1 *Biolinguistic concerns*

1.1 Approaching Universal Grammar *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 – i.e. a flourishing *biolinguistics* – 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 to keep us away from the biological 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 do 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 when attempting to delineate the neurobiology of the language faculty. Internal obstacles are more difficult to deal with, for those are habits that were adopted early during the practice of

[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" (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.

linguistics-as-cognitive-science, that rendered initial progress possible, but that can subsequently prove an impediment to further progress. In this contribution I want to examine one such factor, which I will refer to as 'lexicocentrism.'

By 'lexicocentrism' I mean the theoretical stance, shared across many frameworks, that appeals to the 'lexicon' (I put the term in scare quotes because of the many definitions it has received in the literature; for my purposes here, understand lexicon as the repository of elements on which syntax feeds) to account for most, if not all of what many would regard as core properties of the language faculty (detailed illustrations of this stance will follow momentarily, and will be found throughout the book). If one is interested in what Saussure called the arbitrariness of the sign, appeal to the lexicon is, of course, in order. It is a brute lexical fact that Catalan speakers say *gos* to refer to DOG, but French speakers say *chien* to refer to the same thing. But if one is interested in more grammatical facts, such as the ability for a noun to combine with a verb, or even, I will suggest, the very existence of categories like 'noun' and 'verb,' or in the patterns of cross-linguistic variation, then lexicocentric accounts retain their arbitrary character, and leave unanswered – indeed, they often make it hard to ask – certain questions that are well worth reflecting upon.

Take, for example, the currently standard treatment of displacement in transformational generative grammar. In situations like *Who did Mary kiss?*, we learn that *who* appears pronounced where it is (say, SpecCP) because the relevant functional head (C) in English has a particular lexical need that can only be satisfied through the presence of a lexical item with the appropriate lexical specification ([wh]-feature) in its specifier. Such an account is then generalized to the following condition ('principle'): displacement takes place only to satisfy the lexical demands of the host.[2] This in turn leads researchers to posit lexical properties on functional heads (e.g., "EPP" features) just because an element appears to have been displaced in their vicinity. I am certainly not the first to have noticed the arbitrary character of this kind of explanation. Perhaps no one said it as well as George Lakoff in the following passage, quoted in Kibort and Corbett (2010, 31):

> So linguists fudge, just as has been done in the reflexive rule, by sticking on the arbitrary feature +REFL. Such a feature is a fudge. It might as well be called +CHOCOLATE, which would in fact be a better name, since it would clearly reveal the nature of the fudge.

---

[2] In the minimalist literature, this is known as the "Attract" principle (Chomsky (1995, chap. 4)), or sometimes also "Suicidal Greed" (Chomsky (2000a)). Alternative accounts of movement ("Enlightened Self-Interest" (Lasnik (1999)) or "Greed" (Chomsky (1993))) proposed in the literature are equally lexicocentric.

In the early 1980s, Muysken and van Riemsdijk (1986, preface) correctly
pointed out that

> Even basic questions [concerning features] such as 'how many are there?',
> 'what are they?', 'how do they distribute over syntactic structures?' were
> hardly addressed, let alone answered. Nevertheless, it is clear that syntac-
> tic features do play an important role in syntax . . . It would appear to be
> high time, therefore, to examine the theory of syntactic features in a more
> systematic way.

In their introduction to the volume, Muysken and van Riemsdijk write that
"too little is known about [features]." More than twenty-five years later, similar
remarks still apply. In their volume on "features as key notions in linguistics,"
Kibort and Corbett (2010) write that "the impact of features has increased
steadily . . . features [are] essential to how we do linguistics" (p. 2), but quickly
add that "there is much more to be understood about features" (p. 3). Strik-
ingly, though, what one repeatedly finds in the literature is an attempt to
replace one obviously stipulative lexicocentric account with another, perhaps
at first less obviously stipulative, but equally lexicocentric account (see, for
instance, Fukui and Speas' early (1986) treatment of the "Extended Projec-
tion Principle (EPP)" in terms of structural case, or the opposite suggestion in
Marantz (1991)).

