

Clarifying the Notion “Parameter”

Juan Uriagereka

This article aims to reflect on linguistic architecture by re-examining language variation. Three progressively deeper forms of variation are suggested, each of which arguably contributes to this exercise in rather different ways. The discussion that unfolds will then lead to a new twist on the question of whether MP and OT are compatible.

Keywords: learnability; minimalism; optimality; parameter

1. Classes of Variation

I would like to reflect on linguistic variation, which I don't see as a unified phenomenon. In fact I will suggest that there are *three progressively deeper forms of variation*, and moreover that they are all important in understanding the architecture of the language faculty. This will lead me to asking, in the end, whether the Minimalist Program (MP) and Optimality Theory (OT) are compatible.¹

To start the discussion, consider the Polysynthesis Parameter (in the sense of Baker 1996), which gives different values for, say, English and Basque: While the latter case-marks verbal arguments and exhibits them as agreement elements in an auxiliary, the former does neither. Correspondingly, English disallows *pro*-drop (1a), presents a fixed word-order (2), and allows extractions from nominal objects (3a), whereas Basque can drop any argument (1b), allows any of the orders in (2), and disallows extractions from nominal objects (3b); “pied-piping” extractions as in (4) are fine in both languages:

- | | | | |
|-----|----|-----------------------|----------------|
| (1) | a. | * (I) love *(Lucy). | <i>English</i> |
| | b. | Maite nuen. | <i>Basque</i> |
| | | <i>love I.AUX.III</i> | |
| | | ‘I love him/her/it.’ | |

The final version of this talk was presented at WECOL in Fresno (*Western Conference on Linguistics*, October 2006). Aside from appreciating generous comments from the audience, I want to thank the organizers, and very especially Brian Agbayani, for offering me the opportunity to present these ideas in that environment.

¹ I will be using a couple of case studies which, I'm sure, could be interpreted differently. But I do this for illustration purposes, as I'm also sure that other examples could be used to raise the same points I will make.



- (2) a. Nik maite nuen Lucy.
lit. 'I love Lucy.' (OK in English)
- b. Lucy nik maite nuen.
lit. 'Lucy I love.' (OK in English only with topicalization)
- c. Nik Lucy maite nuen.
lit. 'I Lucy love.' (Permutations OK too, all * in English)
- (3) a. Who has John seen [pictures of *t*] ?
- b. *Noren ikusi ditu [*t* argazkiak] Jonek?
who.GEN see III.AUX.III pictures.ABS Jon.ERG
'Of whom has Jon seen pictures?'
- (4) a. [Pictures of whom] has John seen *t* ?
- b. [Noren argazkiak] *t* ikusi ditu Jonek?
who.GEN pictures.ABS see III.AUX.III Jon-ERG
'Pictures of whom has Jon seen?'

This is the expected situation in genuine parametric choices, which typically correspond to low-level morphological facts (case, agreement) and have vast consequences for the grammar at large (in terms of syntactic conditions).

To start considering markedness issues next, keep in mind that in situations whereby a set of structures in language *L* is a subset of a larger one in language *L'*, we assume that the language acquisition device (LAD) must hypothesize that it is acquiring that aspect of language *L*, unless presented with direct positive evidence for a structure in the superset. Had the LAD hypothesized, in the absence of such confirmation, that it was learning the relevant aspect of the language corresponding to the larger set, the only way it could retreat from a mistaken assumption is by way of analyzing negative data.

Now, which is the larger set of structures related to (and therefore, which can set) this particular parameter — the Basque or the English one? If we go with the evidence in (2), the Basque fragment is the superset (more grammatical combinations with the same words and identical grammatical relations are possible in this language); but if we go with the evidence in (3) and (4), the opposite is the case. So the LAD cannot decide which is the unmarked option for this particular language fragment. That is not problematic, so long as robust positive evidence exists for each option of the parameter, which of course is the case for this aspect of Basque and English. *In the presence of robust evidence for both settings, learning either is trivial.* As a consequence, there is no logical need to postulate an unmarked option. We may think of this as a *core* parameter.

Not all parametric situations are like that. Compare languages, like Spanish, which allow clitic-doubling, and languages like English that don't:

- (5) a. Juan la está viendo (a María). *Spanish*
Juan CL is seeing to María
'Juan is seeing María.'
- b. John is seeing'er (*Mary). *English*

Let’s call whatever is involved in this difference the Clitic Parameter, without attempting to argue for it or characterize it deeply. Inasmuch as doubling is optional (5a), the set of structures it sanctions is a super-set of the set of structures associated with absence of doubling (5b). One could argue that, just as (3b) is impossible in languages with object agreement, so is (6):

- (6) * De quién las está viendo [a amigas t] ?
 of whom CL is seeing to friends
 ‘Who is he/she seeing friends?’

