<u>Language, Chaos and Entropy</u>

Diego Gabriel Krivochen, University of Potsdam

Institut für Slavistik

Abstract:

In this paper we will try to provide a formalization of some characteristics of language as a chaotic (physical) system aiming at *justificative adequacy*: what architectural properties license the occurrence of certain phenomena. We will claim that the derivational dynamics that can be found in language (and other systems of the mind) should be analyzed from the perspective of chaos theory, as a system which is hypersensitive to initial conditions: a small change in a certain state of the system may have drastic effects on the output, focusing on the consequences this approach would have for linguistic derivations and ontogenesis.

Keywords: Chaos; Entropy; Recursion; Physics; Language

1. Some Introductory Considerations

To begin with, we will introduce the basic framework within which we will conduct our inquiry. Radical Minimalism is a program which analyses language as a physical system. Since language is part of the natural world in a non-trivial way (that is, it not only occurs in the Universe, but is a system that can be assimilated to others in that Universe), as Chomsky and mainstream Generative Grammar have claimed over the years (see Chomsky, 1965, 1986, 2005), it is assumed to be ruled by the same principles. The program strives to answer questions concerning the integration of what we refer to as *language* in a system of interacting cognitive capacities and furthermore advocates that language as a *physical system* should therefore not be studied in isolation, but rather in interaction with other systems. As this interaction occurs in the so-called 'natural world', it is constrained by physical laws which are in turn particular instantiations of mathematical possibilities. Considering this scenario, Radical Minimalism proposes the following tenets:

(1) a. Language is part of the 'natural world; therefore, it is fundamentally a physical system.

b. As a consequence of (a), it shares the basic properties of physical systems and the same principles can be applied, the only difference being the properties of the elements that are manipulated in the relevant system[1].

c. The operations are taken to be very basic, simple and universal, as well as the constraints upon them, which are determined only by the interaction with other systems.

d. Points (b) and (c) can be summarized as follows:

Strong Radically Minimalist Thesis (SRMT):

[1] Such a pretention of universality regarding physical explanation is somehow reinforced by the advances in M-Theory and, more humbly, on string theory and field theory (Greene, 1999, among others). At this point, however, it is a desideratum, but one that, if pursued seriously, could lead to significative improvements in our understanding of the relations between physical systems. A reviewer pointed out that the space-time continuum and particle physics refute our thesis, since the former is a continuum whereas the latter is discrete: notice that the characteristics of the elements are irrelevant, since our focus is put on the operation that generates complexity in the physical world. In any case, a quantum perspective would in turn refute the reviewer's objection: particles are not discrete, but a possible state of a quantum field, or, in string-theoretical terms, a certain pattern of vibration. There is nothing "discrete" in particles, if seen from these perspectives.

> All differences between physical systems are 'superficial' and rely only on the characteristics of their basic units [i.e., the elements that are manipulated], which require minimal adjustments in the formulation of operations and constraints [that is, only notational issues]. At a principled level, ***all physical systems are identical***, make use of the same operations and respond to the same principles.

SRMT licenses the possibility of looking for biological, physical and mathematical properties of mental computations (i.e., syntax), *without its being a metaphor* but a true account in three levels: description, explanation and justification. This claim is essential for our argument: we are not saying that language "is like" a physical system and, therefore, a mathematical structure. We are saying that language *is* a physical system. As it is, SRMT is nothing that can be proved right or wrong (it is really a methodological premise one can choose to follow or not), and even if it could, it would be only trivially. However, interesting empirical predictions derive from SRMT, and those can be observationally contrasted (we are consciously adopting a form of Carnap's 1966 model here) in order to determine the descriptive, explanatory and justificative power of the model. The description is 'the what', the explanation is 'the how', and finally, the justification is 'the why'. The latter has been either taken for granted or done in a truly non-minimalist way both substantively and methodologically (e.g., triggering operations by means of *ad hoc* features, e.g., Matushansky, 2006 and feature cyclicity stipulations). Our effort, then, will focus on trying to set a radically minimalist alternative of justification, taking into account that a theory of the physical universe must address all three: in this methodological and substantive integration of the study of language within the more general study of physical objects lies the main difference between Radical Minimalism and other approaches. Attempting justification is what we understand as the ultimate goal of going 'beyond explanatory adequacy'. This has a direct consequence for the research program we are proposing: it is simply impossible to attempt to explain a problem from just one perspective, but we need to understand the multiple facets of a phenomenon: its computational dimension, the biological properties of the system that allows such a computation (i.e., *genotypic-phenotypic dynamics*) and the physical principles that license this biological configuration, expressed as a mathematical structure. From a methodological point of view, then, our work has consequences for the foundations of the "biolinguistic enterprise" (Chomsky, 2007; Boeckx, 2010; Di Sciullo & Boeckx, 2011) and its place within linguistic inquiry.

1.1 *How do we derive?*

In this paper we will focus on one possibility for a mathematical formalization of the biological dimension of language, provided that the physical principles that license a biological configuration are, in turn, particular instantiations of mathematical structures: we will pursue the claim that the accurate mathematical model for this formalization is *chaos theory*. It must be noticed that the use of mathematical frameworks to formalize properties of language is not a new idea, see for example Di Sciullo & Isac (2008), who work within a *set-theoretical* framework in which Merge is driven by *proper inclusion* between feature bundles of lexical items, that is, the set of intensional properties of item A must properly contain that of item B for them to Merge and thus build a more complex structure; assuming, as orthodox Minimalism does, that a lexical item is nothing but a bundle of such features -like Case, Person, Number, Tense, etc.- some of which enter the derivational workspace valued (and are, therefore, legible) and other must be valued throughout the derivation (see Chomsky, 1995 et. seq. for details). However, we will pursue a different venue, since the set-theoretical approach to derivations Di Sciullo & Isac propose (also see Panagiotidis, 2010 for a different yet related proposal) require a rich feature system and an *Agree* operation to ensure the generation of fully legible units for the systems of sound and meaning which interface with the generative engine in language derivation. For example, Di Sciullo & Isac's approach require a subset relation between lexical items to enter into a Merge relation, which requires a highly specified feature matrix for each lexical item (including categorial features, person, number, and other formal features like [Wh-] for interrogative elements and so on) and also principles determining interpretability and valuation conditions: if a feature is to be defined as a valued dimension (e.g., + velar, in phonology, means "positive value for the dimension 'velar'"). The theory grows in complexity without improving descriptive or explanatory adequacy, since free-Merge systems (like Boeckx's 2010b, or Chomsky's 2004) work just as fine, we will attempt to demonstrate. We will also work with a *Free Generator* (i.e., Free blind Merge) and no features (this is, no [*value*-F] primitives and no Agree as the driving force of the syntactic component), all constraints being determined by third-factor interface constraints, which have to be formulated for each interface (a general approach will be provided by our *Dynamic Full Interpretation* and *Analyze*). The general architecture follows an Optimality Theory-like model, a GEN(erator) – EVAL(uator) dynamics. That said, our GEN algorithm in charge of structural complexity in language is an *n*-ary *Concatenate* operation, formulated as follows:

1) *Concatenation* defines a *chain* of coordinates in *n*-dimensional generative workspaces W of the form $\{(x, y, z\ldots n) \subset W_X \ldots (x, y, z\ldots n) \subset W_Y\ldots(x, y, z\ldots n) \subset W_n\}$ where $W_Y \equiv W_X \equiv W_n$ or $W_Y \neq W_X \neq W_n$.

This generator dispenses with featural requirements for Merge, such as Pesetsky & Torrego's (2007) *Vehicle Requirement on Merge* or Di Sciullo & Isac's proper subset requirement. Free Merge has already been argued for in, for example, Chomsky (1995, 2004), but in a different venue. Moreover, our derivational model is closer to that proposed by Putnam (2010) insofar as it is only interface-constrained:

as we will see below, we propose an extremely local evaluation mechanism (Optimality Theory's EVAL) *Analyze*$_{IL}$ (that is, *Analyze* from the interface level IL) that applies after each instance of *Concatenate* and determines whether the object is fully legible by the relevant interface; in the case of language, Conceptual-Intentional for meaning and Sensory-Motor for sound. Echoing Boeckx (2007, 2010), *Analyze* can be seen as the interpretative systems (whichever they turn out to be) accessing the syntactic workspace in which Concatenation applies, and taking the minimal unit they can read. Merge, as we take it, is a completely free operation that can apply as long as the objects to which it applies have the same format, motivated by interface conditions (this is, {α} is trivial at the interface levels, while {α, β} is not[2]). In the Faculty of Language FL, we have lexical items[3], and we can say that they have the same format (be them "lexical categories" or "functional categories") since they share a nature, they are linguistic instantiations of elements that, *per se*, are not manipulable by $C_{(HL)}$ like generic concepts (used by the visual capacity, for example). The only attribute of Merge would be putting things (lexical items, atoms, sounds, concepts…see Jackendoff, 2002; Talmy, 2000 for examples of conceptual structures that could be generated via Merge) together, without any restriction by principle as regards the nature or number of objects, since it would be a stipulation. An example of the derivational dynamics we are proposing is the following:

2) Concatenate (α, β) = {α, β}
   Analyze$_{IL}$ {α, β} [is {α, β} fully interpretable by IL?]
   (Transfer {α, β]} to IL if Analyze$_{IL}$ results in convergence at IL)

The idea of "invasive interfaces" is a natural result of separating interpretative and generative systems. If generation is restricted to a single operation *concatenate*, which is the optimal scenario (since there are no *a priori* Agreement restrictions, Cf. Pesetsky & Torrego 2004) and it occurs in an *n*-dimensional

[2] Boban Arsenijevic (p.c) claims that "*{{a}} is non-trivial in at least one faculty: the arithmetic capacity. Hence, output conditions can't be that bare to favor a binary merge*". However, our position is that if Merge is considered to be an operation and we assume also a *dynamic* version of Full Interpretation that states that *any derivational step must be interface-justified*, that is, the application of any operation must lead to a legible object the application of Merge to a single object is trivial in any faculty, as it does not contribute in any way to legibility. If {{a}} is legible for the interface system the arithmetic faculty has to interact with, then why apply Merge in the first place, to apply Merge to a single object is trivial in any faculty. If {a} is already legible in the relevant interface level, then why apply Merge in the first place? It would be computationally redundant, and therefore far from Minimalist. We maintain that binary Merge is the minimal-maximal non-trivial option. We therefore reject any proposal of *unitary Merge* on interface grounds.

[3] As a matter of fact, we have *roots* semantically defective and procedural features that make them manipulable by the computational system, but we will use the term "lexical items" for the time being. See *infra*.

workspace (another null hypothesis: restricting dimensions to 2, as in traditional Kynean tree diagrams, is stipulative, derived from considerations of “unambiguous paths” and linearization issues, see Kayne, 1984, 1994), it is only natural that the operation cannot *read* or evaluate what it has built. On the other hand, it is also only natural that the EVAL function (in OT terms, *Analyze* in ours) is not separated from the interfaces but in itself be the set of Bare Output Conditions (or legibility conditions) each cognitive interface has. Then, if we depart from considerations of computational simplicity like “maintain as few structure at once in W as possible”, that is, “transfer as soon as you can” (bearing an obvious resemblance to Pesetsky’s 1995 *Earliness Principle*), this “as soon as you can” is determined not by internal syntactic conditions (as they do not exist in our proposal), but by *Analyze*. In this way, the generative workspace is wiped clean several times along a derivation (without this being pre-determined by the presence of a certain element –a phase head-) thus liberating working memory without the concomitant problems of defining, for example, endocentric transfer domains (i.e., chomskyan *phases*).

