Defeating lexicocentrism

Outline of Elementary Syntactic Structures

Part A

Cedric Boeckx, ICREA/UAB
Cedric.Boeckx@uab.cat

July 2010—*First draft*
October 2010—*Second draft*

Contents

0 Part A: Preamble

The present work represents the draft of the first part of a much longer essay in progress, entitled *Elementary Syntactic Structures*. I have decided to circulate this draft at this point because I think that, although it is still incomplete, it nonetheless gives a clear idea of the direction I want to articulate the issues discussed in the longer work. Moreover, I find myself referring more and more to this work in other publications, and hope that by circulating this draft, readers will be able to see where I am heading. Parts B (dealing with variation), C (dealing with interfaces), and D (dealing with brain issues) will be made available soon.

The complete draft will contain appropriate acknowledgements to the many people who helped me with this research. For now, let me single out my students Dennis Ott, Bridget Samuels, Hiroki Narita, Adriana Fasanella-Seligrat, and Carlos Rubio, as well as Paul Pietroski, Juan Uriagereka, Massimo Piattelli-Palmarini, Noam Chomsky, Ángel Gallego, and Marc Richards, all of whom influenced the present work in ways inadequately expressed in the pages that follow.

1 Approaching UG *in medias res*

The road leading theoretical linguistics beyond explanatory adequacy,[1] that is, towards a naturalistic, biologically-grounded, better-integrated cognitive science of the language faculty, is chock full of obstacles. One can distinguish between external and internal obstacles. External obstacles are very familiar to the student of modern cognitive science. They are the remnant of behaviorist proclivities, the result of our seemingly innate bias towards dualism, and of our traumatic encounter with grammatical prescriptivism, to say nothing of our extrapolations based on our failed attempt to master another language as an adult. All of these factors invariably tend

[1] Chomsky defines "beyond explanatory adequacy" thus: "... we can seek a level of explanation deeper than explanatory adequacy, asking not only *what* the properties of language are but *why* they are that way." (Chomsky, 2004, 105).

Much earlier, Chomsky (1965, 63) defined explanatory adequacy by contrasting it with descriptive adequacy, as follows:

> ... a grammar that aims for descriptive adequacy is concerned to give a correct account of the linguistic intuition of the native speaker; in other words, it is concerned with the output of the device; a linguistic theory that aims for explanatory adequacy is concerned with the internal structure of the device; that is, it aims to provide a principled basis independent of any particular language, for the selection of the descriptively adequate grammar of each language.

to keep us away from the true nature of the language faculty, making us believe that this thing called language is a cultural invention you've got to learn painstakingly, full of non-sensical arbitrariness, nothing like the language next door. Old habits really die hard. Although I think that compelling arguments can be made (and have been made) against overly empiricist, cultural views of language, these views are part of our nature, and one has to be aware of them (and keep them in check) at all times. Internal obstacles are more difficult to deal with, for those are habits that emerged during the practice of linguistics-as-cognitive-science, that were once successful, but that can, at other times, prove an impediment to further progress. In this contribution I want to examine one such factor, which I will refer to as "lexical design" or "lexicocentrism."

Is it really true, as M. Polinsky and I. Caponigro (In press) state matter-of-factly, that "the lexicon of a language is expected to shape its syntax"? Most linguists, I suspect, would endorse this view, certainly those pursuing a standard minimalist approach. This is in fact the view enshrined in most textbooks (see, e.g., Adger (2003), Hornstein et al. (2006)), the view that lies behind such notions as "Last Resort" and "triggered Merge", and that makes it possible to claim that "labels can be eliminated" (Collins (2002)), that "syntax is crash-proof" (Frampton and Gutmann (2002)), and also the one that facilitates the endorsement of the "Borer-Chomsky Conjecture" concerning parametric variation. It is *the* cornestone on which our modern conception of Principles-and-Parameter rests. A quick examination of other frameworks reveals that it is also an aspect of language design that is widely shared. (Accordingly, critics of minimalism, of which there are quite a few, would be ill-advised to use what follows as evidence for the superiority of their own view of syntax.) And, finally, it should be said that *prima facie* the view expressed by Polinsky and Caponigro makes a lot of intuitive sense: if syntax studies how words (equivalently, morphemes/categories) are combined, it's sort of a no-brainer to start with the very units whose combinations we want to discover, isn't it? And yet, I will argue that lexicocentrism is wrong. Deeply wrong. In fact, it may be the biggest (internal) obstacle that lies on the road to "beyond explanatory adequacy".

A common criticism of minimalist syntax is that it simplifies syntax by dumping everything it cannot deal with or does not like onto the external systems with which it interfaces. But I think that minimalist syntacticians commit an even bigger mistake—one that is rarely if ever highlighted (perhaps because it's shared across frameworks and also because it's so deeply intuitive)—by coding virtually everything they should explain as lexical traits, better known as features. Although it is true that minimalist syntacticians relegate a lot of standard syntactic phenomena to post-syntactic components, I do not think that this is necessarily a bad thing, given

that we are finally coming to terms with the fact that these systems have powerful resources (cf. Hauser, Chomsky, and Fitch (2002) on "FLB"). I think that a lot of what makes minimalist analyses unconvincing, and certainly what makes them fall short of going beyond explanatory adequacy is that by the time such analyses begin, all the action has already taken place, as it were. It has been carefully pre-packaged into lexical entries. And once in the lexicon, it's taken for granted. It's not derived, it's not constructed. It is simply assumed as a matter of virtual conceptual necessity. But I take it that Epstein and Seely are right when they say that "if you have not "grown" [i.e., derived, constructed] it, you have not explained it." (Epstein and Seely, 2006, 7). Instead of approaching syntax (and UG) from below, minimalist syntacticians approach it *in medias res*, and as such they do not depart much from pre-minimalist practice.

Newmeyer (2004, 226n.10) is certainly right when he points out that the lexicon is all-important in the minimalist program (MP):

> . . . [I]n no framework ever proposed by Chomsky has the lexicon been so important as it is in the MP. Yet in no framework by Chomsky have the properties of the lexicon been as poorly investigated.

But I do not agree that the lexicon is more important in minimalism than before. It may be more conspicuous, due to the constant appeal to lexical features, but the lexicon has virtually always be as central. The problem is that if minimalism is to genuinely seek to move beyond explanatory adequacy (i.e., if minimalism is to do what makes it worth doing), it will have to explain, as opposed to encode, most of the properties that it now assumes as "given by the lexicon." It will have to break free of a long tradition of linguistic practice.

It is clear that minimalism suffers from featuritis (to borrow a term from computer science that nicely conveys the *ad hoc* character of feature-creation), and often syntacticians hide away all the interesting problems by convincing themselves that (as the saying goes) it's not a bug, it's a feature.[2] These days, we have features for everything: structure-building features/merge-features (a.k.a edge-features), agree-features (a.k.a. unvalued/uninterpretable features), move features (a.k.a. EPP-features), to say nothing of all the fine-grained featural distinctions ('flavors') brought about by the intensive cartographic projects that currently dominates syntactic inquiry. The problem is clear: in the absence of any realistic, grounded, cognitively

[2]Bridget Samuels observes (p.c.) that the problem with features extends beyond the realm of syntax, which I concentrate on here. In the context of semantics, one is reminded of the featural decomposition of concepts that, according to Fodor, got cognitive science on the wrong track. As for phonology, the beginning paragraph of chapter 9 of Chomsky and Halle (1968) says it all. For valuable discussion of the use and abuse of features in phonology, see Samuels (2009, chap.3).

sound, biologically plausible theory of what's a possible feature, and a possible lexical entry (read: feature-bundle, or category), it is too easy to come up with a feature that will do the job. But it should be clear that features and the way we manipulate them syntactically are the problem, not the solution. It's where the investigation should end, not where it should start.

Our blind reliance on the lexicon has had serious detrimental effects. This is nowhere as clear as in the context of the logical problem of language acquisition ("Plato's problem"). Consider the following quotes:

> "Parametric variation is restricted to the lexicon" (Chomsky, 2001, 1)
> "The availability of variation [is restricted] to the possibilities which are offered by one single component: the inflectional component [of the lexicon]" (Borer, 1984, 3)
> "A parameter is an instruction for a certain syntactic action expressed as a feature on a lexical item and made operative when the lexical item enters syntax as a head. ... In this conception, the size of the set of parameters is not determined by the number of principles, but by the size of the (functional) lexicon." (Rizzi, 2010)

Although such statements give the appearance of a very restrictive theory of language variation, in the absence of a theory of "lexical item", "the (functional) lexicon", or its "inflectional component", they amount to little more than statements of ignorance, or wishful thinking.[3] This feeling is reinforced when we consider the fact that virtually throughout the generative period, the lexicon has been taken to be "really an appendix of the grammar, a list of basic irregularities." (a conception already expressed in Chomsky (1965), and reiterated in Chomsky (1995); a conception ultimately going back to Bloomfield.). If that is the lexicon, surely we cannot claim to have understood the nature of variation.

But the gap between our understanding of the lexicon and the intensive use we make of it is by no means limited to the domain of parametric variation. It is equally damaging on the other side of the 'Principles-and-Parameters' model, in the domain of principles. As I pointed out in above, all syntactic operations are currently assumed to be feature-driven. That is to say, as Epstein (2003[2007], 43) has correctly pointed out, that "the most fundamental operation ceases to be structure-building (Merge) and becomes structure-licensing (Check/Agree/Value)." Epstein's

[3]Consider the notion of "head" in Rizzi's statement. Why do we regard bundles of features as a (minimal) head, and other collections of features as a (maximal) phrase? Where is the dividing line?

statement makes it clear that contrary to the rhetoric often used in minimalist circles, it is not true that "all there is, is Merge." In fact, Merge is subordinated to the lexicon. Short of features triggering it, Merge (Agree, etc.) can't apply. No wonder Pinker and Jackendoff (2005) criticized Hauser, Chomsky, and Fitch (2002) for being silent on the evolution of the lexicon. Pinker and Jackendoff had primarily the size of the human lexicon in mind as a fundamental attribute of the language faculty. But minimalist practice, if taken seriously, indicates than, more than the size, it is the nature of the features making up the lexicon that was the key evolutionary event (I return to this at length below). But, just like in the context of parametric variation, no hypothesis concerning how the lexicon came about is forthcoming. Worse, the very little we know about the lexicon, as currently used, makes its emergence as the key evolutionary event widely implausible (to put it mildly).

Perhaps the clearest example of how much the lexical properties dictate syntactic behavior is to be found in Adger (2010). Adger (2010) (alongside Adger and Svenonius (In press)) is the first and most explicit discussion of the nature of lexical entries (/pre-syntactic feature bundles) within minimalism that I know of. The paper opens with a challenge. Its goal is to "explore the consequences of the idea that structure embedding in human language is only ever syntactic (that is, that there is a single engine for the generation of structure and the engine is the syntax" (an idea familiar in recent works by Marantz, and Borer, and directly related to the proposal in Hauser, Chomsky, and Fitch (2002))". As Adger correctly observes "if structure embedding is only syntactic, then the feature structures that are the basic atoms of syntax (i.e., lexical items) cannot involve embedding of one feature inside another." In so doing, Adger notes, this minimalist approach "contrasts rather starkly with work in other approaches which take lexical items to have rich featural structure" (all feature-unification frameworks, such as HPSG, and LFG). A more accurate statement would be 'in so doing, this minimalist approach *would* contrast rather starkly with other frameworks, for, as a matter of fact, Adger ends up being forced to propose a fair amount of embedding inside his minimalist lexical entries. Consider his final proposal concerning the shape of feature bundles:

> a. A lexical item is a set $\{K, F_1, \ldots, F_n\}$,
> (i) K is an ordered pair ⟨Cat, N⟩
> (ii) Cat is drawn from the set [known as the functional sequence or Cinque-hierarchy–CB]{C, D, T, Num, Asp, ...}–∅, and N is drawn from the set of natural numbers above 0
> b. F_1 is a pair ⟨Att, Val⟩, where Att is drawn from a finite set of M[orpho]S[yntactic]-features and Val from a finite set of values [say, +/–]
> c. Hierarchies of Projections: these are sequences of Ks whose second

member is ordered by the relation $\prec$ [see, e.g., Cinque (1999)].

The above would, for example, correspond to the following concrete lexical entry:

(1) *was* = {T, #8 [on the functional sequence], tense:past, pers.:3, num:sing, case:nom}

I am sure that the reader recognizes a fair amount of structuring here: ordered pairs are nothing other than embedded structures ($\langle\alpha,\beta\rangle = \{\alpha,\{\alpha,\beta\}\}$). There is a clear sense in which an attribute stands in a subset/superset (i.e., embedded) relation. In his paper, Adger is silent about the EPP/strength property of some features. If this property is thought of in the standard way (e.g., Pesetsky and Torrego (2001), Carstens (2005)), as a feature of a feature, it will require a further enrichment of lexical entries; i.e., further embedding.

The reason for this irreducible lexico-structural complexity is clear: Adger takes as its starting point the view that "much like categorial grammars, minimalist grammars can be seen as lexically driven combinatory systems" (a point of view well articulated in his textbook; Adger (2003)).[4]

Adger in fact recognizes the limits of his own proposal. In the concluding section of his paper, he writes: "I should stress that the theory is certainly not as 'minimal' as one might like . . . " But I think that Adger's theory of feature structure is probably as minimal as it could be given the set of assumptions he adopts—in particular his claim that it is "features [that] drive the various syntactic operations. We can think about this in an essentially Fregean way: the features have some property that needs to be satisfied." As an example, consider the fact that Adger takes Merge, as he does in Adger (2003), to be licensed by a matching of categorial features on the selector and the selectee, where the feature of selector is unvalued. Accordingly, for lexical entries of lexical categories, one cannot get around distinguishing between a valued categorial feature, and an unvalued categorial feature.

The outline of my general conclusion should already be clear at this point: the best-case scenario for minimalism (i.e., for an approach that seeks to approach UG from below) is for there to be no need for feature-bundles driving syntax, for bundles are structures, and structures demand syntax. Pre-syntactic lexical entries must be flat. The objects syntax manipulates must be atomic.[5] It should be clear that for

[4] Adger claims that "minimalism is unlike other 'lexicalist' theories in that almost all the interesting structure is syntactic," but then quickly adds "although the information which leads to the building of that structure is entirely lexical", thus recognizing the lexicocentric character of minimalist syntax.

[5] This is the view expressed in Marantz (1996, 3): "The syntax starts with simple, non-branching constitutents." A view that is, unfortunately, never really put into practice, even in post-syntactic,

this to be the case, Merge must be trigger-free, Agree as a syntactic process must be rethought, functional hierarchies must be emergent. Various other properties must be contextual/configurational, and some even outsourced completely. If all of this obtains (yes, I know, it's a big IF), then the lexical input to Narrow Syntax can be kept to a bare minimum, and Adger's challenge (essentially a biolinguistic challenge, as it ultimately relies on evolutionary plausibility) against embedding inside lexical items can be met. I will call this approach "Merge α", for, as many will probably recognize, it is reminiscent of the spirit behind the "Move α/Affect α" that dominated Government-and-Binding approaches (see Chomsky (1977), Chomsky and Lasnik (1977), Lasnik and Saito (1992)).

What this amounts to is a call for a Gestalt shift regarding syntax and its relation with the lexicon. It will also require a re-conception of the relation between syntax and grammar, as we will see. The key issue was already identified by Adger in the paper I have been quoting from:

> When two items merge, is there some constraint on what they can/must be? Empirically, we clearly want to say that there is, to distinguish, for example, *the many men* from **many the men*, so the question is how this kind of distributional constraint is modelled theoretically (p. 15)

Adger is right, constraints of this sort abound in natural languages. But how strong is the argument for a (lexico-)syntactic treatment of these constraints? The unacceptability of *the many men* is a fact about a particular state of the language faculty ("English"). Even if universal, the constraint need not necessarily be viewed as a fact about the faculty of language in the narrow sense. It could be a fact about the faculty of language in the board sense, the result of many components, an interaction effect. It is true that constraints of this sort have been at the heart of cartographic projects (witness the adverb ordering restrictions discussed in Cinque (1999), or the facts discussed in Rizzi (1997)), where they are reduced to lexical (c-selectional) restrictions. But this type of reduction does not mean explanation. Saying that it is *the*P that c-commands *many*P (or DP, QP), and not the other way around, is not very different from saying that *the many men* is ok but *many the men* is not. Fortuny (2008, 112f.) is certainly correct when he writes:

> [T]he empirical results of the cartographic project do not lead us to a

realizational, exo-skeletal approaches such as Distributed Morphology or Borer's (2005) parallel morphology. As Marantz (1996, 1) points out, "The lexicon as input to the computational system is always considered generative in some sense. Even within DM, ...". Nanosyntacticians also rely heavily on an organized pre-syntactic lexicon, as they crucially assume a language-specific functional sequence.

> primitive element of the syntactic component, but rather to the study of the C-I system[s] [...] [These ordering restrictions are] clearly relative to the levels of interpretation of the C-I system[s], and hence one may be skeptical about coding them in the theory of grammar in the form of derivational devices or in the form of universal hierarchies [...] The theory of the syntactic component of the faculty of language need not — and therefore must not — encode devices that translate the kind of external requirements to be satisfied: the syntactic component does not have to be defined to avoid the generation of [structures violating vacuous quantification]; similarly, if cartographies derive from the Full Interpretation condition, cartographies do not reflect the knowledge of grammar [I would prefer the term 'syntax' here instead of 'grammar' —CB].

In the words of Harbour (2006): "Combinatorial restrictions have to be explained, not reified." To drive the point home, let me illustrate this by means of the following paradigm.

(2) a. They have a ton of money
b. They have tons of money
c. *They have seven tons of money

Watanabe (2009) seeks to explain the badness of the (c) example, and in so doing commits what I like to call "the grammarian's fallacy." His explanation is cartographic in nature, as he accounts for the fact that the presence of a numeral is incompatible with the idiomatic "very large [vague] quantity" by saying that "Vague quantity expressions are merged in the Spec of QP, which is above #P." But this structural generalization, if true, is not what accounts for the badness of the example in question in a deep way (in a way that goes beyond explanatory adequacy). It is what is in need of explanation (why exactly is QP above #P?). I wager that such an explanation lies in the fact that lexical items serve as instructions to activate certain concepts, which reside in dedicated (core) knowledge modules, and that activating "very large vague quantity" (appealing to one of the component of our number sense (Dehaene (1997), Boeckx (2009b), viz. large approximate quantity estimation) by means of *tons (of)* clashes with the instruction to activate a different component of our number sense, the one dealing with precise quantities, by means of numerals like *seven*.

Fortuny (2008), who I quoted above, suggests an approach along similar lines to account for various ordering restrictions observed by Cinque. Although a lot of work

remains to be done, this kind of interface-oriented approach has the dual benefit of diminishing the appeal to pre-syntactic lexical design (lexicocentrism) and of forcing linguists to pay attention to properties of other cognitive modules, thereby enhancing the quality of interdisciplinarity that is badly needed to advance the biolinguistic program. Short of that, it is hard to disagree with Koster's (2010) "dissatisfaction with current Minimalism":

> My concerns are not about Minimalism as a program. On the contrary, I subscribe to the overall goal to construct a theory that makes grammar look as perfect as possible and that relegates as much as it can to "third factor" principles. My dissatisfaction is about how this program is carried out in practice. Others disagree, but my personal feeling is that little *theoretical* progress has been made since the 1980s. I emphasize *theoretical*, because empirically speaking the progress has been impressive. One can hardly think of any topic nowadays of which it cannot be said that there is a wealth of literature about it. All of this progress, I claim, is mainly "cartographic" and therefore compatible with pre-minimalist generative grammar and even certain forms of pre-generative structuralism. Part of the theoretical stagnation is due to the fact that some key problems of earlier versions of generative grammar, as they arose for instance in the GB-period, are either unresolved or ignored. But there are deeper problems, it seems, that involve the very foundations of the field.

The term 'biolinguistics' has now appeared a few times in the preceding paragraphs, and I think it's time to make explicit that my attempt to defeat lexical design/lexicocentrism is ultimately motivated by biological plausibility. I know (as Dick Lewontin never tires of telling me) that plausibility is not science, but it can be a good source of scientific motivation. It can serve as a good intuition pump.

A growing number of passages such as those that follow can now be found in the linguistics literature.

> How much should we ask of Universal Grammar? Not too little, for there must be a place for our unique ability to acquire a language along with its intricacies and curiosities. But asking for too much won't do either. A theory of Universal Grammar is a statement of human biology, and one needs to be mindful of the limited structural modication that would have been plausible under the extremely brief history of Homo sapiens evolution. (Yang, 2010, 1160)

> Given this, consider a second fact about F[aculty of]L[anguage]: it is of recent evolutionary vintage. A common assumption is that language

> arose in humans in roughly the last 50,000-100,000 years. This is very rapid in evolutionary terms. It suggests the following picture: FL is the product of (at most) one (or two) evolutionary innovations which, when combined with the cognitive resources available before the changes that led to language, delivers FL. This picture, in turn, prompts the following research program: to describe the pre-linguistic cognitive structures that yield UG's distinctive properties when combined with the one (or two) specifically linguistic features of FL. ... The short time scale suggests that the linguistic specificity of FL as envisaged by GB must be a mirage. FL must be the combination of operations and principles scavenged from cognition and computation in general with possibly small adventitious additions. In other words, despite appearances, FL is "almost" the application of general cognitive mechanisms to the problem of language. The "almost" signals the one or two innovations that the 50,000–100,000 year time frame permits. The minimalist hypothesis is that FL is what one gets after adding just a little bit, a new circuit or two, to general principles of cognition and computation. (Hornstein, 2009, 4)

These passages express well the point of view at the heart of Hauser, Chomsky, and Fitch (2002), who, more than anything, want to draw attention to the richness of the Faculty of Language in the Broad Sense, and the many difficulties of assuming a high degree of linguistic specificity (a rich Faculty of Language in the Narrow Sense). This is the point where (as Jackendoff and Pinker correctly pointed out) the Hauser-Chomsky-and-Fitch vision meets minimalism, which takes a deflationist stance on Universal Grammar ("approaching it from below"):

> At the time [pretty much throughout the history of generative grammar], it seemed that FL must be rich, highly structured, and substantially unique. ... Throughout the modern history of generative grammar, the problem of determining the character of FL has been approached "from top down": How much must be attributed to UG to account for language acquisition? The M[inimalist] P[rogram] seeks to approach the problem "from bottom up": How little can be attributed to UG while still accounting for the variety of I-languages attained? (Chomsky, 2007, 2, 4)

The clearest expression of this "rich, highly structured, and substantially unique" hypothesis regarding UG, its apogee, was the modular view at the heart of Government-and-Binding (Chomsky (1981)). There, UG was assumed to consist of a variety of internal modules, of many distinct components that interact in complex ways. (This

view remains at the heart of parametric models such as Baker's (2001) parameter hierarchy.) The problem with such a view is that this amount of internal modularity, and its complex organization can, biologically speaking, only have one source: natural selection. Everyone agrees on this much, from ultra-darwinians (Dawkins, Pinker) to wiser heads (Gould, Fodor). Fodor (1998b) puts it best when he writes:

> If the mind is mostly a collection of innate modules, then pretty clearly it must have evolved gradually, under selection pressure. That's because . . . modules contain lots of specialized information about problem-domains that they compute in. And it really would be a miracle if all those details got into brains via a relatively small, fortuitous alteration of the neurology. To put it the other way around, if adaptationism isn't true in psychology, it must be that what makes our minds so clever is something pretty general . . .

So, on the assumption that complexity requires natural selection and that natural selection requires time to work its magic (everyone's best guess, from Darwin onwards, is that it requires lots of it: say on the order of (at least) millions of years), the rapid rise of language in humans does not allow for this kind of complexity to develop. It is true that one may question the assumption that complexification requires a lot of time. After all, complex systems theorists demonstrate on a daily basis that complexity can emerge rapidly, but it is important here to distinguish between the kind of complexity that they study (which is complexity that emerges from simplicity, and is always generic in character) and the sort of complexity that we are talking about (module-internal, highly-domain-specific complexity). Of the latter, we only have good examples that required a lot of time to come about.

One could, of course, also question the assumption that the human language faculty of extremely recent vintage. If, contrary to what most scientist think, the language faculty has a long history, the internal complexity/modularity ascribed to it in previous models, may have had enough time to emerge. So, how good is the evidence for this recent emergence?

There is no proof, of course, but it's a fairly good conjecture. In fact, I tend to think that the evidence is now stronger than what many suspect. It's now generally agreed upon, based on converging evidence, that the guess we are talking about is as good as any: the emergence of new tools, cultural artifacts, signs of trade, cave paintings, and so on, that we find in the archeological record, first in Africa and then in Europe (McBrearty and Brooks, 2000, Mellars et al., 2007) point to a significant evolutionary transition. I tend to agree with Diamond (1992), Tattersall (1998), and many others that it is hard to imagine the emergence of these artifacts and signs of

modern human behavior in the absence of the language faculty. But, as I pointed out in Boeckx (In pressb), we can now make an even stronger case for the suddenness of the emergence of the language faculty. Recent genetic data suggest that Homo Sapiens split into two sub-populations around 150,000 years ago, which remained separated for about 100,000 years (Behar et al. (2008)). If this interpretation of the data is correct, it suggests that the language faculty was already in place 150,000 years ago. Now, if we combine this with the *FOXP2* evidence that suggests something significant happened, at the latest, 200,000 years ago (Piattelli-Palmarini and Uriagereka (2005)), the time of appearance of *homo sapiens*, and you get a window of time of just about 50,000 years. So all in all, I think we should really try to keep the evolutionary add-ons for language to a bare minimum, with virtually no internal modularization to speak of.

I am not the first to make an argument against language-internal modules in the context of minimalism. Hornstein (2001), I think, was the first to see this issue clearly. The real target of minimalism is not government, Spec-head, or whatever. It should be language-internal modules. But although Hornstein expressed this admirably, it seems to me that he has not seen that this means we should reduce our appeal to lexical features as much as possible. Curiously, Hornstein (2001, 215–6) addresses this issue briefly, and says that replacing modules by features (say, a Θ-module by θ-features requiring checking) is an advance. I think he is wrong, for features are nothing more than nano-modules. Features have the very same degree of language-internal specificity and informational encapsulation as modules. All the talk about modules listed in Chomsky (1981) may have disappeared from the minimalist literature, but modules themselves live on disguised as features. It is easy to find featural counterparts to all the modules postulated in *GB*, which, to me, illustrates that much of minimalism today is really still GB in disguise. If, as I am trying to argue, going beyond explanatory adequacy means going beyond features/modularity, we have not begun to explore the space of possibilities afforded by minimalism. It is time for minimalism to evolve away from its precursors and become something new and distinctive on the theoretical linguistic scene. It is (high) time we take minimalism for what it is, a program to explore, not a theory to stick to, especially if it's one that assumes all the answers instead of asking the questions, as all lexicocentric approaches invariably do.

Of course, even programs have to be operationalized if one is to explore them. And at the beginning of minimalism, features did just that. They made it possible to identify properties of the computational system such as Last Resort and other economy principles. But appealing to case-features and Greed to explain the contrast between (3) and (4) is one thing. Appealing to EPP-features to account for the

EPP effect, or to scrambling features to account for scrambling, is, I'm sure everyone will admit, another.

(3) a. John seems [t to have left]
b. It seems [John has left]

(4) a. *seems [John to have left]
b. *John seems [t has left]

From the use of features, minimalists quickly moved to the abuse of features. Cartographic representations (even fine-grained ones such as those assumed by proponents of nano-syntax; see Starke (2010)) encode notions that cannot possibly be taken as primitives: "ProcessP" or "EvaluativeMoodP" are almost certainly the output of syntax-dependent interpretive processes, not pre-syntactic primitives. They are the product of an interaction among various cognitive components, not the cause of syntactic operations. In this sense, features in syntax have gone the way of genes in biology. From a plausibility argument to deal with certain facts ((3), (4) in the case of features, Mendelian data in the case of genes), they have become all powerful to the point of relegating syntax and the organism to the level of a mere vehicle of lexical/genetic expression—the selfish gene and (as one might call it) the selfish lexeme view. And just like geneticists hope to understand the organism by sequencing genomes, cartographists hope to understand syntax (and the syntax-semantics interface) by examining functional (lexical) sequences. Both clearly suffer from an unacceptable degree of preformationism and Platonism. It's curious that linguists who have written about the evolution of language are quick to identify the severe limitations of the selfish-gene-world view, but are not so quick to identify the very same limitations in their own linguistic practice. Shouldn't they take out the beam of wood from their own eyes before removing the speck that they see in their neighbor's?

Just like geno-centrism led to the demise of the organism and development, lexicocentrism is leading to the end of syntax, to models where "syntactic structures are not generated by lexicon-independent rules ... but as the spelling out of the contextual properties of lexical items ("valency")" (Jan Koster[6]), models where phrase structure (i.e. syntax) cease to exist, as in Collins and Ura (2001), who explicitly advocate a Word Grammar-style grammatical model, where lexical valency is the most important component (see also Carnie (2008)).

The alternative, as is now clear in biology with the revival of embryology under

[6] `http://odur.let.rug.nl/~koster/resume.htm`

the rubric of "Evo-Devo",[7] the emphasis on developmental and phenotypic plasticity, epigenetics, and the emergence of theories such as niche construction, is one that stresses organismic processes as opposed to genetic blueprints or programs, interactions (the interactome as opposed to the genome; the triple helix as opposed to the double helix). As such it seems tailor-made for minimalist explorations, especially once these discard lexical blueprints or programs (i.e., numerations), and truly explore interface-based explanations. Much like the emerging expanded synthesis in biology, linguistics will have to embrace pluralism, get rid of isolationist (i.e., modular, self-sufficient) tendencies, and revisit the works of old foes to treat them as friends.[8] Much like modern biology, modern linguistics will have to soften its stance of various issues, especially those touching on specificity and innateness (Massimo Piattelli-Palmirini would talk about this in terms of leaving behind the (necessary) age of specificity; see Piattelli-Palmarini (2010)). The range of processes explored are likely to be more abstract (less-task-dependent) and generic. My own view is that this is the only way linguists have to take on Poeppel's challenge.[9]

Like many others, Poeppel would like to relate mind and brain somehow, and like many he is not impressed with how far we've gotten until now. It's not just the linguists' fault, of course. But I tend to think it's in part our fault. It is fair to say that GB is cognitively exceptional in that its principles and operations are cognitively *sui generis* and very specific to language. Eric Reuland once said that GB

[7]Being a program, "Evo-Devo" is pursued along many lines, some of which are still heavily gene-centered. The present essay is more in line with more epigenetic pursuits within Evo-Devo. For discussion, see Boeckx (2010e), Benítez-Burraco and Longa (To appear).

[8]For linguistics, this means taking into account the fact that many works of biologists of which generative grammarians approve are closely associated with psychological models that generative grammarians have prematurely rejected *in toto*. Witness the influence of Waddington on Piaget, or the influence of McCulloch and connectionism on Stuart Kauffman. For more on this point, see Boeckx (2010h).

[9]Fitch (2009, 298) formulates a similar challenge in his prolegomena to a science of 'biolinguistics':

> We need to distill what we know from linguistic theory into a set of computational primitives, and try to link them with models and specific principles of neural computation ...
>
> Thus we need linguistic models that are explicit about the computational primitives (structures and operations) they require, and that attempt to define linguistic problems at a fine enough grain that one can discuss algorithmic and implementational approaches to their solution. We need a list of computations that linguistic theorists deem indispensable to solve their particular problem (e.g., in phonology, syntax, or semantics).

principles are too good to be false, but too (domain-)specific to be true. As Poeppel and Embick (2005) observe, this is a serious problem for those aiming to find brain correlates for the primitives of FL. They dub this the granularity problem. The aim is to find those primitive operations that are at once empirically grounded and that could be embodied in neural wet-ware. Given this, the goal for the linguist will be to find a class of very basic primitive operations that plausibly underlie linguistic computations for consideration as candidates for possible neural circuits.

As Poeppel notes elsewhere (Poeppel (2005)), if anything is localized in nervous tissue, at least at the level of cellular ensembles or columns, it will be elementary computational functions. Poeppel goes on to say (that's his challenge to the linguistic community),

> Linguists and psycholinguists owe a decomposition (or fractionation) of the particular linguistic domain in question (e.g. syntax) into formal operations that are, ideally, elemental and generic. The types of computations one might entertain, for example, include concatenation, comparison, or recursion. Generic formal operations at this level of abstraction can form the basis for more complex linguistic representation and computation.

I think it's a great challenge. And it's for everybody. But I think minimalists in particular should take this challenge to heart because, as I will try to show below, by their emphasis (in theory, if not yet in practice) on approaching UG from below, they are better equipped than most to do something about it. So, they would be missing a great opportunity if they didn't try. I have argued elsewhere (Boeckx (2006, 2010a, 2009b), Di Sciullo and Boeckx (In press)) that minimalism has contributed substantially to the re-emergence of interdisciplinarity in linguistics and thus to the return of biolinguistic discourse. The minimalist program forces linguists to reformulate previous findings in terms of elementary units, operations, and interface conditions; those that, to our very best bets, have the character of conceptual necessity (those that are, in Fitch's term (quoted above), "indispensable"). Many of these, a few of us anticipate, will have such a generic flavor to them (combine, map onto a linear sequence, etc.) that they are plausibly not specific to the language faculty. This should be very good news to researchers in other areas, as the concepts articulated by minimalists may find an equivalent in their own field, or be more readily testable using familiar techniques, something that modern ethologists like Marc Hauser and Tecumseh Fitch fervently desire. At the same time, some of these generic operations will make it more plausible to entertain 'descent with modification' scenarios concerning the evolution of language, and also, I suspect, reconnect with experts on

psycholinguistics and language acquisition, who currently view our isolationist (read: modular) stance as counterproductive.

Some aspects of lexicocentrism have been pointed out by Distributed Morphologists like Alec Marantz, or advocates of exoskeletal approaches to word structure like Hagit Borer, who correctly implicate syntax more. Nevertheless there is a residue of generativity in their pre-syntactic lexicon (as Marantz (1996) acknowledges, cf. footnote 5), as they do not seek to derive (syntactically) the functional lexemes (and the hierarchies they form) that they crucially rely on to deflate the content of lexical categories.

Ray Jackendoff, too, has identified part of the problem with lexicocentrism in various publications (Jackendoff (1997, 2002, 2005, 2010)), by drawing attention to the constant projectionist models of syntax, and also, like Marantz and Borer, to the phrasal nature of much of our lexical knowledge (what Construction Grammarians call "Constructions"), but I suspect that he too retains a lexico-centric view of syntax. Unfortunately, Jackendoff is not explicit at all here. He (in my view, correctly) takes language-specific lexical items to be mappings between different representations they are (small) interface rules ("correspondence rules"; see already Marantz (1984)), but nowhere (as far as I know) does Jackendoff tell us how these different representations come about. Jackendoff claims that the key grammatical process is Unify (as opposed to Merge), but in order for Unify to apply, different representations must be constructed. Here too, we begin *in medias res*. In the context of Jackendoff's "Simpler Syntax" model, this is an important point, for the antidote to lexicocentrism is more syntactocentrism, not less (just the opposite of what Jackendoff claims): if one wishes to explain, one should be a Constructing Grammarian, not a Construction Grammarian. One should adopt exo-lexicalism *in toto*, that is, do away with any remnant of lexicalism. Marantz (1995) is wrong, minimalism (seen as the attempt to go beyond explanatory adequacy) is not the end of syntax, it's the end of the all-powerful lexicon.