However, that would be an unfair comparison. This is because (7a) is as good in Spanish — crucially, without a clitic — as its English version in (7b):

- (7) a. ? De quién está viendo [amigas t] ?
 of whom is seeing friends
 ‘Who is s/he seeing friends of?’
 b. ? Who is he/she seeing [friends of] ?

In a language with obligatory agreement, like Basque, one cannot build a version of (7a) *without* agreement. That is one of the differences between clitics and agreement markers: The latter are not dropped. Note the consequence of this state of affairs: A grammatical version of (6) exists in Spanish, so in this instance there arguably is *no structure that the English version of the Clitic Parameter allows which Spanish doesn’t* — and hence English is, in this particular data region, a genuine subset of Spanish. I would like to think of the relevant parameter ensuing here as a *sub-case* parameter, which presupposes a distinction between a marked and an unmarked value.²

Incidentally, the sub-case situation just described doesn’t entail that this portion of Spanish (the marked option) will take significantly longer for a child to learn. To see this, consider first the fact that sub-case conditions are in principle relevant only with a finite set of options — indeed, a small such set. A conservative learner may never find a crucial missing data piece if relevant sets are large. Unfortunately, that is a possibility for first-order syntactic data, which can be unbounded in principle. This entails, so far as I can see, that, *if sub-case situations are to be of any use in syntax, they must involve second-order data analysis* — phrasal type combinations as opposed to mere word token combinations;³

² I’m calling the options “sub-cases” instead of “sub-sets” to avoid E-language talk, an issue that is at right angles with my concerns here.

³ I mean this talk of orders of complexity in syntax in the customary sense these notions have in logic. Following work by Cornell & Rogers (2000), I will take an appropriate characterization of phrasal units of the customary sort (not just in MP, but also in other versions of the Principles-and-Parameters system, broadly characterized) to require not just operations over objects-in-the-lexicon, but moreover quantificational statements over functions of those — for instance, conditions involving contextually defined Case assignments — that cannot be coded as more complex predicates, no matter how artificially this is done. It should be obvious that I don’t mean any of this critically: Using higher-order devices for syntactic analysis has been very useful at least since the introduction of the notion “filter” in the late 1970s.

moreover, to be very small such sets must be limited to characteristic cycles in a manner I return to. Now, if we allow the child access to second-order grammatical descriptions of available data — which David Lightfoot calls “cues” (Lightfoot 1999) and Janet Fodor “triggers” (Fodor 1998) — then the evidence cueing *even the marked option* of the parameter ought to be readily available for a child to use. That highlights the difference between the current Principles-and-Parameters model (Chomsky 1995 *et seq.*) and a more traditional proposal (e.g., the *Aspects* model of Chomsky 1965) in which learners always compare grammars in terms of the first-order linguistic structures that they license (see fn. 3). In genuine parametric terms, a set comparison of the sort alluded to may well be relevant only in situations of a learning conflict, where lower-order evidence leads to ambiguous analyses (I return to this possibility). Still, the scenario outlined in the previous paragraph is important in principle, and arguably central in fact for situations of language change.

But there has to be more to linguistic differences than mere parametric settings, be they of the core or the sub-case sort. Chomsky is very explicit in 1981 about the role of idealization and how that relates to the notions we are considering. He says that:

[W]hat are called “languages” or “dialects” or even “idiolects” will [not conform — JU] to the systems determined by fixing the parameters of UG [...]. [E]ach actual “language” will incorporate a periphery of borrowings, historical residues, inventions, and so on [...]. (Chomsky 1981: 7-8)

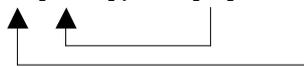
Nonetheless, Chomsky also emphasizes that “outside the domain of core grammar we do not expect to find chaos. Marked structures have to be learned on the basis of slender evidence too, so there should be further structure to the system.” While that is reasonable, it is difficult to pin down the nature of that further structure. Chomsky continues on the same page:

[W]e assume that the child approaches the task [of language acquisition — JU] equipped with UG and an associated theory of markedness that serves two functions: [I]t imposes a preference structure on the parameters of UG, and it permits the extension of core grammar to a marked periphery. (Chomsky 1981: 8)

The first of these functions is obvious, and has been discussed already; but the second one is less so, and clarifying it has interesting consequences.

2. A Case Study

In many languages argumental operators like *who* trigger, upon fronting an information question, an ancillary verb movement, involving an auxiliary (as in the English (8a)) or the main verb itself (as in the Spanish (8b)):

- (8) a. [Who [has [John [t [seen t]]]]] ?

 (cf. *Who John has seen t ?)

- b. [A quién [vio [Pedro [t [t t]]]]] ?
 to whom saw Pedro
-
- (cf. *A quién Pedro vio?)

