, before speculating over their potential significance in our discussion if correct.

[211] While it 'relocates' features, the operation of inheritance is not itself counter-cyclic in a meaningful sense: its 'downward' action is analogous to the operation of probe-goal agreement (see Chomsky 2007b: fn.26).

interface. As noted in §4.1.4, the duality of semantics seems to be a CI-IC (or at least CI-concern) which is met by the property of displacement. This was discussed as potentially an optimal means of meeting this requirement, avoiding the need for repeated operation of *External Merge* and related complications. It does however require that there be available locations for displacement identified with discourse-level semantics – that is, an A/A'-distinction within NS. The A/A'-distinction is captured by having one PH, associated with edge semantics, and one NPH, [212] associated with core semantics. Feature inheritance can plausibly be thought of as an optimal way of providing a second site for *Internal Merge*, since it avoids the need for instantiating a separate, fully-specified head and an independent *External Merge* operation introducing it.

The particular nature of this inheritance relationship and its consequences suggests other principled motivations (cf. M. Richards 2007). Feature inheritance imposes *cyclical Spell-Out* of NS derivation once a PH's features are fully specified; the results of NS are therefore made available to the mappings to PF and the CI-interface at various points within the overall derivation. Chomsky (2007b: 18-9) notes that the valued probe features of NPHs must be removed when transferred to the CI-interface, since they remain semantically uninterpretable. However, their removal can take place no later than the phase-level at which they are valued, since at the next phase-level they will be indistinguishable from interpretable features, and so will remain undeleted, causing a crash. *Transfer* / removal of valued uninterpretable features must therefore be associated with the locus of their valuation, i.e. their source PH, because it is only here that the *Transfer* operation can "know" which valued features are interpretable and which are uninterpretable. The inheritance of uninterpretable probe-features therefore imposes phasal Spell-Out.

I make the first salient point about this proposal in advance of presenting its empirical basis. Phasal Spell-Out demands the *Earliness Principle* of Pesetsky (1989) and Pesetsky & Torrego (2001): unvalued features must be valued at the earliest possible opportunity or phasal Spell-Out will cause the derivation to crash (Chomsky 2005b: fn.12). This is computationally optimal, since featural needs ('unfulfilled promises') would otherwise have to be held over in WM, in breach of *MinCID*.

There is significant empirical evidence for the piece-meal *Transfer* of NS derivation. The classic NS-internal evidence comes from the successive-cyclicty of A'-movement, under which an A'-

[212] Both plausibly further articulated, in accordance with the results of cartographic enterprise discussed in §3.2.2.1.

raising SO passes through the edge of every phase, avoiding *Transfer* before it can be realised as the specifier of the relevant EF-bearing PH.

(82) [Which of the papers that he_i gave to $Mary_j]_k$ did every $student_i$ [*v*P t_k ask her_j [to read t_k carefully]]?
(Legate 2003: 507)

The *wh*-phrase in (82) must reconstruct to a position below *the student*, such that *every student* binds *he* (under standard c-command conditions on bound pronouns); this position must also be above *her*, such that *Mary*, an R-expression, is not c-commanded by a co-referent pronoun (under Binding Principle C). 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 phasal Spell-Out.

Strong evidence of phasal Spell-Out also comes from effects manifest in mapping-to-PF, shown in work by *i.a.* Ishihari (2003), Marvin (2003) and Newell (2008). A sample of Marvin's findings is presented in the following paragraph for expository reasons. Other support is found in studies of the mapping to the CI-interface, in particular work by Marantz (2001) and Cecchetto (2004), which we will not present here.

Assuming a syntactic model of morphology (Halle & Marantz 1993, Marantz 1997), Marvin (2003: 51-4) discusses the action of the stress assignment and vowel reduction rules of English. The details of the stress assignment rule need not concern us: it is sufficient to recognize 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 effect of phases in phonological derivation is best seen with respect to an example, e.g. the word *governmental*, with the structure shown in (83) (movement rules, which we will not show, establish the correct linear order):

(83) [*aP* [*a* al] [n_1*P* [*n* ment] [*vP* [*v* Ø] [√govern]]]]]
(Marvin 2003: 52)

Let us consider each stage 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. 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, giving the SR [gʌvərnmɛntəɬ], stressed on the [ʌ] and the [ɛ]. 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.

The cyclic *Spell-Out* of phases means phonology can "forget about" (Chomsky 2004c: 107) the derivational results of previous phases, clearly reducing the space complexity of these algorithms, in accordance with *MinCCD*.

As pointed out by Nissenbaum (2000), the same cannot be true of the mapping to the CI-interface, since there are semantic operations which refer to the unitary output of NS derivation – such as the licensing of variable binding or Binding Principle C – and which are sensitive to (sometimes long-distance) c-command. Take Binding Principle C for instance, holding that an R-expression is not c-commanded by a co-referent pronoun. In (84), *John* cannot be co-referent with *he* since it is c-commanded by *he* at the interface.

(84) $[_{IP}$ He INFL $[_{vP}$ *v* $[_{VP}$ thinks $[_{CP}$ that $[_{IP}$ Mary INFL $[_{vP}$ $[_{VP}$ saw John]]]]
(Cecchetto 2004: 364)

To compute the relevant c-command relationship, it is necessary to have a complete representation of (84) at the interface, and so SOs individually shipped to the CI-interface cannot be dismissed from WM in the same way they can in the phonological computation.

A separate question arises as to whether phasal Spell-Out will give rise to 'impenetrability' of the contents of a phase passed in syntactic derivation with respect to higher syntactic operations. The standard understanding is that it does, yielding the *Phase Impenetrability Condition* (PIC). The PIC holds that only the previous PH and its edge are visible to higher syntactic structure. If this is the case, then the syntactic derivation manifests the effect of *MinCCD* in the same way as that mapping to PF does (Chomsky 2004c).

Chomsky has proposed two different versions of the PIC:

(85) *Phase Impenetrability Condition*$_1$ (Chomsky (2000): PIC$_1$)
In phase α with head H, the domain of H is not accessible to operations outside α; only H and its edge are accessible to such operations.

(86) *Phase Impenetrability Condition*$_2$ (Chomsky (2001): PIC$_2$)
(Given structure [$_{ZP}$ Z … [$_{HP}$ α [H YP]]], with H and Z both PHs:)
The domain of H is not accessible to operations at ZP; only H and its edge are accessible to such operations.
(adapted from M. Richards (2012: 136-7))

The initial formulation of the PIC was modified in accordance with Icelandic examples such as (87), so that the complement of the lower phase (*v*P in this case) only becomes inaccessible (is transferred to the interfaces) when the next PH is merged (C in this case), allowing T to agree with the complement of V.

(87) Mér **T** þóttu $t_{mér}$ [**þær** vera duglegar]
Me-$_{DAT}$ thought-$_{\mathbf{3pl}}$ **they-**$_{\mathbf{NOM}}$ to-be industrious
'I thought them industrious'
(*ibid.*: 137)

As it is, there also seem to be constructions which breach PIC$_2$, summarized in M. Richards (*ibid.*: 138), including those from Chukchee noted in Chomsky (2008: fn.25), in which there seems to be agreement between a matrix probe and an argument in an embedded CP:

(88) ənan qəlɣilu ləŋərkə-nin-et iŋqun Ø-rətəm'ŋəv-nen-at qora-t
he-inst regrets-**3pl** that 3sg-lost-3-pl reindeer-**pl**.NOM
'He regrets that he lost the reindeer.'
(Bošković 2007a: 57)

Chomsky (2008: 143) points out that "probe into an earlier phase will almost always be blocked by intervention effects *[i.e. the effects of the MLC]*" and accepts that there may in fact be true exceptions to any PIC. M. Richards (2012) argues that the second set of alleged counter-examples have alternative syntactic analyses, not in breach of PIC$_2$, and goes on make a suggestion as to how to reconcile PIC$_1$ with examples such as (87). I am not in a position to present and comment on these suggestions, merely noting that there is an open empirical issue as to whether domains of

syntactic derivation are dismissed from WM on their transfer to the interfaces, in accordance with *MinCCD*.

Finally, I would like to discuss a further dimension of the computational optimality of phasal Spell-Out, associated with the way it relates to the different components of the FL. When a computational task is divided into sub-tasks, the output of one element becomes the input of the next. *Pipelining* is an implementation technique that allows serial processing elements to operate in parallel. Each stage of the *pipeline* will periodically pass on a completed section of computation to the next stage, which can then perform its processing alongside the continuing computation of the previous stage. In turn, this stage 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 at the same time. While it does not decrease the time each stage takes to perform a given amount of computation, it increases throughput of the series as a whole, and, as such, the rate at which tasks are completed (Hennessy & Patterson 2006: A2-A77).

A precise analogy to this situation is afforded in the FL. Linguistic cognition consists of a series of processing elements performing different sub-tasks: NS provides input to the mapping-to-CI and to the mapping-to-PF, which in turn release information to the interpretive and phonetic systems. Phasal Spell-Out allows computation to take place in a piece-meal fashion, with each stage in the pipeline periodically availed input by the previous stage. As such, syntactic, semantic, phonological, interpretive [213] and phonetic processing may operate in parallel, improving the throughput of the 'linguistic task' as a whole: the pairing of form and meaning, and subsequent interpretation and/or production, takes place more efficiently. Phasal Spell-Out implicates *Optimize Scheduling* as a principle of computational optimality alongside the narrowly 'algorithmic' principles of Chapter 2's framework. [214]

[213] It might be supposed that the role of a unified syntactic representation in the mapping-to-the-CI precludes simultaneous interpretive computation, but this neglects the fact that most interpretation is "local in nature (consisting of function application or predicate modification, which take place between two sister nodes)." (Cecchetto 2004: 364) In other words, interpretation may be concurrent with syntactic computation, but cannot be completed until the final piece of NS computation has been transferred.

[214] Chomsky (2002: 158) makes some suggestive comments in this regard, albeit not in relation to phases. He asks "why should biology be set up so that there is one fixed point in the computation at which you have an interface?", suggesting "[i]nterpretation could be "on line" and cyclic"".

In conclusion, we have seen that phase-based derivation and phasal Spell-Out are of great interest from a Minimalist perspective. Feature inheritance is a principled breach of substantive and computational optimality, motivated by its economy in encoding an independently-demanded A/A'-distinction. It also results in phasal Spell-Out, which renders phonological computation maximally 'forgetful', in accordance with the *MinCCD*, and potentially does the same for the syntactic computation. Phasal Spell-Out was also seen to be consistent with the role of *MinCID* in NS. Finally, we observed that phasal Spell-Out cashes out the principle of *Optimize Scheduling* in the overall design of the FL. On the grounds that they underpin phasal Spell-Out, therefore, the symbols and operations associated with unvalued interpretable features in the NS can be considered principled breaches of optimality (see fn.179).

5.2 Case and the Activity Condition

In the Probe-Goal-*Agree* system, abstract Case has a more marginal role than it did in previous accounts of phrase structure. It no longer serves to motivate movement, but instead is considered to 'activate' interpretable goals for the purpose of *Agree*.

The Activity Condition (AC) developed in Chomsky (2000, 2001) takes it that for *Agree* to obtain between a probe and a goal they must both be active. Probes are considered active by virtue of bearing unvalued features. Goals, being sets of (valued) interpretable features, remain active under the force of uninterpretable features designated to keep them so: for φ-agreement, (abstract) Case-features. Case is not involved in operations itself, but rather deletes under the matching of φ-features. In the first formulation of the AC, a probe could only delete the uninterpretable Case-feature of the goal if it was "φ-complete" and "non-defective" (Chomsky 2001: 6). This assumption was abandoned under later conceptions of the syntactic phase in which "only completeness at the phase level has an effect" (Chomsky 2008: fn.27), and, at any rate, is refuted by empirical phenomena discussed in Frampton & Guttman (2004) and Nevins (2004).

Under the AC, Case deletion and φ-feature valuation are directly related, capturing *Inverse Case Filter* (ICF) effects (cf. Bošković 2002c, Rezac 2004). That is, the AC accounts for the ungrammaticality of certain derivations in which features of the probe go unvalued on account of prior Case-valuation of the goal. For example, (89) is prohibited because the potential goal has already been involved in φ-agreement and so had its Case-feature deleted by a nearer probe:

(89) *It seems [~~it~~ was told Mary [that Bill is a liar]]
(M. Richards 2008b: 183)

Similarly, the AC alone yields 'defective intervention' effects of the kind discussed in §4.3.3 (including fn.201).

With the role of the AC / Case well-evidenced in NS, questioning turns to its motivation, since Case is uninterpretable at the interface – its instantiation and deletion breach substantive and computational optimality. There is no immediately obvious motivation for the phenomenon at the interfaces, but I (tentatively) suggest that Case-features afford computational benefits. It is possible that Case / the AC has utility in minimizing the burden placed on WM in NS, in accordance with *MinCCD*. Of course, fully interpretable arguments cannot be 'paged out' as soon as they lose any unvalued features, since they remain part of a syntactic derivation which must be presented to the interfaces coherently. There is some evidence, however, that WM has more 'structure' than we assume, such that the burden of fully interpretable arguments on WM can be reduced in a way compatible with coherent presentation.

Gibson (1998: 15), in his discussion of parsing, proposes that the "memory" cost of postulating a predication relationship increases as "a monotone increasing function" of "the number of new discourse referents […] processed" since the syntactic source of predication has passed. He presents evidence for this "distance-based integration hypothesis" (*ibid.*: 19) from experiments by King & Just (1991) and Gibson & Ko (1998). These experiments test the prediction that "integration times at points of higher memory cost will be increased relative to integration time at points of lower memory cost" (Gibson 1998: 20) by measuring the pace of reading throughout two relative clause constructions: "The reporter who the senator attacked admitted the error" (an object-extracted relative clause) and "The reporter who attacked the senator admitted the error" (a subject-extracted relative clause). The prediction of Gibson's model is that "reading times [will] be fast throughout both relative clause constructions, with the exception of (1) the matrix verb in each, where a long distance integration occurs and (2) the embedded verb in the object-extraction construction, where two substantial integrations take place" (*ibid.*: 21). The results of the experiments accord with these predictions (see Gibson (*ibid.*: 22-3) for graphical results).

Gibson (1998) interprets these results in line with ideas presented in *i.a.* Just & Carpenter (1992): operations and recall from storage rely on the relevant items achieving a particular level of 'activation' – a property which fades with further cognition, minimizing the burden on WM (see

fn.27). Increased levels of activation impose a greater burden on WM – the 'structure' of WM alluded to above. There is no reason to assume WM in NS computation operates any differently to that elsewhere in cognition. As such, we can think of Case as a feature which (when deleted) allows elements which will not be needed for future syntactic operations to reduce their burden on WM, being held 'just hard enough' that they can be presented to the interfaces. SOs with undeleted Case-features on the other hand will remain highly activated for ease of retrieval by future *Agree* operations. In short, Case / the AC could well be the machinery of *MinCCD*, given the nature of WM.

This chapter seems to implicate domain-general cognitive principles minimizing cache load in the design of the FL. It also demonstrates the FL involvement of a 'super-algorithmic' principle of cognition, optimizing the scheduling of the consecutive algorithms within a computational task. We also argued for the integration of a subtler understanding of the nature of WM into Minimalist argument.

6. Minimalism and the Parser

Chapter 6 continues to track the cache burden imposed by the linguistic computation, but focuses on an area of the FL's activity largely neglected in Minimalist enquiry – parsing. Central results of parsing research are introduced, and an overall model of the parser is composed for assessment in Minimalist terms. The chapter goes on to explore the role of this model in a (partial) account of the linearisation of abstract syntactic structure, in keeping with a suggestion by Chomsky (2007b: 9). Finally, the chapter suggests an account of the *Final-over-Final-Constraint* (Biberauer, Holmberg & Roberts 2014) on surface word order.

6.1 The Parser

In assigning meaning to a sentence, it is generally acknowledged that language users must (at least in non-idiomatic cases) pass through its syntactic and thematic structure. (See J.A.Fodor, Bever & Garrett (1974) and Levelt (1974) for discussion.) The computational process which takes an unstructured string and somehow applies the rules of grammar to it, producing syntactically and thematically structured output, is known as the parser. In order to assess the parser for evidence of 'motifs' of optimality, it is necessary to commit to a particular account of its algorithm(s).

6.1.1 Direction and delay

A central question concerning the working of the parser is how much input it receives from the modules of perception before beginning operations. The parser could either seek to incorporate words into the phrase marker as it goes, incrementally, or it could 'buffer' (store) input up to certain point, before parsing it in either a 'left-to-right' or 'right-to-left' fashion. [215]

[215] Taken completely literally this is a false dichotomy of course; we are discussing the overall approach only here. A parser can take an incremental approach and still have to buffer input to a limited extent. For instance, if the parsing algorithm is incremental *and* 'head-driven' (cf. Pritchett's (1992a, b) model discussed in §6.1.4), then any modifiers preceding a head cannot be incorporated until that head is perceived and are buffered until that point.

Research points strongly to the (perhaps unsurprising) conclusion that parsing is (systematically) incremental and left-to-right; that is, it appears that we try to integrate each word into a unified syntactic/thematic representation at the earliest opportunity. Several studies (*i.a.* Rayner, Carlson & Frazier 1983; Ferreira & Clifton 1986; Trueswell, Tanenhaus & Garnsey 1994; Clifton *et al*. 2003) show that language comprehenders encounter processing difficulties with locally ambiguous inputs before the end of the sentence. For instance, the parsing of (90) is disrupted (i.e. slows down) immediately *by the lawyer* is perceived.

(90) The defendant examined by the lawyer turned out to be unreliable.
(Ferreira & Clifton 1986: 352)

This sentence becomes (locally) ambiguous at *examined*, since this form could be either: (a) a passive past participle, indicating a reduced relative clause structure, in which *the defendant* holds a theme theta-role; or (b) the active preterite, indicating a main clause structure, in which *the defendant* is assigned the agent theta-role. The processing difficulty experienced at *by the lawyer* suggests that the main clause analysis has been adopted by this point, and, as a result, the parser is forced to 'reconsider' its previous analysis, slowing processing.

A different line of experimentation, pursued by Marslen-Wilson & Welsh (1978), points to the same conclusion. In these experiments, participants had to "shadow" a variety of continuously spoken sentences, that is, repeat them back as quickly as possible. The sentences contained deliberate errors in the form of mispronounced words; however the participants were not informed these would be present, simply being instructed to reproduce what they heard precisely. In approximately 50% of cases the participants made "fluent restorations" of the mispronounced words, producing "tragedy" for "travedy". One of the variables found to be predictive of such restorations was the context of the mispronunciation in the sentence, i.e. the 'likelihood' of the mispronounced word being the continuation of previous input. For example, "cigarette" is a highly probable continuation of the string "still, he wanted to smoke a …", whereas in a context such as "it was his … that they were stationary", "misfortune" would be less predictable. In highly-constraining contexts participants made fluent restorations even in the face of significant phonological deviation. This implies that the contexts determining fluent restorations must already have been syntactically/thematically structured by the stage of mispronunciation. In summary, the parsing algorithm seems to operate incrementally, forming a syntactic and thematic analysis at the earliest possible opportunity relative to the perception of the string.

This hardly seems a surprising result: intuitively, a parser will deal with input as it receives it. The basis of this intuition, however, deserves closer scrutiny: does it access notions of externalization-oriented optimality, substantive/computational optimality, or both? As is anticipated after the discussion of §2.1, it is significantly easier to contemplate the issue of substantive/computational optimality, so we reserve discussion of incrementality and concern for (informative) externalization until the following section.

An incremental approach to parsing is computationally preferable to a systematically buffering one on trivial grounds. Any 'buffer' must be stored in WM until parsing has begun, and is only uncached on a piece-by-piece basis. Burdening WM in this way is largely avoided under an incremental approach (cf. discussion in *i.a.* Weinberg (1988: 36*ff.*) and Just & Carpenter (1992: 124)). Plausibly, this implicates a domain-general cognitive principle reducing space complexity, *MinCUI*.

With respect to time complexity, however, it might appear that an incremental parsing algorithm is less preferable than a buffering one. An incremental parser frequently has to perform additional operations in response to local ambiguities which arise when approaching input in such a fashion (see §6.1.4 below). *Prima facie*, we might consider that this is a breach of the algorithmic principle barring operational redundancy; but this would be to neglect the necessary 'blindness' of derivation. As discussed in §2.4.2, 'look-ahead' has no place in theories of computational optimality, because it entails that the results of later operations must be known to an earlier operation. The length of the delay has no bearing on the strategy adopted by a parser once the input is available, and so incremental and buffering parsing algorithms are equivalent with respect to principles of algorithmic design minimizing time complexity.

6.1.1.1 An aside on optimality of function

An attractive argument is that incremental parsing is advantageous with respect to informative externalization, since it allows us to reach a thematic analysis of the input in the shortest possible time, taking place more or less contemporaneously with perception. This is not an entirely straightforward suggestion, however, as it neglects the pervasive ambiguity of input under an incremental approach, which frequently results in processing difficulty and the need for costly reanalysis. In such cases, a complete analysis (up to that point) is reached no faster than would have been possible by waiting for the disambiguating information in the first place. Nonetheless, in systematically buffered parsing, analysis may *only ever* begin after the buffer has been

perceived, even when no ambiguity would obtain; and so, in as far as an input string is non-ambiguous, or a local ambiguity can resolved before the end of the buffer, incrementality may afford a real improvement in the average speed with which accurate analysis takes place. [216] Even this is somewhat speculative of course, since we cannot make an empirical comparison: perhaps the perception time saved by an incremental approach is, on average, exceeded by the time taken to reanalyse the ambiguous input, making the speed of communication superior with a buffered parser.

A further difficulty with this argument is that it relies on the assumption that the *mean* average time to correct analysis is the appropriate way to assess the parser's optimality with respect to externalization. Another possible measure is the degree of accuracy of the first analysis, the *mode* average time to correct analysis. The increased possibility of a misunderstanding / underspecified analysis, even temporarily, could have been particularly costly in the context in which information-oriented externalization evolved, overriding concern for the mean time to correct analysis. In the absence of anything other than 'just so' stories about how externalization might initially have been advantageous, or, which is partly related (see §2.2), a coherent notion of what it means to be optimal in that regard, it is difficult say any more than this.

As it is, an incremental parsing algorithm is clearly serviceable from the point of view of externalization, and, depending on what account of optimality we accept as relevant in this regard, may be better than that. Incremental parsing is (trivially) optimal from a computational standpoint, instantiating *MinCUI*. This is consistent with the anticipation of the third Minimalist proposal, which suggests that the modules of externalization will satisfy communicative conditions in ways that maximize the use of existing cognitive resources (approaching cognitive optimality). [217]

[216] Note that there is no problem with 'forward-thinking' explanations of this kind in the context of adaptive design.

[217] It is important to remember at this point that none of the Minimalist proposals is incompatible with adaptive explanation: even the fourth Minimalist proposal anticipates it with respect to the modules of externalisation. It should also be noted that there is no constraint on externalization-oriented optimality and substantive/computational optimality pointing in the same direction, and so adaptive pressure may be to maintain underlying principles of cognition.

6.1.2 Parsing domains

When analysing an input string, it is generally assumed that the parser systematically 'closes off' units of syntactic structure whose properties are complete, before 'paging them out' to less pressured secondary storage; this reduces space complexity in accordance with *MinCCD*. (Cf. Kimball (1973) and Church's (1980) observation that "the parser cannot keep an unbounded number of sites open for potential attachments because of the bounded size of working memory" (Gibson 1998: 28).)

Studies suggest that closure takes place at the end of a clause. Evidence is adduced from a variety of sources, including studies of recall in interrupted discourse (Jarvella 1971) and the systematic misperception of extraneous noises as occurring at clause boundaries (*i.a.* Garrett, Bever & J.A.Fodor 1966; Caplan 1972; Chapin, Smith & Abrahamson 1972). Furthermore, languages seem to display a strong preference for local attachment when there is (global) ambiguity over VP-attachment sites for adjuncts (*i.a.* Mitchell 1994, Gibson 1998):

(91) (a) Juan dijo que Bill se murió (#morirá) ayer. [Spanish]
(b) John said Bill died (#will die) yesterday.
(Gibson 1998: 28)

In both versions of (91), the parser has a strong tendency to adjoin the adverb to the more recent VP. This preference persists even when the (future) tense of that VP is incompatible with the adverb, suggesting that the higher clause attachment site is much less accessible. [218]

The parallel between the phasal Spell-Out of NS and the clause-based closure principle of parsing is unmistakeable. Like phasal Spell-Out, domain-based closure arguably has a 'pipelining' motivation, allowing 'post-semantic' computational tasks (pragmatic enrichment, reasoning, etc.) to take place in concert with remaining parsing, in accordance with *Optimize Scheduling*.

[218] Ackema & Neeleman (2002) argue that this explains the clause-internal locality restriction on rightward movement (the so-called *Right Roof Constraint* (Ross 1967)): establishing a filler-gap dependency over a greater distance would require clause-based closure to be suspended.

6.1.3 The Active Filler Strategy

An incremental approach to parsing militates in favour of fillers preceding gaps. Postulating the existence of a gap in a string is not a trivial matter, since "a gap is a nothing" and its identification "requires a rather intricate chain of inference based on a variety of facts about" the properties of words in the sentence (J.D.Fodor 1978: 429). [219] For example, "doubtful" gaps arise in parsing: many verbs are optionally ditransitive, or homophonous with other verbs. If *what* had been postposed rather than fronted in (92a/b), left-to-right parsing would have to have been suspended, pending disambiguation of the verb's theta-grid:

(92) (a) What$_i$ do you want Mother to sing $(\Delta)_i$ to Mary for her treat?
(b) Which student$_i$ did the teacher walk $(\Delta)_i$ to the cafeteria?

The precedence of fillers does not, however, speak to the strategy adopted to establish filler-gap dependencies. Early models all involved the initial identification of a gap, which then triggered the search for an antecedent. Two models of this kind ("last-resort" and "first-resort" models of gap finding) are discussed and dismissed in J.D.Fodor (*ibid.*: 433-7), on the basis of consideration of strings containing *true* doubtful gaps, such as (92a/b), and of strings containing *false* doubtful gaps, as in (93):

(93) What$_i$ do you want mother to sing (Δ) to Mary about Δ_i?

The relevant considerations cannot usefully be reviewed here, but J.D.Fodor (*ibid.*: 437*ff.*) observes that a "lexical-expectation", gap-driven model of establishing filler-gap dependencies allows for consistent and accurate parsing in the face of these ambiguous strings. This approach (cf. also Wanner & Maratsos (1978)) can be formalized as:

(94) Gap-as-second-resort strategy (GASRS)
Check the lexical string for an expected item before postulating a gap (i.e. rank the option of a gap below the option of a lexical phrase).
(Frazier & Flores D'Arcais 1989: 332)

[219] Cf. discussion in §4.3.1 concerning the deletion of lower copies created by displacement.

Under this approach, the parser predicts the existence of syntactic elements in advance only when a string fails to manifest an expected phrase:

(95) Who(m) did you find ___?

In processing (95), the GASRS parser first checks the local string for a possible DP, postulating an empty DP once the parser fails to find one. Only then would the filler *who(m)* be sought and checked, and a second copy inserted in the gap.

Subsequent work has pointed out that while the GASRS model offers a satisfactory account of the distribution of difficulties in establishing filler-gap dependencies, it is inconsistent with the other observations. A different, *filler-driven* approach is suggested, proposing that the parser posits a gap once a potential filler (e.g. a phrase in a non-argument position, such as SPEC-CP) is identified. In particular, Frazier (1987) and Frazier & Flores D'Arcais (1989) suggest the parser follows an *active filler strategy* (AFS), which seeks to "[a]ssign an identified filler [to a gap] as soon as possible" (*ibid.*: 332).

For instance, certain studies (*i.a.* Crain & J.D.Fodor 1985, Stowe 1986, Frazier & Clifton 1989) measured self-paced reading times of sentences such as:

(96) (a) My brother wanted to know if Ruth will bring us home to Mum at Christmas.
(b) My brother wanted to know who$_i$ Ruth will bring (*__$_i$) us home to __$_i$ at Christmas.
(Stowe 1986)

Participants read the object of the transitive verb more slowly in (96b), where there is a syntactic filler in the context, than in (96a), where there is not. This suggests a tendency to associate the filler *who* with the direct object position of *bring* as soon as it is encountered, a postulation which immediately has to be undone upon encountering *us*, resulting in slowed processing.

Further evidence for the AFS is provided by the possibility of adjunct movement (i.e. movement of unexpected phrases) (Ackema & Neeleman 2002: fn.10). For example, there is nothing in the string "you think Mary fixed the bike" to suggest that it contains a gap, and so, in parsing "how

do you think Mary fixed the bike?", postulation of a gap must therefore depend on the presence of the fronted *how*.

An AFS is computationally preferable to a GASRS. Under an AFS, the syntactic/thematic properties of a gap and its filler (and the associated properties the gap/filler's c-selector) can be assigned as soon as the gap is identified. Under a GASRS, however, these unfulfilled properties must be held in WM until the filler is identified, in breach of *MinCID*. *MinCID* may in fact be doubly implicated, in that an AFS postulates a gap "as soon as possible" (cf. Gibson 1998: 55). While this 'urgency' will sometimes lead to incorrect assertions and the need to recapitulate derivations, this is not a breach of *MinVO*, for the familiar reason that *MinVO* is a locally-assessed, 'blind' principle of computational optimality. A GASRS, on the other hand, does breach *MinVO*, postulating search operations which may or may not develop relevant structure.

Again it seems plausible that the parser, a 'module of externalization', employs principles of cognitive optimality, in accordance with evo-devo logic.

6.1.4 The parser and ambiguity

Under an incremental approach to parsing, strings are often locally ambiguous. For instance, (90), repeated below, gives rise to both reduced relative and main clause readings at *examined*, with the main clause reading being favoured for reasons discussed shortly. However this commitment is inconsistent with the perception of *by the lawyer*, giving rise to minor processing difficulty and *automatic* reanalysis.

(90) The defendant examined by the lawyer turned out to be unreliable.

Other cases of local ambiguity lead to more pronounced processing difficulty. In classic garden-path strings such as (97), reanalysis cannot be performed automatically, only consciously and with some confusion. We reach the successful parse only after some consideration.

(97) ¿The horse raced past the barn fell.

To understand the distribution of these different types of local ambiguity (in the absence of disambiguating information – a matter returned to below), we must understand what drives the choice of initial parse in the first place.

One proposal which makes a range of accurate predictions is Pritchett's (1992a, b) *head-driven* parsing model, which holds that the parsing is driven by the local application of grammatical principles upon identification of a head. He formalizes this in the following principle:

(98) *Generalized Theta Attachment* (GTA): Every principle of the Syntax attempts to be maximally satisfied at every point during processing. (Pritchett 1992b: 138)

A (head-driven) GTA-based algorithm is one formulation of the incremental parsing. Unlike other formulations (discussed shortly) however, it incorporates a second notion of incrementality – with respect to fulfilling the featural 'needs' of the syntactic elements incorporated. Projection is postulated at the earliest possible opportunity (relative to the identification of a head) and in such a way that it leaves as few features unsatisfied as possible.

Let us demonstrate Pritchett's model using an unproblematic sentence such as (99):

(99) Fred found the students.

Fred is encountered first and identified as a DP. No TP is projected, however, since the incorporation of arguments is driven by syntactic properties of the head; the DP and its features must therefore be held over in WM.

On encountering *found*, the parser identifies a T-head, [220] associated with *i.a.* the EPP and a Case-feature. In accordance with maximal satisfaction of these unfulfilled commitments, the cached DP is incorporated as Spec-TP, in which configuration the EPP and Case-deletion can be immediately discharged. Similarly, the parser will also now be able to identify a V-head, [221] associated with *i.a.* a theta-grid of two theta-roles (θ_1 and θ_2), a Case-feature, and c-selection features. In accordance with maximal satisfaction of theta-assignment and c-selection, a second copy of the cached DP is incorporated as Spec-VP: the DP immediately receives the agent theta-role, θ_1, and the theta-assigner immediately discharges part of its theta/c-selection grid. [222]

[220] Abstracting away from phase structure.