The solution is not to let syntax blindly follow lexical instructions, but rather to let syntax construct the lexicon, for it is only by constructing things that one can hope to explain them. The Inclusiveness guideline suggested by Chomsky (1995) was a way to prevent the reification of relations in featural terms (e.g., indexing in the context of binding relations). This was not meant to be a ruling against emergent phenomena in syntax. Getting out as much (in syntax) as one puts in (the lexicon) simply begs the question of what syntax is for.[10]

Let me conclude by stressing that lexicocentrism really is endemic, and threat-

[10]For a very good example of lexical overspecification, consider the treatment of locality in Müller (2010), where lexical entries consist of not one, but multiple structured stacks of features.

ens the minimalist goal of moving beyond explanatory adequacy, or indeed the very project of "biolinguistics." The more we rely on the lexicon, the more inescapable Koster's conclusion that linguistics is at best applied biology appears to be. As Koster notes,[11]

> In his [Koster's] overall theory of language [which resembles the mainstream more than he claims—CB], the capacity for language, even in its narrowest sense, is not seen as a matter of biology but as applied biology, i.e., a technology belonging not primarily to individuals but to their shared culture. Invented words rather than syntax are at the essence of language in this view, while recursive syntax is seen as a successful extension of the properties of the cultural objects in question ("words"). The combinatorial potential of words is as cultural and non-individual as the words it belongs to and therefore first and foremost public property that individuals seek to adopt from the day they are born into a community.

If one wants to avoid this conclusion, as I think we should, we must try very hard to limit the scope of lexical supervision, and seek a level of representation of syntax that is appropriate for biolinguistic inquiry. The following is an attempt in this direction.

2 A metaphor not to live by

By any measure the mental organ called the language faculty is very complex. Moreover, as Hauser, Chomsky, and Fitch (2002) emphasized, our capacity to develop (internalize/cognize) a language is far from monolithic, consisting as it almost surely does,[12] of many ingredients, many of which of distinct — and non-linguistic — origin. It is because of this complexity that Chomsky, beginning with his *Logical Structure of Linguistic Theory* in 1955, insisted on distinguishing various levels of 'representation' (equivalently, 'structure') to be studied (as a first pass) in relative isolation. These levels correspond to traditional subdisciplines within linguistics: phonology, syntax, semantics, morphology, etc. To this day, the approach guides theoretical linguistic practice, as well as the practice in neurolinguistics, where researchers attempt to locate these various levels in brain wet-ware. By focusing on the concept of interface, minimalism has turned what everyone passively acknowledged (that these various levels must function in tandem) into an active area of research, to the point

[11] Taken from `http://odur.let.rug.nl/~koster/resume.htm`; see also Koster (2009).

[12] I view this recognition of a-monolithicity as a biological imperative. Denying this would make the language organ unlike everything else we know about ow biological structures come about.

that the autonomous status of some levels is being questioned (witness Marantz's (1995) "end of syntax" pronouncement). Certainly, the distributed character that some levels have acquired (as is for example the case with morphology) is indicative of this shift in perspective, which promises to be more fruitful in the context of mind/brain unification, given the generally distributed character of higher-level brain processes. What needs to happen now is linking this distributed character to the Hauser-Chomsky-Fitch question of which aspect (if any) is specifically linguistic. Marcus's (2006) descent-with-modification approach to modules is a step in this direction. As a matter of fact, if, as I have claimed above, features are nothing more but nano-molules, a descent-for-modification is called for in this domain as well. This will inevitably entail a softening of the 'specificity' stance that is generally adopted in generative circles. Almost certainly, what are now considered highly specific linguistic properties will reduce to a conjunction of various non-linguistic, generic factors—ultimately, the recognition of a vastly underspecified linguistic code, and of the importance of emergence.

The Gestalt-shift just mentioned will be felt most clearly in the context of syntax, traditionally the level of representation taken to be most specific and unique (for good reasons, see Anderson (2004) for a good overview based on comparative ethology). It is in the domain of syntax that an "approach from below" is most urgently needed, one that genuinely begins the analysis from the ground up.

As Dick Lewontin has often pointed out (the following is taken from Lewontin (2000, 3)),

> It is not possible to do the work of science without using a language that is filled with metaphors. ... While we cannot dispense with metaphors in thinking about nature, there is a great risk of confusing the metaphor with the thing of real interest. ... The result is that the properties we ascribe to our object of interest and the questions we ask about it reinforce the original metaphorical image and we miss the aspects of the system that do not fit the metaphorical approximation.

In this section I would like to cast doubt on a metaphorical image that continues to dominate syntactic studies (and, in my view, vitiates them), and suggest that the metaphor that it is usually contrasted with may, in fact, be more conducive to progress (in the direction of 'beyond explanatory adequacy').

Consider now the following passage, quoted in full, from the beginning of Baker (1999).

> Why are there so many languages in the world—on the order of 5000, not counting many others that have become extinct in historical times?

This question would be fairly easy to answer if languages differed only in relatively superficial ways, such as which set of sounds is conventionally used to refer to a particular kind of concrete objects, or the details of how that set of sounds is produced phonetically. The word for any given is clearly arbitrary in most cases, so one choice is as good as another, and speakers in different parts of the world can be expected to make different choices. Similarly, there are different ways of accommodating sounds to one another in the phonology and phonetics that decrease effort while maintaining contrasts: again it is not surprising that languages make different choices in these matters. However, linguistic diversity goes far beyond these local and superficial matters. This is harder to explain in a priori terms, and it places an important boundary condition on the task of constructing an adequate linguistic theory.

In this paper, I present and defend the view that one important reason why there are so many different languages in the world is because all human beings are equipped with a detailed and rigid Universal Grammar. At first, this sounds paradoxical to most people. Indeed, the discoveries of linguistic typological research are often taken to tell against Chomskian notions of Universal Grammar. However, a strong case can be made that these facts show just the opposite.

To appreciate the logic of the situation, consider whether human languages are more like piles of sand or Swiss watches. Many people think of languages as loose connections of many more or less unconnected words, morphemes, inflections, and syntactic constructions. These various pieces come into languages and pass out of them in a quasi-continuous fashion by the various processes of historical change. If this picture is more or less correct, then languages are similar to piles of sand, since those are a paradigm case of objects made up of many small parts without any rigid or prespecified relationships to each other. Now one striking fact about piles of sand is that they all look pretty much the same. Sometimes beach-side artists work hard to produce striking counterexamples, but after a few hours even the most impressive dragons, mermaids, and castles revert to looking like a basic mound. From this viewpoint, distinguishing traits like English Exceptional Case Marking structures or Dyirbal ergative Case marking would be expected to come and go, with their presence or absence having little effect on the language as a whole, and all languages tending toward a relatively homogeneous mush.

On the other hand, suppose that language is more like a Swiss watch,

> which has a complex and rigid internal structure. One important property of this kind of system is that small changes in one localized part can be passed on, even magnified, so that they have large effects on the behavior of the whole. For example, a small twist of the knob on one side can change the configurations of the hands on the face. A small change in the tension of the mainspring could change the speed at which those hands turn, A small gear removed or added anywhere in the watch could cause the whole system to stop. My claim is that this is similar to what happens with human language: small changes in the properties of a single word or class of words, or the addition of a single grammatical principle can have large scale repercussions on the entire language. This happens precisely because everything is bound together into a structured whole by a relatively complex and rigid unified universal grammar. That is why there are so many languages, and why those languages are sometimes so different in their grammars.

The watch *vs.* the sandpile is a useful contrast for the message that I want to convey. Baker sides with the clock model, insisting on a "detailed and rigid" Universal Grammar, one that (in contrast to the sandpile) "prespecifies relationships" (among lexical units), one where different languages are, in a deep sense, quite different from one another. A Universal Grammar that has a "complex and rigid" internal structure, of the sort defended by Mark Baker, corresponds to the view of UG approached from above (in the sense of Chomsky (2007)), and is very reminiscent of the view put forth in Chomsky (1981), as we will have opportunity to examine more closely in section 6.

The image of the watch is also an image made famous in biology by Paley, the author of *Natural Theology*, who greatly influenced Darwin.[13]

> In crossing a heath, suppose I pitched my foot against a stone, and were asked how the stone came to be there; I might possibly answer, that, for

[13]Darwin has this to say about his reading of Paley in his *Autobiography*.

> In order to pass the B.A. examination, it was, also, necessary to get up Paley's Evidences of Christianity, and his Moral Philosophy ... The logic of this book and as I may add of his Natural Theology gave me as much delight as did Euclid. The careful study of these works, without attempting to learn any part by rote, was the only part of the Academical Course which, as I then felt and as I still believe, was of the least use to me in the education of my mind. I did not at that time trouble myself about Paley's premises; and taking these on trust I was charmed and convinced of the long line of argumentation.

> any thing I knew to the contrary, it had lain there for ever: nor would it perhaps be very easy to show the absurdity of this answer. But suppose I had found a watch upon the ground, and it should be inquired how the watch happened to be in that place; I should hardly think of the answer which I had before given, that, for any thing I knew, the watch might have always been there. Yet why should not this answer serve for the watch as well as for the stone?

Paley is now remembered for offering the watch metaphor (the image of intricate complexity) as an argument in favor of the existence of a watchmaker decreased, which Darwin destroyed. Darwin did so by appealing to adam Smith's metaphor of the invisible hand. But it is important to stress that although Darwin eliminated the need for a watchmaker, he did not eliminate the watch. He argued for a blind watchmaker (to use Richard Dawkins' apt characterization), but left the watch intact. That is to say, Darwin did not deny the existence of intricate complexity. And modern biologists have followed him in this respect. In fact, most of them would subscribe to the view that "the theory of evolution by cumulative natural selection is the only theory we know of that is in principle capable of explaining the evolution of organized complexity" (Dawkins (1996)). The designer has changed, but the design stayed the same.

Very occasionally, however, a few brave souls have attacked the appropriateness of the watch metaphor and the idea of intricate, built-in complex design. David Hume, in his *Dialogues Concerning Natural Religon*, writes the following:

> The world, says he, resembles the works of human contrivance; therefore its cause must also resemble that of the other. Here we may remark, that the operation of one very small part of nature, to wit man, upon another very small part, to wit that inanimate matter lying within his reach, is the rule by which Cleanthes judges of the origin of the whole; and he measures objects, so widely disproportioned, by the same individual standard. But to waive all objections drawn from this topic, I affirm, that there are other parts of the universe (besides the machines of human invention) which bear still a greater resemblance to the fabric of the world, and which, therefore, afford a better conjecture concerning the universal origin of this system. These parts are animals and vegetables. The world plainly resembles more an animal or a vegetable, than it does a watch or a knitting-loom. Its cause, therefore, it is more probable, resembles the cause of the former. The cause of the former is generation or vegetation. The cause, therefore, of the world, we may infer to be something similar

or analogous to generation or vegetation.

A more explicit attack of the watch metaphor, one that I find particularly compelling, was made some 200 years later by Alan Turing. To linguists, Alan Turing means computation (The Turing machine) and, perhaps also, the role of language in cognition (the Turing test). But I would like to suggest that to biolinguists, Turing should be first and foremost remembered for his contribution to the field of morphogenesis (a term Turing himself introduced).

In his seminal (1952) paper, Turing addressed what he clearly saw as a (if not, *the*) central problem of biology, namely, how the cell (the zygotic cell of conception) manages, through strictly chemical and physical means, to grow into the far more complex structures of the fetus, the baby, and the mature organism, creating all along *new* information and structure. Turing went on to propose a mechanism, a reaction-diffusion process, that is at the heart of much current work on biophysics. Turing describes his approach thus:

> Unless we adopt a vitalistic and teleological conception of living organisms, or make extensive use of the plea that there are important physical laws as yet undiscovered relating to the activities of organic molecules, we must envisage a living organism as a special kind of system to which the general laws of physics and chemistry apply. And because of the prevalence of homologies of organization, we may well suppose, as D'Arcy Thompson has done, that certain physical processes are of very general occurrence. (Turing and Wardlaw (1953 [1992], 45))

In a remark to Robin Gandy, Turing explicitly states that his new ideas were intended to "defeat the argument from design" (cited from Hodges (1983, 431)). Turing here was not alluding to Paley. He was, I'm sure, assuming that Darwin had put the conscious designer to rest. As Leiber (2002, 86) correctly notes, Turing, rather, endorsed the D'Arcy Thompson view that the teleological "evolutionary explanations" endemic to (neo-)Darwinian adaptationist biology are (to quote Leiber) "non-fundamental, fragile, misdirected, and at best mildly heuristic." Or as Saunders interprets Turing's remark, "The primary task of the biologist is to discover the set of forms that are likely to appear [for] only then is it worth asking which of them will be selected." (Saunders (1992, xii)) Turing was thus addressing the question that Darwin did not address, viz. the origin of species (or, more generally, the origin of forms, the variety on which selection feeds). Interestingly, Turing wanted as much as possible to show that this central question could be studied by appealing to "strictly chemical and physical means." Furthermore, these were anticipated to be "of very general occurrence." His was thus an inquiry into the explanatory power of "third

factor" principles (to use Chomsky's (2005) terminology).

Turing was approaching complexity, not as a *fait accompli*, but as something to construct, and explain by simple means. He was not denying the existence of an additional layer of historical baggage (an extra layer of complexity), but he clearly saw that not all complexities are equal. Some are more understandable than others. Only the complexities that can be explained by strictly physical and chemical means, i.e. those complexities that can be studied in a lawful, a-historical, universal and generic fashion, can be the subject matter of a science called biology. The rest must be left to the historians.

In so doing, Turing was outlining[14] a research program that has grown into an important field of study: Complex Systems, where themes of self-organization, emergence, chaos, and dynamics figure prominently. I like to think of this field as "population physics" (to be contrasted[15] with "population biology", the fundamentally novel perspective introduced by Darwin, according to Mayr (1959, and much subsequent work)), as it studies the behavior of large numbers of entities over which it averages (for a very useful survey, see Ball (2006)). The field of Complex Systems initiated by Turing has grown and diversified. It received key contributions from big thinkers influenced by Turing such as Ilya Prigogine,[16] who named Turing structures "dissipative structures", Brain Goodwin, Stuart Kauffman, Stuart Newman, R. Solé, and many others.

The various approaches to Complex Systems all explore the conditions under which structure and order can form, spontaneously, out of a homogeneous medium, avoiding (very locally, and temporarily) the inevitable grasp of the second law of thermodynamics. One of these approaches, pioneered by Per Bak (see Bak et al. (1988), Bak (1996)), has focused on the phenomenon of 'self-organized criticality' and interestingly, has used the sand pile (one of the images used by Mark Baker in the quote above) to convey its essential properties. Bak et al. (1988) showed how a simple cellular automaton could produce several characteristic features observed

[14]In so doing, Turing was also harking back to a research tradition that (neo-)Darwinians had relegated to the furthest margins of biology, the Rationalist Morphology tradition of Goethe, Geoffroy, Owen, and D'Arcy Thompson, to whose *On Growth and Form* Turing alludes in the quote above. I discuss this tradition in Boeckx (2009a), which I relate to the rationalist cognitive science tradition that Chomsky (1966) called "Cartesian Linguistics". Both of these traditions inform some current Evo-Devo trends in biology, and the renewed focus on Biolinguistics in current linguistics. For more on this, see Boeckx (2010f).

[15]On the contrast between physics and biology I am alluding to here, see Boeckx (2006, chap.4).

[16]Prigogine attended some of Turing's lectures on morphogenesis, and is reported to have said that he spent a day in vigorous conversation with Turing (see Hodges (1983, 564) and Dupuy (2009, 130), who cites a paper by Isabelle Stengers which I have not been able to access).

in natural complexity (fractal geometry, *1/f* noise and power laws) in a way that could be linked to critical-point phenomena. Importantly, they demonstrated that the complexity observed emerged in a robust manner, that it, it crucially did not depend on finely-tuned details of the system: variable parameters in the model could be changed widely without affecting the emergence of critical behaviour (hence, the term *self*-organized criticality). The image of the sand pile nicely conveys their idea: keep adding grains of sand onto a sand aggregate and at some point you will reach a critical sand mass that will cause an avalanche (or a suite of avalanches) until the system returns to a 'barely stable' state, one where avalanches are likely to be triggered soon afterwards if sand is being added.

The image of the sand pile in the context of self-organized criticality is important when one considers Baker's quote above. Yes, sand pile appears to be made up of (quasi-)identical grains of sand, loosely connected, and without any pre-specified pattern. But Bak and colleagues showed that out of this homogeneous state patterns (of the sort Baker alluded to when he is talking about the Swiss watch) do emerge, with "small changes passed on, and magnified, so that they have large effects on the behavior of the whole." What this means is that complex-wide repercussions triggered by small, seemingly insignificant changes robustly emerge without requiring detailed, pre-specificed, rigidly encoded instructions ("grammar" or "lexicon"). The complex whole is the output of simple processes and interactions, rather than the result of complex specifications.

I believe that this was the essence of what Turing wanted to convey in his 1952 paper, and I think that this is the message that (bio)linguists (and most biologists) have not yet assimilated. It is this failure to assimilate Turing's lesson (perhaps because it appears to conflict with the table of instructions that is part of the computational vision that Turing brought forth in his famous 1936 paper, and that is at the heart of classical cognitive science, the so-called computational theory of mind?) that invariably draws us back to preformationism, (thinly) disguised as genocentrism (for biologists) and lexicocentrism (for linguists). Sand piles (think of hour glasses) and Swiss watches may perform the same function, but depending on which metaphor one chooses, it may lead to irreducible complexity or not. Since simplicity is the only road to scientific progress (i.e. progress in explanation and understanding), we should should favor the model that enabes us to built complexity simply, even if it means leaving out of the explanatory scheme quite a few irreducibly complex phenomena.

The model of syntax that I will propose in the next section will draw heavily from Complex Systems approaches, and will inspire certain non-classical conclusions regarding language variation (and its acquisition) and language evolution, as well

as some attempt to relate theoretical constructs to brain entities. In light of the difficulties in grounding extent linguistic theories into biology (be it neurobiology, evolutionary biology, or developmental biology) beyond what Lenneberg (1967) and Chomsky achieved (i.e., "biology at a certain level of abstraction"), I believe that the rather non-conventional (generative, minimalist) approach that follows has a better chance of providing a naturalistic account of some key aspects of the human language faculty. Because of its connection with (and dependence on) concepts from dynamical systems, and the sort of pluralism that I will advocate in the context of language acquisition (section 6 [Part B]), I think that the approach rests on a firmer biological (or biophysical) foundation.

3 Natural Language Syntax as a dissipative system

Minimally, what do we expect from the syntactic component of the grammar, the generative engine of the language faculty? It seems to me that the answer is very clear to everyone: minimally, syntax must combine. The technical word for 'combine' is *Merge*, and that's the term that I will adopt here. (To repeat, the dichotomy between Merge and Unify, so often mentioned by Jackendoff, is not a real one. As far as I know, Jackendoff never says how the various representations that are unified are constructed. Minimalist syntacticians focus on this construction process, and all of them assume something like Merge as part of their answer. Until Jackendoff and others tell us how constructions are constructed, comparing Merge and Unify is like comparing apples and oranges.) I take it that Merge is a very generic set-formation operation. It takes on its linguistic specificity when we consider the units it combines, how many of these there are, and in which context this combination operates. Traditionally, these units are called lexical items, and are taken to consist of bundles of features. For reasons discussed above, I would like to abstract away from these features as much as possible, not because they don't exist at some level of description, but because they have come to be used so much that they prevent us from shedding light on pure syntax and the minimal amount of information required to capture the essence of the language faculty. In short, they are too language-specific, in the sense that they are too specific to particular languages. They are too idiosyncractic. They are also too 'Fregean' in character: these features that serve as triggers for syntactic processes are all of the to-be-saturated kind: once saturated, they cease to exist, or be active for syntactic purposes. As such, they fail to exhibit the necessary property of reusability that is at the heart of recursion in language.

Accordingly, we need to look elsewhere to find that which provides the necessary defining characteristic of lexical items. I don't think we have to look very far, for Chomsky (2008, 139) already identified it here:

> For a L[exical] I[tem] to be able to enter into a computation, merging with some [syntactic object], it must have some property permitting this operation. A property of an LI is called a feature, so an LI has a feature that permits it to be merged. Call this the edge-feature (EF) of the LI.

What this passage says is that minimally for an item to qualify as a lexical item, this item must be merge-able, not just once, but as often as syntax wants. Note that this property, which Chomsky calls the edge feature,[17] is something that by definition all lexical items must have. So, from this perspective, all lexical items are alike. I stress that it is only as far as Merge/syntax is concerned that all lexical items are alike, for semantically, lexical items likely consist of many distinct properties. They can also be differentiated to a large extent by means of the PF-index that they are associated with: they 'sound' different. But, unless we insist on ascribing to them various syntactic features other than the edge feature, all lexical items are indistinguishable syntactically. This, of course, does not prevent them from Merging — which is all that one wants (minimally). So, as far as pure form is concerned, the edge feature is both necessary and sufficient, or so I would like to claim. That is, I would like to claim that for syntax to get started, all that is needed ('the initial condition') is a pool of syntactically undifferentiated, completely uniform lexical items. Because 'lexical' is such a loaded term, I will try to avoid the term 'lexical item' to refer to the pre-syntactic units on which syntax operates, and will call them *conceptual addresses*. Such addresses do what addresses typically do: they point to entities (in this case, concepts). Using an idea of Paul Pietroski's (see Pietroski (To appear)), we can think of these addresses as instructions to "fetch" or activate concepts. Like addresses generally, conceptual addresses are not meant to mirror the entities they point to. Think of snailmail or email addresses: they are silent on the physical specifications of the person they are tagging, or the arrangement of the furniture. They have a flattening, atomizing effect. This is what I want conceptual addresses to do: I want them to activate concepts, but they are not concepts themselves, they are only one (efficient) way to get to concepts. It's important to stress that even if, as I am claiming here, conceptual addresses are atomic, I am not making any claim regarding

[17] I suspect Chomsky called this property the edge feature because "merge is always at the edge" (Chomsky (2005, 13), Chomsky (2007, 11)), in accordance with the Extension condition (Chomsky (1993))/No Tampering Condition (Chomsky (2000)), which forces Merge (be it external or internal) to always target the root.

the atomic character of concepts themselves. I personally think it quite likely that concepts are highly structured, not just in human minds, but also in other animals', but it means that as far as human language syntax is concerned, we will be able to find precious little about the structure of concepts. As far as the combinatorial character of language is concerned, concepts are as Fodor (1998a) claim they are, i.e., completely opaque.

Conceptual addresses in the sense used here are close to the $\sqrt{Roots}$ in Distributed Morphology, or the lexemes of Borer (2005). But, as will become clear momentarily, I claim that syntax only manipulates (combines) roots/lexemes/conceptual addresses. As far as I am concerned, there is no functional lexicon (pre-syntactically); the latter must be constructed too (something which Borer or Marantz or the nano-syntacticians don't do). The pool of conceptual addresses—henceforth, POCA—is completely uniform and homogeneous. Syntax recognizes these addresses as forming a (structurally) natural class. Quite possibly, each address points to distinct conceptual realms, but — I hypothesize — syntax is not sensitive to that. Much like a telephone can dial any number from the address book (in virtue of its format), so syntax can combine any two conceptual addresses in virtue of their common format. The edge feature is the least common denominator among conceptual addresses, and that is key to understanding the infinite character of syntactic combination. Conceptual addresses retain their format, and since that is what enables Merge, Merge can use and re-use conceptual addresses at will. Thus, Merge can be external and internal. As Chomsky (2004, 110) said, "Merge is unconstrained, therefore either *external* or *internal*." It would take a special condition (stipulation) to restrict Merge. I am claiming that no such syntactic condition exists. The constraints come in post-syntactically, at the level of interpretation (on both sides of the grammar).

One way of looking at the present proposal is to view it as a drastic reduction of the number of (morpho-)syntactic features, arguably the most drastic reduction there can be: for syntax to exist, only one syntactically relevant property is needed: the edge feature, the existence of conceptual addresses. In his first explicitly minimalist paper, Chomsky (1993, 44) states as an axiom that "derivations are driven by morphological [i.e., featural] properties to which syntactic variation of languages is restricted." I am denying this, because I think that it invariably leads to derivations that are peculiar to specific languages. In fact, it leads to the view expressed by Polinsky and Caponigro that I quoted above, it leads to the expectation that the lexicon shapes syntax. I am claiming that syntax is not so dependent on lexical details. It requires some specific initial condition (the existence of POCA), but beyond that it is free to operate at will. We'll see that this lifting of lexical restrictions will open the door to a greater use of generic principles ("third factor principles" in the

sense of Chomsky (2005)), thereby enhancing the explanatory quality of syntactic analyses. This will come at a cost, of course: loss of language-specific (in the sense of specific languages), morphological detail. But if the goal is to understand the fabric of the language faculty, the fundamental processes behind the generalizations that make up Universal Grammar, this loss must not only be tolerated, but should in fact be welcomed. The explanatory goal, from a biolinguistic perspective at least, is not the final state of knowledge of language known as I-language, but the forces that make the attainment of this final state possible.

Why am I claiming that the final state is not the explanatory goal from a biolinguistic perspective? Because I-languages are a mixed bag, they are not only biological entities, but also the products of cultural processes. This is clearly express in the following quote, taken from Chomsky (1995, Introduction, p. 11n.6):

> Thus, what we call "English" or "French" or "Spanish" and so on, even under idealizations to idiolects in homogeneous speech communities, reflect the Norman conquest, proximity to Germanic areas, a Basque substratum, and other factors that cannot be regarded as properties of the language faculty.
> ... Pursuing the obvious reasoning, it is hard to imagine that the properties of the language faculty — a real object of the natural world — are instantiated in any observed system. Similar assumptions are taken for granted in the study of organisms generally.

Koster (2009) is right in saying that the capacity for language, even in its narrowest sense, is not a matter of biology but of applied biology, i.e., a technology belonging not primarily to individuals but to their shared culture, if the focus is on I-languages. But Koster errs in thinking that this invalidates the goals of biolinguistics. It simply means that as far as the biolinguistic enterprise is concerned, a level of representation at which the morphosyntactic properties of words don't matter must be found. This is what I am trying to achieve in this work. I am essentially asking readers to entertain the following thought-experiment: imagine syntax without morpho-lexical baggage, syntax freed of lexical influence. How does that syntax look like? Is it good enough (or better) to capture (in a naturalistic, explanatory fashion) the basic properties of the syntactic component (the goal of the minimalist program)? I will try to indicate why I believe that the answer is positive.

3.1 Humaniqueness

At this point readers may well wonder why I introduced this seemingly costly distinction between concepts and conceptual addresses. For reasons I have elaborated

on elsewhere (Boeckx (2009b, In pressc)), I think that the distinction is extremely important because it lies at the heart of why other species lack full-blown combinatoriality ("syntax" in the usual sense), despite the fact that many (arguably, most) ingredients of the language faculty are present in them. It's the root of what Hauser (2008, 2009b,a) has called "humaniqueness". Although I tend to side with Fodor (1998a) in thinking that we know very little about concepts, one thing we know is that other species have a lot of them too. There is massive evidence in favor of the existence of rich mental life in other species. Whatever concepts are, everyone agrees that without them, there can't be any mental life. If to think is to compute, you need representations (conceptual representations, i.e., concepts) to think with (and through). Another thing that's quite clear is that these concepts inhabit, or cluster into mental modules. Many cognitive scientists would say that this is a far more controversial thesis, but I don't think so. The existence of modules isn't at issue, I think. What is controversial is whether Fodor's (1883) specific take on the modularity of mind is correct. But this is a debate that I do not want to open here. All I mean here by modularity is the existence of mental organs, the source of specialized mental computations. This much should really be uncontroversial (see Boeckx (2009b), drawing heavily on reflections by Gallistel (2007)). That modules exist I take to be correct. How many there are, how they arise, what their structure is, and so on is a separate set of issues. Just like concepts, modules are not an asset of *homo sapiens* alone. The ethology literature is replete with evidence that throughout the animal kingdom creatures are equipped with specialized behaviors, many of which require a certain amount of highly specific triggering experience, which crucially transcend the limits of any behaviorist Stimulus-Response schema. I follow Gallistel, Hauser, Marler, Cheney and Seyfarth, and many cognitive ethologists in claiming that animals are equipped with learning organs, that is, modules (a.k.a. "core knowledge systems"). For present purposes, the most remarkable aspect is that animal thought, as can be inferred from behavior in the wild or in the lab, seems to be confined to a few of these modules. That is, there does not seem to be much cross-modular thought. Humans, in fact, appear to be uniquely endowed with an ability to consistently go beyond the limits of these modules and engage in systematic cross-modular combinations (i.e., cross-modular thought). Cross-modular thought, in fact, appears to be the seat of humaniqueness.

It is commonly assumed that the key evolutionary step that gave us our distinctness is cognitive in nature. (Needless to say, it is perfectly possible that our special cognitive features are the result of more basic anatomical changes. See section 8 below.) Accordingly, the quest for humaniqueness amounts to identifying the factor (or factors) that make(s) human cognition special (and ultimately the change(s) at

the level of the brain that made these possible). In the words of Hauser (2008),

> [a]nimals share many of the building blocks that comprise human thought, but paradoxically, there is a great cognitive gap between humans and animals. By looking at key differences in cognitive abilities, [we hope to] find the elements of human cognition that are uniquely human. The challenge is to identify which systems animals and human share, which are unique, and how these systems interact and interface with one another.

As I pointed out in Boeckx (In pressc), the program defined by Hauser can be seen as an extension of Hauser et al. (2002), from FLN to HCN (Human Cognition in the Narrow sense; that which is specific and unique to human cognition), or (building on Fodor (1975)), LOTN (Language of Thought Narrow). A very natural extension, in fact, for, as we will see below, it's quite possible that the nature of LOTN is intimately related to FLN.

Hauser presents four evolved mechanisms of human thought that give us access to a wide range of information and the ability to find creative solutions to new problems based on access to this information:

1. the ability to combine and recombine different types of information and knowledge in order to gain new understanding;

2. to apply the same "rule" or solution to one problem to a different and new situation;

3. to create and easily understand symbolic representations of computation and sensory input; and

4. to detach modes of thought from raw sensory and perceptual input.

Details of formulation aside, Hauser's hypothesis is a very familiar one. The essence of Hauser's claim really goes back to the Descartes' fascination with human cognitive flexibility, its fluidity, its detachment from perception, and its unbounded character—in short, its creative character. This is what led the Cartesians to claim that Man has no instinct, by which they meant that Man's cognitive faculties rise above the *hic and nunc*. This too was clear to Konrad Lorenz, who said that "man is a specialist in not being specialized." (Lorenz (1959)) As Hauser likes to put it, while other animals display, in the terms of Cheney and Seyfarth (1990), "laser-beam intelligence" (i.e., highly precise specificity), humans intelligence is floodlight-like (generalized specificity) in character. Tattersall (1998, 197) calls it "the human noncondition," and writes:

> . . . [O]ver millenia now, philosophers and theologians have made something of an industry of debating the human condition. Even if inevitable, it is rather ironic that the very species that apparently so much enjoys agonizing over its own condition is, in fact, the only species that doesn't have one—or at any rate, whose condition, if any, is most difficult to define. Whatever condition is, it is surely a lot easier to specify it in the case of an amoeba, or a lizard, or a shrew, or even a chimpanzee, than it is in our own.

Elsewhere (p. 207), Tattersall notes that in our case, "natural selection has gone for "flexibility" instead of specificity in behavior" (something which one may attempt to relate to Gould's 1977 discussion of 'neoteny').

To be sure, scientists have found that some animals think in ways that were once considered unique to humans. For example, some animals have episodic memory, or non-linguistic mathematical ability, or the capacity to navigate using landmarks. As I pointed out above, animals have a rich mental life, full of modules or what Liz Spelke calls 'core knowledge systems.' What Man seems to have in addition is the ability to systematically transcend the boundaries of modular thought and engage in cross-modular concept formation.

I would like to claim that this ability of building bridges across modules is directly related to language. Specifically, I would like to claim that it amounts to having a concept vs. forming a conceptual address for that concept. Concepts, I will claim, have limited combinatorial capacity (a fact I will attribute to their adicity, or "selectional properties"), whereas conceptual addresses are far more promiscuous.

I am by no means the first to speculate along these lines. Spelke (2003), Carruthers (2006), Pietroski (2007), Tattersall (1998), Chomsky (2005), and, to some extent, Mithen (1996), all agree with Descartes that language plays a significant role in human cognition. Darwin himself appears to be in agreement when he writes in *The Descent of Man*,

> If it could be proved that certain high mental powers, such as the formation of general concepts, self-consciousness, etc., were absolutely peculiar to man, which seems extremely doubtful, it is not improbable that these qualities are merely the incidental results of other highly-advanced intellectual faculties; and these again mainly the result of the continued use of a perfect language. (p. 126)

The emergence of conceptual addresses was the sort of perfect storm that gave Man his niche. Once concepts as it were became dissociated from their conceptual sources by means of an address, the mind truly becomes algebraic, and stimulus-free. With

language, creativity emerged, understood (as did Arthur Koestler) as "the sudden, interlocking of two previously unrelated skills or matrices of thought," an almost limitless capacity for imagination, metaphorical extension, etc.

Note that the present hypothesis departs from Hauser (2008) positing four distinct mechanisms to account for humaniqueness. One key event (the emergence of POCA. Going back to Hauser's four ingredients for human specificity listed above, we can now claim that by means of conceptual addresses, humans are able to "detach modes of thought from raw sensory and perceptual input," and lexicalize at will ("create and easily understand symbolic representations of computation and sensory input"). Via Merge, humans have "the ability to combine and recombine different types of information and knowledge in order to gain new understanding, and apply the same rule or solution to one problem to a different and new situation."

Thanks to POCA and concomitant Merge, the human mind became capable of true Swiss-Army-Knife-style cognition. Before that the tools at the animal's disposal were exquisitely tuned to their tasks, but too isolated. Their effects could only be combined sequentially; they could not be seamlessly and smoothly integrated with one another. With language, the human mind developed into a key ring, where all keys (concepts) can be combined and available at once, thanks to the hole/format (edge feature) that they all share.

One could say that the ability to endow a concept with an edge feature (i.e. the ability to forma conceptual address) was, to paraphrase Armstrong, a relatively small step for a man, but a giant leap for mind-kind (and mankind). As Dennett (1996, 17) puts it (in agreement with the intuition behind Cartesian dualism), "perhaps the kind of mind you get when you add language to it is so different from the kind of mind you can have without language that calling them both minds is a mistake."

Later on in this monograph [Part D], I will return to the question of what brain changes may have made the rise of conceptual addresses possible, but for now, let me remain at the cognitive level, and say that these conceptual addresses gave Man a truly general language of thought, a *lingua franca*, where before there were only modular mutually incomprehensible dialects/(proto-)languages of thoughts. It significantly altered Man's conceptual structures—how humans think the world. By merging lexicalized concepts, Man was able to hold in mind concepts of concepts, representations of representations, and associations of associations. *Homo* became *Homo combinans*.