A question that may arise in this point is whether *binarity* in the symbolic objects generated by Merge is an interface requirement or it is derivable from interface conditions imposed by interpretative systems (for an extensive discussion, see Krivochen, 2011a). Our answer is ‘yes and no’. Binarity is the *simplest-non-trivial* combination of elements, and syntax (in the broad sense) is fundamentally economical: there is simply no point in applying Merge to {α}, thus generating {{α}}, if the latter object is in no way “more legible” than the former. Merge is thus driven by what we call “Dynamic Full Interpretation”, a strict *interpretative interface* condition:

3) ***Dynamic (Full) Interpretation*** (DFI): *any derivational step is justified only insofar as it increases the informational load and/or it generates an interpretable object.*

By “informational load” we simply mean the amount of data that is read from a structure.
Having a set S of units S = {α, β, γ…}, free Merge (or *Merge-α*) allows unbounded *n*-ary combination, binarity being required for interface purposes, among them, *label recognition* (as described in Gallego, 2010): sticking to a strict version of the Inclusiveness Condition, we reject labels *in the syntax* because, if the syntactic component is purely generative (that is, strictly blind to the characteristics of the objects it manipulates), they fulfill no role. At the Conceptual Intentional (C-I) interface the situation is very different, however. “Labels”, symbolic markers which indicate how a structural unit is to be interpreted for the purpose of further computation, are read off a structure, following Hale & Keyser’s 1997 claim that “if we know the meaning, we know the structure, perforce, because we know the meaning *from* the structure”. The label of a syntactic object, that is, the instructions as to how that object is to be

manipulated for the purposes of further computations after Transfer, is then *recognized* from a construal (crucially, not *created*, since it would violate the Inclusiveness Condition, a *faithfulness constraint* in OT terms) at the semantic interface, following considerations of legibility and convergence. However, if there is no *interpretation*, DFI does not apply (for example, when deriving mathematical structures, which have no meaning and therefore no halting algorithm, as Tegmark, 2007 points out), which is already a strong and interesting generalization about the functioning of the mind: free manipulation of elements in *n*-dimensional workspaces[4]

1.2 *What do we derive with?*

We will now explicit the assumptions we make regarding the objects Merge manipulates in natural language derivations. We assume only two types of elements in linguistic derivations (following Wilson & Sperber, 2003 and Escandell & Leonetti, 2000):

a) <u>Conceptual elements</u>: i.e., roots, semantically underspecified, malleable and generic meaning. *Roots* are pre-categorial linguistic instantiations of a-categorial generic concepts. Generic concepts are "severely underspecified", since they are used by many faculties, and therefore cannot have any property readable by only some of them. Roots convey generic conceptual instructions, and their potential extension is maximal (expressible by the superset that properly contains all referential sets), given their semantic underspecification: bare roots have no (spatio-temporal) anchor.

b) <u>Procedural elements</u>: provide instructions as to how to interpret a root (anchor it to a specific point in Space or Time) or a relation between roots -including complex structures-. Procedural elements, for expository purposes, can be identified with Functional Categories, but there is a fundamental difference: the procedural character of a node is of no relevance to syntax, as Merge-α, we have said, is *blind*. "Procedurality" is thus recognized and relevant *at the semantic interface*, not before. Following and expanding on Escandell & Lonetti, 2000, we will identify D, T, *v*, C, P as procedural elements. The instructions conveyed by procedural elements play two main roles:

- Restrict reference in terms of a (finite) proper subset of the root. Each element restricts the set of the root's intensional characteristics in different ways. For example:

[4] See D'Espósito (2007) and particularly Baddeley (2003a, b) for an overview of the importance of the notion of workspace (or 'working memory') for the study of the biology of language.

$\surd = \{\alpha, \beta, \gamma, \delta \ldots n\}$[5]

$(X, \surd) = \{\alpha, \delta\}$

$(Y, \surd) = \{\beta, \gamma\}$

- Provide instructions as to:
  - Where to retrieve information? (that is, considering the mind to be massively modular, in which module should the parser look for the relevant information?)
  - What kind of information to retrieve? (that is, eventive, sortal, causative, locative…?)

Therefore, *procedural elements convey locative meaning* in the sense that they relate a *figure* (i.e., the root) to a *ground* (a set of intensional properties), and they are thus predicators (i.e, functors).
Formally, we have the following situation in the case of a lexical item LI:

4) A lexical item LI is a structure $\{X \ldots \alpha \ldots \surd\} \in W$, where X is a procedural category (Determiner, Time, Preposition) specified enough as regards distribution, α is an *n* number of non-intervenient nodes for category recognition purposes at the semantic interface, √ is a root and W an *n*-dimensional workspace.

If the root within the nominal domain is an element to be read off at the semantic interface as N, the semantics of α is predication.
Now, we will not assume a monolithic Lexicon, as in orthodox versions of the Minimalist Program, but work with the concept of *type-Array*. In such an Array, there are no numeric subindexes, nor phonological features or formal features: there are just *types* to be instantiated as *tokens* in a derivation according to interface requirements. There is no *a priori* limit to the number of times a type can be instantiated as a token, but those established by C-I interface conditions, particularly Dynamic Full Interpretation.
Let us assume we have an *Array* consisting on root and a procedural node D. Assuming that the label of {α, β} must be ***either α or β*** (which seems to be the simplest option, as it does not include a new element in the derivation) for the purposes of further computations and interface recognition, the derivation could go either of the following ways (5) or (6):

[5] Bare roots have maximal extension, in the following sense: the bare root √APPLE denotes all the apples in the Universe that have existed, that exist, that will exist and also those that exist in non-factual contexts. In this sense, procedural elements restricting such maximal extension become indispensable for manipulation when deriving semantic representations (i.e., Logical Forms). It is because of this maximal extension, also, that bare roots cannot be manipulated by C-I, which affects the label recognition algorithm.

5) *Narrow Syntax:* Merge (√, D) = {√, D}
6) *C-I:* Label {√, D} = {√, {√, D}}

Or

7) *C-I*: Label {√, D} = {D, {√, D}}

(6) seems *a prima facie* to collapse at the semantic interface, but, is it possible to derive this from interface conditions without appealing to, for example, Distributed Morphology's *Categorization Assumption* (see Embick & Noyer, 2004)? In our model, in the line of Panagiotidis (2009) and Fábregas (2005), roots are way too semantically underspecified to undergo *referent assignment* (see Escandell & Leonetti 2000 for a Relevance-Theoretic perspective on the D-N relations), and thus an explicature (i.e., a full propositional form) cannot be built. On the other hand, if we let D be the "label" at the semantic interpretative component, the whole structure is interpreted as a *specified entity*, because of the rigidity of D's procedural instructions: conceptual content can be narrowed or widened[6], but procedural content cannot. "Ill-formations", therefore, are *interface-determined*; NS has nothing to do with them. Let us consider a more extreme case:

8) *Narrow Syntax*: Merge (D, T) = {D, T}
9) *C-I*: Label {D, T} = {D, {D, T}} / {T, {D, T}}

Both labeling alternatives collapse, as it is to be expected. There is no way of building an explicature out of that structure, no matter how C-I tries to interpret it, assuming that every input conveys the assumption of its own Optimal Relevance: there is no substance to delimit, no maximal extension to restrict. *Optimal Relevance cannot possibly be achieved*, in other words. It is obvious as well that there is nothing wrong with {D, T} in the NS, as Merge is blind, free and unbounded[7]. Any restrictions are interface-imposed, third-factor legibility principles. This is a way of giving principled status to Grimshaw's (1991, 2000) claim (traceable to Abney's influential PhD thesis) that *extended projections* always have a *lexical head*,

[6] For example, √BANK can be widened to include ATMs as well as financial institutions on the one hand, and narrowed to specify a small-scale bank instead of the World Bank.

[7] This is why there is no point in positing instructions that "*apply at the level of syntactic computation*" (Leonetti & Escandell Vidal, 2011: 3; see also Chomsky, 1995): syntax (i.e., Merge) is purely generative, not interpretative. Any attempt to codify instructions for the syntactic component would lead to a constructivist system of the kind Frampton & Gutmann (2002) and Lasnik, Uriagereka & Boeckx (2005) argue in favor of.

bearing in mind, as we do, that *headedness is an epiphenomenon*, of no relevance to syntax (see Chomsky's discussion with Cedric Boeckx, 2009).
Furthermore, let us analyze the following cases:

10) *NS* Merge (α, β, γ) = {α, β, γ}

In the Narrow Syntax, everything would be fine, since Merge is blind and NS is ***not an interpretative component***, but just a dynamic workspace in which computations are performed. But, at the interface, problems would arise. Let us assume that α = √ and β and γ are procedural categories, say, D and T respectively.

11) *C-I* Label {√, D, T} = ??

Having two procedural categories results in crash at the explicature level, there is no way of labeling a structure where two elements could "guide" the interpretation in different directions. The same happens if the numbers are changed, say, two roots and one procedural category: even if we think that one of the roots may be "categorized" under the scope of the procedural element, there would still be an uninterpretable element, namely, an uncategorized root, uninterpretable in LF because of severely semantic underspecification[8]. Binary-distinct Merge, then, is interface-required, no special conditions are imposed over Merge itself. In any of the cases, it must be said, the application of Merge involving Ø (e.g., {α, Ø}, see De Belder & van Craenenbroeck, 2011) equals Self-Merge for interface purposes, thus being rejected in our proposal as violations of DFI.

Let us make some further clarifications on our notion of *Type-Array* and *Token-Merge*:

12) A *type* is an abstract element in a physical system Φ.
13) A *token* is an occurrence of a *type* within a workspace W_X. There are no *a priori* limits to the times a *type* can be instantiated as a *token* but those required by Interface Conditions IC.

The number of tokens required is determined by interface conditions, so that the *minimal number* of tokens leading to a convergent (i.e., fully interpretable) object is used, provided that the notion of

[8] Maximal extension entails semantic underspecification in the sense that the root is too vague to be manipulated by the semantic component: no single entity can be identified (either sortal or eventive).

"convergent object" does not arise from look ahead (the syntactic component looking at the legibility conditions of the interface systems, which would lead us to a "bad" crash-proof system), but rather from the interfaces "peering into" (i.e., "invading") the syntactic workspace (something similar to the proposal by Boeckx, 2007, although the consequences he draws are completely different from ours) and *Analyzing* whether a syntactic object is ready to be transferred. Boeckx (2010) points out, and we agree, that feature bundles cannot be driving syntax, since these bundles are structures and that structure is (must be) syntactic in very much the road taken by Distributed Morphology (Halle & Marantz, 1993, Embick & Noyer, 2004, Panagiotidis, 2010), where it is considered that roots and "f-morphemes" (to use the old DM term) are combined syntactically, with the same constraints that apply to an "s-syntactic" representation. The elements the syntax manipulates should optimally be *atomic*, and Merge should be taken, in his view, as a free-triggered unbounded operation, *Merge α*. The whole argumentation of Boeckx's aims at "defeating lexicocentrism", that is, the presence of a pre-syntactic instance where fully-fledged lexical items are taken from. Contrarily to Chomsky's (2005) and Hauser, Chomsky & Fitch's (2002) claims, the "great leap forward" (that is, the qualitative evolutionary difference between humans and non-humans) would be *the emergence of conceptual addresses, not Merge itself*, since it is really a recombination of pre-existing processes, Concatenate and Label (see also Hornstein & Pietroski, 2009 for a similar view). Our perspective, however, turns out to be simpler in that we have formalized Concatenate as the only generative operation to apply freely in workspaces and characterizing Label as a result of reading the structure that is transferred to the interpretative interfaces[9].

We have so far sketched the formal framework that will lead our inquiry, whose main claim is that language is an object with no particular idiosyncrasy within the natural world: it is, as any other portion of the Universe, a physical system in the technical sense. Since we consider language to be a physical system, we will start by characterizing it using the tools chaos theory offers us, as this framework has proven very useful for the analysis of complex systems and their development. The fact that natural language is in fact such a system is for the time something we take on faith, but we will put to test throughout the paper. Given this scenario, the first question to be addressed is, then, "is language a chaotic system?"

2. Language as a Chaotic System?

[9] As Gallego (2010) puts it, this implies replacing "label creation" by "label recognition": if {V, DP} is to be labeled {VP}, it is because of interface requirements and not because a new element is introduced to the derivation or the arbitrary application of categorial feature percolation.

The main proposal of this paper is that natural language *is* indeed a chaotic system in the technical sense of the expression (which, in turn, has far-reaching consequences for a theory of the mind, as we will see below). Our discussion, even though related to that in Uriagereka (1998, 2000, 2002), has different foundations and consequences, since our focus is not put on language "warps" and we make no reference to the "paradigm" / "syntagm" tension that is essential in his (2002) argumentation, but on more general properties of physical systems, that happen to be visible in language. Our definition of *Concatenation* should suffice to make our discussion different from Uriagereka's as far as generation is involved, since he assumes the traditional conception of Merge, even though it must be noticed that the "warp" notion of folding dimensions is perfectly compatible with our own proposal, and expressible in its terms.