   Readers may well wonder why lexicocentrism has dominated linguistic the-
orizing for so long. I think that part of the answer lies in the fact that for all
its explanatory limitation lexicocentrism has revealed important grammatical
facts and patterns and has made it possible to formulate interesting general-
izations that (in the words of Eric Reuland) are "too good to be false." We all
have to start somewhere. But I think there is another reason why lexicocen-
trism has maintained such a strong foothold in theoretical linguistics. For all
the emphasis on the biological substrate of the language capacity in Chom-
sky's writings since the beginning of the so-called cognitive revolution of the
mid 1950s, most theoretical linguists, even those of a Chomskyan persuasion,
remain fundamentally interested in languages as objects of study,[3] and for
languages it makes sense to start with the lexicon since this is clearly the most

---

[3] As evidence for this assertion, consider the following samples of replies provided by
alumni/visitors of the MIT linguistics program on the occasion of its 50th anniversary, when
they were asked "What was the broad question that you most wanted to get an answer to during
your time in the program?" (http://ling50.mit.edu/category/replies):

> Since the beginning of my student career in the early 1970s I had been fascinated
> with the issue of how aspects of grammatical diversity cluster across languages
> (G. Longobardi).

4   *1 Biolinguistic concerns*

distinctive thing about them. But what if the focus of inquiry is the language faculty (that which makes the acquisition of specific languages possible)? Here I think the dominance of lexicocentrism is down to the weight of tradition, of thinking of Universal Grammar as a grammar, organized like the traditional grammars of specific languages, where syntax is thought to be the study of how words are put together. Of course, if that is what syntax is, it makes eminent sense to start with the words, just like in semantics, one starts with lexical meaning to capture compositionality. As Caponigro and Polinsky (2011) state matter-of-factly, "the lexicon of a language is expected to shape its syntax."

1.2   Infrequently asked questions

Most linguists, I suspect, would endorse Polinsky and Caponigro's statement, 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" in current minimalism, and that makes it possible to claim that "labels can be eliminated" (Collins (2002)),[4] that "syntax is crash-proof" (i.e., driven

I was preoccupied to know what should be the correct relationship between linguistic theory and language description (A. Salanova)

As a student who had been strongly attracted by grammars of L (= English, Latin, German, Greek, French) and holder of a mathematics MA, what attracted me to the MIT program, via Chomsky's writings, was the sense that at least preliminary explicit formulations of these grammars of *L* were in sight—not during my stay at MIT, but in say a couple of decades.

With almost everyone else, I was convinced from the first of '... the necessity for supplementing a "particular grammar" by a universal grammar if it is to achieve descriptive adequacy.' (Aspects of the Theory of Syntax: 6). Thus, I understood,

(1) Grammar of L = UG + G_1 (= a Particular Grammar of L_1)

These grammars, supplemented by UG, were to generate all and only grammatical sequences of the L_1. So, the broad question had two parts: what was UG, perhaps the hardest part, and what were the (formalized, explicit) Particular Grammars, a supposedly easier question. Nonetheless, the second part also seemed intriguing and puzzling, since, beyond some generalities, exact aspects of e.g. English and French grammars had little in common. (Kayne's dissertation, his later *French Syntax*, didn't seem to be a book about English grammar.) Thus in addition to UG, "the broad question for which I most wanted to get an answer to" was:

(2) What exactly is the form of particular grammars that UG can then 'supplement'? [J. Emonds]

Felix (2010) contains a deeply insightful discussion of how these goals differ from Chomsky's (biolinguistic) motivation, and of the unfortunate consequences this can lead to.

[4] Collins's work on labels is often misunderstood, I think. He crucially did not argue for the elimination of labeling or headedness, but rather for a representational change: a replacement

by featural needs of 'pivots'; Frampton and Gutmann (2002)), and also the one that facilitates the endorsement of the "Borer–Chomsky Conjecture" concerning parametric variation (the idea that all variation reduces to lexical choices). It is indeed *the* cornerstone on which our modern conception of Principles-and-Parameters 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 whom 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 yet, I will argue that lexicocentrism is wrong. Deeply wrong. In fact, it may be the biggest (internal) obstacle that lies on the road towards a level of explanation that Chomsky has referred to as "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 (see Hauser *et al.* (2002) on "the Faculty of Language in the Broad sense (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 (pre-merged) 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" (2006, 7). Instead of "approaching syntax (and UG) from below" (Chomsky (2007)), minimalist syntacticians approach it from the lexicon, *in medias res*, and as such they do not depart at all from pre-minimalist practice (or, for that matter, from the practice of traditional grammarians).

of standardly labeled nodes by a lexically/featurally defined set of asymmetric prominence relations. See Adger (2013b) for an unusually clear statement of Collins's approach. See also Boeckx (2008b).

Newmeyer (2004, 226 n.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 been seen 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. In this sense, minimalism has so far failed to distinguish itself from previous transformational accounts, which relied on a principle made explicit in the Government-and-Binding era: the Projection Principle. As is clear from popular textbooks such as Haegeman (1994), "the projection principle: i.e. the idea that all syntactic structure is projected from the lexicon" was taken as "a basis for the organization" of the grammar. It is the central dogma regarding the flow of information in the grammar. Whereas minimalist syntacticians insisted on the demise of "government" as proof of the difference between minimalism and the models preceding it (see, e.g., Bošković and Lasnik (2007)),[5] they remained extremely conservative when it comes to the relationship between syntax and the lexicon (arguably, a more fundamental architectural property of grammar than 'government').