The result of the emergence of FLN was a creative, cultural explosion well attested in the archeological record (art, symbol, music, notation, feelings of mystery, mastery of diverse materials, true innovation in toolmaking, sheer cleverness), a "quantum leap," as Tattersall (1998) calls it (Diamond has called it the 'Great Leap Forward').

I agree with Tattersall (1998, 171) that "it is very hard to avoid the conclusion that articulate language is quite intimately tied up with all the other mysterious and often unfathomable aspects of modern human behavior."[18] Tattersall (1998, 186, 228) notes further,

> Almost all of the unique cognitive attributes that so strongly characterize modern humans—and that undoubtedly also distinguished our fellow *Homo sapiens* who eliminated the Neanderthals—are tied up in some way with language. Language both permits and requires an ability to produce symbol in the mind, which can be reshuffled and organized by the generative capacity that seems to be unique to our species. Thought as we know it depends on the mental manipulation of such symbols, which are arbitrary representations of features belonging to both the internal and outside world.
> . . . virtually any component of our ratiocinative capacities you can name—from our sense of humor to our ability to entertain apocalyptic visions—is based on those same mental abilities that permit us to generate language

Through conceptual addresses, we became the symbolic species (cf. Deacon (1997))—a transformative experience that is directly reflected in the archeological records. As Camps and Uriagereka (2006) have noted, aspects of *Homo sapiens*'s toolmaking seem to require the kind of mental computation that are distinctive of FLN (the discussion in Mithen (1996, 76) can also be interpreted in this way).

Monboddo—one of the forefathers of evolutionary thought—was clearly correct in his belief that language is "necessarily connected with an[y] inquiry into the original nature of Man." As Tattersall (1998, 58) writes, "universal among modern humans, language is the most evident of all our uniqueness." Tattersall goes on to note (p. 68) that our closest relatives "do not display "generativity," the capacity that allows us to assemble words into statements or ideas into products." It seems to me that the edge feature of conceptual addresses is a very good candidate for the source of this very generativity, and humaniqueness.

So, to go back to the issue raised at the beginning of this section, the distinction between concept and conceptual address is really important. It is basically at the heart of FLB vs. FLN, for the transition from concept to conceptual address is

[18]Tattersall is quite careful, as is Chomsky (see Chomsky (2008)) to distinguish between language [FLN] and speech, which he regards as an "adjunct" (p. 186) to language. It is quite possible that this adjunction was a recruitment of old circuits specialized in externalization of the sort we find in songbirds. For relevant discussion, see Piattelli-Palmarini and Uriagereka (In press) on the possible role of *Foxp2* in this respect. I return to this theme in sections 7 [Part C] and 8 [Part D].

akin to endowing a concept with a certain 'inertia', a property that makes a "lexical item" active, allowing it to engage in endless Merge-relations. When I say 'endow' I should perhaps stress that the formation of mergeable conceptual addresses was not so much an addition as a lifting of combinatorial restrictions. It was a proces of demodularization (the lifting of a constraint of the type explored by Evo-Devo biologists). Hence my attack on lexicocentrism, for it essentially reintroduces selectional restrictions that obscure the fact that, with conceptual addresses, Merge became unconstrained (and acquired its most famous characteristic, unbounded recursion). Hence also the importance of the elasticity, flexibility of lexicalization discussed by Borer (2005) and Pietroski (To appear), who clearly show the loss of modular selectional restrictions thanks the process of 'lexicalization' (understood as 'formation of conceptual addresses'). When it comes to the concept/conceptual address distinction, Inclusiveness is massively violated. For *homo sapiens*, this was a good thing (it provided the spark that in the fullness of time made the language faculty possib;e). For the linguist, it made lexical semantics a nightmare, as it opacified/atomized concepts, making it impossible to infer anything about the innards of concepts from morpho-lexical decomposition. It effectively invalidated Generative Semantics. It also rendered reference-based semantic theories of language futile, as conceptual addresses resolutely point inside the mind, not to anything in the outside world to which they might refer.[19]

3.2 Regulating Merge α

The concept looming (very) large at this point is symmetry: the pre-syntacticl lexicon, which I have called POCA, is completely homogenous and uniform; all the conceptual addresses inside POCA are, for syntactic purposes, identical; completely interchangeable. And the unique process discussed to so far is Merge, which, in its simplest form, amounts to a symmetric set-formation procedure. For symmetry to prevail is highly desirable from an explanatory point of view: good explanations are those that tend not to beg further questions (at least immediately). Bad explanations are those that end in 'But why?'. Symmetry is so valued in the context of explanation precisely because unlike its opposite (a-symmetry), it does not beg immediate "but why?" questions. Invariably, symmetry is taken to be the default at the explanatory level. If an asymmetry exists, the urge to ask "but why did the asymmetry go this way, as opposed to that way?" is irrepressible. Since by definition symmetries don't go one way as opposed to another, no question asked, no request

[19]Concepts themselves may have perceptual, hence, ultimately, external roots. But that is independent of the language faculty.

for a deeper explanation. So the fact that symmetry rules is good for the minimalist project that seeks to go beyond explanatory adequacy. But, it's not so good as a representation of I-languages: there, to judge from the literature, asymmetries abound, at all levels: in the lexicon (features, which are nothing more than expressions of differences/asymmetries), in the syntax (think of traditional accounts of extraction or binding asymmetries, or asymmetric c-command, etc.), in the semantics (scope asymmetries, and so on), and in the phonology (incorporation asymmetries, prosodic prominence asymmetries, ...). Everywhere you look, it seems that Saussure was right in saying that "A language is a system of differences with no positive terms." Of course, Saussure is talking about specific states of the language faculty (John's English as opposed to Jürgen's German), not about the initial state that eventually develops into an I-language. If one takes a more dynamic, developmental perspective, one quickly realizes that asymmetric output need not entail asymmetric input. The mistake at the heart of lexicocentrism is the incurable urge to build in asymmetries right from the start to render the developmental process deterministic. Asymmetries do not convert into symmetries. As a matter of fact, asymmetries tend to amplify. So, if one puts asymmetries at stage 0, it's a fair (comforting) guess that it will show up at stage n ($n > 0$). But this is not the only recipe for asymmetries to obtain. There is a more emergentist method (which Turing pointed to) that lets asymmetries arise (and grow) out of symmetries. Symmetry-breaking is by now a well-established, and much-studied process, central at fundamental levels of physical explanations, and essential in the context of complex dynamical systems. I intend to put it to use in the context of the language faculty to help me bridge the gap between the desirable symmetry already achieved and the asymmetries that are need to capture essential generalizations about I-languages.[20]

Consider what the syntactic component can produce on the basis of what I have proposed so far: giant sets like (5).

(5) {A, {B, {C, {D, {E, {F, ...}}}}}}

Such sets don't look so inadequate until we remind ourselves that the As, and Bs, of the set just written down are indistinguishable from one another. Accordingly, (5) is the same as (6).

[20]The use of symmetry-breaking advocated here can be seen as a radical extension of Moro (2000), who was the first to investigate the possibility of symmetry-breaking in syntax, but who limited its use to situation of small clauses (merge partners bearing an identical label), and who resorted to movement to break the symmetry of Merge—an option that we should be suspicious of, if move really (internal) merge.

(6) {A, {A, {A, {A, {A, {A, ...}}}}}}

In turn, (6) equals {A}. All this Merge activity achieves very little. It is clear what the problem is. I assume that the external systems interacting with syntax have resources of their own, among which are distinctions ('features'), be they PF-distinctions or LF-distinctions. But such resources cannot be put to use (in a systematic fashion) in the absence of anchor points in the syntax or instructions as to what to distinguish where and when. If syntax produces structures like (6), interpretively, it boils down to {A}. Think of this situation in information-theoretic terms: since Shannon's famous characterization of information, we know that if I say A, then I say A again, then I say A another time, very quickly (infant studies run by psychologists indicate that this is extremely quick), my utterances will become uninformative. They will have become too predictable. In terms of informational addition, their value is nil. To avoid this habituation effect, psychologists know that they have to introduce unexpected elements, which is to say, they have to introduce differences, or choice points to avoid the degradation of the signal (the irremediable effect of entropy). The same strategy must be used if we want to avoid (6). But for reasons that I outlined in the previous sections, I refuse to do this pre-syntactically. Instead, I will let the syntactic derivation, the dynamics of Merge, do it. I will do so by invoking cyclic transfer (also known as cyclic spell-out or derivation-by-phase). As I suggested in Boeckx (2009c), points of transfer, distributed over the entire derivation, will function as anchor points on which to graft the various differences that the external systems want to impose on syntactic expressions to make the most of them, interpretively speaking (a version of Chomsky's (1986b, 1995) principle of Full Interpretation, as it were). Cyclic Spell-Out will, as it were, regulate the Merge α procedure to avoid massive information loss at the interface.

Conceived of in these terms, cyclic spell-out (CSO) does dynamically what labels did representationally, or what features did lexically (since lexical features are micro-representations that project at the maximal node level as labels, labels and features can be used interchangeably. There is no gain using one or the other, contra Collins (2002)), but the key point here is that they do so dynamically: they do not encode the interpretive differences, that is, they do not redundantly express differences that are necessarily present post-syntactically. Instead they construct points at which differences that independently exist get expressed linguistically. They render a Merge α model informative (that is, meaningful/fully legible) from the perspective of the (mind-internal) external systems.

Let me add in the present context that because I will resist the temptation to encode these interpretive asymmetries in the syntax/lexicon, it is critical that CSO

deliver information to both SEM and PHON at the same time (contra Marušič (2005), Richards (2008), Narita (In progress)). Failure to do so would result in irreparable discrepancies between sound and meaning.[21] Only representational devices (labels, features) could resolve mismatches arising from asynchronous transfer (much like traces did in GB).

The last remark in favor of synchronizing CSO provides an encouraging sign of the restrictive character of a Merge α model. Although Merge itself is not triggered by anything, and is unconstrained (insensitive to lexical content), once it is supplemented with CSO, analytical options that are available in accounts relying on a richer conception of UG cease to exist. Without features, there cannot be any talk of feature percolation, feature copying, feature checking, feature deletion, etc. in syntax. The range of analytical options is in fact dramatically reduced: all asymmetries deemed relevant must be related to points of transfer. From here on the task is clear: all lexical (pre-syntactic) and all syntactic asymmetries must be relegated to points of transfer or to post-syntactic interpretive processes that make reference to these transfer points. Everything else will be deemed non-kosher. Needless to say, the task is too large for any single book, but I will show in the next section that some of the most salient properties of the language faculty can be captured (moreover, in a theoretically more satisfying way) along the lines discussed here.[22]

Before turning to core aspects of language, let me make a few final remarks concerning the model of grammar outlined here.

First, the combination of (free) Merge and CSO is belongs to a family of systems that self-organize—avoiding the form-deteriorating effect of entropy by staying at the edge of chaos. To this family belong: the reaction-diffusion mechanism that Turing envisioned, Prigogine's dissipative structures, and the Clock-Wavefront model of Cooke and Zeeman (1976). It has all the characteristics of Bak's self-organized criticality: a mechanism that creates elements that will eventually leads to its destruction. Think of the sand pile: grains of sand accumulate to form a pile, and in so doing produce a structure that leads to its destruction (avalanche). If the system is connected to a constant source of energy/matter (constant sand addition), the accumulation will begin again after the avalanche, leading to another avalanche, and so on.

The Conjunction of Merge and CSO, connected to a constant source of "con-

[21] As Bridget Samuels points out (p.c.), within the present approach, interface synchronization is not only "the best-case scenario" (Chomsky (2004, 107)), it is in fact the only viable option. Synchronization here is not just a question of optimizing computational efficiency, as in Chomsky's discussion, it is a matter of convergence.

[22] For a companion work, focused on phonology, see Samuels (To appear).

ceptual matter/energy" (POCA), has the same characteristic: the more you merge, the faster you reach a phasal point, a point of transfer that destroys the structure achieves so far, only to serve as the basis for the formation of a new structure, to be transferred/destroyed again, etc. The general properties of such a system can be related to process in chemistry, biology, social structures, and so on (see Ball (2006)) because these systems have been viewed from a perspective that seeks to identify its universal, generic characteristics. It is precisely because all these processes are insensitive to the fine details of what they manipulate that they can be compared and related to one another. And it is precisely because I have reduced syntax's sensitivity to lexical information so much that the generic properties of the syntactic component can be made visible.

A second point that I would like to make is that in a way the conjunction of Merge and CSO vindicates the assertion I made in Boeckx (2009d) that Merge in in some sense not enough. In Boeckx (2009d) I claimed that Merge had to be decomposed into Combine and Label (analyzed there as copying), that is as a combination of a symmetric and asymmetric process.[23] This idea is maintained in the present framework, but a more natural characterization of the asymmetric step is offered here in terms of CSO. In so doing, we no longer need to say (as I did in Boeckx (2009d)) that Merge has to be decomposed. Rather, as Noam Chomsky noted in a remark reported in Boeckx (2009d, 52f.), a labeling algorithm has to be added to Merge. What I am proposing here is a natural labeling algorithm (without actual syntactic labeling)[24] in terms of CSO, the details of which I am about to turn to.[25]

4 Converting mysteries into problems via phases

How can we (re)construct syntax from the ground up, with such a minimal set of (lexical) instructions? The first thing to do is get over the fear of over-generation, and the desire to capture every detail. As Borer (2003, 33) points out,

[23]For a related proposal, using different sets of primitives, see Hornstein (2009).

[24]Thinking of Borer (2005) one could speak of exoskeletel labeling, where endocentricity is derived, not built in lexically. In the present system, endocentricity is the result of the internal domain of the phase being spelled out.

[25]Chomsky (2007, 2008) himself has proposed a labeling algorithm, which boils down to "the label is the head contained in the Merge set." Chomsky's proposal faces numerous difficulties: (i) what counts as a head if lexical items are bundles of features (i.e., phrasal)?; (ii) the need for syntactic label, and the format of the labeling algorithm is (as Chomsky (2008) admits, "carried over from X-bar-theoretic approaches" (especially, it stipulates that the target of movement counts as the label). For additional discussion, see Richards (2009b). Other labeling algorithms proposed in the minimalist literature are equally stipulative; see Boeckx (2008b, 80–84).

> It is in the nature of things that an endo-skeletal approach, with its ability to associate idiosyncratic as well as unpredictable syntactic properties with atomic listed lexical items, is both less restricted and more redundant, but also, potentially, more capable, at least prima facie, of describing the wealth of phenomena attested in natural language

As I claimed in the previous section, and as I intend to show in detail in this section, using phases judiciously can go a long way towards curbing the power of unconstrained Merge. I say 'judiciously' because there exist proposals according to which "every phrase is a phase" (see Epstein and Seely (2002, 2006), Müller (2010), Bošković (2007), Narita (In progress)), which I think takes away most of the explanatory potential the notion of phase has to offer. According to these authors, every instance of Merge is followed by Transfer. So, every output of Merge is a phase. Two major motivations have been offered in favor of this view:[26] (i) the fact that every application of Merge corresponds to an interpretive unit suggests that every output of Merge is transferred (at least to the SEM-component; Narita, Epstein and Seely); (ii) the fact that every maximal projection can serve as an intermediate landing site for movement suggests that every phrase is a phase, if successive cyclicity is tied to the concept of derivation by phase (Müller, Bošković). Treating every application of Merge as phasal would be problematic for the present proposal. If everything is a phase, my claim that Merge is regulated by phases cannot be correct. If there are too many phases, their roles as anchor points is lost. Phases would lose their information-carying role: a consecutive string of phases would be no different from a long list of undistinguishable Merge applications. Phases only become useful if there are non-phases too. As I will argue below, it is the delicate balance (the 'Goldilocks solution') between phases and non-phases that take the concept of phase beyond explanatory adequacy.

This is not to deny that every application of Merge corresponds to an interpretive unit, or that every phrase can be the target of (intermediate) movement. But it suggests (as I have argued elsewhere, in fact; see Boeckx and Grohmann (2007), Boeckx (2008d, 2009e, 2010d, Submitted)) that the relationship between phases and successive cyclicity (or locality, more generally) should be loosened.[27] It also requires us to

[26] A third motivation may be added: the fact that no one has yet succeeded in justifying on minimalist grounds why some heads as opposed to others should qualify as phase heads.

[27] Roughly speaking, I have argued that in the context of successive cyclicity (/locality) one must carefully distinguish between actual movement steps taken in narrow syntax, and the footprints of these steps seen through the lenses of the interfaces. The fact that we seem to find evidence for 'punctuated paths' is simply telling us that the syntactic derivation is accessed in a punctuated fashion. It says nothing about the punctuated or uniform nature of these paths in syntax.

recognize that the concept of phase, as articulated by Chomsky since 2000, consists of two parts: a phase edge (the non-transferred part) and a phase complement (the transferred part). It is the transferred part that must correspond to an interpretive unit, not the phase as a whole.[28] Accordingly, if anything, the fact that every Merge output corresponds to an interpretive unit suggests that every Merge output is a phase complement, not a phase.

4.1 Cyclic Spell-Out as the mother of all linguistic asymmetries

The decomposition of a phase into a phase edge and a phase complement is a very important one. Assuming that phases define the size of the active memory of the computation (sometimes called the 'locus' or the 'workspace'), and that once spelled out, the relevant elements are taken out of this active memory, it must be the case that not everything that is part of the phase is spelled out for this would effectively mean that syntactic computations are Markovian. I take it to be a foundational result of early generative grammar that syntactic computation goes beyond the bounds of finite-state automata (see Chomsky (1957, 1959), Lasnik (2000)), and thus conclude that at the very least one element of the phase must escape spell-out at the phase level. For Chomsky, this element is the phase head and everything adjoined to or in the specifier of that head. Chomsky calls everything that is not transferred at the phase level the phase edge. In Chomsky (2000), v and C are the designated phase heads: TP and VP are thus the transferred units (phase complements). Subsequently, other functional elements such as D and P were categorized as phase-heads. Root-categorizing heads in Distributed Morphology, such as n or a (on a par with v), were also said to be phase heads (Marantz (2008)). Needless to say, the exponential growth of the functional lexicon in the context of cartographic approaches also begs the question of whether other heads also qualify as phase heads, a question rarely addressed explicitly (but see van Craenenbroeck and van Koppen (2002)).

The difficulty in identifying—and much more importantly in a minimalist context, in justifying—which of the heads of the lexicon count as phase heads belies the absence of a theory of categories in syntax, an absence that the recent attention on labels and labeling algorithms has obscured. Baker (2003) has very explicitly stated that even for 'basic' lexical categories, we are still very much in the dark. The

[28]Epstein (2003[2007]) was the first to make this point explicitly. Reading Chomsky can be confusing in this context, as he often talks of phases as "having a natural characterization in terms of I[nterface]C[ondition]" they should be semantically and phonologically coherent and independent" (Chomsky (2004, 124)). For additional relevant quotes, see Gallego (In press, 54–55).

situation is even worse in the functional domain, where positing new heads is as easy as positing new features. As is well-known, the strength of theories is measured in large part by the sort of things they deem impossible. Weak theories don't exclude very much. In the context of categories, our theories barely exclude anything. To the extent we deem a category impossible, we rely on the raw data to tell us what seems to exist and what doesn't.[29] It should be clear to everyone that positing "*a*", "*n*" and "*v*" merely begs the question of why we have these category-forming heads, and not others. In this domain, we have not advanced much since Chomsky's (1970) [±N; ±V]. Baker (2003) is perhaps the most comprehensive attempt in recent years to shed light on the nature of (lexical) categories, but there are reasons not to endorse his conclusions: Baker takes Nouns to be the things that bear a referential index, verbs to be the categories that project/license specifiers, and adjectives to be neither verbs nor nouns. In a minimalist context, it has become clear that specifiers do not have a special status, they are merely complements on the other side of the head as it were. Any condition that treat specifiers *sui generis* is a stipulation. As for Nouns, is it because they bear a referential index that they are Nouns, or is it because they are Nouns that they bear a referential index? If the explanatory flow is from syntax to semantics, as Hinzen (2006, 2007) urge us to assume, we should be suspicious of Baker's stance. Finally, it is now clear that adjectives are composite categories (most of those focusing on this issue bet on an adposition + (incorporated) nominal combination; see Mateu (2005), Amritavalli and Jayaseelan (2003), Kayne (In press)), so treating them as default, as Baker does, does not strike me as promising.

4.2 Emergent categories

In the framework being developed here, the pre-syntactic lexicon is category-less: every presyntactic lexical item (i.e., conceptual address) is a root-like element. Categories, like everything else 'lexical' (in the sense of language-specific) must emerge from the dynamics of the syntactic computation and the way it interfaces with the external systems. A good first step, then, is to ask what sort of structures the syntactic component could deliver to the external systems, and what sort of structures these external systems can digest. A good example to bear in mind here is the perspective on argument structure advocated by Hale and Keyser (1993, 2002). Recall that for Hale and Keyser, two things must hold for argument structure to emerge: there must be a well-defined structural template mapping onto the external systems

[29]Witness, e.g., Fukui's (2006) Visibility Guideline, that relies on overt morphology alone to accept the existence of functional projections in a given language.

(for Hale and Keyser, this was the $\overline{X}$-bar template), produced by the syntax,[30] and also there must be available at the interface, coming from the SEM-component, a suite of thematic notions (agent, patient, goal, etc.) to graft onto the structural template. To paraphrase Kant, structural templates without conceptual notions are empty[31]; conceptual notions without structural templates are blind.

To formulate an explanatory account of linguistic categories, we must pay attention to the format of the information delivered to the external systems, and to the conceptual primitives by means of which the structures can be interpreted. From the point of view of syntax, the structures produced are the result of Merge. How much of the structure is transferred depends on when Transfer takes place. So far the only thing we know is that not everything that is built by Merge can be transferred at once: there must be a non-transferred residue, called the edge. If the account is to go beyond explanatory adequacy, we do not want to stipulate what the edge consists of. Transfer should be an operation as free as Merge is. That is, we should think of Transfer as Transfer α: bad transfer choices should be filtered out post-syntactically. We do not even have to stipulate that there must be a non-transferred portion (phase edge): let the transfer-all option exist. If everything is transferred, the syntactic computation comes to an end. Arguably, this is a good solution to the persistent problem of what to do with the edge of the highest phase (root CP), in a framework like Chomsky (2000): the absence of a phase edge serves as an indication that the sentence has been completed.[32] Arguably, this very indication could serve as an instruction to take the object thus transferred as the one to be judged as True or False (if the speakers intends to do so), if Hinzen (2007) is right about Root ForceP/CP being the point of judgment. Assuming Transfer α also means that what Chomsky takes to be the phase head may be transferred alongside its complement, leaving but the specifier(s) in the (non-transferred) edge. This very possibility is suggested by Ott (To appear) in his analysis of Free Relatives. Ott points out that transferring the phase head would in many cases lead to a selectional failure (e.g., if C is transferred, it becomes impossible to state that a higher V takes the entire CP as complement, since C and V are not phase-mates), hence is 'prohibited' (again, I beg the reader to understand 'prohibited' as a descriptive term, not as a constraint on the operation Transfer: everything is possible; what is 'prohibited' is a post-hoc

[30]Strictly speaking, for Hale and Keyser, the $\overline{X}$-bar template was produced by lexical phrase structure rules. They did not in any way attempt to derive/construct it. Their perspective was still very much projectionist. For insightful discussion, see Fasanella-Seligrat and Rubio (2010).

[31]This is to me the strongest argument against Hinzen's reduction of semantics to syntax advocated in Hinzen (2006, 2007).

[32]Obata (2010) has independently reached the conclusion that Root C is included in the last transferred portion.

evaluation based on what led to post-syntactic filtering). In some cases, though—for example, in cases where selection is not an issue—transferring the phase head as part of the phase complement is the desired result. Ott argues that this is exactly what is needed to understand the nature of free relatives. Later on in this work, when I discuss the nature of feature valuation in a feature-less framework like this one, we will see that not even the transfer of "unvalued features" ("Feature Inheritance") can be enforced (contra Richards (2007), Chomsky (2007), Gallego (In press)). (I will argue that if (what amounts to) unvalued features don't get transferred, they will be interpreted. In the context of agreement features, this will provide the basis for clitics, which, I will argue, are nothing more than interpreted agreement markers.)

On the basis of these considerations, we can reach the following conclusions: in circumstances where we want the syntactic computation to continue, and two elements (conceptual addresses), say α and β, have been merged, forming $\{\alpha, \beta\}$, either one of them can be transferred, assuming that the external systems can "digest" (perceive/interpret) singletons. It is a fair guess that they can. In the context of visual cognition, Pylyshyn (2007) has argued that the mind connects with (i.e., perceives, filters, frames, categorizes, interprets, conceives of) the world at least minimally by resorting to "bare demonstratives" that identify (instantiate) individuals (this is what Pylyshyn calls "Fingers of Instantiation" or FINSTs), forming structures like THING(x), where x is the element perceived. Building on Pylyshyn, I would like to argue that when single units (conceptual addresses) are transferred, they receive a nominal interpretation: "if {x} is transferred, interpret x as thing", or more accurately, "if {x} is transferred, categorize x as N." This rule of construal has the same effect as Marantz's root-categorization rule: "interpret $\sqrt{x}$ as N if $\sqrt{x}$ merges with a phase head of type n." I do not allow myself the use of specific heads like n, but there is an easy way to reconstruct them from below: given $\{\alpha, \beta\}$, if β alone is transferred (and thus interpreted as N), $\alpha = n$, exoskeletally, as it were. In virtue of being linked (via Merge) to β at Transfer, α acquires an identity too, as a consequence of derivational history (the memory of the derivation). That is to say, although α is not transferred, it is irreversibly linked to the fate of β: as the derivation grows, symmetries formed by Merge break down (spontaneously), and the asymmetries thus formed constitute the basis for further asymmetries. Layers of complexity accrue, as syntactic elements acquire ever larger contexts of identity. Differentiation comes about in the course of (derivational) history—modification through descent.

Can the external systems interpret more than singletons? (After all, nothing forces immediate transfer of either α or β, perhaps both of them could be transferred at the same time once $\{\alpha, \beta\}$ merges with γ.) The answer seems to be yes; and, again, I am taking Pylyshyn (2007) as a source, if only because I take visual

cognition to be a very good basis of conceptual primitives. Visual cognition is much more ancient, evolutionary speaking, than linguistic cognition, and it is not inconceivable that some of its interpretive properties could have been recruited (recycled) in the linguistic domain (forming part of FLB).[33] Pylyshyn argues that in addition to connecting (visually) with the world via "things", we can also connect via "places": we can not only individuate objects, we can also locate them, place them on some referential axis. Alongside bare demonstratives, we also have bare locative: PLACE(x,y), where x is really THING(x). Bare locatives form the basis of Figure-Ground articulations, which have been recognized as central to much of cognition. Pylyshyn's places and things are only well-motivated cognitively, they also relate in an obvious way to the neurologically-motivated Two-Streams hypothesis concerning visual processing, according to which as visual information exits the occipital lobe, it follows two main channels, or "streams" (see Mishkin and Ungerleider (1982), Ettlinger (1990)). The ventral stream (also known as the "what pathway") travels to the temporal lobe and is involved with object identification. The dorsal stream (or, "where pathway") terminates in the parietal lobe and process spatial locations. This is all to say that I believe I am on firm bio-cognitive ground when I posit the existence of a capacity external to narrow syntax that can categorize inputs as objects and locations.[34]

[33]For a similar point, taken in a different direction, see Arsenijević (2008).

[34]Let me note that the inventory of basic categories is thus, from the purely linguistic point of view, an accident. As Dennis Ott notes

> A differently structured C-I system (one that "digests" different kinds of information) would yield different phases, i.e. P-completeness (Chomsky's "propositionality") is relative to the structure of C-I, hence not a syntactic notion at all. (Ott, 2008, 1)
> The particular phases we find in human syntax is thus not a matter of necessity; if the C-I system were structured differently, different structures would be 'picked out' for concept formation. It is perfectly coherent to imagine a C-I system that can only deal with local theta-structure configurations but has no use for CP-like structures. The structure of Expressions is thus not determined by narrow syntax, but by C-I properties.[footnote omitted] (Ott, 2009, 360)

(Ott's remark echoes Uriagereka's (2008, chap. 1) well-taken observation that the existence of two interface levels (LF and PF) in the early minimalist model of Chomsky (1993) is a contingent fact, an accident of our species' cognitive make-up prior to the emergence of the language faculty. A similar point is made in Chomsky (2007, 15), where Hinzen's project of deflating semantics (Hinzen, 2006, 2007) is being discussed. Chomsky approves of the attempt to "reduce the primacy of CI, [al]though satisfaction of CI conditions cannot be entirely eliminated. CI must have some range of resources that can exploit the properties of the generated expressions." In terms of Chomsky (2005), the kinds of phases or categories one finds is a first factor effect, not a third factor effect, only the fact that we find phases in Narrow Syntax at all is I claim a third factor effect.

In a linguistic context these categories map straightforwardly onto proto-Nouns and proto-adpositions.[35] I use terms like "proto-noun" and "proto-adposition" (inspired by Dowty's (1989) "proto-roles") because nouns and adpositions are loaded terms, morphologically. I will argue below that most of the functional elements standardly assumed in the literature actually reduce to adpositional elements. This set includes adjectives, complementizers, determiners, conjunctions, and even varieties of v. And yet, at some level, such as morphology, all these categories are quite distinct in many languages. So by "proto-adposition" I intend to draw the reader's attention to the fact that I am not using terms familiar from parochial paradigmatic systems, and also to the fact that morphology can be misleading. Having said this, the literature on grammaticaliation reveals patterns that cannot be ignored: as Heine and Kuteva (2007) show, most categories found cross-linguistically can be traced back to nouns and what they call verbs, although I think that 'adposition' would be a better term for the latter, as Heine and Kuteva explicitly state that the verbs they have in mind are non-inflected for tense (a defining property of adpositions). Be that as it may, what is important in Heine and Kuteva (2007) is that patterns of grammaticalization reveals that a two-category system is sufficient to generate all the other categories. Heine and Kuteva make the claim in the domain of diachrony, but I want to extend this to the domain of derivational dynamics (intentionally blurring the synchrony-diachrony distinction sacrosanct since Saussure; see also Longobardi (2003) for skepticism regarding the hard dividing line between synchrony and diachrony), and say that morphological categories other than nouns and adpositions must emerge post-syntactically as more complex versions of these two basic classes.

To repeat, syntactically, there is no notion of Noun or Adposition, or any other category. What matters there are Spell-Out patterns: what gets transferred gets categorized, and this in turn is dictated by when Spell-Out happens: after first merge, or somewhat later. What we find in syntax then is what can descriptively be called intransitive and transitive phase patterns: transferred singletons and transferred

[35] One consequence of the present system is that structures typically described as Noun-Noun compounds must be analyzed as being mediated by a non-noun forming category, a covert p, as in (i) or (ii).
(i) $[[_{nP}\ n\ [\sqrt{N}]]\ p\ [\sqrt{P_{null}}\ [[_{nP}\ n\ [\sqrt{N}]]]]]$
(ii) $[[p\ [\sqrt{P_{null}}\ [[_{nP}\ n\ [\sqrt{N}]]]]\ n\ [\sqrt{N}]]$
The existence of a null P may account for the semantic underspecification of many N-N compound relations (e.g., *banana box*: box for bananas? box made out of bananas? box in the shape of a banana?), for the adjectival/modifying role of one of the Ns, and also may provide a way to understand the link between availability of productive N-N compound and double object structures (Snyder (1995, 2001)), where covert Ps have also been posited (see, among others, Pesetsky (1996)).
For valuable discussion of compounding in the context of minimalism, see Bauke (In progress).

pairs, respectively. Arguably, these provide the basis for the distinction between non-relational and relational elements, with the non-relational elements corresponding to nouns and the relational elements to everything else. Predicates and arguments[36] would also be good descriptive approximations if it weren't for the fact that I believe Pietroski (2005, 2007) to be right in regarding semantic computation to have a far less Fregean character than usually assumed, and in arguing in favor of a model of semantic composition that is based on a single semantic type, that of (monadic) predicate. I will have more to say about this in section 7 [Part C] (see also Boeckx (2009e)). For now suffice it to say that Neo-Davidsonian representations seem to make do with two types of categories only: X(e) (e.g., $\exists$(e) or Verb(e)) and X(Y,e) (e.g., Θ(Y,e) or Adposition(Y,e)).

So far we have seen that Transfer need not take place as soon as it can (right upon first Merge), but can be delayed, for a larger transferred portion still corresponds to units that the external systems can process. As a matter of fact, I think that there is at least one reason to believe that Transfer must be delayed along the lines just discussed. Recall that upon Transfer it is not only the transferred portion that acquires a specific identity, it is also the sister of the transferred portion, the residue/edge (in traditional terms, the phase-head) that acquires an identity (a history). Now imagine that upon merging α and β and spelling out β, we merge α to γ and then spell-out α. This instance of Transfer would not, or should not result in a recategorization of α: in a cyclic system, we certainly want to avoid having to erase and rewrite information gained at an earlier stage in the derivation (see Watanabe (1995) on this point). What we want then if for each transferred domain to contain one previously uncategorized element. This means that after the first instance of Merge (and Transfer), Spell-Out should be delayed until the transferred domain contains at least two elements: an uncategorized unit and the sister of the portion transferred earlier. This effectively ensures that syntactic derivations conform to an alternating pattern of (using more traditional terminology) Phase-head–Non-Phase-Head–Phase-Head–Non-Phase-Head, a fluctuation that I described in Boeckx (2009c) as a syntactic (structural) rhythm or beat that will form the basis of numerous asymmetries at the interfaces.[37]

[36]The present system also relates in an obvious way to Mateu's (2005) distinction between relational and non-relational categories which Mateu shows is enough to capture all the necessary distinctions in Hale and Keyser's (1993, 2002) theta-theory.

[37]The Phase–Non-Phase alternation was first noted by Richards (To appear), who, however, went on to segment derivations as Non-Phase–Phase–Non-Phase–Phase, which I take to be wrong if derivations proceed bottom up. A combination of Richards's segmentation and mine may help us understand where left-to-right parsing meets bottom-up derivations, as I intend to explore in future work.

It is interesting to note that the claim I just made that each transferred domain contains one previously uncategorized element, if strengthened as 'each transferred domain contain only one previously uncategorized element' has interesting empirical consequences. It makes the prediction that transferred domains should not contain two elements of the same category. This, of course, is but a specific reflex of the need to asymmetrize at the interfaces. There is indeed massive evidence that the external systems do not tolerate symmetry/ambiguity (see Boeckx (2008b) for extensive discussion). If phases are to function as the anchors of asymmetries, as I have argued here, we want to eliminate sources of ambiguity/symmetry at the phase level. Evidence of a lack of tolerance of ambiguity/symmetry at the phase level comes from the work of Andrea Moro on small clauses. As Moro (2000, 2007) shows, there are local domains, such as small clauses, which cannot contain two elements of the same type/category, as represented in (7).