Since Uriagereka's articles are considered the first application of chaos theory and entropy to language within a generative framework (but see Shannon, 1948 and García Mayoraz, 1989 for earlier and more physics / mathematics oriented discussions), they deserve some comments as our proposal greatly differs from his.

Uriagereka assumes the following tenets (2000: 865):

> *(7) Within local derivational horizons, derivations take those steps which maximize further convergent derivational options.*
>
> *(8) All else being equal, derivations involving fewest steps outrank their alternatives.*

We find some problems with these assumptions and their theoretical consequences, acknowledging that the framework was devised with the tools in hand at that point. To begin with, the concept of *local derivational horizons* is left undefined. To our knowledge, two options arise:

1) Chomskyan endocentric phases (Chomsky, 1998, 1999, 2005)
2) Each application of a transformational rule (Epstein & Seely, 2002)

This is, in any case, problematic, since not only the identity of the local domains is unknown, but also where they are defined. In our case, domains are exclusively interface-defined by *Analyze*. Therefore, stipulations are replaced by (third-factor, i.e., non-faculty specific), interface conditions. Of course, this could be taken as a rhetoric exercise with no impact on the theory itself, but notice that DFI is a condition that optimally applies to all mental systems. If this is so, then little does it help to call it 3rd factor or 1st factor (after all, if our constraint is as basic as we believe it is, it must be part of the genetic endowment), it is not specifically linguistic, period. Since locality has been so far analyzed within linguistic structures in mainstream studies, we believe this is a step forward.

Uriagereka's (8) is, however, more conflictive from a contemporary perspective as it overtly builds on Chomsky's (1995) proposal about reference sets (then defined to apply the Minimal Link Condition). His derivations assume the existence of a reference set $R = \{D_1, D_2 \ldots D_n\}$, where D is a derivation, each built by taking "decisions" that, in turn, allow different possibilities. Derivations can be compared, in this system, with respect to the number of possibilities each derivational decision (e.g., Merge / Move) allows. According to Uriagereka, if an operation Σ allows an n number of continuations and an operation Σ' allows a number m of continuations, if $m < n$, then Σ increases derivational entropy and Σ' is to be preferred. Problems with this approach arise from the very beginning of the argumentation. To begin with, the system is syntactically based and centered, with the interfaces playing no role at all in determining cost / "derivational entropy", in consonance with a strongly constructivist system which needs syntax to be partly interpretative in order to read previous derivational steps and, what is more, allow massive *look ahead* so that all possibilities have to be *fully derived* to be compared. In our opinion, this does not imply any optimization of the mechanism but, on the contrary, overloads the computation as there is no *a priori* way of limiting, say, m. Theoretically, m could be *infinite*, as there is no element in the Array that necessarily conveys an instruction to end the derivation, and no halting algorithm has been proposed within generativism beyond stipulations about phase heads and complement transfer.

Returning to the ground problem of *reference sets*, a *crash-proof* model like that argued for by Stroik & Putnam (in press) and Krivochen & Kosta (in press) has a reference set that is *unary*: the system generates (i.e., derives) one and only one object, an optimal candidate. In our framework, if each and every derivational step is Analyzed and, moreover, operations are subjected to DFI, the computational system only generates what can be read. In our proposal, like in Survive Minimalism, there is no point in talking about competing derivations or degrees of optimality: if something reaches the interfaces, it is as optimal as it can be, something that has a very close parallel in biological and physical systems with the concept of 'non-genomic nativism' (see Cherniak, 2009 for details).

The main proposal of this paper is that natural language is indeed a chaotic system in the technical sense of the expression. Let us now review some characteristics of chaotic systems according to Bernárdez (1995, 2001), Baranger (2004), Udagonkar (2001) and Boccara (2002), and then proceed to the discussion of whether language is one of them:

Following Bernárdez's characterization, chaotic systems are:

14) Chaotic systems are:

- *Open* to external influence[10]
- *Complex* (i.e., contain subsystems)
- *Dynamic* (i.e., change over time)

Natural languages seem to fulfill all of these intensional requirements: they are *open* because constraints to apply to an output[11] are not abstracted by a mind unless in contact with data in the phenomenological world and it is also a methodological mistake to study language in the mind-brain completely isolated from other systems insofar as those systems impose conditions upon language design; they are *complex* because *syntax* (the *generative* component) interfaces with other two components of interpretative nature and language comprises all three (FLB in Hauser, Chomsky & Fitch, 2002); and they are *dynamic* because the number of constraints that apply to the output of the generative component changes over time, as they are subsumed to other, more fundamental constraints until reaching the optimal scenario: a crash-proof system with only one constraint. A caveat is in order here: the more constraints we have, the more stable the system will be: after a certain number of constraints, a "threshold", the change is suddenly perfectly ordered and predictable. If we have only few constraints, the result will be a system tending to infinite complexity instead of achieving internal balance after a certain period of time. In the end, we cannot fully dispense with constraints upon the output of the generative component (say, for example, an OT-like EVAL), but basic economy considerations warn us against keeping a number n of constraints when $n > 1$ (as it happens in most, if not all, OT-related research), so we will subsume all constraints to our *Dynamic Full Interpretation*. Moreover, as we have suggested in Krivochen (2012b), acquisition can be expressed in terms of constraint abstraction from the input on an early stage and constraint elimination by subsumption in later stages of acquisition until reaching the optimal $n = 1$.

We have used Bernárdez's (1995, 2001) approach to chaos in language because it is the best attempt of which we are aware of making a chaotic model of (externalized) language. However, should we take other characterizations, for example, the widely accepted approach of Boccara's (2002), it works just as well. Consider Boccara's claim that complex systems show the following characteristics (2002: 3):

- They consist of a large number of interacting *agents*

[10] A note is in order here: if we say that chaotic systems are hypersensitive to initial conditions, we need a factor that produces a perturbation in the system. Such a factor is not a part of the system itself, from here follows the concept of "open" we are managing.

[11] Whenever we talk about *constraints*, we have in mind an OT-like architecture, and the definition of *constraint* that follows from that architecture. For further discussion, see Krivochen (2012b).

- They exhibit *emergence*
- The emergent behavior does not arise from the existence of an external *controller*

Notice that the use of the word "agent" does not presuppose volition: Boccara himself analyses the stock market as a self-organizing complex system without taking the intentions of the stock buyers-sellers into account. In language, we have a certain number of elements to manipulate, the characteristics and behavior of the collective structure does not directly follows from the individual characteristics of the interacting elements (that is to say, the *meaning* of a syntactic structure is not only in the elements, but also in the way they organize, something that cannot be predicted from the elements alone), what is called *emergence*[12]; and finally, this emergence is not the result of the existence of a *central processor* as in Fodor's (1983) modular model, since we follow Massive Modularity premises (Carruthers, 2005), according to which there is no "controller" but only modules (or "vertical faculties") and *bare output conditions* (Chomsky, 1995 et. seq.). In one way or another, then, language can (and, we claim here, should) be seen as a chaotic system.

DFI implies very local *evaluation* (although, *contra* Heck & Müller, 2007 and Müller, 2011; no *optimization* is implied). *Analyze*, on the other hand, applies to each output of *concatenation* (i.e., Merge) to check whether a given unit is fully interpretable by the interfaces. The scenario we have sketched so far, then, consists on a free generator, in the sense that it is not constrained by intra-syntactic filters and an evaluator, which in its "mature" state consists on a single constraint, DFI. The application of an extremely local evaluation procedure gives *balance* to the system, what is normally called "negative feedback": after an external perturbation in the system (for the initial state of a mental faculty, say, the perturbation would be the contact with the "external physical reality"), it balances itself, thus getting to the stable state we are interested in analyzing. System self-organization is expressed by Bernárdez (2001) using Shannon's (1948) equation for entropy augmentation[13] (taken from Bernárdez, 2001: 9):

---

[12] Interestingly, the concept of *emergence* can also be applied to *locality principles*: "*Emergent properties* are large-scale effects of a system of locally interacting agents that are often surprising and hard to predict" (Boccara, 2002: 23). If language displays chaotic characteristics, it is to be expected that it also displays locality effects, since locality is, apparently, a condition for complex systems.

[13] The use of this equation implies, as we will see later, that Language is a *system of locally increasing entropy* (i.e., $E > 0$) and its value increases as the system develops. We consider this an essential property of the system, both from an ontogenetical point of view and from a *derivational* point of view. The notion of *redundancy* Shannon uses as a tool for reducing entropy in information transmission is perfectly applicable here: in an expression like

i) Las sillas

15)

$$H = -\sum_{i=1}^{N} p_i \log p_i$$

where N is the number of individuals (i.e., constraints) and *p* the possibility of finding *i* of the N individuals in active state. Interestingly, when the number of individuals and frequency of interactions between then increase, the system becomes more predictable, following patterns that traditional mathematics were unable to appraise. *Chaos Theory*, initially developed as a meteorological model by E. Lorentz (among many others in different fields) is an alternative mathematical model that precisely deals with complex systems in a simple way, allowing us to understand the dynamics of an open system, hypersensitive to initial conditions. Our thesis is the following: initial conditions correspond to the *genotype*, and perturbations that lead to drastic effects on the output (i.e., the *phenotype*) are Copy-Modify operations (Watson, 2005) in crucial regulating genes, as we will see below. This means that a modification in a regulating gene, slight though it might be (say, 3 aminoacids), can produce a significant mutation with noticeable effects on the final state.

To begin with our argument, we will define the concept of *entropy*, which will be the *leitmotif* of the present paper:

16) *Entropy is a measure for disorder*

Whereas (16) is uncontroversial in modern physics, it is also of little informative value: we still do not have a definition of "disorder" (or even of "order", for that matter), and we lack also a circumscribed field for the definition to be scientifically useful. Let us enrich the definition, then:

17) *Entropy is the measure of the free energy that is released and cannot be used in a changing physical system, provided that the system is free and in normal temperature.*

The$_{\text{FEM PL}}$ chairs$_{\text{FEM PL}}$

Morphological doubling of φ-features can be posited to occur in order to make them more *salient*, and therefore counter entropy. The index of entropy of each language (E_L) would vary following the following equation:

ii) E_L = total number of morpheme tokens in a Syntactic Object SO / number of morpheme tokens that convey the same information in SO

This theory is still to be fully developed within a wider biological-mathematical framework.

This definition is more useful, since we have a definition that is both general enough to be applied to different physical systems and specific enough so that it is scientifically interesting. We will work under the hypothesis that the aforementioned definition is valid for *every physical system*, and, ***since biological systems are particular instantiations of physical systems, they are also analyzable from an entropic perspective***[14]. We claim that biological systems are particular instantiations of physical possibilities since, for example, neurological optimization ("non-genomic nativism", Cherniak, 2009) is restricted to the possibilities licensed by the physical organization of matter and the configurations it can adopt with basis on the physical laws that rule the phenomenological world. There is already a considerable amount of work within Biological Physics regarding the concept of entropy in biological systems, but we will try to give another turn to the screw by analyzing entropy in the genotype-phenotype relation and within derivational procedures, taking "derivation" in a wide sense to be the successive application of *concatenate* (and thus being available in every physical system). Let us formulate the working hypothesis we will manage in the course of the paper:

18) *Language is a* ***chaotic*** *physical system, hypersensitive to initial conditions*

This first formulation will do for the time being. Let us see how the system would work.

3. Language as an Entropic System:

How and when does entropy enter the game? Let us consider the model of global system entropy increase (i.e., $dS > 0$) presented in García Mayoraz (1989), where S = entropy:

19) $\Delta S_{\text{SYSTEM}} = \Delta S_{\text{ENVIRONMENT}} + \Delta S_{\text{PROCESS}}$

In our case, we have worked with a *biological* system (recall that the generative approach considers language to be a biological system, a state of a specific faculty in the mind-brain of a speaker) in *normal temperature* (which can be replaced by any other *ceteris paribus* condition, like the absence of language impairment), with an environment that favours some phylogenetic outcomes over others (a weak version of so-called "natural selection"), and Copy-Modify (C-M) operations as the relevant process affecting the genes responsible for the development of areas of the brain related to linguistic functions (FOXP2 and Broca's area, for example):

[14] That is, the characteristics a biological system can adopt are constrained by the physical principles ruling the environment in which that biological system is to develop. Udagonkar (2001) is a good introduction to some problems of physical biology.