[221] Again abstracting away from phase structure.

[222] The existence of labels (§4.1.8) means that a specifier position can be postulated before a head has projected under *Merge*-r with its complement. This avoids the need for specifier-head *Merge* operations which would only

Next, *the students* is detected, which, identified as a DP, is immediately attached as the direct object of *found*, receiving the theme theta-role, θ_2. This step exhausts theta-role assignment, c-selection, and Case-assignment, and the string concludes; the parse is complete.

Now consider (100):

(100) Fred found the students' lecturers.

Up until *the students*, parsing proceeds as above, resulting in (101):

(101) [$_{TP}$ [$_{DP}$ Fred] [$_{T'}$ T [$_{VP}$ found the students]]]

This structure is syntactically complete. However, once *lecturers*, the true head of the direct object, is perceived, it becomes clear that the initial analysis was incorrect, and so reanalysis must take place – *automatically* in this case.

(102) [$_{TP}$ [$_{DP}$ Fred] [$_{T'}$ T [$_{VP}$ found [$_{DP}$ the students' lecturers]]]]

This reanalysis, in which a preceding DP becomes the possessive specifier of a following DP, can be iterated without conscious processing difficulty. "Fred found her students' lecturers' books" is equally unproblematic.

Under Pritchett's account, reanalysis of local ambiguity is automatic when the *On-Line Locality Constraint* (OLLC) is observed:

(103) *On-Line Locality Constraint* (OLLC):
> The target position [of the reanalysed SO] must be *governed* or *dominated* by its source position […], otherwise attachment is impossible […].
> (Pritchett 1992b: 101)

When parsing (97) using Pritchett's model, the main clause analysis of *the horse raced* is initially favoured by GTA over the reduced relative reading because all the syntactic properties of the unergative verb and T-head can be satisfied immediately, along with the Case-feature and

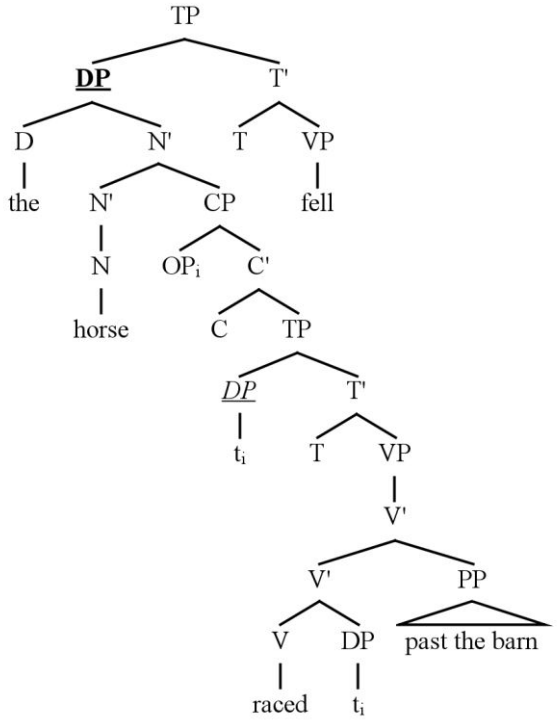
---

have to be undone once a complement of the head is encountered; the parser does not violate the *No Tampering Condition* in this way.

thematic properties of *the horse*. The Case-feature of *the horse* cannot be deleted by a passive form of a transitive verb, [223] and the DP is not assigned its main clause theta-role; furthermore, the agent theta-role of the passive verb cannot be assigned until the presence/absence of a *by*-phrase can be established. Reanalysis of (97) therefore involves the reinterpretation of the subject of *raced* (an embedded verb) – found in the 'source position' – as the subject of *fell* (the main verb) – incorporated in the 'target position'. Consider the necessary reanalysis. As the correct final analysis, shown in (104) and (105), illustrates, the target position (**emboldened**) of reanalysis is neither governed nor dominated by the source position (*italicized*), predicting a garden-path effect.

(104) [$_{TP}$ [**$_{DP}$** The [$_{N'}$ horse [$_{CP}$ OP$_i$ [$_{C'}$ C [$_{TP}$ [*$_{DP}$* t$_i$] [$_{T'}$ T [$_{VP}$ raced [$_{DP}$ t$_i$] [$_{PP}$ past the barn]]]]]]]] [$_{T'}$ T fell]]

(105)



(adapted from Mulders (2002: 34))

[223] Without getting ahead of ourselves, it is clear that there is a real question over whether the original and the correct analysis involve the same LIs.

There is, however, a set of reanalyses which are automatic despite apparently breaching the OLLC. These are associated with strings in which a subsequent non-theta-assigner demonstrates the impossibility of a previous GTA-motivated parse, usually on word order grounds. Consider (106) for instance:

(106) The spaceship destroyed in battle disintegrated.
(Pritchett 1992b: 92)

Under GTA, the parser will initially analyse *the spaceship* as the subject of *destroyed*. But the PP proves that this postulate is incorrect, since direct objects are almost invariably adjacent to transitive verbs in English: "*The spaceship destroyed in battle the aliens" is ungrammatical. The source position and target position in this example have the same relationship to each other as in (104) and (105), yet no conscious difficulty accrues on reanalysis, seemingly a breach of the OLLC. Pritchett accounts for this by suggesting that when a *non-theta-assigner* is the element to discredit a parse, then a process of *rebuffering* takes place, whereby the parser sends perceived forms back to storage, for the parse to 'start again' when the next head is encountered (in this case, *disintegrated*). This avoids any breach of the OLLC because source and target positions no longer exist; essentially, the parse begins again. As will be clear, *rebuffering* is also powerful enough to explain OLLC-consistent automatic reanalysis, so there is some redundancy in Pritchett's model, and so grounds for suspicion under methodological minimalism. [224]

Pritchett's formulation of an incremental parser, driven by the recognition of heads and maximal satisfaction of principles of syntax, contrasts sharply with 'classic' models, which rely on extraneous, *configurational* factors to determine attachment. Such approaches involve "structural complexity metrics" (Phillips 1995: 108) of one sort or another to discriminate between competing parses. The best-known of these are metrics of "global formal simplicity" (*ibid.*), such as the *Minimal Attachment* (MA) principle of Frazier & J.D.Fodor (1978). MA holds that the preferred transition to the next parse state is the one adding the fewest new nodes to the current phrase marker. When applied to classic garden-path sentences, the main verb analysis, (107), is favoured over the reduced relative one, (108), because the latter includes many more phrase structure nodes:

[224] Rebuffering is consistent with the grammaticality of heavy-*NP*-shift sentences – e.g. "The spaceship destroyed in battle the giant Kzinti cruiser which had been pursuing it for weeks" – as the parser can still construct the correct structure from the rebuffered elements once the whole clause has been perceived (Pritchett 1992b: 95).

(107) [$_{TP}$ [$_{DP}$ The [$_{N'}$ horse]] [$_{T'}$ T [$_{VP}$ raced]]] …

(108) [$_{DP}$ The [$_{N'}$ horse [$_{CP}$ OP$_i$ [$_{C'}$ C [$_{TP}$ [$_{DP}$ t$_i$] [$_{T'}$ T [$_{VP}$ raced [$_{DP}$ t$_i$]]]]]]] …

This is the basis of the MA account of the distribution of conscious processing difficulty.

A second set of formal complexity metrics emphasizes locality in the attachment of incoming elements, requiring them to attach "at the lowest possible existing node in the phrase marker" (Phillips 1995: 110). Although not a homogeneous set, these metrics include Kimball's (1973) principle of *Right Association* (RA) and Frazier's (1978) principle of *Late Closure* (LC). When applied to classic garden-path sentences, these approaches favour the main verb analysis, (107), over the reduced relative clause analysis because of the shorter distance (in terms of nodes) between T' and *the horse*. Locality metrics have typically been positioned as complements to simplicity metrics (or constraints based on grammatical satisfaction); in actual fact, there is significant empirical overlap between metrics of simplicity and locality, in ways discussed by Pritchett (1995: 111*ff.*). For our purposes however, nothing is lost by considering their predictions independently.

Phillips (1995) argues for an idiosyncratic, locally-attaching model, in which parsing is motivated by the same "economy condition" active in his model of NS. He contends that linear order is established as part of NS operations and that "new items must be attached to the right edge of the structure" (*Merge Right*) (*ibid.*: 24). (He argues that only a *Merge Right* approach to NS provides a solution to problems of contradictory constituency (*ibid.*: ch.2).) The matching parsing principle, *Branch Right* (BR), imposes "the most right-branching attachment of the incoming item by choosing the attachment that creates the shortest path through the phrase marker from the preceding item to the incoming item" (*ibid.*: 30).

(109) BRANCH RIGHT (BR)

Metric: select the attachment that uses the shortest path(s) from the last item in the input to the current input item.

Reference set: all attachments of a new item that are compatible with a given interpretation. [225]

(*ibid.*: 111)

[225] Here Phillips makes explicit the fact that simplicity metrics involve radically vacuous computation of multiple representations and relevant comparisons, of the kind discussed in §2.4.2. This contrasts with Pritchett's *locally-assessed* parsing algorithm.

The operation of BR in parsing forces the postulation of right-branching structures, just as RA and LC do. If the previous item parsed was X and the fresh input is Y, then BR insists that of the two possible parses (110a) be chosen over (110b):

(110) (a) [Z [X Y]]
(b) [[Z X] Y]

There is a range of empirical reasons to favour Pritchett's head-driven, GTA-based instantiation of incremental parsing over configurationally-oriented ones, which project structure in advance of perception. First, a head-driven approach correctly predicts the systematic difference between head-initial and head-final languages in terms of processing difficulties. A 'top-down' parsing algorithm predicts rampant local ambiguity and processing problems in languages such as Japanese, which is not attested (see Mazuka *et al*. (1989)). Taking an example from Pritchett (1992b: 151), the Japanese string "*NP*-ni *NP*-ga" is ambiguous between a dative-nominative analysis and a double subject analysis, yet constructions such as (111a) and (111b) do not cause processing difficulty, suggesting that 'speculative' parsing does not take place.

(111) (a) *Dative -ni*
Rex-ni John-ga hanasita
Rex-$_{DAT}$ John-$_{NOM}$ spoke
'John spoke to Rex'
(b) *Subject -ni*
John-ni nihongo-ga wakaru
John-$_{DAT}$ Japanese-$_{NOM}$ understand
'John understands Japanese'

Pritchett's model makes a second, more subtle prediction about the range and scope of reanalytic processing difficulties. As discussed above, the process of rebuffering permits automatic reanalysis when a non-theta-assigner demonstrates the impossibility of a previous, GTA-motivated parse. This leads us to expect that the processing difficulty in such cases will be commensurate with the amount of parsing before 'rebuffering' is motivated and the amount of structure reconstructed. This is precisely what we seem to find. In (106), for instance, where incompatibility is established early and rebuffering is limited, the processing difficulty is negligible. (112), however, presents a much more significant problem – weakly conscious processing difficulty, but without the marked confusion/deliberation of a true garden-path effect.

(In Mazuka & Itoh's (1995) terms, the distinction is between "sentences that leave the impression of conscious reanalysis" (p. 305) and "sentences that cause an even stronger garden-path effect" (p. 309), such as (97) and Japanese examples discussed shortly.)

(112) ¿Yoko-ga kodomo-o koosaten-de mikaketa takusii-ni noseta [Japanese]
Yoko-$_{NOM}$ child-$_{ACC}$ intersection-$_{LOC}$ saw taxi-$_{DAT}$ put-in
'Yoko put the child on the taxi she saw at the intersection'
(*ibid.*: 305)

The initial analysis of (112) is as a main clause, with: *yoko-ga* made the subject of *mikateka*; *kodomo-o* made the object of *mikateka*; and *koosateten-de* made the locative modifier of *kodomo-o*. When *takusii-ni* is encountered however, this analysis has to be revised to parse a relative clause (a second object isn't consistent with *mikaketa* as a main verb), and so a significant number of thematic/syntactic commitments have to be abandoned and SOs rebuffered. This lengthy abandoned parse and extensive rebuffering process would be expected to result in greater processing delay, as is borne out. [226]

The classic evidence in favour of the thematic/syntactic 'preoccupation' of parsing, as expressed by GTA, comes from the preference for argumental readings of PPs also interpretable as adjuncts, i.e. a preference for assigning a theta-role and deleting a Case-feature over not. The attachment of the adjunct phrase *for a month* to *considered* in (113a) presents greater processing difficulty than the attachment of the argument phrase to *employee demands* in (113b):

(113) (a) The company lawyers considered employee demands for a month but they did not act.
(b) The company lawyers considered employee demands for a raise but they did not act.
(Schütze & Gibson 1999: 417)

These findings are problematic for both simplicity-based (MA) and locality-based (RA/LC/BR) approaches. The introduction of a PP as an argument of a NP or as an adjunct of a VP introduces

[226] Cf. Mazuka & Itoh's (1995: 323) *Tentative Attachment Strategy*, under which "reanalysis of each decision will have a psychologically measurable cost … but any single reanalysis will not be costly enough to cause conscious processing difficulty [unless] combined with other complexities [including] *multiple reanalyses* [at which point reanalysis] may become conscious" (emphasis my own).

one extra node in each case, so MA cannot discriminate. While locality metrics produce the correct result in (113), they fall down in analogous cases where ambiguity exists over whether a PP is an argument of a VP or an adjunct of a NP, such as (114):

(114) Rex slapped the man with the board.
(Pritchett 1992b: 145)

RA/LC/BR predict that non-argumental NP-attachment will be the favoured reading in such cases, contrary to fact.

A Pritchett-style model of parsing also anticipates the indeterminate distribution of processing difficulty when local ambiguity is not resolved by GTA. This includes cases in which there is a surplus of possible theta-assigners available for a single argument. Consider (115):

(115) A waitress poured the visitors who were drinking red wine.
(Sadeh-Leicht 2007: 54)

On encountering *red wine*, theta-assignment can be maximally satisfied in two different ways: *red wine* can either take the internal theta-role of *poured* or the internal theta-role of *drinking*. GTA offers no grounds to favour either analysis, and so anticipates that they will be pursued with approximately equal frequency. The ambiguity is resolved in favour of *red wine* bearing the internal theta-role of *poured* once the clause ends. If the other analysis had initially been pursued, then reanalysis will be necessary. The indeterminacy of initial analysis ought, therefore, to manifest as indeterminacy in the experience of processing difficulty within a pool of subjects' responses to a theta-surplus construction, and likewise within an individual's responses across a range of such constructions. The prediction of MA is identical, whereas RA/LC/BR models anticipate uniform processing difficulty over theta-surplus constructions of this type, as the attachment to the immediately preceding phrase will always be favoured (cf. Sadeh-Leicht (2007: 58-9)).

Data bearing on this matter are comparatively limited, but appear to sustain the GTA/MA prediction of equivocation. Without giving details of the study, Pritchett (1992b: fn.111) cites an informal survey of Harvard students' responses to English theta-surplus constructions, which reveals marked variability in processing responses. Sadeh-Licht (2003a, 2003b) replicates these

findings for Hebrew in a formal, off-line study: 64% [227] of the subjects noted processing difficulty interpreting theta-surplus sentences (compared with 95% in response to garden-path constructions, and 2% with unambiguous sentences). [228]

Cuetos & Mitchell (1988) make findings along similar lines in studies on the attachment of relative clauses. In sentences such as (116), neither Spanish nor English locates relative clauses consistently, with English demonstrating a weak preference for low attachment, and Spanish a moderately strong tendency toward high attachment.

(116) Un alumno insultó a [$_{NP}$ los profesores de [$_{NP}$ las clases]] [que no gustaron a los estudiantes].
The student insulted [$_{NP}$ the professors of [$_{NP}$ the courses]] [that were disliked by the students].
(Gibson 1998: 26)

Again, these results are inconsistent with the predictions of LC/RA/BR-based models, but show the equivocation anticipated by GTA- or MA-based approaches.

In summary, there is a significant empirical basis for supposing a head-driven, non-configurational, GTA-based parsing algorithm. The same cannot, however, be said for the OLLC constraint on automatic parsing. Mazuka & Itoh (1995) identify examples of strings predicted to be garden-path constructions under the OLLC but which present no conscious processing difficulty whatsoever. In (117), for instance, reanalysis is forced by a theta-assigner, so there can be no appeal to rebuffering (under Pritchett's strictures), and the target position of the reanalysed SO is neither dominated nor governed by its source position.

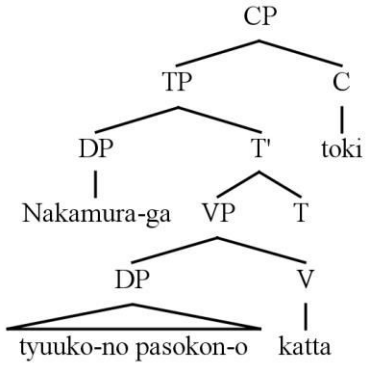
[227] As found in Pritchett's survey, the deviation from 50% is not statistically significant.

[228] Sadeh-Leicht (2007: 69) is concerned that his initial off-line study may not necessarily be an accurate reflection of "actual human performance during real-time parsing". He goes on (*ibid.*: 70*ff.*), therefore, to demonstrate that a different off-line methodology ("magnitude estimation") mirrors the results of methods directly recording on-line responses (eye-tracking procedures), and so can be considered indicative of real-time processing effects. Using this methodology, he replicates the findings of the previous studies, only now with Dutch speakers: the mean estimation of complexity for theta-surplus constructions lies in between that associated with garden-path constructions and that associated with unambiguous sentences, and differs from both with statistical significance.

(117) Nakamura-ga Ø tyuuko-no pasokon-o katta toki syuuri-site-kureta
Nakamura-$_{NOM}$ second-hand PC-$_{ACC}$ bought when repaired (for me)
'When I/(he)/(she) bought a second-hand PC, Nakamura repaired it for me'
(*ibid*.: 302)

Until *toki* is perceived, (117) is analysed as a subordinate clause, (118):

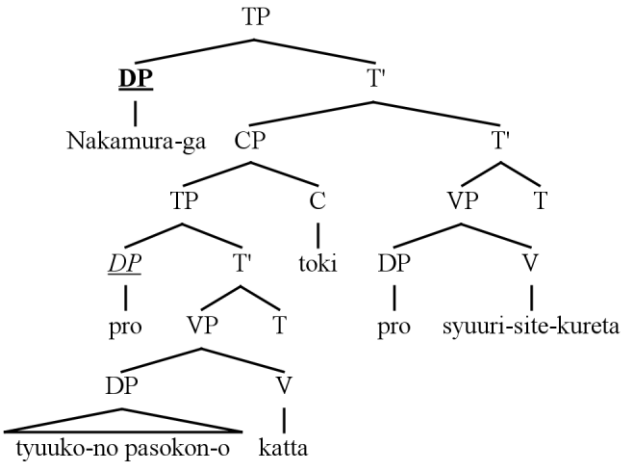
(118)



(adapted from Mulders (2002: 126))

However, when *syuuri-site-kureta* is perceived, it becomes apparent that (118) must be incorrect. The verb *kureru* can only be used when the receiver is the speaker or in the speaker's 'in-group', and it means 'give to me' when used independently, and 'do something for my benefit' (non-self-reflexively) when used as part of a compound verb (Mazuka & Itoh 1995: 302-3). For this reason, *syuuri-site-kureta* cannot take a *pro* subject interpreted as the speaker. Therefore, *Nakamura-ga* must be reinterpreted as the subject of the matrix clause, and the subject position of the embedded clause taken to be a *pro* (Mulders 2002: 127). As (119) makes clear, the target position (**emboldened**) of reanalysis is neither governed nor dominated by the source position (*italicized*), predicting a garden-path effect under the OLLC, which does not obtain.

(119)



(adapted from Mulders (*ibid.*))

This of course leaves a head-driven, GTA-based account with the problem of how to explain garden-path phenomena. Mulders (2002: 136*ff.*) dedicates extensive efforts toward revising and saving the OLLC in a way I will not discuss here, since there is a simple, more satisfactory alternative. As alluded to above, restructuring is in fact redundant: there is no need for any process in automatic reanalysis other than rebuffering. Furthermore, such empirical ground as was covered by the OLLC can be explained by independently-motivated factors: the pertinent observation is that GTA, restated here for convenience, does not determine the LIs over which it operates.

(98) *Generalized Theta Attachment* (GTA): Every principle of the Syntax attempts to be maximally satisfied at every point during processing.

As observed by Ford, Bresnan & Kaplan (1982) (cf. also MacDonald *et al.* (1994) and Trueswell (1996)), an erroneous lexical commitment seems to lie at the heart of every true garden-path construction, with mistaken syntactic commitments secondary to that. In the classic example, (97), the preterite/passive ambiguity involves two verbs with different semantics and different theta-grids: the preterite reading involves an intransitive (unergative) verb meaning something like ‘to move very quickly’; the passive reading involves a transitive verb, assigning agent and theme theta-roles, and meaning ‘to ride’ / ‘to make X race’. The confusion in parsing (97) arises

because the former LI is far more common and so assumed in first parsing; the reconsideration required at *fell* reaches 'right back to first principles'.

This account predicts that garden-path constructions will arise in any language, including rigidly head-final ones, in as far as ambiguity of this kind arises. This expectation is fulfilled by otherwise inexplicable examples from Japanese presented by Mazuka & Itoh (1995). To take just one, consider (120):

(120) Ginkoo-no torisimariyaku-ni tuita bakari-no sokutatu-o watasita
bank-$_{GEN}$ director-$_{DAT}$ arrived just-$_{GEN}$ express mail-$_{ACC}$ handed
[appointed]
'(I) handed the express mail that has just arrived to the director of the bank.'
(*ibid*.: 310)

Tuita can mean either 'became / got a position' or 'arrived'. *Torisimari-yaku* can mean either 'the position a director holds' or 'a person holding a directorship'. The initial reading of *tuita* is as meaning 'became', taking a *-ni* marked object, assuming a goal theta-role; the agent theta-role awaits assignment. The parse up to *bakari-no* is therefore: 'X just became director of the bank'. However, when *sokutatsu-o watasita* 'express mail-$_{ACC}$ handed' is perceived but no other *-ni* marked indirect object able to adopt *watasita*'s goal theta-role is apparent, then reanalysis is required. The 'arrived' reading of *tuita* is as an unaccusative, intransitive verb that takes a *-ga* marked theme, which is absent from the string; this forces a relativized interpretation of an accusative-marked noun to provide its subject, with *ginkoo-no torisimariyaku-ni* becoming its goal. The conscious difficulty with parsing this sentence is associated with trying to impose the existing arguments onto the theta-grids of verbal LIs which do not satisfy them, and must therefore be modified.

Given we are proposing a revision to the GTA-based, head-driven account, we would do well to examine the motivation for the OLLC:

(103) *On-Line Locality Constraint* (OLLC):
The target position [of the reanalysed SO] must be *governed* or *dominated* by its source position […], otherwise attachment is impossible […].

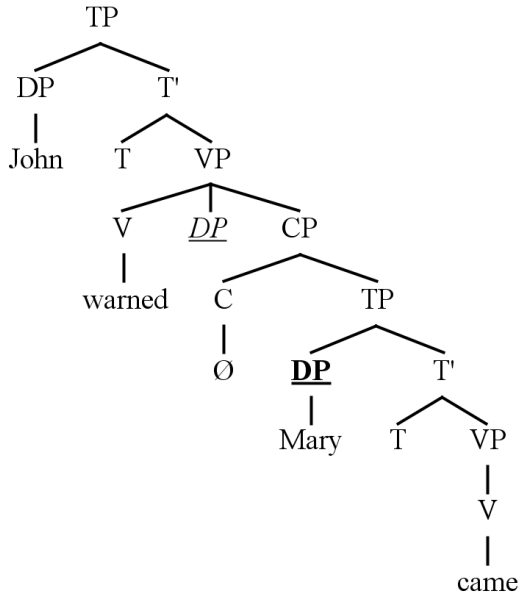
Evidence for the OLLC is adduced from the *lack* of processing difficulty associated with reanalysis between strings such as (101) and (102), in which the source position dominates the target position. Similarly, it is argued that the OLLC is necessary to account for the garden-path effect in ditransitive constructions such as (121):

(121) ¿John warned Mary came.

(Mulders (2002: 23); cf. Pritchett (1992b: 19-20, 105-6))

Under GTA, when *Mary* is encountered it will be attached as the internal argument of *warn*, taking the goal theta-role. However, when *came* is encountered *Mary* must be reanalysed as the subject of a clause, taking the proposition theta-role of *warned*. Following Pritchett's (*ibid.*: 105) assumption of ternary branching for this construction, we see in (122) that the target position (**emboldened**) of reanalysis is neither governed nor dominated by the source position (*italicized*), predicting a garden-path effect under the OLLC. [229]

(122)



(adapted from Mulders (2002: 24))

[229] The correct, binary-branching representation of (122), detailed in §4.1.7, still has the target position being dominated by the source position, so we have not shown the extra detail here.

However, it will be clear that, again, lexical commitments change in this reanalysis: the verb *warn* has been reanalysed from a LI bearing a ditransitive theta-grid and intrinsically benefactive semantics, to a LI with a single internal argument and lacking intrinsically benefactive semantics. A garden-path effect is predicted by the overwhelmingly more frequent attestation of the cautionary, ditransitive LI compared to the reportative, 'monotransitive', clause-selecting LI, and the need to change to the latter in reanalysis. The OLLC account is therefore redundant: rebuffering and correction of the LI-selection will suffice. [230]

The next logical question is what factors determine the selection of the LIs comprising the material available to the parsing algorithm. We have already suggested that the relative frequency of different LIs will significantly influence the experience of conscious processing difficulty. As Trueswell (1996: 567) puts it, "[t]he information computed when a word is recognized is used to define the set of syntactic … possibilities … relevant for evaluating possible interpretations", continued by Ford, Bresnan & Kaplan's (1982: 745) observation that "the various lexical forms … have different 'strengths' or 'saliences,' and the strongest form somehow determines the preferred syntactic analysis." MacDonald (1994) systematically demonstrates that the existence of LIs with different argument structures but the same phonetic form is predictive of increased mean processing difficulty in the parsing of reduced relative/preterite-ambiguous strings: "verb argument structure frequencies … inhibit competitor interpretations" (*ibid.*: 166). Many studies show that, in determining initial analyses, other factors are considered alongside structural preferences inherent in the parsing algorithm itself, e.g. issues of semantic and pragmatic plausibility (*i.a.* McRae, Spivey-Knowlton & Tanenhaus 1998, Ni, Crain & Shankweiler 1996). [231] However, these studies have focused on the role of such factors in resolving non-garden-pathing local ambiguities; little attention has been paid to their role in determining the initial choice of LIs themselves. The reason for this is probably because the issue is somewhat trivial. In the context of "the dog led past the door stumbled; the sheep driven past the gates shuffled; but ….",

[230] Both the strings discussed actually refer to the dominance component of the OLLC alone. The role of government as part of the OLLC is only necessitated by (a) ternary branching analyses of different ditransitive constructions, shown in §4.1.7 to be inaccurate, and (b) effects actually covered solely by GTA's preference for argumental readings of PPs over adjunctive ones. Given the OLLC has already been fully undermined, there is nothing to be gained from detailing the failure of these arguments. The reader can straightforwardly justify this to him or herself from Mulders' (2002) (47) and (48) respectively if desired.

[231] In the above discussion, sentences are treated in isolation precisely to allow the abstract algorithm and its associated structural concerns to be examined. We should not lose sight of the fact that syntactic and thematic structure must be constructed to give a sentence meaning, and so an abstract algorithm must still be postulated even within a properly 'interactive' account.

"the horse raced past the barn fell" clearly presents less of a processing challenge. Similarly, if *the boat* in "the boat floated down the river sank" is replaced by *the cargo of gold*, then the transitive reading of *floated* will be more "salient". Based on their own examples of Japanese garden paths, Mazuka & Itoh (1995: 313) observe that "[s]emantic and pragmatic information is used to narrow down the search space for possible following lexical items … enhancing and reducing garden-path … effects." [232]

There is a significant consequence to our findings regarding the disambiguation of garden-path constructions and syntactic reanalysis. By demonstrating the redundancy of restructuring operations and the OLLC, and the sufficiency of rebuffering and reselection of LIs, we have removed the necessity of *tampering* operations within the parsing algorithm. Reanalysis as advocated by Pritchett (see, for instance, (102) and (103)) involved undoing previous *Merge* operations and/or featural commitments, in breach of "informational monotonicity" (M. Marcus, Hindle & Fleck 1983). Rebuffering (independently-motivated, even in Pritchett's model) merely consists of the ability to restart a parse with the same phonetic forms. It can be conceived of as involving the break-up of an existing parse and 'adjusting' LIs (i.e. rampant tampering), but there is no reason to suppose that anything other than abandonment is necessary. In summary, there are suggestive, if not conclusive, findings which implicate the *NTC* in the parsing algorithm.

The similarity between the proposed parsing algorithm and the proposed syntactic algorithm is now very close indeed. Parsing is driven by fulfilment of the syntactic needs of heads, without reference to extraneous, configurational factors, or considerations of a 'look-ahead' variety, very much akin to phase-head-based, feature-driven, locally-assessed BPS, which itself eschews X-Bar Theory postulates and representational considerations. As Phillips (1995: 271) points out, GTA is directly comparable to the *Earliness Principle* of syntactic derivation introduced in §5.1, both shown below for comparison:

(98) *Generalized Theta Attachment* (GTA): Every principle of the Syntax attempts to be maximally satisfied at every point during processing.

(123) *Earliness Principle*: Featural needs must be met at the earliest possible time.

[232] Interestingly, prosodic factors do not seem to play a significant role in disambiguating garden-path constructions (J.D.Fodor 2002), as discussed in Wagner & Watson (2010), although they do seem to play a significant role in parsing more generally (again see Wagner & Watson (2010)).

Both implicate the cognitive principle of *MinCID* and both operate over the formal featural content of LIs. Likewise, we have seen that both syntax and parsing observe the *NTC* (cf. §4.1.2), and therefore extension conditions, although in the parser's case this involves *Merge* to labels (nonetheless SOs), rather than directly to LIs. This seems to be the only relevant distinction at the *abstract* level. (In *practice*, parsing computes over a range of non-syntactic considerations, discussed briefly above, and must also confront ambiguities: which is merely to say that it takes its input from actual utterances, rather than mind-internal Numerations.)

This contrasts sharply with the standard positon that the parser and the grammar must be largely distinct. Phillips (1995: ch.5) presents the full range of arguments offered in support of this claim, including: (a) the fact that there are many sentences which are perfectly grammatical but essentially impossible to parse, including centre-embedding sentences (see §3.1.2); and (b) that there are many sentences which are easy to understand despite being ungrammatical, including those violating restrictions on double-object constructions ("*John donated the museum a painting", etc. – cf. §1.3.3.4). We follow Phillips' overall refutation of these proposals, presenting his counter-arguments to (a) and (b) as a sample of his reasoning. The objection to claim (a) is that it does nothing more than observe that parsing and syntactic computation aren't exactly the same thing, which is not the matter at hand. Interpretation of production is bounded by WM in a way that production isn't: the parsing and grammatical algorithms may be similar / identical, merely operating in different contexts. The objection to (b) is that there is a distinction between successful parsing and successful comprehension, which may typically involve parsing, but is by no means exhausted by it.

On the basis of the above, we tentatively suggest a very strong formulation of the 'Grammar-Performance Correspondence Hypothesis' (see fn.24) with respect to (the abstract elements of) parsing: an identity relationship formulated as something like the *Grammar is Parsing* (GIP). This reverses Phillips' suggestion that *Parsing is Grammar* (PIG) – (a) to draw a distinction between the argument above and his, which is based on a model of syntax and parsing we reject, and (b) because it makes it clearer that our position does not take parsing capacities to be exhausted by grammar. [233, 234]

[233] In fn.24 we also observed the correspondence between I-phonology and perceptual abilities. Whether this is in any sense a relationship of identity is, to the best of my knowledge, unknown.