Why the verb movement is required used to be a topic of much discussion. Not to get too technical about it, let's say that a sentence boundary (a CP) is (in most languages) a barrier, and the "barrierhood" of this category disappears if the CP is in construction with a lexical head:⁴

(9) *L(exical)-marking Convention*

XP is not a barrier if it is in construction with a lexical head Y, where X is in construction with Y if X = Y or X is selected by Y.

By the L-marking Convention (LC) in (9), an embedded CP does not require verbal inversion of this sort. Thus, compare (8a) to (10a):

- (10) a. [I wonder [CP who [John [has [seen t]]]]]
 b. ... [CP C⁰ [IP John [has [VP seen who]]]]

(10b) goes back in the derivation to the point prior to the displacement of *who*. Suppose all XPs along the way are potential barriers for this displacement. VP is in construction with a lexical head, its own; as a result, VP is not a barrier for the movement of *who*. Of course, by this sort of reasoning, all categories headed by a substantive head will not be barriers to displacement. If the I-, or T(ense)-, head of IP also counts as substantive, then IP will not be a barrier either. Now CP is headed by the abstract C⁰, not a substantive element. But is it in construction with a selecting element? It is, if the CP is selected by *wonder*. Then CP is not a barrier either, though not for intrinsic reasons (its own head), but by way of its contextual properties. This is the step missing in (8a), where nothing selects the relevant CP. But this CP may cease to be a barrier if it incorporates a neighboring lexical head, in which case the combined projection will be in construction with an appropriate L-marker. That's what head movement arguably achieves:

- (11) [CP has-C⁰ [IP John [t [VP seen who]]]]
-

Observe that the displacement of the verb in (11) does not cross the CP but adjoins instead to C⁰. In contrast, *who* must cross CP; however this category is no longer a barrier after verb movement, in the manner indicated. The reasoning is rounded up by the assumption that the mechanism is, in some relevant sense,

⁴ It is immaterial for my purposes here what the best analysis is of this phenomenon, and why it is not universal, although it certainly is very common. I expressed my own take on the matter both in earlier work (Uriagereka 1988, 1999).

costly, which is why the grammar does not undertake it if it is not necessary, in the usual minimalist fashion; so inversion in the circumstances in (10) is unacceptable.

The question then arises about structures involving adjunction to CP itself, which thus *should not cross this element to begin with*. As Rizzi (1990) indicated, this situation arises for causal modifiers, and therefore for a corresponding *why*. Questions involving one level of embedding should not trigger verb preposing; however, they certainly should if they involve two such levels. In other words, (12a) should be good, alongside with (12d), while both (12b) and (12c) should be bad; please hold your judgments:

- (12) ⓐ a. Why [_{CP} John has seen Mary] *t* ?
 ⓑ b. * Why has [_{CP} John seen Mary] *t* ?
 c. * Why [_{CP} you have thought [_{CP} John has seen Mary] *t*] ?
 d. Why have [_{CP} you thought [_{CP} John has seen Mary] *t*] ?

In (12a) *why* does not have to cross CP, thus moving *has* as in (12b) should be unnecessary to void the barrierhood of this CP. In contrast, although *why* in (12c) does not have to cross the embedded CP it modifies, it does have to move across the matrix CP in its displacement to the clausal periphery; hence this time ancillary verbal displacement to the C-head is justified. Standard speaker judgments for (12c) and (12d) accord with the theoretical prediction; however, those for (12a) and (12b) are backwards, as the unhappy faces indicate.⁵

So is the theory wrong? Possibly, of course, but there was something intuitively right about it, and it did seem to work for arguments as in (8)/(11); it is when extending our reasoning to adjuncts — correctly it would seem — that things start to fail. Intriguingly, Crain & Thornton (1998) report observations, which I first was told by Tom Roeper and Jill deVilliers in the mid-1980s, that English-speaking children (some into their late primary school years) provide judgments as in (12). Some dialects of Spanish, too, present almost the same paradigm:

- (13) a. Por qué [_{CP} Juan vio a María] *t* ?
why Juan saw to María
 ‘Why Juan saw María?’
 b. Por qué vio [_{CP} Juan a María] *t* ?
 ‘Why did Juan see María?’
 c. * Por qué [_{CP} tú pensaste que [_{CP} Juan vio a María] *t*] ?
why you thought that Juan saw to María
 ‘Why you thought that Juan saw María?’
 d. Por qué pensaste [_{CP} tú que [_{CP} Juan vio a María] *t*] ?
 ‘Why did you think that Juan saw María?’

⁵ Again, I present this particular paradigm here, with the nuances I’m about to report, solely for illustrative purposes. Other relevant examples come to mind, although they are not so easy to illustrate.