(7) a. **pro* copula [$_{SmallClause}$ DP DP]
b. *sono molte foto del muro la causa della rivolta *(Italian)*
are many pictures of-the wall the cause of-the riot
'many pictures on the wall are the cause of the riot'

As can be seen in the next example, Moro shows that one of the two elements of a small clause (it does not matter which one) must vacate the relevant local domain.

(8) a. molte foto del muro sono la causa della rivolta
b. la causa della rivolta sono molte foto del muro

Moro's observation connects directly to the issue of phases in the present context thanks to the work of Richards (2010), who discusses numerous situations in a wide range of languages where adjacent elements of a same category are prohibited (see also van Riemsdijk (2008)). For example, Richards observes the following paradigm:

(9) a. "It's cold" said [John]
b. "It's cold" said [John to Mary]
c. *"It's cold" told [John Mary]

As Richards notes, it is not the case that any instance of two adjacent elements sharing a category leads to unacceptability. After all, English allows double object (double NP) constructions:

(10) John gave [Mary a book]

According to Richards, the right generalization, which fits perfectly with the categorization-by-phase mechanism advocated here, is that no two elements of the same category may be contained in the same transferred portion (phase complement). (Acceptable instances of adjacent elements sharing a category are only apparent cases of structural adjacency. Richards shows that in all these cases the two elements are separated by a phase boundary.) In the present context, Richards' generalization can be reinterpreted as a prohibition against two elements requiring labeling being contained in the same labeling domain (phase complement)—a ban on symmetry (what van Riemsdijk (2008) call Anti-Identity Avoidance) imposed at the interface.[38]

It is precisely because one wants to limit the number of uncategorized element within each spell-out domain to one that one must not delay transfer beyond the second application of Merge. Delaying spell-out would simply mean allowing more instances of Merge before Transfer, meaning adding more uncategorized elements to the transfer domain. That is to say, there are intransitive and transitive phase patterns, but no ditransitive or supertransitive phase patterns. This fits well with the idea that cyclic spell-out keeps the size the the active workspace to a minimum (cf. Chomsky (2000) on derivation by phase and computational efficiency). In this, like in all minimax situations, one must reach a compromise. Spelling out too much too soon is prohibited (even if it would reduce the size of the active workspace to a bare minimum), but delaying spell out too much is equally illicit. Quick spell-out prevents the formation of irreparable increase in symmetry. As I said earlier, a judicious use of phases can regulate wild-type, unconstrained Merge. Note that this ban on too much symmetry does not conflict in any way with repetition patterns that are standardly used to illustrate recursion in language (*John said* **that** *Bill believed* **that** *Sue thought* **that** ...). It is true that the effect of the ban on local symmetry/Indentity Avoidance prevents one from witnessing the recursive nature of Merge 'naked' as it were (what I have called 'Deep' or 'Dense' recursion in Boeckx (2010g)), and imposes patterns that Tom Roeper has called 'Indirect Recursion' (called 'Sparse' or 'Surface' recursion in Boeckx (2010g)). But it would be wrong to conclude, as Arsenijević and Hinzen (2010) do, from these surface patterns of category distribution that there is nothing like true recursion in language. Merge is truly recursive, but the external systems processing the output of narrow syntax imposes conditions that obscures this fact, much like the system processing languages obscures the legitimacy of center-embedding in syntax, as Miller and Chomsky (1963) already concluded over forty years ago.

[38] van Riemsdijk (2008) relates this Anti-Identity Avoidance ban to Relativized Minimality, which, interestingly, (as Richards argues for Anti-Identity) also appears to hold at the phase-level; cf. Chomsky (2001).

4.3 Too many phases?: A digression

A major way in which the present system differs from Chomsky's phase-based approach is that it not only dispenses with pre-defined phasal categories (phase heads) (thus avoiding altogether the so far intractable problem of why C, and not T, is a phase head), but also it allows for more than two phase boundaries within a clause. Chomsky restricts himself to just two phasal nodes at the clausal level: C and v (which begs the question of how question words in SpecCP_{root} are ever interpreted since they are never transferred). The present system says that the more elements a clause contains, the more asymmetries will have to be anchored (the more conceptual addresses/roots will have to be categories), hence the more phase boundaries will be needed.

In recent presentations, Ángel Gallego has criticized this move, for two reasons: (i) he claims that adding phase boundaries in a derivation renders the computation more complex, and (ii) having more phases than C and v forces one to abandon the link Chomsky thought to establish between his two phase heads and the "duality of semantics" (an idea going back to the distinction between Deep Structure semantics, which deals primarily with thematic properties, and Surface Structure semantics, which deals with discourse properties). The reasons offered by Gallego are not grounds for rejecting the present approach. If anything, they are reasons to adopt it. For one thing, it is true that the present proposal says that as the size of the derivation grows, the more phase boundaries will have to be posited. But for the syntactic component to self-organize in this way is precisely what we expect if good design plays any role in constraining the language faculty. In fact, the increase of phase boundaries proportional to the size of the derivation appears to be a specific instance of what Chomsky would likely call a "third factor" principle; in this case, a specific instance of Menzerath-Altmann law. Menzerath-Altmann law is a general law, first formulated in the context of human language, but later on extended to non-linguistic domains, such as the organization of the genome (see Ferrer i Cancho and Forns (2009)),that says that the increase of a linguistic construct results in a decrease of its constituents, and vice versa. So for instance, the longer a word, the shorter its syllables (i.e., the larger the number of syllables). Menzerath-Altmann law strikes me as very intuitive: the longer the sequence to memorize, the more one chunks it. Gallego is wrong to believe that an increase in phase boundaries leads to greater computational complexity. Just the opposite, in fact. Phases reduce computational complexity (by reducing the size of the active workspace). We therefore expect more phases (to reduce the burden on computational memory) if the size of the derivation grows.

As for Gallego's criticism related to the duality of semantics, it too fails to cast

doubt on the present proposal. It is true that CP and vP neatly map onto the domains of discourse structure and argument structure, respectively, but they do so only in the most basic of circumstances. As soon as one looks into the issue in more detail, the neat cut evaporates: for example, we find the same duality of semantics expressed in nominal phrases, and one can posit a similar C-v articulation in nominals (indeed, many have done so), but note that the C-layer of nominals often enters into the v-layer of the clause, disrupting the neat mapping. Similarly, clausal CPs function as arguments, again causing the two semantic layers to intersect. Moreover, it is now a well-established finding of the Cartography literature that the edge of the v layer hosts elements that contribute to discourse articulations (topics, foci, etc.) (Belletti (2004)), which causes the two kinds of semantic relations to intersperse ($\overline{A}$–A–$\overline{A}$–A). Finally, it is becoming increasingly tempting to analyze Surface semantics as just another instance (a higher-order expression) of Deep semantics, that is to say, to view discourse structure as a special instantiation of argument/event structure, namely, an articulation of the speech event, with discourse functions corresponding to higher-order theta-roles (discourse/speech-event participants). Evidence in favor of this view comes from analyses of Quantification (as reprojection/second-order monadicity) (Hornstein and Uriagereka (2002), Pietroski (2003)), Focus articulation (Herburger (2000), Irurtzun (2007)), evidential markers (Speas and Tenny (2003)), and the nature of person-marking (Sigurdsson (2004)). Such works point to the treatment of the duality of semantics as the result (as opposed to the cause) of the existence of the two dimensions on which Merge can be formulated:[39] external and internal; First Merge and Re-merge; First-order Merge and Second-order Merge—event and speech event. I will have more to say about this in section 7 [Part C], when I discuss the character of neo-Davidsonian representations, in which it is clear that event structure is built by adding participants as arguments, and then taking the event structure thus built and reusing it (recycling it) as an argument. To repeat, the *linked* (or *coiled*) duality of semantics is the result of recursion and the possibility of internal merge (afforded by the unerasable, reusable character of the edge property of conceptual addresses), not the cause of movement (an old idea going back to Jackendoff (1972)), or the rationale behind C and v. (This perspective fits well with Hinzen's (2006, 2007) key idea that semantic properties that are specific to human cognition should be explained in terms of narrow syntax, not the other way around.)

Let me conclude this brief digression on Gallego's criticism by stating that al-

[39]Occasionally, a third type of Merge is distinguished: Parallel Merge (Citko (2005)), or Pair-Merge (Chomsky (2004)), but it would be wrong to view this as a third dimension on which to formulate Merge. Pair-Merge is just familiar Merge (internal and external) on a new plane. On the existence of both external Pair-Merge and internal Pair-Merge, see Richards (2009a).

though a typical syntactic derivation is likely to have more phases in it under the present approach than under Chomsky's, but it is important to bear in mind that we are talking here about phase tokens. When it comes to phase types, the present system only recognizes two phases: transitive and intransitive. In the absence of a theory of (functional) categories in Chomsky's system, it is unclear how many types of phases there are: for me, C and v are of the same type (in line with what I just said, C is just a higher-order v/p). I do not know how Chomsky would relate C, v, and the other phase heads he occasionally recognizes (see Chomsky (2007)), such as P, D, and n.

4.4 More on what makes basic categories different

Let us now return to the theory of categories I advanced above, and examine a few more properties of the system based on transitive and intransitive phases. The first point I would like to make is a comparison with Baker's (2003) system. Recall that for Baker there are three basic (lexical) categories: N, V, and A. (Baker regards P as functional.) What makes V special is its ability to license a specifier; N, its ability to bear a referential index; and A, neither of these abilities. A is then the default. It is interesting that in passing Baker (2003) suggests that VPs bottom out into As, which would lead to the conclusion (which Baker does not explore) that V is a functional category, on a par with P. In the present system, every category is functional in some sense. There is no lexical category (the pre-syntactic lexicon is a-categorial). V (more accurately, v) is on a par with P (more accurately, p). A results from the conflation of P and N (APs are pPs with (morphologically incorporated) Ns). Interestingly, in the present system, all categories can license a specifier (the latter term used descriptively, of course). What distinguishes N from the rest is the fact that it is not transferred with a complement (the complement of n consists of a bare root; that's what makes this phase type intransitive). Note that it would be wrong to speak of an inability on the part of N to license a complement. Since categories here are defined post-syntactically (exo-skeletally): nouns can license complements, but if they do the system calls (categorizes) them differently.

The intransitive character of the nominal phase leads me to address the distribution of ϕ-features (which I will relate to Baker's notion of "referential index"). Φ-features are the prototypical instance of uninterpretable/unvalued features in minimalism. These are the features that require checking, trigger movement, force inheritance, etc. They are at the heart of current syntactic theory. Below I will address the question of how a feature-less system like the present one could accommodate their existence. But for now all I want to do is focus on the architectural constraints they

impose. Whatever they are, unvalued features will require valuation (i.e., the formation of a 'checking' configuration). Even if the actual valuation takes place outside of narrow syntax (as I will argue below), the configuration leading to valuation must be something that the syntax must not only construct (in accordance with the idea that narrow syntax is the sole generative engine), but syntax must also be the component identifying that configuration as special. Already at this point it should be clear that phasal boundaries will be crucial since 'special' configurations mean 'different' — it entails an asymmetry (checking configurations vs. non-checking configurations), and, as I hypothesized above, all asymmetries must build on the "ur"-asymmetry of Spell-Out. Interestingly, implicating phases in valuation is something that Chomsky (2004) already proposed. Chomsky's reasoning is very interesting, and will take us back to the differences among categories in the present system, so let me summarize it briefly here (for more extensive discussion, see Epstein and Seely (2002)).

Our starting point will be (un)interpretability and the notion of valuation. It appears to be a fact about natural languages that not all the features that appear on lexical items are interpreted. Think of ϕ-features on finite verbs. These features are often morphologically interpreted, but semantically, they are inert. Chomsky (1995) calls these 'uninterpretable'. The very fact that they are realized morphologically suggest either that they are introduced post-syntactically, on the PF-side of the grammar, or else that they are present in the syntax (feeding morphology), but are eliminated in the course of the syntactic derivation, prior to interpretation. Chomsky assumes that the latter is the case. (Marantz (1991), Bobaljik (2008) and others have pursued the possibility that at least some uninterpretable features do not enter syntax at all.) In Chomsky (2000) Chomsky characterized (un)interpretability in terms of valuation, the reason being that barring undesirable look-ahead it is not clear why syntax should care about interpretability: why should it react (triggering movement, checking, etc.) to a notion that is only relevant post-syntactically? Chomsky proposed we understand (un)interpretability as the semantic counterpart of a lexical property to which syntax could be sensitive, viz. valuation. According to Chomsky, some features on lexical items come unvalued (they are stored lexically without a specific value), and unvalued features correspond one to one to uninterprable features.[40] Syntax does not tolerate unvalued features because (by hypothesis)

[40]Pesetsky and Torrego (2007) and Bošković (In press) question the link between uninterpretability and lack of value. They argue in favor of a double dissociation between (un)interpretability and (lack of) value so that some lexically valued features can lack an interpretation, and some lexically unvalued features can receive an interpretation. Clearly, this is the worst possible state of affairs in a minimalist context (especially one that argues against the proliferation of features), and so only the strongest kind of empirical argument should force us to go in that direction. I will return to some of the cases they discuss below, showing that they do not pose a problem for Chomsky's

lack of value leads to illegitimate outputs at both SEM and PHON. (Think of a value as an instruction for the external system to perform a certain action. Lack of value means lack of instruction, and assume that the external systems do not like points of indecision.) So, syntax must see to it that all features on lexical items come to be valued by Spell-Out. The problem is that once valued all features will look alike: it will be impossible for the interpretive components to pick out those features that it should not interpret (the original uninterpretable features in Chomsky (2000)). So, although we do not want syntax to have access to a post-syntactic notion like interpretability, it seems like we need the semantic component to have access to a pre-syntactic (lexical) notion like "lexically (un)valued." In Chomsky (2004), Chomsky proposed an ingenuous way of having our cake and eat it too. Essentially, Chomsky proposed that valuation (the point at which the relevant distinction fro the semantic component disappears) takes place at the point of transfer, when the external systems interfaces with syntax. The idea is that by synchronizing valuation and transfer, the semantic component "sees" the valuation process, and is thereby able to tell which feature got valued (i.e., which feature was lexically unvalued). Once the semantic component sees this, Chomsky reasons, it is able to use this fact as an instruction not to interpret the relevant feature.

Chomsky concluded from this that since uninterpretable/unvalued features are distributed across the syntactic computation (they are not all concentrated on the last stage of the derivation), the presence of these features forces cyclic spell-out. If Transfer only took place once (as in the earliest minimalist models like Chomsky (2000)), valuation and Transfer could not be synchronized, valuation requiring a very local domain.

Now let us look at the process of valuation more closely. Valuation is achieved by pairing the unvalued feature of the lexical item α with a matching (lexically valued) feature on another lexical item β that is connected to α. The pairing/valuation process is called *Agree* in Chomsky (2000). The bearer of the unvalued feature is called the *Probe* and the bearer of the matching feature providing a value the feature on the other element is called the *Goal*. As Chomsky (2004) remarks, we do not want Agree to take place under Merge because Merge happens before Transfer does, and we want to synchronize Agree (valuation) and Transfer. (Recall from the above discussion that we do not want to synchronize Merge and Transfer for this would not exploit all the interpretive possibilities of the interfaces: it would only create nominal structures in terms of the present proposal; or if the entire output of Merge is transferred, it would force syntax to behave like a finite-state automaton.) To achieve this (i.e., to delay transfer, and thus dissociate it from Merge), Chomsky

characterization.

(2008) hypothesizes that the unvalued feature always originates on a phase head (i.e., only phase heads are probes), with the goal located in the complement domain of the phase head. But now we face another problem. If the unvalued feature sits on the phase head, it will not be transferred upon valuation, for only the phase complement is transferred. The problem is clear: we must find a way for the unvalued feature to be introduced after the goal is merged, but we must also make sure that it is transferred alongside the goal. The solution, proposed by Richards (2007) and adopted in Chomsky (2007), is to invoke a process of (unvalued) Feature Inheritance,[41] according to which in the course of Transfer (and valuation!), the unvalued feature on the phase-head (probe) is inherited by the non-phase head sitting next to the goal in the complement domain of the phase (the domain that is transferred).

The important point for us is that if Richards's and Chomsky's reasoning about feature inheritance is correct, it restricts valuation to transitive phase domains. With intransitive phases, only one element gets transferred, and this is not enough for feature inheritance to take place, since the relevant feature must end up on a Merge partner of the goal. To put it differently, only transitive phases are proper domains for phenomena like agreement. Arguably, this is why nouns (the outcome of intransitive phase spell-out in the present systems) are the only categories according to Baker (2003) that reliably support the interpretation of a referential index (which I take to be equivalent to interpreted ϕ-features). If one finds ϕ-features on nouns, these cannot have been the result of valuation; they must therefore by interpreted.

The present proposal converges with a key idea in Kayne (In press), viz. that nouns are not associated with unvalued features.[42] Kayne's proposal is similar in spirit to the present one, as it also takes Nouns to be singleton sets. But we differ in how this singleton set is formed. For Kayne, this is the result of self-merge, which is motivated by the fact that in his system (roughly that of Kayne (1994)), two heads cannot merge directly (the result would be too symmetric, given his stance on the a(nti)symmetry of syntax). Therefore, one of the heads must be turned into a phrase. The simplest way to achieve this is via self-merge. Kayne calls the category formed by self-merge a "noun" (or "nominal root"). As Kayne notes, in his system, there is no need of n; Nouns are the only lexical (open-class) category of the human language lexicon. All other categories are closed-class, functional items. I depart from Kayne

[41] In Chomsky (2008) Chomsky contemplates a generalized process of feature inheritance, but restricts it to unvalued features in Chomsky (2007).

[42] Kayne pursues the stronger hypothesis that non-Nouns are necessarily associated with unvalued features, but I think this is too strong, as is the idea that phase heads necessarily bear unvalued features (Gallego (In press)). (Gallego attributes—incorrectly in my opinion—the latter hypothesis to Chomsky (2008).

in taking all categories to require a phase-envelope. For me, the only open-class is POCA, the set of roots/conceptual addresses, which lack categorial information (to repeat, for Kayne, all roots are of category N):[43] Categorial distinctions only make sense when embedded in a system of oppositions.

4.5 The emergence of even more complex categories

I hope that the guiding intuition behind the present model is clear when it comes to the formation of linguistic categories: Narrow Syntax merges freely, and Spell-Out patterns (transitive or intransitive phases) determine, in conjunction with conceptual resources, the range of post-syntactic categories: (proto-)Ns and (proto-)Ps and their distributions.[44] In the present system, there is no other option available. Needless to say, many doubt that such a minimal(ist) system can produce the full range of categories that linguists have identified. If all that can be produced are Ns and Ps, where do categories like C, V, v, D, T, etc. come from—to say nothing about the various 'flavors' that all these categories take in cartographic approaches (C_{Force}, C_{Fin}, P_{Path}, P_{Place}, ...)? But it is important to remember that the justification for most of these categories is morphological in nature, and that they may have much in common with one another syntactically. The very term 'flavor' used in the cartography literature in fact suggests that there is an underlying commonality behind the myriad instantiations one finds.

My strategy here is to pursue a line of inquiry I began in Boeckx (2008b, chap.4) and treat morphological categories that depart from N and P as instances of duplication and divergence, a process well-known in biology, another realm where surface variation abounds. To put it differently, I would like to suggest that finer-grained categories will be defined 'configurationally' (i.e., contextually), according to a schema

[43] I also insist in roots lacking semantic types (contra Marantz (2010), Levinson (2007)). Recall that for me, roots are conceptual addresses, they are not concepts. In particular, they do not preserve the adicity of the concepts they relate to (see also Pietroski (2007)). The concepts themselves may well be (and most likely are) of different types. Roots acquire their semantic 'types' (roles), syntactically, like everything else that one may be tempted to include as part of their pre-syntactic lexical entries. Marantz claims that rots can serve as modifiers of entities, states, and events. But these clearly are post-syntactic semantic uses of roots, based on the syntactic configurations they are in.

[44] It is often suggested that adpositions bottom out into nouns (see, e.g., Collins (2007)), but I think this is incorrect: like all other categories, adpositions bottom out into roots (although we may have to resort to nouns to refer to these roots, but this is a meta-linguistic fact, not a true property of narrow syntax). Evidence for this may come from Mayan languages, where locatives are the only categories that resist association with classifiers (Carme Picallo, p.c.). If classifiers are of the n type, and if you can't get N without n, locatives cannot be nominal at heart.

familiar from phonological rules. For example:

(11) $p \rightarrow v$ / {T,{__}}

As the derivation grows, so does the number of Ps. As the syntactic context for these Ps grows with each instance of Merge, it can be exploited morphologically. T, C, v, and so on become akin to what geneticists call paralogs (genes related by duplication within a genome. Whereas so-called orthologs retain the same function in the course of evolution, paralogs evolve new functions, even if these are related to the original one).

This effectively means that many categories will be born alike, but become distinct, i.e., functionally specialized post-Spell-Out (at the point of recombination/vocabulary insertion). They emerge configurationally, much like theta-roles in Hale and Keyser's (1993, 2002) theta-theory. (Recall that I have argued above that the "duality of semantics" is to be understood as an instance of domain duplication.)

Although this approach to the fine details of cartography may look unfamiliar, I would like to point out that many of the ideas needed to make it work are already available, and in some cases already quite standard. Take, for example, Pylkkänen's (2002, 2008) approach to applied arguments, where the same category (ApplP, a flavor of v/p) becomes specialized depending on its merge-site, hence the existence of low applicatives, medial applicatives, high applicatives, super-high applicatives, etc. (see Tsai (2010)). Likewise, as I will suggest below, the nominal category Person may be regarded as high n (low n corresponding to Gender/ [+animate] Class).[45] Similarly, the distinction between P_{Place} and P_{Path} plausibly reduces to high p and low p. Ditto for Force and Finiteness: high C and low C. (As a matter of fact, Rizzi (1997) conjectures that Force and Finiteness form a syncretic category, and becomes split only under the presence ("activation") of the Topic/Focus field.) The more analytic the pattern of category expression, the more specialized each occurrence of the category in question will become. One could in fact speak of descent/transfer with modification, a process that I take to underlie the patterns of grammaticalization studied by typologists.

It stands to reason that the richer the language is morphologically, the more salient these functional specialization of category occurrences will become. In morphologically poor languages, the inventory of category may reduce to the most basic one: N and P. It is in fact quite significant that in well-studied languages like English, loss of overt inflection blurs the distinction between C and P (think of non-finite

[45]For relevant material, see Picallo (2006, 2008).

complementizer *for*),[46] T and P (think of non-finite T *to*), Aux and P (the auxiliary *of* discussed in Kayne (1997)) — all of these revealing the P-nature of all the other categories, exactly as the present system predicts. (If C is really what gives T its identity (much like p labels a root), as Chomsky (2007, 2008) has argued,[47] isn't it quite natural to treat C as an element providing a location on a time line for an event (Ditto Mood, which locates an utterance in a possible world).)[48]

The same logic extends to the relationship between P and V (or p and v): Perhaps the strongest, most robust argument that Svenonius (2007) comes up with to tease the two categories apart is morphological Tense-marking (which morphological Ps resist). Non-finite verb forms, such as participles and gerunds have been argued to

[46]On C as P, see already Emonds (1985).

[47]Here are relevant passages:

> . . . [F]or T, ϕ-features and Tense appear to be derivative, not inherent: basic tense and also tenselike properties (e.g., irrealis) are determined by C (in which they are inherent: "John left" is past tense whether or not it is embedded) or by selecting V (also inherent) or perhaps even broader context. In the lexicon T lacks these features. T manifests the basic tense features if and only if it is selected by C (default agreement aside); of not, it is a raising (or ECM) infinitival, lacking ϕ-features and basic tense. So it makes sense to assume that Agree and Tense features are inherited from C, the phase head. (Chomsky, 2008, 143–4)
>
> There are further reasons for expecting that TP is not a phase. T has the basic properties of uninterpretable features. It may yield a phonetic reflex, but its ϕ-features are determined by the context, so it should enter the lexicon without values for these features. T bears these features if and only if it is selected by C, hence it should inherent these from C (. . .). The biconditional holds of embedded clauses, but it would make no sense to hold that in root clauses T has different properties. It therefore follows that root clauses must have C, even if it is unpronounced . . . .
>
> What is true of agreement features appears to hold as well for tense: in clear cases, T has this feature if and only if it is selected by C, though C never (to my knowledge) manifests Tense in the manner of ϕ-features in some languages. If that is basically accurate, then there are two possibilities. One is that Tense is a property of C, and is inherited by T. The other is that Tense is a property of T, but receives only some residual interpretation unless selected by C (or in other configurations, e.g., in English-like modal constructions). [footnote omitted] One advantage of the latter option is that T will then have at least some feature in the lexicon, and it is not clear what would be the status of an LI with no features (one of the problems with postulating AGR or other null elements). Another advantage would be an explanation for why C never manifests Tense in the manner of ϕ-features (if that is correct). (Chomsky, 2007, 20)

[48]Contra Chomsky (2007, 21). Chomsky's claim that T has the basic properties of uninterpretable features does not seem quite accurate. Uninterpretable features are not interpreted, whereas T is, albeit its interpretation depends on the functional structure around it.

be P-like (see Gallego (2009b), Masullo (2008) ; see also Emonds (2008) for relevant discussion). The event/aspect structure and adpositional structures have been argued to match perfectly (Ramchand (2008), Tungseth (2008); see also Cuervo (2003) on applicative (prepositional) classes matching verbal aspectual classes). Finally, it is interesting to note that in so-called highly analytic, morphologically isolating languages, it is hard, and perhaps syntactically irrelevant, to distinguish between serial verbs as V-V compounds and serial verbs as (light) P-V combinations, as Aboh (2009) shows.

All in all, if we treat coordinators (say, *and*) as morphologically disguised adpositions (say, commitative *with*), and (as already suggested above) adjectives as adpositions as well (*angry* = *with anger*),[49] then virtually all major categories reduce to adpositions, once stripped of their morphological (historical) baggage.

These remarks point to the fact that the (post-syntactic) morphological component of the human language faculty, like natural selection, constantly tinkers with the spare resources made available to it (by the generator, narrow syntax), recycling[50] the same categories and adding morphophonological flavors to them, which have sidetracked linguists into thinking that these emergent classes of categories are primitives.

I would like to close this section on categories by briefly examining the nature of the functional category D, a category I have not discussed yet, and one that, if the theory sketched here is correct, must be regarded as an instance of p (transitive phase). The present system takes N/n to be the core nominal category ("which is what we intuitively always wanted to say" (Chomsky, 2007, 25-26)). But what is the role of D in this system? Although widely adopted, and deeply implicated in matters such as argumenthood (see Longobardi (1994)), the DP-hypothesis has always been problematic for selection and for concord (see Bruening (2008) for a careful survey of problems; see also Fukui and Zushi (2008)). The selection problem is particularly salient (although not confined to the nominal domain, as I will discuss momentarily; see Shlonsky (2006)): we want the Verb/Preposition to select N, not D. But if D dominates N, as in the DP-hypothesis, why doesn't it block selection?

Being forced to view D as a special occurrence of P, and taking advantage of the hypothesized identity between P and C, I would like to argue that D is in fact a kind

[49]Adverbs would have to be treated as adjectives with an extra P layer: angrily being [$_{pP}$ [$_{pP}$ with [anger]]-ly], with *-ly* in fact being historically derived from the adposition *like*. I return to the double P-layer of adverbials in section 4.10 below, as it bears on their opaque character in the context of extraction.

[50]The idea of morphological recycling is not new; see Longa, Lorenzo, and Rigau (1996, 1998). I would like to add that the tinkering character of morphology may have interesting implications for phylogenetic studies of the human language faculty.

of C. Although taking D to be a complementizer is not completely unheard of (cf. Szabolcsi (1984), among others), I would like to further argue that D is a relative complementizer, and adopt Kayne's (1994) revival of the raising analysis of relative clauses to allow for N/n to raise from within the DP and reproject an NP/nP layer, rendering selection by V/P straightforward, as represented in (12).[51]

(12) $[_{nP}\ [_{DP}\ \text{D}\ \ldots]\ [n\ [\ \sqrt{N}\ ]]]$, where DP = $[_{DP}\ \text{D}\ \ldots [_{nP}\ n\ [\ \sqrt{N}\ ]]]$

Note that the derivation in (12) not only solves the selection problem, it also allows us to treat concord (controlled by N/n, not by D) as a case of standard agreement, established in a manner similar to agreement at the clausal level. Upon raising internal to what is a relative clause, nP established agreement/concord, much like a subject nominal triggers verbal agreement inside a clause. The proposal also captures the oft-noted dependency between D and Person (see, e.g., Longobardi (2006) for the claim that DP is really PersonP), if Person corresponds (as suggested above) to the higher occurrence of n, the one that reprojects out of the relative clause, the existence of Person (a higher n) entails the presence of C (the head of the relative clause out of which n reprojects). If PersonP is really the reprojected nP, it follows that DP-internal concord cannot involve person agreement, a robust generalization (see Baker (2008)), one that is in fact often taken to militate against a single mechanism for concord and agreement.

In the context of Person and a second occurrence of n (over D/P), it is worth noting the presence of a P-element (such as Spanish *a*) in the context of [+person/animacy] objects. Martín (In progress) in fact explores the relationship between Datives and Locatives, and finds the presence of locative elements in a variety of person contexts, such as person inflection (Martin suggests that the final *y* of Spanish 1^{st} person singular verb forms like *soy* '(I) am', *voy* '(I) go', etc. is really a locative element, cognate of French *y* and Catalan *hi*), clitic composition (Latin 1^{st} person singular dative clitic *tibi* as *t-ibi* 'you-there') clitic syncretism (Catalan locative *hi* used in dative contexts), and so on.

It is worth noting in the context of the present proposal that the representation in (12) captures the well-established fact (see Heine and Kuteva (2007)) that determiners tend to grow out of (grammaticalize out of) demonstratives, which in turn tend to emerge from locatives. If D is C, and C is just a high occurrence of P, then D is ultimately P, a locative. Leu (2008) has recently formalized this link between demonstratives and locatives by taking D(em.) to contain a locative element (which Leu takes to bottom out into an adjective, i.e., a PP for me). Leu's representation

[51] On NP-reprojection, from a different perspective, see Georgi and Müller (2010).

is reproduced for *this man* is as follows:

(13) $[_D$ the HERE man]

Leu in fact suggests that phrases like *this man* contains a two-DP layer (roughly: [[*this* HERE] THE *man*], with HERE and THE phonetically null in most, but not in all languages. One could straightforwardly adapt Leu's proposal in the present framework, with the locative element forming a low occurrence of P/D, and the definite determiner forming a high occurrence of *n*. Treating (some) determiners as the spell-outs of the higher *n* may help us understand why in some languages classifiers (traditionally corresponding to a low projection, the most natural candidate being *n* in the present context, given that classifiers are light nouns) can function as high (definite) determiner categories (see Cheng and Sybesma (1999, 2005), Simpson (2005)). The present proposal in fact anticipates differences among so-called classifier languages, which appears to exist (see Jeong (In progress), Ochi (2010), Jenks (2010), Saito et al. (2008))—an issue I return to in the section on parameters (building on Jeong (In progress)), as Jeong (In progress) shows that it bears in important ways on the issue of DP/NP-languages raised in recent works by Željko Bošković (see, e.g., Bošković (2005), Bošković (2008)).

4.6 Solving the selection problem

In the previous section, I noted that the reprojection of *n*P over DP could solve the selection problem that has plagued the DP hypothesis since its original formulation (Abney (1987)). It is in fact tempting to extend the reasoning to all the selection-related problems caused by the proliferation of functional categories, particularly in the context of cartographic proposals.

The problem is obvious, but rarely addressed in the literature. The only exception I know of is Shlonsky (2006), who discusses the uneasy relationship between cartography and minimalism (on the latter, see Boeckx (2008b), Rubio (2010), Boeckx (2010c)). Let me quote the relevant passage in full:

> The maps drawn by Cartography require some rethinking of the traditional division of the clause into a v/VP, TP and CP domains. This is particularly relevant in light of the relatively simple structures that Chomsky's Minimalism works with. In part, this is a division of labor: Minimalism focuses on mechanisms of computation (Merge and Search) and the role of uninterpretable features, while the cartographic enterprise is primarily concerned with the inventory of interpretable features.

Hence, under this view, minimalism needs an abbreviated structure, the C-T-v-V system, while cartography explores the full representation (see (Chomsky (2001, n.8)) for a comment to this effect).
In practice, however, Minimalist research has adopted the C-T-v-V system not merely as an "expository convenience" (Rizzi (2004, 7)), but as a substantive hypothesis for clause structure. The tension between Minimalism's impoverished structures and the richness of cartographic representations is a real one.
Thus, phases (CP and vP) and their edges (i.e., their heads and specifiers) play a key role in the computation of locality in minimalist syntax. It is far from clear how to integrate these notions into the structural maps of cartography, in which the clause is typically seen as a homogenous hierarchy of projections. In Cinque's system, for example, T dissolves into two distinct projections (Past and Future). Each should, in principle, have a specifier but which one corresponds to T? Similarly, what does "little v" [v] correspond to in a cartographic articulation of lower aspect and event-type? Which one of these lower heads should be taken to constitute the edge of vP? The problem is just as acute in the CP domain, where the edge of CP is its (outer) specifier, but in a cartographic perspective, should it be equated with Spec/Fin, Spec/Force or perhaps Spec/Focus (which, according to Rizzi (1997), hosts wh operators)?
Perhaps an even more nagging problem is that of selection. Minimalism inherits from previous approaches the view that selection is carried out under sisterhood. Thus, C selects T and V selects C. How is selection satisfied in e.g., an indirect question, if the head bearing the interrogative feature is Foc or Int (cf. Rizzi (2001)) and thus not a sister to V? Or take the familiar problem of how subjunctive features on an inflectional head can be selected by a higher predicate, given the number of intervening heads between V and the relevant mood head? The locality of selection plays a major role in Chomskys most recent work (Chomsky (2008); see also Richards (2007)), in which C transfers Case features to T. Which C? Which T?
The desirable goal of integrating the research agendas of Minimalism and Cartography requires, so it seems, modifications in the way structure, in the cartographic sense, is manipulated by the computational system. One possible direction would be to formally implement the notion of "abbreviated structures" by construing the cartographic structures not as a homogeneous hierarchy of equiponent projections, but as a structure com-

> posed of delimited "domains" or "extended projections", in the sense of Grimshaw (1991), a vP domain, a TP domain and a CP domain. Such a delimitation of structure is necessary for good empirical reasons, as well, since one needs to explain not only the clustering of similar features in the structure but also the delimitedness of verb-movement, NP movement and other operations which depend, in Minimalism, on uninterpretable features like Case or agreement. Why, one may ask, do these features typically fail to extend to the heads of left periphery? Cartographic works have, for the most part, implicitly assumed delimited structures or spaces but have not provided a formal implementation of domains.

Shlonsky's remarks are spot on. Take the nominal domain. Cartographic arguments have led to the explosion of the traditional NP into numerous projections, all dominated by a D-like layer (say, DP for the sake of concreteness; other labels would do just as well). This resulted in the material heading the traditional NP (the substantial Noun) being related to the most deeply embedded projection inside the DP, and yet everyone agrees that of all the elements contained in DP, it is the N that must be accessible for selection by, say, a verb. This is as clear a locality problem as one can find in syntax. Moreover, as Shlonsky correctly points out, cartographic representations have led to the scattering of traditional heads into several single-feature-based projections, making it hard for a particular head to be selected, for what one wants to say is that it is an ensemble of heads that is selected, a "domain", as Shlonsky calls it.