20) $\Delta S_{\text{BIOLOGICAL}} = \Delta S_{\text{BIOSPHERE}} + \Delta S_{\text{COPY-MODIFY}}$

Considering the fundamental premise that genes encode *information*, the equation for entropy variation in information is also relevant, since C-M affect qualitatively the conveyed information to be interpreted by specific cellular organs. The entropy variation equals the variation in the quantity of information over the time it takes to transmit it, mathematically,

21)

$$\frac{dI}{T} = dS$$

Given the following definition of *information* (García Mayoraz, 1989):

22) $$\frac{1}{P} \log_2 = I$$

Where P expresses the probability for an object to occur (and 1/P is the inverse of that probability). We use a logarithm with base 2 because we express information in BITs, that is, *binary digits*, which can adopt two values (for the time being, but see section 5 below). Arguably, in genetics, the information equation should include a $\log_4$, since there are four nucleotides (A, G, T and C), but since only two combinations are possible (i.e., A-T and G-C), the $\log_2$ will be maintained for the time being as far as DNA is concerned (we will come back to this below, in any case). Taking this approach as valid, if a genome has multiple copied genes, the possibility for certain nucleotide chains to recur is high, so the "fine tunings" performed by C-M operations are fundamental to assure the informative value of the genome, as modifications are *never repeated*. This means that P is very low, possibly < 1/20.000 (as modifications do not alter a whole sequence, but some nucleotides, as in the case of FOXP2).

This is very relevant when considering a system in which there is the possibility of *copying*, since it becomes a *redundant* system, and redundancy is a feature that is present in language as morphological doubling on the phonological side (morphemes, clitics) and remerge of predicates on the semantic side[15] (recall that Merge, as an operation, is driven by DFI). Redundancy is a tool for countering entropy: the more redundant a system is, the less information is left behind (never lost or destroyed, as information is energy). Redundancy increases the value of P for a certain object, with which the value of I decreases.

The relevant formula for relative redundancy is usually represented as follows (where S is entropy):

[15] For example, Negation raising, Quantifier raising, etc. In sum, all those processes that require that semantic scope be different from phonological precedence relations.

23)

$$1 - \frac{\text{actualS}}{\text{maximumS}} = R$$

In perfect order –in a binary BIT system-, R = 1 and S = 0, but the system is not informative anymore since it becomes highly predictable regarding both the units and their combination. If we were dealing with a linguistic derivation to be interpreted at C-I, a perfectly ordered system would imply no positive (i.e., extra) cognitive effects when trying to build a propositional form: in a word, we would learn nothing. Such a linguistic stimulus would not be *relevant* in the technical sense.

In greatest disorder, R = 0 and $S_m = 1$, but total informativity is not possible in practice, since at least the code is already known, and the system is inherently redundant. Moreover, such a state or affairs like total disorder would not be relevant either since every informational unit (say, a proposition) is computed in a *context*, a set of known informational units that enter a context-sensitive deductive process to derive new propositions, *positive cognitive effects* (Wilson & Sperber, 2003). If nothing is known, nothing is possibly relevant.

The advantage of redundancy-allowing systems in practical terms is that there are more possibilities that the information conveyed by the objects whose possibility of occurrence P represents in (21) will be eventually recovered without much trouble (in mental terms, "cognitive effort" using the terminology from Sperber & Wilson, 1986). What the system loses with entropy, it gains in faithfulness.

3.1. *Entropy in linguistic derivations:*

A *derivation* will be taken here to be a successive application of an algorithm. In more concrete terms, a *derivation* is the set S containing the objects created by the successive application of the *concatenation* function we defined as follows:

24) *Concatenation* defines a *chain* of coordinates in *n*-dimensional generative workspaces W of the form $\{(x, y, z\ldots n) \subset W_X \ldots (x, y, z\ldots n) \subset W_Y \ldots (x, y, z\ldots n) \subset W_n\}$ where $W_Y \equiv W_X \equiv W_n$ or $W_Y \neq W_X \neq W_n$. If $W_X \neq W_Y$, they must be *isodimensional*[16].

Concatenation is a free operation, blind to internal characteristics of the objects it manipulates (but *format*, see below). Nothing impedes, then, that concatenation applies indefinitely, which gives us infinite

[16] This means, not only having the same number of dimensions, but also *the same dimensions* (e.g., a Calabi-Yau figure has six specific dimensions, therefore, only a W with those very six dimensions will be taken as *isodimensional* to a Calabi-Yau figure).

use of (in-)finite media, in Humboldt's terms. However, when the result of a number X of applications of concatenation has to be interpreted by an interface system (which is not the case in arithmetic, for example), the derivation is driven by Dynamic Full Interpretation, which we repeat here for the reader's comfort:

25) *Dynamic (Full) Interpretation*:

Any derivational step is justified only insofar as it increases the information and/or it generates an interpretable object.

Let us see what this implicates for the entropy of derivations. If entropy reaches its maximum level when all elements are equally probable, any restriction regarding the set of elements that can be merged in the following derivational step (see Putnam's 2010 definition of "Soft crash", which includes this notion of *local derivational unit*) will make entropy decrease, since the system is asymptotically tending towards an *ordered state*. Let us see an example. A derivation starts when some items are drawn from the Lexicon, the full set of conceptual and procedural elements represented in a speaker's mind. Those items are *types*, and each of their instantiations in a derivation is a *token*. The first step in a derivation, then, is to conform a *type-array*, containing types of all the elements that will be used to generate a sentence. For example:

26) *Type-Array* = {D, P, $\sqrt{\alpha}$, $\sqrt{\beta}$}

Of course, any of these elements could enter the workspace W first (if no stipulations are assumed), so the first step in a derivation has the maximum entropy. Let us assume $\sqrt{\alpha}$ enters the derivation first, as would be expected in a 2-D tree-like representation, but completely irrelevant in an *n*-dimensional syntax. This makes the entropy decrease, since not all elements in the initial Array may enter the derivational space and undergo concatenation with $\sqrt{\alpha}$ ***while respecting DFI***. Depending on the conceptual structure that this syntactic derivation is to embody, the possibilities for P vary, but assume we want to build a fully-fledged prepositional phrase, containing two {D} constructions related in a central / terminal coincidence manner by means of P. If this is the case, the possibilities for P = 0, since the conceptual structure requires entities (i.e., DPs or, in our terms, {D, $\sqrt{}$} structures); and the possibilities for $\sqrt{\beta}$ = 0 as well since roots are semantically underspecified and cannot be interpreted by C-I unless within a larger structure containing a procedural element (but see De Belder & van Craenenbroeck, 2011 for a different view on root-to-root merge). This means that we have an optimal situation: only one type-candidate satisfies DFI. The following derivational step would be:

27) *Concatenate* (√α, D) = {√α, D}

Notice that should we have introduced D first, the results would have been the same since generation does not care about order or "side" of the tree in which symbols are inserted, unless one is willing to concede that $C_{(HL)}$ is both generative and interpretative. If such a comment applies to traditional Kynean trees, it sure does to our multidimensional symbolic structures (see Krivochen, 2012a for details on this last point). For the purposes of future computations, this structure will be interpreted as a D, this is what we call the *label* of the construction (otherwise, no information would have been *added* strictly speaking since the whole structure would still be interpreted as an underspecified root). Having thus a {D} structure, the situation changes: even though merging another D token is not possible if DFI is to be satisfied, we still have, *in principle*, three candidates, with which the possibilities for each rise from 25% to 33, 33…%. This increases the entropy again and, if we had no way to determine what must come next, the derivation could not continue. In other works we have argued in favor of a pre-linguistic conceptual structure very much on the line of Jackendoff (1997, 2002) and Mateu (2000a) that is instantiated linguistically via the *Conservation Principle*[17] (hereafter, ConsP):

28) ***Conservation Principle****: information cannot be eliminated in the course of a derivation, but it must be instantiated in the relevant system in such a way that it can be read and it is fully preserved.*

If this pre-linguistic conceptual structure was *locative*, then a {D, {P, D}} construction is to be built in W (remember that {D} = {D, √}). However, we can also derive the better candidate from purely post-syntactic interface conditions, which is another (perhaps less controversial) option. Assuming Brody's (1995) *Radical Interpretation Thesis*, every element must be receive an interpretation and, we will add, if the element conveys procedural instructions as to how to relate conceptual elements, this information must be represented syntactically, so as not to lose information (and therefore incur in a violation of the ConsP). We have {P} in the Type-Array, which conveys *locative* procedural instructions, in terms of *central-terminal coincidence relation* between a Figure and a Ground (Talmy, 1983, 2000; Hale &

[17] Lasnik, Uriagereka & Boeckx (2005) borrow *Conservation Principle* from physics, and state the following law:

1st Conservation Law:

All information in a syntactic derivation comes from the lexicon and interpretable lexical information cannot be destroyed.

The problem with this law is that it makes use of lexical information taken from a pre-syntactic and monolithic lexicon, which is the norm in orthodox Minimalism, but with which we will not work. Moreover, this formulation is limited to linguistic structures, whereas ours is wider in scope.

Keyser, 1993, 2002; Mateu, 2000a, b). If we introduce any other token into the derivation the procedural information conveyed by {P} will be lost since there will be no available instruction to relate two structures, typically of the type {D, √}, at the semantic interface: any other element, even though procedural, would not be relational. So, the ConsP favors the introduction of a {P} token in W[18]. This way, the derivation continues with variable levels of entropy, basically responding to the following dynamics: in a Cartesian system, X = derivational steps (i.e., applications of Token-Merge), and in the Y axis 0 = S and 1 = R (see above), the function tends asymptotically towards 1 and 0 in a sinoidal manner, such that entropy increases from X = 0 tending towards 1 (i.e., perfect order) as a domain is built and the possibilities for each element to appear change. Once a domain is recognized by the relevant interface as complete, entropy tends towards complete disorder again, since there is no non-stipulative way (unless the reader accepts, as we do, the idea that syntactic structure in the narrow sense embodies a syntactic conceptual structure and the mapping, though not uniform, is ruled by ConsP) of determining which type will be introduced in W as a token.

We see that entropy behaves in a chaotic way, but, crucially, this is so only *within a local domain*, where *locality* is defined according to interface conditions and by no means as an *a priori* concept: a local domain is the minimal fully interpretable unit for an Interface Level IL. This has the advantage of defining "phases" in real time and without resorting to assumptions of Endocentricity which have proven problematic ever since put forth. Assume new domains as defined above are created departing from separate (or, even, parallel, since there is no *a priori* limit to the number of W activated simultaneously but a cost-benefit ratio) Type-Arrays: each time a new domain is built, the entropy value is maximum again since all elements are equally probable. This means we have to consider entropy dynamics in two scales: global and local. *Global* entropy would be the specific value for a domain *d*, whereas *local* entropy is the value of entropy at every point in the derivation (a rough equivalent to *harmonic serialism* in OT-like models of syntax, but including both a biological and a mathematical dimension).