The parallel is not total, since both (13a), involving no verbal displacement, and (13b), involving it, are possible. Nonetheless, facts are similar enough for Crain & Thornton (1998) to make their point: Children acquiring a language L must be assuming a variant present in some other language L'. Supposing this is correct, a question remains: What is responsible for the English pattern in (12)? Or related to this question, why is the Spanish in (13b) — involving what looks like a costly and unnecessary option — also an option alongside the predicted (13a)? Actually, it is well-known that *vernacular versions of English present the pattern in (12) as well*, and upon closer examination, the Spanish in (13a) arguably belongs to a more relaxed register than in (13b). Is it possible that the verb preposing in (12b) or (13b) is a “peripheral invention,” somehow achieved on analogy with instances of verb preposing where it is needed in order to eliminate a barrier by way of the LC in (9)?⁶ That would explain why children continue to use the pattern predicted by the theory well beyond normal stages of acquisition, as do “uneducated” speakers.

If we allow for that kind of variation, it clearly will be neither of the core or the sub-case sort. The prestige adult pattern is, I believe, psychologically real (in the sense that one has intuitions about it), but its acquisition constitutes a genuine instance of training of some sort, and as such is different from whatever is involved in more elementary parameters. Thus core parameters recall growing, by fixating structure through elementary information, in much the same way, I would suggest, that epigenesis works in biology; and, of course, sub-case parameters involve the customary untrained learning, via unconscious analytical processes that allow the child to compare second-order chunks of grammars. In this instance, in contrast, a form not predicted by the core grammar is acquired under peripheral conditions, presumably involving such things as peer or adult pressure, and similar, as of now, unclear mechanisms.

As Chomsky asked nearly 30 years ago:

How do we delimit the domain of [...] marked periphery? [...] [E]vidence from language acquisition would be useful [, but is] insufficient [...]. We are therefore compelled to rely heavily on grammar-internal considerations and comparative evidence, that is, on the possibilities of constructing a reasonable theory of UG and considering its explanatory power in a variety of language types [...].
(Chomsky 1981: 9)

I have little to add to that: I'm just proposing that we take it seriously, assuming that micro-variations like the one I have examined point towards the existence of a *systematic Periphery*, of a sort that seems quite different from whatever is involved in the constitution of I-language. Aside from “cleaning the variation act,” I believe this may have rather intriguing architectural consequences.

3. Considerations about Syntactic Change

So far I have argued: (a) that there are three sorts of systemic variations (core, sub-case, and micro-variations) and also (b) that sub-case parameters must

⁶ I haven't seen this particular position taken in the literature, at least for this case.

involve data of a second-order sort (concretely, statements involving category types). This is already slightly different from current assumptions of the “Three Factors” sort, in Chomsky’s recent sense: genetic endowment, experience, and physico-computational laws. These invite the inference, explicit in Chomsky (2005), that variation is restricted to the second factor. In my view, in contrast, variation starts actually in the very first factor, the genetic endowment — and following Piattelli-Palmarini & Vercelli (in press) — I take this variation to be quite literally of an epigenetic sort. It is slightly misleading to think of it as fixed by experience, in any classical sense of the term “experience.” This parametric fixation is as structurally fateful and blind as whatever happens to a bee larva being fed on the crucial protein that royal jelly involves, thereby growing, structurally and behaviorally, into a queen-bee.⁷ Moreover, variation of the classically experiential sort comes, I am suggesting, in two varieties. There is, first of all, unconscious learning, geared by sub-case considerations and deploying second-order data analysis. But there has to be some room, also, for more or less conscious training, for lack of a better term.⁸ In what follows I will suggest that this is extremely restricted, in particular to first-order data analysis in what amounts to the left-periphery of parsed sentences. But I find no reason to doubt that this kind of apparently low-level phenomenon can have — with some probability at least — serious architectural consequences.

To make my case, I will conjecture that *variations we encounter correspond to levels of the Chomsky hierarchy of grammars* available to grammatical description.⁹ If sub-case parameters correspond to the sort of (phrasal) objects that enter context-free relations, formally simpler finite-state relations should be the locus of peripheral variation, while formally more complex context-sensitive relations should be involved in core parametric variation. I will not focus on the latter claim here, but it is surely true. Core parameters, starting with the Polysynthesis Parameter one discussed above, certainly involve (mild) context-sensitive processes of the Agree sort, together with all the nuances we associate to them (for instance, whether languages of one or the other type allow for hyper-raising, more or less local A-movement, possessor raising, and so on).

I will concentrate, instead, on the former claim, which is intuitively obvious though rarely emphasized: Analogies and similar sociological processes typically happen across low-level domains where adjacency (between “grammaticalized” forms) is typically presupposed by the theorist. Needless to say, adjacency is the finite-state notion *par excellence*.