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 (an imperfection), it's a feature. These days, we have features for everything: structure-building features/merge-features (aka edge-features), agree-features (aka unvalued/uninterpretable features), move features (aka EPP-features), to say nothing of all the fine-grained featural distinctions ('flavors') brought about by the intensive cartographic projects that currently dominate syntactic inquiry. The problem is clear: in the absence of any realistic, grounded, cognitively

---

[5] In the end, the lack of appeal to "government" turned out to be a relatively superficial difference, given that the Agree-relation (Chomsky (2000a)) is quite similar to it.

sound, biologically plausible theory of what counts as a possible feature, 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. As Šimík (2011) correctly states

> It is commonly assumed that formal features on syntactic categories are essentially descriptive devices—remnants of construction-based approaches to grammar. They are often introduced into the grammatical model in order to bridge the gap between the empirical facts we face and the assumptions which we believe to follow from independently motivated principles. In that respect, the postulated inventory and properties of formal features provide a useful overview of what we do not understand.

(In their volume on features, Kibort and Corbett (2010) seem to share Šimík's view when they write that "features are fundamental to linguistic description" (p. 1), but then puzzlingly add that "linguists frequently turn to them as they try to understand . . . the complexity of natural language." I say puzzlingly because I agree with Šimík that features obscure understanding, they label our ignorance.)

The problem with lexicocentrism is in fact even more severe once we realize that the basic units manipulated by syntax (lexical items/categories) are defined not as single features but as "feature-bundles" (see, among many others, Sprouse and Lau (2013): "we believe it is fair to say that there is some degree of consensus that the basic units are bundles of features"). Chomsky (2007, 6) makes it very explicit in the following passage: "In addition to Merge applicable without bounds, UG must at least provide atomic elements, lexical items LI, each a *structured* array of properties (features) to which Merge and other operations apply to form expressions" (my emphasis).[6] Such bundles are nothing more than little syntactic trees. How such treelets are constructed is left unaddressed, and in fact asking the question quickly leads to a paradox: if such treelets are built by merge, much like the regular syntactic trees that they resemble so much, why is merge at the sentential level said to require featural triggers, but merge at the lexical level is not? As we will see in detail shortly, featural triggers are nothing other than feature bundles, but then we are stuck

---

[6] The idea that syntactic features are internally organized is far from new or exclusive to minimalism. Although the references just given in the text are recent ones, many syntacticians adopted this idea long ago. As Muysken and van Riemsdijk (1986, 19) pointed out, "eventually, one . . . expect[s] there to be a full-fledged theory in which features are grouped into hierarchically-ordered classes and subclasses, like in phonology."

in the following loop: merge requires feature bundles to apply, and feature bundles require merge to exist! One could, of course, appeal to a new operation to construct feature bundles – call it Bundle – but, all else equal (I return to this issue in Chapter 3), this would duplicate entities in a way that flatly violates minimalism's favorite tool, Occam's razor. If Bundle constructs structures that look like those constructed by Merge, if Bundle swims like Merge, and quacks like Merge, then Bundle is Merge. The more so, if Merge operates on features, as Chomsky compellingly argued for in Chomsky (1995) when he introduced the notion of Feature-movement.

Tellingly, although the argument for moving 'just' features was a rather bold one, Chomsky did not go all the way. Although he recognized that moving just F (F a feature) was the most natural hypothesis within the framework he was considering, as the following passage reveals:

> So far I have kept to the standard assumption that the operation Move selects α and raises it, targeting K, where α and K are categories constructed from one or more lexical items. But on general minimalist assumptions, that is an unnatural interpretation of the operation. The underlying intuitive idea is that the operation Move is driven by morphological considerations: the requirement that some feature F must be checked. The minimalist operation, then, should raise just the feature F.   (1995, 262)

Chomsky asked in the following paragraph "when F is raised to target K, why does F not raise alone . . .?" He went on to write:

> The answer should lie in a natural economy condition.
> (26) F carries along just enough material for convergence.
> The operation Move, we now assume, seeks to raise just F. Whatever "extra baggage" is required for convergence involves a kind of "generalized pied-piping" . . . For the most part—perhaps completely—it is properties of the phonological component that require such pied-piping.