3.1.1 *A Sample Derivation:*

What follows is a fully explicit derivation of a ditransitive clause (i.e., caused and with a locative node). The examples are literal translations of each other:

a) John put the book on the table (English)
b) Juan puso el libro en la mesa (Spanish)

[18] This could be regarded as a *weak formulation* of the theory. A *stronger formulation*, in favor of which we will not argue here –but we have done so in previous works- would claim that {P} is the *only* token whose introduction in W would satisfy ConsP and DFI.

c) Iuluis librum mensā posuit (Latin)
d) John legte das Buch auf den Tisch (German)
e) Σοκράτες άφησε το βιβλίο στο τραπέζι (Ancient Greek)
f) Ivan položil knigu na stol (Russian)

In this derivation, Merge is to be understood as *Monotonic Merge* unless explicitly indicated, that is, atomic elements conform a single *command unit* (see Uriagereka, 1999 for details).
We will also resort to parallel derivational spaces, which have been discussed elsewhere, see Uriagereka (1999) and Krivochen (2012b). We will maintain that mechanic here, since C-I also appears to be able to derive several representations in parallel when deriving *higher-level explicatures* and *implicatures* (Wilson & Sperber, 2003). The availability of parallel syntactic workspaces has also been justified within Radical Minimalism from a mathematical point of view:

In our formal definition of Merge we already introduced the concept of *n*-dimensional workspace. As a reminder, W is an *n*-dimensional *generative* workspace. Taking two distinct workspaces W_X and W_Y, either

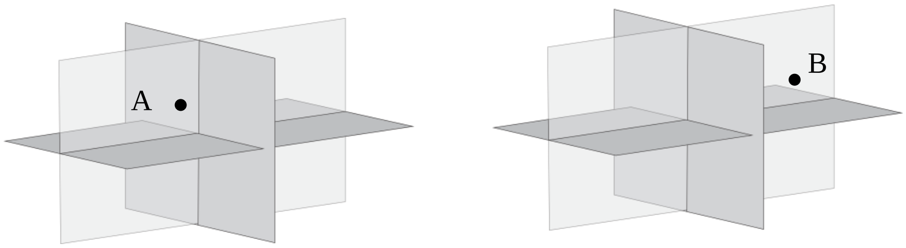
a) $W_X \equiv W_Y$ *iff* $\forall(x) \mid x \in W_X, x \in W_Y \wedge \nexists(x), x \in W_X, x \notin W_Y$

b) $W_X \neq W_Y$ *iff* $\nexists(x) \mid x \in W_X, x \in W_Y$

c) $W_X \cong W_Y$ *iff* $\exists(x) \mid x \in W_X, x \in W_Y$

This allows us to define the relations of *identity*, *difference* and *similarity* between syntactic objects in set-theoretical terms. *Identity* holds between W_X and W_Y if and only if every element of W_X is also an element of W_Y and vice versa (formalized in (a)). If this condition obtains, there is also *logical equivalence* between W_X and W_Y in all relevant contexts. *Logical equivalence* is entailed by *identity*, as it is to be expected. *Difference* holds if and only if there is no common element between W_X and W_Y, which means that they are not set-theoretically related (formalized in (b)). *Similarity* is a relation in which there are common elements *at least* -but not necessarily *only*- one, between W_X and W_Y (formalized in (c)).

29)



Workspace 1 (W1) Workspace 2 (W2)

Merge (A_{W1}, B_{W2}) by *Binary distinct Merge* = {A, B}

i.e., A ∈ W1, B ∈ W2

We will just build a *complete thematic domain* (what is traditionally called a Verb Phrase or VP) *in a ditransitive structure*, leaving the derivation of higher nodes (Modality, Aspect, Tense) to the interested reader. We believe the object we will build serves our purposes of showing derivational local entropy. We will also assume that every instance of (monotonic or non-monotonic) Merge generates an interface-legitimate object, something that is not necessarily the case in a *restrictivist system* with free Merge (Cf. Frampton & Guttman, 2002). In any case, the generation of "momentarily" illegible structures can be tolerated because of Putnam's (2010) definition of *Soft Crash*, as we assume there is a "local derivational unit" to repair the violation:

> <u>Soft Crash:</u> *If a syntactic object α cannot be interpreted at an IL in any and all of its features, α is neither useable nor legible at IL, iff α cannot be combined with another local derivational unit that repairs the violation(s) of α.*

The concept of "local derivational unit" is here restricted to the very next derivational step, provided that DFI holds.

Regarding "categorial interpretations", that is, how "categories" arise in the derivational dynamics we have proposed, we will assume that the local relation between the procedural element D collapses the root's categorial potential to Noun (i.e., sortal entity) without excluding [cause] (e.g., in derived nominals from causative construals), T collapses it to Verb (extending-into-time perspective) and P collapses it into Adjective (see Mateu, 2000a, b for adjectives seen as locative predicates). The reference restriction process we mentioned above also applies to categories, as they are not primitives of syntactic theory but arise in the interaction between the syntactic component and the interpretative interfaces. Common sense may dictate that the primitive *cause* appears only in verbal (i.e., eventive) structures, but there is an aspect of the C-I/syntax interface that we have mentioned elsewhere and is essential to this: *this interface (and possibly, all other interfaces) is not transparent* (i.e., there is no exact correlation between a Relational Semantic Structure and its syntactic realization, as well as there is no exact isomorphism between the representations manipulated by two modules, even respecting ConsP). Consider the following example (lexico-semantic structures follow Jackendoff, 2002 and Mateu, 2000):

30) A. They destroyed the city = [$T_{PastPerf}$ [EA [cause [event GO [location CITY [TO [√DESTRUCTION]]]]]]]
B. The destroyer of the city = [D_{DefSg} [EA [cause [event GO [location CITY [TO [√DESTRUCTION]]]]]]]

The correlations we have established follow from strict distributional properties, and only distributionally specified enough elements are capable of generating a categorial interpretation (that is why we excluded *cause*). A lexical item, in our proposal, is nothing more than the interface-read local relation between a procedural node, distributionally specified enough and a root, allowing an *n* number of non-intervenient nodes in between. Formally,

31) A lexical item LI is a structure {X…α…√} ∈ W_X, where X is a procedural category (Determiner, Tense, Preposition), α is an *n* number of distributionally underspecified non-intervenient nodes at the semantic interface, and √ is a *root*.

On the same configurational vein, Case is taken to be an epiphenomenon of the local relation between {D, …√} constructions and procedural elements. These interpretations will follow Krivochen (2012b) ***Case Sphere Theory***, which we very briefly summarize here:

Nominative: *read off* from a {Time, {D}} local relation, and *interpreted thematically* (in the explicature building process, see Sperber & Wilson, 2003) as Agent / Force

Accusative: *read off* from a {Cause, {D}} local relation, and *interpreted thematically* as Theme, the object (Figure) located in / moving towards, etc. a Ground

Dative: *read off* from a {P, {D}} local relation, and *interpreted thematically* as Location, the Ground in Talmy's terms.

We have argued against the feature system in previous works, now we will present a summary of our proposal. The reason we do this is because we will present a derivation driven by invasive interfaces, not by the need to value or check formal features, and such a position must be justified. The basic claim underlying all Agree machinery, as we understand it, is "things establish relations to one another" (Zeijstra, p.c.). If that is the intuition, our system is completely compatible with it. Let us first summarize what is needed in an Agree-based syntax (Chomsky, 1999, Pesetsky & Torrego, 2000, 2004, 2007):

a) Dimensions

b) Values

c) An unvalued instance of a dimension [D] that acts as a *probe*, searching for a valued counterpart

d) An operation to relate a valued and an unvalued instance of the same dimension

Given DFI, the orthodox claim that some features enter the derivation unvalued and are therefore uninterpretable (Chomsky, 1999) and must thus be valued in a probe-goal relation (this process driving the derivation) is not an option: as already pointed out by Epstein & Seely (2002), an Agree-driven system requires the syntactic component to be sensitive to feature value. This implies that the computational component, which we have reduced to a single algorithm, is (at least) partly interpretative, since it must have access to inner substantive properties of the elements it manipulates, going beyond the identification of their format. In Krivochen (2011b) we have argued in favor of a system in which [val-F] features are replaced by semantically interpretable dimensions comprising *in abstracto* all possible outcomes, what we have called "quantum dimensions" given this proposal's obvious parallel in Quantum Mechanics. The architecture we have outlined above points towards a theory in which there is no superfluous information in interface terms (i.e., no EPP, no formal Wh-, no undeletable Edge Features in so-called *phase heads*, see Chomsky, 2005b), but only elements with a range of possible interpretations (as in our brief discussion of theta-roles above, expanded in Krivochen, 2012b), the actual interpretation being determined only contextually *at the interface level*. Since Merge is blind, it cannot but put things together.

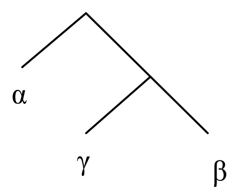
From an interface point of view, thus, *matching theory* (and its *valuation* alternative) *is unteneable* since formal features do not add anything informationally (that is, at the interfaces) as they must be *eliminated* before reaching them, and we therefore stick to the concept of *collapse*:

32) *Collapse* is an interface reading of a *local dependency* in which the interpretation of an element is defined by another, structurally higher, whose distribution is specified enough.

In order for this to work, we need to assume that elements bear an inner potentiality for *category* and, in the case of *nominal constructions*, also Case in the sense defined above, as a purely semantically relevant dimension recognized at the interface level as a relation between symbols (occasionally Spelled-Out, epiphenomenally).

The tree-like 2-D structure would look like follows:

33)



There is a locality requirement in this licensing process. A *dependency* is *Local* if and only if there is no intervenient object γ (of arbitrary complexity) such that: (i) the relation between α and γ is equivalent to that between α and β for interface purposes (ii) α, β and γ belong to the same W and (iii) γ is structurally closer to α than β, that is, *γ is not an intervenient object*. Such a condition is a rephrasing of Relativized Minimality, but interface-based. It is not the "type" of element (A-position, A'-position, head) that determines its being intervenient in each case but the interface effects (i.e., "drastic effects on the output", in Grohmann's (2003) terms) the possible configurations generate. Notice that there is an added condition on workspaces, which is absent in a polystratal theory that works with transformations upon phrase markers in a single workspace.

Now, let us turn to the derivation analyzing the variations in entropy as the process unfolds:

<u>Type-Array:</u> {D, P, √α, √β, √ε, *event, cause*}

This is the point of maximum entropy, as all elements, either procedural (D, P, event, cause) or conceptual (roots) are equally likely to enter the W.

1) *NS* Merge (D, √α) = {D, √α} Either D or √α could have entered the W first (for simplicity purposes, we will not consider P / *cause* / *event*, but notice that, being relational elements, their introduction would make the derivation to crash since they would have nothing to relate.

2) *C-I* Label Recognition {D, √α} = {D, {D, √α}} *This* {D} *will be taken as a unit for the purpose of future operations*. Incidentally, {D, √α } "categorizes" √ as N, following our definition.

3) *C-I* Analyze: not fully interpretable unit: {D} has to be in such a configuration that a Case interpretation can be generated, otherwise, an explicature will not be built since crucial information will be missing.

4) *NS* Merge (P, {D}) = {P, {D}} P*'s* locative *procedural instructions collapse the inner case potentiality on* {D} to the DAT sphere.

5) *C-I* Label {P, {$D_{[DAT]}$}} = {P, {P, {$D_{[DAT]}$}}}

6) *C-I* Analyze: {D}'s referential properties depend on the cumulative influence of Time, Aspect and Modality, if it is a common name. Proper names are taken to be inherently manipulable by C-I (see Krivochen, 2012b for arguments and examples). Not fully interpretable yet. Moreover, the locative relational element P requires another element (a *figure*) to relate to the *ground*.

7) *NS* Merge (D, √β) in parallel to (1) = {D, √β} Labeling and Analyzing also take place. No procedural head can collapse {D}'s Case potentiality, so the structure is not yet fully interpretable.

8) *NS* Merge by *Structural Format* ({D, √β}, {P, {P, {D, √α}}}) = {{D}, {P, {P, {D}}}}

9) *C-I* Label {{D}, {P, {P, {D}}}} = {P}: if either of the D types had been recognized as a label, the locative procedural value of P would have been lost. A referential expression {D} is legible at the interface, but in that case P would be a superfluous element since its contribution to the explicature building would be null for the purposes of further computations.

10) *C-I* Analyze: {D, √β} has a Case potentiality still uncollapsed and is thus not fully interpretable. Therefore, the whole domain P ∋ {D, √β} is not interpretable either. This leads us to the following principle:

A domain d is interpretable by IL iff $\forall(x), x \in d$*, x is interpretable by IL.*

At this point, *entropy* goes back to the maximum level, since DFI cannot stipulate what should come next. However, if we accept Jackendoff's (2002) –among others- idea of a pre-syntactic conceptual structure, which we have developed in our own terms, the presence of an eventive primitive in the conceptual structure will require the introduction of *event* in the derivational space.

11) *NS* Merge (*event*, {P}) = {*event*, {P}}

12) *C-I* Label {*event*, {P}} = {*event*, {*event*, {P}}}

13) *C-I* Analyze: idem (10)

Here, the disorder of the system decreases again, since there is a closer conceptual relation between *cause* and *event* that between *location* and *event*: there can be caused events without locative nodes (e.g., *unergatives*) and locations with eventive-clausal flavor without an *event* node nearby (e.g., central coincidence relations). However, *cause* cannot appear if *event* is not there in the conceptual structure, since there is nothing to cause. This dynamics of varying entropy, regulated by DFI and the ConsP, are crucial for the analysis of derivations within Radical Minimalism.