⁷ This is not the place to defend epigenesis, but innumerable other examples can be found in the recent literature, ranging from body conditions to behavior, and involving “input data” as varied as proteins, temperature, or information. Of course, one can call all of that, by definition, “experience,” as it is not part of the genome. But that’s clearly not what is normally meant by the term, particularly when we are talking about information that manages to penetrate into cellular levels, instead of staying at the “surface” realm of cellular networks (like neuronal ones).

⁸ Of course, one could get technical and call the process Bayesian or some related notion, which I don’t object to but have little to add to here.

⁹ I will not defend this hierarchy here, and simply assume the defense mounted in Lasnik & Uriagereka (forthcoming), summarizing — and adapting to I-language purposes — much important work by the MIT, UPenn, and UCLA groups, whose results owe so much to Aravind Joshi’s important insights.

To be concrete, reconsider adjunct questions involving a verbal inversion. If this innovation is peripheral, speakers should fall into it *under finite-state conditions involving adjacency*. This is plausible: The moved Wh-element is in the specifier of CP, while the putative verb movement carries the verb to the C-head, thus to a position which, in phonetic form, ends up being right-adjacent to the specifier. This is the sort of window that a surfacey finite-state analysis can see through. The consequence is interesting. Although, as we saw, Universal Grammar (UG) would not require the displacement of the verb to C⁰ in relevant adjunct questions, *a costly movement is allowed in order to meet the string syntax of other structures involving (argument) wh-elements and adjacent verbs*. In other words, speakers are not aware of why they move a verb to the domain of adjacency of an argument *wh*-phrase; however, they can, in effect, be made aware, due to some form of more or less subtle societal pressure, that the relevant output has been achieved and that it ought to generalize. That conclusion is not demanded by UG — but it is consistent with it (disregarding a preference for absence of move-ment).

It is interesting to couple “peer-pressure under merely finite-state conditions” with familiar considerations emphasized by Lightfoot (1999) regarding Degree-0 learnability (data analysis by children focused on overwhelmingly more abundant main clauses). This, in the end, leaves very little room for sociological drift. Matters are even more constrained if considerations about phase-impenetrability (in the sense of Chomsky’s recent work) are involved: Degree-0 may then mean nothing but, in effect, the last CP phase, perhaps even its edge. If so, sociological drift would arguably be limited to the left periphery of main clauses, a very small window for variation. A limiting condition of this sort is necessary to ensure the usefulness (or psychological plausibility) of sub-case parameters, as already observed. If, as I suggest next, non-trivial variation must always start in terms of sociological drift, then the window for any kind of language change will be drastically reduced, as desired — or languages would change too often.

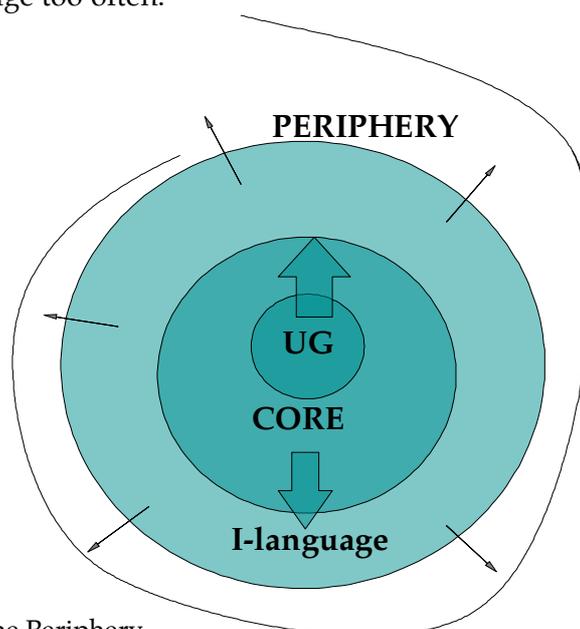


Figure 1: From I-language to the Periphery

The diagram in Figure I is intended to convey this fundamental difference (essentially, a comparison among growth, learning and training) between the combinatorial systems of language and those that are based on an ultimately sociological exchange.