But the pages following the passage just quoted make it clear that the situations Chomsky is considering here are situations where phonological (and perhaps) semantic features must raise along with syntactic features. Crucially, for purposes of the present discussion, Chomsky never in fact considered breaking the syntactic feature bundle to raise just F (F a syntactic feature). As he writes on p. 265:

> When the feature F on the lexical item LI raises without pied-piping of LI or any larger category α, as always in covert raising, does it literally raise alone or does it automatically take other formal features along with it? There are strong empirical reasons for assuming that Move F automatically carries along FF(LI), the set of formal features of LI. We therefore understand the

operation move F in accord with (28), where FF[F] is FF(LI), F a feature of
the lexical item LI.
(28) Move F "carries along" FF[F].

Although unfortunately he does not elaborate on which "strong empirical rea-
sons" he has in mind, Chomsky here makes it clear that he never in fact fully
entertained the most natural hypothesis (Move *just* F, F a feature). Although
he argued for splitting the syntactic feature bundle from the phonological and
semantic feature bundles that together make up a lexical item, he never went
as far as breaking the syntactic feature bundle itself.

By not doing so, Chomsky kept the nature of this bundle shrouded in mys-
tery. In effect, bundles in minimalism retain the status of constructions in
non-transformational frameworks: they are templates whose origins one is not
supposed to discuss.[7]

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)

---

[7] The only explicit passage regarding this issue that I have been able to find in Construction-
friendly approaches is the following, from Jackendoff (2011, 602), who, after four or five
books and many articles praising the superiority of "Unify" over Merge as the central operation
in the grammar (including in the very article from which this passage is drawn!), acknowl-
edges the limitations of such an operation (one can only unify structures that have been created
beforehand, but what is responsible for this?; see Boeckx and Piattelli-Palmarini (2007)):

> I should make clear that Unification alone cannot create constituent structure: it only
> creates a Boolean combination of pre-existing features and structures. In order to build
> structure, one needs a skeletal constituent structure that can be unified with two or more
> items. Such a skeleton is of course already richly present in cognition: the part-whole
> schema. One formal realization of this schema is a set {x, y} with variable elements
> *x* and *y* as parts. This can be unified with specific elements A and B to form the set
> {A, B}—in effect the output of Merge. Similarly, a linearly ordered constituent [A∧B]
> can be licensed by the unification of A and B with a linearly ordered schema [x∧y],
> which is also ubiquitous in nonlinguistic cognition. Thus the effects of Merge can be
> constructed from Unification and one of these schemas.

> One might say then that these schemas are nothing but constraint-based counterparts of
> Merge, and this would be partly correct.

This passage makes it clear that Unify and Merge both fail to get to the heart of the matter,
since they assume the existence of preformed structures. Jackendoff hints at the possibility that
such structures may have non-linguistic origins, but if that is the case, then why are we the only
species that has the kind of syntax we do?

> 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) (for a similar statement, see Rizzi (2009))

Although such statements give the appearance of a very restrictive theory of
language variation – indeed, they have been argued to provide such a restrictive
theory – in the absence of a theory behind terms like "lexical item," "the (func-
tional) lexicon," or its "inflectional component," they amount to little more
than disguised statements of ignorance, or wishful thinking.[8] 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 gram-
mar, a list of basic irregularities" (a conception already expressed in Chomsky
(1965), and reiterated in Chomsky (1995); a conception ultimately going back
to Bloomfield and the structuralists). If that is the lexicon, surely we cannot
claim to have understood the nature of variation by placing it (variation) there
(in the 'lexicon').

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. In minimalism, all syntactic
operations are currently assumed to be feature-driven. That is to say, as Epstein
(2003 [2007], 43) has correctly pointed out, "the most fundamental opera-
tion ceases to be structure-building (Merge) and becomes structure-licensing
(Check/Agree/Value)." Epstein's statement makes it clear that contrary to the
rhetoric often used in minimalist circles, it is not true that "all you need
is Merge" (see Berwick (2011)). As Chomsky himself makes it clear in a
passage already quoted above, "[i]n addition to Merge applicable without
bounds, UG must at least provide atomic elements, lexical items LI, each a
structured array of properties (features) to which Merge and other operations
apply to form expressions" (Chomsky, 2007, 6). But saying that Merge merely
"applies" to LIs does not capture the fact that Merge is subordinated to the
lexicon. Short of features triggering it (the "vehicle requirement" imposed
on Merge in Pesetsky and Torrego (2007)), Merge can't apply. No wonder

[8] Consider the notion of "head" in Rizzi's statement. Why do we regard bundles of features as
(minimal) heads, and other collections of features as (maximal) phrases? Where is the dividing
line? (The problem is of course even more severe in a Bare Phrase Structure framework like
Chomsky (1995).)