[234] The same idea is explored in different ways by *i.a.* Phillips (1995) (see above), Gorrell (1995) and Weinberg (1999). And of course by Pritchett (1992a, b) himself, who notes that "the core of parsing is in essence simply the local application of global grammatical principles" (Pritchett 1992b: 4).

Not only does it appear that the parsing algorithm makes extensive use of domain-general principles of optimal cognition, but it also appears that it fully exploits the symbols and COs of the grammar (at least syntactic grammar) in accordance with *TLTB*. Both these findings are consistent with evo-devo logic in the proposed narrative of the evolution of language's externalisation: both domain-general and language-specific ancestral cognition seem to be implicated in the design of putatively less ancestral components of the FL.

6.2 The linearisation of syntactic structure

A common assumption in theories of grammar is that syntactic computation deals with only the hierarchical relations between SOs, and their linear/temporal order is imposed in the phonological computation. Under this conception, syntactic structure is underlyingly *symmetric*. The classic argument against the need for linear relations within syntax comes from Reinhart's (1976, 1983) observations regarding Binding Principle C. Under a 'precedence-and-command' view of binding, object pronouns preceding R-expressions (and within the same binding domain, the clause (see §5.1)) ought to bind them and rule out co-reference, like subject pronouns. (124) and (125), however, reveal this is not the case:

(124) (a) Rosa is kissing him$_i$ passionately in Ben$_i$'s high school picture.
(b) People worship him$_i$ in Kissinger$_i$'s native country.
(Reinhart 1976: 79)

(125) (a) *She$_i$ is riding a horse in Rosa$_i$'s high school picture.
(b) *He$_i$ was killed in Hoffa$_i$'s home town.
(*ibid*.: 68)

With clause-level adjuncts, as in (124a) and (124b), the R-expression is not c-commanded by the pronoun, and co-reference is possible, whereas in (125a) and (125b) they are, and co-reference is prohibited. (124a) and (124b) contrast with other constructions containing R-expressions within rightward constituents – VP-adjuncts shown in (126a) and (126b).

(126) (a) *I can't even find him$_i$ in your picture of Ben$_i$.
(b) *The gangsters killed him$_i$ in Hoffa$_i$'s hometown.

In these cases, the R-expressions are c-commanded by the object pronouns and co-reference is impossible. Precedence of the object pronoun seems to be irrelevant, with binding dictated by the height of adjunction and resulting c-command relationships.

Kayne (2013) makes an argument for the role of precedence in establishing binding relations, adducing co-reference possibilities between antecedents and pronouns in "configurations of non-c-command", such as (127) and (128):

(127) The fact that $John_i$ is here means that he_i's well again.
(128) The fact that he_i's here means that $John_i$ is well again.
(*ibid.*: 16)

He and *John* do not c-command each other in these examples, but are interpretable as co-referent. "Forward pronominalization" (FP) of the kind seen in (127) is apparently universal cross-linguistically. However, "backward pronominalization" (BP) seems to be impossible in certain languages, such as Haitian Creole and Jacaltec, and problematic in others, such as Danish and Chinese. While Kayne interprets this asymmetry as evidence that precedence might well be part of syntax "in a sense to be made precise" (*ibid.*: 17), he does, however, also concede that this only applies "[t]o the extent that the backward vs. forward pronominalization question is one of … syntax". On the basis of the available evidence, this seems unlikely. First, even in languages in which BP is robustly available, it is not obligatory: in (128), the reading that *he* and *John* are not co-referent is also possible – *John*'s presence might mean, for instance, that his brother has recuperated. This strongly suggests that the process conferring co-reference in BP is contextual/interpretive, rather than grammatical. The same argument is made by the existence of languages in which BP is merely problematic, implying that other properties of these languages make interpreting subordinate clauses as relevant context difficult. In languages in which BP is systematically impossible, a natural conclusion is that these independent issues have become insurmountable. In the case of Haitian Creole, Kayne (*ibid.*: fn.39) even speculates as to what this independent issue might be – the lack of rightward displacement of heavy-*NP*s. This would presumably condition (the interpretive modules of) the language user never to consider rightward context as equivalent to leftward context, always a possible source of co-reference for unbound pronouns. [235] The asymmetry between FP and BP has a natural explanation under a contextual/pragmatic understanding of how co-reference is afforded in "configurations of non-c-

[235] Cf. fn.24 for discussion of grammar plausibly conditioning performative modules.

command": leftward string is existing context, whereas rightward string is not. Under an incremental approach to parsing in which contextual information is available for interpretation *ab initio*, this leads to the expectation that FP will be favoured over BP. There seems little reason to admit precedence a role in syntax.

The seeming absence of linear order from syntax accords with the domain-general cognitive principle of *MinRedup*. It avoids stating in syntax what must be independently restated at PF (Chomsky 1995c: 340). It is also consistent with the possibility that externalisation was not a pressure motivating the development of syntax (part of the fourth Minimalist proposal). Of course, the question of how linear order is imposed on symmetric structure in the mapping to PF remains open.

There is no consensus on this issue among generative linguists. This is at least partly attributable to the fact that our theory of syntactic structure is itself a work in progress, confounding efforts to determine how it is linearised. Various models of linearisation have been proposed, varying (in as far as they are specific) with respect to: (i) the underlying order (or orders) they consider to be basic, (ii) their mechanisms, (iii) their grammatical location, and (iv) the theories of syntax they assume. Furthermore, among the proposals which are explicit with respect to the latter three issues, many are formulated in ways which do nothing to preclude their adaptation into different frameworks. Reviewing these proposals, their empirical bases, and their difficulties is not a feasible task here; I merely note that none of them is unproblematic, including the one pursued briefly below. [236]

A particularly influential theory of linearisation is Kayne's (1994) *Linear Correspondence Axiom* (LCA), which gives rise to a SPEC-H-COMP underlying word order:

(129) Linear Correspondence Axiom (LCA) (informal)
If α asymmetrically c-commands β then α precedes β.

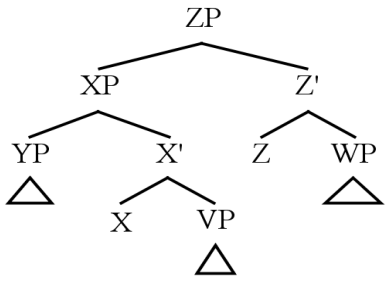
The original LCA assumes X-bar-theoretic syntax, fraught with empirical and methodological difficulties (see §4.1.3), and the imposition of linear order within syntax, problematized above.

[236] For just a flavour of some of the rich debate, see Biberauer's (2008) discussion of various head-parametric approaches; Broekhuis's (2008: 73-96) discussion of proposals by Haider ((1997a, b), 2000, 2003) and Barbiers (2000); Takita's (2009) discussion (including his self-criticisms) of Lin's (2006) adaptation of Kayne's (1994) proposals; discussion of Kayne (1994) below; etc.

As Chomsky (1995c: 34) points out, however, there is nothing to prevent the LCA from being recast as a principle of Spell-Out acting on the symmetric output of BPS. Various issues relating the LCA's ability to impose linear order still persist under this relocation however.

Following the presentation of these problems in Sheehan (2013b), we note that in complex structures such as (130) (shown in X-bar notation for ease of reference only), contradictory ordering commands are given by the LCA. Z' asymmetrically c-commands YP, X', X, and VP, but XP also asymmetrically c-commands Z and WP. To address this problem, it is necessary to incorporate a distinction between categories and segments into the definition of c-command, (131), so that specifiers are treated as adjuncts in the sense of May (1985) – an idea which can be incorporated under both X-bar-theoretic and BPS assumptions (Chomsky (1995c: 242, 437), (2001: 40)).

(130)



(131) Definition of c-command for the LCA:

X c-commands Y iff X and Y are categories and X excludes Y and every category that dominates X dominates Y.

(Kayne 1994: 18)

By removing Z' as a c-commander in this way, however, we introduce underdetermined ordering instructions. For instance, there is no asymmetric c-command between Z and VP in either direction. This requires that Kayne rely on the notion of dominance as well as asymmetric c-command in LCA-based linearisation. This is presented in Uriagereka's (1999) two-step characterization of LCA computation:

(132) Uriagereka's (1999: 252) formulation of the LCA:

(a) Base step: If α asymmetrically c-commands ß, then α precedes ß.

(b) Induction step: If α precedes ß and α dominates γ, then γ precedes ß.

This ensures that a complex specifier like XP (including VP and YP) precedes Z (and WP).

Under a synthesis of the LCA and BPS, a further issue arises, concerning the incomplete ordering of the 'base pair' of the derivation when a head selects a head (e.g. a verb taking a pronominal object) on account of mutual c-command (see Nunes (1999), Moro (2000), M. Richards (2004a, b)). Moro (2000) describes this relationship as "a point of symmetry". Various suggestions have been made as to how to render points of symmetry orderable by the LCA, including: obligatory movement of one of the elements (Moro 2000) [237] and the cliticization of one element to the other by head-adjunction (Chomsky (1995c: 337), Uriagereka 1998, Nunes 1999). However, these suggestions clearly require 'look-ahead', and so global computation of great complexity. Another proposal, by M. Richards (2004a, b) (cf. Epstein *et al.* 1998), involves systematically ignoring half of all computed precedence relationships, and so is also rampantly redundant – questionable on methodological grounds.

Here I pursue a suggestion of Chomsky's (2007b: 9) that the *fixed* word order properties of languages might be best explained in terms of the nature of the parser, in the same way as *non-rigid/discourse-level* word ordering phenomena. He implies that by determining the nature of parser and considering the syntactic/thematic properties of, and relationships between, the *abstract syntactic positions* – H, COMP, SPEC, ADJ [238] – we might gain some insight into the linearisation of syntactic structure. In particular, I suggest that the LCA can be reduced to the principles of SPEC-before-COMP, H-COMP adjacency, and SPEC-H adjacency, each of which plausibly reflects the action of a computationally-optimal parser on the output of syntax. This

[237] The origin of the "dynamic antisymmetry" approach co-opted to render SOs labellable (see §4.1.8).

[238] For the purposes of this section (and the rest of the dissertation), we take upper-case lettering to refer to the abstract, relationally-defined *positions* in the syntactic structure, rather than the elements occupying (or originally occupying) those positions. H refers to the head position, COMP to the complement position, SPEC to the specifier position, and ADJ to the adjunct position. The latter three are defined solely in terms of initial *set-Merge*, non-initial *set-Merge*, and *pair-Merge* respectively. In particular, we do not take SPEC to encode any information about the endocentricity of the phrase (cf. Chomsky (2010, 2013, 2014) and the discussion of labelling in §4.1.8 above).

gives a principled account of SPEC-H-COMP underlying word order, therefore maintaining the empirical coverage of LCA-based approaches, but avoiding their stipulations and difficulties.

This is not the place to discuss at length the evidence for and against underlying SPEC-H-COMP linearisarion, not least because this section is partly just a "proof of concept" in regard of the utility of parsing explanations; to some extent, it anticipates its redundancy in light of a better understanding of syntactic algorithms and a fuller analysis of the linearisation data (co-dependent issues). It is, however, worth reiterating the point made in M. Richards (2004b: 186) (regarding the LCA) that "the 'Spec-Head-Comp' 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, evidence in favour of SPEC-initial [239] (a fact, at any rate, largely unchallenged in the literature) has been taken as evidence in favour of an LCA-based approach to linearisation, extrapolating beyond the data. However, there is some suggestive evidence for H-COMP linearisation also. In particular, in canonical head-final word ordering the head and complement are commonly non-adjacent, suggesting the order has been derived by the movement of the complement over an intervening SO, as discussed in §3.2.2.2 (cf. Zwart 1993). Also persuasive in this regard is Carstens' (2002) finding that serial verb constructions are cross-linguistically uniform in ordering the higher verb before the lower (complement) verb. There is, however, one obvious repercussion of assuming SPEC-H-COMP linearisation: namely, all head-final ordering must be derived by movement out of COMP. In the case of phrases headed by non-phase-heads, this is observed to be problematic (see Abels (2003) and Chomsky (2008: fn.31)), a potential difficulty for such an account. Assuming, however, that the data do reflect SPEC-H-COMP linearisation, we now consider the action of the parser over abstract syntactic structure, and whether this offers any explanation of the finding.

We first observe that another seemingly uncontested fact is underlying H-COMP adjacency. Under the model of parsing discussed above, there is an entirely straightforward account for this: H-COMP adjacency minimizes the carry-over in WM of either the complement, pending arrival of a head, or of unfulfilled syntactic properties of the head, pending the arrival of a complement. This is consistent with *MinCUI* or *MinCID*. [240] Assuming (pending further discussion) that there

[239] For instance, the positioning of floating quantifiers before VP, marking the base-generated position of subjects (Hornstein, Nunes & Grohmann 2005: 86-8); analyses of VOS languages (Paul 1997, Pearson 2001, Aldridge 2004, Coon 2010); etc.

[240] Within his functionalist framework, Hawkins (*i.a.*2004: 252) takes H-COMP adjacency as consistent with *MinD*.

is only one underlying order imposed on syntactic structure and that phrases do not overlap, there are six possible orders of H, COMP and SPEC, of which the constraint of H-COMP adjacency eliminates two, leaving: SHC, SCH, HCS and CHS.

We now simply make the equivalent argument regarding the adjacency of SPEC and H. SOs base-generated in SPEC may fulfil theta-assignation-features, c-selection features, extended projection requirements (i.e. the EPP), etc. associated with their head. SOs raised into SPEC may fulfil the head's Case-features, EPP-features, EF-features, etc., and 'matching' features may be satisfied in return. [241] Again therefore, a head-driven parser either has to cache the specifier or unfulfilled properties of the head, and this burden is minimized by linearising SPEC and H adjacently, in accordance with *MinCUI* or *MinCID*. Of the four remaining orders, this suggests that SCH and HCS will be disfavoured by optimal functioning of the parser discussed above, leaving SHC and CHS.

It will be noted, of course, that SPEC is the target position for (phrasal) *Internal Merge*, i.e. the position to which displaced phrases are moved. While some SPEC positions are projected under *External Merge* (e.g. external arguments, expletives, indirect objects, etc.), every case of *Internal Merge* uncontroversially part of syntax projects a SPEC position, [242] and such displacement of course results in a *filler-gap* relationship. For reasons discussed in §6.1.3, computationally-optimal, incremental parsing insists that fillers precede gaps. For the same reason, it is anticipated that the abstract location of displaced elements, SPEC, will precede their source domain, COMP, i.e. that movement will be underlyingly leftward. [243] This narrows the remaining underlying word order options to one: SPEC-H-COMP.

If our logic follows, the parser imposes a single underlying word order (and does not allow phrases to interlace their parts) in linearisation. A 'Universal Base' seems independently preferable, on the grounds that a parser which must entertain several possible underlying schemas

[241] Exactly which depends on assumptions about the syntactic algorithm, but at least Case-features. Under strict cartographic models of the left periphery (e.g. Rizzi 1997), a head will also provide information about the identity of its specifier.

[242] As mentioned in fn.149, the status of head-movement remains an open question.

[243] There is some suggestion that certain sign languages exhibit underlyingly rightward movement in *wh*-contexts (Cecchetto, Geraci & Zucchi 2009). However, Abner (2011) proposes a leftward-moving derivation of many of the same facts. If a true possibility, this phenomenon is presumably related in some way to the different parsing challenge presented by the signed modality.

(cf., for instance, Abels & Neeleman (2009)) will have a commensurately more complex computational task, introducing unnecessary string-ambiguity and redundant exercises in structure-building. This is somewhat similar to the concern for avoiding computation of reference sets discussed above.

And so, if the assumption of a SPEC-H-COMP underlying order is indeed apposite, it seems plausible to suggest it results from the optimal functioning of the parser over syntactic structure. This also avoids the need for the independent symbols, operations, or computations associated with the LCA, in further accordance with substantive and computational optimality. There is no obvious need for the LCA and its associated machinery (c-command, domination, a category/segment distinction, procedures to deal with points of symmetry, etc.) in order to begin explaining empirical findings: independently-motivated devices seem to produce the same result.

This approach also offers some insight into the linearisation of the ADJ position, which, under the approach to adjunction outlined in §4.1.5, falls outside the remit of the LCA, being invisible to c-command. [244] As discussed, SOs in ADJ positions have no syntactic impact on the SO to which they are adjoined: they are introduced on a "separate plane". This immediately lessens/removes any constraints on their linearisation imposed by caching unfulfilled syntactic features associated partners, although not constraints imposed by the caching of the adjunct itself. This offers a suggestive account of the possibility of ADJ-finality: *MinCID* will not be a relevant concern in establishing the relationship between an adjunct and its partner. [245]

The parser nonetheless needs to postulate a *pair-Merge* relationship between ADJ and its partner to recover semantic composition. The standard adjacency of ADJ and its partner can therefore be seen as the reflex of *MinCUI* in the operation of parsing in the case of ADJ-XP ordering, or, in the case of XP-ADJ orderings, as a reflex of the limits on WM, which minimize the distance over which syntactic/semantic relationships (even optional ones) will be established (see §1.3.3.1). [246] This (emergent) principle of linearisation will operate over a grammatical hierarchy of adjunct

[244] The Kaynean approach to adjunction treats adjuncts as akin to specifiers, failing to make the distinction discussed in §4.1.5.

[245] Under Kaynean approaches to adjunction (cf. *i.a.* Kayne 1994, Cinque 1999), adjunct-finality, like head-finality, must be derived by movement, since, as adjuncts are structurally identical to specifiers, they are obligatorily left-branching.

[246] The limits on WM are also relevant in the derivation of amenable surface word orders from underlying ones through movement (see §1.3.3.1).

attachment (see §3.2.2.1, and Cinque (1994, 1999, 2010)), and so the partner of an ADJ may have already taken part in a previous *pair-Merge* operation. The option of ADJ-finality, the principle of *pair-adjacency*, and the rigid hierarchy of *pair-Merge* operations will together preserve a relative order of adjuncts – in a left-to-right fashion before a mutual partner, and in right-to-left fashion afterwards (cf. Koster 1974, Pearson 2000, Rackowski & Travis 2000, Svenonius 2005) – and do so without necessarily appealing to 'roll-up' movement accounts, an obligatory part of the explanation under Kaynean approaches to adjunction (see Svenonius (2005) for illustration).

Other properties are known to condition the linearisation of adjuncts, including their semantic 'orientation' (see Jackendoff (1972) and Ernst (2003)), but discussion of this issue (and its interaction with SPEC-H adjacency and *MinCID*) takes us too far afield, and we bring an end to our discussion of the parser and linearisation here.

6.3 *The Final-over-Final Constraint*

A robustly attested generalisation over possible surface word orders is the so-called *Final-over-Final Constraint*.

(133) The Final-over-Final Constraint (FoFC)
Where α and β are members of the same extended projection: if α is a head-initial phrase and β is a phrase immediately dominating α, then β must be head-initial; if α is a head-final phrase, and β is a phrase immediately dominating α, then β can be head-initial or head-final.
(adapted from Biberauer, Holmberg & Roberts (2014))

An extended projection is a "spine" of elements which are *categorially identical* and select each other. The 'clausal' spine consists of a lexical VP and projections of *v*, T and C (and any other projections vindicated by cartographic research). The 'nominal' spine consists of (something like) a lexical NP and projections of *n*, Num, D and Adposition (cf. §3.2.2.1). More formally, (133) can be stated as:

(134) The FoFC (formal)

*[$_{\beta P}$ … [$_{\alpha P}$ … α γP] β …]

where

a. αP is immediately dominated by a projection of β, and

b. α and β have the same value for [±V].

(*ibid.*: 199)

(134) is motivated by facts such as the absence of VO languages with final complementizers (Hawkins (1990: 256-7), Dryer (2009: 199-205), Kayne (2000: 320-1)). The configurations which could result in this surface order are:

(135) (a) [$_{CP}$ [$_{TP}$ [$_{VP}$ V O] T] C]

(b) [$_{CP}$ [$_{TP}$ T [$_{VP}$ V O]] C]

(135a) violates (134) with β=T and α=V, while (135b) violates (134) with β=C and α=T.

Similarly, in the nominal domain, Finnish manifests mixed headedness, having both prepositions and postpositions, and both N-Comp and Comp-N orders. The one permutation of these possibilities that is not attested is the FoFC-violating *[[N-Comp] PostP] order. This and a wide range of other data, both synchronic and diachronic, are adduced in support of the FoFC in Biberauer, Holmberg & Roberts (2014).

Various explanations have been proposed for the FoFC. They include: Biberauer, Holmberg & Roberts' (2014) account based on the spread of movement diacritics associated with 'spine-defining' features down an extended projection; Sheehan's (2013a, b) account in terms of a "copy theory" of labelling; Hawkins' (2013) suggestion that FoFC is not motivated by the data (see Sheehan (2013b) and Biberauer, Holmberg & Roberts (2014: online appendix) for discussion); Philip's (2013) account within Optimality Theory (again, see Biberauer, Holmberg & Roberts (2014: online appendix) for discussion). It is not my intention to detail or assess these proposals, merely to make an alternative suggestion.

This alternative shares the observation underlying Cecchetto's (2013) functionalist account of the FoFC, where he notes that the relationship between higher and lower heads of the same extended projection is analogous to the relationship between fillers and gaps. In each case, the relationship is "asymmetric" in two ways. First, the antecedent entails the consequent: the parser knows that a

filler has a matching gap; similarly, as observed in fn.138, "in a rigid, conceptually-determined hierarchy, the 'route' to the higher functional categories is through the lower ones". For instance, it is a grammatical / conceptual fact that complementizers operate over events and are otherwise not projected, etc. (see §3.2.2.1). Secondly, the consequent element relies upon the antecedent element for specification of its relevant properties. A gap is a "nothing", and only when a copy of the filler is posited in its stead can syntactic/thematic properties be established. Similarly, the lower member of a conceptual hierarchy relies on the higher member for specification of its number, definiteness, finiteness, mood, etc. At this point the analogy breaks down somewhat however, since a gap entails a filler, whereas a lower functional head does *not* entail a higher functional head. Lower functional material can project without enrichment from higher material. Examples include, say, bare-TPs found in ECM infinitives ("I thought her to have left") (Cecchetto (*ibid.*: 67)) and raising-to-subject constructions ("She seems to have left"), or the bare-*n*Ps discussed in §3.2.2.1.

The FoFC amounts to an adjacency requirement on members of a fixed conceptual hierarchy if presented 'out of order' in left-to-right parsing. The difficulty of this "inverted dependency" (*ibid.*: 68) (not an entirely accurate characterization as just seen) and the reason "it is acceptable only if the dependency is very local" (*ibid.*) can be understood in two parallel ways. If a lower functional head is perceived first, its relevant conceptual specification in terms of the functional hierarchy cannot be established, and this unfulfilled predicate-argument relationship must be held over in WM, in breach of *MinCID*. Alternatively, this can be thought of as delaying answering the question of whether a lower functional head is an argument of a higher predicate at all, since, as we have seen, lower heads may operate as predicates in their own right, with 'generic' semantics relative to higher functional material. This suspends left-to-right parsing, since it postpones perfectly valid interpretations in light of an awareness of the possibility of alternatives – a breach of *MinCUI*. This issue (presented in two different ways) is avoided if the higher functional material is reached first, since the conceptual template then entails existence of the lower predicate-argument. The adjacency requirement which holds when an "inverted dependency" obtain – i.e. the FoFC – can therefore be considered a reflection of concern for WM in parsing, reducing the inevitable carry-over to the *absolute minimum*, under the force of *MinCID* / *MinCUI*.

This completes our discussion of linguistic Minimalism and the (adult) parser.

7. Minimalism and Acquisition

Chapter 6 attempted to bring certain word order effects within Minimalist research, and this chapter attempts to do the same for certain aspects of FLA. In particular, it investigates the possibility that acquisitional cognition employs domain-general, third-factor principles. This chapter follows on from the previous one in a more concrete sense also: it construes acquisition as intimately related with efforts by LA-ers to parse their PLD, using over-rich, immature grammars.

The evidence assessed will be drawn both from studies of acquisition in real-time and from studies of language change. The use of diachronic findings assumes an acquisitional basis for language change (Paul 1920), which falls out from the indirect transmission of grammars across generations. "[G]rammars 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 [anew] by the individuals of each generation" (Lightfoot 1979: 391). This entails that "diachronic change can represent crucial information on those factors that learners rely on to select hypotheses" (Clark & Roberts 1993: 302).

7.1 Statistical learning

As discussed in §2.3, the logic of evo-devo anticipates the existence of third-factor principles which "enter into all facets of [the] growth and evolution [of cognitive systems]" (Chomsky 2007b: 3). Chomsky presents the following characterization of these third factors:

(24) (a) *principles of data analysis that might be used in acquisition;*
(b) *principles of structural architecture and developmental constraints;*
(Chomsky 2005a: 6)
(c) *properties of the human brain that determine what cognitive systems can exist.*
(Chomsky 2007b: fn.4)

The cognitive principles of substantive and computational optimality focused on above are found in (24b), while the nature of neural matter is found in (24c). (24a), however, includes domain-general principles of learning which LA-ers might use to converge upon adult grammars (in conjunction with innate principles of language structure).

One domain-general principle of learning, implicated across a range of tasks (see §2.3), is a means of tracking of probabilities associated with particular analyses. In accordance with the expectations of the third Minimalist proposal, there is clear evidence of its involvement in FLA, found in work by: Yang (*i.a.* 1999, 2000, 2002, 2004) on syntactic development; Legate & Yang (2007) on morphological development; Legate & Yang (2012) on phonological development; and Shulka, White & Aslin (2011) (partly reviewing previous work) on speech segmentation.

Following here the presentation given in Yang (2002), we see that discrete views of syntactic acquisition run into difficulty with the facts of development. In *transformational* models (TMs) of parameter setting (*i.a.* Chomsky 1981, Hyams 1986, Gibson & Wexler 1994), LA-ers change hypotheses in an all-or-nothing fashion, with grammatical options turned 'on and off' in response to particular PLD. We follow Yang (2002: 18) in taking Gibson & Wexler's (1994) TM as representative, under which the child will "switch" parametric settings one at a time until the data is accurately characterized. [247] However, this conflicts with the finding that child language development seems to be gradual. Assuming that LA-ers' productions are indicative of the grammars they currently assume, TMs anticipate that "abrupt changes in linguistic expressions [will] be observed" (*ibid.*: 20), a prediction apparently not borne out. For instance, a longitudinal study of the V1 and V2 productions of Dutch LA-ers (Haegeman (1995), adapted in Yang (2002: 106-7)) indicates that the percentage of V2 productions rises gradually from c.50% use at 2 years 4 months old to c.80% at 3 years old.

To remedy the empirical insufficiency of TMs, Yang (*i.a.* 1999. 2000, 2002) introduces a probabilistic element to parametric learning. He suggests that:

[247] Whether this is a meaningful way of construing grammatical decision-making in light of discussion in §3.2 is a matter returned to in the following section.

(136) For an input sentence *s*, the child:

(a) with probability P_i selects a grammar G_i

(b) analyzes *s* with G_i

(c) – if successful, reward[s] G_i by increasing P_i

– otherwise, punish[es] G_i by decreasing P_i

(Yang 2010: 1162).

This *variational* method has alternative grammars co-existing (cf. Roeper 2000), with the target grammar gradually becoming more prominent over time, while non-target grammars fade to extinction. Another example of gradual change in FLA is afforded by the development of compulsorily overt pronominal subjects in non-null-subject-language: Bloom (1993) failed to discover any sharp change in the frequency of overt subject use in the data from Brown's (1973) study of the "null subject stage" of two American children.

It seems that a domain-general principle of data analysis acts in FLA, in concert with innate grammatical knowledge and PLD "converted to experience that selects one or other language from within a narrow range". [248] This is consistent with the implication of *TLTB* in FLA and the logic of the third Minimalist proposal.

7.2 'Parsing-acquisition'

This section begins by presenting an important theoretical consideration bearing on Yang's variationist model, turning away from the empirical argument.

Berwick & Niyogi (1996) demonstrate that, even using their three parameter space, Gibson & Wexler's (1994) deterministic model of learning runs into difficulties with *local maxima* – "non-target grammars from which the learner can never reach the target grammar" (Yang 2002: 18). As Kohl (1999) demonstrates, this is a problem which becomes more serious as the parameter space grows. Of course, this would not be a problem if LA-ers were equipped with information about the order in which parameters were to be set. And, indeed, it does seem that FLA shows a particular order of acquisitional steps (Brown 1973). Within transformational frameworks, this

[248] As discussed in §2.3, a (24a)-principle of data analysis will reflect constraints imposed by "principles of structural architecture", (24b), in turn informed by "properties of the human brain", (24b). How this is expressed in this case is currently a matter for speculation alone.

has led to speculation regarding the need for innate mechanisms delaying the setting of particular parameters.

One approach, proposed by Dresher & Kaye (1990) and Dresher (1999) (cf. Lightfoot (1999) for a modified version in the syntactic domain), is that parameters are associated with *cues*, pieces of input which unambiguously determine the values of the parameters in a language. Dresher & Kaye (1990) suggest that LA-ers innately possess both knowledge of the cue associated with each parameter and a default value for each parameter. The presence of the cue in the PLD will cause the LA-er to set the parameter accordingly. The cues are also taken to be innately ordered, resolving the problem of local maxima.

Proposing another TM, a different approach is taken by J.D.Fodor (1998) and Sakas & J.D.Fodor (2001). Their *Structural Triggers Learner* (STL-er) will try to parse PLD in order to understand it. On encountering a sentence not licensed by the current grammar, parsing will break down at some point in the string. The STL-er's assumed parsing device will then look for ways to patch up the parse, by drawing on whichever parameter values are made available by innate knowledge, but absent from the current grammar. It then adopts whichever values are appropriate to render the current grammar compatible with the input. This parsing device possesses the STL-er of the ability to determine which sentences are ambiguous with respect to certain parameters; the values of these parameters are left unchanged and only determined when completely unambiguous PLD is encountered, resolving the local maxima problem.

Without presenting the detailed discussion (see Yang (2002: 102-9)), we note that a secondary effect of Yang's variationist model is that, when it is applied to real PLD bearing on Gibson & Wexler's parameter space, [249] the correct order of parametric setting arises naturally. Some parameters accumulate decisive probabilistic support before others, depending on the frequency of the different types of unambiguous evidence providing such support, and their relationship/overlap with each other. The local maxima problem can therefore be resolved in a way compatible with empirical facts.

The most obvious problem with Yang's model is that it has no account of *how* the grammar G_i selected "analyzes *s*" so that its probability can be adjusted accordingly. The cue-based approach

[249] Gibson & Wexler's (1994) three-parameter space consists of: finite verb raising, as seen in French, acquired relatively early (Pierce 1989); obligatory subject use, as in non-null-subject-languages, acquired relatively late (Valian 1991, Bloom 1993); and V2 word order, acquired relatively late (Haegeman 1995).

discussed above more or less trivializes the learning process, to such an extent that Dresher (1999) (cf. Lightfoot 1991) even suggests that 'error-driven' parsing can be dispensed with entirely, given a sufficiently developed system of (innate) cues. However, this is inconsistent with empirical arguments against TMs. The cue-based mechanism of parameter setting is reformulated in the concept of *parameter expression* introduced by Clark & Roberts (1993), expressible as follows:

(137) Parameter expression:

A substring of the input text S expresses a parameter p_i just in case a grammar must have p_i set to a definite value in order to assign a well-formed representation to S.

(Roberts & Roussou 2003: 15)

With a trigger being defined as:

(138) Trigger:

A substring of the input text S is a trigger for parameter p_i if S expresses p_i.