This model of grammar allows for very little change. Suppose a sociological drift takes place in some peripheral construction (e.g., an augmentation of displacement processes in the left periphery, as in (13)). That — if stable enough to last — will constitute a *bona fide* instance of linguistic change (presenting the characteristic “S”-shaped curves of such smooth transitions). But such a curve may end up having architectural consequences with regards to the data. A learner doesn’t consciously discriminate between second-order or first-order data. However, we already saw that the first and second-order interpretation is quite different: In one instance sets of word tokens are relevant, whereas in the other, instead, what counts are grammatical structures. Similarly, core settings are interpreted by the system differently from peripheral trainings. The very same datum could be used in setting core parametric values (first of the cueing sort, next in terms of set evaluation), and eventually in whatever mechanics are involved in establishing peripheral patterns.¹⁰ This hierarchy is crucial in the dynamics for language change. An innocent-looking drift cannot be kept from affecting those aspects of the data that may tilt the otherwise stable part of the equation. For instance, a simple frequency change of the peripheral sort can, no matter how rarely, affect the second-order sub-case conditions for a learner to set the paradigmatic value(s) of a sub-case dimension like the Clitic Parameter. A change of the latter sort may in turn, if sporadically, imply the emergence of further sorts of evidence which, when analyzed at an even higher order (involving context-sensitivity), result in different conditions for a learner to set a core dimension like Polysynthesis.

So in this view the course of syntactic change involves a hierarchy:

(14) Peripheral change > Sub-case change > Core change

A disturbance emerges in the periphery of a language which manages to cascade through interesting consequences for the first-order analysis of the data available to a language learner, and eventually a higher-order analysis. This course of action is, of course, not necessary: A disturbance in the Periphery may simply stay there, and even if it manages to affect an I-language, it may never trickle down to Core parametric options. Nonetheless, *if* a radical language change is to take place, the logic examined dictates that it proceed in the manner assumed in (14) — unless it is directly imposed by outside forces (e.g., a foreign invasion, enslaving, acculturation, etc.). That logic does not allow us to expect a *direct* change to occur either in the I-language or in the Core. While it is possible for a generation of *adults*, following societal pressures, to change their first-order data, it is arguably impossible for them to engage in more elaborate manipulations. It

¹⁰ This is akin to saying that the very same protein can have very different roles in the development of an organism: it can serve to trigger a genetic regulation if it interacts early on, and at a cellular level, with the organism, or it may merely serve to contribute to metabolism, if the interaction is at a different level.

is equally impossible for children to change anything drastic altogether: They simply analyze data. In other words, while the cause for radical (internal) language change may be children, the trigger must be adults (they are the ones changing sociological patterns).

Needless to say, things could be more complicated. To repeat, there could be drastic data disturbances of an invasive sort, or it could be that all individuals essentially entertain multiple grammars at once, as explored by Anthony Kroch in the past and more recently by Charles Yang (e.g., Kroch 2000, Yang 2002). My point, however, is more modest: Even within the narrow conditions of *no external influence and purely monolingual speakers*, languages could significantly drift due to peripheral factors (which William Labov stressed in other domains; cf. Labov 1994); and more to the point of my concerns, with some probability such minuscule changes in the language frontier could have significant architectural consequences, rearranging sub-case and even core parameters.¹¹

4. Where Does This Leave Minimalism and Optimality?

MP is a “third factor” conjecture about the architecture under discussion — that it may have arisen as an optimal solution to interface demands, when biologically relating an internal system of thought to externalization mechanisms. OT is, in turn, a proposal about the plastic manifestation of the system — that externalized structures be conceived as landscapes emerging from re-ranking soft constraints; in effect, it is an enriched Markedness Theory. Due to their focus, each proposal has undertaken different tasks. For the most part, MP has worried about ascertaining the validity of the “economy” conception, exploring minimization processes to account for universal conditions. In contrast, OT has been applied to the description of phenomena in their subtle variants, to examine under what circumstances the observed diversity can be made to follow from the plastic architecture.

Whatever the ultimate answer is to why the linguistic computational system is of the mild context-sensitive sort (in Aravind Joshi’s sense), it pertains to something that OT takes for granted: the GEN function. Natural language doesn’t seem to be built on a random set-theoretic object: Its combinatorial possibilities have roughly the usual shape. That said, the formal object we are dealing with is characteristically unstable — something that must be part of the explanation — and OT is a theory precisely about that instability. However, we already saw *several types* of instability in the faculty of language, from core-variations to “micro-parametric” ones. In the latter realm it is not obvious to me what it buys us to speak of “parameters”: Nothing shifts in the system with

¹¹ I know of at least one careful study by Irene Moyna in Rio de la Plata Spanish that all but demonstrates a drift of the expected sort, taking place during the eighteenth and nineteenth century and involving, in the adult population, precisely the left periphery of clauses, as argued here (Moyna 2007). It would be interesting to study whether that peripheral change has had an influence on deeper parametric options of this dialect, known to exist in the present generation.

consequences for the system at large with each variation.¹² OT, in contrast, provides a way to sieve through these micro-variants, particularly if we think of it as enriching the Theory of Markedness that sub-set parameters demand. As we saw, the size of the variation doesn't make it any less important to the system, even systemically so. The architecture requires both macro and micro-variation: Without each we either cannot get transitions in the system to take place, or the necessary drift to trigger them.