14) *NS* Merge (*cause*, {*event*}) = {*cause*, {event}} Procedural instructions on the *cause* semantic primitive can collapse the Case potentiality on the closest {D} structure to the ACC sphere.

15) *C-I* Analyze: is {P} now fully interpretable? Let us assume a central coincidence P [WITH], which gives the P domain a clausal flavor since the analysis of Double Object Constructions (see Harley, 2002 among others) show that P [WITH] is semantically equivalent to V [HAVE]. P is then a fully interpretable object as there are no potentialities to collapse that would make the structure too underspecified for C-I to build a (partial) explicature.

16) *C-I* Transfer {P}: the interpretative systems *take* what they can read, instead of having a generative component that *sends* information. This proposal avoids the look-ahead problem that arises with the traditional conception of Spell-Out or, more recently, Transfer (Chomsky, 1998, 2005).

17) *C-I* Label {*cause*, {*event*}} = {*cause*, {*cause*, {*event*}}}

18) *C-I* Analyze: two procedural instructions will cause collapse, since there is no √ to provide the "semantic substance" needed for an explicature to be built. The semantic primitive *cause* licenses an "external position"[19], forcing the system to "wait one more turn".

19) *NS* Merge (D, √ε) in parallel = {D, √ε}. Idem (7).

20) *NS* Merge by *Structural Format* ({D}, {*cause*, {*cause*, {*event*}}}) = {{D}, {*cause*, {*cause*, {*event*}}}}.

21) *C-I* Label {{D}, {*cause*, {*cause*, {*event*}}}} = {*cause*, {{D}, {*cause*, {*cause*, {*event*}}}}}

Needless to say, {D} in (21) also has a Case potentiality, but that will be collapsed to the NOM sphere by finite T procedural dimensions. Raising-to-subject, that is, Token-Merge of {D} in a position licensed by T will take place if {D} is *thematic*, following Krivochen (2011a, c).

3.2 *Generation, Interpretation and Entropy:*

Let us begin this section with a quote from Folia et. al. (2011):

34) "*It is uncontroversial that any physically realizable computational system is necessarily finite with respect to its memory organization and that it processes information with finite precision (e.g., due to the presence of internal noise or architectural imprecision; Turing, 1936a, 1936b; Minsky 1967; Savage 1998; Koch 1999).*" Folia et. al. (2011)

"Finite precision" can be identified with the inevitable amount of free energy (i.e., energy that is capable of performing a work) that is lost (though not destroyed) is any derivation, i.e., processing of information. If our claim that biological systems are systems of increasing entropy, then the term "architectural imprecision" gains a whole new physical dimension: entropy is built within the system, as an inevitable part of it, just as uncertainty in quantum mechanics is not a result of instrumental limitations but of the very design of physical systems (Heisenberg, 1999). An important distinction must be made at this point: it would be inaccurate to regard the syntactic component as "imprecise", since there is no possible definition of precision in free generation via a conceptually necessary concatenation function. A generative workspace is not "precise" or "imprecise" in any non-trivial sense, mainly because *precision is a measure of interpretation*, not of *generation*. Therefore, we conclude that relevant issues of entropy arise, when it comes to derivations, *at the interpretative systems*. As a reminder, we will define

[19] Since there are no bar-levels, there is nothing to be "external" of. In our *n*-dimensional (n > 2) model of syntax, the traditional concept of external and internal makes no sense.

Generative Systems and Interpretative Systems (and their relations) as we understand them (Krivochen, 2011c):

35) *Generative Systems*: Generation equals *Concatenation* (i.e., *Merge*), a free, unbounded, blind operation that takes elements sharing either *ontological* or *structural* format[20] and puts them together. Let us take W to be an *n*-dimensional *generative* workspace. Taking two distinct workspaces W_X and W_Y, either

a) $W_X \equiv W_Y$ *iff* $\forall(x) \mid x \in W_X, x \in W_Y \wedge \nexists(x), x \in W_X, x \notin W_Y$

b) $W_X \neq W_Y$ *iff* $\nexists(x) \mid x \in W_X, x \in W_Y$

c) $W_X \cong W_Y$ *iff* $\exists(x) \mid x \in W_X, x \in W_Y$

36) *Interpretative Systems*: Interface systems, they have to read structural configuration build up by generative systems. In turn, an *Interface Level* IL is an *interpretative system* that has access to W_X and establishes legibility conditions regarding its output.

A provisional (but promising) conclusion is that *entropy* occurs only when there is a *generative-interpretative* interaction. If there is no *interpretative system* to read a structure, and evaluate "imprecision" with the measure of the Conservation Principle, there is no point in introducing *entropy.* A stronger claim is that only when there is *information* there can be *entropy*. Obvious as this may seem, it has far-reaching consequences: consider the following linear function:

37) $f(x) = 2x$

This simple function generates the series of even numbers until infinite. Is any information lost, or any disorder created as the values of *x* increase? If the set of natural numbers is infinite, then the possibility for each number to occur is $1/\infty$, asymptotically tending towards maximum disorder, but never reaching P

[20] A brief definition is the following (Krivochen, 2011a): "(…) *Ontological format refers to the nature of the entities involved. For example, Merge can apply ("ergatively", as nobody / nothing "applies Merge" agentively) conceptual addresses (i.e., roots) because they are all linguistic instantiations of generic concepts. With ontological format we want to acknowledge the fact that a root and a generic concept cannot merge, for example. It is specially useful if we want to explain in simple terms why Merge cannot apply cross-modularly: ontological format is part of the legibility conditions of individual modules.*
Structural format, on the other hand, refers to the way in which elements are organized. If what we have said so far is correct, then only binary-branched hierarchical structures are allowed in human mind. The arguments are conceptual rather than empirical, and we have already reviewed them: Merge optimally operates with the smallest non-trivial number of objects. Needless to say, given the fact that ontological format is a necessary condition for Merge to apply (principled because of interface conditions, whatever module we want to consider), the resultant structures will always consist on formally identical objects (…)"

= 0 (see (23)). Since there is no variation with respect to the possibilities of each element to appear, entropy remains stable: the second principle of Thermodynamics does not apply to mathematical systems, because there is pure generation. Within the mind-brain, *concatenate* is not subjected to entropy, unless a generative W interfaces with an interpretative system, for the aforementioned reasons. The interaction between the first and second principles of Thermodynamics thus proves essential for an accurate understanding of the dynamics of biological systems within the wider framework of physics.

4. A note on Recursion within Physical Systems:

To begin with, we will explicit the definition of recursion that underlies most (if not all) of current syntactic research:

38) *Recursion is an operation that applies to its own output*[21]

With basis on that definition, phrase structure based grammars can account for discrete infinity, mainly Chomskyan oriented theories. In these frameworks, *Merge* is taken to be recursive is the sense of (38) as it applies to the product of the previous application of Merge:

39) Merge (a, b) = {a, b}
Merge ({a, b}, c) = {c, {a, b}}

and so on indefinitely. Kitahara (1997), for example, formalizes Merge in the following terms:

40) Applied to two objects α and β, Merge forms the new object K by concatenating α and β.

Input: α, β

Concatenate α and β, forming K

Output: K

[21] We will not consider here the early EST definition, in which recursion meant that a symbol contained itself in some point of the rewriting function, e.g. (in bold):

i) **S** -> NP Aux VP

ii) VP -> V (**S** / NP / PP / AP / AdvP)

Notice that the definition of Merge (taken in turn from Chomsky, 1995: 243) is circular, since Merge is defined in terms of Concatenation, without formalizing either. Apparently, Merge is in itself a complex operation, as it comprehends Concatenate (α, β), Label (α, β) as K and Erase (α, β) (Kitahara, 1997: 5). It is difficult to see how such a complex operation could come for free in a Minimalist framework. Moreover, the generatie engine is enriched by adding Label, which we have already characterized as an interface-driven process, which involves *recognition* at the C-I interface rather than *creation* in the syntactic component, as in Kitahara' characterization. In our system, there is not even *bare phrase structure*, neither *phase structure* (Boeckx, 2008), but just *n*-dimensional *structure*. Following the definition of *recursion* in (38), Merge could in turn apply to K and L, forming M and eliminating by substitution both K and L. This requires the system to maintain K active in the working space, which is wiped clean when the input of the operation is eraser by substitution.

There is a fundamental question we have to ask ourselves when considering a derivation like that pictured in (40): how does the operation that its input in a derivational moment T_2 is *identical* to its output in a previous derivational point T_1? That is, how can an *operation* that applies in W be *interpretative* and *label-sensitive*?

The only answer, without resorting to any additional stipulations, is that it simply cannot. Therefore, *recursion* must be redefined within this stricter conception of purely generative workspaces, without including operations that fall within the range of action of interpretative systems. Our proposal is the following:

41) *Recursion is the possibility for an operation to apply more than once in the course of a derivation*.[22]

(38) was constrained by the ontology of the output (i.e., an *identity* relation had to hold between output_1 and input_2), whereas (41) is perfectly compatible with free-Merge frameworks, like Radical Minimalism. Notice that if we only have *concatenation*, the definition will be restricted to it, but this is a consequence of the specific RM theory, not of the programmatic definition. Moreover, if the operation does not need to be substance-sensitive (just format-sensitive), it can apply to many more objects in the natural world. If we combine (41) with our claim that *concatenation* is the *only* operation in the natural world, then it

[22] It is time to address another reviewer's comment: "(…) *the relation of concatenation to recursion as defined in the paper is mathematically flawed given that recursion and generativity as Chomsky formulated the concepts refine concepts of computability—a process ontologically independent of physical (biological) phenomena*." It is precisely against this independency that we argue. No operation or relation should be formulated if the physical-biological properties of the concrete substrate are not researched as well, since computational properties arise from the possibilities licensed by such substrate.

makes no sense to talk of *recursion* as a language-specific feature in a non-trivial sense, as any *structure* is actually *recursive*, only atomic units are not (in language, for example, *roots*. In the physical world, possibly *strings*). Some may object that, if we claim that any structure is recursive, then there is no interest in the concept of recursion, since it can apply to a very wide set of objects, without giving us any particular distinctive characteristic of system X. However, our belief is that specificity cannot rely on just one characteristic, just like a given phenotype is not a monolithic state but a set Φ including different characteristics, some of which may be shared with other phenotypes. The specificity of a given phenotype, then, resides on the *specific combination of characteristics* $\{\varphi_1, \ldots \varphi_n\}$, it is naïve, in our opinion, to look for a single characteristic to load with the whole bargain of being the distinctive feature in which all uniqueness of language resides. Besides, and by no means less importantly, if recursion is the sense of (41) is indeed found in *all* and *every* physical system (which requires experimentation, needless to say, but also a well-grounded framework to begin with), then we will have a very important generalization about the structure of the natural world at all scales. We follow Laka (2009) in her claim that, for a property / element to belong to UG (that is, for *x* to be the uniqueness-bearing characteristic) in a non-trivial way, it must fulfill two conditions:

42)

a) Innateness: the relevant property must be present in the genotype

b) Specificity: the relevant property must be *exclusive* of a certain faculty F

Radical Minimalism only acknowledges the conceptual necessity of a combinatory operation and something to combine for language to arise within the mind. Our point is that neither fulfills both of Laka's criteria, and therefore, UG is an *empty set*. Let us start analyzing the combinatory operation, as we have defined it above:

43) Merge is a *free unbounded operation* that applies to *two* (smallest non-trivial number of elements) distinct (see below) objects *sharing format*, either ontological or structural. Merge is, on the simplest assumptions, *the only generative operation in the physical world*.

If Merge is available in other mental faculties (as seems to be the correct view since compositionality can be found in visual perception, mathematical thought, conceptual semantics and other domains), then it fulfills (a) but not (b), thus not being a suitable candidate for UG. This, even orthodox Generative linguists recognize, for example, Boeckx (2010) adopting Gould's (2002) term *Umbildung*, that is, *recombination*, using things already available in separate biological systems to form something new. If Merge is just "putting things together", with no labeling operation involved (Cf. Boeckx's *Copy*, Hornstein & Pietroski's *Label*), then this is a strong biological point in favor of our idea. Now, regarding the elements combined, recall we only have *roots* (i.e., instantiations of *generic concepts*) and *procedural elements*. They share ontological format, so, for the purposes of this argumentation, they will be grouped as "elements combined".