(Roberts 2007: 232)

This formulation of the role of PLD in parameter setting dissociates cues from parameters and relocates the 'causative experience' to the world, rather having it as part of mental representation. Despite justification for this (see discussion in Roberts (*ibid.*: 242-5)), [250] and appropriate clarification of the related notion of the *P-ambiguity* associated with a substring (*ibid.*: 233), the process a LA-er goes through in interpreting a substring as expressing/not expressing a parameter remains obscure. While an essential tool in understanding diachronic change, appeal to 'error-driven' setting still begs the question.

As Yang (2002: 42) correctly observes the STL model (as well as being a TM) "seems to introduce computational cost that is too high to be realistic". The special version of the parsing device posited has to be able to simultaneously countenance and compare the effects of all the different parametric options made available by innate grammar. Nonetheless, the STL model does at least confront the issue of how a LA-er might go about interpreting PLD as indicative of particular parametric settings. I believe the intuition that a LA-er seeks to parse and understand

[250] As Roberts (2007: 236) points out, cue-based and expression-based approaches are not mutually exclusive and may co-exist.

PLD *ab initio* is an important insight, which when appropriately framed allows us to address the question of 'process'.

The computational complexity of J.D.Fodor (1998) and Sakas & J.D.Fodor's (2001) suggestion is intimately related with their (and Yang's) conception of parameters. They rely on the "switchboard metaphor" under which a grammatical system must be fully-specified in order to be 'large' enough to allow parsing of an input string; this results in comparison of operations across a reference-set of grammars. As discussed in §3.2, a more accurate conception of a fully-specified grammatical system is that it is *smaller* than initial grammatical competence. An underspecified system is one in which no constraints are imposed on the bundling of syntactic features on LIs / head-movement options available / the distribution of movement-diacritics / etc. The process of FLA in fact consists of winnowing down the options available to the parser, rather than providing it with the instructions it needs to function at all; an underspecified grammar is *larger* than a mature grammar. The same is true outside the syntactic domain, with innate grammar supplying a range of possible phonemes / a range of phonological features / an algebraic rule format, etc., without constraint on how they are exploited / which rules can be entertained / etc.

Under this more accurate conception of the immature grammar, the role of parsing as the process behind FLA can be afforded new life, and becomes compatible with a variationist model of learning. The LA-er can attempt to parse PLD (syntactically / phonologically / morphologically etc.) straight away, using exactly the same parser (and reanalytic machinery) as employed by an adult language-user, only having access to more grammatical options. Needless to say, parsing accurately will be significantly more challenging with a larger grammar, with less 'guidance' on a possible parse. This is consistent with the data, which show that grammatical options exist side-by-side for a prolonged period in LA-ers' I-languages. Under this approach then, the *parser-acquirer* (cf. J.D.Fodor 1998, Sakas & J.D.Fodor 2001) will attempt to parse PLD (in various ways), with successful results increasing the probability associated with the options exploited in the parse, and diminishing those associated with options not exploited – in accordance with probabilistic tracking. The reverse situation obtains in response to an unsuccessful parse. As certain bundlings / distributions of movement diacritics / phonological features etc. become associated with higher probabilities, they will be become increasingly favoured as 'first choices' in the parsing process, and, if part of the 'PLD grammar', these choices will give rise to fewer reanalyses, further increasing associated probabilities, as part of a positive feedback loop. Other grammatical options will become increasingly disfavoured, until excluded from consideration

altogether. The probabilities associated with grammatical postulates at each stage will be reflected in the productions of the LA-er.

Yang's variational model relies on the simplest, linear instantiation of 'competition'-based learning (see Yang (2002: 29) for details). He (*ibid*.: 28) stresses "the *passiveness*" of "the 'dumb' learner in language acquisition". He justifies this with (a) the observation that successful FLA is possible "irrespective of 'general intelligence'", and (b) with "an argument from ignorance": the observation that "we simply don't have a theory of children's cognitive/computational capacities to put into a rigorous model of acquisition". Objection (a) is answered by the suggestions made here, which rely on the reflexive ability of parsing, known to dissociate from general intelligence. Similarly, this ability provides the "theory of children's … capacities" his model lacks, responding to objection (b). The variationist parser-acquirer is "dumb" in a wide range of relevant senses however: (i) not altering postulates within a given parse; (ii) not engaging in complex operations of comparison; (iii) not possessing knowledge of surface strings associated with parametric settings or default parametric settings; (iv) not ordering acquisition in accordance with innate specification or the foresight afforded by a highly sophisticated parsing device.

This section suggests a synthesis of domain-general statistical learning with a key insight from a model of transformational learning. It preserves the key empirical and theoretical advantages of variationism, while addressing a major lacuna in the model – simply by introducing an accurate understanding of immature grammatical competence. It will be seen in the following section that this new focus on 'process' allows us to replace problematic explanations of various phenomena in language change and extend Minimalist reasoning to FLA still further.

7.3 'Conservative' acquisition

As discussed in §2.3, much attention has been paid to the 'conservatism' of LA-ers in the linguistic literature, with studies suggesting that they tend to characterize their PLD using the smallest set of postulates possible. There is persuasive evidence that this is a domain-general principle, following an observation in Vaux (2009).

Vaux (2009) notes that there is a parallel between the conservatism evident in the acquisition of phonological rules (see §7.3.3 below) and classic results from studies into the nature of

associationist conditioning. In particular, he points to studies by Reynolds (1961) and Kamin (1969), which show that when two "to-be-conditioned" stimuli (CSs) are associated with an unconditioned stimulus (US) in a training protocol, as "redundant predictors" (Gallistel 2003: 2-3), then "conditioning to one CS tends to block conditioning to the other" (Gallistel 1990: 456) (in Kamin's terms – "overshadowing"). In Reynolds (1961), for instance, two pigeons were trained to peck a red key with a central white triangle for a food reward; pecking a green key with a central white circle did not yield a reward. When Reynolds tested the four stimulus components (redness, greenness, a white triangle, and a white circle) in isolation, it was found that one bird had been conditioned to the white triangle, whereas the other had been conditioned to the red background. The finding of these studies was that when "there is more than one solution to [a] contingency problem, because the animal's experience is inherently ambiguous" (Gallistel 2002: 164), "an information-theoretic consideration … dictates a preferred solution … fastening onto one carrier or the other [such that] subjects respond to the minimum set of cues that conveys all the available information about [the] US" (*ibid*.: 160-4). [251]

As discussed in §2.3, in as far as data analysis involves computational operations, such as postulating symbolic structure to parse an input, it may be constrained by principles of efficient computation, being rendered domain-general by more basic third-factor constraints. A plausible interpretation of the domain-general finding of 'conservatism' is the implication of *MinVO* in the activity of the parsing processes through which a learner seeks to interpret its input. [252] In the context of FLA, my suggestion is that acquirers will parse their PLD in a computationally optimal fashion, postulating only such operations / symbols of grammatical competence as are motivated, and no more, minimizing the operational (and, indirectly, cache) load imposed by parsing. In other words, *markedness* (Jakobson 1941) in acquisition is associated with economic data processing.

This contrasts crucially with what has standardly been assumed about markedness since the work of Chomsky & Halle (1968) (following earlier work in Halle (1961, 1962) and Chomsky (1965)) concerning a distinctive-feature system for phonology. For Chomsky and Halle, the markedness

[251] The classic interpretation of these results was in terms of selective attention (Mackintosh 1975, Pearce & Hall 1980). For extensive and persuasive argument in favour on an information-theoretic interpretation of results in conditioning studies, which we cannot present here, see *i.a.* Gallistel & Gibbon (2000, 2002) and Gallistel (2002, 2003).

[252] If true, this would be a rare concrete example of the activity of plausibly ancestral principles of cognitive optimality in a non-linguistic domain (see §2.3).

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). For Chomsky and Halle, therefore, 'conservative' acquisition was a matter of capturing generalisations which were significant and natural, i.e. of addressing the PoS, not merely a case of simplifying the grammar. However, 'feature-counting' metrics of this kind can only be applied using "astronomical calculation" (Chomsky 2005b: 7), involving reference-set computation of many derivations and comparative operations, such that there is simply "no computationally feasible way of going from data to finding the optimal instantiation of the format" (Chomsky 2012: 82). [253, 254]

Kiparsky (1965) adapts Chomsky and Halle's "simplicity metric" from its original use in capturing linguistically significant generalisations (as principle of UG) to a literal preference on the part of the LA-er for "*simplification* [emphasis PK's own] of the grammar" (*ibid.*: 53), still assessed in terms of counting the number of features involved. Kiparsky (1965) takes this preference to motivate (among other things) the loss of featural content in phonological rules (see §7.3.4). Again, however, this approach requires intractable 'look-ahead', and, in this case, without obligatory motivation: for Chomsky and Halle simplification was part of overcoming the PoS – a learner 'must' – whereas for Kiparsky it is merely a learner preference.

Following the suggestion by Clark & Roberts (1993: 342) that "children avoid grammars that create inelegant representations", Roberts & Roussou (2003) adapt Chomsky & Halle's notion of markedness to the syntactic domain in a way compatible with the P&P framework. They propose that diachronic syntax reflects LA-ers' "general preference for simplicity of representations" (Roberts 2007: 208), evaluated in accordance with their feature-counting metric:

(139) *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), simplified from Roberts & Roussou (2003: 201))

[253] The same is true of Gallistel's (2002) explanation of the preference for simplicity shown in over-shadowing studies. He suggests that parsimony is preferred "because it is the most powerful solution, where power is measured by the average amount of information conveyed per CS" (*ibid.*: 164).

[254] With respect to syntactic acquisition, the "simplicity metric" *as a principle of UG* was rendered unformulable by the P&P framework, under which acquisition is merely selection from among pre-determined parametric options, no longer making reference to a "format" for acquisition.

In a crucial difference from Kiparsky's approach, Roberts & Roussou's metric is associated with the representation that is assigned to a substring of PLD, rather than the grammar itself, and so is more in keeping with the approach pursued here. Nonetheless it relies on reference-set computation, and so, if accurate, cannot be considered a reflection of computational optimality.

7.3.1 An aside on explicit markedness

Innate markedness conventions have sometimes been construed as a necessary design feature, allowing FLA in the case of pervasive ambiguity in PLD. The suggestion is that grammatical knowledge encodes some kind of 'default' settings that a LA-er can revert to when external evidence does not provide sufficient grounds for selecting one option 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" which exist "*in addition* to universal principles of syntax" (Bickerton 1988: 282) [my emphasis]. (Cf. also Clark & Roberts (1993: 302), Chomsky (1993: fn.6), Chomsky (2004c: 104).)

Some initial plausibility is granted this position by the observation that creoles disproportionately assume particular grammatical options, including: lack of V-to-T movement, the absence of referential null subjects, head-initial word order, and preverbal tense/mood/aspect particles (Roberts 2007: 406-19). However, these are only 'default' options in the sense that they are properties of immature grammatical competence, with their 'marked' counterparts requiring sufficiently robust support from PLD. V-to-T movement (on top of obligatory V-to-*v* movement) is expressed by rich tense morphology (Biberauer & Roberts 2010); in its absence, V will remain '*in situ*' in its position in the functional hierarchy (see §3.2.2.1). Referential null subjects are supported either by rich φ-morphology or bare-*n*P LIs (see §3.2.2.1); in their absence, null DP-clusters will not be posited. Head-final grammar is evidenced by *canonical* head-final word order; in the absence of sufficient evidence for this, the movement operations disrupting SPEC-H-COMP underlying word order will not be posited (see §6.2). Verbal conjugation will rely on its reasonably consistent attestation in PLD; in the absence of this, these features of the verb's extended projection will be marked analytically in their position in the functional hierarchy (again, see §3.2.2.1). [255] As discussed above, it is a misrepresentation to suggest that an underspecified grammar is a defective grammar, and the support of explicitly-stated default

[255] An example supporting the classic position that simplicity in one part of grammar will be associated with complexity in another.

parametric settings is unnecessary once one considers: the coherence of underlying competence, the way in which 'marked' settings are derived, the evidence required for these derivations, and the impoverishment of the PLD (cf. Roberts 2007: 418). Similarly, there is no need to appeal to any preference for simplicity on the part of the LA-er (Roberts 2007), or economic data processing (Mobbs 2008) to account for creole facts. 'Markedness' in this sense is simply a property of underlying competence and its relationship with PLD.

Returning now to the main track, we observe that the approach to non-explicit, metrically-determined markedness pursued in the literature is inconsistent with the view of cognitive economy we adopt here. The logic of the third Minimalist proposal holds that such economy is more plausibly considered a property of *computation itself* than of its *output*. There is no principled reason to suppose that LA-ers will conduct complex computation in order to simplify their grammar or simplify the representations they assign to substrings.

7.3.2 Grammaticalisation

We will now present the evidence suggesting that *MinVO* is implicated in the process of parsing which converts PLD into evidence relevant for determining grammatical options, plausibly giving rise to 'conservative' FLA.

Our first observation is the prevalence of *grammaticalisation* in the face of ambiguous PLD (Hopper & Traugott 1993), that is, the reanalysis of a member of a lexical category as a member of a functional category (or of a member of a functional category as a member of a different functional category). [256] Grammaticalisation pathways can typically be characterised by the loss of a node in a phrase marker or the loss of functional layers associated with a SO (cf. Roberts & Roussou (2003) and (139) above).

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 (140a), with the structure of DP as in

[256] Note: we are not commenting here on the prevalence of grammaticalisation *relative* to degrammaticalisation (the falsifiable claim of *unidirectionality*). Phonological and conceptual considerations underlie this (distinct) finding.

(140b). 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 (140c).

(140) (a) *... 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'
(Foulet 1990: 268)

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

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

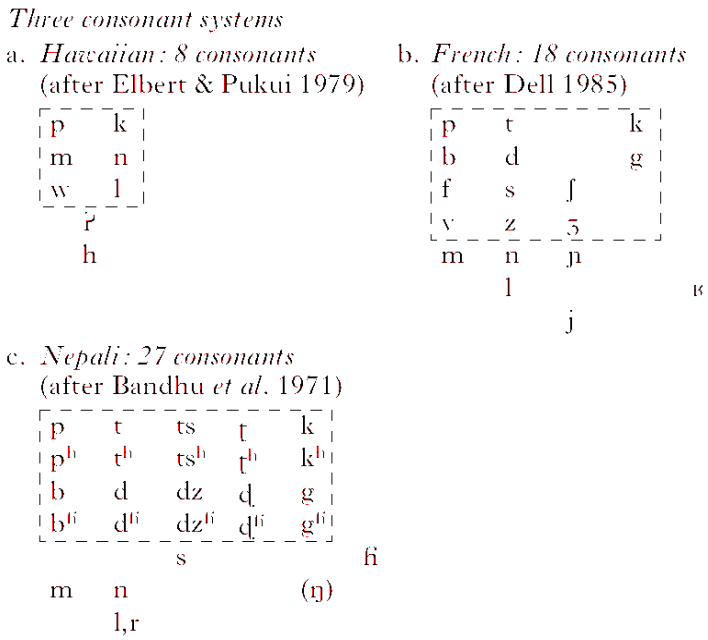
(*ibid*.: 196)

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 minimization of formal content postulated by the parser-acquirer, since the NP of (140c) is mono-phrasal, rather than bi-phrasal. The *Merge* operation between *point* and its NP complement, forming the NP of the extended projection, is no longer necessary, with the NP complement instantiating the base of the extended projection in its own right.

7.3.3 Featural economy

Clements (2003, 2009) demonstrates that sound systems tend to use featural contrasts with "maximal efficiency", expressing "featural economy". This can be demonstrated using the consonant inventories of three different languages:

(141)



(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 "decreasing the number of features while holding the number of sounds constant" (Clements 2009: 28), as reflected in "the frequent historical elimination of "isolated" sounds that do not fall into regular patterns of correlation with other sounds" (*ibid*.: 28-9), such that "the feature that previously characterized them becomes redundant." (*ibid*.: 29) An example of this is provided by a development in the realisation of Zulu plosives:

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

(a) stage 1

p’	t’	k’
p^h	t^h	k^h
b	d	g
ɓ		k

(b) stage 2

p’	t’	k’
p^h	t^h	k^h
p	t	k
b		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 /b, d, g/ become devoiced and /k/ became voiced (at least inter-vocalically). Subsequently (see Louw (1962) for evidence of the intermediate stage), in “an apparent pull-chain effect” (Clements 2003: 317), /ɓ/ shifted to /b/, forming a voiced series with /g/, shown in the last row. The motivation for this “apparent pull-chain effect” may plausibly be ascribed to the tendency of LA-ers to characterize phoneme inventories using the fewest number of features: the feature that previously distinguished the implosive from its plain voiced counterpart, [-obstruent], need no longer be postulated at all (Clements 2003: 317-8). This is consistent with the force of *MinVO* in the acquisitional ‘parsing’ of sound systems: the number of distinctive phonemes is correctly characterized, but without postulating an additional feature.

7.3.4 The acquisition of phonological rules

As discussed above, one of the seminal works on conservative acquisition, Kiparsky (1965), makes reference to the tendency of phonological rules to lose featural content over time. A subset of these changes is represented by the general trend of phonological rules toward extended contexts. Classic examples include: the American /æ/-tensing rule (Labov (1981, 1986), Kiparsky

1995) and English /u:/-shortening (Dickerson 1975, Kiparsky 1995). [257] Following Vaux & Nie (2013: 36), we present here the example of English /r/-insertion, which developed as follows (cf. discussion in Wells (1994)):

(143) Expansion of contexts triggering /r/-insertion:

Stage 1: /ə/, including centring diphthongs

Ø → r / [–high, +back, –low, –round, +ATR] __]$_\sigma$ V

Stage 2: /ə, ɑ:/

Ø → r / [–high, +back, –round] __]$_\sigma$ V

Stage 3: /ə, ɑ:, ɔ:/

Ø → r / [–high, +back] __]σ V

As will be clear, the /r/-insertion rule becomes increasingly general, with the number of features necessary to define its context decreasing. This is consistent with the force of *MinVO* in parsing-acquisition, since the previous contexts of the rule's application will be correctly captured at each stage, but without featural postulations which contribute no further to that aim. [258]

[257] Kiparsky (1995) construes both these examples as evidence for a *grammar*-simplifying LA-er, which as we have discussed, is a problematic claim on 'look-ahead' grounds. It is perhaps for this reason that Kiparsky (1995: fn.4) makes "no commitment to any formal evaluation measure", merely suggesting that "any theory which characterizes [loss of context] as … optimization" will suffice.

[258] It is worth pointing out that this directly opposes the claim of the *Subset Principle* (SP) (Berwick 1985, Wexler & Manzini 1987, Hale & Reiss 2008), which holds that, in response to a given set of PLD, the LA-er will hypothesize the grammar with the smallest extension including the observed data. The conceptual motivation for this claim is that LA-ers do not have access to (direct) negative evidence. Knowledge of the ill-formedness of certain parts of PLD is unavailable/unused, and so, the logic goes, a LA-er may fall into a "superset trap", positing a grammar which generates a language which is a superset of the target languages, so that no positive evidence can disconfirm it – unless they observe the SP. The most obvious objection to this proposal is the "issue of how the learning device "knows" which are the subset and superset" settings (Biberauer & Roberts 2009: fn.2) – the familiar problem of 'look-ahead'. For discussion of further difficulties associated with the reasoning and predictions of the SP, see Vaux (2009) and Vaux & Nie (2013).

7.3.5 The development of fusional morphology

Schleicher (1861-2: 4, 342-3) proposes a general pathway of morphological change followed by languages – from isolating, to agglutinating, to inflectional. [259]

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)). To the extent that a language's morphology demonstrates this trend diachronically, directionality at the level of the individual grammatical changes is indicated – in this case, the prevalent fusion of two morphemes into a single morpheme rendering the morphosyntactic features of both. This would accord with the force of *MinVO* in the morphemic parsing of a word, since every different aggregation of features that must be postulated will be associated with an additional parsing operation. Syncretism minimizes this effect, and so would be the favoured result.

To take a concrete example of an individual change implicating *MinVO* 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.

Table 3: The paradigm of Proto-Slavic **nesti* 'to carry' in the present tense:

	SING.	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*

[259] Although Schleicher's pathway itself may be significantly unidirectional, its effect on language typology is balanced by the development of analytic morphology directly from inflectional morphology through the phonetic attrition of inflectional distinctions and their subsequent loss. Expressive demands for the new realization of functional morphology may then be met by processes of grammaticalisation, feeding Schleicher's pathway. This is a very approximate description of the development of analytic Modern English morphology from inflectional Old English morphology.

A phonological change in Proto-Slavic nasalized the vowel [o] when preceding a nasal consonant, as in *nes-ō-mь* and *nes-o-nti*. This created ambiguity with regard to the presence of 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 phonological perspective it is no longer clear in these two cases whether 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, or rather by having one morpheme instantiate the thematic and inflectional features, as shown in Table 4.

Table 4: The paradigm of **nesti* 'to carry' in the present tense (later version):

	SING.	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-õtъ*

(reproduced from Długosz-Kurczabowa & Dubisz (*ibid.*))

While this reanalysis is contrary to the evidential force of the rest of the paradigm, it is in accordance with the force of *MinVO* in acquisitional computation, reflecting complete characterization of the morphosyntactic input without an operation associated with postulating a second morpheme.

In this section we have seen that there are highly plausible grounds to suppose that various central results of the diachronic literature are explicable in terms of the role of *MinVO* in the process of parsing which underlies FLA. The implication of *MinVO* in analogous acquisitional 'parsing' tasks, found in other domains, conceivably underlies the status of 'conservatism' as a third-factor principle of data analysis.

7.4 'Liberal' acquisition

As noted in §2.3, there is also some evidence that a third domain-general principle of data analysis is involved in FLA. Studies of language learning tasks suggest that both adults and children may regularize inconsistency present in PLD (Hudson & Newport 2005, Reali &

Griffiths 2009, Smith & Wonnacott 2010). Studies of FLA itself reveal the same tendency, such as the well-known over-regularization errors in child morphology (*i.a.* G. Marcus *et al.* 1992). Similar findings are made in non-linguistic domains. In experiments in which participants track the relative frequencies with which different lights flash, *non-veridical* learning is found, with subjects projecting the system as more regular than it really is, overestimating the probability associated with the most frequently flashing light (Gardner 1957).

There is persuasive evidence to believe that this domain-general tendency toward over-regularization (*Generalisation of the Input* (GofI) in Roberts' (2007) terms) is motivated by the limits of WM capacity – a "principle of structural architecture" informed in some way by "properties of the human brain". A general finding is that children are more likely to regularize probabilistic input, both in the linguistic (Hudson Kam & Newport (2005, 2009)) and non-linguistic domains (*i.a.* Weir 1964, Bever 1982), interpreted by Hudson Kam & Newport (2005, 2009), Hudson Kam & Chang (2009) and Ferdinand *et al.* (2013) as suggestive of the involvement of computational capacities. This conclusion is lent greater credibility by studies in which over-regularization is seen to increase with the number of variables tracked (Gardner 1957, Ferdinand *et al.* 2013). Similarly, Kareev, Lieberman & Lev (1997) found that differences between individuals in WM capacity (determined by performance in a digit-span test) predicted their over-regularization behaviour in probability-matching experiments – those with weaker WMs being more inclined to over-produce the more common variant. [260]

7.4.1 Morphological analogy

IG plausibly underlies the prevalence of analogical change in the diachrony of morphology. Analogical levelling is the process by which some members of a morphological paradigm change to match other forms within the paradigm. An example is the development of the strong class II conjugation of verbs between Early New High German (ENHG) and New High German (NHG). The <eu>~<ie> alternation was lost in favour of <ie> consistency:

[260] For alternative approaches to the reduction in computational burden achieved by the postulation of productive rules over variable data, see Yang (2006) and Mobbs (2008). For an account in which learners are motivated by the speed with which they can characterize a data set, see Roberts (2007).

Table 5: Loss of /iu/~/ie/ alternation in early NHG (adapted from Albright (2008: 145))

'to fly'	ENGH	NGH
1sg	fl**eu**ge	fliege
2sg	fl**eug**st	fliegst
3sg	fl**eu**gt	fliegt
1pl	fliegen	fliegen
2pl	fliegt	fliegt
3pl	fliegen	fliegen

This change can be interpreted as difficulty / 'reluctance' on the part of the LA-er in tracking the different stem forms within the paradigm, implicating IG.

In other ENHG verbs, stem alternations were not lost, but rearranged. Strong class V verbs such as *geben* 'give' changed their 1st pers. sing form to match their plural form, generalizing the *umlaut* pattern of alternation seen in verbs such as *graben* 'dig'.

Table 6: Rearrangement of /i/~/ë/ alternations in ENHG (adapted from Albright (*ibid.*))

'to give'	ENHG	NHG	Following pattern of 'dig'
1sg	g**i**be	gebe	grabe
2sg	g**ib**st	gibst	gräbst
3sg	g**i**bt	gibt	gräbt
1pl	geben	geben	graben
2pl	gebt	gebt	grabt
3pl	geben	geben	graben

Again this change is consistent with the role of IG in FLA, interpretable as difficulty on the part of the LA-er in tracking which paradigm is associated with which verb.

Kiparsky (1965 *et seq.*) interprets analogical changes of this kind as demonstrating the LA-er's preference for simplicity of the grammar. Again, this implausibly requires that learners "first correctly acquire the target grammar (so they can evaluate its complexity), and then … replace the acquired grammar with a simpler one" (Reiss 2003: 150). Furthermore, it is demonstrably false that all analogical change simplifies. Taking the classic example, the rule of plural

formation by <-s> suffixing was generalised across the different declensions of Old English. In this way, the noun *brother* acquired a new, regular plural form, *brothers*. However, it (at least initially) retained its archaic plural *brethren* also, and so analogy in fact increased the complexity of the grammar, creating a new nominal class alongside that class of nouns with nasal plurals only, e.g. *child~children* and *ox~oxen*. [261] On the other hand, we have seen that analogical change is entirely consistent with the implication of a domain-general principle of data analysis in FLA, generalising across inconsistent PLD.

7.4.2 Cross-categorial harmony

As discussed in §1.3.3.2, a classic typological observation is that harmonic word order systems, those demonstrating consistence of head-complement order across categories, are disproportionately represented relative to disharmonic systems (Greenberg 1963). A case in point is the correlation between VO/OV and PrepN/NPost ordering, in which 94.3% of the attested pairings are harmonic.

(9) Correlation between VO/OV and PrepN/NPost in the 981 languages showing dominant order for both, as surveyed in Haspelmath *et al.* (2013):

OV & NPost	472	(48.1%)
OV & PrepN	14	(1.4%)
VO & NPost	41	(4.2%)
VO & PrepN	454	(46.2%)

A convincing explanation for this phenomenon is afforded by the performative preference for locality in the assessment of syntactic and semantic relationships, minimizing the burden placed on WM in the parsing process. [262] For reasons discussed in §1.3.3.2, Gibson's (1998) "capacity-

[261] Kiparsky (2014: 73) acknowledges that his account of analogy faces "a real obstacle" in overcoming the existence of "messy intermediate stages". See Kiparsky (2010) for discussion of his Optimality Theoretic model of how to maintain a grammar-simplification-oriented approach in the face of such "bumpy rides".

[262] It is important to note that a subset of disharmonically-ordered pairs is precluded by the *Final-over-Final Constraint* (FoFC); the skew toward typological harmony is to a certain extent determined by synchronic factors. (See §6.3.) The FoFC also predicts that harmony will be stronger *within* categorial domains than across them. This seems to be accurate, but the data demonstrating this are too extensive to present here and require 'problematizing' in many cases. Compare, however, the finding that c.33% of V-Comp/Comp-V and D-NP/NP-

constrained" account of this pressure (recapitulating classic insights of Chomsky (1961) and Miller & Chomsky (1963)) is preferred to Hawkins' (1994 *et seq*.) account.

This preference motivates utterances in which dependent heads are maximally adjacent. For instance, the Heavy-NP shift version, (6b), of an otherwise equivalent construction is overwhelmingly preferred in performance, minimizing the maximum distance over which *waited* and a modifier head must relate to each other in parsing.

(6) (a) [I $_{VP}$[waited $_{pp1}$[for the man with the ginger beard who had bought me a coffee earlier and really made my day] $_{pp2}$[in the late afternoon sun]]]
(b) [I $_{VP}$[waited $_{pp1}$[in the late afternoon sun] $_{pp2}$[for the man with the ginger beard who had bought me a coffee earlier and really made my day]]]

As clarified by Newmeyer (2001, 2003, 2004), preferences in performance may be interpreted as reflecting grammatical constraints by LA-ers and so give rise to fixed word order effects encoded in grammar (cf. discussion in §§4.3.3 and 4.3.4). A preference for local dependencies in performance, working through this mechanism of change, predicts harmonic Head-Comp ordering. Canonical disharmonic ordering will create greater distances between dependent heads and lead to extraposition in performance in order to minimize these distances; this may subsequently be reanalysed as underlying harmony. For instance, many accounts of the shift from OV to VO word order in the history of English point to the role of extensive rightward extraposition of head-initial complements of verbs in feeding reanalysis (*i.a.* Stockwell 1977, Lightfoot 1979, van Kemenade 1987).

(144) forðam ðe he hine ætbræd fram flæsclicum lustum
because he himself withdrew from fleshly lusts
'… because he withdrew himself from fleshly lusts.'
(van Kemenade 1987: 33)

In (144), a prepositional phrase has moved rightward from its pre-verbal, base-generated position, minimizing the distance between the verbal head and the adpositional head, minimizing the

---

D pairs are disharmonic (Dryer 1992: 104), with the finding that only *c*.11% of C-VP/VP-C and V-Comp/Comp-V pairs are disharmonic (Dryer 2009: 199). (The particularly strong cross-domain correlation between VO/OV and PrepositionN/NPostposition word order can plausibly be explained by the commonly-followed grammaticalisation pathway from verbs to adpositions (Dryer 2005: 387)).

caching of unintroduced items in head-driven parsing. It is plausible that sufficient PLD of this kind led to reanalysis of Old English as canonically V-initial. Similarly, in language systems in which a head-initial phrase dominates a head-final phrase, performance pressure for locality might result in leftward extraposition of the dominated phrase, with the resultant ambiguity in PLD giving rise to reanalysis of the dominating phrase as head-final. [263]

Albeit adopting a different position regarding the status of innate competence and functionalist motivation, the above is essentially Hawkins' (2004 *et seq*.) current account of harmonic skewing in the typological record. In earlier work, Hawkins (1979, 1983) proposed a different account. He suggested that the disproportionate representation of harmony was the reflex of a principle of acquisition, *Cross-Categorial Harmony* (CCH), motivated by LA-ers' preference for simpler, more *usable* grammars:

(145) *Cross-Categorial Harmony* (CCH):
There is a quantifiable preference for the ratio of preposed to postposed operators within one phrasal category […] to generalize to others.
(Hawkins 1983: 134)

The difficulty with this account is familiar from §1.3.3.3: it remains unclear how an understanding of complexity and efficiency in performance is expected to shape acquisitional postulates. 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, Complementizer-initial, VP-Aux, 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.

Kiparsky (1996) adopts a similar position within a formalist framework, suggesting that "harmonization of the direction of complementation" reflects a preference for "simplification of

263 This account would be expected to reinforce the strength of intra-domain harmony relative to cross-domain harmony (cf. fn. 264). Pairs of heads which are part of the same extended projection are more frequently in a relationship of dependency than those which are not: for instance, it is only on rare occasions that a VP-phrase is not embedded in a CP, yet it is comparatively frequent for a V not to select a DP object.

the grammatical system" as part of FLA. This account runs into the recognized problem of the tractability of forward-looking computation, and can be discounted for this reason. [264] Most relevantly for our purposes, however, Kiparsky's proposal, (a), eschews any difficulty with the persistence of disharmonic systems in the face of functional deficiency, and, (b), allows a natural account of their persistence in spite of whichever force does disfavour them in FLA.