Such state of affairs is not even surprising, if the language faculty exists on an internal domain (leading to LF) and an external one (leading to PF). We don't expect genuine internal variation, for it would be virtually impossible for infants to acquire it. What crucial information would set it? But by the very same reasoning, variation in the external domain is expected, indeed even natural if the system, like much else in basic biology, doesn't specify its full structural details. The only issue is what the nature of that variation ultimately is, and how connected it is to the internal conditions. If implied at the core, we should see massive consequences for the emergent system, not just isolated surface manifestations. Otherwise, we should witness massive variation, but not swinging in tandem with anything else. A rich theory of Markedness then, is in fact a necessary development.

But remember, the field has used the Core/Periphery distinction, and a corresponding theory of Markedness, with systematic ambiguity: Either as a way of distinguishing pre-set values in a parameter from those requiring evidence, or as a way of separating merely individual (i.e. psychological) behaviors from also historical (i.e. sociological) ones. I see no reason to reject either interpretation of the distinction, and it is then an interesting question which of those two dimensions of the problem OT is addressing. Needless to say, OT could in the end be the wrong sort of Markedness theory — just as MP could be wrong-headed. To decide on this, part of what we need to figure out is which theory models which sort of variation best. Interestingly, inasmuch as both approaches systematically seek optimizations, the fates of the programs would seem to be abstractly linked. Indeed it would be surprising if MP-style optimization is entirely wrong while the one in OT is perfectly right, or vice-versa.

I don't want to end without a final reflection on what all of this could mean, naturally, if remotely on track. The issue boils down to how seriously we want to take the idea that language is central to human existence in the full sense, involving creatures that are — well, alive, animals, and societal. As merely alive, we have to obey the properties of our genetic code, but now we know that much of that depends on early developmental factors that are plainly not genetically encoded. This is not news any more, and the only issue is whether language

¹² I find this virtually a defining characteristic of what a parameter is, at least in systems outside linguistics. In other words, if a variation is entirely restricted to a domain (say clitics or even third-person clitics, etc.), then it simply doesn't seem like a core phenomenon. In contrast, classical parameters were meant as much more abstract and fateful. For example, the sorts of apparently unrelated correlations that Snyder (2007) studied in for various VP nuances, which sharply divide languages in terms of whether they present resultatives, "X-way" idioms, systematically ambiguous interpretations (dynamic and static) for verbs of motion, and so on. The parameter cannot even be stated in a simple feature like "clitic" or "person."

participates on it or not. If it does non-trivially, some parametric options will have to be fixed at cellular levels.

Second, again by the sheer logic of being animals involved in complex acquired behaviors (i.e. “intelligent”), humans must participate in the sorts of learning nuances that, say, (some) song-birds do. In that case it is not even controversial that relevant acquisition circuits require delicately balanced input data, whose structure may well present more or less marked characteristics — this being the locus of much investigation, for instance, in Stephanie White’s lab (<http://www.physci.ucla.edu/research/white>). I personally don’t find it all that surprising that the brain circuits responsible for the acquisition in the bird case appear to be partly regulated by the one gene that research has told us is very probably implicated in language: FOXP2 — and that they correspond to entirely analogous circuits in the human brains, where FOXP2 is patently present as well (Jarvis 2006), perhaps for the sorts of reasons that Michael Ullman has conjectured, involving procedural memory (see, e.g., Ullman & Pierpont 2005). Moreover, if any of this is right, it wouldn’t be surprising if, again following Ullman (and colleagues)’s research, in both the birds and the humans there is critical regulation of the relevant circuitry that involves hormonal controls, and hence obeys characteristic criticality.

But by the same sort of reasoning, if we have learned so much from biology and neuro-psychology in understanding what language is, why can we not, also, learn something from sociology? It is hard to argue with some of the results our colleagues are showing us that “language” (in whatever broad sense one cares to characterize it) is a “dynamical” entity. The issue of course is old, and in modern linguistics was wrestled with explicitly by Saussure, who put it to the side for purely practical reasons: It was too tough, until now, to make full sense of the diachrony of language, until we had a better understanding of its synchronic properties. Surely that prejudice paid off, but now that we *have* a decent understanding of what’s going on — and that we can no longer “blame” biology for complete stativity! — it may be time, again, to reconsider the virtues of dynamism. But with utmost care.

That last aside is important in two respects. One has already been mentioned: Almost certainly matters are more complex than I have implied. Certainly all humans are, in some sense or another, multi-lingual, and in short we have no clue, really, what that means, and how or where (or when...) various grammars are represented (if that’s the word) in the human brain. Lacking a good understanding of all of that, it’s really very hard to tease apart the (mono-lingual) possibility I presented here from the role played by multi-linguism — yet another reason to keep multilinguism alive.