In Krivochen (2011a) we appealed to the distinction between *generic concepts*, pure a-categorial semantic substance which is used by more than one faculty, and *roots*, linguistic instantiations of those generic concepts, underspecified and pre-categorial. If *roots* are linguistic instantiations of *generic concepts*, then they are *specific*, thus fulfilling condition (b). Now, are they innate? Our answer will be *no*. Let us think of the mind-brain in uniform terms: this means that whatever faculties we find, will have *analogous* ontogenetic development, which seems to be the simplest option and, as "*evolution is Minimalist*" (Bickerton, 2009), biologically plausible. If this is correct, then C-I must have an initial state, determining genotypic possibilities and a final state, a phenotypic expression of this potentiality. The initial state, we will call a *Construal Acquisition Device* (CAD) which, in contact with the phenomenological world, gives as a result conceptual eventive templates, which we could summarize as follows (Mateu, 2000a, b):

44)

i) Unergative: caused, without location (i.e., [EA [*v* [V]: run, vomit, sleep, speak)

ii) Unaccusative / Ergative: uncaused, with location (i.e., [V [Figure [[P] Ground]]: arrive, leave, go, come)

iii) (Di-)Transitive: caused, with location (i.e., [EA [*v* [V [Figure [[P] Ground]]]: give, put, take, show)

In our system, reminiscent of that of Talmy (2000) and Pylyshyn (2007) among others, information enters the mind chaotically, but it is organized in very much a Kantian way, into categories that exist as *a priori* forms in CAD: mainly, *Space* (if we accept that *Time* is conceptualized as a metaphor of space, for which there is neurocognitive evidence, see Talmy, 2000 and references therein), activating *temporal* and *parietal* lobes. However, the existence of more complex conceptual templates must be posited, as our awareness and understanding of the phenomenological world around us does not limit to *things* in a *location* (be it concrete or abstract). The interaction between CAD and the phenomenological world also licenses the different *Aktionsarten*, and it is here, perhaps, where the possibilities of different phenotypic outcomes may be clearer: let us imagine a "lightspeed world": in such a phenomenological environment, there would be *no achievements*, since, for example, a bomb could "be exploding" in imperfective progressive grammatical aspect up to the point of being a state if we stretch Time enough (the same would happen if space is stretched enough as in the horizon of a black hole, which is evidence in favor of the idea that Time is a metaphor of Space). In our phenomenological world, conditions of pressure, gravity in a "local" level, and the physical principles that rule the Universe (Relativity and Quantum mechanics, at different scales) determine the development of a species-uniform (within genotypical limits) C-I, and the "delimitation" of the conceptual continuum (recall Hjiemslev's continua for form and meaning) gives us those *generic concepts* we were talking about above.

If the reasoning above is on the right line, then we can say that *roots are specific, but not innate*, since they are the instantiation of generic concepts that result from the interaction between CAD and the phenomenological world. We have thus proved that, as far as Laka's criteria are concerned, UG is an *empty set*, and specificity lies on the combination of faculty-shared individual characteristics.

5. Chaotic language in a Quantum Mind

So far, within linguistics, it has been a fairly common assumption (either overtly or covertly) that the human mind works computationally, and that those computations are to be made explicit. Following this enterprise of explicitation of mental computations, Watumull (2012a, b) proposes that the formalism for a model of the mind is that of a Turing machine, performing serial computational steps one-at-a-time. In other words, a Turing-like mind (TM) assumes linear processes and, crucially, leaves little, if any, place for parallel computations and multidimensional workspaces. A TM works like a digital computer, with two basic activation values: 1 and 0, which are mutually exclusive. Filtering, if it occurs (that is, if there is *generation-interpretation* interplay), takes place one step at the time, and, crucially, *one constraint at a time*. The reason is that, if our generator function has produced n candidates to be evaluated, and we have c constraints to apply to each of the candidates to pick the one and only optimal output, in a single workspace and deriving serially there is no other option but to order the candidates according to some hierarchy and apply the constraints (which are actually "well-formedness" conditions) following another hierarchy, existent in OT models but equally unclear as to its grounds and justification (see Krivochen, 2012b for detailed discussion). Turing computation is "condemned" to be serial, understanding seriality as follows:

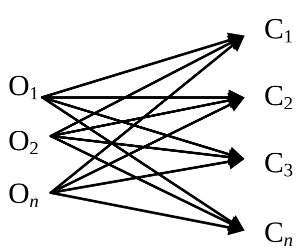
45) A process p is an ordered series of steps [x ^ x+1 ^ … x+n], where ^ is string ordering; and, if x is subjacent to x+n via x+y, then x+y is a *sine qua non* condition for the linear derivation to reach x+n departing from x.

Let us imagine the following situation: we have *n* candidates and *x* constraints, where both *n* and $x > 1$. For all natural languages NL and all constraints *x*, we assume that:

46) $\forall(x)\ \&\ \forall(\mathrm{S}) \mid \mathrm{S} \in \mathrm{NL},\ x(\mathrm{S})$

(46) is a crucial assumption for OT syntax, since it guarantees universality as well as the application of the constraints to all sentences belonging to a certain NL (that is why we used the predicate-argument notation). We now face the following problem: how and when do those constraints apply? One possibility is the Turing machine procedure, the serial and linear application of the EVAL algorithm, along the lines of (45). However simple this option may seem, we find it difficult to implement it in the kind of system we have been arguing in favor of in the present paper: even if statistical learning can be codified in a Turing machine (and thus, neural network learning), learning timing seems to pose a counterargument difficult to overcome for a linear theory. The EVAL function, which is assumed to have neural basis, is determined by the characteristics the hardware imposes over the software. Another, quite different possibility is that evaluation takes place not only in simultaneous workspaces but also *non-linearly in each workspace*. Given candidate representations and constraints, multiple workspace evaluation would look as follows:

47)



The problem is, how do we make our network process multiple candidates *simultaneously*? Assume the following candidates:

O_1 = What did you buy from the store?

O_2 = You did buy what from the store?

O_3 = What you bought from the store?

and the following EVAL component[23]:

[23] These constraints, or equivalent, can be easily found in OT literature. See Müller (2011) for multiple examples.

C_1 = Wh-CRIT (Wh- appear in Spec-C)

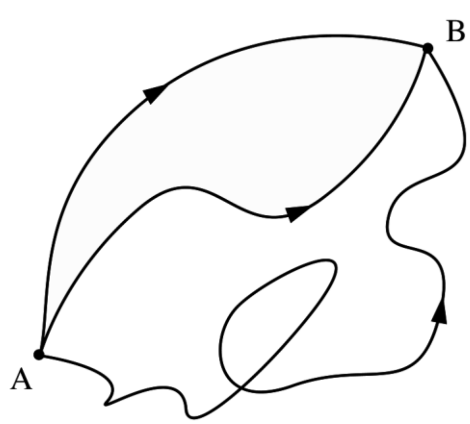
C_2 = θ-ASSIGN (Elements appear in the site in which they receive a theta-role)

C_3 = Full Interpretation (no superfluous elements in representations: traces, expletives, etc.)

C_4 = Subjacency[24] (movement cannot cross more than one bounding node)

This state of affairs is a very limited representation, but it will do for our purposes. The complex network of interactions between GEN and EVAL depicted in (47) should be performed in real time, so, how does the system select O_1 among its competitors? Our proposal, to be confronted with linear Turing-machine processors, is to think of this situation in the same terms in which Feynman thought about particle movement: the so-called *path integral formulation* (e.g., Feynman, 1948). The rationale is as follows: instead of *measurement* in charge of determining the specific location of (say) a photon in T, being its previous location a *wave function*; what we have is a functional integral over an infinity of possible trajectories to compute a quantum amplitude. The formulations are equivalent within quantum mechanics, and they both follow from Schrödinger's Equation, thus being incompatible with traditional particle models. A common representation of Feynman's approach is the following:

48)



Of course, the indicated paths are just three of infinite possible paths. The question now is how to implement this in a model of language, in which "infinity" is a notion strictly restricted to the outputs of a generative algorithm, slightly modifying Humboldt's approach. Mental entities are ultimately electricity,

[24] Depending on the reader's preference, Subjacency could also be expressed in terms of the Phase Impenetrability Condition or Relativized Minimality, among others. RM has the advantage of not defining *a priori* the identity of the intervenient nodes, whereas in Subjacency and PIC, the nodes are pre-defined (S and NP in the case of Subjacency and *v* and C in the case of the PIC).

obtained through neurological stimulation/inhibition as a function of the presented input: as they are ultimately *energy*, they cannot be created nor destroyed, just transformed (what we expressed via ConsP); and in a neural network the GEN-EVAL dynamics are to be represented as the translation from A to B in (48), being A and B points in a specific network within the phase space of all possible connections, of which the contact with the phenomenological world reinforce only one, which we can call an "optimal network" given input and output conditions[25]. Being A the set of candidates and B the EVAL component, each candidate is *simultaneously* applied all constraints by EVAL, based on the rationale behind the path integral formulation. Such computational method is consistent with the formulations of the *Quantum Human Computer Hypothesis* (QHC), whereas it is dubious whether it can be applied, maintaining the theoretical advantages and empirical prospects, in a Turing-machine model of language like that argued for in Watumull (2012a).

The QHC predicts that the mind will have the power to perform complex computations simultaneously and in interaction, a prediction that better suits "multi-component" perspectives on mental faculties (Baddeley, 2003b: 837) and working memory, as well as providing a completely new perspective on the interactions between mental capacities in an *n*-dimensional *logische Raum* only constrained by interface conditions on legibility. Essentially, as we have pointed out in previous works, the generative component can manipulate elements which have multiple potential outcomes (e.g., Case, which can be interpreted as NOM, ACC or DAT depending on the configuration. See Krivochen, 2012b for discussion and details), as it is blind to the characteristics of the manipulated elements. This situation, we have called the "ψ-state" of an element (e.g., a {D} construction), which collapses to one of the possible outcomes in a local relation with a distributionally specified procedural node (e.g., NOM = {T, …{D}}) *only* at an interpretative interface, should there be one. This means that the generative component can manipulate elements in their ψ-state for as long as transfer is not required by any interface. The points of maximum entropy, arguably, coincide with those in which there is "ambiguity", that is, derivational moments in which there are elements in their ψ-state and a suitable procedural node has not yet been merged. That is, for instance, our step (10) above, which we repeat here as (49):

49) "*C-I* Analyze: {D, √β} has a Case potentiality still uncollapsed and is thus not fully interpretable. Therefore, the whole domain P ∋ {D, √β} is not interpretable either. This leads us to the following principle:

[25] In this sense, we are assuming a generalized-delta rule in a supervised learning model. Nevertheless, the model can be applied in a non-supervised learning model without major adjustments.

A domain d is interpretable by IL iff $\forall(x), x \in d$, *x is interpretable by IL.*

At this point, *entropy* goes back to the maximum level, since DFI cannot stipulate what should come next."

is also a point in which, as we have said, there is an element with an uncollapsed potentiality. Therefore, the whole structure cannot be transferred, as *Analyze* does not recognize it as a fully interpretable object (i.e., a minimal-maximal transfer domain).

The basic idea, then, is that *if language is indeed a chaotic system, then its biological basis must be a quantum mind-brain*, which licenses the computational properties we can analyze in more empirically-oriented linguistics. Our hypothesis has proven useful when considering, for example, pragmatic processes, and the reformulation of linear semantic-pragmatic theories (e.g., Speech Act Theory) in a non-linear way (see Krivochen, 2012c), following the model of a biologically-based, third-factor oriented Relevance Theory, as presented in Krivochen (2012b). We have no doubt it could also prove useful when considering and revisiting other phenomena, both *theoretical* (e.g., ultrametrical distances in syntactic representations and the nature of phrase structure, see for example Roberts, 1998; or the nature and interpretation of Case) and –consequently- *empirical* (e.g., the parallel derivation of explicatures and implicatures in Relevance Theory; or the "syntactic" effects observed in mathematics and music and their comprehension, as well as language acquisition and impairment).

While many scientists consider such claims as the ones we have put forth here uninteresting given their generality, we believe they open the door for interdisciplinary study and provide interesting information about foundational aspects of formalizations of the Universe, possibly giving tools to simplify theories and make scientific exchange easier (see Uriagereka, 1998, for a detailed take on the issue). That is one of the goals that the present programmatic piece has set itself. Whether we have been successful to some extent or not, only time will tell.