I would like to propose that all such selectional problems be solved by means of reprojection (head-raising analysis of relative clause), and that the multiple occurrences of the (re)projected head delimit the size of a domain. Put differently, I propose we treat the potentially offending material for selection as being contained inside a relative clause, off the main projection line. To give an example, in addition to my reformulation of the DP-hypothesis, I would like to argue that when CP splits into ForceP and FinitenessP (i.e., two occurrences, and subsequent functional specializations, of CP), it is indeed, as Rizzi (1997) conjectured, due to the activation of the Topic/Focus field. The latter I would like to treat as relative clause material. Interestingly, Irurtzun (2007) has independently argued, following interpretive-baed proposals by Herburger (2000), in favor of Focus-activation forcing reprojection of FinitenessP. Irurtzun's specific analysis is a bit too complex for me to go into here, as his treatment of focus, following Herburger, is not standard (though fully compatible with the interpretive properties discussed in section 7 [Part C]), and would require too much of a digression. I merely want to note that Irurtzun offers an explicit rationale for why activation of the Topic/Focus field (Rizzi's insight) may lead

to reprojection. I expect that other instances of split categories (e.g., the discussion of the splitting or bundling of vP and VoiceP in Pylkkänen (2008, chap.3)) will be linked to the presence of intervening material that 'forced' reprojection. (Once again, let me stress that in a Merge α framework, nothing is forced. Failure to reproject will lead to selectional problems and filtering at the interfaces.) I also therefore predict that somewhere inside the intervening material one is likely to find relative-complementizer-like elements (ultimately, adpositions). This may well be how applicative morphemes, topic markers, and the like should be treated.

Splitting of a syntactic category via reprojection leads us to expect that languages that consistently do so will develop distinct morphological markers for the various occurrences of the category in question, much like the various occurrences will acquire distinct interpretive roles. At the interpretive level, we expect these conceptual addresses to reach below what one may call Spelke-Object or cover level, and force the activation of finer-grained conceptual distinctions: Person and Gender as opposed to Class. Ditto at the morphological level, where finer-grained morphological classes may fossilize (grammaticalize) over time. This ties in with the discussion in Béjar (2003), where values/attributes of higher-order feature types (say, Person and ϕ, respecitvely) are not treated as different, ontologically speaking. Attributes and values are simply finely-cut features.

Before closing this section, I would like to go back to the notion of domain brought up by Shlonsky. This is a topic I already addressed in Boeckx (2008b, chap.4), and that I would like to return to. In Boeckx (2008b) I noted that at both the paradigmatic and syntagmatic levels, the various features that head the fine-grained cartographic representations form recognizable informational/conceptual clusters of a limited size, typically forming triplets (± 2). Think of the ranges of person, spatial deictis, tenses, illocutionary moods, evidential categories, and so on. All these domains recall the argument structure templates (limitations) discussed by Hale and Keyser (1993). In Boeckx (2008b) I speculated that the limited size of these domains was not a fact about Narrow Syntax, but a constraint imposed by the external systems, and how they chunk (categorize) information transferred to them (an idea adopted in Speas (2010)). The limited size of these domains (which I extended to the structure of chains) must be telling us something about the representation power of memory systems, which appear to follow a very general beginning-middle-end structure (hence the pervasiveness of triplets). The informational diversity of these triplets appears to be the result of a very general, semanticallly/conceptually improverished structure connecting to different core knowledge domains (modules). This proposal ties in with Harbour's program (see Harbour (2006, 2009)), who has argued that aspectual systems, number systems, pronominal series, and more can be explained solely on the

basis of very primitive (and, by hypothesis, non-human-specific) concepts like singularity,[52] addition, and augmentation, applies to the structures generated by Merge and restricted by the representational power of memory systems. Importantly, for this program to have a fighting chance, it is imperative to regard linguistic categories (number, person, case, etc.) as superficial manifestations of far more abstract, and conceptually more general categories. In Harbour's terms:

> a. In the domains I've explored (person, deixis, location, number), there turn out not to be any person/number/... features per se, but, rather, that the relevant features have a more general semantics of which person, etc., represent a usage that arises under a particular lattice embedding. (I.e., there's a degree of semantic underspecication in the predicates that the features denote, which permits the features to be applied to different semantic domains.) Particularly interesting case: there's near identity between aspectual features of Krifka (1992) and the 'number' features, and 'locative' features.
> b. Language design exhibits a type of economy envisaged by the Minimalist Program: not only do such traditional grammatical labels as inclusive and exclusive (dual and trial) dissolve into combinations of more abstract features (no news there), but so too do the categories of person and number themselves.
> c. [Feature] Geometries are otiose. Harbour (2006)

As I suggested above in the context of Chomsky's notion of the "duality of semantics", it is becoming clear that much of the richness, diversity, and specificity of our mental ontology should be approached in the way Darwin understood the richness, diversity, and specificity (adaptiveness) of life (bio-physical ontology): as the result of descent with modification (duplication and divergence) from primitive forms. Paraphrasing him, one could say that from so simple conceptual beginnings, endless features most beautiful have evolved.

4.7 Unvalued 'Features'

In previous sections I have referred to objects like "unvalued features" and processes like "Agree" and "valuation", despite the fact that I have insisted from the beginning on the need to resist 'feature-talk' as much as possible when discussing the workings of narrow syntax. The fact that I have found it impossible to avoid using terms like "unvalued features" and "Agree" is indicative of the centrality of featural transactions

[52] On the non-human-specific character of the concept 'singular', see Barner et al. (2008).

in minimalist syntax. It is now incumbent upon me to show how the effects of Agree and the role of unvalued features can be assimilated in the present, feature-free model.

The first thing to bear in mind is that the seemingly innocuous notation *u*F ('unvalued F') is an abbreviation for a feature lacking a value ([–F]), which is a property of a lexical item that (at the very minimum) bears at least one valued feature. So, when one talks about a lexical item X bearing *u*F, we should have the following (minimal) structure in mind:

(14) $\mathrm{X}_{uF} = \{\{\text{–F}\}, \{\text{+F}\}\}$

But what is, really, *u*F? It's a feature lacking a value, lacking content (hence it's lack of interpretation at the semantic interface). I propose we represent this by means of an empty set: {{}F}. Now, in a model like the present one, the most plausible interpretation of this structure is for F (a functional element, such as Case, or ϕ) to correspond to a phase 'head', taking as its complement an empty set, instead of a root (hence the lack of content). Put differently, I am proposing we take unvalued features as functional categories lacking a (contentful) root.[53] Moreover, I suggest we follow Chomsky (1995) in taking unvalued material to be added to a lexical item. Now, Chomsky takes this addition operation to take place upon selection, at the level of the numeration. Since there is no numeration in the present model, I propose we take this addition operation to be a case of syntactic adjunction, taking place not in the "lexicon", but in the course of the syntactic derivation.

Contrary to Pesetsky and Torrego (2007) and Bošković (In press), I believe it is important to maintain the link between pre-syntactic valuation and interpretability, so I reject outright the possibility of there being "lexically valued but semantically uninterpreted" features active in narrow syntax (such as Gender for Bošković (In press)). But given that I recognize the existence of a rich post-syntactic "lexicon" (morphological component à la DM) I allow for the possibility of late (that is, post-syntactic) feature (i.e. category) insertion, which will by definition mean that these features/categories will not be interpreted, since they are not part of the computation feeding semantic interpretation. At this point it is perhaps important to stress that the term "uninterpretable" is perhaps not the most adequate (although it is now well-established), as Brody (2003) was the first to recognize: all features are grounded, hence interpretable, but lack an interpretation in a given context: Class features are interpretable on nouns, but Gender features are typically purely mor-

[53]The move from feature to category is axiomatic in various current approaches to morphology, such as Starke (2010), Manzini and Savoia (2005), Kayne (2005), and is quite natural in the context of, e.g., clitics, which are nothing but ϕ-features functioning as independent syntactic categories.

phological. There is no need to identify a class of strictly uninterpretable features. What is needed is a way for the syntax to provide the right instructions to the external systems. If one does not want the (neuter) class/gender feature on (say) German Noun *Mädchen* 'girl' to be interpreted, we have to make sure that the relevant feature/category is inserted post-syntactically, otherwise the semantics will blindly follow the interpretive path carved by the syntactic computation, and will seek to assign interpretation to all the categories it is being fed.

Pesetsky and Torrego (2007) also address the possibility of features lexically unvalued yet interpretable (in a sense Chomsky (2008)) himself discusses this possibility in the context of T, and its (Tense) interpretation in the context of C[54]). For me these are roots, which lack interpretation in isolation, and only acquire identity in the course of the derivation.

In sum, the double dissociation that Pesetsky and Torrego (2007) seek to establish between interpretation and valuation is not enough for us to give up Chomsky's (2001) conjecture ("uninterpretable iff pre-syntactically unvalued"), especially if we recognize the possibility of late insertion, and (contra Pesetsky and Torrego (2006)) do not enforce a "vehicle requirement" (feature checking) on Merge (which leads to the proliferation of syntactically active feature, and the need to enrich the typology of syntactic features).

Let me now turn to a few architectural consequences that follow from the existence of "unvalued features" (which one may characterize as rootless phase heads). The first one I want to discuss pertains to an operation that Chomsky (2008) called Feature Inheritance, which I discussed briefly above. Feature Inheritance is the name of the process whereby in the course of Transfer (and valuation!), the unvalued feature on the phase-head (probe) is inherited by the non-phase head sitting next to the goal in the complement domain of the phase (the domain that is transferred). Chomsky (2007), Richards (2007) claim that Feature Inheritance forces Transfer to take place "as soon as possible", that is, immediately after the introduction of the unvalued feature.[55] In other words, the presence of unvalued feature on the phase head imposes a Spell-Out pattern known as "PIC_1" (so-called because it corresponds to the first definition of Cyclic Transfer put forth by Chomsky, viz. Chomsky (2000)):

(15) *Phase Impenetrability Condition* ("PIC_1"; Chomsky 2000, 108)
Spell-Out the Complement of Ph(ase) as soon as Ph is completed

[54]Another example that comes to mind is anaphors, although I confess to not being clear about how binding relations are established syntactically.

[55]This timing corresponds to the conception of the Strict Cycle that is known as the "virus theory". For relevant discussion, see Uriagereka (1998), Bošković and Lasnik (1999), Lasnik (2006).

In the literature, "PIC_1" is often contrasted with "PIC_2", the definition of Cyclic Transfer put forth in Chomsky (2001):

(16) *Phase Impenetrability Condition* ("PIC_2"; Chomsky 2001, 14)
Spell-Out the Complement of $Ph(ase)_1$ as soon as Ph_2 is merged

The difference between the two version of Cyclic Transfer pertains to when the complement domain of a given phase is transferred: "as soon as possible" or after the next phase has been introduced. As It happens, the present model, with its treatment of "unvalued features" as "rootless phase heads" allows one to eliminate the distinction between the two versions of PIC, for if unvalued features really are in fact (rootless) phase heads, the adjunction of unvalued features on a phase head boils down to the insertion of a new phase head, meaning that the spell-out pattern known as "PIC_1" is sort of an optical illusion: it really is a "PIC_2" pattern, albeit one where the introduction of the higher phase head follows so closely upon the introduction of the lower phrase head that the distance is barely perceptible, resulting in the impression that spell out of the complement domain of the lower phase is immediate.

The effect shared by "PIC_1" and "PIC_2" should be clear: both seek to keep to a minimum the size of the workspace at any given time. As the number of phase heads increases, the size of the workspace grows. As the definition of "PIC_2" makes clear, Cyclic Transfer is a way to avoid keeping in active memory ("workspace") too many levels of embeddings: Let's eliminate (transfer) the complement domain of Phase head #1 since the presence of Phase head #2 entails the presence of another, more encompassing complement domain.

With this in mind, consider the effect of the presence of unvalued features (understood as phase heads (of a special kind, viz. rootless)). As soon as unvalued features are added, the workspace contains two consecutive, adjacent phase heads, and (by the very same logic as "PIC_2"), it must discharge (/transfer) some material to keep the size of the workspace to a minimal constant. Now we are in a position to understand the process of Feature Inheritance. Contrary to what is being claimed in the literature, it is not necessary to view Feature Inheritance as a completely new process (nor is it correct to view it as an instance of Feature sharing, contra Gallego (In press)): feature inheritance is simply a way of discharing (transferring) material that overwhelms the workspace: it is simply a process of re-adjunction (re-attachment), from the phase head to an element that will be transferred. Which element? It can't be the goal, since unvalued features and valued features, upon valuation, will be indistinguishable at PF; accordingly, adjunction of unvalued features to the goal would lead to an *XX violation (Richards (2010), van Riemsdijk (2008)). Adjunction must be to the root, which is distinct from the unvalued feature, and will stand in

between the unvalued features and the goal, allowing the external system to keep them distinct.

Importantly, nothing of the above must be stipulated. The PIC_1 pattern of Transfer follows if PIC_2 holds and if unvalued features are treated as phase heads of sort. Feature Inheritance does not require a new process, not do we need to specify which element inherit the unvalued features. At each point if "wrong choices" are made, the derivation will fail to yield a convergent output post-syntactically.

As the reader may have noticed, any sort of adjunction of a phase head to another phase head will almost automatically lead to a PIC_1 Transfer pattern, a conclusion anticipated by Raposo (2002), who proposed that adjunction triggers immediate transfer of the adjoined material. We will see in section 4.10 below that this property shared by unvalued feature domains and adjunction domains may provide the basis for a possible unification of the two opaque domains that Huang's (1982) "Condition on Extraction Domain" (CED) generalization sought to capture.[56]

Let me conclude this section on unvalued features by highlighting two architectural consequence of the existence of agreement. First, the phenomenon of agreement really is inevitable in a Merge α framework, as merging an element with the empty set (member of any set) cannot possibly be triggered. In this sense, the present approach provides a much better rationale for the existence of unvalued features. Second, given that more material must be transferred in the presence of unvalued features, the phenomenon of agreement requires the independent existence of transitive phases (i.e., it requires the development of adpositions, or verbs as categories). Put another way, it requires more computational memory than the existence of nouns impose. In other words, the emergence of agreement requires a certain degree of grammaticalization.

4.8 Remarks on cliticization

In the previous section, I wrote, following Chomsky (2007), Richards (2007), that feature Inheritance forces Transfer to take place "as soon as possible" (imposing a PIC_1 Transfer pattern), but as always in the present monograph, the term "force" should be understood as a descriptive shorthand for "is necessary to achieve a particular outcome post-syntactically," but without intending any look-ahead. Strictly speaking, the presence of unvalued features/feature inheritance does not force any particular Spell-Out pattern; it only imposes a PIC_1 pattern if one wants to preserve the link between lack of value and lack of interpretation conjectured in Chomsky (2001). There is, I think, no debate to be had about the necessity of this link in the

[56]Crucially, for this unification to be possible, the role of "agreement"/unvalued features must be taken into account (which Huang did not do), as I stressed in Boeckx (2003, 2008b).

context of agreement: we do not want agreement markers to be interpreted. Or do we?

I can think of at least one context where we would not necessarily reject the possibility of interpreted "agreement markers" or unvalued features, viz. clitics. There is a long tradition within generative grammar treating clitics and agreement markers as a natural class. Recently, Roberts (2010) gave a new twist to this tradition. Roberts's point of departure is the problematic status of head-movement (of which cliticization is a special instance) within minimalism. For various reasons that I won't review here, head-movement as we know it is at odds with several minimalist guidelines, so much so that some authors, including Chomsky and myself (see Chomsky (2001), Boeckx and Stjepanović (2001)), have argued that head-head dependencies don't belong to narrow syntax. However, Roberts points out that, according to Chomsky (2000), Agree (feature-valuation) is a head-head dependency that is recognized as syntactic in that work. True, it is not a head-to-head movement dependency, but Roberts observes that some analyses of feature-valuation as feature-sharing or feature-copying make the outcome of Agree quite similar to movement viewed as remerge or a merge + copy operation. What keeps Agree and head-movement distinct under these analyses is that feature-valuation targets a subset of features internal to a lexical item, whereas head-movement targets the entire feature bundle. As Roberts correctly notes, in the case of featurally-impoverished categories (for example, in situation where an element consists of unvalued features and nothing else), Agree and head-movement may become indistinguishable as the target of feature-valuation would amount to the feature feature bundle). According to Roberts, this situation obtains in the case of clitics, which have standardly been treated as radically impoverished categories.[57] For clitics, then, the mere fact of entering into an Agree relation may give rise to the illusion of syntactic (head-)movement.

Setting aside empirical issues pertaining to specific cliticization patterns, Roberts's account strikes me as highly desirable, for it would, for the first time ever, immediately provide a non-stipulative rationale for why clitics, as opposed to their non-clitic nominal counterparts, never fail to move.[58] Whereas this asymmetry has always been

[57]Roberts discusses other cases of featureally-deficient categories, some of which strike me as less compelling than cliticization. One situation worth looking into, and which is not discussed by Roberts, is anaphors: if anaphors are as featurally impoverished as many have argued over the years, perhaps they too will pattern like clitics. This does not seem implausible: the link between reflexives and clitics has been discussed since at least Chomsky (1986a), and recently, Reuland (2001) has offered a comprehensive treatment of anaphora in terms of valuation/agreement.

[58]In work in progress, I also argue in favor of Roberts's analysis in the context of well-known clitic-cluster constraints, such as the Person Case Constraint: Roberts's account predicts the formation of clitic clusters if we accept the idea of Multiple Agree. But such clitic clusters will only obtain

problematic for frameworks that take movement to be driven (what is it about clitics that force them to move?), Roberts's analysis immediately explains it: it is not that clitics move, it is just that due to their featurally impoverished status, whenever they agree (something which by hypothesis they share with their full nominal counterparts) they necessarily give rise to the illusion of movement at the PF-interface, fooling PHON as it were.

The problem for Roberts's analysis is that it was formulated before theoreticians recognized the need for feature inheritance in the context of agreement, and when the concept of inheritance is taken into account, Roberts's intuition runs into technical problems, threatening its adequacy. For one thing, if the agreement marker that is to become a clitic is inherited from the phase head to some element inside the complement of the phase being transferred, the clitic should not be interpreted, which runs counter to the subtle, but real semantic effects of cliticization (Uriagereka (1995)). For another, if the clitic is spelled out in the same phase as the goal that valued it, there is bound to be a spell-out problem of the sort Richards (2010) has exemplified: two elements of the same kind must be separated by a phase boundary. The latter problem would not arise if feature inheritance had not taken place, as the clitic and its valuer would be separated by a phase boundary (assuming the clitic to have been adjoined, like other instances of agreement markers, to the phase head; see Boeckx and Gallego (2009) for empirical reasons in favor of this, based on the distribution of citics in a variety of contexts in Romance). Likewise, the interpretive problem would not arise, as it is inheritance that ultimately serves as the instruction not to interpret the agreement marker.

As we can seen, if feature inheritance is not forced,[59] we can maintain Roberts's insight and regard cliticization as the illusion of movement under Agree at the PHON interface (presence of the valued agreement marker at the edge of the phase will lead to deletion of the 'lower copy' (the goal), and will also lead to the interpretation of the clitic). The optionality of inheritance, predicted under Merge α also receives em-

if, as it were, both goals can value non-overlapping parts of the agreement marker, a condition I relate to the idea of Distributed Deletion explored for movement in the context of the copy theory (Fanselow and Cavar (2002)). I show that this non-overlap condition accurately describes known clitic cluster restrictions.

[59] Although I have insisted on Feature Inheritance being optional, some readers may have noticed that I argued in a previous section that agreement induces a PIC_1-Spell-Out pattern (through Inheritance) to keep the edge of the phase simple (avoiding the presence of two consecutive phase heads). But if inheritance does not take place under cliticization, how can we simplify the edge of the phase and avoid the derivation continuing with two phase heads adjacent to one another. They key observation here is that valuation splits the phase-head sequence (filling in the space left open by the absence of a root associated with lack of value), allowing the derivation to continue.

pirical justification (it predicts the existence of this phenomenon called cliticization). And, finally, both proponents and critics of the link between clitics and agreement markers can be said to be right: the same mechanism underlies both cliticization and agreement, but the optionality of feature inheritance can be used to keep both phenomena apart.

The present account of cliticization has a few consequences that are worth highlighting. First, all else equal, it predicts that agreement markers interpreted as clitics will tend to be linearized in a higher position than 'regular' agreement markers are. Empirically, this appears to be correct, especially in the context of procliticization. Moreover, this prediction receives support from the growing number of studies that indicate that clitics, unlike agreement markers, are more vulnerable in situation of agrammaticism (see Martínez-Ferreiro (2010)), often interpreted as the result of higher portions of the syntactic representation being more fragile than lower portions (the so-called Tree Pruning hypothesis; see Friedmann (2002), Martínez-Ferreiro (2010)).

Second, the present account predicts that if for some independent reason feature inheritance were forced to take place, cliticization would be ruled out, and only agreement would be possible. One domain where the prediction appears to be correct is adpositions. As Abels (2003) demonstrates, adpositions are domains that resist cliticization of their complements. Witness the following example from Serbo-Croatian (reproducible in many languages):

(17) Prema *joj/njoj trče
Toward her$_{clitic}$/her$_{pronoun}$ run
"They are running towards her."

Abels's explanation for this fact appeals to Anti-locality: assuming that there is something that forces clitics to move, and assuming that adpositions are phases, clitics that are complements of adpositions are ruled out, as their clitic nature forces movement, but since the phasal nature of P imposes a movement step to the edge of the phase ("PIC"), cliticization is ruled out as this would be an instance of movement from the complement to the specifier of the same head, a derivational step that Abels argues is illicit ("Anti-locality"). Abels's explanation is unavailable in a Merge-α framework, where no movement is deemed illicit in principle (more on this in section 4.10). But the present approach to cliticization can capture the generalization Abels was concerned with, once we recognize that the domain Abels concentrates on is one of adjunction. If, as I claimed above, the entire adjoined structure is transferred at Spell-Out (see also Raposo (2002), Uriagereka (1999)), this will in effect cause the clitic/agreement-marker to be inherited, as the phase head hosting the marker will be

transferred along with its complement. As a result, the clitic/agreement-marker will find itself in the same transferred domain as its valuer, leading to a case *XX-filter violation. The ban on cliticization discussed by Abels thus falls under Anti-Identity Avoidance, not under Anti-locality.

4.9 On the "obligatory" nature of some instances of displacement

A leitmotiv of the present work is that all syntactic operations are optional ("Merge α"). If this line of inquiry is adopted, it begs the question of what it is about certain instances of displacement that make these "obligatory." For many cases, it will be possible to say that if the relevant element had not moved, some post-syntactic filter would have been violated. Ideally, such filters should be independently motivated. Such a motivation is particularly difficult to find for those instances of movement that have failed to receive an interface explanation over the years. Perhaps the best-known (and most recalcitrant) illustration of this is the so-called original "Extended Projection Principle (EPP)".

It is well-known that natural languages impose a structural requirement forcing the presence of "subjects"; in more technical terms, the specifier positions of certain functional heads such as T_{finite} must be filled. This observation is illustrated here:

(18) It seems that John is here

(19) *___ seems that John is here

No one (within a broad spectrum of theoretical persuasion) denies that the EPP captures some true property of natural languages like English (perhaps all natural languages at some level of analysis). Put differently, no one seriously denies that there are EPP-effects. However, over the past ten years or so, several authors (see Castillo et al. (1999, 2009), Epstein and Seely (2006), Boeckx (2008a), Bošković (2002)) have questioned the theoretical status of the EPP as a principle in its own right.

Conceptually, no one can disagree with these authors: the EPP is not explanatory, and has never been. Despite its name, it is not an extension of anything (certainly not in a minimalist context, which dispenses with the projection principle). It is a mere description. To the extent that one is interested in pursuing a minimalist program for syntactic theory, one that asks why the language faculty is the way it is, the EPP ought to be abandoned as a primitive principle or feature.

This, of course, is easier said than done. After all, the explanatory weakness

of the EPP hadn't escaped GB-practicioners. Fukui and Speas (1986) already proposed that the effects of the EPP follow from a more general requirement, that a Case assigner must assign/check its Case (what Howard Lasnik aptly called the Inverse Case Filter). Under this view, EPP violations are the result of the relevant functional head failing to assign its Case (or have its Case feature checked). Within the minimalist studies cited above, Case has also been recruited in some fashion or other to account for the EPP (see, e.g., Epstein and Seely (2006), Bošković (2002)). But there are reasons to doubt that Case considerations alone are behind the EPP. After all, what is it about Case that forces displacement? The question is particularly salient in a model that recognizes the existence of Agree, which makes it possible for features (including case and phi-features) to be checked at a distance, without actual displacement. Invoking Case is very tempting indeed, as Case has been used since Vergnaud's 1977 influential suggestion to account for the distribution of overt DPs. But despite its impressive empirical coverage, the nature of the Case Filter, and Case more generally, remains poorly understood. As a matter of fact, Case stands out as a sore thumb in current minimalism: it is the only feature that is uninterpretable on the Probe and on the Goal, the only feature that is never interpreted at the interfaces on any element. Clearly, this begs the question of why Case should be part of the system in the first place, especially if that system is optimally tuned to interface demands, as the (strong) minimalist thesis contends. Faced with the challenge posed by Case, various researchers have tried to 'eliminate' Case, by either reducing it completely to a morphological notion (McFadden (2004), Landau (2006), or by showing that Case is indeed interpretable (Pesetsky and Torrego (2001)), or by claiming that Case-checking is parasitic on agreement (Chomsky (2000)). The net effect of any of these approaches is to minimize the role of Case in syntactic derivations, making it virtually impossible to formulate an analysis where the (Inverse) Case filter causes displacement. (It is significant that Marantz (1991), the first approach to try to relegate Case to PF, replaces the syntactic role of Case by the EPP.)

The alternative I would like to entertain here is the idea that the EPP is the result of a conspiracy, caused by the fact that syntactic derivations are phase-based in the way I have explored in this work. Under the account to be developed in this section, the EPP is nothing more than the inevitable result of cyclic computation, and the requirement this imposes on feature-valuation at the interfaces.

Recall that in the presence of "unvalued features" (rootless phase heads), Transfer of the phase complement follows a PIC_1 pattern, where "Feature" Inheritance (re-adjunction of the rootless phase head), valuation (/probing) , and determination of interpretation are synchronized. In Chomsky's terms, these suboperations happen "simultaneously". Assuming this much, I would like to claim that the EPP-effect

(the seemingly obligatory nature of movement of a nominal element to the specifier position of T) is one of only two ways of achieving this synchronization of operation without violating any other filter. (The other way of achieving the same result is moving T outside of the transferred domain, which I discuss below.)

Let's examine why this is the case in detail. In Probe-Goal/valuation terms, the EPP boils down to the following requirement: at the point of transfer, the Goal must be the sister of the element bearing the feature(s) valued by the Goal; crucially, the Goal cannot be structurally lower than the element bearing the feature(s) being valued. I resort to the longer phrase "element bearing the unvalued features" instead of the Probe because, following Chomsky (2007, 2008), I take it that the real Probe is the phase-head. The element bearing the unvalued features at the point of transfer is not the Phase head (the Probe), but the head of the phase-head complement, the recipient of the inherited features, the Probe-by-Proxy, as it were. And this is crucial, I think. If we take the idea of simultaneity seriously, valuation is established in the presence of a "dual" Probe as it were: the original host of the unvalued features, and the Probe-by-proxy. To determine the correct interpretation of features, it must be the case that upon Transfer the external systems (especially, C-I) see this dual probe. This turns out to be what I claim causes the EPP-effect: upon Transfer this dual Probe situation gives rise to a Relativized Minimality situation, where three (crucially, not two) elements bear the same (relevant) feature set: the Goal/valuer, the Probe, and the Probe-by-proxy. Now, if by the time Transfer happens, the Goal had not exercised its option to move to the specifier position of the probe-by-proxy, the configuration at Transfer would be one where the Probe-by-proxy would disrupt the Probing relation between the original Probe (phase head) and the Goal: a straightforward case of Intervention/Relativized Minimality. If, however, movement of the Goal has taken place by the time Transfer happens, the minimality effect is avoided: both Probes can be valued at once. The EPP-compliant configuration is thus a way to avoid an instance of Intervention. Put another way, displacement of the Goal turns out to be (*a posteriori*!) the design feature required to avoid a minimality effect caused by the simultaneous character of Probing/Valuation and Inheritance.

Let me note that it is hard for me to see how the present analysis, which treats the EPP as a way to circumvent a minimality effect, can be reproduced in a non-phase-based framework. Ángel Gallego and Norbert Hornstein independently pointed out to me that López (2002, 2007) suggests an account similar in spirit to mine: for him, movement of the subject to SpecTP is a way to void the barrierhood of T. The intuition that locality is behind the EPP is indeed similar, but the two accounts are crucially different, in a minimalist context: whereas Relativized Minimality appears to be a well-grounded, "third factor" principle, I do not know of any non-stipulative

account of barrierhood. In addition, the reason why I can appeal to Minimality is the direct result of Feature Inheritance, an operation that is crucially tied to a spare use of "features" (a one-to-one correspondence between valuation and interpretation) and to phase-based computation. One could of course try to dispense with these and stipulate that C and T share features, but a cyclic-transfer and an impoverished pre-syntactic lexicon go further in deriving this fact. In this sense, I agree with Noam Chomsky's (p.c.) assessment that the analysis of the EPP presented here, if correct, strengthens the phase-based model, as the current analysis crucially relies on the cyclic nature of computation..

As I just argued, the EPP is to be understood not as a primitive or something special, but as another example of minimality circumvention.[60] In the literature on this topic, two kinds of circumvention strategies are often distinguished (see McGinnis (1998), Anagnostopoulou (2003), Bissell-Doggett (2004), Jeong (2007)): (i) movement of the goal around the intervener (what is known as "leapfrogging" or "hurdling"), and (ii) movement of the would-be intervener outside of the Probing domain.[61] Both strategies are schematically illustrated here: (All three elements represented are taken to share relevant features.)

(20) $[\alpha \ldots [\gamma \ldots [\beta \ldots [t_\gamma\ ]]]]$

(21) $[\beta\ [\alpha \ldots [t_\beta \ldots [\gamma\ ]]]]$

The EPP-effect discussed above relied on strategy (i) (leapfrogging). This is what gives rise to the appearance of obligatory movement of the Goal. But nothing excludes situations where strategy (ii) is at work. Either strategy is fine, the availability of each depending on the (post-syntactic) lexicalization resources of specific languages. The difference between the two is that in one case (strategy (i)), it is movement of the Goal that will appear obligatory; in the other (strategy (ii)), it is movement of the Probe-by-proxy that will have this character. For the purpose of the EPP, though, situations where strategy (ii) is at stake may have been described as situations where the EPP is not enforced, as the formulation of the EPP tends to focus of presence of movement of the Goal. But if we stop taking the EPP as a primitive requirement, and if instead we take the key requirement to be successful valuation, then the minimality effect caused by the Probe-by-proxy leads to the pre-

[60]It may be important to point out that such circumvention effects are quite local, limited (as we expect) to a single phasal domain, which fits well with Chomsky's (2001) idea that minimality is evaluated at the phase level, and also with the claim made here that phasal domains are those on which post-syntactic features—necessary for minimality—are grafted.

[61]Strategy (ii) is perhaps better known as the idea that "traces" don't block movement or don't intervene, an idea going back to Uriagereka (1988) and adopted in Chomsky (2001).

diction that some languages will not require the subject to move in case T raises to C.[62] This prediction corresponds precisely to the description given by Wurmbrand

[62]This may be a good point to note that although 'head-movement' is reinstated as an operation within narrow syntax under the present proposal, there is no room for a distinct head-to-head adjunction operation (the classic form of head-movement) here. In fact, the form of head-movement takes in a framework like this one is very similar to the one advocated by Vicente (2007), Matushansky (2006), where heads move (remerge) into 'specifier' positions like XPs do, with the difference between X^0- and XP-movement being a matter of morphology (essentially a matter dependent on the affixal status of the moved element and/or its host). The one important effect of allowing heads to move within narrow syntax is that in case of phase-heads moving their spell-out domain (complement) domain may get extended—what Gallego (2007b) dubbed "phase sliding", although unlike Gallego, I take all the consequences on this domain extension to be of a post-syntactic nature.

One interesting consequence of the present approach is that it may recapture an asymmetry that Lasnik (1999b, 2000) took to be established as a parameter in the pre-syntactic lexicon. I have in mind the famous French vs. English verb placement rules discussed by Emonds (1978), Pollock (1989) and many others since. The basic facts are well-established: French finite verbs appear higher than English finite verbs (except auxiliaries). To capture the basic asymmetry, Lasnik offered a complex, "hybrid" morpho-syntactic framework where, as a matter of parametric option, verbs may combine with their finite inflectional markers in the syntax (if verbs are "bare" in the lexicon), or else as a matter of feature checking (if verbs are "inflected" in the pre-syntactic lexicon). The present approach need not resort to such architectural complexities: all that needs to be said is that what Lasnik called "bare" verbs in English are roots ($\sqrt{V}$), whereas inflected verbs are lexicalized combinations of these roots with higher heads (υ) and T, made possible by the option of moving $\sqrt{V}$. Assuming that auxiliaries universally lexicalize υ (as proposed in Den Dikken (1994), Bošković (2001)), all auxiliaries in all languages will be lexicalized in the same transferred portion as T. There is no way for them to remain bare. Because English main verbs are bare roots within their spell-out domain, they depend on the operation of phase-chunk recombination to find themselves adjacent to their inflections. French verbs combine with their inflections much earlier in the derivation, as do English auxiliaries.

The present framework may be flexible enough to capture not only the main-verb vs. auxiliaries in English, or the French vs. English contrast, but also the well-known generalization that Italian verbs appear higher than French verbs do (Belletti (1990)). This latter fact led to a proliferation of landing sites for verbs in the IP field, but no such move is needed here. The contrast could fall out of the movement options left open under Merge α: once $\sqrt{V}$ has raised out of its origination site, it can remain as a sister of υ, or else both υ and $\sqrt{V}$ can raise past $\sqrt{T}$. The latter movement is what gives the Italian pattern. As in Cinque (1999), verb placement boils down to lexicalization choices.