Second, every time I give this talk I typically encounter a somewhat parochial reaction. Everyone wants their particular turf to be the most important, hence dismissing all other forms of structuring as trivial. I find this a bit obtuse. One of the lessons of contemporary science, it seems to me, has been how “*raffiniert*” *die Frau Mutter Natur* turns out to be, to adapt Einstein’s quip.

Thus to force a fully OT analysis on all of this, or a complete minimalist take on it — aside from boring at conference parties — would be disappointing, indeed a concession that language is not as *natural* as we think it is.

References

- Baker, Mark. 1996. *The Polysynthesis Parameter* (Oxford Studies in Comparative Syntax). New York: Oxford University Press.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1981. *Lectures on Government and Binding: The Pisa Lectures*. Dordrecht: Foris.
- Chomsky, Noam. 1995. *The Minimalist Program* (Current Studies in Linguistics 28). Cambridge, MA: MIT Press.
- Cornell, Thomas L. & James Rogers. 1999. Model theoretic syntax. In Lisa L.-S. Cheng & Rint Sybesma (eds.), *The First Glot International State-of-the-Article Book* (Studies in Generative Grammar 48), 171-198. Berlin: Mouton de Gruyter.
- Crain, Stephen & Rosalind Thornton. 1998. *Investigations in Universal Grammar: A Guide to Experiments on the Acquisition of Syntax and Semantics* (Language, Speech, and Communication). Cambridge, MA: MIT Press.
- Fodor, Janet Dean. 1998. Unambiguous triggers. *Linguistic Inquiry* 29(1), 1-36.
- Jarvis, Erich D. 2006. Evolution of brain structures for vocal learning in birds: A synopsis. *Acta Zoologica Sinica* 52(Supplement), 85-89.
- Joshi, Aravind K. 1985. How much context-sensitivity is necessary for characterizing structural descriptions — Tree Adjoining Grammars. In David Dowty, Lauri Karttunen & Arnold M. Zwicky (eds.), *Natural Language Processing: Theoretical, Computational, and Psychological Perspectives* (Studies in Natural Language Processing), 206-250. Cambridge: Cambridge University Press.
- Kroch, Anthony. 2000. Syntactic change. In Mark Baltin & Chris Collins (eds.), *The Handbook of Contemporary Syntactic Theory* (Blackwell Handbooks in Linguistics), 629-739. Malden, MA: Blackwell.
- Labov, William. 1994. *Principles of Linguistic Change*, vol. II: *Social Factors* (Language in Society). Oxford: Blackwell.
- Lasnik, Howard & Juan Uriagereka. Forthcoming. Structure. To appear in Martin Stokhof & Jeroen Groenendijk (eds.), *Handbook of the Philosophy of Linguistics* (Handbook of Philosophy of Sciences). Amsterdam: Elsevier.
- Lightfoot, David. 1999. *The Development of Language: Acquisition, Change, and Evolution* (Blackwell-Maryland Distinguished Lecture Series on Language and Cognition 1). Malden, MA: Blackwell.
- Moyna, Maria Irene. 2007. *Voseo/Tuteo* replacement in Rio de la Plata Spanish: Saliency and learnability effects on paradigms. Paper presented at the *18th International Conference on Historical Linguistics (ICHL 18)*, Montreal. [Université de Québec à Montréal, 6-11 August 2007.]
- Piattelli-Palmarini, Massimo & Donata Vercelli. In press. Language in an epigenetic framework. In Massimo Piattelli-Palmarini, Juan Uriagereka & Pello Salaburu (eds.), *Of Minds and Language: The Basque Country Encounter with Noam Chomsky*. New York: Oxford University Press.
- Rizzi, Luigi. 1990. *Relativized Minimality* (Linguistic Inquiry Monograph 16). Cambridge, MA: MIT Press.
- Snyder, William. 2007. *Child Language: The Parametric Approach*. New York:

- Oxford University Press.
- Ullman, Michael T. & Elizabeth I. Pierpont. 2005. Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex* 41(3), 399-433.
- Uriagereka, Juan. 1988. On government. Storrs, CT: University of Connecticut dissertation.
- Uriagereka, Juan. 1999. Minimal restrictions on Basque movements. *Natural Language & Linguistic Theory* 17(2), 403-444.
- Uriagereka, Juan. In press. *Syntactic Anchors: On Semantic Structuring* (Cambridge Studies in Linguistics 118). Cambridge: Cambridge University Press.
- Yang, Charles. 2002. *Knowledge and Learning in Natural Language*. New York: Oxford University Press.

Juan Uriagereka
University of Maryland
Department of Linguistics
1401 Marie Mount Hall
College Park, MD 20742
USA
juan@umd.edu