6. Bibliography:

Baddeley, A. (2003a) "Working Memory and Language: An Overview". *Journal of Communication Disorders* 36 (2003) 189–208

(2003b) "Working Memory: Looking Back and Looking Forward". *Nature Reviews Neuroscience*, 4 (10): 829-839.

Baranger, M. (2004) *Chaos, Complexity and Entropy*. Ms. Cambridge, Mass. MIT.

Bernárdez, E. (1995) *Teoría y Epistemología del Texto.* Madrid, Cátedra.

(2001) De Monoide a Especie Biológica: Aventuras y Desventuras del Concepto de Lengua. Ms. Universidad Autónoma de Madrid.

Bickerton, D. (2009) *Adam's Tongue*. Hill and Wang.

Bierwick, R. (2011) "Syntax Facit Saltum Redux: Biolinguistics and the Leap to Syntax". In Anna Maria Di Sciullo & Cedric Boeckx, editors (2011) *The Biolinguistic Enterprise*. *New Perspectives on the Evolution and Nature of the Human Language Faculty*, 65-99. Oxford, OUP.

Boeckx, C. (2007) *Understanding Minimalist Syntax: Lessons from Locality in Long-Distance Dependencies*. Oxford: Blackwell.

(2008) "Treelets, not Trees: Phase Structure, not Phrase Structure" Talk presented at BCGL 3 — Trees and Beyond. May 21–23, 2008

(2010a) *Defeating Lexicocentrism*. lingBuzz/001130

(2010b) *Language in Cognition*. Oxford, Blackwell.

Boccara, N. (2002) *Modeling Complex Systems*. Springer. [2nd edition: 2010]

Carnap, R. (1966) *Philosophical Foundations of Physics*. Martin Gardner, ed. Basic Books.

Cherniak, C. (2009) "Brain Wiring Optimization and Non-Genomic Nativism". In M. Piattelli-Palmarini, et. al., eds. *Of Minds and Language* (Oxford, 2009) 108-119.

Chomsky, N. (1995) *The Minimalist Program* Cambridge, MA. MIT press.

(1998) *Minimalist Inquiries. The Framework.* MIT Occasional Papers in Linguistics 15.

(1999) *Derivation by Phase.* MIT Occasional Papers in Linguistics 18.

(2004) *Beyond Explanatory Adequacy*. En Belleti (ed.) "Structures and Beyond", Oxford, OUP.

(2005a) "Three Factors in Language Design". In *Linguistic Inquiry* 36. 1-22.

(2005b) "On Phases". Ms. MIT.

(2007) *Approaching UG from below*. Ms. MIT.

(2009) “Opening Remarks”. In Piatelli Palmarini et. al. *Of Minds and Language*. Oxford, OUP.

De Belder, M. & J. Van Craenenbroeck (2011) “How to Merge a Root”. LingBuzz 001226.

D’ Espósito, M. (2007) “From Cognitive to Neural Models of Working Memory”. Phil. Trans. R. Soc. B 29 May 2007 **vol. 362 no. 1481** 761-772

Di Sciullo, A. M. & Isac, D. (2008). The Asymmetry of Merge. In Biolinguistics Vol 2, No 4.

Embick, D. & R. Noyer (2004) “Distributed Morphology and the Syntax-Morphology Interface”. Draft: 25/10/2004

Epstein, S. & T. D. Seely eds. (2002) *Derivation and Explanation in the Minimalist Program*. Oxford: Blackwell.

Escandell Vidal, M.V. & Leonetti, M. (2000), “Categorías conceptuales y semántica procedimental”. In *Cien años de investigación semántica: de Michél Bréal a la actualidad* Tomo I, Madrid, Ediciones clásicas, 363-378.

Folia, V. et. al. (2011) “Implicit Artificial Syntax Processing: Genes, Preference, and Bounded Recursion”. In *Biolinguistics*, 5.1-2. 105-132.

Frampton, J. & S. Gutmann (2002) “Crash-proof syntax”. In *Derivation and explanation in the minimalist program*, S. Epstein & T.D. Seely (eds.), 90–105. Oxford: Blackwell.

Friederici, A. et. al. (2011) “The Neural Basis of Recursion and Complex Syntactic Hierarchy”. In *Biolinguistics*, 5.1-2. 87-104.

Fukui, N. (1996) “On the Nature of Economy in Language”, *Cognitive Studies* **3**, 51–71.

Gallego, A. (2010) *Phase Theory*. Amsterdam, John Benjamins.

García Mayoraz, J. E. (1989) *Entropía y Lenguajes*. Buenos Aires, Hachette.

Gould, S. J. (2002) *The Structure of Evolutionary Theory*. The Belknap of Harvard University Press, Cambridge, MA & London.

Greene, B. (1999) *The Elegant Universe*. Kopf Doubleday Publishing Group.

Grimshaw, J. (1991) “Extended Projection”. Ms. Brandeis University.

(2000) “Locality and Extended Projections”. In Coopmans, P., Everaert, M. and Grimshaw, J. (eds.). *Lexical specification and lexical insertion*. 115-133. Amsterdam, John Benjamins.

Hale, K. & S. J. Keyser (1993) “On argument structure and the lexical expression of Syntactic Relations”. In *The view from Building 20: Essays in honor of Sylvain Bromberger*, ed. by Kenneth Hale and Samuel Jay Keyser, MIT Press.

(2002) *Prolegomenon to a Theory of Argument Structure*. MIT Press.

Halle, M. & A. Marantz (1993) “Distributed Morphology and the Pieces of Inflection”. In Hale & Keyser (eds.). 111-176.

Harley, H. (2002) "Possession and the double object construction," *Yearbook of Linguistic Variation* 2, pp. 29-68.

Harley, H. & R. Noyer (1999) “State-of-the-Article: Distributed Morphology”. In *GLOT 4.4*. April, 1999. 6-9.

Hauser, M.D., N. Chomsky & W.T. Fitch (2002) “The Faculty of Language: What Is It, Who Has It, and How It Evolve?” In *Science* 298, (5598): 1569–79

Hawking, S. & L. Mlodinov (2005) *A Briefer History of Time*. Bantam Books.

Heisenberg, W. (1999) *Physics and Philosophy*, New York: Prometheus Books

Jackendoff, R. (1997) *The Architecture of the Language Faculty*. Cambridge, Mass. MIT Press.

(2002) *Foundations of Language*. OUP.

Jenkins, L. (2011) “Biolinguistic Investigations: Genetics and Dynamics”. In Di Sciullo, A-M & C. Boeckx (eds.) 2011 *The Biolinguistic Enterprise*. OUP. 126-134.

Kayne, R (1984) *Connectedness and Binary Branching*, Foris, Dordrecht.

(1994) *The Antisymmetry of Syntax*, MIT Press.

Kitahara, H. (1997) *Elementary Operations and Optimal Derivations*. Linguistic Inquiry Monographs 31. MIT Press.

Krivochen, D. (2011a) “An Introduction to Radical Minimalism I: on Merge and Agree (and related issues)”. In *IBERIA Vol 3 n° 2*. Pp. 20-62.

(2011b) “An Introduction to Radical Minimalism II: Internal Merge Beyond Explanatory Adequacy”. Submitted.

(2011c) The Quantum Human Computer Hypothesis and Radical Minimalism: A Brief Introduction to Quantum Linguistics. Published in *International Journal of Linguistic Studies* Vol. 5 n° 4, pp. 87-108.

(2012a) Towards a Geometrical Syntax. Ms. Universität Potsdam. Under Review. http://ling.auf.net/lingbuzz/001444

(2012b) *The Syntax and Semantics of the Nominal Construction*. Potsdam Linguistic Investigations 8. Frankfurt am Main, Peter Lang.

(2012c) “The Quantum Human Computer (QHC) Hypothesis andits Formal Procedures: Their Pragmatic Relevance”. To appear in International Journal of Language Studies (IJLS). Special Issue on "Interface Perspectives on Pragmatics" (ed. Diego Gabriel Krivochen). http://www.academia.edu/2334645/_in_press_The_Quantum_Human_Computer_QHC_Hypothesis_and_its_Formal_Procedures_Their_Pragmatic_Relevance

Ladd, R. et. al. (2008) “Languages and Genes: Reflections on the Nature-Nurture Question”. In *Biolinguistics*, 2.1. 114-126.

Laka, I. (2009). What is there in Universal Grammar? On innate and specific aspects of language. In M. Piattelli-Palmarini, J. Uriagereka, & P. Salaburu (Eds.) *Of Minds and Language: A dialogue with Noam Chomsky in the Basque Country*, pp.329-343, Oxford: Oxford University Press.

Lasnik, H; J. Uriagereka & C. Boeckx (2005) *A Course in Minimalist Syntax*. Oxford, Blackwell.

Mateu Fontanals, J. (2000a) Why Can’t We Wipe the Slate Clean? A Lexical-Syntactic Approach to Resultative Constructions. Universitat Autònoma de Barcelona. Departament de Filologia Catalana

(2000b) *Universals of Semantic Construal for Lexical Syntactic Relations*. Universitat Autónoma de Barcelona. Departament de Filología Catalana.

Mendívil Giró, J. L. (2004) *Biolingüística: qué es, para qué sirve y cómo reconocerla*. Ms. Universidad de Zaragoza.

Matushansky, O. (2006) “Head Movement in Linguistic Theory”. Linguistic Inquiry 37, 1. 69-109.

Panagiotidis, P. (2009) *Four Questions on Categorization of Roots.* Ms. University of Cyprus.

(2010) *Functional Heads, Agree and Labels*. Ms. University of Cyprus.

Pesetsky, D. & E. Torrego (2007) Probes, Goals and Syntactic Categories. In *Proceedings of the 7th annual Tokyo Conference on Psycholinguistics* (Keio University, Japan)

Philosophy of Quantum Mechanics MIT course Spring 2005. Available at http://ocw.mit.edu/courses/linguistics-and-philosophy/ Downloaded on 16/7/2008

Pylyshyn, Z. W. (2007) *Things and places: How the mind connects with the perceptual world* (2004 Jean Nicod Lectures). Cambridge, MA: MIT Press.

Putnam, M. (Ed.) (2010) *Exploring Crash-Proof Grammars*. LFAB Series, edited by Pierre Pica & Kleanthes K. Grohmann. Amsterdam, John Benjamins.

Roberts, M. (1998) Ultrametric Distance in Syntax. http://arXiv.org/abs/cs.CL/9810012

Shannon, C. (1948) *A Mathematical Theory of Communication*. Bell System, Technical Journal.

Sperber, D. & D. Wilson (1986) “Sobre la definición de Relevancia”. In L. Ml. Valdez Villanueva, *En Búsqueda del Significado*. Madrid, Tecnos.

(1986b/1995) *Relevance: Communication and Cognition*. Oxford. Blackwell.

Talmy, L. (1983) “How language structures space”. In Herbert L. Pick, Jr. & Linda P. Acredolo (eds.) *Spatial orientation: Theory, research, and application*. New York: Plenum Press. 225-282.

(2000) *Toward a cognitive semantics.* Cambridge, MA: Massachusetts Institute of Technology.

Tegmark, M. (2007) *The Mathematical Universe*. In “Foundations of Physics”, November 2007, 116.

Udagonkar, J. (2001) “Entropy in Biology”. In *Resonance,* September, 2001, 61-66.

Uriagereka, J. (1998) *Pies y Cabeza*. Madrid, Visor.

(1999) “Multiple Spell-Out”. En N. Hornstein & S. Epstein (eds.), *Working Minimalism*, Cambdridge (MA), MIT Press, 251-282.

(2000) “On the Emptiness of the “Design” Polemics”. *Natural Language & Linguistic Theory* **18:** 863–871.

(2002) *Derivations: Exploring the Dynamics of Syntax*. London, Routledge. (Chap. 15: “Warps: Some Thoughts on Categorization”)

(2011) *Spell Out and the Minimalist Program*. OUP.

Watson, J.D. (with A. Berry) (2003) “*DNA. The Secret of Life*” Random.

Watumull, J. (2012a) A Turing Program for Linguistic Theory. In *Biolinguistics,* 6.2. 222-245.

(2012b) The Computability and Computational Complexity of Generativity. COPiL 11. 311-329.

Wilson & Sperber (2003) “Relevance Theory”. In (Horn, L.R. & Ward, G. eds.) *The Handbook of Pragmatics*. Oxford: Blackwell, 607-632.