As Kiparsky (1996: 151) points out, discussing the shift to VO in Old English, "a preference which is general … only manifests itself when conditions allow – for example, when the evidence for the old structure … becomes attenuated." In other words, the mechanistic demands of reanalysis must be met: in the absence of ambiguity in the input, reanalysis may not occur (cf. the *Inertia Principle* of Keenan (2002) and Longobardi (2001)). In terms of the account favoured here, non-canonical word orders produced by performative pressure must be numerous enough to mask the underlying word order; that is, there must therefore be sufficient parsing difficulty to motivate a significant amount of extraposition.

A second force serves to stabilize disharmonic (and harmonic) language systems against word order change. The FoFC discussed in §6.3 constrains the structural positions at which word order change is possible at any given time. Under the 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 (146) for the verbal domain:

(146) (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, Newton & Sheehan 2009: 6)

In change from head-final to head-initial, (146a), if change were to start at VP, the resulting word order would have a head-final CP/TP dominating a head-initial VP, violating FoFC. If TP-order 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-order to change first, such that a head-initial CP dominates a head-final TP. Thereupon, TP-order may

[264] Various issues arise in connection to the claim that harmonization reflects simplification. They include: (i) the fact that under common assumptions about linearisation (see §6.2), head-final orders are inherently "more complex", (ii) the issues discussed in fn.261 only translated to Head-Comp ordering, and (iii) the difficulty of treating Head-Comp ordering as a coherent grammatical property for the purposes of analogical assessment (see discussion in relation to Roberts' (2007) suggestion immediately below).

change, and only then is VP-order 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 the FoFC; and so on. (See Biberauer, Newton & Sheehan (2009) and Biberauer, Sheehan & Newton (2010) for discussion and data.)

At any given stage therefore, reanalysis of 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 in spite of the ambiguity created in PLD by preferences in performance.

Roberts (2007: 273-5) reinterprets CCH as an immediate reflex of IG acting in acquisition (this is the source of my term for the principle). In his case, IG is motivated by the LA-er's 'instinct' to characterize a set of data as quickly as possible, and so "to exploit pieces … of input to the full" (*ibid*.: 275). I will not present the details of his account here; I merely wish to observe that there is potentially a difficulty with assuming 'comparability' of variables when tracking surface word orders, a premise on which an IG account rests. For example, if we assume SPEC-H-COMP linearisation, VO word order can result from a number of different derivations, including, for example: (i) V-in-*v* and object-in-situ (e.g. English); (ii) V-in-T and object-in-situ (e.g. French); (iii) sometimes V-in-*v* and object-in-situ, but other times involving V2 and (if the object is pronominal) object extraction (e.g. Danish); (iv) *v*P-fronting (sometimes, and sometimes not, involving prior object extraction) (e.g. Chol (Coon 2010), Niuean (Massam 2000, 2001)). While a tempting explanation of harmonic skewing, it is not clear that IG has any explanatory purchase here given the lack of a stable parameter of variation – not a problem encountered by approaches motivated by surface word order processing effects.

One final point that needs addressing is the approximately even number of head-final-type and head-initial-type languages. Assuming a SPEC-H-COMP underlying order, a prediction of 'conservative' acquisition of the kind discussed in §7.3 is that head-initial orders will be preferred by LA-ers in the face of ambiguity. Head-final orders involve the parser's postulating movement operations of one kind or another, and so extra nodes in the phrase marker (cf. (139) above and discussion). This suggests that an opposing force is also indicated, promoting the introduction of movement into a language system. The natural place to look for this is the motivation for dislocation itself – the duality of semantics (see §4.1.4). Paraphrasing suggestions by Kiparsky (1996) and Simpson (2004), optional movement might be postulated in syntax to generate a discourse or scopal effect before becoming obligatory through the subsequent conventionalisation

and loss of this effect. As Roberts (2007: 276) points out, this is in keeping with "the traditional intuition that the drive for expressivity" and "the drive for simplicity" "create an overall equilibrium". The section has presented a wide-ranging, if cursory account of the issue of cross-categorial harmony in the typological record, demonstrating (again – see §§4.3.3 & 4.3.4) the appropriate place for functionalist explanation in formal work. This is an alternative to a potential account of the phenomenon in terms of a domain-general principle of data analysis.

In this chapter we have discussed the various different ways Minimalist concerns might be reflected in the process of FLA. In particular, we have seen suggestive evidence that FLA invokes domain-general principles of data analysis, plausibly informed by more fundamental third-factor principles and the limits of computational capacities. In doing so, we introduced a new appreciation of how PLD is interpreted as relevant in determining grammatical options. We concluded with an extended discussion of how to account for a typological result superficially amenable to explanation in terms of a third-factor principle of data analysis; in doing so, we reiterated the proper place for functionalist explanation in formalist work.

8. The Linguistic Condition

This final substantive chapter is very brief, picking up §3.1.1's discussion of the evolution of the FL and its broader role in human cognition.

8.1 The evolution of lexical semantics

The fourth Minimalist proposal suggests that syntactic abilities were selected on grounds of vastly expanded capacity for abstract thought (the 'human capacity'). A range of support was adduced for this position. First, it is clear that a mechanism of unbounded, embedding recursion affords internal representations of greater semantic complexity, so allowing reason to operate in new arenas. Secondly, studies of sign language show that core properties of language are independent of modality, unexpected if language evolved under pressure for externalization. Thirdly, the artifactual record manifests the simultaneous (and sudden) emergence of symbolic activity and a range of advantageous 'thought-behaviours', including systematic technological innovation. Finally, various properties of the concrete output of syntax seem inconsistent with adaptive pressure for external usage, including: ineffability, wide variation in communicative abilities, easily comprehensible ungrammatical constructions, unparseable grammatical constructions, etc.

It was pointed out, however, that abstract thought supervenes not just on an infinity of hierarchically-organized, embedding structure, but also on (non-reflexively produced) LIs with complex, non-referential semantics: thought is not limited to referential items or abstract items of one 'conceptual type'. In §1.1, the 'conceptually eclectic', mind-internal semantic structure of LIs was illustrated through discussion of the (single) LI *book*. Animal 'LIs' are afforded both meaning and form by a particular perceptual-conceptual system of the mind, and so demonstrate a one-to-one relationship with (an aspect of) the environment. On the other hand, human LIs are disconnected from these perceptual-conceptual systems and are able to draw on meaningful content from several of them at once, instantiating the "promiscuity of the interfaces". These

accretions of instructions to the CI-systems are memorized as part of a lexicon, each associated with (indirect) instructions to the AP-system granting them form. [265]

As has been pointed out, "[t]hese properties of lexical items seem to be unique to human language and thought, and have to be accounted for somehow in the study of their evolution" (Berwick & Chomsky 2011: 40). What's more, in the absence of any evidence for their earlier instantiation or immediate stages of their development, [266] the evolutionary chronology of language discussed argues for the catastrophic emergence of human LIs every bit as much as it does for the catastrophic emergence of unbounded *Merge*. It has been observed that "no one has any idea" (*ibid.*) how LIs evolved and there are "no sensible origins" (Chomsky 2012: 40). I tentatively suggest here that this is not the case.

The capacity for unbounded, embedding recursion is necessary to ground the detailed semantic representations which allow for advantageous 'thought-behaviours' such as complex planning. As discussed in §2.3, it is not known for certain whether the FL's capacity for building structure in this way evolved *de novo*, or was co-opted from another faculty. Either way, the capacity for unbounded, embedding recursion is a compound one, consisting of both a *Merge* operation and an *undeletable* EF (the symbol which identifies the mental representations over which *Merge* can operate). [267]

There is plausibly some adaptive utility to a *Merge*/undeletable EF conglomerate operating within a single conceptual domain – this is the intuition behind suggestions that recursion may have been exapted by the FL from another faculty, as discussed in §2.3). However, productive, abstract thought-behaviours of the kind discussed above and in §3.1.1 require that it be possible to compose concepts from across a range of conceptual systems. There is, therefore, significant pressure in the context of infinite, embedding recursion for conceptual promiscuity to develop. The manner in which infinite, embedding recursion is instantiated offers a straightforward

[265] On account of their complexity human LIs usually relate to the world in a number of different ways, taking on many meanings; consequently, a particular meaning can usually be given by more than one LI.

[266] The latter possibly being inherently nonsensical in light of the manner of their instantiation (pending the following discussion): any intermediate stage in evolution (for which there is no evidence) would involve the imposition of a constraint on conceptual "promiscuity" – the end of an evolutionary process being reached before the middle. Cf. fn.92.

[267] The EF allows new LIs to be introduced, capturing the *unboundedness* of recursion; its undeletability captures the particular kind of recursion found in human language – embedding recursion – permitting "free *Merge* to the edge, indefinitely" (Chomsky 2008: 144). Cf. §2.3.

account of this development. An undeletable EF marks an object for unbounded use in *Merge* but is plausibly 'blind' to the conceptual contents of the object itself, both in terms of its source and its 'volume'. [268] If infinite, embedding recursion evolved in the FL initially, then the promiscuity of the interfaces 'comes for free'. If recursion emerged within a different faculty and was subsequently co-opted by the FL, then either: (a) a constraint on the 'location' of *Merge* was lost, with the EF already being entirely 'generalized', or (b) a constraint on the conceptual source (and perhaps volume) of the contents of objects marked with EF was lost, with *Merge* itself already being unconstrained in terms of 'location'. Either way, the genetic innovation would have been minor, with non-referentiality coming very cheaply, if not for free, from non-linguistic infinite, embedding recursion – 'easily' motivated by the selective advantage of productive, abstract thought.

Consistent with the intimate relationship between the machinery of unbounded recursion and non-referential semantics is the observation from §3.1.1 that it is far from clear that the ability to form an isolated mental representation with 'cross-conceptual' semantics is of any utility at all. This is revealed by the fact that only by using examples combining LIs with similar semantic objects are we able to reveal the subtleties of their meaning. Similarly, it is not clear that conceptual systems can relate to each other in anything other than this indirect way, through external composition and subsequent 're-presentation' at the end of syntax. The promiscuity of the interfaces is often casually alluded to as a trivially realizable property of human minds/brains underlying the uniqueness of human thought; the assumption seems to be that the different conceptual systems can somehow 'converse with each other' in human brains in way they cannot in other animals' brains. For instance, Hauser (2013: 4) observes that "the difference between human and animal brains" crucially involves "interconnected modules". He suggests that while several "authors emphasize that language was essential in forging connections between modules … language itself is based on interconnected modules, including those dedicated to phonology, semantics, and syntax. It is thus more likely that the connections were in place before language, providing benefits in thinking that went far beyond the parochial style of other animals." [269] Hauser seems to assume that the promiscuity of the interfaces is realized as true *cross-modularity*. This is a questionable notion at best, since it relies on the mysterious ability of one core conceptual domain to interpret the output of another; if true, this would constitute a very rich

[268] Linguistic LIs are (at least typically) abstract *and* complex. See §1.2.

[269] This is a misrepresentation of grammatical modules, which are merely 'standard', consecutive tasks, rather than interconnected ones, but that is not our primary concern here.

evolutionary conjecture indeed. [270] This assertion seems to rely on an over-literal interpretation of "interconnectedness".

The account given of the evolution of non-referential semantics here shares with Ott (2009) and Boeckx's (2011: 50-4) account the observation that the EF is central to "demodularizing" concepts. It differs from it in crucial ways, however. Ott and Boeckx both claim that the development of linguistic LIs, which they take as synonymous with the emergence of an EF, was the "key evolutionary novelty", rendering possible both the operation of linguistic *Merge* (and therefore syntax) and conceptually-eclectic meaning. Four important corrections need to be made to this understanding however. First, *Merge* and an EF rely on each other for any significance (the latter being the symbol demanded by the former, an operation), and so they cannot meaningfully be dissociated in the evolutionary narrative, regardless of recursion's origin. Secondly, the existence of an EF and the existence of linguistic LIs are not equivalent: linguistic LIs require a *generalized* [271] EF, and so it is the freedom (both in terms of location and amount) of "the process that associates concepts with [LIs]" [272] that affords non-referential, complex semantic items. Thirdly, when linguistic lexicalisation (LL) is properly construed, the *Merge*/EF amalgam has potential independent utility, as speculation over its origins reveals, and so LL may plausibly have been an independent, *secondary* evolutionary event. Fourthly, even assuming a FL origin for *Merge*, the non-referential semantics made possible under LL is of no obvious utility without *Merge*'s existence (see §3.1.1), and so conceiving LL as somehow more 'significant' than linguistic *Merge* makes little sense. In short, while I agree that the EF is plausibly a central feature of the evolutionary narrative of non-referential semantics, I favour an account in which the generalized EF which grounds it is either provided wholesale by the machinery of unbounded, embedding recursion, or is provided very 'cheaply' under selective pressure. If there is anything to this proposal, it would not just be interesting because of the dearth of plausible accounts of the evolution of lexical meaning, but because it suggests unbounded, embedding recursion and abstract semantics might be part of one of the 'channels' in evolutionary space of interest to evo-devo (and those pursuing the third Minimalist proposal). Recursion and abstraction may be in same relationship of "correlation and balance" in humans as blue eyes and deafness are in white cats (Darwin, 1856 letter to W.D. Fox, cited in Berwick & Chomsky (2011: 24)).

[270] The closest thing to empirical support comes from the highly peripheral phenomenon of synaesthesia (Baron-Cohen & Harrison (eds.) 1997).

[271] And undeletable.

[272] Ott's (2009: 261) incomplete formulation of linguistic *lexicalisation*.

8.2 Metaphorical, social, and personal cognition

We have already discussed the origins of non-reflexive, non-referential symbolism and linguistic computation, and their reflection in aspects of behavioural modernity, including complex planning and problem-solving, art and music, interpretation of past patterns in the world, etc. We have seen that human mental creativity includes the capacity to imagine novel situations, ideas, or other representations of potentially infinite complexity by recursively combining a range of LIs 'tailored' to fit human concerns. This allows us to reason over new topics of enquiry and contingencies. It allows us, for instance, following Hauser's (2008: 137) example, to reason that if a surface is wet, a pencil will not satisfy the requirements on a writing implement.

We have not, however, properly discussed various other properties of behavioural modernity / human-uniqueness reliant on a slightly different, but related ability of human cognition. Our creativity goes beyond complex abstraction, but allows us to fundamentally 'reconceptualize' situations (real or imagined). This capacity for *metaphorical* thought allows us to extend, neglect, or substitute properties of mental representations, and so change the way we look at (and so are able to reason over) the world. This may include mentally 'adjusting' a thought so it can be construed of in terms of familiar concepts which we have thought about before, or over which we have stronger intuitions. For instance, in thinking about syntactic computations, we tend to substitute in the notion of 'branching' (approximately, [space] + [division] → [unity]) *in place of* 'set-formation' (approximately, [multiplicity] + [inclusion] → [unity]), because the two are identical for the *relevant* property. This makes syntax significantly easier to reason over, because we can conceive of matters using our rich spatial / navigational cognition, rather than our limited mathematical intuitions.

Metaphorical thought is premised on the conceptual compositionality and flexibility of LIs. Emphasis is placed on "in place of" in the above paragraph to stress that such thought involves the actual transfer (and subsequent restoration) of conceptual properties rather than some vague, problematic notion of inter-connectedness or cross-modality discussed above. Metaphor is at its heart literality.

The malleability of conceptualisation under lexicalisation may even allow us to think about notions for which there is no appropriate conceptual domain of the mind. Through picking up on familiar properties of otherwise mysterious notions, we may 'suspend' their unknown/unknowable properties and conceive of them entirely in meaningful terms. This is the

basis of much scientific progress, including any made within biolinguistics. Itself quoting, §1.1 suggested that "the language faculty is a mental "organ" [which] may be abstracted for special investigation because [its] apparent internal integrity and special properties". In other words, the FL shares the property of a physical organ which allows it to be construed as a circumscribed object of enquiry, and we rightly treat it as such. However, we currently understand practically nothing about its physical coherence or unified operation: its "integrity" is merely "apparent". As long as we remember to substitute the concept "subset of cognitive activity" back in for "physical object" within the word *organ*, then we have extended the domain of things over which we can reason beyond the 'thinkable'. Metaphorical thought is a significant feat of imagination beyond complex abstraction, and of clear utility in scientific enquiry and beyond. The nature of LIs is a double-fold source of creativity. [273]

I would also like to suggest that lexically-grounded, metaphoric thought underlies humans' unique personal and social cognition in interesting ways. A range of studies has demonstrated that humans conceive of others and themselves in a different way to other animals, including other great apes. Taking the issue of other-regarding cognition first, [274] a familiar result of the literature is that only humans are able to understand and reason over *false belief*. This suggests that while other animals may be able to attribute perceptual-states in domain-specific ways (see §3.1.3), only humans can attribute belief-states, doing so in a domain-general way (Kaminski, Call & Tomasello 2008). [275] Similarly, studies have revealed pronounced differences between humans

[273] It will be noted that for all this discussion of the creative opportunities afforded by language, we offer no suggestion as to how the human brain is able to exploit them. This is the internal reflex of the problem of "the 'creative' aspect of language use" (Chomsky 1972: 11) noted in §4.1.8. There is something which allows language users to produce and understand "an indefinite number of expressions which are new to [their] experience" (*ibid*.: 100), but "appropriate to situations [and] not controlled by stimulus considerations" (Chomsky 1980: 222). This allows language to serve "as an instrument … of thought … available for use in whatever contingencies our thought processes can comprehend" (*ibid.*). I have no suggestions to make in this regard, merely noting Chomsky's (1983) observation that "questions about the causation of behaviour, the exercise of will, choice, and so on" may well be more properly characterized as "mysteries", rather than as "problems", possibly "lying beyond our cognitive grasp" entirely.

[274] We deliberately avoid the term 'theory of mind', since as Tomasello, Call & Hare (2003: 239) observe, it is actually a "generic label … covering a wide range of processes in social cognition". For reasons of space we will only discuss *differences* in other-regarding cognition here, but see §3.1.3 for discussion of ancestral aspects of the understanding of others' perceptual/knowledge-states and intentions.

[275] See fn.123 for discussion of whether animals' attribution of perceptual states enables them to understand others' knowledge within limited domains, or whether such understanding supervenes on the attribution of belief-states.

and animals in understanding the *emotions* of others. For instance, Cheney & Seyfarth (2007) studied the stress levels (measured by glucocorticoid concentration in the blood) in baboon females. They found that glucocorticoid levels are elevated after the death of a close relative, after the arrival of a potentially infanticidal new male, and when the normally rigid social hierarchy of females is unstable. On the other hand, stress levels did not appear to be elevated in response to male rank instability or the death of a cohort female's infant. This is strongly suggestive of an inability to project over conspecifics' emotional states, as found with belief-states.

Moving on to self-regarding cognition, it will be clear that experimental studies are harder to conceive. The failure of great apes to consistently pass classic *mirror self-recognition* tests (Gallup 1970) implies that animals lack the same sense of self as humans, but various problems have been pointed out with the protocol (see Parker, Boccia & Mitchell (eds.) (1994) and the papers within for a discussion of these issues). Other studies have monitored the uncertainty of subjects regarding their own mental states – the argument being that if they are able to reason over this uncertainty, this demonstrates self-regarding 'metacognition' (*pace* Carruthers (2008)). The results of these studies show, at best, inconsistent awareness of this kind on the part of non-human animals (see Shettleworth & Sutton (2006) for a review), suggesting that their sense of self is systematically different from our own.

In as far as a robust theory of others and (secondary) self-awareness are distinctive properties of humans, I suggest that they are grounded in: (a) the existence of some (primary) concept of [self] (simply whatever unconsciously motivates self-interest, fears, desires, etc.) and some primitive concept of [conspecific] (trivially manifest in a range of behaviours, including breeding); [276] (b) the lexically-based ability to interchange these freely. The ability to ascribe belief-states and emotion relies on the ability to project the concept [self] over the concept [other], so that we are 'seeing something from someone's perspective'. A range of thought-behaviours clearly supervenes on empathy, including: sympathy (empathic concern), gratitude, resentment, forgiveness, manipulation, etc. Self-awareness requires the ability to project [other] over [self], knowing ourselves "even as also [we] are known". The ability to reflect on ourselves underlies thought-behaviours such as: vanity, pretence, insight into our motivations, efforts at self-

[276] The finding that action perception in animals involves mirror neuron systems (Di Pellegrino *et al*. 1992) also argues strongly for these concepts. (For dismissal of the putatively *causal* role of mirror neurons in *understanding,* see Hickok (2009) and Catmur (2014); for counter-arguments, see references within these two papers and Michael *et al.* (2014).)

improvement, a sense of mortality, concern for meaning, and an awareness of self-determination. Some of these ways of thinking have been discussed under the rubric of "the human condition". If the proposals here are along the right lines, then the FL is even more fundamental to what it means to be human than already assumed; it could be said that we live in a 'linguistic condition'. Furthermore, the continuity of this section's suggestions with the previous section's serves to further elaborate the putative 'channel' in evolutionary space occupied by the human brain, plausibly extending the reach of evo-devo logic in the explanation of human evolution. [277]

[277] The development of a robust theory of others and self-awareness is of central importance in explaining the development of the increased social and personal sophistication seen in the archaeology of early anatomically-modern humans. Evidence of co-operative living emerges in the artifactual record around the same time as evidence of symbolic activity, consistent with the above account. For instance, around 70,000 years ago, we begin to find non-local sourcing of raw materials involved in tool manufacture and the appearance of shell beads, indicative of inter-group trade and gift-giving practices (Ambrose 2002). After the emergence from Africa, the archaelogical record is more complete and demonstrates a great dichotomy between the social complexity of European Cro-Magnon man (from 35,000 years ago) and contemporary Neanderthals, evidenced by:1 more intensely used living spaces (Bar-Yosef 2000), larger group size with social stratification (Mellars 1996, Vanhaeren & d'Errico 2005), long-distance trade networks (Adler *et al.* 2006), etc. See Lewis-Williams (2002) for discussion of artifactual evidence that Cro-Magnon men reflected on themselves within "a spiritual realm" (*ibid*.: 92).

9. Reprise

Chapter 1 of this dissertation demonstrated that we have every right to investigate language from a biological perspective, and that this entails many meaningful questions. To do so, it was necessary to outline reasons to disregard a range of empiricist accounts of language acquisition, seeking to trivialise biolinguistic questions.

The following two chapters sought to clarify the sometimes abstruse content, logic and interdependencies of various proposals appropriate to a biolinguistic commitment – the Minimalist proposals, which together comprise the argument of linguistic Minimalism. Chapter 2 characterized two core notions of the Minimalist Program for Linguistic Theory (expressed by the first three Minimalist proposals) – the interface conditions and cognitive optimality. In the former case, we introduced the idea of a useability/'inheritability' interface; in the latter case, we problematized the issue as necessary and co-opted principles of algorithmic science to present a (partial) framework for Minimalist enquiry. The logic of evolutionary developmental biology was also introduced.

Chapter 3 discussed the two remaining Minimalist proposals. It was seen that a range of language-internal and language-external considerations are consistent with the fourth Minimalist proposal, which holds that core computational abilities of the FL were selected for their role in grounding a recursive 'language of thought', with language's externalization being a secondary concern. It was necessary to illustrate the ancestral nature of key features of the conceptual-intentional interface.

The second half of the chapter gave empirical and theoretical argument (largely) supporting the claim that language variation is highly restricted in the syntactic domain, but freer in domains relating to externalization – the fifth Minimalist proposal. In particular, I argued that such unambiguous syntactic variation as there is tallies strikingly with the predictions of the fourth Minimalist proposal, offering a slightly richer characterization of the fifth Minimalist proposal. I presented a simple approach to the phenomenon of *discourse pro-drop*, arguing that it coat-tails on the presence of the wide range of null lexical items associated with 'bare' nominal syntax, a uniform feature of discourse *pro*-drop languages.

Chapters 4 and 5 began the task of investigating the FL for evidence of optimal cognition and its principled breach by interface conditions. Chapter 4 focused on principles of optimality minimizing operational load in computation, whereas Chapter 5 focused on cache load. Both chapters dealt with the principles of substantive optimality and the interface conditions. Chapter 4 discussed evidence from: the Copy Theory of Movement, the Extension Condition, Bare Phrase Structure, Probe-Goal agreement and movement, adjunction, the Minimal Link Condition, binary branching, labelling, lexical categorization, the realization of syntactic copies at PF, strict cyclicity in phonology, patterns of vowel harmony, and the rules and representations of phonology. Novel arguments were interspersed. The findings were consistent with the implication of optimal cognition, the interface conditions, and the primacy of the CI-interface in the design of the FL.

Chapter 5 presented a novel approach to the value of phase-based Spell-Out, suggesting that it improves the throughput of linguistic computation as a whole by optimizing the scheduling with which consecutive algorithms release their output. More familiar arguments concerning its role in reducing the space complexity of linguistic algorithms were also presented. A novel account of the computational role of abstract Case was proposed, arguing that when we adopt an appropriately nuanced understanding of the structure of working memory, it becomes clear that it reduces cache load. Again, results were suggestive of the role of optimal cognition and interface conditions in the design of the FL.

Chapter 6 attempted to bring word order effects within Minimalist enquiry. In particular, core aspects of the parser were assessed in terms of how they affected the burden on working memory and the scheduling of linguistic tasks. A modified version of Pritchett's thematically-oriented, head-driven model of parsing was advocated. The proposed revision suggests the abstract parser is 'monotonic', in keeping with operational optimality, and highlights its close identity with the grammar itself, in keeping with substantive optimality. This chapter then proposed an alternative to the Kaynean account of the SPEC-H-COMP linearisation of abstract syntactic structure, couched in terms of the nature of the parser and principles of optimal computation. Finally I suggested that the Final-over-Final-Constraint falls out from concern for minimizing cache load in the mapping of surface strings onto the functional hierarchies of syntactic structure.

Chapter 7 attempted to bring certain aspects of first language acquisition within Minimalist enquiry. I argued for the implication of domain-general principles of data analysis: statistical learning; 'conservative' / parsimonious acquisition; and 'liberal' / analogical acquisition. The

second of these was argued to emerge from concern for computational optimality in (syntactic, morphological, and phonological) parsing of primary linguistic data by language acquirers, using their over-complete grammars. It was suggested that this explains various diachronic phenomena. 'Liberal' data analysis was taken to reflect the limits of working memory and to motivate the prevalence of analogical extension and levelling in the diachrony of morphology. I also discussed the clustering of word order systems around 'harmonic' poles in typology, and the problems associated with adopting an analogical account of this phenomenon. In doing so, I illustrated the appropriate place for functional explanation in formal linguistics.

The final substantive chapter, Chapter 8, offered accounts of the evolution of semantic abstraction, metaphorical thought, and humans' rich social and personal cognition. In particular, it argued that 'the linguistic condition' represents a 'channel' in natural selection of the kind of interest to evo-devo biologists (and linguistic Minimalism).

In conclusion, when pursued appropriately, the proposals of linguistic Minimalism are very powerful tools for investigating the design of the FL (including its broader role in human cognition). In as far as there are concrete Minimalist hypotheses, the early findings of research are supportive. Given the continuity of linguistic Minimalism with standard assumptions of scientific practice and biological thought, this is perhaps not a surprising result. Furthermore, it suggests that we can sensibly compare biolinguistics with other branches of biology at the same stage of their development, anticipating insights and discoveries beyond current contemplation – possibly even meaningless under current premises. To this end, it is essential that we continue scrutinizing current analyses for unappreciated assumptions and remain open to insights from related sciences.

Bibliography

Abe, K. & D. **Watanabe**. (2011). 'Songbirds possess the spontaneous ability to discriminate syntactic rules', in *Nature Neuroscience*, 14, pp. 1067-74.

Abner, N. (2011). 'Right where you belong: right peripheral WH-Elements and the WH-question paradigm in American Sign Language'. Master's thesis: UCLA.

Abels, K. (2003). *Successive Cyclicity, Anti-locality, and Adposition Stranding*. Ph.D. dissertation: University of Connecticut.

Abels, K. & A. **Neeleman**. (2009). 'Universal 20 without the LCA', in J. Brucart, A. Gavarró & J. Solà (eds.), *Merging Features: Computation, Interpretation, and Acquisition*, pp. 60-79. Oxford: Oxford University Press.

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. 219-56. Dordrecht: Kluwer.

Adler, D., G. **Bar-Oz**, A. **Belfer-Cohen** & O. **Bar-Yosef**. (2006). 'Ahead of the game: Middle and Upper Paleolithic hunting behaviours in the Southern Caucasus', in *Current Anthropology*, 47, pp. 89-118.

Ahn, H.-D. & S. **Cho**. (2012). 'Fragments vs. null arguments in Korean', in S. Müller (ed.), *Proceedings of the 19th International Conference on Head-Driven Phrase Structure Grammar*, pp. 369-87. Stanford: CSLI Publications.

Albright, A. (2008). 'Explaining Universal Tendencies and Language Particulars in Analogical Change', in J. Good (ed.), *Language Universals and Language Change*, pp. 144-81. Oxford: Oxford University Press.

Aldridge, E. (2004). *Ergativity and Word Order in Austronesian Languages*. Ph.D. dissertation: Cornell University.

Alexiadou, A. & E. **Anagnostopoulou**. (1998). 'Parameterizing AGR: word order, V-movement and EPP checking', in *Natural Language and Linguistic Theory*, 16, pp. 491-539.

Ambrose, S. (2002). 'Small Things Remembered: Origins of Early Microlithic Industries in Sub-Saharan Africa', in R. Elston & S. Kuhn (eds.), *Thinking Small: Global Perspectives on Microlithic Technologies*, pp. 9-29. Washington: American Anthropological Association.

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 Linguistic – Volume I: General Papers*, pp. 111-36. Amsterdam: John Benjamins.

Andrews, K. (2012). *Reading People: Folk Psychological Pluralism for Humans and Other Apes*. Cambridge, MA.: MIT Press.

Asbury, A., J. **Dotlačil**, B. **Gehrke** & R. **Nouwen** (eds.). (2008). *Syntax and Semantics of Spatial P*. Amsterdam: John Benjamins.

Axel, K. (2005). 'Null Subjects and Verb Placement in Old High Germanic', in S. Kesper & M. Reis (eds.), *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*, pp. 27-48. Berlin: Mouton de Gruyter.

Baayen, R. (2003). 'Probabilistic Approaches to Morphology', in R. Bod, J. Hay & S. Jannedy (eds.), *Probabilistic Linguistics*, pp. 229-87. Cambridge, MA.: MIT Press.

Baker, C. (1979). 'Syntactic theory and the projection problem', in *Linguistic Inquiry*, 10, pp. 533-81.

Baker, M. (1996). *The Polysynthesis Parameter*. Oxford: Oxford University Press.

Baker, M. (2003). *Lexical Categories*. Cambridge: Cambridge University Press.

Baker, M. (2008). *The Syntax of Agreement and Concord*. Cambridge: Cambridge University Press.

Baker, C. & J. **McCarthy** (eds.). (1981). *The Logical Problem of Language Acquisition*. Cambridge, MA.: MIT Press.

Barbiers, S. (2000). 'The Right Periphery in SOV Languages: English and Dutch', in P. Svenonius (ed.), *The Derivation of VO and OV*, pp. 181-218. Amsterdam: John Benjamins.

Barbosa, M. (2011). 'Partial pro drop as Null NP Anaphora'. Unpublished ms.: University of Minho. [As online as http://ling.auf.net/lingbuzz/001717].

Barbosa, M. (2013). '*Pro* as a minimal NP: towards a unified theory of *pro*-drop'. Unpublished ms.: University of Minho. [As online at http://ling.auf.net/lingbuzz/001949].

Barner, D., D. **Thalwitz**, J. **Wood**, S. **Yang** & S. **Carey**. (2007). 'On the relation between the acquisition of singular-plural morpho-syntax and the conceptual distinction between one and more than one', in *Developmental Science*, 10, pp. 365-73.