The following summarizes the relevant structural patterns (with curly brackets indicating Spell-Out boundaries):

(i) $\{\sqrt{T} \ldots \upsilon\}\{\sqrt{V} \ldots\}$: English
(ii) $\{\sqrt{T} \ldots \sqrt{V}\ \upsilon\}\{t_{\sqrt{V}} \ldots\}$: French
(iii) $\{\sqrt{V}\ \upsilon\ \sqrt{T} \ldots t'_{\sqrt{V}}\ t_{\upsilon}\ \}\{t_{\sqrt{V}} \ldots\}$: Italian

(2006) and Haider (1993) for EPP-effects (and lack thereof) in German. I suspect that other languages (perhaps, many VSO patterns) will fall into this pattern.[63] In general, the present account leads to the (testable) expectation that if the subject of a finite clause (one where ϕ-features have to be valued on some high functional head like T) remains in situ (what is often described as lack of EPP-effects, or optional EPP-effects[64]), it is either because the Probe-by-proxy (ϕ-feature bearer) has moved beyond the position of the actual Phase-probe, or else (and this, of course, is where careful analysis is required) because the pronounced position of the subject is a lower copy (that is, movement is the subject took place within narrow syntax, but linearization obscures it).[65]

Let me conclude this discussion of the EPP by pointing out that the reasoning developed so far extends to all instances of Probing.[66] That is to say, there is nothing specific to subjects, C, and T. Put another way, if all the relevant conditions obtains, I predict the existence of EPP-effects with non-subject arguments. But I should stress one important point here, regarding objects. Following Chomsky (2007), I take it that object agreement is on a par with subject agreement in that it involves, in addition to a Goal nominal, a Probe (v for objects) and a Probe-by-proxy ($\sqrt{V}$). But there is an independent, but in the context of the EPP, very important difference between subject and object agreement/EPP-effects. In the case of subjects, the Goal originates in a projection lower than that of the Probe-by-proxy: there is a clear asymmetric c-command relation between the two. By contrast, in the case of

[63]See McCloskey (2001) on the issue of the EPP in Irish.

[64]I mention "optional" EPP effects because recall that no movement is obligatory in the present model, which also means that no movement is prohibited within narrow syntax. Accordingly, the subject is not required to remain in situ even if the Probe-by-proxy has moved, and there is no minimality effect left to circumvent.

[65]I suspect that this latter situation obtains in the context of so-called Locative Inversion in English, where lower copy pronunciation is likely due to discourse-prosodic factors. Such a masked movement of the subject in Locative Inversion contexts would account for why there is no EPP-effect in these contexts in a language where otherwise EPP effects are always felt, and also for why no *do*-support is necessary.

[66]The account could also extend to EPP-effects in the $\overline{A}$-domain (for *wh*, or some instances of complementizer agreement), of the sort discussed in Miyagawa (2010). These would require the existence of an extra phase (in effect a SplitCP). An analysis of this sort is developed in van Craenenbroeck and van Koppen (2002). Such a proposal could readily capture the oft-noted connection between the Force projection encoding [+Interrogative] and the lower FocusP hosting *wh*-phrases (see, e.g., Rizzi (1997), Aboh (2004)). This connection could be thought of as on a par with the C-T, or v-V pairs. If Irurtzun (2007) is correct in his analysis of restricted quantification over events, the Force-Focus pair is an instance of C-layer iteration ('reprojection'). A similar iteration (of the v-V pair) may be required in constructions involving complex tenses/auxiliary sequences, as in *John may have been swimming*, where agreement iterates in some languages (Carstens (2001)).

simple, transitive objects, the Goal and the Probe-by-proxy are merge-mates: they originate in the same projection, and are in a mutual c-command relation. This is crucial in the case of the EPP, which I claim reduces to a Relativized Minimality situation. The latter, ever since Rizzi (1990), is crucially defined upon asymmetric c-command relations. If there is no asymmetric c-command relation, no intervention effect is expected. As a result, in the case of simply transitive object, no minimality effect has to be circumvented, meaning that we will no detect any EPP-effects there. When the object originates in a projection lower than the Probe-by-proxy, we predict such EPP-effects for the same reason we predict them for subjects. The prediction appears to be correct (cf. Bošković (1997)): displacement of the object appears to be as obligatory as that of subjects only when the object starts very low, as is the case in ECM-contexts, ditransitives/applicatives, perhaps also in the case of some particle verbs, restructuring environments, and so on.[67] If this is true, the present account derives this fact straightforwardly.

4.10 Locality conditions

The approach to Narrow Syntax pursued in this work is one that is guided by a single key hypothesis: that Merge is essentially "free" (unconstrained). Any type of Merge can happen, to any element endowed with an edge property. A consequence of this perspective is that constraints in movement (what Chomsky (1973) called "Conditions on transformations"), typically thought of as part of Narrow Syntax, cannot in fact be narrowly syntactic. Internal Merge ('movement') is just as free as any other type of Merge. To the extent constraints on movement exist (and I take it that they do), they must be construed as arising from problems at the external systems, or as emerging from the way syntactic information is passed onto these external systems. I would like to explore the latter possibility here, stressing that the possibility for some constraints on movement to be of the former type is very real indeed. For example, (some aspects of) the Coordinate Structure Constraint may reduce to a Parallelism condition on conjuncts, as argued by Fox (2000), Kato (2007), and others. Likewise, many 'Relativized Minimality'/intervention effects analyzed in the literature require a fairly large inventory of semantic or pragmatic features (witness Starke (2001)). It is quite possible, therefore, that at least some of these effects come about once the features to which Narrow Syntax is blind become active (this is in fact one way of

[67] Lasnik (1999a, 2001a) suggests that even in those contexts, movement of the object is optional, but there are good empirical and conceptual reasons to doubt his conclusions, and to stick to his original (1999b) claim concerning the obligatoriness of the EPP for objects; see Boeckx (2008a), Bošković (2002) for discussion.

reading Szabolcsi (2006). In a sense, the position advocated for here is a return to the (minimalist) logic of Miller and Chomsky (1963): keep the theory of competence maximally simple, and find the source of unacceptability in the ways in which linguistic information is digested by the performance systems.

Quite independently of the set of assumptions made here, the recent literature on resumption and ellipsis (see Boeckx (2003), Merchant (2001) and references therein) suggests that it is simply false to hold onto the traditional slogan that movement cannot take place across islands. There is empirical evidence reviewed in the works just cited that indicates that movement has taken place across nodes that many would regard as defining island domains, with acceptability determined only once a variety of other, post-syntactic factors has been taken into account.[68] In a phase-based system, it is actually quite hard (short of *ad hoc* moves[69]) to prevent movement from crossing phases. Quite apart from the fact that constraining internal merge would depart from Merge α, some 'communication' must be allowed to take place across phases. This is the role of phase edges. But as soon as some movement is allowed to target phase edges (as it must, to capture the phenomenon of long-distance movement dependencies), this escape hatch can be used to bypass any blocking effect one may be tempted to encode in phase-heads.[70]

To the extent one wants to involve phases in some aspects of locality (and I agree with everyone else that the kind of elements assumed to be phasal matches the traditional island typology so well that it can't be an accident), the only natural option is to implicate phases indirectly. Dependencies across phases (via Internal Merge) can be formed, but perhaps something about some of these dependencies, *when spelled out cyclically*, causes the external systems to 'gag' (and in many cases, with no option for 'recovery'). I like to call this the logic of the garden path. As is well-known, the status of sentences like *the horse raced past the barn fell* is degraded.

[68]Notice that it would be wrong to conclude from this literature that island effects have nothing whatsoever to do with syntax. The most one can conclude from this is that 'bounding' nodes do not block movement, but it may well be the case that (for some reason to be made precise) movement across bounding nodes leads to unacceptability unless something else (deletion, resumption, etc.) happens (cf. Lasnik's (2001b, 2005) notion of 'repair by ellipsis').

[69]The clearest (but by no means the only) example of this is to be found in Müller (2010), where a very explicit way to capture CED-effects is proposed, which requires no less than all the following conditions: that (i) all instances of merge be feature-driven, (ii) all phrases be phases, (iii) features on lexical items be ordered (hierarchically organized), (iv) edge features to be added in the course of the derivation.

[70]I am referring to phase heads because these are the only heads that stand out in the sort of model pursued here, and also because many have pointed out that the typology of phase heads adopted in the literature is reminiscent of the traditional bounding nodes (see Boeckx and Grohmann (2007) among many others).

The standard explanation for this fact is that the reduction of the relative clause ('raced past the barn') causes the language processor to parse 'the horse raced past the barn' as a main clause, leaving 'fell' as a puzzling afterthought. Had the relative clause been spelled out in full ('the horse that was raced past the barn'), the full sentence would have been unproblematic. The hypothesis concerning (some) islands that I would like to entertain here is that the cyclic character of Spell-Out imposes a certain mode of processing syntactic chains (movement dependencies) that in some cases (i.e., when island effects obtain) leads to the perception, on the part of the external system, that the dependency has been completed, when in fact it hasn't. Put differently, cyclic spell-out can be misleading, and can lead the external system down the garden path, into an island.[71]

Irrespective of island effects, it is clear that Cyclic Spell-Out has an effect on how syntactic chains are processed. Take the now well-established[72] phenomenon of successive cyclic movement. Abels (2003) drew attention to two competing views on successive cyclicity: among those who assume that long-distance dependencies are the result of a conjunction of local dependencies, some hold that a moving element stops by each and every projection along the movement path, while others argue that the moving element stops by a proper subset of all the projections along the movement path, the subset being defined by some version of the bounding theory (see already Chomsky (1973)). Abels calls the first analytical option the uniform path hypothesis, and the second, the punctuated path hypothesis. Empirical evidence in favor of one or the other option is very hard to come by. As Abels notes, standard tests, such as presence of reconstruction effects in intermediate landing sites, is bound to be inconclusive since presence of reconstruction only indicates that some intermediate position acted as a landing site, but the sort of evidence we need to distinguish between the two hypotheses should be evidence that some position along the movement path was not used as a landing site. Abels argues that the absence of reconstruction effect in some position would be more revealing. Unfortunately,[73] this is only true if we assume that every copy left by movement can be reconstructed, which we know independently is untenable (see the discussion in Fox (2002), Takahashi (2010), Takahashi and Hulsey (2009) on the need for copies to be modified to feed reconstruction, and why some copies of some categories can't be so modified). So, the evidence in favor of the punctuated path hypothesis presented in

[71]It is well-known that repeated exposure to garden path sentences improves their status. Perhaps this is how one should understand the island-alleviating effects reported in the literature on islands (Snyder (2000), Hiramatsu (2000)); although see Sprouse (2007) for a dissenting view on satiation.

[72]For a review of the evidence, see Boeckx (2008d).

[73]As already pointed out in Boeckx (2008d), Boeckx and Hornstein (2008).

Abels (2003), Abels and Bentzen (2009), Lahne (2008), based on lack of reconstruction effects in some position, is inconclusive. It seems to me that in a framework that takes movement to be (internal) merge it would take an extra constraint to ban movement from not stopping by certain positions: internal merge should be as free as external merge. Uniform paths should not be forced, but they should not be ruled out either. That said, the idea advocated here that Transfer happens not at every Merge-step,[74] but in a punctuated fashion, may yield a reconstruction pattern that gives the illusion that paths are punctuated—not because movement has not proceeded in a uniform fashion, but because chains are read off in chunks.

Another case where an alleged derivational constraint may simply be the reflex on how chains are processed in a Cyclic Spell-Out model concerns Anti-locality. Various researchers (Bošković (1994), Grohmann (2003), Abels (2003), Kayne (2005), Boeckx (2008d), among many others; see Grohmann (In press) for a comprehensive overview) have appealed to a ban prohibiting movement to relate two positions within the same phrase (e.g., movement from the complement of X to the Spec of XP, or adjunction of the complement of X to X). Such a ban has no place in a Merge α framework, but its effects can be recaptured if we say (as seems plausible) that the external systems interfacing with syntax lacks the wherewithal to distinguish occurrences of the same element within a single phrase. If occurrences within a chain are defined relative to their sisters, and if phrase structure is bare (labelless), Complement of X and Specifier of X are indistinguishable. Similarly for Complement of X and adjoined to X. Internal Merge may take place, but chains read off at the interfaces will necessarily be trivial (one-membered). The only case where Anti-local movement may be detected is in those phrases where complement and edge of a phrase are separated by Spell-Out: through phasal nodes. Abels (2003) offers arguments against movement of the complement of a phase head (TP/VP) to the edge of the same phase, appealing to Anti-locality.[75] But the same movements can be ruled out in a different

[74]Dennis Ott correctly points out (p.c.) that the data discussed by Abels (2003), Abels and Bentzen (2009), Lahne (2008) is very problematic for advocates of a 'Spell-Out-as-You-Merge' model such as Epstein and Seely (2006).

[75]Abels (2003) extends his analysis to the adpositional domain, and resorts to Anti-locality and phases to analyze P-stranding, and lack thereof in many languages. If the present approach is on the right track, Abels' analysis of P-stranding can't be maintained. There is at least one independent reason for doubting Abels' analysis. To capture the ban on P-stranding in many languages, Abels is forced to say that adpositional phrases in such languages are structurally less complex than in English, but most of the evidence for complex functional structure in the literature comes from languages that ban P-stranding. As an alternative one might explore the possibility that P-stranding is possible only in those languages that allow the complement of P to value *u*F on *p* without displacement, whereas those languages that prohibit stranding are those where the object of the adposition has to move *p*P internally to value *u*F. This would mirror my treatment of

way: if pied-piping (understood as a PF-instruction to linearize as a group; a PF-effect at the level of phasal chunk recombination) requires a phase-head serving as an address (for constituency), (branching) complements of phases will be unable to be pied-piped. This approach appears empirically superior to Abels', as it predicts that complements of intransitive phase heads (n-heads), since they are non-brancing, will be able to move leaving the phase-head behind. Taking Root-movement to amount to syntactic incorporation, our analysis makes the apparently correct prediction that only Nouns can be the target of syntactic incorporation (Baker, 1996).[76]

The bottom-line of this discussion is that some instances of displacement/some occurrences created by Internal Merge may be invisible at the interfaces, due to the specific way Transfer works (Cyclic Spell-Out). Metaphorically speaking, in the case of successive cyclicity, it's as if the external systems caused a wave to be perceived as a series of particles—in the case of language, an internally merged element is perceived as a chain consisting of at least two copies. In the case of Anti-locality, it's as if the external systems were unable to see movement because every time they look the element seems to them to be back to square one. If this is on the right track, specific conditions on Internal Merge, as part of Narrow Syntax, need not be posited. Merge can retain its "unrestricted" character. It is this logic that I would now like to apply to core cases of islands.

The minimalist literature on islands is vast, and I will not attempt to review it here (Boeckx (In preparation)). I will instead focus on what is I think a growing consensus in that literature: that traditional CED-effects, previously unified under the umbrella of proper government (Huang (1982)), should not in fact be brought together: subjects (or displaced arguments more generally) and adjuncts pattern differently. The robust impression[77] gathered from decades of cross-linguistic and cross-constructional investigation is that extraction from adjoined domains is consistently bad, whereas extraction from displaced arguments is subject to variation, and in some cases perfectly acceptable (an impression reinforced by more recent studies indicating different psycholinguistic profiles for the 'subject island' and 'adjunct island;' see Sprouse (2007) for valuable discussion).[78]

---

the subject/object freezing asymmetry in text below.

[76]Alleged instances of Non-Noun incorporation must reduce to morphological compounding or pseudo-incorporation, as Baker himself suggests (Baker, 1996, 295). For relevant discussion, see also Kamali (2009).

[77]See Stepanov (2001a, 2007) for clear statements. See also Richards (2001, 187).

[78]To be fair, though, comparing the robustness of the subject island and of the adjunct island must, if it is to be compelling, ensure that the meaning of the term 'subject' in defining the 'subject island' is kept constant. After all, 'subject' is not a primitive notion in generative grammar: there are low (*in situ*) subjects, high subjects, agreeing subjects, non-agreeing subjects, and more.

The general strategy for dealing with the adjunct island in the minimalist literature is to say that adjoined domains are, at the point where movement is to apply (when the target of movement is introduced), not yet part of the workspace (Stepanov's (2001b) obligatory Late-Merge of adjuncts analysis), or occupy a different dimension (Chomsky (2004)), or lack a label that can be probed (Hornstein (2009), Hornstein and Nunes (2008)), or can't take part in an Agree-relation (hence preventing movement, if move requires Agree) (Boeckx (2003)), or some combination of these propositions. All of these proposals are intended to capture the intuition that adjoined domains are syntactically inert.

With regard to the subject island (or more generally, the displaced argument island), some syntacticians (Richards (2001), Boeckx (2003, 2008b), Rizzi (2006), Rizzi and Shlonsky (2007)) have sought to unify the difficulty of extracting (displaced) subjects (/arguments) (cf. *that-t* effects) and the difficulty of extracting from (displaced) subjects (/arguments), appealing to a version of the freezing condition (Wexler and Culicover (1980)). Though some advocates of this position take freezing to be a primitive (e.g., Rizzi's 'Criterial Freezing'), others have tried to derive the freezing effect some some other property of the grammar. Richards (2001) characterized as 'freezing' those positions that are "strong," where strength is understood as in the sense of Chomsky (1993): an instruction to pronounced the copy associated with that position. Once an element reaches a strong position, it can't move (and be pronounced) into another (equally strong) position. Boeckx (2003, 2008b) and Gallego (2007b, In press) propose to tie strength (and freezing) to case- and ϕ-feature checking.[79] Boeckx and Gallego point out that tying freezing to features whose distribution is known to vary cross-linguistically may account for why the subject/displaced argument island is not as robust as the adjunct island.

With this much backgound, let us now consider what the present framework could say about the phenomenon of islandhood.

Consider the adjunct island first.[80] No domain can be said to be adjoined 'pre-

Perhaps part of the cross-linguistic variability of the subject island is the direct result of the many meanings of 'subjects.' It is interesting to note in this respect that low adjuncts are not as impervious to extraction as prototypical (high) adjuncts (see Truswell (2007), Boeckx (2003), Uriagereka (To appear); see already Uriagereka (1988), Browning (1987)). It is also worth noting that the psycholinguistic evidence against the robustness of the subject island is still a matter of debate (see Phillips (2006), Jurka (2009)). Uriagereka (To appear) makes use of such considerations to advocate for a uniform treatment of CED-effects, of the sort he originally put forth in Uriagereka (1999). For a related (uniform) treatment of CED-effects, see Narita (2008). For different proposals to reunify the CED, see Boeckx (2008b), Müller (2010).

[79]On the freezing effect of Agreement, see also Donati (2006).

[80]Since virtually nothing is forced within Narrow Syntax in a Merge α-framework, it is impossible to force adjuncts (in the sense of optional material) to be introduced via Pair Merge. Merging

syntactically'. That is to say, no element heading the adjunct can be categorized as a bounding node inherently. Some researchers, such as Raposo (2002) have tried to claim that adjuncts are fully saturated/convergent expressions,[81] hence should be spelled out independently of the main brain of the derivation (a variant of Uriagereka (1999)), thereby trapping all the elements they contain. The idea of adjuncts being transferred as a whole is a good one, but I do not know of any compelling definition of what counts as "fully saturated/convergent."[82] Instead I would like to say that it is the operation of adjunction as such that renders an adjunct opaque. What is adjunction? Chomsky (2004) takes adjunction to be the formation of a Pair: $\langle \alpha, \beta \rangle$. We need not worry about the representational format of this pair. Adjunction, *qua* Pair-Merge, expresses the idea that a certain constituent (say β) is added to a structure already formed (say α). It is an asymmetric process, and as such, in accordance with the idea pursued here that the locus of Asymmetry is Transfer, adjunction must take place at the point of Spell-Out; that is, adjunction must take place at the phase-level,[83] and must involve a phase head, since we want to be appeal to labeled structures to be able to express that a certain kind of constituent β is adjoined to an already existing structure. This way of characterizing adjunction makes clear that the adjunct must be reduced to a 'point,' a unique lexical item, a simplex edge so that it (as a whole) is added to a set formed by Merge. This is, I claim, the true meaning of Pair Merge.[84] It is the forced reduction of an entire constituent to a single element that can be added to a structure built in parallel. (Since the adjunction process is asymmetric, no such reduction is imposed on the constituent being adjoined to.) Notice that there is no need for a special pair-merge operation; the immediate spell-out pattern forms a pair (an asymmetry). Merge itself remains

adjuncts is as optional as merging arguments—everything is subject to interpretive coherence. We therefore expect extraction from some adjuncts (those introduced below the phase-level) to be possible (an empirically desirable result). These low adjuncts, because they will be merged below v, will abide by Truswell's 2007 single-event restriction.

The idea that adjunction is not a uniform phenomenon, or that more than one representational format is needed (Pair Merge and Set Merge), is not unique to this analysis. For relevant discussion, see Hornstein and Nunes (2008), Hornstein (2009), Gallego (2007a).

[81]Richards (2001, 187) takes adjuncts to always be associated with a strong position, which is not only *ad hoc*, but also problematic since adjuncts as a whole can form chains, which Richards would (incorrectly) predict are too strong.

[82]Epstein (2003[2007]) can be read as making a similar point, in a different context.

[83]This may also account for late-merge effects associated with adjuncts (Chomsky 1993, Lebeaux 1988, Stepanov 2001b, and work on scrambling/extraposition)

[84]We can think of this as a way of cashing out Chomsky's (2004) SIMPL operation. The operation of feature inheritance discussed in section 3 is closely related to this analysis. Here the adjoined structure is pushed into the Transfer portion of the phase.

set-formation. The reduction of the constituent being adjoined means that all the elements, save the phase head, inside the adjoined phrase must be transferred prior to adjunction.[85] It follows that no extraction will be able to take place from within the adjoined structure. Another way of phrasing what has just been discussed is that the formation of an adjoined structure forced the PIC_1-pattern of Spell-Out upon the adjunct, rendering it opaque. In this sense (but in this sense only) adjunction is similar to the presence of unvalued features on a phase head, which also requires "immediate spell-out" (PIC_1) of a portion of the tree, as we saw above.

We could capitalize on this result and attempt to unify two aspects of the CED: adjunction and valuation are the source of islandhood. But notice that the cause of islandhood would still be distinct. Whereas the adjoined domain is opaque, the domain where valuation takes place is not. In the case of adjunction, we have a PIC_1-Spell-Out pattern tied to the need for a simplex edge restriction. In the case of valuation, the phase-edge hosting the unvalued feature is capable of supporting internally merged phrases. The island-inducing effect of PIC_1 in this case is more indirect, and more specific: it does not freeze the entire complement domain of the phase. The agreeing/valuing subject can still move to the edge of the phase, but I would like to claim that because of this valuing function the chain formed by a valuer is more susceptible to freezing/islandhood. Here is what I mean.

Consider freezing first. Why are subjects (and previously displaced objects) subject to freezing, while in-situ objects aren't, as the following examples attest:

(22) a. *Who did John say that ⟨Who⟩ caused the riot?
b. *Who did John say that Peter gave ⟨Who⟩ the book?

(23) Who did John say that Peter kissed ⟨Who⟩?

Assuming that objects value the ϕ-features on v the same way subjects value the ϕ-features on C, we can conclude from these examples that valuation per se is not the cause of freezing. Instead, what seems to cause freezing is the fact that the subject (as well as any other previously displaced argument) heads a non-trivial chain at the point of valuation, whereas the (direct) object doesn't. So it's valuation *cum* displacement *cum* cyclic spell-out that underlies the contrast under discussion. Why should this be so? Consider the (boxed) valuation configurations for displaced subject and *in situ* objects, respectively.

[85] There is one exception. Material adjoined to the adjoined constituent would not be part of this spell-out step. This could capture the resumption pattern inside adjuncts discussed by Boeckx (2003), if the antecedent of the resumptive pronoun is treated as adjoined to the adjunct (hence capable of extraction). For more discussion, see Boeckx (2008b).

(24) $[_{\text{CP}}\ \text{C}_{\phi}\ \boxed{[_{TP}\ \mathbf{XP}\ \text{T}_{\phi}\ \ldots [\ \langle \text{XP} \rangle\ \ldots \upsilon\ ]]}\ \ldots]$

(25) $[\text{XP}\ \upsilon_{\phi}\ \boxed{[_{VP}\ \text{V}\ \langle \text{XP} \rangle]}\ \ldots]$

I would like to claim that a chain headed by a valuer that is transferred is interpreted as a complete chain at the interfaces. If extended in Narrow Syntax, the chain will be doubly-headed, as it were. It will be too strong in Richards' (2001), or ambiguous in Boeckx's 2003, 2008b sense, and as such it will cause processing[86] problems for the external systems, the same way garden path sentences do.[87] No such problem arises with *in situ* valuers. At the heart of the freezing phenomenon lies the fact that the external systems only seem to be able to handle two kinds of information (Chomsky's "duality of semantics"): that's why they can only see two types of phases, and only two units/occurrences per chain: a head and a tail. The presence of a chain upon valuation means that any extension of that chain will result in there being three relevant unit for that chain: an $\overline{A}$-occurrence, a ϕ-valuing occurrence, and a tail—one too many. In the case of in-situ arguments, the ϕ-valuing occurrence and the tail are collapsed, hence do not pose any processing difficulty. It is as if once the external systems receive a chunk of structure with a valuer heading a non-trivial chain, they consider that chain closed,[88] and treat any extension as unexpected (deviant by default, I assume).[89]

Arguments susceptible to freezing render the elements they contain susceptible to freezing as well, which gives rise to one aspect of the CED. But notice that the ban on subextraction pertains not to displaced elements per se, it is only tied to displaced elements that form a non-trivial chain at the point of valuation. So valuation or displacement per se do not give rise to freezing/islandhood. We therefore expect what Müller (2010) call "melting effects" (extraction out of scrambled, non-agreeing/valued intermediate positions), and the oft-noted transparent character of

[86] I understand 'processing' in a broader sense than the way it is used in the psycholinguistic literature. Readers uncomfortable with my use of chain-processing in this sense could use 'chain-scanning' instead.

[87] Epstein and Obata (In press) explore the hypothesis that Improper Movement boils down to a side effect of how valuation proceeds in phase-based derivations. Their study, coupled with Boeckx's (2008b) argument that Improper Movement forms a natural class with island effects, could serve as another argument in favor of implicating islands in the explanation of constrains on internal merge (displacement).

[88] An earliness (interface) condition of chain completion as it were; not too dissimilar from what one finds in the psycholinguistic literature on filler-gap dependencies or early attachment phenomena.

[89] Kotzgoulou (2007) presents an account based on linearization by phase that is similar in spirit to what I do here.

in situ/post-verbal subjects (see Gallego (2007b, In press)).

Going back to the issue of whether CED-effects should be unified, the answer we are led to in the present framework is: no. First adjuncts and subjects are not the right units of analysis. It should be pair-merged units and displaced valuers. Second, even if both types of phrases are associated with a PIC_1-Spell-Out pattern, the PIC only has a direct influence on subextraction in the case of pair-merged units. In the case of valuers, the character of Spell-Out relevant only if valuation is associated with displacement. True, the need for displacement (the "EPP") is ultimately tied to Inheritance and the PIC, but the link with islandhood is much more indirect than in the case of Pair-Merge. This is the result we in fact want. We want a lot more (sub)extraction flexibility in the case of arguments. One could say that slandhood is definitional of adjunction/Pair-Merge (for Pair Merge to exist, the adjoined material save the phase head must be transferred), but islandhood is not definitional of either valuation or displaced domains (it only arises when valuation and displacement are tied).

What emerges from this discussion of islandhood is that while islands remain tied to syntactic structure,[90] at least some of the traditional islands (those especially subject to variation) are better thought of as representational, interface conditions (post-syntactic filters, in the sense of Miller and Chomsky (1963), Chomsky and Lasnik (1977)), with phases providing the mode of interfacing.

4.11 Intermediate Summary

Let me try to summarize what I have done so far. Taking as my starting point the explosion of unexamined feature structures pre-syntactically, I have argued for a shift of perspective that I view as the legacy, and radicalization, of Hale and Keyser's (1993, 2002) configurationalism and Borer's (2005) exoskeletality. This I view as the (only?) antidote against featuritis in linguistic theory, and a necessary step in the direction of integration of theoretical linguistics within biolinguistics.

I have argued that when it comes to the Faculty of Language in the Narrow Sense, the key factor is not recursion, it is the existence of a universal, uniform, homogeneous lexicon, which gives us the capacity to abstract away from selectional restrictions and the varying valency of concepts. The existence of enough of these conceptual addresses (itself, I think, a by-product of brain expansion; see Part D) pushed syntax (at bottom, simple set-formation) into the realm of Turing patterns/dissipative

[90] Goldberg (2006) suggests that some islands arise at the level of discourse representation, but since the latter depends on syntactic structure, there is no reason to assume that islandhood is divorced from syntax.

structures. Recursion as we know it in language comes for free once we reach a critical mass of homogeneous lexical atoms.

Perhaps the most immediate consequence of this approach is that morphology is radically divorced from syntax. Likewise, given the uniformity of the pre-syntactic lexicon, the complete uniformity of syntax is guaranteed. There cannot be any syntactic (or pre-syntactic) variation (on this, see Boeckx (2008c, In pressa, 2010i), in addition to Part B).

Along with Merge, the single most important mechanism within syntax appears to be Cyclic Spell-Out (phase-based derivations). Some may be surprised to see me argue in the present work for the necessity of phases in light of my lengthy critique of phases in Boeckx and Grohmann (2007). There is no contradiction here: the critique in Boeckx and Grohmann (2007) is aimed at models that combine standard minimalist tenets such as Last Resort and feature-driven operations with phases. By contrast, the type of framework adopted here, which develops themes found in Chomsky (2004), make phases a virtual necessity. Thus, one way of reading Boeckx and Grohmann (2007) is to say that phases only make sense in a radically minimalist framework of the sort explored here, at the intersection of Chomsky (2004) ('Merge α'), Distributed Morphology (universal late insertion), and Hale & Keyser's (1993, 2002) theory of argument structure, generalized to all other aspects of lexical entries. This is at variance with much existing literature.

The overall point of this contribution is that Free (or, as I prefer to call it, 'wild-type,' 'chaotic') Merge need not be seen as giving rise to too unconstrained a theory of narrow syntax, as the (independently well-motivated) cyclic nature of the mapping to SEM and PHON can be put to use in order to, as it were, regulate wild-type Merge. Indeed, the hypothesis pursued here is very restrictive. As I pointed out in Boeckx (2009c), it leaves very few analytic options open to capture the range of facts that $\overline{X}$-constructs or labels were used for.

Take, for example, the phenomenon of pied-piping. Traditionally, pied-piping is said to require some form of percolation along a projection path (see Heck (2004) for review and discussion of various modes of implementation), but more recently Cable (2007) has argued that pied-piping is the result of the way in which a *Q*-morpheme (the element that enters into a Probe-Goal relation with the interrogative C)[91] is added to an already existing structure (containing a *wh*/indenterminate-element). If *Q* is adjoined (pair-Merged; cf. Chomsky 2004), no pied-piping will take place. If *Q* is added by regular Merge (set Merge), pied-piping takes place (movement of the *wh*/indeterminate-element to the edge of the pied-piped unit depends on whether

[91]The exact identity of this element will be left open here. For arguments that Q is in fact Foc(us), see Narita (2008) and Rizzi (1997).

or not Q enters into an Agree-relation with it). In the event of, say, clausal pied-piping (of the sort one finds in Basque or Quechua), one need not assume that the [wh]-feature on the *wh*/indetermine-element must percolate up the clausal spine for pied-piping to take place. Rather, according to Cable, one must entertain the possibility that the Q-morpheme has been added (via set-Merge) to the whole clause. Under the latter scenario, pied-piping takes place under Q-movement for the same reason that the D-head in $\{$*the man*$\}$ pied-pipes its complement under raising. I will not be able to review all the evidence that Cable adduces in support of this theoretical understanding of pied-piping. I simply want to point out that in the absence of any notion of label other than what can be derived from the distribution of phase heads, Cable's analysis is the only one available in the present framework. This in turn means that the constraints on pied-piping identified by Cable ought to reduce to constraints on phase-distribution.[92]

Another illustration of the restrictiveness of the perspective adopted here comes from the domain of ellipsis. Traditional descriptions of ellipsis processes crucially refer to labels[93] and projections (how else could they define, say, *VP*-ellipsis?). According to the hypothesis pursued here, the only constituents that can be referred to at the interfaces are those defined by phase-heads. So, we predict that only phase-complements will be subject to deletion[94] (when interpretive factors such as focus/contrast are met, of course[95]). The fact that the the best-studied cases of

[92]On the face of it, this seems to be the correct result. Cable (2007) claims that no Q-morpheme can be inserted via set-Merge between a functional head and the lexical layer it immediately dominates (D and its NP complement, e.g.). If, as some (Borer, 2005, Marantz, 2008) have argued, lexical layers require a functional layer to be labeled, Cable's constraint may follow from the fact that the insertion of Q between a phase (label) and its (unlabeled) complement would give rise to a ⟨Phase, Phase, Non-Phase⟩ pattern that deviates from the alternating Phase-Non-Phase pattern taken to be the core structural rhythm in syntax. To go through, this reasoning requires Q to be a phase-head, but this is independently required by the logic of valuation, if Cable (2007) is right in arguing that Q may carry an unvalued feature. For further examination of the nature of pied-piping within the framework pursued here, see Narita (2008).

[93]Equivalently, *ad hoc* features like Merchant's (2001) [+E].

[94]Deletion could be construed as lack of lexicalization in a late insertion/Distributed Morphology context, as suggested by Holmberg (2001).

[95]Clearly not any phase-complement will be deletable. Conditions of recoverability and other discourse factors must be met. My point here is that it is a necessary (but not sufficient) condition that elided constituents be phase-complements, because only these are 'labeled' in the present system.

The present approach must also take into account the fact that phase heads, like any LI, may undergo movement, leading to the enlargement of its complement domain, hence of the ellipsis site (the idea behind 'phase-sliding' in Gallego (2007b)). Instances of TP-ellipsis in Romance may perhaps be analyzed as ellipsis of the complement domain of v raised to/past T. I return to the

ellipsis (VP-ellipsis and TP-ellipsis [a.k.a. sluicing]) correspond precisely to phase-complements is, from this perspective, very encouraging.[96]

My point in the previous paragraph is that if it is indeed the case that the units that syntax manipulates are atoms and that Narrow Syntax connects with the external systems via phases, these systems will only be able to track/exploit properties of the syntactic derivations through phase-heads: every constituent will be defined indirectly, by reference to those elements that 'stand out.' As we have seen, though dramatically impoverished, such interface representations have a certain empirical plausibility. At the very least, they impose certain limits on the range of analysis one can entertain for certain phenomena, a restrictiveness that should be welcome.

If phases prove sufficient to keep wild-type Merge in check, the radically minimal take on UG argued for here is in turn strengthened. It suggests that the notion of labeling is indeed "a dispensable notion," "a convenient notational device," "playing only an explository role" (Chomsky (2007, 8, 23)), and that the recent trend of 'exoskeletality' (Borer's 2005 apt term) found in many works on the lexicon should be extended to core syntactic properties such as headeness. As we have seen in previous sections, recent works argue that instead of packing all the relevant information inside lexical entries, lexical information should be distributed across various components of the grammar (especially those outside narrow syntax), with the effect that many properties once thought to be part of narrow syntax emerge dynamically, as interface (epi)phenomena. The route pursued here amounts to treating endocentricity as one such epiphenomenon. Instead of letting headedness inhere in the definition of Merge (as in Chomsky (1995, 243, 244, 246), or via feature-proxies (Collins (2002)), I claim that asymmetric labeling should be defined outside of the product of Merge, in the context of phases, pretty much like lexical categories are defined contextually in Distributed and Parallel Morphologies (Borer, 2005, Marantz, 2008). The basic idea behind endocentricity is captured, but the label of a phrase (at least that of the the phrase corresponding to the phase-complement) is defined from without, no longer from within—syntactically, not lexically. In this system, labeled structures are as sparse as in Brody's 2000, 2003) Mirror system. Phrase Structure collapses

idea of head-movement below.