Baron-Cohen, S. & J. **Harrison** (eds.). (1997). *Synaesthesia: Classic and Contemporary Readings*. Oxford: Blackwell.

Barss, A. & H. **Lasnik**. (1986). 'A note on anaphora and double objects', in *Linguistic Inquiry*, 17, pp. 347-54.

Bar-Yosef, O. (2000). 'A Mediterranean Perspective on the Middle/Upper Palaeolithic Revolution', in C. Stringer, R. Barton & J. Finlayson (eds.), *Neanderthals on the Edge*, pp. 9-18. Oxford: Oxbow Books.

Bar-Yosef, O. & J.-G. **Bordes**. (2010). 'Who were the makers of the châtelperronian culture?', in *Journal of Human Evolution*, 59, pp. 586-93.

Battistella, E. (1996). *The Logic of Markedness*. Oxford: Oxford University Press.

Beckers, G., J. **Bolhuis**, K. **Okanoya** & R. **Berwick**. (2012) 'Birdsong neurolinguistics: songbird context-free grammar claim is premature', in *Neuroreport*, 23, pp. 139-45.

Berent, I. & S. **Pinker**. (2008). 'Compound formation is constrained by morphology: a reply to Seidenberg, MacDonald & Haskell', in *The Mental Lexicon*, 3, pp. 176-87.

Bernstein, J. (1991). 'DPs in Walloon: evidence for parametric variation in nominal head movement', in *Probus*, 3, pp. 101-26.

Berwick, R. (1985). *The Acquisition of Syntactic Knowledge*. Cambridge, MA.: MIT Press.

Berwick, R. & N. **Chomsky**. (2011). 'The Biolinguistic Program: the Current State of its Development', in A. Di Sciullo & C. Boeckx (eds.), in *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*, pp. 19-41. Oxford: Oxford University Press.

Berwick, R., A. **Friederici**, N. **Chomsky** & J. **Bolhuis**. (2013). 'Evolution, brain, and the nature of language', in *Trends in Cognitive Sciences*, 17, pp. 89-98.

Berwick, R. & P. **Niyogi**. (1996). 'Learning from triggers', in *Linguistic Inquiry*, 27, pp. 605-22.

Berwick, R., P. **Pietroski**, B. **Yankama** & N. **Chomsky**. (2011). 'Poverty of the stimulus revisited', in *Cognitive Science*, 35, pp. 1207-42.

Bever, T. (1970). 'The Cognitive Basis of Linguistic Structures', in J. Hayes (ed.), *Cognition and the Development of Language*, pp. 279-362. New York: Wiley & Sons.

Bever, T. (1982). 'Regression in the Service of Development', in T. Bever (ed.), *Regression in Mental Development: Basic Properties and Mechanisms*, pp. 153-88. Hillsdale, NJ.: Lawrence Erlbaum.

Biberauer, T. (2008). 'Introduction', in T. Biberauer (ed.), *The Limits of Syntactic Variation*, pp. 1-72. Amsterdam: John Benjamins.

Biberauer, T. (2010). 'Semi Null-Subject Languages, Expletives and Expletive *pro* Reconsidered', in T. Biberauer, A. Holmberg, I. Roberts & M. Sheehan (eds.), *Parametric Variation: Null Subjects in Minimalist Theory*, pp. 153-99. Cambridge: Cambridge University Press.

Biberauer, T., A. **Holmberg** & I. **Roberts**. (2014). 'A syntactic universal and its consequences', in *Linguistic Inquiry*, 45, pp. 169-225. (Appendix online at http://www.mitpressjournals.org/doi/suppl/10.1162/ling_a_00153).

Biberauer, T., A. **Holmberg**, I. **Roberts** & M. **Sheehan**. (2010). *Parametric Variation: Null Subjects in Minimalist Theory*. Cambridge: Cambridge University Press.

Biberauer, T., G. **Newton** & M. **Sheehan**. (2009). 'Limiting synchronic and diachronic variation and change: The Final-over-Final Constraint', in *Language and Linguistics*, 10, pp. 699-741.

Biberauer, T. & I. **Roberts**. (2009). 'The Return of the Subset Principle', in P. Crisma & G. Longobardi (eds.), *Historical Syntax and Linguistic Theory*, pp. 58-74. Oxford: Oxford University Press.

Biberauer, T. & I. **Roberts**. (2005). 'Subjects, Tense and Verb-movement', in T. Biberauer, A. Holmberg, I. Roberts & M. Sheehan (eds.), *Parametric Variation: Null Subjects in Minimalist Theory*, pp. 263-302. Cambridge: Cambridge University Press.

Biberauer, T., M. **Sheehan** & G. **Newton**. (2010). 'Impossible Changes and Impossible Borrowings: The Final-over-Final Constraint', in A. Breitbarth, C. Lucas, S. Watts & D. Willis (eds.), *Continuity and Change in Grammar*, pp. 35-60. Amsterdam: John Benjamins.

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*, pp. 267-84. Cambridge: Cambridge University Press.

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

Bloom, P. (1993). 'Grammatical continuiy in language development: the case of subject-less sentences', in *Linguistic Inquiry*, 24, pp. 721-34.

Bock, K. & K. **Miller**. (1991). 'Broken agreement', in *Cognitive Psychology*, 23, pp. 35-43.

Boden, M. (2005). *Mind as Machine: A History of Cognitive Science*. Oxford: Oxford University Press.

Boeckx, C. (2006). *Linguistic Minimalism: Origins, Concepts, Methods, and Aims*. Oxford: Oxford University Press.

Boeckx, C. (2011). 'Some Reflections on Darwin's problem in the Context of Cartesian Biolinguistics', in A. di Sciullo & C. Aguero (eds.), *Biolinguistic* Investigations, pp. 42-64. Cambridge, MA: MIT Press.

Bolhuis, J. & M. **Everaert**. (2013). *Birdsong, Speech and Language: Exploring the Evolution of Mind and Brain*. Cambridge, MA.: MIT Press.

Borer, H. (1984). *Parametric Syntax*. Dordrecht: Foris.

Borer, H. (1986). 'I-Subjects', in *Linguistic Inquiry*, 35, pp. 431-52.

Borer, H. (2005a). *In Name Only: Structure Sense: Vol. I*. Oxford: Oxford University Press.

Borer, H. (2005b). *The Normal Course of Events: Structuring Sense: Vol. II*. Oxford: Oxford University Press.

Bošković, Ž. (1997). *The Syntax of Non-Finite Complementation: An Economy Approach*. Cambridge, MA.: MIT Press.

Bošković, Ž. (2002a). 'On multiple *wh*-fronting', in *Linguistic Inquiry*, 33, pp. 351-83.

Bošković, Ž. (2002b). 'Clitics as non-branching elements and the Linear Correspondence Axiom', in *Linguistic Inquiry*, 33, pp. 329-40.

Bošković, Ž. (2002c). 'A-movement and the EPP', in *Syntax*, 5, pp. 167-218.

Bošković, Ž. (2007a). '*Agree*, phases, and intervention effects', in *Linguistic Analysis*, 33, pp. 54-96.

Bošković, Ž. (2007b). 'On the locality and motivation of Move and Agree: an even more minimal theory', in *Linguistic Inquiry*, 38, pp. 589-644.

Bošković, Ž. (2008). 'What will you have, DP or NP?', in E. Elfner & M. Walkow (eds.), *Proceedings of NELS 37*, pp. 101-4. Amherst: GLSA Publications.

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.

Bowers, J. (1993). 'The syntax of predication', in *Linguistic Inquiry*, 24, pp. 591-656.

Bresnan, J. (1982). *The Mental Representation of Grammatical Relations*. Cambridge, MA.: MIT Press.

Brody, M. (1997). 'Perfect Chains', in L. Haegeman (ed.), *Elements of Grammar*, pp. 139-67. Dordrecht: Kluwer.

Broekhuis, H. (2008). *Derivations and Evaluations: Object Shift in the Germanic Languages*. Berlin: Mouton de Gruyter.

Brown, R. (1973). *A First Language*. Cambridge, MA.: MIT Press.

Bruening, B. (2001). 'QR obeys superiority: frozen scope and ACD', in *Linguistic Inquiry*, 4, pp. 275-343.

Burke, A., C. **Nelson**, B. **Morgan** & C. **Tabin**. (1995). '*Hox* genes and the evolution of vertebrate axial morphology', in *Development,* 121, pp. 333-46.

Bybee, J. (2001). *Phonology and Language Use.* Cambridge: Cambridge University Press.

Cable, S. (2007). *The Grammar of Q*. Ph.D. dissertation: MIT.

Cable, S. (2010). *The Grammar of Q: Q-Particles, Wh-movement and Pied-Piping*. Oxford: Oxford University Press.

Caplan, D. (1972). 'Clause boundaries and recognition latencies for words in sentences', in *Perception & Psychophysics*, 12, pp. 73-6.

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. Berlin: Mouton de Gruyter.

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

Carroll, S., J. **Grenier** & S. **Weatherbee**. (2001). *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Malden, MA.: Blackwell.

Carruthers, P. (2006). *The Architecture of the Mind*. Oxford: Oxford University Press.

Carruthers, P. (2008). 'Meta-cognition in animals: a sceptical look', in *Mind and Language*, 23, pp. 58-89.

Carstens, V. (2002). 'Antisymmetry and word order in serial constructions', in *Language*, 78, pp. 3-50.

Catmur, C. (2014). 'Unconvincing support for role of mirror neurons in "action understanding": commentary on Michael *et al.* (2014)', in *Frontiers in Human Neuroscience*, 8, p. 553.

Cecchetto, C. (2004). 'Explaining the locality conditions of QR: consequences for the theory of phases', in *Natural Language Semantics*, 12, pp. 345-97.

Cecchetto, C. (2013). 'Backward Dependencies Must Be Short', in T. Biberauer & I. Roberts (eds.), *Challenges to Linearization*, pp. 57-92. Berlin: Mouton de Gruyter.

Cecchetto, C., C. **Geraci** & S. **Zucchi**. (2009). 'Another way to mark syntactic dependencies: the case for right-peripheral specifiers in sign languages', in *Language*, 85, pp. 278-320.

Cecchetto, C. & C. **Papagno**. (2011). 'Bridging the Gap between Brain and Syntax: A Case for a Role of the Phonological Loop', in A. Di Sciullo & C. Boeckx (eds.), *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of Human Language*, pp. 440-60. Oxford: Oxford University Press.

Chapin, P., T. **Smith** & A. **Abrahamson**. (1972). 'Two factors in perceptual segmentations of speech', in *Journal of Verbal Learning & Verbal Behavior*, 11, pp. 164-73.

Cheek, D. (2001). *The Phonetics and Phonology of Handshape in American Sign Language*. Ph.D. dissertation: University of Texas.

Cheney, D. & R. **Seyfarth**. (1990). *How Monkeys See the World: Inside the Mind of Another Species*. Chicago: University of Chicago Press.

Cheney, D. & R. **Seyfarth**. (2007). *Baboon Metaphysics: The Evolution of a Social Mind*. Chicago: University of Chicago Press.

Cheng, L. (2009). '*Wh*-in-situ, from the 1980s to now', in *Language and Linguistics Compass*, 3, pp. 767-91.

Cherniak, C. (1995). 'Neural component placement', in *Trends in Neuroscience*, 18, pp. 522-7.

Cherniak, C., Z. **Mokhtarzada**, R. **Rodriguez-Esteban** & K. **Changizi**. (2004). 'Global optimization of cerebral cortex layout', in *Proceedings of the National Academy of Sciences, USA*, 101, pp. 1081-6.

Chierchia, G. (1998). 'Reference to kinds across languages', in *Natural Language Semantics*, 6, pp. 339-405.

Chomsky, N. ([1955], 1975). *The Logical Structure of Linguistic Theory*. New York: Plenum.

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

Chomsky, N. (1959). 'A review of B.F. Skinner's *Verbal Behaviour*', in *Language*, 35, pp. 26-58.

Chomsky, N. (1961). 'Some methodological remarks on generative grammar', in *Word*, 17, pp. 219-39.

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. (1970). 'Remarks on Nominalization', in R. Jacobs & P. Rosenbaum (eds.), *Readings in English Transformational Grammar*, pp. 184-221. Waltham: Ginn.

Chomsky, N. (1972). *Language and Mind*. 2nd edition. New York: Harcourt Brace Jovanovich.

Chomsky, N. (1973). 'Conditions on Transformations', in S. Anderson & P. Kiparsky (eds.), *A Festschrift for Morris Halle*, pp. 232-86. New York: Holt, Rinehart & Winston.

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

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

Chomsky, N. (1983). 'The Psychology of Language and Thought', in R. Rieber (ed.), *Dialogues on the Psychology of Language and Thought*. New York: Plenum.

Chomsky, N. (1986a). *Knowledge of Language*. New York: Praeger.

Chomsky, N. (1986b). *Barriers*. Cambridge, MA.: MIT Press.

Chomsky, N. (1987). 'Reply', in *Mind & Language*, 2, pp. 178-97.

Chomsky, N. (1990). 'On formalization and formal linguistics', in *Natural Language and Linguistic Theory*, 8, pp. 143-7.

Chomsky, N. (1991). 'Some Notes on Economy of Derivation and Representation', in R. Freidin (ed.), *Principles and Parameters in Comparative Grammar*, pp. 417-54. Cambridge, MA.: MIT Press.

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

Chomsky, N. ([1995], 1998). 'Some Observations on Economy in Generative Grammar', transcript of talk delivered at a conference held at MIT, May 21st 1995, in P. Barbosa (ed.), *Is The Best Good Enough?*, pp. 115-27. Cambridge, MA.: MIT Press.

Chomsky, N. (1995a). 'Language and nature', in *Mind*, 104, pp. 1-61.

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

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

Chomsky, N. (1996). 'Silent children, new language'. Interview given to the BBC. [Online at http://www.chomsky.info/interviews/1996----.htm].

Chomsky, N. (2000a). '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. (2000b). *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University 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. (2002). *On Nature and Language*. Cambridge: Cambridge University Press.

Chomsky, N. (2004a). 'The Generative Enterprise Revisited', in R. Huybrechts, H. van Riemsdijk, N. Fukui & M. Zushi (eds.), *The Generative Enterprise Revisited*, pp. 147-87. Berlin: Mouton de Gruyter.

Chomsky, N. (2004b). 'Language and Mind: Current Thoughts on Ancient Problems', in L. Jenkins (ed.), *Variation and Universals in Biolinguistics*, pp. 379-405. Amsterdam: Elsevier.

Chomsky, N. (2004c). '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). '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 at Stony Brook University, New York, October 14th 2005. [As online at http://www.punksinscience.org/kleanthes/courses/MATERIALS/Chomsky_Evo-Devo.doc].

Chomsky, N. (2005b). 'Three factors in language design', in *Linguistic Inquiry*, 36, pp. 1-22.

Chomsky, N. (2007a). 'Symposium on Margaret Boden, *Mind as Machine: A History of Cognitive Science*', in *Artificial Intelligence,* 171, pp. 1094-1103.

Chomsky, N. (2007b). 'Approaching UG from Below', in U. Sauerland & H. Gärtner (eds.), *Interfaces + Recursion = Language?: Chomsky's Minimalism and the View from Syntax-Semantics*, pp. 1-29. Berlin: Mouton de Gruyter.

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*, pp. 133-66. Cambridge, MA.: MIT Press.

Chomsky, N. (2009a). 'Comments', in M. Piattelli-Palmarini, J. Uriagereka & P. Salaburu (eds.), *Of Minds & Language*, pp. 219-20. Oxford: Oxford University Press.

Chomsky, N. (2009b). 'Conclusion', in M. Piattelli-Palmarini, J. Uriagereka & P. Salaburu (eds.), *Of Minds & Language*, pp. 379-401. Oxford: Oxford University Press.

Chomsky, N. (2010). 'Poverty of the stimulus: unfinished business'. Transcript of a presentation given at Johannes-Gutenberg University, Mainz, March 24th 2010. [As online at http://www.stiftung-jgsp.uni-mainz.de/Bilder_allgemein/Mainz_transcript_edited.pdf].

Chomsky, N. (2011). 'Language and other cognitive systems: what is special about language?', in *Language Learning and Development*, 7, pp. 263-78.

Chomsky, N. (2012). *The Science of Language: Interviews with James McGilvray*. Cambridge: Cambridge University Press.

Chomsky, N. (2013). 'Problems of projection', in *Lingua*, 130, pp. 33-49.

Chomsky, N. (2014). 'Problems of projection: extensions'. Unpublished ms.: MIT.

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

Chomsky, N. & H. **Lasnik**. (1993). 'Principles and Parameters Theory', in J. Jacobs, A. Von Stechow, W. Sternfield & T. Vennemann (eds.), *Syntax: An International Handbook of Contemporary Research*, pp. 506-69. Berlin: Mouton de Gruyter.

Chung, S. (2012). 'Are lexical categories universal? The view from Chamorro', in *Theoretical Linguistics*, 38, pp.1–56.

Church, K. (1980). 'On memory limitations in natural language processing'. M.A. thesis: MIT.

Cinque, G. (1994). 'On the Evidence for Partial N-movement in the Romance DP', in G. Cinque, J. Koster, J.-Y. Pollock, L. Rizzi & R. Zanuttini (eds.), *Paths Towards Universal Grammar: Studies in Honor of Richard S. Kayne*, pp. 85-110. Washington: Georgetown University Press.

Cinque, G. (1999). *Adverbs and Functional Heads*. Oxford: Oxford University Press.

Cinque, G. (2010). *The Syntax of Adjectives*. Cambridge, MA.: MIT Press.

Cinque, G. (2013). 'Cognition, universal grammar, and typological generalizations', in *Lingua*, 130, pp. 50-65.

Cinque, G. & L. **Rizzi**. (2009). 'The Cartography of Syntactic Structures', in *The Oxford Handbook of Syntactic Structures*, pp. 51-66. Oxford: Oxford University Press.

Clark, A. & R. **Eyraud**. (2007). 'Polynomial time identification in the limit of substitutable

context-free languages', in *Journal of Machine Learning Research*, 8, pp. 1725-45.

Clark, A., R. **Eyraud** & A. **Habrard**. (2008). 'A Polynomial Algorithm for the Inference of Context-free Languages', in A. Clark, A Coste & L. Miclet (eds.), *Grammatical Inference: Algorithms and Applications, Lecture Notes in Computer Science 5728*, pp. 29-42. New York: Springer.

Clark, R. & I. **Roberts**. (1993). 'A computational approach to language learnability and language change', in *Linguistic Inquiry*, 24, pp. 299-345.

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

Clements, N. (2009). 'The Role of Features in Phonological Inventories', in E. Raims & C Cairns (eds.), *Contemporary Views on Architecture and Representations in Phonological Theory*, pp. 19-68. Cambridge, MA.: MIT Press.

Clifton, C., M. **Traxler**, M. **Mohamed**, R. **Williams**, R. **Morris** & K. **Rayner**. (2003). 'The use of thematic role information in parsing: syntactic processing autonomy revisited', in *Journal of Memory and Language*, 49, pp. 317-34.

Cole, M. (2009). 'Null subjects: a reanalysis of the data', in *Linguistics*, 47, pp. 559-87.

Collins, C. (1997). *Local Economy*. Cambridge, MA.: MIT Press.

Coon, J. (2010). 'VOS as predicate fronting in Chol', in *Lingua*, 120, pp. 354-78.

Coppock, E. (2007). 'Towards a true theory of the periphery'. Presented at the 33rd Annual Meeting of the Berkeley Linguistic Society 2007. [As online at http://eecoppock.info/CoppockBLS07.pdf].

Coppola, M. & E. **Newport**. (2005). 'Grammatical Subject in home sign: abstract linguistic structure in adult primary gesture systems without linguistic input', in *Proceedings of the National Academy of Sciences, USA*, 102, pp. 19249-53.

Crain, S. & J.D. **Fodor**. (1985). '**How can Grammars help Parsers**?', in D. Dowty, L. Karttunen & A. Zwicky (eds.), *Natural Language **Parsing,*** pp. 94-127. Cambridge: Cambridge University Press.

Crain, S. & C. **McKee**. (1985). 'The acquisition of structural restrictions on anaphora', in S. Berman, J.-W. Choe, & J. McDonough (eds.), *Proceedings of NELS 16*, pp. 94-111. Amherst: GLSA Publications.

Crain, S. & M. **Nakayama**. (1987). 'Structure dependency in grammar formation', in *Language*, 63, pp. 522-43.

Crain, S. & P. **Pietroski**. (2002). 'Why language acquisition is a snap', in *Linguistic Review*, 19, pp. 163-83.

Crockford, C., I. **Herbinger**, L. **Vigilant** & C. **Boesch**. (2004). 'Wild chimpanzees produce group-specific calls: a case for vocal learning?', in *Ethology*, 110, pp. 221-43.

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

Cuetos, F. & D. **Mitchell**. (1988). 'Cross-linguistic differences in parsing: restrictions on the use of the late closure strategy in Spanish', in *Cognition*, 30, pp. 73-105.

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

Cutler, A., D. **Oahan** & W. **van Donselaar**. (1997). 'Prosody in the comprehension of spoken language: a literature review', in *Language and Speech*, 40, pp. 141-201.

Cutting, J. (1982). 'Plucks and bows are categorically perceived, sometimes', in *Perception and Psychophysics*, 31, pp. 462-76.

Cutting, J. & B. **Rosner**. (1974). 'Category boundaries in speech and music', in *Perception and Psychophysics*, 16, pp. 564-70.

Dally, J., N. **Emery** & N. **Clayton**. (2006). 'Food-caching Western Scrub-Jays keep track of who was watching when', in *Science*, 312, pp. 1662-5.

D'Errico, F., M. **Vanhaeren**, C. **Henshilwood**, G. **Lawson**, B. **Maureille**, D. **Gambier**, A. **Tillier**, M. **Soressi** & K. **van Niekerk**. (2009). 'From the Origin of Language to the Diversification of Languages: What Can Archaeology and Palaeoanthropology Say?', in F d'Errico & J.-M. Hombert (eds.), *Becoming Eloquent: Advances in the Emergence of Language, Human Cognition, and Modern Cultures*, pp. 13-68. Amsterdam: John Benjamins.

De Saussure, F. ([1915], 1971). *Cours de Linguistique Générale*. Paris: Payot.

Dickerson, W. (1975). 'Variable rules in the speech community', in *Studies in Linguistic Sciences*, 5, pp. 41-68.

Di Pellegrino, G., L. **Fadiga**, L. **Fogassi**, V. **Gallese** & G. **Rizzolatti**. (1992). 'Understanding motor events: a neurophysiological study', in *Experimental Brain Research*, 91, pp. 176-80.

Długosz-Kurczabowa, K. & S. **Dubisz**. (2001). *Gramatyka Historyczna Jezyka Polskiego*. Warszawa: Wydawnictwo Uniwersytetu Warszawskiego.

Donati, C. (2006). 'On *Wh*-head Movement', in L. Cheng & N. Corver (eds.), *Wh-Movement: Moving On*, pp. 21-46. Cambridge, MA.: MIT Press.

Dresher, E. (1999). 'Charting the learning path: cues to parameter setting', in *Linguistic Inquiry*, 30, pp. 27-68.

Dresher, E. & J. **Kaye**. (1990). 'A computational learning model for metrical phonology', in *Cognition*, 34, pp. 137-95.

Dryer, M. (1992). 'On the Greenbergian word-order correlations', in *Language*, 68, pp. 81-138.

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.

Dryer, M. (2009). 'The Branching Direction Theory Revisited', in S. Scalise, E. Magni & A. Bisetto (eds.), *Universals of Language Today*, pp. 185-207. New York: Springer.

Dyer, F. & J. **Dickinson**. (1994). 'Development of sun compensation by honeybees: how partially experienced bees estimate the sun's course', in *Proceedings of the National Academy of Sciences, USA*, 91, pp. 4471-4.

Embick, D. (2010). *Localism versus Globalism in Morphology and Phonology*. Cambridge, MA.: MIT Press.

Embick, D. (2012). 'Roots and features (an acategorial postscript)', in *Theoretical Linguistics*, 38, pp. 73-89.

Emery, N. & N. **Clayton**. (2004). 'Western Scrub-Jays (*Aphelocoma Californica*) use cognitive strategies to protect their caches from thieving conspecifics', in *Animal Cognition*, 7, pp. 37-43.

Epstein, S., E. **Groat**, R. **Kawashima** & H. **Kitahara**. (1998). *The Derivation of Syntactic Relations*. Oxford: Oxford University Press.

Erlewhine., M. & H. **Kotek**. (2014). 'Intervention in focus-piedpiping', in H.-L. Huang, E. Poole & A. Rysling (eds.), *Proceedings of NELS 43*, pp. 117-30. Amherst: GLSA Publications.

Ernst, T. (2003). 'Semantic Features and the Distribution of Adverbs', in E. Lang, C. Maienborn & C. Fabricius-Hansen (eds.), *Modifying Adjuncts*, pp. 307-44. Berlin: Mouton de Gruyter.

Evans, C. & P. **Marler**. (1995). 'Language and Animal Communication: Parallels and Contrasts', in H. Roitblat & J. Meyer (eds.), *Comparative Approaches to Cognitive Science*, pp. 341-82. Cambridge MA.: MIT Press.

Everett, D. (2005). 'Cultural constraints on grammar and cognition in Pirahã: another look at the design features of human language', in *Current Anthropology*, 46, pp. 621-46.

Faarlund, J. (2013). 'The *pro* Cycle', in E. van Gelderen, J. Barðdal & M. Cennamo (eds.), *Argument Structure in Flux: The Naples-Capri Papers*, pp. 257-84. Amsterdam: John Benjamins.

Ferdinand, V., B. **Thompson**, S. **Kirby** & K. **Smith**. (2013). 'Regularization behavior in a non-linguistic domain', in M. Knauff, N. Sebanz, M. Pauen & I. Wachsmuth (eds.), *Bielefeld University Proceedings of the 35th Annual Cognitive Science Society*, pp. 436-41. Austin: Cognitive Science Society.

Ferreira, F. & C. **Clifton**. (1986). 'The independence of syntactic processing', in *Journal of Memory and Language*, 25, pp. 348-68.

Fillmore, C. (2002). 'Mini-grammars of *some time-when* Expressions in English', in J. Bybee & M. Noonan (eds.), *Complex Sentences in Grammar and Discourse: Essays in Honor of Sandra A. Thompson*, pp. 31-59. Amsterdam: John Benjamins.

Fiser, J. & R. **Aslin**. (2001). 'Unsupervised statistical learning of higher-order spatial structures from visual scenes', in *Psychological Science*, 12, pp. 499-504.

Fitch, W.T., M. **Hauser** & N. **Chomsky**. (2005). 'The evolution of the language faculty: clarifications and implications', in *Cognition*, 97, pp. 179-210. [Appendix online at http://www.wjh.harvard.edu/~mnkylab/publications/recent/EvolAppendix.pdf].

Flombaum, J. & L. **Santos**. (2005). 'Rhesus monkeys attribute perceptions to others', in *Current Biology*, 15, pp. 447-52.

Fodor, J.A. (1975). *The Language of Thought*. Cambridge, MA.: Harvard University Press.

Fodor, J.A., T. **Bever** & M. **Garrett**. (1974). *The Psychology of Language*. New York: McGraw Hill.

Fodor, J.D. (1978). 'Parsing strategies and constraints on transformations', in *Linguistic Inquiry*, 9, pp. 427-73.

Fodor, J.D. (1998). 'Unambiguous triggers', in *Linguistic Inquiry*, 29, pp. 1-36.

Fodor, J.D. (2002). 'Psycholinguistics cannot escape prosody' in B. Bel & I. Marlin (eds.), *Proceedings of Speech Prosody 2000 Conference*, pp. 83-98. Aix-en-Provence: Laboratoire Parole et Langage, CNRS / Université de Provence.

Folli, R. & H. **Harley**. (2005). 'Flavours of *v*: Consuming Results in Italian & English', in P. Kempchinsky & R. Slabakova (eds.), *Aspectual Enquiries*, pp. 95-120. Dordrecht: Springer.

Foulet, L. (1990). *Petite Syntaxe de l'Ancien Français*. Paris: Champion.

Ford, M., J. **Bresnan** & R. **Kaplan**. (1982). 'A Competence-Based Theory of Syntactic Closure', in J. Bresnan (ed.), *The Mental Representation of Grammatical Relations*, pp. 727-96. Cambridge, MA.: MIT Press.

Fortuny, J. (2010). 'On the Duality of Patterning', in J.-W. Zwart & M. de Vries (eds.), *Structure Preserved: Studies in Syntax for Jan Koster*, pp. 131-40. Amsterdam: John Benjamins.

Fox, D. (1995). 'Economy and scope', in *Natural Language Semantics*, 3, pp. 283-341.

Fox, D. (2000). *Economy and Semantic Interpretation*. Cambridge, MA.: MIT Press.

Franck, J., G. **Lassi**, U. **Frauenfelder** & L. **Rizzi**. (2006). 'Agreement and movement: a syntactic analysis of attraction', in *Cognition*, 101, pp. 173-216.

Franck, J., U. **Frauenfelder** & L. **Rizzi**. (2007). 'A syntactic analysis of interference in subject-verb agreement', in C. Schütze & V. Ferreira (eds.), *The State of the Art in Speech Error Research: Proceedings of the LSA Institute Workshop: MIT Working Papers in Linguistics 53*, pp. 173-90.

Frazier, L. (1978). *On Comprehending Sentences: Syntactic Parsing Strategies*. Ph.D. dissertation: University of Connecticut.

Frazier, L. (1987). 'Syntactic processing: evidence from Dutch', in *Natural Language and Linguistic* Theory, 5, pp. 519-60.

Frazier, L. & C. **Clifton**. (1989). '**Identifying gaps** in **English sentences', in** *Language and Cognitive Processes*, 4, pp. 93-126.

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.

Frazier, L. & J.D. **Fodor**. (1978). 'The sausage machine: a new two-stage parsing model', in *Cognition*, 6, pp. 291-325.

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

Fujita, K. (2007). 'Facing the logical problem of language evolution: review of Jenkins (2004), *Variation & Universals in Biolinguistics*', in *English Linguistics*, 24, pp. 78-108.

Gallego, A. (2014). 'Deriving Feature Inheritance from the Copy Theory of Movement', in *The Linguistic Review*, 31, pp. 41-71.

Gallistel, C.R. (1990). *The Organization of Learning*. Cambridge, MA.: MIT Press.

Gallistel, C.R. (1997). 'Neurons and Memory', in M. Gazzaniga (ed.), *Conversations in the Cognitive Neurosciences*, pp. 71-89. Cambridge, MA.: MIT Press.

Gallistel, C.R. (1998). 'Brains as Symbol Processors: the Case of Insect Navigation', in S. Sternberg & D. Scarborough (eds.), *Conceptual and Methodological Foundations: An Invitation to Cognitive Science*, pp. 1-51. Cambridge, MA.: MIT Press.

Gallistel, C.R. (2000). 'The Replacement of General-Purpose Learning Models with Adaptively Specialized Learning Models', in M Gazzaniga (ed.), *The New Cognitive Neuroscience*, pp. 1179-91. Cambridge, MA.: MIT Press.

Gallistel, C.R. (2002). 'Frequency, Contingency and the Information Processing Theory of Conditioning', in P. Sedlmeier & T. Betsch (eds.), *Etc. Frequency Processing and Cognition*, pp. 153-71. Oxford: Oxford University Press.

Gallistel, C.R. (2003). 'Conditioning from an information processing perspective', in *Behavioural Processes*, 62, pp. 89-101.

Gallistel, C.R. (2009). 'The Foundational Abstractions', in M. Piattelli-Palmarini, J. Uriagereka & P. Salaburu (eds.), *Of Minds & Language*, pp. 58-73. Oxford: Oxford University Press.

Gallistel, C.R. (2011). 'Prelinguistic thought', in *Language Learning and Development*, 7, pp. 253-62.

Gallistel, C.R. & A. **Cramer**. (1996). 'Computations on metric maps in mammals: getting oriented and choosing a multi-destination route', in *Journal of Experimental Biology*, 199, pp. 211-7.

Gallistel, C.R. & J. **Gibbon**. (2000). 'Time, rate and conditioning', in *Psychological Review*, 107, pp. 289-344.

Gallistel, C.R. & J. **Gibbon**. (2002). *The Symbolic Foundations of Conditioned Behavior*. Mahwah: Lawrence Erlbaum Associates.

Gallistel, C.R. & A. **King**. (2009). *Memory and the Computational Brain: Why Cognitive Science Will Revolutionize Neuroscience*. New York: Wiley-Blackwell.

Gallup, G. (1970). 'Chimpanzees: self-recognition', in *Science*, 167, pp. 341-3.

Gambell, T. & C. **Yang**. (2006). 'Word Segmentation: Quick but not Dirty'. Unpublished ms.: Yale University.

Gardner, A. (1957). 'Probability-learning with two and three choices', in *The American Journal of Psychology*, 70, pp. 174-85.

Garrett, M., T. **Bever** & J.A. **Fodor**. (1966). 'The active use of grammar in speech perception', in *Perception & Psychophysics*, 1, pp. 30-2.

Gelman, R. (1990). 'First principles organize attention to and learning about relevant data: number and the animate-inanimate distinction as examples', in *Cognitive Science*, 14, pp. 79-106.

Gelman, R. (2009). 'Innate Learning and Beyond', in M. Piattelli-Palmarini, J. Uriagereka & P. Salaburu (eds.), *Of Minds & Language*, pp. 223-35. Oxford: Oxford University Press.

Gergely, G., Z. **Nadasdy**, G. **Csibra** & S. **Bíró**. (1995). 'Taking the intentional stance at 12 months of age', in *Cognition*, 56, pp. 165-93.

Gervain, J., F. **Macagno**, S. **Cogoi**, M. **Peña** & J. **Mehler**. (2008). 'The neonate brain extracts speech structure', in *Proceedings of the National Academy of Sciences, USA*, 105, pp. 14222-7.

Gleitman, L., H. **Gleitman** & E. **Shipley**. (1973). 'The emergence of the child as grammarian', in *Cognition*, 1, pp. 137-64.

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

Gibson, E. & K. **Ko**. (1998). 'An integration-based theory of computational resources in sentence comprehension'. Paper presented at the 4th Architectures and Mechanisms in Language Processing Conference, University of Freiburg.

Gibson, E. & K. **Wexler**. (1994). 'Triggers', in *Linguistic Inquiry*, 25, pp. 355-407.

Gilligan, G. (1987). *A Cross-linguistic Approach to the* pro-*drop Parameter*. Ph.D. dissertation: USC.

Giorgi, A. & F, **Pianesi**. (1997). *Tense and Aspect: From Semantics to Morphosyntax*. Oxford: Oxford University Press.

Givón, T. (1985). 'Iconicity, Isomorphism and Non-Arbitrary Coding in Syntax', in J. Haiman (ed.), *Iconicity in Syntax*, pp. 187-219. Amsterdam: John Benjamins.

Goldin-Meadow, S. (2003). *The Resilience of Language*. New York: Psychology Press.

Goodman, C. & B. **Coughlin** (eds.). (2000). 'Special feature: the evolution of evo-devo biology', in *Proceedings of the National Academy of Sciences, USA*, 97, pp. 4424-56.

Gorrell, P. (1995). *Syntax and Parsing*. Cambridge: Cambridge University Press.

Gould, J. & P. **Marler**. (*1987*). 'Learning by instinct', in *Scientific American*, 256, pp. 62-73.

Gould, S. (2002). *The Structure of Evolutionary Theory*. Cambridge, MA.: Harvard University Press.

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

Grice, P. *(1989). Studies in the Way of Words. Cambridge, MA.: Harvard University Press.*

Grimshaw, J. (2005). *Words and Structure*. Stanford: CSLI Publications.

Guardiano, C. & G. **Longobardi**. (2006). 'The underlying unity of reference and quantification'. Paper presented at the XXXIIth *Incontro di Grammatica Generativa*, Florence.

Haegeman, L. (1995). *The Syntax of Negation*. Cambridge: Cambridge University Press.

Haider, H. (1997a). 'Precedence among predicates', in *Journal of Comparative Germanic Linguistics*, 1, pp. 3-41.

Haider, H. (1997b). 'Extraposition', in D. Beermann, D. LeBlanc & H. van Riemsdijk (eds.), *Rightward Movement*, pp. 115-51. Amsterdam: John Benjamins.

Haider, H. (2000). 'OV is more basic than VO', in P. Svenonius (ed.), *The Derivation of VO and OV*, pp. 45-67. Amsterdam: John Benjamins.

Haider, H. (2003). 'V-Clustering and Clause Union: Causes and Effects', in P. Seuren & G. Kempen (eds.), *Verb Constructions in German and Dutch*, pp. 91-126. Amsterdam: John Benjamins.

Hale, K. & J. **Keyser**. (1993). 'On Argument Structure and the Lexical Expression of Syntactic Relations', in K. Hale & J. Keyser (eds.), *The View from Building 20*, pp. 53-108. Cambridge, MA.: MIT Press.

Hale, K. & J. **Keyser**. (2002). *Prolegomenon to a Theory of Argument Structure*. Linguistic Inquiry Mongraph 39. Cambridge, MA.: MIT Press.

Hale, M. & C. **Reiss**. (1998). 'Formal and empirical arguments concerning language acquisition', in *Linguistic Inquiry*, 29, pp. 656-83.

Hale, M. & C. **Reiss**. (2008). *The Phonological Enterprise.* Oxford: Oxford University Press.

Halle, M. (1961). 'On the role of simplicity in linguistic description', in R. Jakobson (ed.), *Structure of Language and its Mathematical Aspects, Proceedings of the XIIth Symposium in Applied Mathematics,* pp. 89-94. Providence: American Mathematical Society.

Halle, M. (1962). 'Phonology in generative grammar', in *Word*, 18, pp. 54-72.

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.

Harley, H. (1995) *Subjects, Events, and Licensing*. Ph.D. dissertation: MIT.

Harris, Z. (1951). *Methods in Structural Linguistics*. Chicago: University of Chicago 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**. (2013). *The World Atlas of Language Structures Online*. [http://wals.info/].

Hauser, M. (2006). *Moral Minds: How Nature Designed our Universal Sense of Right and Wrong*. New York: Harper Collins.

Hauser, M. (2008). 'The Seeds of Humanity'. The Tanner Lectures of Human Values. Delivered at Princeton University, November 12th 2008. Transcripts available online at http://tannerlectures.utah.edu/_documents/a-to-z/h/Hauser_08.pdf.

Hauser, M. (2013). *Evilicious: Cruelty = Desire + Denial*.

Hauser, M., D. **Barner** & T. **O'Donnell**. (2007). 'Evolutionary linguistics: a new look at an old landscape', in *Language Learning and Development*, 3, pp. 101-32.

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.

Hauser, M. & D. **Glynn**. (2009). 'Can free-ranging rhesus monkeys (*Macaca Mulatta*) extract artificially created rules comprised of natural vocalizations?', in *Journal of Comparative Psychology*, 123, pp. 161-7.

Hauser, M, E. **Newport** & R. **Aslin**. (2001). 'Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins', in *Cognition*, 78, B53-B64.

Hauser, M., C. **Yang**, R. **Berwick**, I. **Tattersall**, M. **Ryan**, J. **Watumull**, N. **Chomsky** & R. **Lewontin**. (2014). 'The mystery of language evolution', in *Frontiers in Psychology*, 5, pp. 1-12.

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. (1990). 'A parsing theory of word order universals', in *Linguistic Inquiry*, 21, pp. 223-261.

Hawkins, J. (1994). *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.

Hawkins, J. (2001). 'Why are categories adjacent?', in *Journal of Linguistics*, 37, pp. 1-34.

Hawkins, J. (2004). *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

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

Hawkins, J. (2009). 'Language Universals and the Performance-Grammar Correspondence Hypothesis', in M. Christiansen, C. Collins & S. Edelman (eds.), *Language Universals*, pp. 54-78. Oxford: Oxford University Press.

Hawkins, J. (2013). 'Disharmonic Word Orders from a Processing Efficiency Perspective', in T. Biberauer & M. Sheehan (eds.), *Theoretical Approaches to Disharmonic Word Orders*, pp. 391-406. Oxford: Oxford University Press.

Hawkins, J. (2014). *Cross-Linguistic Variation and Efficiency*. Oxford: Oxford University Press.

Heck, F. & G. **Müller**. (2007). 'Derivational optimization of *wh*-Movement', in *Linguistic Analysis*, 33, pp. 97-148.

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*. 4th edition. San Francisco: Elsevier / Morgan Kaufmann.

Hickok, G. (2009). 'Eight problems for the mirror neuron theory of action understanding in monkeys and humans', in *Journal of Cognitive Neuroscience*, 21, pp. 1229-43.

Higham, T., R. **Jacobi**, M. **Julien**, F. **David**, L. **Basell**, R. **Wood** & C. **Ramsey**. (2010).

'Chronology of the grotte du renne (France) and implications for the context of ornaments and human remains within the châtelperronian', in *Proceedings of the National Academy of Sciences, USA*, 107, pp. 20234-9.

Hinzen, W. (2006). *Mind Design and Minimal Syntax*. Oxford: Oxford University Press.

Hinzen, W. (2008). 'Prospects for an explanatory theory of semantics', in *Biolinguistics*, 2, pp. 348-63.

Hinzen, W. & M. **Sheehan**. (2013). *The Philosophy of Universal Grammar*. Oxford: Oxford University Press.

Hockett, C. (1960). 'The origin of speech', in *Scientific American*, 203, pp. 88-96.

Holmberg, A. (2000). 'Deriving OV Order in Finnish', in P. Svenonius (ed.), *The Derivation of VO and OV*, pp.123-52. Oxford: Oxford University Press.

Holmberg, A. (2005). 'Is there a little pro? Evidence from Finnish', in *Linguistic Inquiry*, 36, pp. 533-64.

Holmberg, A. (2010). 'Null Subject Parameters', in T. Biberauer, A. Holmberg, I. Roberts & M. Sheehan (eds.), *Parametric Variation: Null Subjects in Minimalist Theory*, pp. 88-124. Cambridge: Cambridge University Press.

Holmberg, A. & T. **Hróarsdóttir**. (2004). 'Agreement and movement in Icelandic raising constructions', in *Lingua*, 114, pp. 651-73.

Holmberg, A. & I. **Roberts**. (2013a). 'The syntax-morphology relation', in *Lingua*, 130, pp. 111-31.

Holmberg, A. & I. **Roberts**. (2013b). 'Parameters and three factors of language design'. Unpublished ms.: Newcastle University / University of Cambridge.

Holmberg, A. & M. **Sheehan**. (2010). 'Control into Finite Clauses in Partial Null-Subject Languages', in T. Biberauer, A. Holmberg, I. Roberts & M. Sheehan (eds.), *Parametric Variation: Null Subjects in Minimalist Theory*, pp. 125-52. Cambridge: Cambridge University Press.

Hopper, P. & E. **Traugott**. (1993). *Grammaticalization*. Cambridge: Cambridge University Press.

Hornstein, N., J. **Nunes** & K. **Grohmann**. (2005). *Understanding Minimalism*. Cambridge: Cambridge University Press.

Hornstein, N. & A. **Weinberg**. (1981). 'Case theory and preposition stranding', in *Linguistic Inquiry*, 12, pp. 55-91.

Hróarsdóttir, Þ. (2000). *Word Order Change in Icelandic: From OV to VO*. Amsterdam: John Benjamins.

Huang, C.-T. (1982). *Logical Relations in Chinese and the Theory of Grammar*. Ph.D. dissertation: MIT.

Huang, C.-T. (2013). 'On Syntactic Analyticity and Parametric Theory', to appear in C.-T. Huang, A. Simpson & A. Li (eds.), *Handbook of Chinese Linguistics.* Oxford: Wiley-Blackwell.

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

Hudson Kam, C. & A. **Chang**. (2009). 'Investigating the cause of language regularization in adults: memory constraints or learning effects?', in *Journal of Experimental Psychology*, 35, pp. 815-21.

Hudson Kam, C. & E. **Newport**. (2005). 'Regularizing unpredictable variation: the roles of adult and child learners in language formation and change', in *Language Learning and Development*, 1, 151-95.

Hudson Kam, C. & E. **Newport**. (2009). 'Getting it right by getting it wrong: when learners change languages', in *Cognitive Psychology*, 59, pp. 30-66.

Hunt, R. & R. **Aslin**. (2001). 'Statistical learning in a serial reaction time task: simultaneous extraction of multiple statistics', in *Journal of Experimental Psychology: General*, 130, pp. 658-80.

Hurford, J. (2007). *The Origins of Meaning*. Oxford: Oxford University Press.

Hyams, N. (1986). *Language Acquisition and the Theory of Parameters*. Dordrecht: Reidel.

Inoue, K. (2006). 'Nihongo-no jookenbun-to shubun-no mdartii' ['Conditional construction and modality in the matrix clause in Japanese'], in *Scientific Approaches to Language*, 5, pp. 9-28.

Ishihari, S. (2003). *Intonation and Interface Conditions*. Ph.D. dissertation: MIT.

Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. Cambridge, MA.: MIT Press.

Jackendoff, R. (1977). *X-bar Syntax: A Study of Phrase Structure*. Linguistic Inquiry Monograph 2. Cambridge, MA.: MIT Press.

Jackendoff, R. (1992). *Languages of the Mind*. Cambridge, MA.: MIT Press.

Jackendoff, R. & S. **Pinker**. (2005). 'The nature of the language faculty and its implications for the evolution of language', in *Cognition*, 97, pp. 211-25.

Jacob, F. (1982). *The Possible and the Actual*. New York: Pantheon.

Jakobson, R. (1941). *Kindersprache, Aphasie, und allgemeine Lautgesetze*. Uppsala: Almqvist & Wiksell.

Janik, V., L. **Sayigh** & R. **Wells**. (2006). 'Signature whistle shape conveys identity information to bottlenose dolphins', in *Proceedings of the National Academy of Sciences, USA*, 103, pp. 8293-7.

Janik, V. & P. **Slater**. (1997). 'Vocal learning in mammals', in *Advances in the Study of Behavior*, 26, pp. 59-99.

Jarvella, R. (1971). 'Syntactic processing of connected speech', in *Journal of Verbal Learning and Verbal Behavior*, 10, pp. 409-16.

Joos, M. (1957). *Readings in Linguistics: The Development of Descriptive Linguistics in America since 1925.* Washington: American Council of Learned Societies.

Jusczyk, P. (1997). *The Discovery of Spoken Language*. Cambridge, MA.: MIT Press.

Jusczyk, P. & R. **Aslin**. (1995). 'Infant's detection of the sound patterns of words in fluent speech', in *Cognitive Psychology*, 46, pp. 65-97.

Jusczyk, P., A. **Cutler** & N. **Redanz**. (1993). 'Preference for the predominant stress patterns of English words', in *Child Development*, 64, pp. 675-87.

Jusczyk, P., K. **Hirsch-Pasek**, D. **Kemler Nelson**, L. **Kennedy**, A. **Woodward** & J. **Piwoz**. (1992). 'Perception of acoustic correlates of major phrasal units by young infants', in *Cognitive Psychology*, 24, pp. 252-93.

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

Kamin, L. (1969). 'Selective Association and Conditioning', in N. Mackinstosh & W. Honig (eds.), *Fundamental Issues in Associative Learning*, pp. 42-64. Halifax: Dalhousie University Press.

Kaminski, J., J. **Call** & M. **Tomasello**. (2008). 'Chimpanzees know what other know, but not what they believe', in *Cognition*, 109, pp. 224-34.

Kareev, Y., I. **Lieberman** & M. **Lev**. (1997). 'Through a narrow window: sample size and the perception of correlation', in *Journal of Experimental Psychology*, 126, pp. 278-87.

Katz, J. & D. **Pesetsky**. (2009). 'The Identity Thesis for Language and Music'. Unpublished ms.: MIT.

Kayne, R. (1984). *Connectedness and Binary Branching*. Dordrecht: Foris.

Kayne, R. (1989, [2000]). 'Notes on English Agreement', reprinted in *Parameters and Universals*, pp. 187-205. Oxford: Oxford University Press.

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

Kayne, R. (2000). *Parameters and Universals.* Oxford: Oxford University Press.

Kayne, R. (2005a). *Movement and Silence*. Oxford: Oxford University Press.

Kayne, R. (2005b). 'Some Notes on Comparative Syntax, with Special Reference to English and French', in G. Cinque & R. Kayne (eds.), *The Oxford Handbook of Comparative Syntax*, pp. 3-69. Oxford: Oxford University Press.

Kayne, R. (2013). 'Why are there no directionality parameters?', in *Studies in Chinese Linguistics*, 34, pp. 3-37.

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

Kimball, J. (1973). 'Seven principles of surface structure parsing in natural language', in *Cognition*, 2, pp. 15-47.

Kimball, J. & J. **Aissen**. (1971). 'I think, you think, he think', in *Linguisic Inquiry*, 2, pp. 242-6.

King, J. & M. **Just**. (1991). 'Individual differences in syntactic processing: the role of working memory', in *Journal of Memory and Language*, 30, pp. 580-602.

Kinsella, A. & G. **Marcus**. (2009). 'Evolution, perfection, and theories of language', in *Biolinguistics*, 3, pp. 186-212.

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

Kiparsky, P. (1973). ''Elsewhere' in Phonology', in S. Anderson & P. Kiparsky (eds.), *A Festschrift for Morris Halle*, pp. 93-106. New York: Holt, Reinhart & Winston

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. (1995). 'The Phonological Basis of Sound Change', in J. Goldsmith (ed.), *The Handbook of Phonological Theory*, pp. 640-70. Oxford: Blackwell.

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. (2010). 'Dvandvas, blocking, and the associative: the bumpy ride from phrase to word', in *Language*, 86, pp. 302-21.

Kiparsky, P. (2014). 'New Perspectives in Historical Linguistics', in C. Bowern & B. Evans (eds.), *The Routledge Handbook of Historical Linguistics*, pp. 64-102. Abingdon: Routledge.

Kirkham, N., J. **Slemmer** & S. **Johnson**. (2002). 'Visual statistical learning in infancy: evidence of a domain general learning mechanism', in *Cognition*, 83, B34-B42.

Kitahara, H. (2003). 'Acoustic perception in human newborns & cotton-top tamarins', in *Reports of the Keio Institute of Cultural & Linguistic Studies*, 35, pp. 31-41.

Klein, R. (2009). *The Human Career*. Chicago: Chicago University Press.

Kobuchi-Philip, M. (2010). 'Indeterminate numerical quantifiers, 'some' and 'many' readings, and questions in Japanese', in *Korean Journal of Linguistics*, 35, pp. 503-30.

Kohl, K. (1999). 'An analysis of finite parameter learning in linguistic space'. M.A. thesis: MIT.

Koopman, H. (2006). 'Agreement Configurations: in Defense of "Spec-Head"', in C. Boeckx (ed.), *Agreement Systems*, pp. 159-200. Amsterdam: John Benjamins.

Koster, J. (1974). 'Het werkwoord als spiegelcentrum', in *Spektator: Tijdschrift voor Neerlandistiek*, pp. 601-18.

Koster, J. (1978). *Locality Principles in Syntax*. Dordrecht: Foris.

Koster, J. (1987). *Domains and Dynasties: The Radical Autonomy of Syntax*. Dordrecht: Foris.

Krämer, M. (2012). *Underlying Representations*. Cambridge: Cambridge University Press.

Kurafuji, T. (2004). 'Plural morphemes, definiteness, and the notion of semantic parameter', in *Language and Linguistics*, 5, pp. 211-42.

Labov, W. (1981). 'Resolving the Neogrammarian controversy', in *Language*, 57, pp. 267-309.

Labov, W. (1986). 'The exact description of the speech community: short /a/ in Philadelphia'. Unpublished ms..

Landau, I. (2006). 'Chain resolution in Hebrew V(P)-fronting', in *Syntax*, 9, pp. 32-66.

Landau, B. & L. **Gleitman**. (1985). *Language and Experience: Evidence from the Blind Child.* Cambridge, MA.: Harvard University Press.

Lappin, S. & D. **Johnson**. (1999). *Local Constraints vs. Economy*. Stanford: CSLI Publications.

Lappin, S., R. **Levine** & D. **Johnson**. (2000a). 'The structure of unscientific revolutions', in *Natural Language and Linguistic Theory*, 18, pp. 665-71.

Lappin, S., R. **Levine** & D. **Johnson**. (2000b). 'The revolution confused: a response to our critics', in *Natural Language and Linguistic Theory*, 18, pp. 873-90.

Lappin, S., R. **Levine** & D. **Johnson**. (2001). 'The revolution maximally confused', in *Natural Language and Linguistic Theory*, 19, pp. 901-19.

Larson, R. (1988). 'On the double object construction', in *Linguistic Inquiry*, 19, pp. 335-91.

Lasnik, H. (2003). *Minimalist Investigations in Linguistic Theory*. London: Routledge.

Lasnik, H. & M. **Saito**. (1991). 'On the subject of infinitives', in L. Dobrin, L. Nichols & R. Rodriguez (eds.), *Papers from the 27th Regional Meeting of the Chicago Linguistic Society*, pp. 324-43.

Leben, W. (1973). *Suprasegmental Phonology*. Ph.D. dissertation: MIT.

Legate, J. & C. **Yang**. (2002). 'Empirical re-assessment of stimulus poverty arguments', in *The Linguistic Review*, 19, pp. 151-62.

Legate, J. & C. **Yang**. (2005). 'The richness of the Poverty of the Stimulus'. Hand-out from talk given at The Golden Anniversary of *The Logical Structure of Linguistic Theory* Conference, organized by the LSA Institute, Harvard University / MIT, July 16th 2005.

Legate, J. & C. **Yang**. (2007). 'Morphosyntactic learning and development of tense', in *Language Acquisition*, 14, pp. 315-44.

Legate, J. & C. **Yang**. (2012). 'Assessing Child and Adult Grammar', in M. Piattelli-Palmarini & R. Berwick (eds.), *Rich Languages from Poor Inputs*, pp. 168-83. Oxford: Oxford University Press.

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

Lewis-Williams, D. (2002). *The Mind in the Cave*. London: Thames & Hudson.

Lewontin, R. (2001). *The Triple Helix*. New York: New York Review of Books Press.

Levelt, W. (1974). *Formal Grammars in Linguistics and Psycholinguistics*. The Hague: Mouton.

Levin, B. & M. **Rappaport-Hovav**. (1995). *Unaccusativity.* 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.

Lightfoot, D. (1999). *The Development of Language*. Oxford: Blackwell.

Lin, T.-H. (2006). 'Complement-to-Specifier Movement in Mandarin Chinese'. Unpublished ms.: National Tsing Hua University.

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

Longobardi, G. (2008). 'Reference to Individuals, Person, and the Variety of Mapping Parameters', in A. Klinge, & H. Müller (eds.), *Essays on Nominal Determination: from Morphology to Discourse Management.*, pp. 189-211. Amsterdam: John Benjamins.

Louw, J. (1962). 'On the segmental phonemes of Zulu', in *Afrika und Übersee*, 46, pp. 43-93.

Lowenstamm, J. (2008). 'On *n*, √, and Types of Nouns', in J. Hartmann, V. Hegedus & H. van Riemsdijk (eds.), *Sounds of Silence: Empty Elements in Syntax and Phonology*, pp. 105-40. Amsterdam: Elsevier

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.

Lutz, B., H. **Lu**, G. **Eichele**, D. **Miller** & T. **Kaufman**. (1996). 'Rescue of *Drosophila* labial null mutant by the chicken ortholog *Hoxb-1* demonstrates that the function of *Hox* genes is phylogenetically conserved', in *Genes & Development*, 10, pp. 176-84.

MacDonald, M. (1994). 'Probabilistic constraints and syntactic ambiguity resolution', in *Language and Cognitive Processes*, 9, pp. 157-201.

MacDonald, M., N. **Perlmutter** & M. **Seidenberg**. (1994). 'Lexical nature of syntactic ambiguity resolution', in *Psychological Review*, 101, pp. 676-703.

Mackintosh, N. (1975). 'A theory of attention: variations in the associability of stimuli with reinforcement', in *Psychological Review*, 82, pp. 276-98.

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

Mandel, D., P. **Jusczyk** & D. **Pisoni**. (1995). 'Infants' recognition of the sound patterns of their own names', in *Psychological Science*, 6, pp. 314-7.

Manser, M., M. **Bell** & I. **Fletcher**. (2001). 'The information that receivers extract from alarm calls in suricates', in *Proceedings of the Royal Society B: Biological Sciences*, 268, pp. 2485-91.

Marantz, A. (1993). 'Implications of Asymmetries in Double Object Constructions', in S. Mchombo (ed.), *Theoretical Aspects of the Bantu Grammar*, pp. 113-50. Stanford: CLSI Publications.

Marantz, A. (1997). 'No escape from syntax: don't try morphological analysis in the privacy of your own Lexicon', *in* Penn Working Papers in Linguistics, 4, pp. 201-25.

Marantz, A. (2001). 'Words and Things'. Handout: MIT.

Marantz, A. (2007). 'Phases and Words', in S.-H. Choe (ed.), *Phases in the Theory of Grammar*, pp. 191-222. Seoul: Dong In.

Marantz, A. (2010). 'Constraints on contextual allosemy'. Handout from talk given at New York University.

Marcus, G., S**. Pinker**, M. **Ullman**, M. **Hollander**, J. **Rosen** & F. **Xu**. (1992). *Over-regularization in Language Acquisition*. Monographs of the Society for Research in Child Development, 57.

Marcus, M., D. **Hindle** & M. **Fleck**. (1983). 'D-theory: talking about talking about trees', in *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, pp. 129-36. Cambridge, MA.: MIT Press.

Marler, P., C. **Evans** & M. **Hauser**. (1992). 'Animal Signals: Motivational, Referential, or Both?', in H. Papousek, U. Jürgens & M. Papousek. (eds.), *Nonverbal Vocal Communication: Comparative and Developmental Approaches*, pp. 66-86. Cambridge: Cambridge University Press.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: Freeman.

Marslen-Wilson, W. & A. **Welsh**. (1978). 'Processing interactions and lexical access during word recognition in continuous speech', in *Cognitive Psychology*, 10, pp. 29-63.

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

Massam, D. (2000). 'VSO and VOS: Aspects of Niuean Word Order', in A. Carnie & E. Guilfoyle (eds.), *The Syntax of Verb Initial Languages*, pp. 97-116. Oxford: Oxford University Press.

Massam, D. (2001). 'Pseudo noun incorporation in Niuean', in *Natural Language and Linguistic Theory*, 19, pp. 153-97.

Matthews, R. (2006). 'The Case for Linguistic Nativism', in R. Stainton (ed.), *Contemporary Debates in Cognitive Science*, pp. 81-97. Oxford: Blackwell.

Mauk, C. (2003). *Undershoot in Two Modalities: Evidence from Fast Speech and Fast Signing.* Ph.D. dissertation: University of Texas.

May, R. (1985). *Logical Form: Its Structure and Derivation*. Cambridge, MA.: MIT Press.

Maynard Smith, J., R. **Burian**, S. **Kauffman**, P. **Alberch**, J. **Campbell**, B. **Goodwin**, R. **Lande**, D. **Raup** & L. **Wolpert**. (1985). 'Developmental constraints and evolution', in *Quarterly Review of Biology*, 60, pp. 265-87.

Mazuka, R. & K. **Itoh**. (1995). 'Can Japanese Speakers be Led Down the Garden Path?', in R. Mazuka & N. Nagai (eds.), *Japanese Sentence Processing*, pp. 295-330. Hillsdale, NJ.: Lawrence Erlbaum.

Mazuka, R., K. **Itoh**, S. **Kiritani**, S. **Niwa**, K. **Ikejiru** & K. **Naito**. (1989). 'Processing of Japanese garden path, center-embedded, and multiply-left-embedded sentences', in *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, 23.

Mazurkewich, I. & L. **White**. (1984). 'The acquisition of the dative alternation: unlearning overgeneralizations', in *Cognition*, 16, pp. 261-83.

McMahon, A. (1994). *Understanding Language Change*. Cambridge: Cambridge University Press.

McRae, K., M. **Spivey-Knowlton** & M. **Tanenhaus**. (1998). 'Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension', in *Journal of Memory and Language*, 38, pp. 283-312.

Mellars, P. (1996). *The Neanderthal Legacy*. Princeton: Princeton University Press.

Michael, J., K. **Sandberg**, J. **Skewes**, T. **Wolf**, J. **Blich**, M. **Overgaard** & C. **Frith**. (2014). 'Continuous theta-burst stimulation demonstrates a causal role of premotor homunculus in action understanding', in *Psychological Science*, 25, pp. 963-72.

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

Miyagawa, S. (2010). *Why Agree? Why Move?*. Cambridge, MA.: MIT Press.

Miyagawa, S. & T. **Tsujioka**. (2004). 'Argument structure and ditransitive verbs in Japanese', in *Journal of East Asian Linguistics*, 13, pp. 1-38.

Mikhail, J. (2007). 'Universal moral grammar: theory, evidence and the future', in *Trends in Cognitive Sciences*, 11, pp. 143-52.

Miller, G. & N. **Chomsky**. (1963). 'Finitary Models of Language Users', in R. Luce, R. Bush & E. Galanter (eds.), *Handbook of Mathematical Psychology, Vol. II*, pp. 419-92. New York: Wiley & Sons.

Mitchell, D. (1994). 'Sentence Parsing', in M. Gernsbacher (ed.), *Handbook of Psycholinguistics*, pp. 375-409. New York: Academic Press.

Mobbs, I. (2008). ''Functionalism', the design of the language faculty, and (disharmonic) typology'. M.Phil thesis: University of Cambridge. [As online at http://ling.auf.net/lingBuzz/000680].

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

Mulders, I. (2002). *Transparent Parsing: Head-driven Processing of Verb-final Structures*. Utrecht: LOT Publications.

Müller, G. (2004). 'Argument Encoding and the Order of Elementary Operations'. Ms.: Universität Leipzig.

Müller, G. (2009). 'Ergativity, Accusativity, and the Order of Merge and Agree', in K. Grohmann (ed.), *Explorations of Phase Theory: Features and Arguments*, pp. 269-308. Berlin: Mouton de Gruyter.

Narita, H. (2011). *Phasing in Full Interpretation*. Ph.D. dissertation: Harvard University

Neeleman, A. & K. **Szendrői**. (2007). 'Radical *pro*-drop and the morphology of pronouns', in *Linguistic Inquiry*, 38, pp. 671-714.

Neeleman, A. & K. **Szendrői**. (2008). 'Case Morphology and Radical *pro*-drop', in T. Biberauer (ed.), *The Limits of Syntactic Variation*, pp. 331-48. Amsterdam: John Benjamins.

Nemoto, N. (2005). 'On mass denotations of bare nouns in Japanese and Korean', in *Linguistics*, 43, pp. 383-413.

Newell, H. (2008). *Aspects of the Morphology and Phonology of Phases*. Ph.D. dissertation: MIT.

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.

Nevins, A. (2010). *Locality in Vowel Harmony*. Cambridge, MA.: MIT Press.

Ni, W., S. **Crain** & D. **Shankweiler**. (1996). 'Sidestepping garden paths: the contribution of syntax, semantics and plausibility in resolving ambiguities', in *Language and Cognitive Processes*, 11, pp. 283-334.

Nissenbaum, J. (2000). *Investigations of Covert Phrase Movement*. Ph.D. dissertation: MIT.

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.

Nunes, J. (2004). *The Linearization of Chains and Sideward Movement*. Cambridge, MA.: MIT Press.