[96]The correspondence between phase-complements and elided sites was independently made by Gengel (To appear), and Gallego and Yoshida (2008); it is explored in more detail in Gallego (2009a). The proposal made above concerning nominal structures suggests a natural extension of the present approach to NP-ellipsis, with *one* being the overt *n*-phase head defining the elided complement. The approach could extend to NP-ellipsis in Romance, if the word/class inflectional marker that appears to license ellipsis in these languages is an instance of *n*, as argued in Bernstein (2001); see also Alexiadou and Gengel (To appear). For discussion of ellipsis in adpositional (phasal) domains, see Gallego (2009a,c).

into Phase Structure.

The profile that phase-based derivations acquire under the present framework is one that guarantees the formation of treelets, minimal constituents that expand via Merge, but quickly contract at the phase-level, via Spell-Out. Phase-heads here are akin to logarithmic functions. They periodically purge derivations from an abundance of material, and just like logarithmic functions, which reduce multiplication to addition and division to subtraction, phase-heads keep derivations maximally simple, monotonic, reducing each derivational step to instances of merging a head with (at most) a phrase containing two units. (Note that this 'restriction' on derivation need not be stipulated, it is a theorem of the way the present system is set up.) As Narita (2008) points out, this in effect reduces derivations within narrow syntax to iterated applications of head-to-head merger.[97] Put another way, pied-piping (XP-movement), as an operation within Narrow Syntax, is well and truly eliminated, its effects emerging at the interfaces where the fragments of syntactic structures spelled out in a cyclic fashion is recombined to form utterances.[98] For a detailed exploration of the consequences of this hypothesis for a wide range of phenomena (coordination, wh-pied-piping, focus projection, etc.), see Narita (2008), where it is (correctly) observed that once labels are eliminated, it is quite natural to eliminate pied-piping, since the former is after all the most local instantiation of the percolation mechanism underlying (in one form or another) the latter.

If indeed all instances of Merge reduce to merging a lexical item (phase head) with (at most) a set of two other units, then the result of Uriagereka's (1999) Multiple Spell-Out hypothesis (i.e., the need to 'pre'-linearize complex specifiers (and adjuncts)) follows immediately, without requiring us to adopt his specific cyclic spell-out model (which is quite different from a phase-based system, since for Uriagereka, there is never any need to spell out complements cyclically).

As in Moro (2000), the points of asymmetry is delayed, but much more so here, and unlike Moro, no special symmetry-breaking role for Internal Merge has to be recognized. Instead of assuming that it is pre-syntactic feature structures that create all the relevant asymmetries, syntax is the generative engine that causes these asymmetries to surface, as Merge takes place.

[97] This conclusion is anticipated by Uriagereka (1999), in his discussion of 'radical multiple spell-out,' although the way it is derived is quite different.

[98] This approach to pied-piping places a lot of emphasis on the role of the edge of phases to be pied-piped. This role of phase-edges as 'addresses' (instructions to pied-pipe) may shed light on the otherwise curious requirement that branching specifiers must have non-phonetically null edges (An, 2007a,b, Landau, 2007). It is tempting to interpret this restriction as saying that phonetically null edges lack the power to instruct PHON to 'pied-pipe' the relevant phase-complements (a plausible hypothesis if phonetic nullness on an element reduces to an instruction for PF to ignore this element.

5 Déjà vu all over again?

The architecture of grammar taking shape in the current work bears some resemblance with models pursued in the wake of Chomsky (1965). It is interesting to note that current trends in generative grammar are in fact reviving ambitions first expressed in light of the prospects discussed in very clear fashion in the first chapter of *Aspects of the theory of syntax*. Thus, we are witnessing a revival of the term Biolinguistics, a term that emerged in the early 1970s (for discussion, see Di Sciullo and Boeckx (In press)). We are also witnessing a revival of the Derivational Theory of Complexity (Marantz, 2005, Boeckx, 2009b), as well as renewed interest for Cartesian antecedent (Boeckx, 2009a, In pressc). Following early suggestions of Chomsky's in *LSLT*, we are currently entertaining seriously the possibility of uniting statistical learning and the (innate) priors (Yang, 2004, Pearl, 2007, Dillon and Idsardi, 2009, Boeckx, 2009b). And last, but not least, the distance between syntactic and semantic representations has been so reduced as to give the impression it may have vanished (see Hinzen (2006), Uriagereka (2008), Boeckx (2009e)), much like it did in the days of Abstract Syntax and Generative Semantics. Of course, current generative works are not merely restating theses first formulated in the years immediately following *Aspects*, but there is a sense in which current work vindicates Mark Twain's statement that "history does not repeat itself, but it rhymes." Some of us feel that we are once again living in exciting times, in which the prospects of interdisciplinarity are bright and the importance of theoretical linguistics, paramount.[99]

I am by no means the first to see the resemblance between some current works in syntax and works in, say, Generative Semantics. Hale and Keyser (1993) acknowledge it, and both Pullum (1996) and Culicover and Jackendoff (2005) insist on the parallelism. It is true that one can see hints of Generative Semantics in (cartographic/nanosyntactic) attempts to semanticize syntax, (antisymmetric) attempts to establish a universal base, (minimalist) attempts to eliminate D(eep)-Structure, as well as the extensive use of late lexical insertion, and constant arguments for the simplest theory, all of which were part and parcel of the Generative Semantics movement. But one can also see similarities between pre-Generative Semantics syntax and minimalism, such as Last Resort and *Syntactic Structures*-style obligatory transformations, the heavy reliance of features, the cross-linguistic component

[99]Those of us that share in this *Zeitgeist* disagree strongly with Jackendoff's (2002) bleak prospects for Chomskyan generative grammar, which is said to have "alienated biology and the rest of cognitive science." For reasons to be highly skeptical of Jackendoff's characterization, see Boeckx and Piattelli-Palmarini (2005, 2007); see also Boeckx (2006, 2005, 2009b), Marantz (2005), and Part D.

of many analyses, the lexical irregularities governing Transformations (as in Lakoff (1970)), and so on. Although such parallelisms are by no means devoid of interest, I will not pursue them here, and instead will concentrate on one of the most detailed (albeit unpublished) discussion of the nature of the lexicon in those days, viz. Otero (1976).

To close this brief excursus on the history of the field, it may be worth mentioning that other frameworks also have a few things in common with Generative Semantics. As Searle (1972, 148) points out, "Those who call themselves generative semanticists believe that the generative component of a linguistic theory is not the syntax ..., but the semantics, that the grammar starts wth a description of the meaning of a sentence and then generates the syntactical structures throuhg the introduction of syntactic rules and lexical rules. The syntax then becomes just a collection of rules for expressing meaning." This is a description that fits Cartographic approaches as well as Culicover and Jackendoff's (2005) "Simpler Syntax" vision — a description that goes in the opposite direction of what I am advocating here. As a matter of fact, I think that recent attempts to deflate the syntactic component of the grammar reflects the functionalism of Generative Semantics,[100] and its fear of pure form (see Koster (1987, chap. 7)). As Chomsky notes (personal communication to M. Brame, 13 August 1975, cited in Brame (1976, 26))),

> I think the point is that they [generative semanticists] are unwilling to stray far, in abstractness, from the "given"—i.e., phonetic fact and semantic fact. Structures that do not directly mirror one or the other are unacceptable. Hence, the virtual demise of syntax.

What is clear from looking back at the Generative Semantics period is that, much as I have done at the beginning of this study, the issue of the nature of the lexicon occupied pride of place then. At the time of the so-called "linguistic wars", the lexicon was the most dangerous mine-field. Hence my interest in Otero's work.

5.1 Otero (1976)

At the heart of Otero (1976) is the need to recognize two distinct grammatical components:[101]

> The grammatical system of a natural language L consists of two autonomous, but closely interrelated, subsystems:

[100] For illustration of this point in the specific realm of control, see Boeckx et al. (2010, chap. 7).

[101] Because Otero (1976) remains unpublished, I decided to use extensive quotes in this subsection to render the reader's appreciation of Otero's argument easier.

> i. A set of *syntagmatic* operations that pairs the pronunciation of each of the constructions it generates directly or derivatively, the set of the constructions in L being *infinite*, and
> ii. A set of *paradigmatic* relations and operations the output of which is the set of all and only the phonologically independent words, the set of words in L being *finite*

As Otero immediately observes, this so-called "two sub-systems hypothesis" has "Immediate consequences". First, "No level of the (syntagmatic) derivation of a *sentence* is represented as a string of "morphemes" " (an idea that has been revived, without acknowledgement, in the context of nanosyntax; see Starke (2010), Caha (2009)). Second, "All processes involved in the determination of "*word* forms" and their corresponding "word contents" belong in the paradigmatic subsystem." And finally, " "Word units" are *inserted* into an abstract phrase marker only *after* all the purely syntactic rules (nontransformational and transformational) have applied."

Otero goes on to note that

> [i]t seems clear that the [Two Subsystems Hypothesis], if basically correct, yields a much improved theory of generative grammar — one with fully differentiated but internally homogeneous components. The syntagmatic subsystem consists of a central component (the syntax) and two interpretive components (the phonetics and the semantics). The syntactic component consists of a recursive set of context-free phrase-structure rules and a transformational subsomponent with root transformations, one nonlocal transformation ("move C") and a set of local transformations in the sense of Emonds (to a great extent language particular), which together generate what might be called "construction forms" (cf. *LSLT*, §33.1), that is, abstract phrase markers including only syntactic category and subcategory feature specifications [...] The "construction forms" will presumably be enough to derive a "logical form" ...; a full interpretation can only be derived after the insertion of phonological matrices of words (in the extended sense) from the paradigmatic subsystem."

Elsewhere (Otero (1983)), Otero clarifies his proposal and notes that his "Dual hypothesis", which distinguishes between a syntagmatic grammar and a paradigmatic grammar,[102] results in an overall grammatical architecture that is "conceptually simpler".

[102] A distinction that harks back to Sassure's view that, next to *Langue/Parole*, and synchrony/diachrony, a theory of syntagms vs. a theory of associations is "the most basic rational division" for the linguist.

Already then, it was clear to Otero that the "lexicon" is to be understood as a family of component, where one ought to distinguish between a "dictionary", a "lexicon in the narrow sense", and an "encyclopedia" — a three-way distinction echoed in Marantz (1996). Otero (1983) observes that "this [paradigmatic system] is the subpart that exerts a special fascination over the minds of some students of language." But it is only "the syntagmatic grammar [that] can be assumed to be a fairly direct reflection of the language faculty of the mind/brain." Citing Chomsky (who echoes Jespersen), Otero notes that "no student of human language ever dreamed of a universal dictionary." For, if

> [a] syntagmatic grammar is essentially universal (biologically given in essence), a paradigmatic grammar is, to a considerable extent, a historically evolving subsystem, burdened with the weight of the past, like other cultural systems. Only a paradigmatic grammar can be fossiliferous. This brings to mind the distinction between "core grammar" and a "periphery" of "borrowings, historical residues, inventions, and so on, which we can hardly expect to — and indeed would not want to — incorporate within a principled theory of UG" (Chomsky (1981, chap. 1))

Otero is quick to add that even if "[e]very paradigmatic grammar is, to a considerable extent, language particular, and to some extent fossilized ... this is not to say that everything in the paradigmatic grammar is language-particular, or that the paradigmatic grammar doesn't have any internal structure. The model of paradigmatic grammar that is emerging in the most recent investigations is itself highly modular ..." (something that we will see again in 7).

Interestingly, Otero (1983) observes that under his model, "there are no processes such as "affix hopping" in the syntagmatic grammar, and plausibly there is no "copying" of agreement features in the course of the syntagmatic derivation", something that is also true of the syntactic component in the present approach.

5.2 Radicalizing Otero's proposal

I have gone at some length to quote passages from Otero (1976, 1983) because, like Otero, I am in favor of a sharp divide between a pure syntactic component, freed from imperfections, and another system which (as Sapir would have said) necessarily "leaks", containing as it does a fair share of accidental properties. It seems to me that this separation is necessary if we are to entertain the strong minimalist thesis seriously. Why this separation has not been made standard yet is something of a puzzle. Already back then, Otero (1976) points out that "[g]iven the theoretical

framework Chomsky had developed in [Aspects], it is somewhat surprising that he did not go on to draw what, from a generative perspective, appears to be a very natural, if not inescapable, conclusion, namely that morphemic representations play no role in the (syntagmatic) derivation of a sentence."

I have found two references to Otero's work on the lexicon in Chomsky's writings. Both are buried in footnotes. The first is in Chomsky (1980, 277n.10), where Chomsky writes that "[o]ne might nevertheless argue that full lexical insertion, including phonological and morphological properties of words, takes place at the level of S-structure, along lines that have been suggested by Carlos Otero (...) and Hans den Besten (...)."[103] The second reference is in Chomsky and Lasnik (1977, 432n18), where Chomsky and Lasnik state that "[i]n fact, there is little reason to suppose that lexical items are inserted in base structures, in this theory. ... We will continue to accept this assumption here for ease of exposition, but everything we say can be translated into an alternative theory in which lexical insertion takes place in surface structure and only abstract features are generated in the base (which is now limited to the categorial component) in positions to be filled by lexical items."

Contrary to what Chomsky and Lasnik claim, I think that there is more at issue than just "ease of exposition". First, keeping the syntagmatic component in the same batch as the paradigmatic component invariably leads, one way or another, to lexicocentrism, and all the problems discussed in previous sections. Segregating the two components, as Otero suggests, leads to a level of representation of the syntactic component that is not only far more explanatory, but also truly invariant, as I will show in the next section. As Otero (1996, 321) already noted, if one adopts the Two Subsystems Hypothesis, one is led to claim that "there is only one language, as the evolutionary biologist would expect." That is to say, for purposes of linguistic description, Chomsy and Lasnik's stance may well be a matter of "ease of exposition." But when it comes to the biolinguistic program, there is far more at stake. Failing to segregate the two subsystems boils down to claiming that syntax is at every level a parametric syntax at best. It is never a truly principled syntax, as it is always relativized to a lexicon tied to a particular language. As a result, the possibility of a truly principled explanation, of the sort demanded by minimalism , is out of reach. It is always a construction-specific syntax. A syntax that is indeed shaped by the lexicon, as Caponigro and Polinsky (In press) assume; a syntax about which it can be said (as Baker does in the metaphor of the Swiss watch discussed at the beginning of this work, that "small changes in the properties of a single word or class of words, or the addition of a single grammatical principle can have large scale repercussions

[103]Chomsky here refers to unpublished work by Hans den Besten to which I have not been able to gain access.

on the entire language". Incidentally, it is also the reason why every characterization of syntax (as an independent generative system) by Jackendoff (and others) deals with issues that are confined to parameters (case, word order, etc.). As the following quote (taken from Culicover and Jackendoff (2005, 22)) reveals, the rationale for syntax is frustratingly highly language-specific (i.e., tied to specific languages):

> Should *all* syntactic structure be slashed away? Our goal, a theory of syntax with the minimal structure necessary to map between phonology and meaning, leaves open the possibility that there is *no* syntax at all: that it is possible to map directly from phonological structure (including prosody) to meaning. Although some people might rejoice at such an outcome, we think it is unlikely. Perhaps this represents a certain conservatism on our part, and someone more daring will be able to bring it off. But at a minimum, we believe that *syntactic categories* such as noun and verb are not definable in purely semantic terms—and that fundamental syntactic phenomena such as agreement and case-marking are based on these categories. And we believe that there are *syntactic constituents* whose categories are determined (for the most part) by the categories of their heads, i.e., that there is something like X-bar phrase structure. We think it is not a matter of phonology or semantics that English verbs go after the subject, Japanese verbs go at the end of the clause, and German inflected verbs go in second position in main clauses but at the end in subordinate clauses. We think it is not a matter of phonology or semantics that English sentences require an overt subject but Italian sentences do not; that English has ditransitive verb phrases but Italian does not; that English has *do*-support but Italian does not)but see Beninca' and Poletto 2004 for a northern Italian dialect that does have *do*-support; that Italian has object clitics before the verb but English does not. That is, we are going to take it for granted that there is some substantial body of phenomena that require an account in terms of syntactic structure. It is just that we think this body is not as substantial as mainstream generative grammar has come to assume. This is why we call our hypothesis "Simpler Syntax" rather than just plain "Simple Syntax".

The architecture I am advocating is in some sense (and with the benefit of hindsight) a radicalization of Otero's proposal (much like my proposal can be seen as a radicalization (and generalization) of Hale and Keyser's (1993, 2002) proposal concerning argument structure, and also a radicalization of Distributed Morphology's late insertion mechanism). Otero still maintains "syntactic category and subcategory feature

specifications" as part of the syntagmatic component. I am doing away with even this much.

There is another reason why I think that there is more to Otero's proposal than just a matter of "ease of exposition." By keeping separate the syntagmatic and paradigmatic systems, Otero makes it possible to view the lexicon has non-distinct from the (paradigmatic) grammar. (By lexicon here, I mean the post-syntactic lexicon. As Piera (1985, 311n.2) already noted, "[t]he term "lexicon" is clearly inappropriate for the complex component envisaged by these theories [cf. Otero 1976].") Over the years, Jackendoff has argued that Chomsky (following Bloomfield) has been wrong in "regarding the lexicon as altogether separate from the grammar." — for reasons that are compelling (roughly speaking, they are the reasons why so much work in morphology assumes a realizational, post-syntactic component that views "words" as phrasal."Otero's separation thesis could be recast (in light of the present proposal) as a separation between syntax and grammar, leaving the door open for the grammar and the lexicon to be one.[104] It can also be seen as a separation between principles and parameters. A separation that is necessary if we want to move beyond explanatory adequacy.

6 Part B: No room left for Parameters

7 Part C: Constructing the interfaces

8 Part D: Building the interdisciplinary bridges

9 General Conclusion

References

Abels, K. 2003. Successive cyclicity, anti-locality, and adposition stranding. Doctoral Dissertation, University of Connecticut.

Abels, K., and K. Bentzen. 2009. Are movement paths punctuated or uniform? *Catalan Journal of Linguistics* 8:19–40.

Abney, S. P. 1987. The noun phrases in its sentential aspect. Doctoral Dissertation, MIT.

[104] A position that can make much better sense of the claims found in the "grammaticalization" literature, as I argue in Boeckx (2010b).

Aboh, E. O. 2004. *The Morphosyntax of Complement-Head Sequences: Clause Structure and Word Order Patterns in Kwa*. Oxford: Oxford University Press.
Aboh, E.O. 2009. Clause structure and verb series. *Linguistic Inquiry* 40:1–33.
Adger, D. 2003. *Core syntax: a minimalist approach*. Oxford: Oxford University Press.
Adger, D. 2010. A minimalist theory of feature structure. In *Features. Perspectives on a key notion in linguistics*, ed. A. Kibort and G. Corbett, 185–218. Oxford: Oxford University Press.
Adger, D., and P. Svenonius. In press. Features in Minimalist Syntax. In *The Oxford Handbook of Linguistic Minimalism*, ed. C. Boeckx. Oxford: Oxford University Press.
Alexiadou, A., and K. Gengel. To appear. Classifiers as morphosyntactic licensors of NP ellipsis: English vs. Romance. In *Proceedings of NELS 39*.
Amritavalli, R., and K.A. Jayaseelan. 2003. The genesis of syntactic categories and parametric variation. In *Generative Grammar in a Broader Perspective: Proceedings of the 4th GLOW in Asia*, 19–41.
An, D.-H. 2007a. Clauses in non-canonical positions at the syntax-phonology interface. *Syntax* 10:38–79.
An, D.-H. 2007b. Syntax at the PF interface: Prosodic Mapping, Linear Order, and Deletion. Doctoral Dissertation, University of Connecticut.
Anagnostopoulou, E. 2003. *The Syntax of Ditransitives: Evidence from Clitics*. Berlin: Mouton/de Gruyter.
Anderson, S.R. 2004. *Doctor dolittle's delusion: Animals and the uniqueness of human language*. New Haven, Conn.: Yale University Press.
Arsenijević, B. 2008. From spatial cognition to language. *Biolinguistics* 2:3–23.
Arsenijević, B., and W. Hinzen. 2010. Recursion as a human universal and as a primitive. *Biolinguistics* 4:165–173.
Bak, P. 1996. *How nature works: The science of self-organized criticality*. New York: Copernicus.
Bak, P., C. Tang, and K. Wiesenfeld. 1988. Self-organized criticality. *Physical review A* 38:364–374.
Baker, M. 1996. *The Polysynthesis Parameter*. Oxford: Oxford University Press.
Baker, M. 1999. On the interplay of the universal and the particular: Case study of Edo. In *Proceedings of CLS 35: The Panels*, 265–289. Chicaho Linguistic Society.
Baker, M. 2001. *The Atoms of Language*. New York: Basic Books.
Baker, M. 2003. *Lexical categories: Verbs, Nouns, and Adjectives*. Cambridge: Cambridge University Press.
Baker, M. 2008. *The syntax of agreement and concord*. Cambridge: Cambridge

University Press.
Ball, P. 2006. *Critical mass: How one thing leads to another*. New York: Farrar Straus & Giroux.
Barner, D., J. Wood, M. Hauser, and S. Carey. 2008. Evidence for a non-linguistic distinction between singular and plural sets in rhesus monkeys. *Cognition* 107:603–622.
Bauke, L. In progress. Doctoral Dissertation, Universität Wuppertal.
Behar, D.M., R. Villems, H. Soodyall, J. Blue-Smith, L. Pereira, E. Metspalu, R. Scozzari, H. Makkan, S. Tzur, D. Comas, et al. 2008. The dawn of human matrilineal diversity. *The American Journal of Human Genetics* 82:1130–1140.
Béjar, S. 2003. Phi-syntax: A theory of agreement. Doctoral Dissertation, University of Toronto.
Belletti, A. 1990. *Generalized verb movement: Aspects of verb syntax*. Torino: Rosenberg & Sellier.
Belletti, A. 2004. Aspects of the low IP area. In *The structure of cp and ip*, ed. L. Rizzi, 16–51. Oxford: Oxford University Press.
Benítez-Burraco, A., and V. Longa. To appear. Evo-Devo. Of course, but which one? *Biolinguistics* .
Bernstein, J.B. 2001. The DP Hypothesis: Identifying Clausal Properties in the Nominal Domain. In *The Handbook of Contemporary Syntactic Theory*, ed. M. Baltin and C. Collins, 536–561. Oxford: Blackwell.
Bissell-Doggett, T. 2004. All else unequal. Doctoral Dissertation, MIT.
Bobaljik, J.D. 2008. Where's Phi? Agreement as a post-syntactic operation. In *Phi theory: phi-features across modules and interfaces*, ed. D. Harbour, D. Adger, and S. Béjar, 295–328. Oxford: Oxford University Press.
Boeckx, C. 2003. *Islands and chains*. Amsterdam: John Benjamins.
Boeckx, C. 2005. Generative grammar and modern cognitive science. *Journal of Cognitive Science* 6:45–54.
Boeckx, C. 2006. *Linguistic Minimalism: Origins, Concepts, Methods, and Aims*. Oxford: Oxford University Press.
Boeckx, C. 2008a. *Aspects of the syntax of agreement*. London: Routledge.
Boeckx, C. 2008b. *Bare syntax*. Oxford: Oxford University Press.
Boeckx, C. 2008c. Did we really solve Plato's Problem (abstractly)? In *Proceedings of the 9th Annual Tokyo Conference on Psycholinguistics*, ed. Y. Otsu. Tokyo: Hituzi Syobo.
Boeckx, C. 2008d. *Understanding Minimalist Syntax: Lessons from Locality in Long-distance Dependencies*. Oxford: Blackwell.
Boeckx, C. 2009a. Cartesian biolinguistics. SOLIFIC lecture, Sophia University, July

2009. Ms., ICREA–Universitat Autònoma de Barcelona.
Boeckx, C. 2009b. *Language in Cognition: Uncovering mental structures and the rules behind them*. Malden: Wiley–Blackwell.
Boeckx, C. 2009c. The locus of asymmetry in UG. *Catalan Journal of Linguistics* 8:41–53.
Boeckx, C. 2009d. On the nature of Merge. In *Of minds and language: A Basque encounter with Noam Chomsky*, ed. M. Piattelli-Palmarini, P. Salaburu, and J. Uriagereka, 44–57. Oxford: Oxford University Press.
Boeckx, C. 2009e. Some notes on the syntax–thought interface. In *Proceedings of the Sophia University Linguistic Society 24*, 92–103. Sophia University Linguistic Society.
Boeckx, C. 2010a. Linguistic minimalism. In *Oxford Handbook of Linguistic Analysis*, ed. B. Heine and H. Narrog, 485–505. Oxford: Oxford University Press.
Boeckx, C. 2010b. On Grammaticalization. Seminar talk, Universitat Autònoma de Barcelona, April 30, 2010.
Boeckx, C. 2010c. The perils of cartography. 2010, ICREA–Universitat Autònoma de Barcelona.
Boeckx, C. 2010d. Phases beyond explanatory adequacy. Ms., ICREA–Universitat Autònoma de Barcelona [Under review for Á. Gallego, ed., *Phase theory: developing the framework*. Berlin: Walter de Gruyter.].
Boeckx, C. 2010e. Research Trends for Biolinguistics: Lessons from the Extended Synthesis in Biology. ICREA–Universitat Autònoma de Barcelona.
Boeckx, C. 2010f. Syntactic order for free. Presented at the 10^{th} European Conference on Complex Systems, Lisbon University Institute, Workshop on 'Modelling the non-separability of a very complex world'.
Boeckx, C. 2010g. Taming recursion. Ms., ICREA–Universitat Autònoma de Barcelona. Ms. based on Presentation at the UMass Conference on Recursion, May 2009.
Boeckx, C. 2010h. The Chomsky-Piaget encounter at Royaumont, 35 years after. CLT Seminar Presentation.
Boeckx, C. 2010i. What Principles *&* Parameters Got Wrong. Ms., ICREA *&* Universitat Autònoma de Barcelona. [Written version of Talk given at the Workshop on Variation in the Minimalist Framework, Barcelona, January 2010.].
Boeckx, C. In preparation. *Syntactic Islands*. Cambridge: Cambridge University Press.
Boeckx, C. In pressa. Approaching parameters from below. In *The biolinguistic enterprise: New perspectives on the evolution and nature of the human language faculty*, ed. A.-M. Di Sciullo and C. Boeckx. Oxford: Oxford University Press.

Boeckx, C. In pressb. The emergence of the language faculty, from a biolinguistic point of view. In *The Oxford Handbook of Language Evolution*, ed. M. Tallerman and K. Gibson. Oxford: Oxford University Press.

Boeckx, C. In pressc. Some reflections on Darwin's Problem in the context of Cartesian Biolinguistics. In *The biolinguistic enterprise: New perspectives on the evolution and nature of the human language faculty*, ed. A.-M. Di Sciullo and C. Boeckx. Oxford: Oxford University Press.

Boeckx, C. Submitted. What does minimalism have to say about syntactic locality? Ms., ICREA–Universitat Autònoma de Barcelona. [Under review for Proceedings volume of the Budapest conference on Minimalist Approaches to Syntactic Locality.].

Boeckx, C., and Á. Gallego. 2009. Clitic Climbing by (Long Distance) Agree. Ms., ICREA *&* Universitat Autònoma de Barcelona.

Boeckx, C., and K.K. Grohmann. 2007. Putting phases in perspective. *Syntax* 10:204–222.

Boeckx, C., and N. Hornstein. 2008. Superiority, reconstruction and islands. In *Foundational issues in linguistics*, ed. C. Otero, R. Freidin, and M.-L. Zubizarreta, 197–225. Cambridge, Mass.: MIT Press.

Boeckx, C., N. Hornstein, and J. Nunes. 2010. *Control as movement*. Cambridge: Cambridge University Press.

Boeckx, C., and M. Piattelli-Palmarini. 2005. Language as a natural object; linguistics as a natural science. *The Linguistic Review* 22:467–471.

Boeckx, C., and M. Piattelli-Palmarini. 2007. Linguistics in Cognitive Science: The state of the art amended. *The Linguistic Review* 24:403–415.

Boeckx, C., and S. Stjepanović. 2001. Head-ing toward PF. *Linguistic Inquiry* 32:345–355.

Borer, H. 1984. *Parametric syntax*. Dordrecht: Foris.

Borer, H. 2003. Exo-skeletal vs. endo-skeletal explanations: Syntactic projections and the lexicon. In *The nature of explanation*, ed. J. Moore and M. Polinsky, 31–67. Chicago: CSLI Publications.

Borer, H. 2005. *Structuring Sense (2 vols.)*. Oxford: Oxford University Press.

Bošković, Ž. 1994. D-structure, theta criterion, and movement into theta positions. *Linguistic Analysis* 24:247–286.

Bošković, Ž. 1997. *The Syntax of Nonfinite Complementation; an Economy Approach*. Cambridge, Mass.: MIT Press.

Bošković, Ž. 2001. *On the nature of the syntax-phonology interface: Cliticization and related phenomena*. Amsterdam: Elsevier.

Bošković, Ž. 2002. A-movement and the EPP. *Syntax* 5:167–218.

Bošković, Ž. 2005. On the locality of left branch extraction and the structure of np. *Studia Linguistica* 59:1–45.
Bošković, Ž. 2007. On the locality and motivation of Move and Agree: An even more minimal theory. *Linguistic Inquiry* 38:589–644.
Bošković, Ž. 2008. What will you have, DP or NP? In *Proceedings of NELS 37*. GLSA.
Bošković, Ž. In press. On valued uninterpretable features. In *Proceedings of NELS 39*. University of Massachusetts, GLSA Publications.
Bošković, Ž., and H. Lasnik. 1999. How Strict Is the Cycle? *Linguistic Inquiry* 30:691–703.
Brame, M.K. 1976. *Conjectures and refutations in syntax and semantics*. London: North Holland.
Brody, M. 2000. Mirror Theory: Syntactic Representation in Perfect Syntax. *Linguistic Inquiry* 29:367–398.
Brody, M. 2003. *Towards an elegant syntax*. New York: Routledge.
Browning, M. 1987. Null operator constructions. Doctoral Dissertation, MIT.
Bruening, B. 2008. Selectional Asymmetries between CP and DP Suggest that the DP Hypothesis is Wrong. ms., University of Delaware.
Cable, S. 2007. The grammar of Q. Doctoral Dissertation, MIT.
Caha, P. 2009. The nanosyntax of case. Doctoral Dissertation, Universitetet i Tromsø.
Camps, M., and J. Uriagereka. 2006. The gordian knot of linguistic fossils. In *The biolinguistic turn*, ed. J. Rossello and J. Martin, 34–65. Barcelona: PPU.
Caponigro, I., and M. Polinsky. In press. Relative embeddings: A Circassian puzzle for the syntax/semantics interface. *Natural Language and Linguistic Theory* .
Carnie, A. 2008. *Constituent Structure*. Oxford: Oxford University Press.
Carruthers, P. 2006. *The Architecture of the Mind*. Oxford: Oxford University Press.
Carstens, V. 2001. Multiple agreement and case deletion: against φ-(in) completeness. *Syntax* 4:147–163.
Carstens, V. 2005. Agree and EPP in Bantu. *Natural Language and Linguistic Theory* 23:219–279.
Castillo, J.C., J. Drury, and K.K. Grohmann. 1999. Merge over move and the extended projection principle. *University of Maryland Working Papers in Linguistics* 8:66–104.
Castillo, J.C., J. Drury, and K.K. Grohmann. 2009. Merge over move and the Extended Projection Principle: MOM and the EPP Revisited. *Iberia* 1:53–114.
Cheney, D.L., and R.M. Seyfarth. 1990. *How monkeys see the world*. Chicago: University of Chicago Press.

Cheng, L. L.-S., and R. Sybesma. 2005. Classifiers in four varieties of Chinese. In *The Oxford Handbook of Comparative Syntax*, ed. G. Cinque and R. Kayne, 259–292. Oxford: Oxford University Press.

Cheng, L.L.S., and R. Sybesma. 1999. Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30:509–542.

Chomsky, N. 1955. *The logical structure of linguistic theory*. Ms. Cambridge, Mass: Harvard University. Published in part in 1975. New York: Plenum.

Chomsky, N. 1957. *Syntactic structures*. The Hague: Mouton.

Chomsky, N. 1959. On certain formal properties of grammars. *Information and Control* 2:137–167.

Chomsky, N. 1965. *Aspects of the theory of syntax*. Cambridge, Mass.: MIT Press.

Chomsky, N. 1966. *Cartesian linguistics*. New York: Harper & Row. [Third edition, with introduction by J. McGilvray, 2009, Cambridge University Press.].

Chomsky, N. 1970. Remarks on nominalization. In *Readings in English transformational grammar*, ed. R. Jacobs and P. Rosenbaum, 184–221. Waltham, Mass.: Ginn and Co.

Chomsky, N. 1973. Conditions on transformations. In *A Festschrift for Morris Halle*, ed. S. Anderson and P. Kiparsky, 232–286. New York: Holt, Rinehart and Winston.

Chomsky, N. 1977. On wh-movement. In *Formal syntax*, ed. P. Culicover, T. Wasow, and A. Akmajian, 71–132. New York: Academic Press.

Chomsky, N. 1980. *Rules and representations*. New York: Columbia University Press.

Chomsky, N. 1981. *Lectures on Government and Binding*. Foris: Dordrecht.

Chomsky, N. 1986a. *Barriers*. Cambridge, Mass.: MIT Press.

Chomsky, N. 1986b. *Knowledge of language*. New York: Praeger.

Chomsky, N. 1993. A minimalist program for linguistic theory. In *The view from Building 20*, ed. K. Hale and S. J. Keyser, 1–52. Cambridge, Mass.: MIT Press.

Chomsky, N. 1995. *The minimalist program*. Cambridge, Mass.: MIT Press.

Chomsky, N. 2000. Minimalist inquiries: the framework. In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, ed. R. Martin, D. Michaels, and J. Uriagereka, 89–155. Cambridge, Mass.: MIT Press.

Chomsky, N. 2001. Derivation by phase. In *Ken Hale: A Life in Language*, ed. M. Kenstowicz, 1–52. Cambridge, Mass.: MIT Press.

Chomsky, N. 2004. Beyond explanatory adequacy. In *Structures and beyond*, ed. A. Belletti, 104–131. New York: Oxford University Press.

Chomsky, N. 2005. Three factors in the language design. *Linguistic Inquiry* 36:1–22.

Chomsky, N. 2007. Approaching UG from below. In *Interfaces + recursion =*

language? Chomsky's minimalism and the view from semantics, ed. U. Sauerland and H.-M. Gärtner, 1–30. Mouton de Gruyter.
Chomsky, N. 2008. On phases. In *Foundational issues in linguistics*, ed. C. Otero, R. Freidin, and M.-L. Zubizarreta, 133–166. Cambridge, Mass.: MIT Press.
Chomsky, N., and M. Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row. Reprinted in 1991. Cambridge, Mass.: MIT. Press.
Chomsky, N., and H. Lasnik. 1977. Filters and Control. *Linguistic Inquiry* 8:425–504.
Cinque, G. 1999. *Adverbs and functional heads*. Oxford: Oxford University Press.
Citko, B. 2005. On the nature of merge: External merge, internal merge, and parallel merge. *Linguistic Inquiry* 36:475–496.
Collins, C. 2002. Eliminating labels. In *Derivation and explanation in the minimalist program*, ed. S. D. Epstein and T. D. Seely, 42–64. Oxford: Blackwell.
Collins, C. 2007. Home Sweet Home. In *NYU Working Papers in Linguistics 1*, 1–34.
Collins, C., and H. Ura. 2001. Eliminating Phrase Structure. Ms., Cornell University and Kwansei Gakuin University.
Cooke, J., and E.C. Zeeman. 1976. A clock and wavefront model for control of the number of repeated structures during animal morphogenesis. *Journal of Theoretical Biology* 58:455–476.
van Craenenbroeck, J., and M. van Koppen. 2002. The locality of agreement and the CP-domain. Presented at 25^{th} GLOW Colloquium.
Cuervo, M.C. 2003. Datives at large. Doctoral Dissertation, MIT.
Culicover, P.W., and R. Jackendoff. 2005. *Simpler syntax*. Oxford: Oxford University Press.
Dawkins, R. 1996. *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe Without Design*. New York: WW Norton & Company.
Deacon, T.W. 1997. *The symbolic species*. New York: Norton.
Dehaene, S. 1997. *The number sense*. Oxford: Oxford University Press.
Den Dikken, M. 1994. Auxiliaries and participles. In *Proceedings of NELS 24*, 65–79. GLSA, University of Massachusetts at Amherst.
Dennett, D. 1996. *Kinds of Minds*. New York: Basic Books.
Di Sciullo, A.-M., and C. Boeckx. In press. Contours of the Biolinguistic research agenda. In *The biolinguistic enterprise: New perspectives on the evolution and nature of the human language faculty*, ed. A.-M. Di Sciullo and C. Boeckx. Oxford: Oxford University Press.
Diamond, J.M. 1992. *The third chimpanzee: The evolution and future of the human animal*. New York: HarperCollins.
Dillon, B., and W. Idsardi. 2009. Investigating Statistical Approaches to Building a

Phonology. Ms., University of Maryland.
Donati, C. 2006. On wh-head-movement. In *Wh-movement: Moving on*, ed. L. Cheng and N. Corver, 21–46. Cambridge, Mass.: MIT Press.
Dowty, D. 1989. On the semantic content of the notion of "thematic role". In *Properties, types and meaning*, ed. G. Chierchia, B. Hall Partee, and R. Turner, 69–129. Dordrecht: Kluwer.
Dupuy, J.-P. 2009. *On the Origins of Cognitive Science: The Mechanization of the Mind*. Cambridge, Mass.: The MIT Press, second edition.
Emonds, J. E. 1978. The Complex V-V′ in French. *Linguistic Inquiry* 9:151–175.
Emonds, J. E. 1985. *A unified theory of syntactic categories*. Dordrecht: Foris.
Emonds, J. E. 2008. Valuing v-features and n-features: what adjuncts tell us about case, agreement, and syntax in general. In *Merging Features: Computation, Interpretation, and Acquisition*, ed. J.M. Brucart, A. Gavarró, and J. Solà, 194–214. Oxford: Oxford University Press.
Epstein, S. D. 2003[2007]. On i(nternalist)-functonal explanaion in minimalism. *Linguistic Analysis* 33:20–53.
Epstein, S. D., and M. Obata. In press. Deducing Improper Movement from Phase-based C-to-T Phi-Transfer: Feature-Splitting Internal Merge. In *Proceedings of WCCFL 27*. Somerville, Mass.: Cascadilla Press.
Epstein, S. D., and T. D. Seely. 2002. Rule applications as cycles in a level-free syntax. In *Derivation and explanation in the minimalist program*, ed. S. D. Epstein and T. D. Seely, 65–89. Oxford: Blackwell.
Epstein, S.D., and T.D. Seely. 2006. *Derivations in Minimalism*. Cambridge: Cambridge University Press.
Ettlinger, G. 1990. "Object vision" and "spatial vision": The neuropsychological evidence for the distinction. *Cortex* 26:319–341.
Fanselow, G., and D. Cavar. 2002. Distributed deletion. In *Theoretical approaches to universals*, ed. A. Alexiadou, 65–107. Amsterdam: John Benjamins.
Fasanella-Seligrat, A., and C. Rubio. 2010. Genuinely label-less syntax. Ms., Universitat Autònoma de Barcelona. Presented at the 2010 Vigo Summer School in Linguistics conference on Language, Nature, and Cognition.
Ferrer i Cancho, R., and N. Forns. 2009. The self-organization of genomes. *Complexity* 15:34–36.
Fitch, W.T. 2009. Prolegomena to a Future Science of Biolinguistics. *Biolinguistics* 3:283–320.
Fodor, J.A. 1883. *The modularity of mind*. Cambridge, Mass.: MIT Press.
Fodor, J.A. 1975. *The language of thought*. Cambridge, Mass.: Harvard University Press.

Fodor, J.A. 1998a. *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.
Fodor, J.A. 1998b. *In critical condition: Polemical essays on cognitive science and the philosophy of mind*. Cambridge, Mass.: MIT Press.
Fortuny, J. 2008. *The emergence of order in syntax*. Amsterdam: John Benjamins.
Fox, D. 2000. *Economy and semantic interpretation*. Cambridge, Mass.: MIT Press.
Fox, D. 2002. Antecedent-Contained Deletion and the Copy Theory of Movement. *Linguistic Inquiry* 33:63–96.
Frampton, J., and S. Gutmann. 2002. Crash-proof syntax. In *Derivation and explanation in the minimalist program*, ed. S. D. Epstein and T. D. Seely, 90–105. Oxford: Blackwell.
Friedmann, N. 2002. Question Production in Agrammatism: The Tree Pruning Hypothesis. *Brain and Language* 80:160–187.
Fukui, N. 2006. *Theoretical comparative syntax*. London: Routledge.
Fukui, N., and M. Speas. 1986. Specifiers and projections. *MIT Working Papers in Linguistics* 8:128–172.
Fukui, N, and M. Zushi. 2008. On Certain Differences between Noun Phrases and Clauses. In *Essays on nominal determination*, ed. H. Høeg Müller and A. Klinge, 265–286. Amsterdam: John Benjamins.
Gallego, Á. 2007a. An L-syntax for adjuncts. Universitat Autònoma de Barcelona.
Gallego, Á. 2007b. Phase theory and parametric variation. Doctoral Dissertation, Universitat Autònoma de Barcelona.
Gallego, Á. 2009a. Ellipsis by phase. Ms., Universitat Autònoma de Barcelona.
Gallego, Á. 2009b. A note on C. Ms., Universitat Autònoma de Barcelona.
Gallego, Á. 2009c. The structure of Phases. Ms., Universitat Autònoma de Barcelona.
Gallego, Á. In press. *Phase Theory*. Amsterdam: John Benjamins.
Gallego, Á., and M. Yoshida. 2008. Phases and ellipsis. Presented at WCCFL 27.
Gallistel, CR. 2007. L'apprentissage de matières distinctes exige des organes distincts. In *Cahier chomsky*, ed. J. Bricmont & J. Franck. Paris: L'Herne.
Gengel, K. To appear. Phases and ellipsis. In *Proceedings of the NELS 37*. Amherst, Mass.: GLSA.
Georgi, D., and G. Müller. 2010. Noun phrase structure by reprojection. *Syntax* 13.
Goldberg, A.E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
Gould, S.J. 1977. *Ontogeny and phylogeny*. Cambridge, Mass.: Harvard University Press.
Grimshaw, J. 1991. Extended projection. Ms., Brandeis University. [Reprinted as Grimshaw 2005. In Words and Structure. CSLI Lecture Notes Number 151, Stan-

ford, Calif.: CSLI Publications].
Grohmann, K. K. 2003. *Prolific Domains. On the Anti-Locality of Movement Dependencies*. Amsterdam: John Benjamins.
Grohmann, K. K. In press. Anti-locality. In *Oxford Handbook of Linguistic Minimalism*, ed. C. Boeckx. Oxford: Oxford University Press.
Haider, H. 1993. *Deutsche Syntax-generativ*. Tübingen: Narr.
Hale, K., and S. J. Keyser. 1993. On argument structure and the lexical expression of grammatical relations. In *The view from Building 20: Essays in Linguistics in honor of Sylvain Bromberger*, ed. K. Hale and S. J. Keyser, 53–110. Cambridge, Mass.: MIT Press.
Hale, K., and S. J. Keyser. 2002. *Prolegomenon to a theory of argument structure*. Cambridge, Mass.: MIT Press.
Harbour, D. 2006. The elimination of geometry. Ms, Queen Mary University of London.
Harbour, D. 2009. The semantics, and generality, of features: or, how not to construct a theory of cognitive evolution. Ms, Queen Mary University of London.
Hauser, M. D., N. Chomsky, and W. T. Fitch. 2002. The Faculty of Language: What is it, who has it, and how did it evolve? *Science* 298:1569–1579.
Hauser, M.D. 2008. Humaniqueness. Talk given at AAAS 2008, Boston. A version to appear in *Nature* as "Mind the gap!".
Hauser, M.D. 2009a. Origin of the mind. *Scientific American* September:44–51.
Hauser, M.D. 2009b. The possibility of impossible cultures. *Nature* 460:190–196.
Heck, F. 2004. A Theory of Pied Piping. Doctoral Dissertation, Universität Tübingen.
Heine, B., and T. Kuteva. 2007. *The genesis of grammar: a reconstruction*. Oxford: Oxford University Press.
Herburger, E. 2000. *What Counts: Focus and Quantification*. Cambridge, Mass.: MIT Press.
Hinzen, W. 2006. *Mind design and minimal syntax*. Oxford: Oxford University Press.
Hinzen, W. 2007. *An Essay on Naming and Truth*. Oxford: Oxford University Press.
Hiramatsu, K. 2000. Accessing linguistic competence: evidence from children's and adults' acceptability judgments. Doctoral Dissertation, University of Connecticut.
Hodges, A. 1983. *The Enigma*. New York: Simon and Schuster.
Holmberg, A. 2001. The syntax of yes and no in Finnish. *Studia linguistica* 55:141–175.
Hornstein, N. 2001. *Move! A minimalist theory of construal*. Oxford: Blackwell.
Hornstein, N. 2009. *A theory of syntax*. Cambridge: Cambridge University Press.

Hornstein, N., and J. Nunes. 2008. Some thoughts on adjunction. *Biolinguistics* 2:57–86.
Hornstein, N., J. Nunes, and K.K. Grohmann. 2006. *Understanding minimalism*. Cambridge: Cambridge University Press.
Hornstein, N., and J. Uriagereka. 2002. Reprojections. In *Derivation and explanation in the minimalist program*, ed. S. D. Epstein and T. D. Seely, 106–132. Oxford: Blackwell.
Huang, C.-T. J. 1982. Logical relations in Chinese and the theory of grammar. Doctoral Dissertation, MIT.
Irurtzun, A. 2007. The grammar of focus at the interfaces. Doctoral Dissertation, Euskal Herriko Unibertsitatea.
Jackendoff, R. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge, Mass.: MIT Press.
Jackendoff, R. 1997. *The architecture of the language faculty*. Cambridge, Mass.: MIT Press.
Jackendoff, R. 2002. *Foundations of language*. Oxford: Oxford University Press New York.
Jackendoff, R. 2005. Alternative minimalist visions of language. In *Proceedings from the 41st Annual Meeting of the Chicago Linguistic Society*, 2, 189–226.
Jackendoff, R. 2010. *Meaning and the Lexicon: The Parallel Architecture 1975–2010*. Oxford: Oxford University Press.
Jenks, P. 2010. Evidence for the Syntactic Diversity of Numeral Classifiers. Ms., Harvard University.
Jeong, Y. 2007. *Applicatives: Structure and interpretation from a minimalist perspective*. Amsterdam: John Benjamins.
Jeong, Y. In progress. On the natures of classifiers. Ms., Universitat Autònoma de Barcelona.
Jurka, J. 2009. Extraction out of subjects in german. Ms., University of Maryland.
Kamali, B. 2009. Phases of Incorporation and a three-way distinction for Direct Objects (in Turkish). Ms., Harvard University.
Kato, Takaomi. 2007. Symmetries in coordination. Ph.D. dissertation. Harvard University.
Kayne, R.S. 1994. *The antisymmetry of syntax*. Cambridge, Mass.: MIT Press.
Kayne, R.S. 1997. The English complementizer of. *The Journal of Comparative Germanic Linguistics* 1:43–54.
Kayne, R.S. 2005. *Movement and silence*. Oxford: Oxford University Press.
Kayne, R.S. In press. Antisymmetry and the lexicon. In *The biolinguistic enterprise: New perspectives on the evolution and nature of the human language faculty*, ed.

A.-M. di Sciullo and C. Boeckx. Oxford: Oxford University Press.
Koster, J. 1987. *Domains and dynasties: The radical autonomy of syntax*. Dordrecht: Foris.
Koster, J. 2009. Ceaseless, unpredictable creativity: language as technology. *Biolinguistics* 3:61–92.
Koster, J. 2010. Language and tools. Ms., Universiteit Groningen.
Kotzgoulou, G. 2007. (Non-)extraction from subjects as an edge phenomenon. Ms., University of the Aegean.
Krifka, M. 1992. Thematic relations as links between nominal reference and temporal constitution. In *Lexical matters*, ed. I. Sag and A. Szabolcsi, 29–53. Stanford, Calif.: CSLI Publications.
Lahne, A. 2008. Where there is fire there is smoke. local modelling of successive cyclic movement. Doctoral Dissertation, Universität Leipzig.
Lakoff, G. 1970. *Irregularity in syntax*. New York: Holt, Rinehart and Winston.
Landau, I. 2006. Severing the distribution of PRO from case. *Syntax* 9:153–170.
Landau, I. 2007. EPP Extensions. *Linguistic Inquiry* 38:485–523.
Lasnik, H. 1999a. Chains of arguments. In *Working minimalism*, ed. S. D. Epstein and N. Hornstein, 189–215. Cambridge, Mass.: MIT Press.
Lasnik, H. 1999b. *Minimalist Analysis*. Oxford: Blackwell.
Lasnik, H. 2000. *Syntactic Structures revisited*. Cambridge, Mass.: MIT Press.
Lasnik, H. 2001a. Subjects, objects, and the EPP. In *Objects and other subjects: Grammatical functions, functional categories, and configurationality*, ed. W. Davies and S. Dubinsky, 103–121. Dordrecht: Kluwer.
Lasnik, H. 2001b. When can you save a structure by destroying it? In *Proceedings of NELS 31*, 301–320. Amherst, Mass.: GLSA.
Lasnik, H. 2005. Review of Jason Merchant, *the syntax of silence*. *Language* 81:259–265.
Lasnik, H. 2006. Conceptions of the Cycle. In *Wh-movement: Moving On*, ed. L. Cheng and N. Corver, 197–216. Cambridge, Mass.: MIT Press.
Lasnik, H., and M. Saito. 1992. *Move α*. Cambridge, Mass.: MIT Press.
Lebeaux, D. 1988. Language acquisition and the form of the grammar. Doctoral Dissertation, University of Massachusetts, Amherst.
Leiber, J. 2002. Philosophy, engineering, biology, and history: a vindication of Turing's views about the distinction between the cognitive and physical sciences. *Journal of Experimental & Theoretical Artificial Intelligence* 14:29–37.
Lenneberg, E.H. 1967. *Biological foundations of language*. New York: Wiley.
Leu, T. 2008. The Internal Syntax of Determiners. Doctoral Dissertation, NYU.
Levinson, L. 2007. The roots of verbs. Doctoral Dissertation, NYU.

Lewontin, R.C. 2000. *The Triple Helix: Gene, Organism, and Environment*. Cambridge, Mass.: Harvard University Press.
Longa, V.M., G. Lorenzo, and G. Rigau. 1996. Expressing Modality by Recycling Clitics. *Catalan Working Papers in Linguistics* 5:67–79.
Longa, V.M., G. Lorenzo, and G. Rigau. 1998. Subject clitics and clitic recycling: locative sentences in some Iberian Romance languages. *Journal of linguistics* 34:125–164.
Longobardi, G. 1994. Reference and proper names: a theory of N-movement in syntax and logical form. *Linguistic inquiry* 25:609–665.
Longobardi, G. 2003. Methods in parametric linguistics and cognitive history. *Linguistic Variation Yearbook* 3:101–138.
Longobardi, G. 2006. Reference to Individuals, Person, and the Variety of Mapping Parameters. In *Essays on nominal determination*, ed. H. H. Müller and A. Klinge, 189–211. Amsterdam: John Benjamins.
López, L. 2002. On Agreement: Locality and Feature Valuation. In *Theoretical approaches to universals*, ed. A. Alexiadou, 165–209. Amsterdam: John Benjamins.
López, L. 2007. *Locality and the architecture of syntactic dependencies*. New York: Palgrave Macmillan.
Lorenz, K. 1959. Psychologie und Stammesgeschichte. In *Evolution der organismen*, ed. G. Heberer, 131–170. Stuttgart: Fischer.
Manzini, M.R., and L.M. Savoia. 2005. *I dialetti italiani e romanci: morfosintassi generativa*. Alessandria: Edizioni dell'Orso.
Marantz, A. 1984. *On the Nature of Grammatical Relations*. Cambridge, Mass.: MIT Press.
Marantz, A. 1991. Case and licensing. In *Proceedings of ESCOL'91*, 234–253.
Marantz, A. 1995. The minimalist program. In *Government and Binding and the Minimalist Program*, ed. G. Webelhuth, 349–382. Malden: Blackwell.
Marantz, A. 1996. Cat as a Phrasal Idiom. Ms., MIT.
Marantz, A. 2005. Generative linguistics within the cognitive neuroscience of language. *The Linguistic Review* 22:429–446.
Marantz, A. 2008. Words and phases. In *Phases in the theory of grammar*, ed. S.-H. Choe, 191–222. Seoul: Dong In.
Marantz, A. 2010. Locality in Morphology: The Domains of Allomorphy and Allosemy. Lecture notes, Universitat Autònoma de Barcelona, May 2010.
Marcus, G. 2006. Cognitive architecture and descent with modification. *Cognition* 101:443–465.
Martín, J. In progress. Datives. Doctoral Dissertation, NYU.
Martínez-Ferreiro, S. 2010. Towards a characterization of agrammatism in Ibero-

Romance. Doctoral Dissertation, Universitat Autònoma de Barcelona.
Marušič, F. 2005. On non-simultaneous phases. Doctoral Dissertation, Stony Brook University.
Masullo, P.J. 2008. The syntax-lexical semantics interface: prepositionalizing motion verbs in Spanish. Ms., University of Pittsburgh.
Mateu, J. 2005. Impossible Primitives. In *The compositionality of meaning and content*, ed. M. Werning, E. Machery, and G. Schurz, 213–229. Heusenstamm: Ontos Verlag.
Matushansky, O. 2006. Head movement in linguistic theory. *Linguistic Inquiry* 37:69–109.
Mayr, E. 1959. Darwin and the evolutionary theory in biology. In *Evolution and anthropology: A centennial appraisal*, ed. J. Meggers, 1–10. Washington D.C.: The Anthropological Society of Washington.
McBrearty, S., and A.S. Brooks. 2000. The revolution that wasn't: a new interpretation of the origin of modern human behavior. *Journal of human evolution* 39:453–563.
McCloskey, J. 2001. The distribution of subject properties in Irish. In *Objects and other subjects: Grammatical functions, functional categories, and configurationality*, ed. W.D. Davies and S. Dubinsky, 157–192. Dordrecht: Springer.
McFadden, T. 2004. The position of morphological case in the derivation. Doctoral Dissertation, University of Pennsylvania.
McGinnis, M. 1998. Locality in A-movement. Doctoral Dissertation, MIT.
Mellars, P., K. Boyle, O. Bar-Yosef, and C. Stringer, ed. 2007. *Rethinking the human revolution*. Cambridge: McDonald Institute Monographs.
Merchant, J. 2001. *The syntax of silence: sluicing, islands, and identity in ellipsis*. Oxford: Oxford University Press.
Miller, G., and N. Chomsky. 1963. Finitary models of language users. In *Handbook of mathematical psychology*, ed. R. D. Luce, R. Bush, and E. Galanter, 419–491. New York: Wiley.
Mishkin, M., and L.G. Ungerleider. 1982. Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural Brain Research* 6:57–77.
Mithen, S.J. 1996. *The prehistory of the mind*. London: Thames and Hudson.
Miyagawa, S. 2010. *Why Agree; Why Move*. Cambridge, Mass.: MIT Press.
Moro, A. 2000. *Dynamic antisymmetry*. Cambridge, Mass.: MIT Press.
Moro, A. 2007. Some notes on unstable structures. Ms, Università Vita-Salute San Raffaele/Harvard University.
Müller, G. 2010. On deriving CED effects from the PIC. *Linguistic Inquiry* 41:35–82.

Narita, H. 2008. Multiple Transfer in service of recursive Merge: Feature percolation eliminated. Ms., Harvard University.
Narita, H. In progress. Full Interpretation of Minimalist Syntax. Doctoral Dissertation, Harvard University.
Newmeyer, F. 2004. Against a parameter-setting approach to language variation. *Linguistic Variation Yearbook* 4:181–234.
Obata, M. 2010. Root, Successive-cyclic and Feature-Splitting Internal Merge: Implications for Feature-Inheritance and Transfer. Doctoral Dissertation, University of Michigan.
Ochi, M. 2010. Classifiers and nominal structure: a parametric approach and its consequences. Ms., Osaka University.
Otero, C. 1976. The dictionary in a generative grammar. Presentend at the 91^{st} Annual Convention of the MLA, New York.
Otero, C. 1983. Towards a model of paradigmatic grammar. *Quaderni di Semantica* 4:134–144.
Otero, C. 1996. Head movement, cliticization, precompilation, and word insertion (comments on Uriagereka's paper). In *Current issues in comparative grammar*, ed. R. Freidin, 296–337. Dordrecht: Kluwer.
Ott, D. 2008. Notes on noun ph(r)ases. Ms., Harvard University.
Ott, D. 2009. The conceptual necessity of phases: Some remarks on the minimalist enterprise. In *Explorations of Phase Theory: Interpretation at the interfaces*, ed. K.K. Grohmann, 253–275. Berlin: Mouton de Gruyter.
Ott, D. To appear. A note on free relative clauses in the theory of phases. *Linguistic Inquiry* .
Pearl, L.S. 2007. Necessary bias in natural language learning. Doctoral Dissertation, University of Maryland.
Pesetsky, D. 1996. *Zero syntax: Experiencers and cascades*. Cambridge, Mass.: MIT Press.
Pesetsky, D., and E. Torrego. 2001. T-to-C movement: Causes and consequences. In *Ken Hale: A Life in Language*, ed. M. Kenstowicz, 355–426. Cambridge, Mass.: Mit Press.
Pesetsky, D., and E. Torrego. 2006. Probes, goals and syntactic categories. In *Proceedings of the 7th Annual Conference on Psycholinguistics*, ed. Y. Otsu, 24–60. Tokyo: Hituzi Syobo.
Pesetsky, D., and E. Torrego. 2007. The syntax of valuation and the interpretability of features. In *Phrasal and clausal architecture: Syntactic derivation and interpretation*, ed. S. Karimi, V. Samiian, and W. Wilkins, 262–294. Amsterdam: John Benjamins.

Phillips, C. 2006. The real-time status of island phenomena. *Language* 82:795–823.
Piattelli-Palmarini, M. 2010. What is language that it may have evolved, and what is evolution that it may apply to language. In *The evolution of human language: biolinguistic perspectives*, ed. R.K. Larson and H. Déprez, V. andYamakido, 148–162. Cambridge: Cambridge University Press.
Piattelli-Palmarini, M., and J. Uriagereka. 2005. The evolution of the narrow language faculty: the skeptical view and a reasonable conjecture. *Lingue e Linguaggio* IV:27–79.
Piattelli-Palmarini, M., and J. Uriagereka. In press. FoxP2: A geneticist's dream; a linguist's nightmare. In *The biolinguistic enterprise: New perspectives on the evolution and nature of the human language faculty*, ed. A.-M. di Sciullo and C. Boeckx. Oxford: Oxford University Press.
Picallo, M.C. 2006. On gender and number. Ms., Universitat Autònoma de Barcelona.
Picallo, M.C. 2008. On Gender and Number in Romance. *Lingue e Linguaggio* 1/2008:44–66.
Piera, C. 1985. On the representation of higher order complex words. In *Selected papers from the XIIIth Linguistic Symposium on Romance Languages, Chapel Hill, NC, 24-26 March 1983*, 287–313. Amsterdam: John Benjamins.
Pietroski, P. M. 2003. Quantification and second-order monadicity. *Philosophical Perspectives* 17:259–298.
Pietroski, P. M. 2005. *Events and semantic architecture*. Oxford: Oxford University Press.
Pietroski, P. M. 2007. Systematicity via monadicity. *Croatian Journal of Philosophy* 7:343–374.
Pietroski, P. M. To appear. *Semantics without truth values*. Oxford: Oxford University Press.
Pinker, S., and R. Jackendoff. 2005. The faculty of language: what's special about it? *Cognition* 95:201–236.
Poeppel, D. 2005. The interdisciplinary study of language and its challenges. Technical report, Jahrbuch des Wissenschaftskollegs zu Berlin.
Poeppel, D., and D. Embick. 2005. Defining the relation between linguistics and neuroscience. In *Twenty-first century psycholinguistics: Four cornerstones*, ed. A. Cutler, 173–189. Hillsdale, NJ:: Erlbaum.
Pollock, J.-Y. 1989. Verb movement, universal grammar, and the structure of IP. *Linguistic inquiry* 20:365–424.
Pullum, G.K. 1996. Nostalgic views from Building 20. *Journal of Linguistics* 32:137–147.

Pylkkänen, L. 2002. Introducing arguments. Doctoral Dissertation, MIT.
Pylkkänen, L. 2008. *Introducing arguments*. Cambridge, Mass.: MIT Press.
Pylyshyn, Z.W. 2007. *Things and Places: How the Mind Connects with the World*. Cambridge, Mass.: MIT Press.
Ramchand, G. 2008. *Verb meaning and the lexicon: a first-phase syntax*. Cambridge: Cambridge University Press.
Raposo, E. 2002. Nominal gaps with prepositional modifiers in Portuguese and Spanish: A case for Quick Spell-Out. *Cuadernos de Lingüística del I. U. Ortega y Gasset* 9:127–144.
Reuland, E. 2001. Primitives of binding. *Linguistic Inquiry* 32:439–492.
Richards, M. 2007. On feature inheritance: An argument from the phase impenetrability condition. *Linguistic Inquiry* 38:563–572.
Richards, M. 2009a. Internal Pair-Merge: the Missing Mode of Movement. *Catalan Journal of Linguistics* 8:55–73.
Richards, M. 2009b. Stabilizing syntax: On instability, optionality, and other indeterminacies. Presented at the Syntax workshop, Universität Stuttgart, March 2010.
Richards, M. To appear. Deriving the edge: What's in a phase? *Syntax* .
Richards, N. 2001. *Movement in language: interactions and architectures*. Oxford: Oxford University Press.
Richards, N. 2008. Lardil "case stacking" and the structural/inherent case distinction. Ms., MIT.
Richards, N. 2010. *Uttering trees*. Cambridge, Mass.: MIT Press.
van Riemsdijk, H. 2008. Identity Avoidance: OCP-Effects in Swiss Relatives. In *Foundational issues in linguistics*, ed. R. Freidin, C. Otero, and M.-L. Zubizarreta, 227–250. Cambridge, Mass.: MIT Press.
Rizzi, L. 1990. *Relativized minimality*. Cambridge, Mass.: MIT Press.
Rizzi, L. 1997. The fine structure of the left periphery. In *Elements of grammar: Handbook in generative syntax*, ed. L. Haegeman, volume 1997, 281–337. Dordrecht: Kluwer.
Rizzi, L. 2001. On the position"Int(errogative)" in the left periphery of the clause. In *Current studies in Italian syntax*, ed. G. Cinque and G. Salvi, 287–296. Amsterdam: Elsevier.
Rizzi, L. 2004. On the cartography of syntactic structures. In *The Structure of CP and IP*, ed. L. Rizzi, 3–15. Oxford: Oxford University Press.
Rizzi, L. 2006. On the Form of Chains: Criterial Positions and ECP Effects. In *Wh-movement: Moving on*, ed. L. Cheng and N. Corver, 97–133. Cambridge, Mass.: Mit Press.

Rizzi, L. 2010. On the elements of syntactic variation. Presented at the workshop on linguistic variation and the minimalist program, Barcelona, January 2010.
Rizzi, L., and U. Shlonsky. 2007. Strategies of subject extraction. In *Interfaces + recursion = language? Chomsky's minimalism and the view from semantics*, ed. U. Sauerland and H.-M. Gärtner, 115–160. Mouton de Gruyter.
Roberts, I.G. 2010. *Agreement and Head Movement: Clitics, Incorporation, and defective Goals*. Cambridge, Mass.: MIT Press.
Rubio, C. 2010. Thesis prospectus. Ms., Universitat Autònoma de Barcelona.
Saito, M., T.H. J. Lin, and K. Murasugi. 2008. N′-ellipsis and the structure of noun phrases in Chinese and Japanese. *Journal of East Asian Linguistics* 17:247–271.
Samuels, B. 2009. The structure of phonological theory. Doctoral Dissertation, Harvard University.
Samuels, B. To appear. *Phonological Architecture: A Biolinguistic Perspective*. Oxford: Oxford University Press.
Saunders, P. T. 1992. *Collected works of A. M. Turing: Morphogenesis*. Amsterdam: North Holland.
Searle, J. 1972. Chomsky's revolution in linguistics. *The New York Review of Books* 18:16–24.
Shlonsky, U. 2006. Extended projection and CP cartography. *Nouveaux cahiers de linguistique française* 27:83–93.
Sigurdsson, H. Á. 2004. The syntax of Person, Tense, and speech features. *Rivista di Linguistica* 16:219–251.
Simpson, A. 2005. Classifiers and DP structure in Southeast Asia. In *The oxford handbook of comparative syntax*, ed. G. Cinque and R.S. Kayne, 806–838. Oxford: Oxford University Press.
Snyder, W. 1995. Language acquisition and language variation: The role of morphology. Doctoral Dissertation, MIT.
Snyder, W. 2000. An Experimental Investigation of Syntactic Satiation Effects. *Linguistic Inquiry* 31:575–582.
Snyder, W. 2001. On the nature of syntactic variation: Evidence from complex predicates and complex word-formation. *Language* 77:324–342.
Speas, P. 2010. The minimal structure of the left periphery. Presented at the Language Design conference, UQAM, Montreal, May 2010.
Speas, P., and C. Tenny. 2003. Configurational properties of point of view roles. In *Asymmetry in grammar Vol I: Syntax and semantics*, ed. A.-M. di Sciullo, 315–344. Amsterdam: John Benjamins.
Spelke, E. 2003. What makes us smart? Core knowledge and natural language. In *Language and Mind: Advances in the study of language and thought*, ed. D. Gen-

tner and S. Goldin-Meadow, 277–311. Cambridge, Mass.: MIT Press.
Sprouse, J. 2007. A Program for Experimental Syntax. Doctoral Dissertation, University of Maryland.
Starke, M. 2001. Move Dissolves into Merge. Doctoral Dissertation, Université de Genève.
Starke, M. 2010. Nanosyntax: A short primer to a new approach to language. *Nordlyd* 36.
Stepanov, A. 2001a. Cyclic domains in syntactic theory. Doctoral Dissertation, University of Connecticut.
Stepanov, A. 2001b. Late Adjunction and Minimalist Phrase Structure. *Syntax* 4:94–125.
Stepanov, A. 2007. The End of CED? Minimalism and Extraction Domains. *Syntax* 10:80–126.
Svenonius, P. 2007. Adpositions, particles and the arguments they introduce. In *Argument structure*, ed. K.V. Subbarao T. Bhattacharya and E. Reuland, 63–103. Amsterdam: John Benjamins.
Szabolcsi, A. 1984. The possessor that ran away from home. *The Linguistic Review* 3:89–102.
Szabolcsi, A. 2006. Strong vs. weak islands. In *The Blackwell Companion to Syntax*, ed. M. Everaert and H. van Riemsdijk, volume 4, 479–531. Malden: Blackwell.
Takahashi, S. 2010. The hidden side of clausal complements. *Natural Language and Linguistic Theory* 28:343–380.
Takahashi, S., and S. Hulsey. 2009. Wholesale Late Merger: Beyond the A/$\overline{A}$-distinction. *Linguistic Inquiry* 40:387–426.
Tattersall, I. 1998. *Becoming Human: Evolution and Human Uniqueness*. Oxford: Oxford University Press.
Truswell, R. 2007. Locality of Wh-movement and the Individuation of Events. Doctoral Dissertation, University College London.
Tsai, W.D. 2010. High applicatives are not high enough: a cartographic solution. Ms., Chinese University of Hong Kong.
Tungseth, M.E. 2008. *Verbal Prepositions and Argument Structure: Path, Place and Possession in Norwegian*. Amsterdam: John Benjamins.
Turing, A. M. 1936. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society* 2:230–265.
Turing, A. M. 1952. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London Series B* 237:37–72.
Turing, A. M., and C. Wardlaw. 1953 [1992]. A diffusion reaction theory of morphogenesis in plants. In *Collected Works of A. M. Turing: Morphogenesis*, ed. P. T.

Saunders, 37–118. Amsterdam: North Holland.
Uriagereka, J. 1988. On Government. Doctoral Dissertation, University of Connecticut.
Uriagereka, J. 1995. Aspects of the syntax of clitics in Western Romance. *Linguistic Inquiry* 26:79–123.
Uriagereka, J. 1998. *Rhyme and reason: An introduction to minimalist syntax*. Cambridge, Mass.: MIT Press.
Uriagereka, J. 1999. Multiple Spell-Out. In *Working minimalism*, ed. S. D. Epstein and N. Hornstein, 251–282. Cambridge, Mass.: MIT Press.
Uriagereka, J. 2008. *Syntactic Anchors: On Semantic Structuring*. Cambridge: Cambridge University Press.
Uriagereka, J. To appear. *Spell-Out and the minimalist program*. Oxford: Oxford University Press.
Vicente, L. 2007. The syntax of heads and phrases. Doctoral Dissertation, Universiteit Leiden.
Watanabe, A. 1995. Conceptual basis of cyclicity. In *MIT working papers in linguistics 27*, 269–291.
Watanabe, A. 2009. Vague Quantity, Numerals, and Natural Numbers. *Syntax* 13:37–77.
Wexler, K., and P. Culicover. 1980. *Formal Principles of Language Acquisition*. Cambridge, Mass.: MIT Press.
Wurmbrand, S. 2006. Licensing Case. *Journal of Germanic Linguistics* 18:175–236.
Yang, C. 2004. Universal Grammar, statistics or both? *Trends in Cognitive Sciences* 8:451–456.
Yang, C. 2010. Three factors in language acquisition. *Lingua* 120:1160–1177.