Nunes, J. (2011). 'The Copy Theory', in C. Boeckx (ed.), *The Oxford Handbook of Linguistic Minimalism*, pp. 143-72. Oxford: Oxford University Press.

Odden, D. (1986). 'On the role of the Obligatory Contour Principle in Phonological Theory', in *Language*, 62, pp. 353-83.

Odden, D. (1988). 'Antigemination and the OCP', in *Linguistic Inquiry*, 19, pp. 451-75.

Oerhle, R. (1976). *The Grammatical Status of the English Dative Alternation*. Ph.D. dissertation: MIT.

Ohala, J. (1994). 'Towards a universal, phonetically-based theory of vowel harmony', in *Third International Conference on Spoken Language Processing*, 94, pp. 491–494. [Available online at http://www.isca-speech.org/isclp_1994].

Oku, S. (1998). 'LF copy analysis of Japanese null arguments', in *Proceedings of the Chicago Linguistic Society*, 34, pp. 299-314.

Ott, D. (2009). 'The evolution of I-language: lexicalization as the key evolutionary novelty', in *Biolinguistics*, 3, pp. 255-69.

Ott, D. (2011). *Local Instability: The Syntax of Split Topics*. Ph.D. dissertation: Harvard University.

Panagiotidis, E. (2011). 'Categorial features and categorizers', in *The Linguistic Review*, 28, pp. 365-86.

Parker, S., M. **Boccia** & R. **Mitchell** (eds.). (1994). *Self-Awareness in Animals and Humans: Developmental Perspectives.* Cambridge: Cambridge University Press.

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

Paul, I. (2000). *Malagasy Clause Structure*. Ph.D. dissertation: McGill University.

Pearce, J. & G. **Hall**. (1980). 'A model for Pavlovian learning: variation in the effectiveness of conditioned by not of unconditioned stimuli', in *Psychological Review*, 87, pp. 532-52.

Pearson, M. (2000). 'Two types of VO languages', in P. Svenonius (ed.), *The Derivation of VO and OV*, pp. 327-63. Amsterdam: John Benjamins.

Pearson, M. (2001). *The Clause Structure of Malagasy: A Minimalist Approach*. Ph.D. dissertation: UCLA.

Perfors, A., J. **Tenenbaum** & T. **Regier**. (2011). 'Poverty of the stimulus: a rational approach', in *Cognition*, 118, pp. 306-38.

Perlmutter, D. (1971). *Deep Structure and Surface Structure Constraints in Syntax*. New York: Holt, Rinehart & Winston.

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.

Pesetsky, D. & E. **Torrego** (2006). 'Probes, goals, and the nature of syntactic categories', in Y. Otsu (ed.), *Proceedings of the Seventh Tokyo Conference on Psycholinguistics*, pp. 25-60. Tokyo: Hituzi Syobo Publishing.

Petitto, L. (1987). 'On the autonomy of language and gesture: evidence from the acquisition of personal pronouns in American Sign Language', in *Cognition*, 27, pp. 1-52.

Petitto, L. (2005). 'How the Brain Begets Language', in J. McGilvray, (ed.), *The Cambridge Companion to Chomsky*. Cambridge: Cambridge University Press.

Petitto, L., M. **Berens**, I. **Kovelman**, M. **Dubins**, K. **Jasinska** & M. **Shalinsky**. (2012). 'The "Perceptual Wedge Hypothesis" as the basis for bilingual babies' phonetic processing advantage: new insights from fNIRS brain imaging', in *Brain & Language*, 121, pp. 130-43.

Petitto, L & P. **Marentette**. (1991). 'Babbling in the manual mode: evidence for the ontogeny of language', in *Science*, 251, pp. 1483-96.

Petitto, L., J. **Zatorre**, K. **Gauna**, E. **Nikelski**, D. **Dostie** & A. **Evans**. (2000). 'Speech-like cerebral activity in profoundly deaf people while processing signed languages: implications for the neural basis of all human language', in *Proceedings of the National Academy of Sciences, USA*, 97, pp. 13961-6.

Philip, J. (2013). '(Dis)harmony, the Head-Proximate Filter, and linkers', in *Journal of Linguistics* 49, pp. 165-213.

Phillips, C. (1995). *Order and Structure*. Ph.D. dissertation: MIT.

Piattelli-Palmarini, M. & J. **Uriagereka** (2004). 'Immune Syntax', in L. Jenkins (ed.), *Universals and Variations in Biolinguistics*, pp. 341-77. Amsterdam: Elsevier.

Pierce, A. (1989). *On the Emergence of Syntax: A Crosslinguistic Study*. Ph.D. dissertation: MIT.

Pinker, S. (1994). *The Language Instinct*. New York: Harper Collins.

Pinker, S. & P. **Bloom**. (1990). 'Natural language and natural selection', in *Behavioral and Brain Sciences*, 13, pp. 707-84.

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

Pinker, S. & A**. Prince**. (1988). 'On language and connectionism: analysis of a parallel distributed processing model of language acquisition', in *Cognition*, 28, pp. 73-193.

Polinsky, M. & E. **Potsdam**. (2013). 'Diagnosing Covert A-Movement', in L. Cheng & N. Corver (eds.), *Diagnosing Syntax*, pp. 210-34. Oxford: Oxford University Press.

Preminger, O. (2012). 'Reversing agreement is untenable: a reply to Zeijlstra'. Unpublished ms.: online at http://ling.auf.net/lingbuzz/001442.

Pritchett, B. (1992a). 'Parsing with Grammar: Islands, Heads and Garden Paths', in H. Goodluck & M. Rochemont (eds.), *Island Constraints: Theory, Acquisition and Processing – Studies in Theoretical Psycholinguistics*, 15, pp. 321-49. Dordrecht: Kluwer.

Pritchett, B. (1992b). *Grammatical Competence and Parsing Performance*. Chicago: University of Chicago Press.

Przezdziecki, M. (2005). *Vowel Harmony and Co-articulation in Three Dialects of Yoruba: Phonetics Determining Phonology*. Ph.D. dissertation: Cornell University.

Pullum, G. & B. **Scholz**. (2002). 'Empirical assessment of stimulus poverty arguments', in *The Linguistic Review*, 19, pp. 9-50.

Pylkkänen, L. (2002) *Introducing Arguments*. Ph.D. dissertation: MIT.

Rackowski, A. & L. **Travis**. (2000). 'V-initial languages: X or XP Movement and Adverbial Placement', A. Carnie & E. Guilgoyle (eds.), *The Syntax of Verb Initial Languages*, pp. 117-41. Oxford: Oxford University Press.

Radford, A. (2004). *English Syntax: An Introduction*. Cambridge: Cambridge University Press.

Ramchand, G. (2008). *Verb Meaning and the Lexicon: A First-Phase Syntax*. Cambridge: Cambridge University Press.

Ramchand, G. & P. **Svenonius**. (2008). 'Mapping a parochial lexicon onto a Universal Grammar', in T. Biberauer (ed.), *The Limits of Syntactic Variation*, pp. 219-245. Amsterdam: John Benjamins.

Ramchand, G. & P. **Svenonius**. (2014). 'Deriving the functional hierarchy', in *Language Sciences*, 46, pp. 152-74.

Ramus, F., M. **Hauser**, C. **Miller**, D. **Morris** & J. **Mehler**. (2000). 'Language discrimination by human newborns and by cotton-top tamarin monkeys', in *Science*, 288, pp. 349-51.

Rayner, K., M. **Carlson** & L. **Frazier**. (1983). 'The interaction of syntax and semantics during sentence processing: eye movements in the analysis of semantically biased sentences', in *Journal of Verbal Learning and Verbal Behaviour*, 22, pp. 358-74.

Reali, F. & M. **Christiansen**. (2005). 'Uncovering the richness of the stimulus: structure dependence and indirect statistical evidence', in *Cognitive Science*, 29, pp. 1007-28.

Reali, F. & T. **Griffiths**. (2009). 'The evolution of frequency distributions: relating regularization to inductive biases through iterated learning', in *Cognition*, 111, pp. 317-28.

Redford, M. (1999). *An Articulatory Basis for the Syllable*. Ph.D. dissertation: University of Texas.

Regolin, L., L. **Tommasi** & G. **Vallortigara**. (2000). 'Visual perception of biological motion in newly hatched chicks as revealed by an imprinting procedure', in *Animal Cognition*, 3, pp. 53-60.

Reinhart, T. (1976). *The Syntactic Domain of Anaphora*. Ph.D. dissertation: MIT.

Reinhart, T. (1979). 'Syntactic Domains for Semantic Rules', in F. Guenthner & S. Schmidt (eds.), *Formal Semantics and Pragmatics for Natural Languages*, pp. 107-30. Dordrecht: Reidel.

Reinhart, T. (1983). *Anaphora and Semantic Interpretation*. Chicago: University of Chicago Press.

Reinhart, T. (2004). 'The processing cost of reference-set computation: acquisition of stress shift and focus', in *Language Acquisition*, 12, pp. 109-55.

Reinhart, T. (2006). *Interface Strategies*. Cambridge, MA.: MIT Press.

Reiss, C. (2003). 'Language Change with Constraint Ranking', in E. Holt (ed.), *Optimality Theory and Language Change*, pp. 143-67. Dordrecht: Kluwer.

Rey, A., P. **Perruchet** & J. **Fagot**. (2012). 'Centre-embedding structures are a by-product of associative learning and working memory constraints: evidence from baboons (*Papio papio*)', in *Cognition*, 123, pp. 180-4.

Reynolds, G. (1961). 'Attention in the pigeon', in *Journal of the Experimental Analysis of Behavior*, 4, pp. 203-8.

Rezac, M. (2004). *Elements of Cyclic Syntax: Agree and Merge*. Ph.D. dissertation: University of Toronto.

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

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

Richards, M. (2007). 'On feature inheritance: an argument from the Phase Impenetrability Condition', in *Linguistic Inquiry*, 38, pp. 563-72.

Richards, M. (2008a). 'Two Kinds of Variation in a Minimalist System', in F. Heck, G. Müller & J. Trommer (eds.), *Varieties of Competition*, pp. 133-62. Linguistische Arbeits Berichte: Universität Leipzig.

Richards, M. (2008b). 'Quirky Expletives', in R. D'Alessandro, S. Fischer & G. Hrafnbjargarson (eds.), *Agreement Restrictions*, pp. 181-214. Berlin: Walter de Gruyter.

Richards, M. (2009). 'Internal Pair-Merge: the missing mode of movement', in *Catalan Journal of Linguistics*, 8, pp. 55-73.

Richards, M. (2012). 'Probing the Past: On Reconciling Long-Distance Agreement with the PIC', in A. Alexiadou, T. Kiss & G. Müller (eds.), *Local Modelling of Non-Local Dependencies in Syntax*, pp. 135-154. Berlin: Walter de Gruyter.

Richards, N. (1997). *What Moves Where When in Which Language*. Ph.D. dissertation: MIT.

Richards, N. (2001). *Movement in Language: Interactions and Architectures*. Oxford: Oxford University Press.

Richards, N. (2004). 'Against bans on lowering', in *Linguistic Inquiry*, 35, pp. 453-63.

Richards, N. (2011). 'A-Bar Dependencies', in C. Boeckx (ed.), *The Oxford Handbook of Linguistic Minimalism*, pp. 173-94. Oxford: Oxford University Press.

Ridley, M. (2004). *Evolution*. 3rd edition. Oxford: Blackwell.

Ritter, E. (1993). 'Where is gender?', in *Linguistic Inquiry*, 24, pp. 795-803.

Rizzi, L. (1982). *Issues in Italian Syntax*. Dordrecht: Foris.

Rizzi, L. (1986). 'Null objects in Italian and the theory of *pro*', in *Linguistic Inquiry*, 17, pp. 501-58.

Rizzi, L. (1990). *Relativized Minimality*. Cambridge, MA.: MIT Press.

Rizzi, L. (1997). 'The Fine Structure of the Left Periphery', in. L. Haegeman (ed.), *Elements of Grammar: Handbook of Generative Syntax*, pp. 281-304. Dordrecht: Kluwer.

Rizzi, L. (2008). 'On delimiting movement'. Paper presented at Cambridge-Connecticut-Hyderabad-Nanzan-Tsingua-Hua Colloquium (Tsing-Hua, Taiwan).

Rizzi, L. (2011). 'Minimality', in C. Boeckx (ed.), *The Oxford Handbook of Linguistic Minimalism*, pp. 220-38. Oxford: Oxford University Press.

Rizzi, L. (2013). 'Locality', in *Lingua*, 130, pp. 169-86.

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

Roberts, I. (2010a). 'A Deletion Analysis of Null Subjects', in T. Biberauer, A. Holmberg, I. Roberts & M. Sheehan (eds.), *Parametric Variation: Null Subjects in Minimalist Theory*, pp. 58-87. Cambridge: Cambridge University Press.

Roberts, I. (2010b). *Clitics, Incorporation, and Defective Goals*. Cambridge, MA.: MIT Press.

Roberts, I. (2012). 'Macroparameters and Minimalism: A Programme for Comparative Research', in C. Galves, S. Cyrino, R. Lopes, F. Sândalo & J. Avelar (eds.), *Parameter Theory and Linguistic Change*, pp. 320-35. Oxford: Oxford University Press.

Roberts, I. & A. **Holmberg**. (2010). 'Introduction: Parameters in Minimalist Theory', in T. Biberauer, A. Holmberg, I. Roberts & M. Sheehan (eds.), *Parametric Variation: Null Subjects in Minimalist Theory*, pp. 1-57. Cambridge: Cambridge University Press.

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

Roeper, T. (2000) 'Universal bilingualism', in *Bilingualism, Language and Cognition*, 2, pp. 169-85.

Rosen, S. & P. **Howell**. (1981). 'Plucks and bows are not categorically perceived', in *Perception and Psychophysics*, 30, pp. 156-68.

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

Rumelhart, D. & J. **McClelland**. (1986). 'On Learning the Past Tenses of English Verbs', in J. McClelland, D. Rumelhart & The PDP Research Group (eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 2: Psychological and Biological Models*. Cambridge, MA: Bradford Books / MIT Press.

Sadeh-Leicht, O. (2003a). 'Parsing of optional Garden path sentences in Hebrew reflects grammatical competence'. M.A. thesis: Tel Aviv University.

Sadeh-Leicht, O. (2003b). 'Sporadic occurrence of the garden path effect', in W. Heeren, D. Papangeli and E. Vlachou (eds.), *Yearbook 2003*, pp. 59-68. Utrecht Institute of Linguistics OTS.

Sadeh-Leicht, O. (2007). *The Psychological Reality of Grammar: The Theta Principle in Parsing*. Utretcht: LOT Publications.

Saffran, J., R. **Aslin** & E. **Newport**. (1996). 'Statistical learning by 8-month-olds', in *Science*, 274, pp. 1926-8.

Saffran, J., E. **Johnson**, R. **Aslin** & E. **Newport**. (1999). 'Statistical learning of tone sequences by human infants and adults', in *Cognition*, 70, pp. 27-52.

Saito, M. (2007). 'Notes of East Asian argument ellipsis', in *Language Research*, 43, pp. 203-227.

Sakas, W. & J.D. **Fodor**. (2001). 'The Structural Triggers Learner', in S. Bertolo (ed.), *Language Acquisition and Learnability*, pp. 172-233. Cambridge, MA.: Cambridge University Press.

Samuels, B. (2009). 'The Third Factor in phonology', in *Biolinguistics*, 3, pp. 355-82.

Samuels, B. (2011). 'A Minimalist Program for Phonology', in C. Boeckx (ed.), *The Oxford Handbook of Linguistic Minimalism*, pp. 574-94. Oxford: Oxford University Press.

Samuels, B., M. **Hauser** & C. **Boeckx**. (In press). 'Do Animals have Universal Grammar? A Case Study in Phonology', in I. Roberts (ed.), *The Oxford Handbook of Universal Grammar*. Oxford: Oxford University Press.

Sandler, W., I. **Meir**, C. **Padden** & M. **Aronoff**. (2005). 'The emergence of grammar: systematic structure in a new language', in *Proceedings of the National Academy of Sciences, USA*, 102, pp. 2661-5.

Santos, L., M. **Hauser** & E. **Spelke**. (2002). 'Domain-specific Knowledge in Human Children and Nonhuman Primates: Artifacts and Foods', in M. Bekoff, C. Allen & G. Burchardt (eds.), *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition*, pp. 205-215. Cambridge, MA.: Bradford Books / MIT Press.

Santos, L., A. **Nissen** & J. **Ferrugia**. (2006). 'Rhesus monkeys, *Macaca Mulatta*, know what others can and cannot hear', in *Animal Behaviour*, 71, pp. 1175-81.

Săvescu Ciucivara, O. & J. **Wood**. (2013). '*Re*-prefixation and Talmy's parameter', in Y. Fainleib, N. LaCara & Y. Park (eds.), *Proceedings of NELS 41*, pp. 193-204. Amherst: GLSA Publications.

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.

Schütze, C. & E. **Gibson**. (1999). 'Argumenthood and English prepositional phrase attachment', in *Journal of Memory and Language*, 40, pp. 409-31.

Seidenberg, M. & L. **Petitto**. (1979). 'Signing behavior in apes: a critical review', in *Cognition*, 7, pp. 177-215.

Seidenberg, M., & L. **Petitto**. (1987). 'Communication, symbolic communication, and language in child and chimpanzee: Comment on Savage-Rumbaugh, McDonald, Sevcik, Hopkins, and Rupert (1986)', in *Journal of Experimental Psychology*, 116, pp. 279-87.

Semaw, S., P. **Renne**, J. **Harris**, C. **Feibel**, R. **Bernor**, N. **Fesseha** & K. **Mowbray**. (1997). '2.5 million-year-old stone tools from Gona, Ethiopia', in *Nature*, 385, pp. 333-6.

Senghas, A. & M. **Coppola**. (2001). 'Children creating language: how Nicaraguan Sign Language acquired a spatial grammar', in *Psychological Science*, 12, pp. 323-8.

Senghas, A., M. **Coppola**., E. **Newport** & T. **Supalla**. (1997). 'Argument structure in Nicaraguan Sign Language: the emergence of grammatical devices', in E. Hughes, M. Hughes & A. Greenhill (eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development*, pp. 550-61. Somerville: Cascadilla Press.

Sheehan, M. (2013a). 'Explaining the Final-over-Final Constraint: Formal and Functional Approaches', in T. Biberauer & M. Sheehan (eds.), *Theoretical Approaches to Disharmonic Word Orders*, pp. 407-44. Oxford: Oxford University Press.

Sheehan, M. (2013b). 'Some implications of a Copy Theory of labeling', in *Syntax*, 16, pp. 362-96.

Sheehan, M. & W. **Hinzen**. (2011). 'Moving towards the edge', in *Linguistic Analysis*, 37, pp. 405-58. [Page numbers here refer to the version available online at http://ling.auf.net/lingbuzz/001300].

Shettleworth, S. (1998). *Cognition, Evolution, and Behavior*. Oxford: Oxford University Press.

Shettleworth, S. & J. **Sutton**. (2006). 'Do Animals Know What They Know?', in S. Hurley & M. Nudds (eds.), *Rational Animals?*, pp. 235-46. Oxford: Oxford University Press.

Shukla, M., K. **White** & R. **Aslin**. (2011). 'Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants', in *Proceedings of the National Academy of Sciences, USA*, 108, pp. 6038-43.

Sigurðsson, H. (2004a). 'Meaningful Silence, Meaningless Sounds', in P. Pica (ed.), *Language Variation Yearbook*, 4, pp. 235-59.

Sigurðsson, H. (2004b). 'The syntax of Person, Tense, and speech features' in *Italian Journal of Linguistics*, 16, pp. 219-51.

Sigurðsson, H. (2011). 'Conditions on argument drop', in *Linguistic Inquiry*, 42, pp. 267-304.

Sigurðsson, H. & A. **Holmberg**. (2008). 'Icelandic dative intervention: number and person are separate probes'. Unpublished ms.. [As online at http://www.uni-leipzig.de/heck/agree08/sigurdssonholmberg.pdf].

Simpson, A. (*2004). '*The EPP, *Fossilized* Movement and Reanalysis', in E. Fuss & C. Trips (eds.), *Diachronic Clues to Synchronic Grammar,* pp. 161-89*.*

Slocombe, K. & K. **Zuberbühler**. (2005). 'Functionally referential communication in a chimpanzee', in *Current Biology*, 15, pp. 1779-84.

Slocombe, K. & K. **Zuberbühler**. (2006). 'Food-associated calls in chimpanzees: responses to food types or food preferences?, in *Animal Behaviour*, 72, pp. 989-99.

Smith, K. & E. **Wonnacott**. (2010). 'Eliminating unpredictable variation through iterated learning', in *Cognition*, 116, pp. 444-9.

Smith, N. (1999). *Chomsky: Ideas and Ideals*. Cambridge: Cambridge University Press.

Spelke, E. (1998). 'Nativism, empiricism, and the origins of knowledge', in *Infant Behavior & Development*, 2, pp.181-200.

Starke, M. (2002). 'The day syntax ate morphology'. Class taught at the EGG summer school, Novi Sad.

Starke, M. (2006). 'Nanoseminar'. Class taught at the University of Tromsø.

Starke, M. (2011). 'Towards an Elegant Solution to Language Variation: Variation Reduces to the Size of Lexically Stored Trees'. Unpublished ms.: Universitat Autònoma de Barcelona.

Stephens, D. & J. **Krebs**. (1986). *Foraging Theory*. Princeton: Princeton University Press.

Stevens, K. (1972). 'The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data', in E. David & P. Denes (eds.), *Human Communication: A Unified View*, pp. 51-66.

Stevens, K. (1989). 'On the quantal nature of speech', in *Journal of Phonetics*, 17, pp. 3-45.

Stockwell, R. (1977). 'Motivations for Exbraciation in Old English', in C. Li (ed.), *Mechanisms of Syntactic Change*, pp. 291-314. Austin: University of Texas Press.

Stowe, L. (1986). 'Evidence for on-line gap location', in *Language and Cognitive Processes*, 1, pp. 227-45.

Svenonius, P. (2005). '1 … 3-2'. Earlier version of a published chapter (which excludes the relevant sections). [As online at http://ling.auf.net/lingbuzz/000200/v1.pdf].

Swisher, M. (1988). 'Similarities and differences between spoken languages and natural sign languages', in *Applied Linguistics*, 9, pp. 343-56.

Szabolsci, A. (2001). 'The Syntax of Scope', in M. Baltin & C. Collins (eds.), *The Handbook of Contemporary Syntactic Theory*, pp. 607-33. Oxford: Blackwell.

Számado, S. & E. **Szathmary**. (2006). 'Selective scenarios for the emergence of natural langauge', in *Trends in Ecology and Evolution*, 679, pp. 555-61.

Takita, K. (2009). 'If Chinese is head-initial, Japanese cannot be', in *Journal of East Asian Linguistics*, 18, pp. 41-61.

Talmy, L. (1985). 'Lexicalization Patterns: Semantic Structure in Lexical Forms', in T. Shopen (ed.), *Language Typology and Syntactic Description*, pp. 57-149. Cambridge: Cambridge University Press.

Tattersall, I. (2002). *The Monkey in the Mirror*. New York: Harcourt.

Tattersall, I. (2010). 'Human evolution and cognition', in *Theory in Biosciences*, 129, pp. 193-201.

Tattersall, I. (2012). *Masters of the Planet*. New York: Palgrave Macmillan.

Terrace, H., L. **Petitto**, F. **Sanders** & T. **Bever**. (1979). 'Can an ape create a sentence?', in *Science*, 206, pp. 891-90.

Thomas, M. (2002). 'Roger Bacon and Martin Joos: generative linguistics' reading of the past', in *Historiographia Linguistica*, XXIX, pp. 339-78.

Thompson, D. ([1917], 1992). *On Growth and Form*. Reprint of 2nd edition, 1942. Cambridge: Cambridge University Press.

Tincoff, R. & P. **Juscyzk**. (1999). 'Some beginnings of word comprehension in infants', in *Psychological Science*, 10, pp. 172-5.

Tomasello, M., J. **Call** & B. **Hare**. (2003). 'Chimpanzees versus humans: it's not that simple', in *Trends in Cognitive Sciences*, 7, pp. 239-240.

Tomioka, S. (2003). 'The Semantics of Japanese Null Pronouns and its Cross-linguistic Implications', in K. Schwabe & S. Winkler (eds.), *The Interfaces: Deriving and Interpreting Omitted Structures*, pp. 321-40. Amsterdam: John Benjamins.

Trubetzkoy, N. ([1939], 1949). *Principes de Phonologie*, trans. Jean Cantineau. Paris: Librairie C. Klincksieck.

Trueswell, J. (1996). 'The role of lexical frequency in syntactic ambiguity resolution', in *Journal of Memory and Language*, 35, pp. 566-85.

Trueswell, J., M**. Tanenhaus** & S. **Garnsey**. (1994). 'Semantic influences on parsing: use of thematic role information in syntactic disambiguation', in *Journal of Memory and Language*, 33, pp. 285-318.

Turing, A. (1950). 'Computing machinery and intelligence', in *Mind*, 59, pp. 433-60.

Turing, A. (1952). 'The chemical basis of morphogenesis', in *The Philosophical Transactions of the Royal Society of London, Volume B*, 237, pp. 37-72.

Ueda, Y. (2006). 'Ninshoo-seigen-to tougo-kouzou' ['Person restriction and syntactic structure'], in *Scientific Approaches to Language*, 5, pp. 161-80.

Uller, C. (2004). 'Disposition to recognize goals in infant chimpanzees', in *Animal Cognition*, 7, pp. 154-61.

Uriagereka, J. (1998). *Rhyme and Reason: An Introduction to Minimalist Syntax*. Cambridge, MA.: MIT Press.

Uriagereka, J. (1999). 'Multiple Spell-out', in S. Epstein & N. Hornstein (eds.), *Working Minimalism*, pp. 251-82. Cambridge, MA.: MIT Press.

Uriagereka, J. (2002). *Derivations*. London: Routledge.

Wagner, M. & D. **Watson**. (2010). 'Experimental and theoretical advances in prosody: a review', in *Language & Cognitive Processes*, 25, pp. 905-45.

Walter, M. (2007). *Repetition Avoidance in Human Language*. Ph.D. dissertation: MIT.

Wanner, E & H. **Maratsos**. (1978). 'An ATN Approach to Comprehension', in M. Halle, J. Bresnan & G. Miller (eds.), *Linguistic Theory and Psychological Reality*, pp. 119-61. Cambridge, MA.: MIT Press.

Watanabe, A. (2001). '*Wh*-in situ Languages' in M. Baltin, M. & C. Collins (eds.), *The Handbook of Contemporary Syntactic Theory*, pp. 203-25. Oxford: Blackwell.

Watumull, J. (2010). 'Merge as a minimax solution to the optimization problem of generativity'. M.Phil. thesis: University of Cambridge.

Watumull, J. (2012). 'The computability and computational complexity of generativity', in *Cambridge Occasional Papers in Lingistics*, 6, pp. 311-29.

Watumull, J., M. **Hauser** & R. **Berwick**. (2014). 'Conceptual and methodological problems with comparative work on artificial language learning', in *Biolinguistics*, 8, pp. 120-9.

Weinberg, A. (1988). *Locality Principles in Syntax and Parsing*. Ph.D. dissertation: MIT.

Weinberg, A. (1999). 'A Minimalist Theory of Sentence Processing', in S. Epstein & N. Hornstein (eds.), *Working Minimalism*, pp. 283-316. Cambridge, MA.: MIT Press.

Weir, M. (1964). 'Developmental changes in problem-solving strategies', in *Psychological Review*, 71, pp. 473-90.

Wells, J. (1994). 'The Cockneyfication of R.P.?', in G. Melchers & N.-L. Johannesson (eds.), *Non-standard Varieties of Language*, pp. 198-205. Stockholm: Almqvist & Wiksell International.

Werker, J. & C. **Lalonde**. (1988). 'Cross-language speech perception: initial capabilities and developmental change', in *Developmental Psychology*, 24, pp. 672-83.

Werker, J. & R. **Tees**. (1984). 'Phonemic and phonetic factors in adult cross-language speech perception', in *Journal of the Acoustical Society of America*, 75, pp. 1866-78.

Wexler, K. & R. **Manzini**. (1987). 'Parameters and Learnability in Binding Theory', in T. Roeper & E. Williams (eds.), *Parameter Setting*, pp. 41-76. Dordrecht: Reidel.

Whitman, J. (2005). 'Preverbal Elements in Korean and Japanese', in G. Cinque & R. Kayne (eds.), *Handbook of Comparative Syntax*, pp. 880-902. Oxford: Oxford University Press.

Wiltschko, M. (2012). 'Decomposing the Mass/Count Distinction: Evidence from Languages that Lack It', in D. Massam (ed.), *Count and Mass Across Languages*, pp. 120-46. Oxford: Oxford University Press.

Wiltschko, M. (2014). *The Universal Structure of Categories: Towards a Formal Typology*. Cambridge: Cambridge University Press.

Valian, V. (1991). 'Syntactic subjects in the early speech of American and Italian children', in *Cognition*, 40, pp. 21-82.

Van Gelderen, E. (2004). *Grammaticalization as Economy*. Amsterdam: John Benjamins.

Vanhaeren, M. & F. **d'Errico**, (2005). 'Grave goods from the Saint-Germain-la-Rivière burial: evidence for social inequality in the Upper Paleolithic', in *Journal of Anthropological Archaeology*, 24, pp. 117-34.

Van Kemenade, A. (1987). *Syntactic Case and Morphological Case in the History of English*. Dordrecht: Foris.

Van Koppen, M. (2011). 'The directionality of Agree'. Invited paper presented at *BCGL* 6.

Vaux, B. (2008). 'Why the Phonological Component Must be Serial and Rule-based', in B. Vaux & A. Nevins (eds.), *Rules, Constraints, and Phonological Phenomena*, pp. 20-61. Oxford: Oxford University Press.

Vaux, B. (2009). 'The Subset Principle vs. bandwidth maximization in phonological acquisition'. Talk given at the London Phonology Seminar, 22nd April 2009.

Vaux, B. & Y. **Nie**. (2013). 'Learning theory and language change: the evolution of phonological scope'. Talk given at the Cambridge Linguistics Forum, 24th October 2013. Handout online at https://www.academia.edu/4888917/Learning_theory_and_language_change_The_case_of_the_evolution_of_phonological_scope.

Vouloumanos, A. & J. **Werker**. (2007). 'Listening to language at birth: evidence for a bias for speech in neonates', in *Developmental Science*, 10, pp. 159-71.

Yang, C. (1998). 'Derivational c-command and the human parser'. Term paper: MIT.

Yang, C. (1999). 'A selectionist theory of language development', in *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pp. 429-35.

Yang, C. (2000). 'Internal and external forces in language change', in *Language Variation and Change*, 12, pp. 231-50.

Yang, C. (2002). *Knowledge and Learning in Natural Language*. Oxford: Oxford University Press.

Yang, C. (2004). 'Universal grammar, statistics, or both', in *Trends in Cognitive Sciences*, 8, pp. 451-6.

Yang, C. (2008). 'The great number crunch', in *Journal of Linguistics*, 44, pp. 205-28.

Yang, C. (2010). 'Three factors in language variation', in *Lingua*, 120, pp. 1160-77.

Yip, M. (2006). 'The search for phonology in other species', in *Trends in Cognitive* Sciences, 10, pp. 442-5.

Zeevat, H. (1995). 'Idiomatic Blocking and the Elsewhere Principle', in M. Everaert (ed.), *Idioms: Structural and Psychological Perspectives*, pp. 301-16. London: Psychology Press.

Zeijlstra, H. (2012). 'There is only one way to agree', in *The Linguistic Review*, 29, pp. 491-539.

Zwart, J.-W. (1993). *Dutch Syntax: A Minimalist Approach*. Ph.D. dissertation: University of Groningen.